FLUORINATED ALCOHOL INDUCED SUPRAMOLECULAR BIPBASIC SYSTEMS IN PROTEOMICS AND LIPIDOMICS

By

MOHAMMADMEHDI AZIZI

DISSERTATION

Submitted in partial fulfillment of requirements

For the degree of Doctor of Philosophy at

The University of Texas at Arlington

May, 2021

Arlington, TX

Supervising committee:

Dr. Morteza G. Khaledi, Supervising Professor

Dr. Daniel W. Armstrong

Dr. Krishnan Rajeshwar

Dr. Jongyun Heo

This thesis is dedicated to my parents.

For their endless love, support, and encouragement

ACKNOWLEDGMENTS

TABLE OF CONTENTS

TABLE OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS

ABC   ammonium bicarbonate

ACN   acetonitrile

BPS   biphasic systems

CAN   acetonitrile

Cer   ceramides

CTAB   cetyl Trimethyl Ammonium Bromide

DCM   dichloromethane

DMMAPS   dimethylmyristylammoniopropane sulfonate

DSC   sodium deoxycholate

DTAB   dodecyltrimethylammonium bromide

DTT   dithiothreitol

FAHFA   fatty acid ester of hydroxyl fatty acid

FAiC   fluoroalcohol-Induced Coacervate

FAiS-BP   fluoroalcohol-induced supramolecular biphasic

FAiC-BPS   fluoroalcohol-induced coacervate biphasic systems

FASP   filter assisted sample preparation

FDR   false discovery rate

GRAVY   grand average of hydropathy

HCD   higher-energy collisional dissociation

HexCer   hexosylceramides

HFMIP	hexafluoro-2-methyl-isopropanol
HFIP	hexafluoroisopropanol
IAA	iodoacetamide
ID	inner diameter
IMP	integral membrane proteins
IPA	isopropanol
LCMS	liquid chromatography mass spectrometry
LFQ	label-free quantification
MGDG	monogalactosyldiacylglycerols
MRP	mitochondrial ribosome proteins
MTBE	methyl-tert-butyl ether
NLC	natural lipid coacervation
PA	phosphatidic acid
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PI	phosphatidylinositol
PS	phosphatidylserine
QUATS	Quaternary ammonium salts
S-BPS	supramolecular biphasic system
SC	sodium cholate
SDS	sodium dodecyl sulfate
SM	sphingomyelins
SMILES	simplified molecular line entry specification
TBAB	tetrabutylammonium bromide
TFE	trifluoroethanol

TFE   trifluoroethanol

TEAB   tetraethylammonium bromide

TOAB   tetraoctylammonium bromide

T.Pentyl.AB   tetrapentylammonium bromide

T.Propyl.AB   tetrapropylammonium bromide

ABSTRACT

FLUORINATED ALCOHOL INDUCED SUPRAMOLECULAR BIPBASIC SYSTEMS IN

PROTEOMICS AND LIPIDOMICS

Mohammadmehdi Azizi, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Morteza G. Khaledi

The Fluoroalcohol induced Supramolecular Biphasic (FA*i*S-BP) systems are composed of a supramolecular phase which is a highly concentrated mixture of a fluoroalcohol and an amphiphile, and a separate aqueous-rich phase. The first part of this study presents new classes of FA*i*S-BP systems and studies their application in identification and coverage enhancement of low-abundance and/or hydrophobic proteins, such as integral membrane proteins. By taking the advantage of FA*i*S-BP systems in proteomics, we can concomitantly extract, fractionate, and enrich proteins; therefore, an additional separation dimension and a simple step of enrichment would be added prior to the LC-MS.

Yeast proteins (Saccharomyces cerevisiae yeast) were fractionated between the two phases of FA*i*S-BP systems and subjected to tryptic digestion and LC-MS/MS analysis. The results of FA*i*S-BP systems were compared to commonly used sample preparation techniques in proteomic as controls. In controls, common solubilizing reagents for proteins, such as urea, sodium dodecyl

sulfate (SDS), and sodium deoxycholate (SDC) were used. All FA*i*S-BP systems of this study showed protein coverage enhancement compared to the controls, interestingly, majority of this improvement was observed at low abundance ranges. Identification of low-abundance proteins (less than 2000 molecule/cell) improved by up to 150%, which was equivalent to detection of additional 177 proteins as compared to the urea control. Interestingly, larger improvements were observed as the protein abundance decreased. Above a certain level of abundance (~5000 molecules/cell), there is little or no difference between protein coverage using the conventional methods and FA*i*S-BP systems. This suggests that the FA*i*S-BP systems are particularly advantageous for detecting low abundance proteins.

In FA*i*S-BP systems, each phase has selectivity towards specific protein groups, this selectivity is generally based on pI and hydrophobicity of proteins. This selectivity can be altered by changing the major components of FA*i*S-BP systems, such as amphiphiles and fluoroalcohols, or by changing the net charge of proteins through modulating the pH that would affect electrostatic interactions in the system. Using quaternary ammonium salts (QUATS) with short alkyl chain lengths, such as tetraethylammonium bromide (TEAB) showed greater electrostatic interactions than other amphiphiles. At moderate basic pH values (pH=8.5), using QUATS with one long hydrophobic alkyl chain, such as cetyltrimethylammonium bromide (CTAB), resulted in considerable sequence coverage enhancement for alpha-helical transmembrane peptides.

The terminology of fractionation of complex mixtures prior to LC-MS analysis was also applied to lipidomics studies. We introduced a novel multiphase system composed of a top aqueous phase, a middle phase which is rich in dichloromethane (DCM), and a bottom phase which is rich in HFIP. Lipids were extracted and enriched in different phases based on their physicochemical properties. Combination of this simple fractionation step with LC-MS-based lipidomics resulted in significant lipidome coverage improvements of about 150%. Generally, lipids fractionate and

enrich in different phases based on their polar headgroups and hydrophobic chains. The most hydrophobic lipids were enriched in the DCM-rich phase, relatively less hydrophobic lipids were enriched in the HFIP-rich phase, and the most hydrophilic lipids were enriched in the aqueous phase. Phosphatidylinositols (PIs), with water-soluble myo-inositol headgroups, were mostly extracted by the HFIP-rich phase. On the other hand, phosphatidylglycerols (PGs) were enriched in the DCM-rich phase. Lyso-phospholipids and Fatty acid ester of hydroxyl fatty acids (FAHFAs) are other classes of lipids that were enriched in the HFIP-rich phase, while ceramides (Cer) and hexosylceramides (HexCer) were enriched in the DCM-rich phase.

CHAPTER 1

INTRODUCTION

In proteomics studies, majority of missing proteins from the list based on gene-coding are very hydrophobic and/or in low abundance.[1] Generally, detection and identification of these types of proteins are challenging.[2,3] Common solubilizing reagents, such as urea and detergents, do not effectively solubilize hydrophobic proteins (*e.g.*, Integral Membrane Proteins, IMP).[4] In bottom-up proteomics, ineffective solubilization of proteins simply leads to low concentration of their tryptic peptides in the digested sample. Sodium dodecyl sulphate (SDS) is another common solubilizing reagent that offers more effective solubilizing power for hydrophobic proteins, however, the drawback associated with SDS is its strong interaction with proteins that results in difficulties of its removal required for effective digestion.[5] Small percentages of SDS, as low as 0.01%, inhibits activity of trypsin significantly.[5] Obviously, incomplete digestion suppresses identification of proteins. In addition, low abundance proteins are underrepresented under the shadow of higher abundance proteins.

Similarly, lipid extracts from biological samples are very complex mixtures and their characterization pose significant challenges. Structural diversity of lipids is not limited to different lipid classes; also, different hydrophobic backbones such as saturated, monounsaturated, and polyunsaturated acyl groups, with different carbon numbers at both sn-1 and sn-2 positions also

affect diversity of lipids.[6,7] In LC-MS-based lipidomics, low abundance lipids are usually underrepresented under the shadow of high abundance lipids. Additionally, coelution of different lipids due the complexity of mixtures would result in losing lipidome coverage.

Sample preparation and separation techniques play crucial roles in the outcome of protein and lipid coverage in proteomics and lipidomics analysis. Recently, this lab has discovered novel approaches to enhance proteome coverage by utilization of Fluoroalcohol induced Supramolecular Biphasic (FA*i*S-BP) systems.[8–10] Supramolecular solvents refer to nanostructured liquids which are produced in colloidal solutions of amphiphilic compounds through the spontaneous and sequential mechanism of self-assembly and coacervation.[11] These supramolecular structures coacervate and form a separate phase from the aqueous media due to due to differences in polarity and density. The amphiphile-rich phase is named as the "coacervate phase", and the other phase as the "aqueous phase".

The fluoroalcohol-induced coacervation of amphiphiles in aqueous media was originally discovered and characterized in this laboratory.[12] Recently, we have investigated the usefulness of FA*i*S-BP systems, mediated by polar fluoroalcohols/fluoroacids, in sample preparation for proteomics.[8] A main category of these FA*i*S-BP systems is Fluoroalcohol-induced Coacervate Biphasic Systems (FA*i*C-BPS) that were first investigated in bottom-up proteomics applications by this laboratory.[8] Addition of fluoroalcohols to the aqueous solutions of different classes of amphiphiles can induce variety of biphasic systems with unique characteristics over a broad range of concentrations.[13] There is a great difference between the conventional coacervates in purely aqueous media and FA*i*C-BPS; formation of conventional coacervates depends strongly on the molecular structure of amphiphiles and occurs over a narrow range of concentrations. Therefore, applications of conventional coacervate systems in sample preparation are limited. In contrast, using FA*i*C-BPS in proteomics applications would allow fractionation of complex protein

mixtures between the aqueous and coacervate phases, effective extraction and solubilization of hydrophobic proteins into the coacervate phase, and enrichment of specific protein groups in one of the phases.

FAiC-BPS have unique properties that cannot be found in other coacervation systems. First, the coacervate phase of FAiC-BPS contains a large concentration of fluoroalcohols and amphiphiles, therefore the coacervate phases in these systems would offer a great solubilization power for hydrophobic compounds, such as IMP.[10] Second, the volume of coacervate phases is only a small fraction of the total volume of system, therefore proteins that are extracted into the coacervate phase are simultaneously enriched. This characteristic is quite helpful in facilitating detection of low-abundance proteins. Third, we can take the advantage of FAiC-BPS for fractionation of complex protein mixtures; more hydrophobic proteins, such as IMP, will be extracted into the coacervate phase while more hydrophilic proteins will remain in the aqueous phase. This feature is helpful in simply reducing the complexity of proteomes in early steps of sample treatment. Finally, FAiC-BPS offer significant selectivity ranges that are close to those in Solid Phase Extraction (SPE). The issue of sample loss is routine in SPE due to protein adsorption on solid phases, but FAiC-BPS is a kind of liquid-liquid extraction methods that offers large sample capacity with no sample loss.

In 2017, this laboratory published the first report on the application of FAiC-BPS in bottom-up proteomics analysis of yeast proteins.[8] FAiC-BPS was induced by addition of hexafluoroisopropanol (HFIP) to the aqueous solution three different types of surfactant, including SDS as an anionic surfactant, cetyltrimethylammonium bromide (CTAB) as a cationic surfactant, and - (N, N-Dimethylmyristyl ammonio) propane sulfate (DMMAPS) as a zwitterionic surfactant. This report also investigated a combination of equimolar oppositely charged CTAB and SDS to form complex coacervation. Afterwards, we investigated coacervate formation of

tetrabutylammonium bromide (TBAB), induced by HFIP or trifluoroethanol (TFE).[9] Opposed to surfactants with long alkyl chains, Quaternary Ammonium Salts (QUATS) with small alkyl chains, such as TBAB, do not form micelles by self-aggregation in aqueous media. Therefore, long chain surfactants denature proteins and the interact with them via hydrophobic effect, on the other hand, the electrostatic interaction is more notable in the systems that contain QUATS with short-length chains such as TBAB. This study focused on fractionation patterns using model proteins, and results showed that the FA*i*C-BPS composed of TBAB and HFIP can be used to fractionate proteins mixtures with more hydrophobic and acidic proteins showing greater affinity for the TBAB coacervate, while the hydrophilic and basic proteins remaining in the aqueous phase.

In this study, we further investigated the capabilities of the TBAB-HFIP system in bottom-up proteomics of yeast (Saccharomyces cerevisiae) to enhance proteins' coverage, especially for membrane and low-abundance proteins, and provide additional evidence about the important role of electrostatic effects imparted by TBAB phases. We further discuss different FA*i*C-BPS, induced by addition of HFIP to the aqueous solutions of different QUATS that have the same quaternary ammonium head group but different hydrophobic alkyl chain lengths. The alkyl chains have hydrophobic interaction with the hydrophobic parts of the proteins, while there is electrostatic interaction between the charged groups of proteins and the positively charged quaternary ammonium headgroup (independent of the solution pH). Protein fractionation patterns change in different FA*i*C-BPS because of the different balances of hydrophobic effects and electrostatic interactions. Selectivity of coacervate phases of FA*i*C-BPS towards proteins can be modulated by judicious alteration of amphiphiles (cationic, anionic, or zwitterionic), such as changing the polar headgroups or altering chain length of hydrophobic tails of amphiphiles. In addition, the effect of protein charge on fractionation pattern in the FA*i*C-BPS was studied by modulating the pH of system. Selectivity of coacervate phases of FA*i*C-BPS towards proteins can also be controlled by

judicious alteration of amphiphiles (cationic, anionic, or zwitterionic), such as changing the polar headgroups or altering chain length of hydrophobic tails of amphiphiles. Further, we investigated the effect of different amphiphiles, such as dodecyltrimethylammonium bromide (DTAB) and CTAB with one long hydrophobic chain, and QUATS with different chin lengths. Additionally, changing the fluoroalcohol would result in having different polarities and hydrogen bonding properties of coacervator that can change the selectivity of phases in FA*i*C-BPS. The next part of this study investigates the effected different coacervators, including HFIP, TFE, and Hexafluoro-2-methylisopropanol (HFMIP) on fractionation patterns of proteins between the two phases. We also compared these FA*i*C-BPS in terms of transmembrane α-helical sequence coverage, and fractionation pattern of different protein groups between the two phases.

The same terminology can be used to develop lipidomics techniques. In lipidomic studies, prefractionation of lipids in multi-phase systems would results in having samples with less complexity and concentrating specific lipid classes with similar physicochemical properties in each phase. Combining this novel technique with LC-MS would result in having a simple additional dimension prior to the LC. Further, we introduce a novel organic three-phase system consists of dichloromethane (DCM), HFIP, and water. The three-phases system consists of an aqueous phase on the top, a the most hydrophobic DCM-rich phase in the middle, and HFIP-rich phase at the bottom with slighter hydrophobicity. This three-phase system can be used as a complementary method for lipidomics studies; lipid can be fractionated between the phases based on their hydrophobicity. Simultaneously, lipids can fractionate base on the lipid classes due to the different molecular interactions between the polar headgroups of lipids and components of phases. This simple fractionation step, combined with LCMS lipidomics approaches, will result in notable improvement in lipid detection.

REFERENCES

(1)   Zhang, Y.; Lin, Z.; Tan, Y.; Bu, F.; Hao, P.; Zhang, K.; Yang, H.; Liu, S.; Ren, Y. Exploration of Missing Proteins by a Combination Approach to Enrich the Low-Abundance Hydrophobic Proteins from Four Cancer Cell Lines. *J. Proteome Res.* **2020**, *19* (1), 401–408. https://doi.org/10.1021/acs.jproteome.9b00590.

(2)   Distler, A. M.; Kerner, J.; Peterman, S. M.; Hoppel, C. L. A Targeted Proteomic Approach for the Analysis of Rat Liver Mitochondrial Outer Membrane Proteins with Extensive Sequence Coverage. *Anal. Biochem.* **2006**, *356* (1), 18–29. https://doi.org/10.1016/j.ab.2006.03.053.

(3)   Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.; Chen, L.; Zhang, Y.; Wen, B.; He, F.; Xu, P. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *J. Proteome Res.* **2016**, *15* (11), 3988–3997. https://doi.org/10.1021/acs.jproteome.6b00390.

(4)   Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography−Tandem Mass Spectrometry. *J. Proteome Res.* **2002**, *1* (4), 351–360. https://doi.org/10.1021/pr0255248.

(5)   Moore, S. M.; Hess, S. M.; Jorgenson, J. W. Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics. *J. Proteome Res.* **2016**, *15* (4), 1243–1252. https://doi.org/10.1021/acs.jproteome.5b01122.

(6)   Shevchenko, A.; Simons, K. Lipidomics: Coming to Grips with Lipid Diversity. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (8), 593–598. https://doi.org/10.1038/nrm2934.

(7)   Shindou, H.; Hishikawa, D.; Harayama, T.; Eto, M.; Shimizu, T. Generation of Membrane Diversity by Lysophospholipid Acyltransferases. *J. Biochem. (Tokyo)* **2013**, *154* (1), 21–28. https://doi.org/10.1093/jb/mvt048.

(8)   McCord, J. P.; Muddiman, D. C.; Khaledi, M. G. Perfluorinated Alcohol Induced Coacervates as Extraction Media for Proteomic Analysis. *J. Chromatogr. A* **2017**, *1523*, 293–299. https://doi.org/10.1016/j.chroma.2017.06.025.

(9)   Koolivand, A.; Clayton, S.; Rion, H.; Oloumi, A.; O'Brien, A.; Khaledi, M. G. Fluoroalcohol – Induced Coacervates for Selective Enrichment and Extraction of Hydrophobic Proteins. *J. Chromatogr. B* **2018**, *1083*, 180–188. https://doi.org/10.1016/j.jchromb.2018.03.004.

(10)   Koolivand, A.; Azizi, M.; O'Brien, A.; Khaledi, M. G. Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies. *J. Proteome Res.* **2019**, *18* (4), 1595–1606. https://doi.org/10.1021/acs.jproteome.8b00857.

(11)   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.

(12)   Khaledi, M. G.; Jenkins, S. I.; Liang, S. Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles. *Langmuir* **2013**, *29* (8), 2458–2464. https://doi.org/10.1021/la303035h.

(13)   Jenkins, S. I.; Collins, C. M.; Khaledi, M. G. Perfluorinated Alcohols Induce Complex Coacervation in Mixed Surfactants. *Langmuir* **2016**, *32* (10), 2321–2330. https://doi.org/10.1021/acs.langmuir.5b04701.

.

CHAPTER 2

IMPROVING IDENTIFICATION OF LOW ABNDANCE AND

HYDROPHOBIC PROTEINS USING FLUOROALCOHOL MEDIATED

SUPRAMOLECULAR BIPHASIC SYSTEMS WITH QUATERNARY

AMMONIUM SALTS

Used with permission from Mohammadmehdi Azizi, Sajad Tasharofi, Amir Koolivand, Armin Oloumi, Halie Rion, Morteza G. Khaledi

ABSTRACT

In this study, a newly discovered Supramolecular Biphasic System (S-BPS) was used in bottom-up proteomics of Saccharomyces cerevisiae yeast. The results were compared to controls which use two common methods in proteomics; in the first control 8 M urea was used as solubilizing reagents, and in the second control SDS was used. With the S-BPS, we identified 3043 proteins as compared to 2653 proteins using the control system. Interestingly, of the additional 390 proteins characterized by the S-BPS, 300 proteins were low abundance (less than 4000 molecules/cell). Remarkably, the identification of proteins at very low abundance (less than 2000 molecule/cell) was improved by 106%. Gene Ontology analysis was conducted to find fractionation pattern of proteins in our two-phase system, and in nearly every gene ontology category, the S-BPS provided greater coverage than the control experiment, i.e., coverage for

8

integral membrane proteins and mitochondrial ribosome proteins are improved by 18% and 58%, respectively. The improvements in proteins coverage for lower abundant and membrane proteins can be attributed to the strong solubilizing power of the amphiphile-rich phase of this S-BPS and its capability for concomitant extraction, fractionation, and enrichment of the complex proteomics samples. Each phase has selectivity towards specific yeast protein groups, this selectivity is generally based on pI and hydrophobicity of proteins. Proteins with greater hydrophobicity and more acidic proteins exhibit greater affinities for the amphiphile-rich phase due to hydrophobic effect and electrostatic interactions.

KEYWORDS: Coacervate, Fractionation, Fluoroalcohols, Hydrophobic Proteins, Low Abundance Proteins, Proteomics.

2.1 INTRODUCTION

Mass spectrometry data has confirmed that there are a number of missing proteins on the list of gene-encoded proteins.[1] These missing proteins tend to have low abundance and/or high hydrophobicity, such as membrane proteins. In proteomic studies using mass spectrometry, low abundance proteins are underrepresented under the shadow of higher abundance proteins; while membrane and hydrophobic proteins are underrepresented due to challenges in their solubilization, enzymatic digestion, and their tendency to aggregate in aqueous media. In this study, we use a Supramolecular Biphasic System (S-BPS), which is recently discovered in this lab, to address these challenges in proteomics studies.

Self-assembly is known as organization of molecules into a stable arrangement by non-covalent forces.[2] Supramolecular solvents is a term to refer to water immiscible liquids consisting of supramolecular aggregates that are generated from amphiphiles through a sequential, self-

assembly process.[3,4] First, three-dimensional aggregates coacervate, and then these coacervates produce a second water-immiscible phase consisting of large supramolecular aggregates dispersed in a continuous phase, generally water.[4] This is why this amphiphile-rich phase is named as the "coacervate phase", and the other phase as the "aqueous phase".

In previous studies, we reported that incorporation of Fluoroalcohol-Induced Coacervate (FA*i*C) Biphasic Systems (BPS), in bottom-up proteomics workflow can lead to higher coverage, especially for membrane proteins and proteins in low-abundance, due to their capabilities to concomitantly extract, fractionate, and enrich proteins in separate phases.[5]

The fluoroalcohol-induced coacervation of amphiphiles in aqueous media was originally discovered and characterized in this laboratory.[6–8] Coacervates are a type of self-assembly of amphiphile molecules in aqueous media that form a separate phase from the bulk aqueous phase. As compared to other kinds of self-assemblies such as micelles, coacervation of surfactants occurs rarely. We first reported that addition of fluoroalcohols such as hexafluoroisopropanol (HFIP) and trifluoroethanol (TFE), to aqueous solutions of amphiphiles such as synthetic surfactants, bile salts, phospholipids, form an unprecedented variety of coacervates and phase separate in aqueous media over a wide range of concentrations.

The coacervate phases have strong solubilizing power, especially for very hydrophobic compounds due to their compositions of high effective concentrations of the amphiphile and the fluoroalcohol molecules. The volumes of the coacervate phases are only a small fraction of the initial total solution volume. Thus, as compounds are extracted into the FA*i*C phase from the top aqueous phase, they are simultaneously enriched. The capabilities of FA*i*C-BPS can be advantageous in proteomics analysis where a complex mixture of proteins with a wide dynamic

range can be solubilized and extracted from biological samples, and concomitantly fractionated and enriched in the two separate aqueous and coacervate phases.

In the first application of FA*i*C-BPS in bottom-up proteomics of yeast samples, we examined HFIP-induced simple coacervates using three types of surfactants with long alkyl chains; anionic sodium dodecyl sulfate (SDS), cationic cetyltrimethylammonium bromide (CTAB), and zwitterionic dimethylmyristylammoniopropane sulfonate (DMMAPS). The FA*i*C-BPS provided greater coverage than the control experiment (no phase separation) that utilized 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 like endoplasmic reticulum and cellular vacuoles.

We also reported a HFIP-induced coacervation of phospholipids in the cell membrane lipid bilayers at high lipids concentrations (e.g. more than 10 mg/mL).[9] Membrane proteins are naturally extracted and enriched in the lipid coacervate phase through the coacervation process, while the hydrophilic proteins reside primarily in the aqueous phase of the two-phase system. This method is detergent-free and obviates the detergent removals steps. Thanks to the capability of such a simple system in fractionation of proteins based on their hydrophobicity, the coacervation of lipid bilayer in natural cell membranes of yeast increased identification of total proteins and membrane proteins by 8% and 13% respectively, as compared to the control urea experiment with no phase separation.[9]

We have recently reported that fluoroalcohols (HFIP and TFE) can also induce coacervation and phase separation in the aqueous solutions of amphiphiles with shorter alkyl chains such as tetrabutylammonium bromide (TBAB).[10] The study of extraction patterns using model proteins showed that the FA*i*C-BPS composed of TBAB can be used to fractionate proteins mixtures with

more hydrophobic and acidic proteins showing greater affinity for the TBAB coacervate, while the hydrophilic and basic proteins remaining in the aqueous phase. The protein fractionation pattern in TBAB-HFIP system seems to stem from a balance of hydrophobic and electrostatic effects than is different from other FA*i*C systems made of long chain surfactants, including C_nTAB, that has the same positively charged quaternary ammonium head group. In this study, we further investigated the capabilities of the TBAB-HFIP system in bottom-up proteomics of yeast to enhance proteins coverage, especially for membrane and low-abundance proteins and provide additional evidence about the important role of electrostatic effects imparted by TBAB phases.

2.2 MATERIALS AND METHODS

2.2.1 PROTEIN SAMPLE

Saccharomyces cerevisiae (strains BY4741, Ward's Science®) was grown and lysed according to the our previous publication;[9] however, the lysis buffer was modified by adding 10 µL pepstatin at the concentration of 1 mg/mL to every 10 mL lysis buffer without adding sodium chloride to the lysis buffer. The lysis solution was made by dissolving 1 Mini Tablet of EDTA free Pierce® Protease and Phosphatase Inhibitor in 10 mL of autoclaved deionized water and adding 10µL of pepstatin solution (Roche® Diagnostics GmbH, dissolved in methanol, 1 mg/mL). We have previously reported that addition of NaCl and NaOH to the coacervation system instigates changes in the phase behavior [10] and the salt-free TBAB/HFIP system exhibited greater selectively toward the more hydrophobic proteins. Addition of NaCl could result in salting-out of more hydrophilic proteins from the aqueous phase to the coacervate phase. The presence of NaOH would change proteins charges and subsequently electrostatic effects. We chose a salt-free coacervate system for extraction of proteins. Finally, the protein concentration of the cell lysate was measured by protein assay (Coomassie Protein Assay Kit, Thermo Scientific®). The cell lysate was kept at -80 °C.

2.2.2 PRECONDITIONING FASP FILTERS

EMD Millipore® Amicon® Ultra-0.5 Centrifugal Filter Units (10K) was used for the filter aided sample preparation (FASP) method. The ultrafiltration membrane in Amicon® Ultra-0.5 devices contain trace amounts of glycerin which may interfere with analysis. To avoid any unfavorable interferences, filters were pre-rinsed with a 500 µL of UTT buffer (5 M urea, 2 M thiourea, 100 mM Tris buffer pH=8.5), and centrifuged at 14,000×g for 40 minutes, a thin layer of the solution must remain in the filter. Once the filters are preconditioned, drying should be avoided and filters must be kept wet.

2.2.3 COACERVATION AND TRYPTIC DIGESTION

Coacervation can be induced over a wide range of TBAB and HFIP concentrations [10]. In this case, the coacervate was formed at 50 mM TBAB and 8% HFIP. In order to form the aqueous/coacervate two-phase system, HFIP was added to the aqueous solution of TBAB, which contained 400 µg protein from the cell lysate, then the sample was centrifuged at 10,000 g for 15 minutes to separate the two phases, the surfactant-rich bottom phase is called the coacervate phase and the water-rich top phase is called the aqueous phase. The aqueous phase was separated using a micro-syringe and placed in a new vial, and protein concentration in each phase was measured by Bradford method [11] with Coomassie Protein Assay Kit (Thermo Scientific), then subjected to the FASP method for TBAB removal and tryptic digestion on the filter [12].

HFIP in the sample interferes with Bradford assay, solubilizes the FASP filter and causes coacervation in the separated aqueous phase; therefore, HFIP was evaporated from the samples prior to FASP workflow. The coacervate phase was dried by exposure to nitrogen gas flow, and the aqueous phase was placed in vacuum centrifuge until half of the volume evaporated. The dried coacervate phase was dissolved in 500 µL solution of 70% IPA and 2M thiourea, then sonicated at room temperature for 5 minutes to form a homogeneous solution. It is important to note that

sonication at lower temperatures, such as 4°C, must be avoided because it causes precipitation. The coacervate solution was loaded to the preconditioned FASP filter and centrifuged at 14,000g for 40 minutes, centrifuged for a longer period if the volume above the filter was more than 20 µL. Then 200 µL IPA was loaded to the filter, thoroughly mixed with the sample by using a micropipette, and centrifuged at 14,000×g for 40 minutes, centrifuged for a longer period if the volume above the filter is more than 20 µL. The same procedure was repeated after addition of another 200 µL of 70% IPA, followed by addition of 200 µL UTT solution.

The aqueous phase was loaded to a preconditioned filter after HFIP evaporation, then centrifuged at 14,000×g for 40 minutes, centrifuged for a longer period if the volume above the filter is more than 20 µL.

20 µL of 250 mM dithiothreitol (DTT) solution was added to each sample and the volume was adjusted to 200 µL by addition of UTT solution, the final concentration of DTT was 25 mM in the solution. Samples were subjected to DTT reduction at 37 ℃ for 30 minutes, then cooled to room temperature and centrifuged at 14,000×g for 40 minutes. 54 mM iodoacetamide (IAA) solution was prepared by dissolving IAA in UTT buffer. 250 µL of 54 mM IAA solution was added to the samples to make the final concentration of 50 mM IAA. Samples were vortexed for 30 sec at 600 rmp, then incubated in the dark for 30 minutes for alkylation. The samples were centrifuged at 14,000×g for 40 minutes. 200 µL of UTT solution was added to each sample, then the samples were centrifuged at 14,000×g for 40 minutes. 200 µL of 50 mM ammonium bicarbonate (ABC) buffer was added to each sample, then the samples were centrifuged at 14,000×g for 40 minutes. The last step was repeated one more time. 150 µL of 100 mM ABC buffer was added to each sample to bring the pH to 7.8, then trypsin was added with the mass ratio of 1 to 25 (trypsin: protein). After addition of trypsin, pH was checked to be between 7 and 8. Filters were sealed with

parafilm, vortexed (600 rpm) for one minute, then incubated in wet chamber at 37 °C for 16 hours for protein digestion. Digested samples on the filters were transferred to new collection tubes and centrifuged at 14,000 g for 40 min. 200 µL 0.5 M NaCl was added to filters and centrifuged at 14,000×g for 40 minutes in the collection tube. Finally, 200 µL of 50 mM NaCl solution was added to the filters, filters were inverted and centrifuged at 1000×g for 2 minutes in the collection tubes to collect all the remaining peptides. The whole workflow is schematically summarized in Figure 2-1. The collected peptides were acidified with TFA to bring the pH down to 1-2. Acidified peptides were desalted based on the published protocol [13], then dissolved in 0.2% formic acid for LC-MS/MS analysis. More details about the two-phase induction and tryptic digestion are presented in the Appendix 2-1.
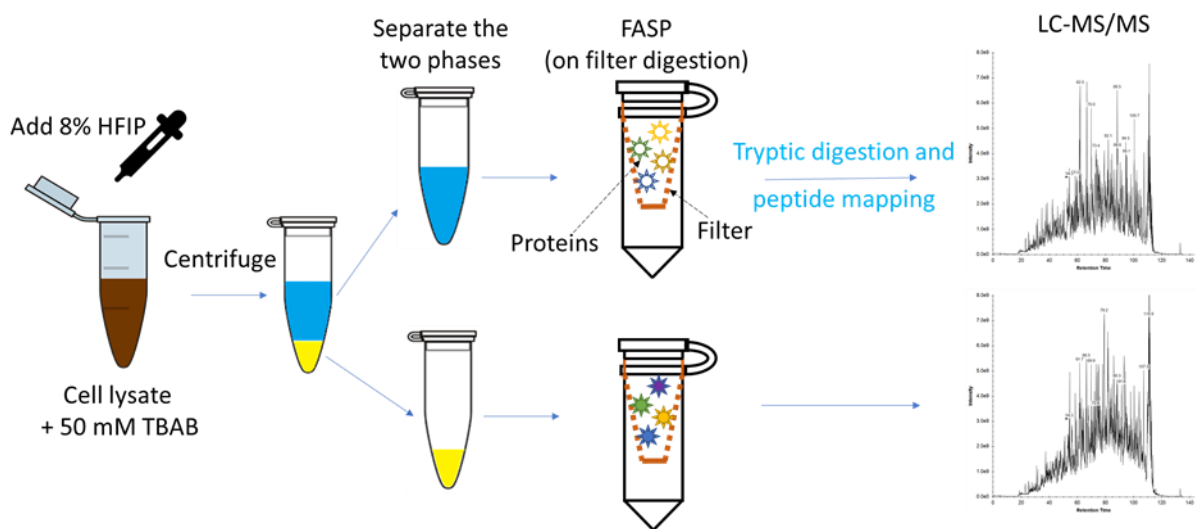
Figure 2-1. Schematic of sample preparation, including coacervate formation, digestion of the two phases using FASP method and LC-MS/MS analysis.

2.2.4 LC-MS/MS ANALYSIS

Digested and desalted proteins were subjected to LC-MS/MS (Ultimate 3000 RSLC-Nano liquid chromatography systems, Dionex; coupled with Orbitrap Fusion Lumos MS®, Thermo

Electron) analysis. 2 µL of each sample was injected into a C18 column (EasySpray column, i.d.: 75 µm, length: 75 cm, particle size: 3 µm, Thermo®) and eluted with the following gradient, 0-90 min: 0-28% B, 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. The mass spectrometer operated in positive mode at the following conditions, source voltage: 2.2 kV, ion transfer tube temperature of 275 °C, resolutions: 120,000; number of MS/MS spectra event: up to 10 for each full spectrum, fragmentation: higher-energy collisional dissociation (HCD) for ions with charges of 2-7, and dynamic exclusion: 25 s after an ion was selected for fragmentation.

2.2.5 DATA PROCESSING

MaxQuant (version 1.1.0.1) was used to process the raw data by using the yeast database from http://www.uniprot.org. The following options were used for protein identification, first search peptide tolerance: 20 ppm, main search peptide tolerance: 4.5 ppm, ITMS MS/MS match tolerance: 0.5 Da, enzyme: trypsin, missed cleavage: 2, fixed modification: carbamidomethyl (C), variable modification: oxidation (M) and acetylation (Protein N-term). False discovery rate (FDR) thresholds were specified at 1% for protein, peptide and modification site. The minimum unique peptide was set to 1. In Label-free quantification (LFQ), iBAQ was selected for quantification in LC-MS/MS analysis. Data analysis were conducted using Perseus proteomics software.[14] Unique peptides and iBAQ were selected for the main analysis, and matrices were filtered to remove the proteins with potential contaminants, identified only by cite, or identified by reverse peptides. The proteins which were identified at least in two replicates out of three were counted as the biological reproducible ones for further data analysis. The list of identified proteins in aqueous and coacervate phases were compared using the http://www.geneontology.org for GO annotation analysis, including biological processes, molecular functions, and cellular components.

GRAVY (grand average of hydropathy) value is a parameter that shows hydrophobicity of proteins and larger GRAVY values show more hydrophobicity. GRAVY values were calculated from yeastmine.yeastgenome.org [15], also available at "gravy calculator" (http://www.gravy-calculator.de)[16] which theoretically calculates GRAVY values based on protein sequence. The isoelectric points of proteins were calculated by yeastmine.yeastgenome.org [15], also available at http://isoelectric.org,[17] which contain a database of pre-computed isoelectric points for proteins from different model organisms. Each isoelectric point value is the average of isoelectric points calculated based on various prediction methods (18 algorithms implemented [18]).

For the proteins which were commonly identified in both coacervate and aqueous phases, partition coefficient was calculated. Partition coefficient (K) for each protein was calculated based on the average iBAQ intensity of that protein in the biological replicates, which is defined by Equation 2-1.

$$K = \frac{Concentration\ in\ the\ coacervate\ phase}{Concentration\ in\ the\ aqueous\ phase} = \frac{(Average\ iBAQ\ intensity)_{CO}}{(Average\ iBAQ\ intensity)_{AQ}} \times \frac{(Volume)_{AQ}}{(Volume)_{CO}} \quad \text{(Equation 2-1)}$$

2.3 RESULTS AND DISCUSSIONS

Addition of 8% (v/v) HFIP to the aqueous solution of 50 mM TBAB resulted in the formation of aqueous/coacervate two-phase system that was used for extraction and fractionation of yeast proteins between the aqueous and coacervate phases. Bottom-up proteomic analysis of each phase was conducted. Table 2-1 and Figure 2-2 show the number of proteins in each of the two phases and the total number of unique proteins identified using the two-phase system as compared to that of the control where 8M urea solution was used (*i.e.,* no phase separation). As can be seen, the proteomic analysis using the two-phase system resulted in identifying 441 additional proteins (237 proteins in the coacervate phase, 164 proteins in the aqueous phase and 40 proteins in both coacervate and aqueous phases) as compared to the control experiment. However, there are 51

unique proteins that were only identified in the control system. Overall, the use of FA*i*C-BPS resulted in identification of 390 additional proteins compared to the control (3043 vs. 2653 or 14.7% improvement).

Table 2-1. The number of identified proteins in all phases and the control.

Number of identified proteins in the coacervate phase (*CO*)	2748
Number of identified proteins in the aqueous phase (*AQ*)	1996
Number of identified proteins after phase separation (*CO + AQ*)	3043
Number of identified proteins without phase separation (NP)	2653
Protein identification improvement $\% = \frac{(CO + AQ) - NP}{NP} \times 100\%$	14.7%
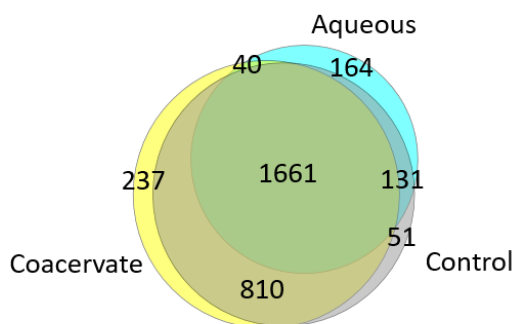
Figure 2-2. Venn diagram of the number of identified proteins in the coacervate phase (CO), aqueous phase (AQ), and control (NP).

2.3.1 ENRICHMENT OF LOW ABUNDANT PROTEIN

Recently, different methods have been reported for determination of the relative proteins abundance; for example, transcriptomic analyses [19], parallel metabolic pulse labelling of genes [20], isotope clusters and stable amino acid isotope labeled peptide pairing [21] and MS techniques. The protein abundance database for Saccharomyces cerevisiae, is available online (https://yeastmine.yeastgenome.org/ [15], curated from Ghaemmaghami et al.). For each protein, different abundances are reported based on the data available in different references. In this study,

we report the protein abundance as the number of molecules per cell. For each protein, the average of the abundance value from different databases was calculated and the results are reported as the database for the protein abundance in this study.

Figure 2-3 shows the number of identified proteins with different ranges of abundance values using the coacervation system (AQ + CO) and in the control system with no phase separation. As shown in Figure 2-3, over 300 more proteins with abundance of lower than 4000 molecules/cell were identified in the coacervate systems. In other words, improvement in the lower abundant proteins coverage (300 proteins) accounts for a significant majority of the overall coverage increase (390 proteins) using the FAiC-BPS vs. the control system (3043 vs. 2653). Figure 2-4 shows the results in terms of the percentage of identification improvement at different abundance ranges, where the identification improvement is calculated by comparing the number of proteins identified in the FAiC-BPS systems with those in the control. As can be seen, the coacervate system can identify significantly larger number of low abundance proteins than the control system.
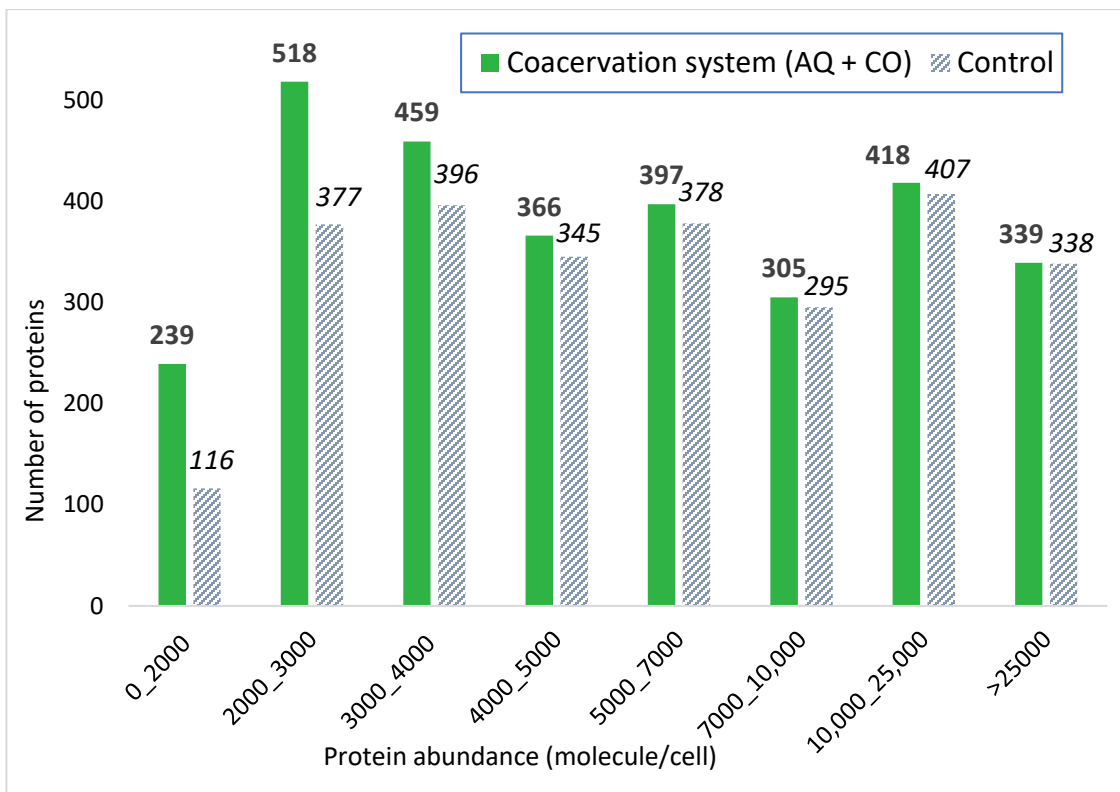
Figure 2-3. The number of identified proteins with different ranges of abundance in; green: using the coacervation system (AQ + CO), gray-striped: the control system with no phase separation.
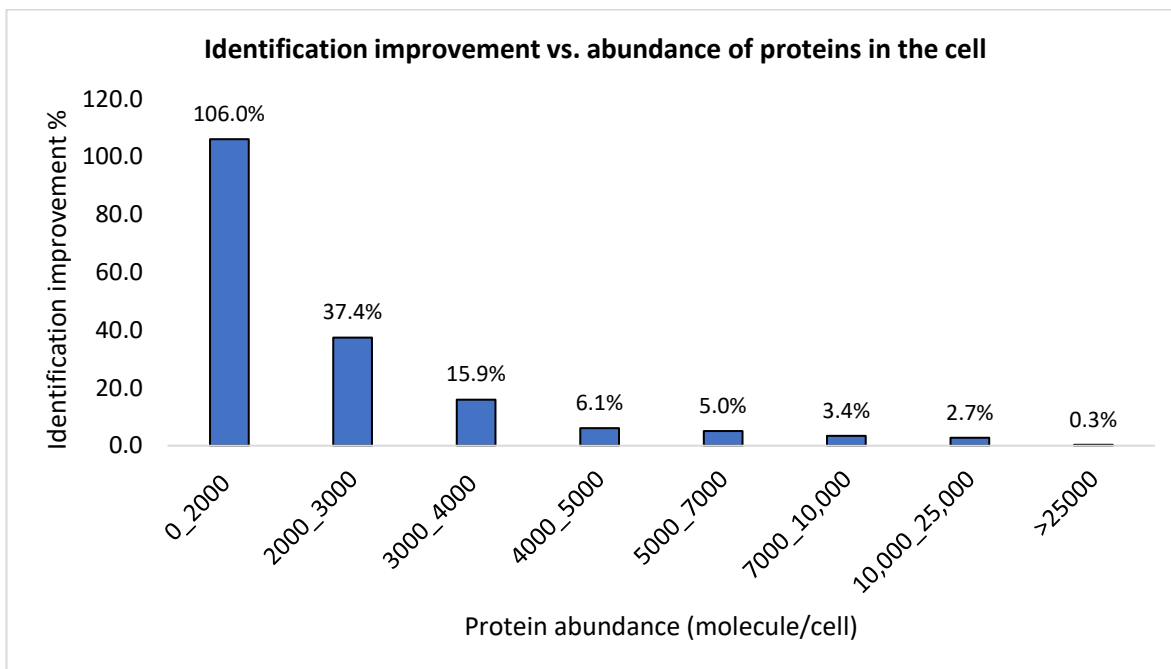
Figure 2-4. Percentage of identification improvement of proteins with different abundances in the cell.

The advantage of this two-phase system to enrich the low abundance proteins in separate phases in FA*i*C-BPS, is illustrated in Figure 2-5. More than 90% of the proteins that are detected only after coacervation (441 proteins) are enriched in either the coacervate or the aqueous phase. Generally, lower abundance proteins are selectively extracted and enriched in either coacervate or the aqueous phase which leads to the great identification improvement for these proteins, while higher abundance proteins are detected in both phases.

Aqueous

164 (37%)

40 (9%)

237 (54%)
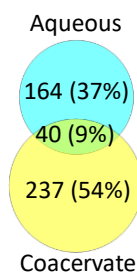
Coacervate

Figure 2-5. Fractionation of the proteins that are identified only after coacervation.

Extraction of solutes into the coacervate phase that has a small volume relative to the starting solution (6.5% of the total volume) results in the enrichment of low abundance proteins. As shown in Figure 2-6, the majority of low abundance proteins are extracted by the coacervate phase with a small volume. For example, 66% (158 out of 239) of the proteins with the abundance of smaller than 2000 molecule/cell are selectively extracted by the coacervate phase. At the same time, the coacervate phase has large concentrations of the amphiphile, TBAB, and HFIP which helps in solubilizing proteins. The high solubilization power of coacervate phases, in addition to their small volumes, make these phases suitable media for enrichment of low abundance proteins.
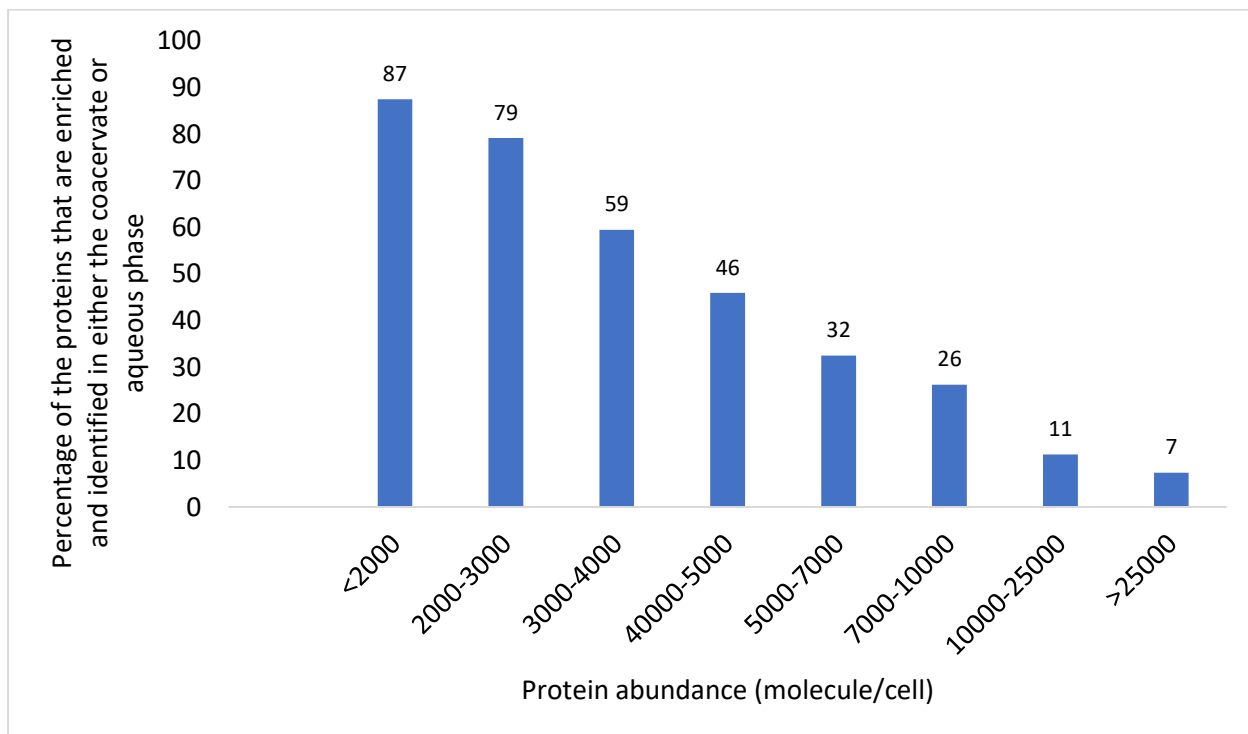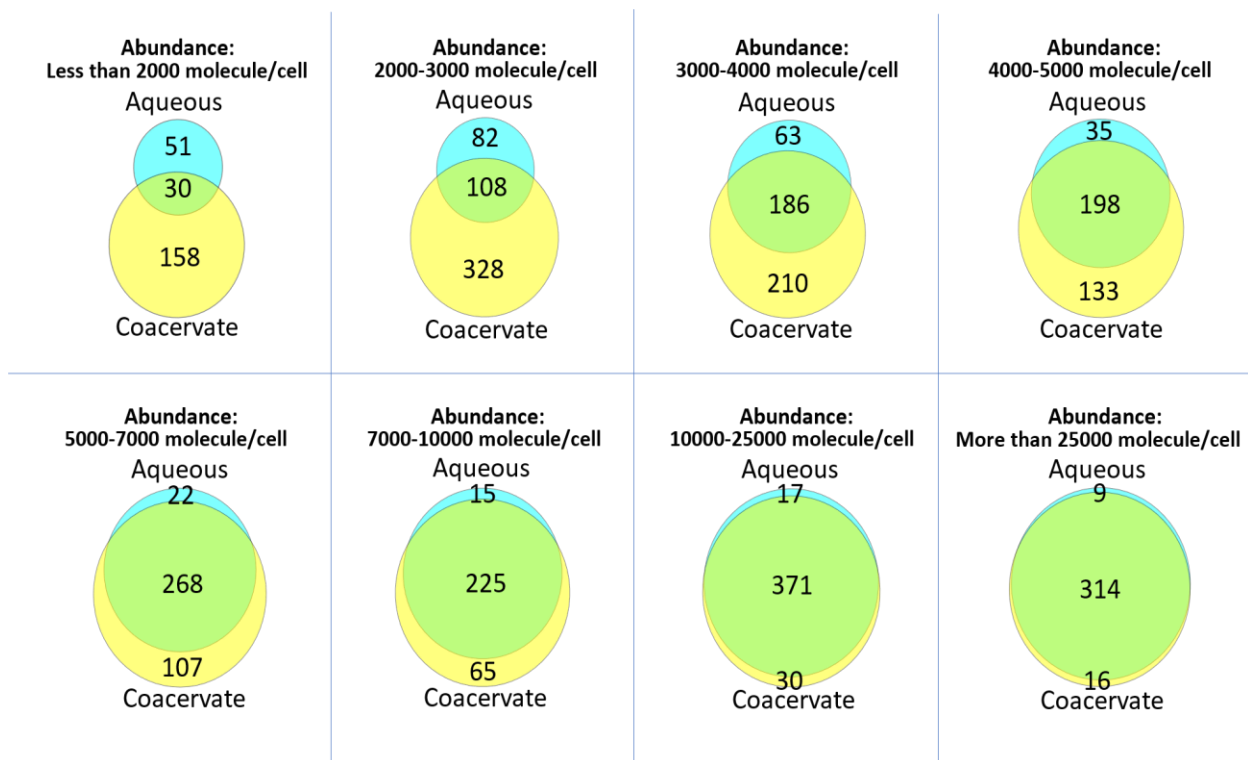
Figure 2-6. Venn diagrams: fractionation of proteins with different ranges of abundance; Bar chart: percentage of the proteins that are enriched and identified in only one phase (blue and yellow parts in the Venn diagrams) vs. their abundance.

22

2.3.2 PROTEIN FRACTIONATION PATTERNS BASED ON CELLULAR LOCATION AND FUNCTIONS

Table 2-2 lists the proteins coverage for each of the two phases and the total number of proteins in the in HFIP-TBAB two-phase system, as compared to that for the control (urea) system. Results show that the number of identified proteins were greater using the two-phase method for nearly all types of proteins classified based on the GO analysis in terms of their cellular locations, biological and chemical functions. Table 2-2 also illustrates that the majority of proteins can be found in the coacervate phase, thus they exhibit higher affinities for the coacervate phase as compared to the aqueous phase.

Table 2-2. Number of identified proteins in the coacervate phase, the aqueous phase, and the sample with no phase separation for selected GOs.

1- Cellular Component	Coacervate (CO)	Aqueous (AQ)	(CO + AQ)	No phase separation (NP)	Identification improvement
Membrane	944	639	1032	903	14%
Integral component of membrane	466	277	513	436	18%
Integral component of organelle membrane	103	74	117	106	10%
Intrinsic component of membrane	483	294	531	453	17%
Integral component of mitochondrial membrane	32	27	39	35	11%
Integral component of endoplasmic reticulum membrane	46	30	49	43	14%
Integral Component of Plasma membrane	27	7	28	20	40%
Cell wall	40	35	43	41	5%
RNA polymerase complex	69	40	70	60	17%
Anchored component of membrane	15	16	16	16	0%
Vesicle	181	105	190	162	17%
Vesicle tethering complex	29	5	31	24	29%
Golgi membrane	111	68	116	108	7%
Golgi apparatus	197	116	205	189	8%
Endosome	99	54	106	91	16%
Extracellular region	39	39	44	37	19%
Ribosome	120	162	172	143	20%

	Coacervate	Aqueous	(CO + AQ)	No phase separation (NP)	Identification improvement
Mitochondrial ribosome	23	65	68	43	58%
Mitochondrial matrix	110	145	165	128	29%
Mitochondrial membrane	192	160	224	193	16%
Mitochondrial outer membrane	76	53	78	71	10%
2- Biological process	**Coacervate**	**Aqueous**	**(CO + AQ)**	**No phase separation (NP)**	**Identification improvement**
Peptide metabolic process	219	247	275	246	12%
DNA duplex unwinding	26	8	26	25	4%
DNA repair	130	68	137	122	12%
DNA metabolic process	193	102	205	177	16%
Lipid metabolic process	174	97	176	162	9%
Membrane lipid metabolic process	40	17	42	35	20%
Lipid biosynthetic process	110	67	112	99	13%
Nuclear DNA replication	18	5	19	18	6%
Regulation of autophagy	29	14	29	25	16%
Amide biosynthetic process	228	248	284	253	12%
Protein targeting to vacuole	49	17	50	42	19%
Sphingolipid biosynthetic process	17	7	18	14	29%
Sphingolipid metabolic process	28	12	29	24	21%
Mitochondrial gene expression	39	75	86	55	56%
Peptide biosynthetic process	195	227	250	221	13%
Mitochondrial translation	32	71	78	50	56%
Translation	191	223	246	217	13%
Organic acid metabolic process	280	225	288	277	4%
Regulation of transcription by RNA polymerase II	216	138	241	207	16%
Response to abiotic stimulus	96	81	110	95	16%
Positive regulation of RNA metabolic process	183	114	204	171	19%
3- Molecular function	**Coacervate**	**Aqueous**	**(CO + AQ)**	**No phase separation (NP)**	**Identification improvement**
DNA-dependent ATPase activity	60	25	61	57	7%
Histone binding	35	16	35	34	3%
Structural constituent of ribosome	87	131	138	114	21%
Structural molecule activity	153	178	208	180	16%
Transcription regulator activity	83	52	96	74	30%
Regulatory region nucleic acid binding	49	32	57	40	43%
DNA binding	234	158	260	211	23%
Sequence-specific DNA binding	80	53	93	70	33%

Drug binding	434	262	439	400	10%
Nucleic acid binding	629	490	698	600	16%
Small molecule binding	590	375	598	522	15%
Nucleotide binding	543	343	551	506	9%
Ion binding	901	637	952	847	12%
Cation binding	403	320	446	389	15%
Anion binding	602	380	610	555	10%
Transferase activity	519	311	541	476	14%
Catalytic activity	1392	952	1462	1322	11%
Transferase activity, transferring hexosyl groups	56	27	57	51	12%
Oxidoreductase activity	215	183	232	217	7%
Unfolded protein binding	65	61	69	67	3%
Catalytic activity, acting on DNA	69	30	70	65	8%

Figure 2-7 depicts the distribution of subcellular proteins between the coacervate and aqueous phases pertaining to gene ontology categories of cellular locations, biological and chemical functions. Figure 2-7 simply illustrates the affinities of various subcellular proteins toward each of the two phases. For instance, integral membrane proteins, having some part of their peptide sequence embedded in the hydrophobic region of the phospholipid bilayer, were mostly extracted by the coacervate phase.[22] Phase separation improved identification of integral membrane proteins by 18% (+77 proteins) as compared to the control system. Similarly, the coacervate phase enriched and mostly extracted the proteins belonging to the gene ontology of "membrane-bounded vesicle", also known as "vesicle". Using the two-phase system resulted in identifying 190 vesicle proteins which shows a 17% increase versus the control with 162 vesicle proteins. Among these 190 vesicle proteins, 181 proteins are identified in the coacervate phase which shows the selectivity of the coacervate phase towards extraction and enrichment of membrane-bounded vesicle proteins. The coacervate phase extracts nearly all (174 out of 176) proteins that are involved in the lipid metabolic process (LMP) and DNA metabolic process (193 out of 205).

Cellular component

Mitochondrial ribosome
Aqueous
45
20
3
Coacervate

Mitochondrial part
Aqueous
89
218
79
Coacervate

Ribosome
Aqueous
52
110
10
Coacervate

Mitochondrial large ribosomal subunit
Aqueous
26
12
Coacervate

Mitochondrial matrix
Aqueous
55
90
20
Coacervate

Integral component of membrane
Aqueous
47
230
236
Coacervate

Intrinsic component of membrane
Aqueous
48
246
237
Coacervate

Golgi membrane
Aqueous
5
63
48
Coacervate

Golgi apparatus
Aqueous
8
108
89
Coacervate

Vesicle
Aqueous
9
96
85
Coacervate

Tethering complex
Aqueous
2
3
26
Coacervate

DNA repair complex
Aqueous
4
7
Coacervate

Endosome
Aqueous
7
47
52
Coacervate

Chromosome
Aqueous
15
105
134
Coacervate

RNA polymerase complex
Aqueous
1
39
30
Coacervate

Molecular function

Structural constituent of ribosome
Aqueous
51
80
7
Coacervate

Structural molecule activity
Aqueous
55
123
30
Coacervate

Histone binding
Aqueous
16
19
Coacervate

Catalytic activity
Aqueous
70
882
510
Coacervate

Catalytic activity, acting on DNA
Aqueous
1
29
40
Coacervate

DNA binding
Aqueous
26
132
102
Coacervate

DNA-dependent ATPase activity
Aqueous
1
24
36
Coacervate

Ion binding
Aqueous
51
586
315
Coacervate

Anion binding
Aqueous
8
372
230
Coacervate

Cation binding
Aqueous
43
277
126
Coacervate

Nucleic acid binding
Aqueous
69
421
208
Coacervate

Nucleotide binding
Aqueous
8
335
208
Coacervate

Drug binding
Aqueous
5
257
177
Coacervate

Small molecule binding
Aqueous
8
367
223
Coacervate

Transferase activity
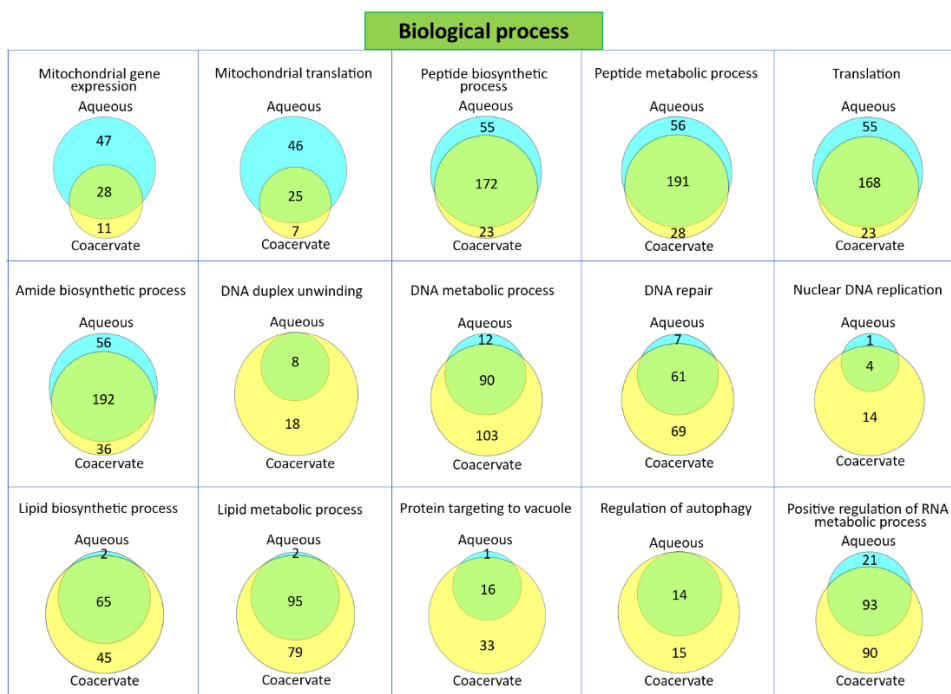Aqueous
22
289
230
Coacervate

26

Figure 2-7. Distribution of proteins between the coacervate phase and the aqueous phase for selected GOs.

The aqueous phase showed selectivity towards certain types of proteins, for example, proteins pertaining to the mitochondrial ribosome and mitochondrial matrix. The total number of identified mitochondrial ribosome proteins (MRP) using the two-phase method was improved by 58% (+25 proteins) as compared to the control, with the majority of the MRP (65 out of 68) extracted into the aqueous phase. Phase separation also improved identification of mitochondrial matrix proteins by 29% (+37 proteins) as compared to the control system. Most of the proteins that belong to the mitochondrial gene expression, mitochondrial translation, peptide metabolic process and peptide biosynthetic process have affinity to the aqueous phase.

2.3.3 ELECTROSTATIC AND HYDROPHOBIC EFFECTS ON DISTRIBUTION PATTERNS

It is important to note that coacervation systems with different amphiphiles and composition can exhibit different selectivity patterns towards various groups of proteins. Two main driving forces that control distribution patterns of proteins between the two phases are hydrophobic effect

and electrostatic interaction. The balance between the two types of interactions can be controlled by changing the chemical composition and nature of amphiphilic molecules in the coacervation system. Therefore, various coacervate systems show different selectivity towards different proteins groups. Although the TBAB/HFIP and the previously reported lipid coacervate systems both extracted the majority of hydrophobic proteins, the protein fractionation pattern between the aqueous and the coacervate phase in the TBAB/HFIP system is different due to electrostatic effects.[9]

For example, mitochondrial ribosome proteins were mostly identified in the lipid coacervate phase, while they were mostly identified in the aqueous phase of the TBAB/HFIP system. More than 97% of the identified mitochondrial ribosome proteins have isoelectric points of greater than 9, so they are positively charged under FA*i*C environments and are repelled by the positively charged tetrabutylammonium in the coacervate phase in the HFIP/TBAB system and are mostly found in the aqueous phase in the HFIP-TBAB two-phase system (Figure 2-8), while these proteins were mostly extracted by the lipid coacervate phase in the NLC system that has a net negative charge [9]. In the natural lipid coacervate (NLC) system reported previously, amphiphilic molecules are negatively charged or zwitterionic phospholipids.

On the other hand, chromosome proteins generally have lower isoelectric points and consequently carry net negative charges and are therefore attracted by the net positive charge of the coacervate phase in the HFIP/TBAB system, while repelled by the negatively charged lipid coacervate phase. Figure 2-8 demonstrates the effect of isoelectric point on their distribution of the two groups of proteins between the aqueous and coacervate phases in the HFIP-TBAB system.
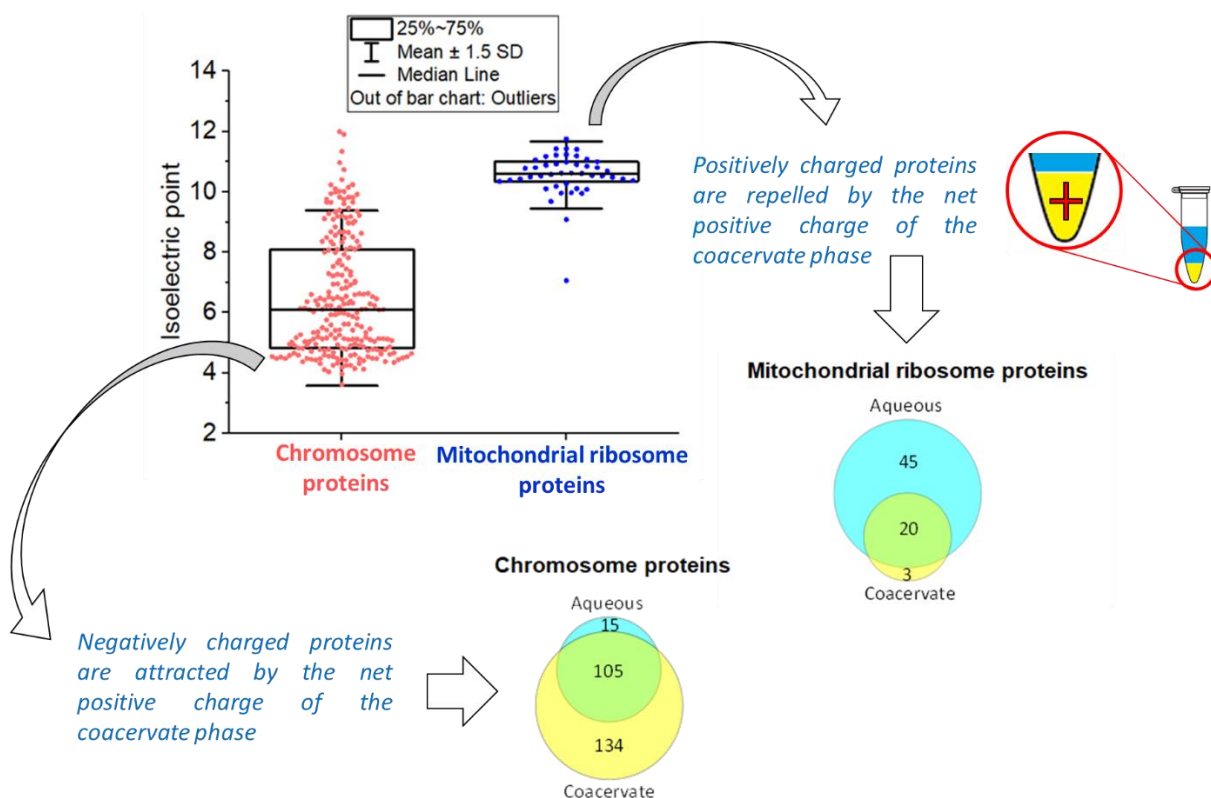
Figure 2-8. The effect of isoelectric point of proteins on their distribution between the two phases.

Figure 2-9(a) shows the distribution of the all unique proteins found in the aqueous and coacervate phase of HFIP-TBAB system as a function of their isoelectric points. The pI of more than 75% of the proteins that were only identified in the coacervate phase were less than 7.67 (median pI = 6.23), while the isoelectric points of more than 75% of the proteins that were only identified in the aqueous phase were higher than 7.55 (median pI = 9.74). The significantly different median pI in the AQ Phase (9.74) vs. HFIP-TBAB coacervate phase (pI=6.23) can be attributed to the electrostatic effects due to interactions of proteins with lower pI (more acidic) with the cationic coacervate. This trend is the opposite of what we have previously seen in the (NLC) system with the anionic lipid-based coacervate [9], which is shown in Figure 2-9(b). Therefore, the net charge of coacervate phase affects the electrostatic interaction and distribution patterns of proteins.
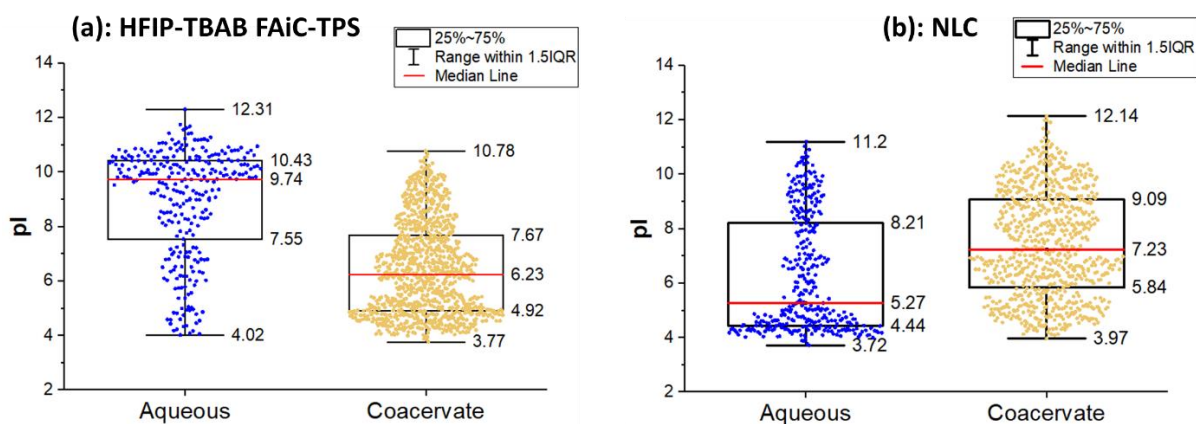
Figure 2-9. Distribution of isoelectric points of proteins which were uniquely identified in the coacervate phase and aqueous phase of (a): HFIP-TBAB FAiC-BPS with net positive charge in the coacervate phase, (b): Natural Lipid Coacervation (NLC) with net negative charge in the coacervate phase. The areas between quartiles show distribution of 25% of data points.

Figure 2-10 shows the comparison between partition coefficients of proteins found in both coacervate and aqueous phases as a function of pI ranges. Proteins were categorized in three groups of pI<5, 5≤pI≤9, and pI>9. Distribution patterns of proteins and their partition coefficients between two phases are a function of many parameters including, but not limited to, their isoelectric point, hydrophobicity, molecular weight, and conformation. However, proteins with larger isoelectric points generally carry net positive charge and therefore they are repelled by the positive charge of TBA in the coacervate phase of HFIP-TBAB FA*i*C-BPS. As a result, more basic proteins have smaller partition coefficients and more acidic proteins have larger partition coefficients. The apparent affinity of the more acidic proteins toward the positively charged TBAB coacervate is indicative of electrostatic effects. The distribution of partition coefficient (Figure 2-10) of proteins with pI<5, 5≤pI≤9, and pI>9 had the median of 50.3, 41.1, and 10.7, respectively.
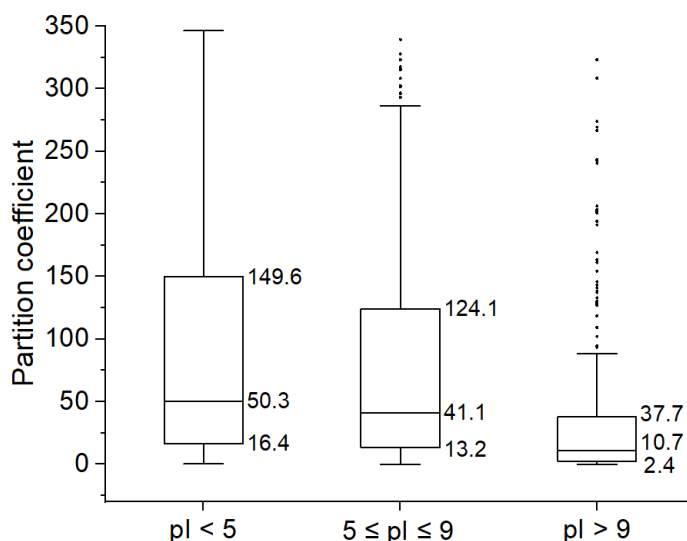
Figure 2-10. Partition coefficient of proteins in three different ranges of isoelectric points.

The distribution patterns of proteins between the aqueous and coacervate phases were also examined with respect to hydrophobicity and molecular mass. The hydrophobicity of proteins can be assessed by their GRAVY [23] scores. Generally, more hydrophobic proteins (i.e., with higher GRAVY scores) are extracted and enriched by the coacervate phase due to its more hydrophobic nature, while hydrophilic proteins having more affinity toward the aqueous-rich phase. Figure 2-11 illustrates the distribution of the 1047 proteins uniquely found in the TBAB-HFIP coacervate phase, the 295 unique proteins found only in the aqueous phase, and the 1701 proteins distributed between the two phases as a function of the GRAVY scores. The average GRAVY value of the proteins identified only in the coacervate phase (median: - 0.35) was higher than the average GRAVY value of the proteins identified in the aqueous phase (median: - 0.75). The proteins with GRAVY values below -1.78 were only identified in the aqueous phase; examples include stress protein DDR48 with GRAVY value of -1.96, which is a DNA damage-responsive protein, and SERF-like protein YDL085C-A with GRAVY value of -1.80.
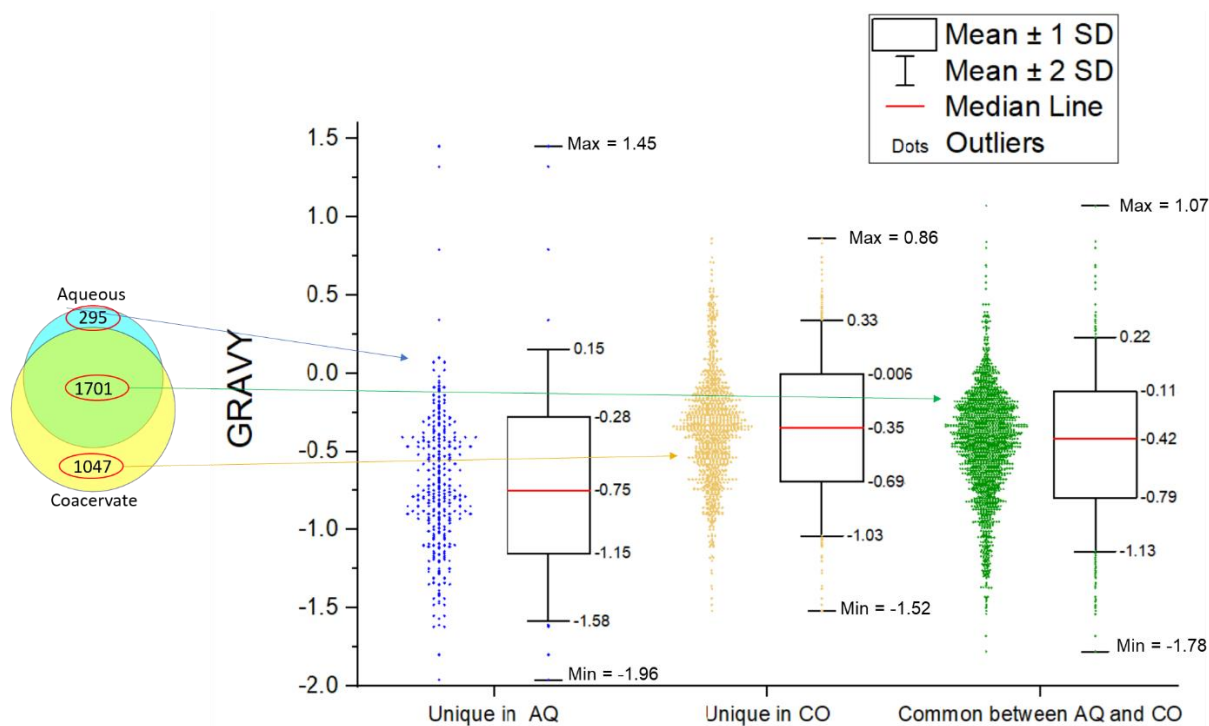
Figure 2-11. Comparison between GRAVY score of proteins that were uniquely identified in aqueous or coacervate phases.

Figure 2-12 shows the distribution patterns for the 1701 proteins that were identified in both aqueous and coacervate phases in terms of their partition coefficients. Proteins with GRAVY values larger than -0.6 (more hydrophobic proteins) generally have larger partition confident than the proteins with GRAVY values lower than – 0.6 (more hydrophilic proteins) due to the hydrophobic effect. The discernible affinity of proteins with larger GRAVY values (more hydrophobic proteins) toward the coacervate phase with hydrophobic nature due the large concentration of HFIP and amphiphiles, is indicative of hydrophobic effects.
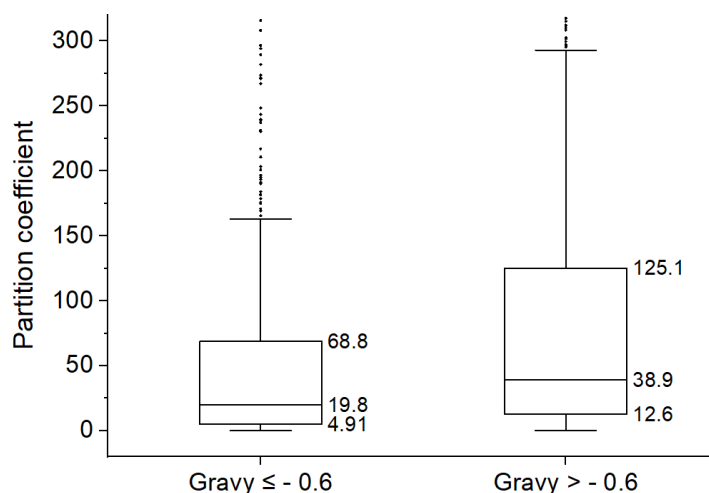
Figure 2-12. Partition coefficient of proteins in two different ranges of GRAVY numbers.

As shown in Figure 2-13, generally proteins with larger molecular weights have greater affinity toward the coacervate phase. One possible explanation is that proteins are denatured due to the high concentration of HFIP and as they unfold, hydrophobic segments of proteins are exposed and interact with the coacervate phase. The larger proteins would on average have a larger number of hydrophobic moieties.
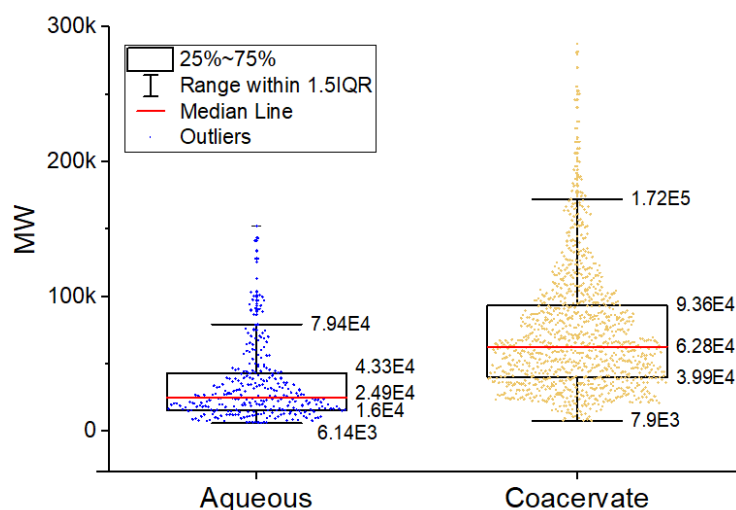

Figure 2-13. Distribution of molecular weights of proteins which were uniquely identified in the coacervate phase and aqueous phase of HFIP-TBAB FAiC-BPS.

While the effects of hydrophobicity and electrostatic interactions on partitioning into the coacervate phase are evident, the relationships with structural descriptors, pI, GRAVY, and MW are not simple or direct. For example, the two most hydrophobic proteins, P61829 (GRAVY= 1.45 and pI= 8.52) and P87284 (GRAVY=1.32 and pI=4.2) were only found in the aqueous phase; and not in the TBAB coacervate phase or in the urea control system. One is a basic protein, and the other is acidic.

A more in-depth analysis of phase distribution – structure relationships and subsequently the underlying molecular interactions and mechanisms would require a better understanding of proteins conformations in different phases. A complicating factor is the impact of HFIP and other chemicals on proteins structures. Fluoroalcohols such as HFIP are known to denature the tertiary structure of proteins due to hydrophobic effect, while stabilizing □-helices and □-sheets due to hydrogen bonding interactions [24]. The presence of an amphiphile such as TBAB would further complicates the picture. Although one cannot rely on the accuracy of structural descriptors such as GRAVY or pI to truly represent effective hydrophobicity and charge of the proteins, the descriptors adequately reveal the impact of electrostatic and hydrophobic effects on proteins distribution patterns in the FA*i*C-BPS.

2.4 CONCLUSIONS

Coacervation improves identification of low abundance proteins. Generally, low abundance proteins are enriched in one of the two phases, coacervate phase or aqueous phase that facilitates their identification. Identification of proteins with the abundance of lower than 2000 molecules per cell was improved by 106% in the FA*i*C-BPS made of HFIP and TBAB as compared to the control system. In addition, coacervation facilitates identification of hydrophobic proteins. Generally, more hydrophobic proteins are extracted and enriched by the coacervate phase due to the hydrophobic effect. In contrast, more hydrophilic proteins are enriched in the aqueous phase.

However, this is not a universal rule and few very hydrophobic proteins with large GRAVY values such as 1.45 and 1.32 were identified only in the aqueous phase. Each phase has selectivity towards specific groups of subcellular proteins; for example, proteins pertaining to integral component of membrane, vesicle part, and chromosome were mostly extracted by the coacervate phase. We were able to identify more integral membrane proteins by 18% (+77 proteins) when TBAB-HFIP coacervation was used in the proteomics workflow. Fractionation patterns between the aqueous and coacervate phases can be rationalized based on molecular descriptors for hydrophobicity (GRAVY), charge (pI), and size (molecular mass).

In conclusion, coacervation can be incorporated into proteomics workflow to improve identification of low abundance and hydrophobic proteins as a result of concomitant extraction, fractionation, and enrichment.

ACKNOWLEDGMENTS

2.5 REFERENCES

(1)   Zhang, Y.; Lin, Z.; Tan, Y.; Bu, F.; Hao, P.; Zhang, K.; Yang, H.; Liu, S.; Ren, Y. Exploration of Missing Proteins by a Combination Approach to Enrich the Low-Abundance Hydrophobic Proteins from Four Cancer Cell Lines. *J. Proteome Res.* **2020**, *19* (1), 401–408. https://doi.org/10.1021/acs.jproteome.9b00590.

(2)   Reinhoudt, D. N. *Supramolecular Materials and Technologies*; John Wiley & Sons, 2008.

(3)   Ballesteros-Gómez, A.; Rubio, S.; Pérez-Bendito, D. Potential of Supramolecular Solvents for the Extraction of Contaminants in Liquid Foods. *J. Chromatogr. A* **2009**, *1216* (3), 530–539. https://doi.org/10.1016/j.chroma.2008.06.029.

(4)   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.

(5)   McCord, J. P.; Muddiman, D. C.; Khaledi, M. G. Perfluorinated Alcohol Induced Coacervates as Extraction Media for Proteomic Analysis. *J. Chromatogr. A* **2017**, *1523* (Supplement C), 293–299. https://doi.org/10.1016/j.chroma.2017.06.025.

(6)   Khaledi, M. G.; Jenkins, S. I.; Liang, S. Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles. *Langmuir* **2013**, *29* (8), 2458–2464. https://doi.org/10.1021/la303035h.

(7)   Nejati, M. M.; Khaledi, M. G. Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis. *Langmuir* **2015**, *31* (20), 5580–5589. https://doi.org/10.1021/acs.langmuir.5b00444.

(8)   Jenkins, S. I.; Collins, C. M.; Khaledi, M. G. Perfluorinated Alcohols Induce Complex Coacervation in Mixed Surfactants. *Langmuir* **2016**, *32* (10), 2321–2330. https://doi.org/10.1021/acs.langmuir.5b04701.

(9)   Koolivand, A.; Azizi, M.; O'Brien, A.; Khaledi, M. G. Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies. *J. Proteome Res.* **2019**, *18* (4), 1595–1606. https://doi.org/10.1021/acs.jproteome.8b00857.

(10)   Koolivand, A.; Clayton, S.; Rion, H.; Oloumi, A.; O'Brien, A.; Khaledi, M. G. Fluoroalcohol – Induced Coacervates for Selective Enrichment and Extraction of Hydrophobic Proteins. *J. Chromatogr. B* **2018**, *1083*, 180–188. https://doi.org/10.1016/j.jchromb.2018.03.004.

(11)   Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72* (1), 248–254. https://doi.org/10.1016/0003-2697(76)90527-3.

(12)   Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6* (5), 359–362. https://doi.org/10.1038/nmeth.1322.

(13)   Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. One-Hour Proteome Analysis in Yeast. *Nat. Protoc.* **2015**, *10* (5), 701–714. https://doi.org/10.1038/nprot.2015.040.

(14)   Perseus http://www.biochem.mpg.de/5111810/perseus (accessed Aug 3, 2018).

(15)   YeastMine: Home https://yeastmine.yeastgenome.org/yeastmine/begin.do (accessed Nov 25, 2019).

(16)   GRAVY Calculator http://www.gravy-calculator.de/ (accessed Nov 4, 2018).

(17)   Kozlowski, L. P. IPC – Isoelectric Point Calculator. *Biol. Direct* **2016**, *11* (1), 55. https://doi.org/10.1186/s13062-016-0159-9.

(18)   Kozlowski, L. P. Proteome- *PI*: Proteome Isoelectric Point Database. *Nucleic Acids Res.* **2017**, *45* (D1), D1112–D1116. https://doi.org/10.1093/nar/gkw978.

(19)   Vogel, C.; Marcotte, E. M. Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses. *Nat. Rev. Genet.* **2012**, *13* (4), 227–232. https://doi.org/10.1038/nrg3185.

(20)   Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global Quantification of Mammalian Gene Expression Control. *Nature* **2011**, *473* (7347), 337–342. https://doi.org/10.1038/nature10098.

(21)   Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372. https://doi.org/10.1038/nbt.1511.

(22)   Stirling, C. J.; Rothblatt, J.; Deshaies, R.; Schekman, R. Protein Translocation Mutants Defective in the Insertion of Integral Membrane Proteins into the Endoplasmic Reticulum. *Mol. Biol. Cell* **1992**, *3*, 14.

(23)   Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157* (1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

(24)   Roccatano, D.; Fioroni, M.; Zacharias, M.; Colombo, G. Effect of Hexafluoroisopropanol Alcohol on the Structure of Melittin: A Molecular Dynamics Simulation Study. *Protein Sci. Publ. Protein Soc.* **2005**, *14* (10), 2582–2589. https://doi.org/10.1110/ps.051426605.

CHAPTER 3

OPTIMIZATION OF FLUOROALCOHOL MEDIATED SUPRAMOLECULAR

BIPHASIC SYSTEMS TO ENHANCE PROTEIN COVERAGE IN

PROTEOMICS: EFFECTS OF QUATERNARY AMMONIUM SALTS AND PH

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

ABSTRACT

The Fluoroalcohol Induced Supramolecular Biphasic (FAiS-BP) systems are composed of a supramolecular phase which is a highly concentrated mixture of a fluoroalcohol (such as Hexafluoroisopropanol, HFIP) and a Quaternary Ammonium Salt (QUATS), and a separate aqueous-rich phase. By taking the advantage of FAiS-BP in proteomics studies, we can concomitantly extract, fractionate, and enrich proteins; thus, an additional separation dimension and a simple step of enrichment would be added prior to the LC-MS. Proteins of Saccharomyces cerevisiae yeast were fractionated between the two phases of FAiS-BP systems and subjected to tryptic digestion and LC-MS/MS analysis. The focus of this study is to investigate the effect of different QUATS, with different chain lengths, on the interactions within the FAiS-BP systems, and controlling the electrostatic interactions by changing the pH of system. The results were compared to the conventional method for solubilization of proteins samples using 8M urea

solutions (single phase) as the control experiment. Fractionation by FA*i*S-BP resulted in enrichment of low-abundance proteins in one of the two separate phases, and consequently improved their identification. In the biphasic system that tetrabutylammonium bromide was used as QUAT, identification of low-abundance proteins (less than 2000 molecule/cell) improved by 104%, which was equivalent to detection of an additional 123 proteins as compared to the urea control system. Interestingly, larger improvements were observed as the protein abundance decreased. In the abundance range of below 5000 molecules/cell, FA*i*S-BP systems resulted in detection of up to 1585 proteins, which is about 330 additional proteins compared to the urea control. Above a certain abundance level (~5000 molecules/cell), there is little or no difference between protein coverage using the conventional methods of solubilizing the protein sample in the urea solution or FA*i*S-BP system in the bottom-up proteomics. This suggests that the FA*i*S-BP systems are particularly advantageous for detecting low abundance proteins. We found that by changing the pH of the fractionation media, we can control the fractionation patterns. Additionally, due to the loss of tertiary structure and increase in α-helix structures at pH≈3, improved sequence coverage of transmembrane α-helices can be achieved in FA*i*S-BP at this pH.

KEYWORDS: Coacervates, Fluoroalcohols, Hydrophobic Proteins, Low Abundance Proteins, Proteomics, Sample Fractionation and Enrichment, Subcellular Proteomics.

3.1 INTRODUCTION

In proteomics analysis, the missing proteins from the list based on gene-coding are often very hydrophobic and or in low abundance. Detection and identification of these types of proteins pose significant challenges.[1–3] Hydrophobic proteins are not effectively solubilized using common solubilizing reagents (e.g. using urea or detergents); lack of protein solubility simply reduces the effective concentration of proteins, and their tryptic peptides in solutions. Another common

solubilizing reagent, sodium dodecyl sulphate (SDS), is known to be an effective in solubilizing hydrophobic proteins, however, the problem with SDS is the strong interaction with proteins that results in difficulties of its removal required for effective digestion. In addition, obviously, incomplete digestion suppresses identification of proteins.[4] Low abundance proteins are underrepresented under the shadow of proteins in higher abundance. We have recently introduced unique Fluoroalcohol induced Supramolecular Biphasic (FA*i*S-BP) systems to address these challenges in proteomics workflows.[5,6]

Supramolecular solvents refer to nanostructured liquids which are produced in colloidal solutions of amphiphilic compounds through the spontaneous and sequential mechanism of self-assembly and coacervation.[7–11] These supramolecular structures coacervate and form a separate phase from the aqueous media due to due to differences in polarity and density. The amphiphile-rich phases in these FA*i*S-BP are also known as the coacervate phases. [12,13] Our lab has previously reported the formation of unique supramolecular solvents, known as Fluoroalcohol-Induced Coacervate Biphasic Systems (FA*i*C-BPS).[9,10] Addition of fluoroalcohols, such as hexafluoroisopropanol (HFIP) and trifluoroethanol (TFE), to the aqueous solution of amphiphilic molecules, results in formation of unprecedented variety of biphasic systems with unique chractristics.[9,14] We have shown the applications of FA*i*C-BPS in proteomics studies for better characterization of low abundance and hydrophobic proteins.[5] Thanks to the unique capabilities of FA*i*C-BPS to concomitantly extract, fractionate, and enrich proteins in separate phases, incorporation of these systems in bottom-up proteomics workflow leads to enhancement of coverage, especially for low-abundance proteins and hydrophobic proteins such as integral membrane proteins.[5,6]

In FA*i*C-BPS, the coacervate phases have very high concentrations of the amphiphiles and fluoroalcohols, therefore they are very powerful solvents for hydrophobic proteins. Additionally,

their volumes are generally a very small fraction of the volume of initial solution. Consequently, when compounds are extracted into the coacervate phases, they get enriched simultaneously. Using FA*i*C-BPS in proteomics applications would allow fractionation of complex protein mixtures between the two phases, effective solubilization and extraction of hydrophobic proteins into the coacervate phase, and enrichment of specific protein groups in one of the phases.

We have previously demonstrated that addition HFIP to the natural phospholipids of the cell membranes results in formation of a FA*i*S-BP at which hydrophobic proteins, such integral membrane proteins (IMP), get enriched and solubilized in the coacervate phase(amphiphile-rich phase).[5] Additionally, previous studies from this lab has reported that addition of HFIP to the solution of tetrabutylammonium bromide (TBAB) leads to formation of a FA*i*C-BPS at which proteins get fractionated between the two phases based on their hydrophobicity's and isoelectric points. More hydrophobic proteins and more acidic proteins were enriched in the coacervate phase, and more hydrophilic and more basic proteins were enriched in the aqueous phase.[6] In this article, we use different FA*i*C-BPS, formed by addition of HFIP to the aqueous solutions of different quaternary ammonium salts (QUATS), in proteomics applications. The QUATS have the same quaternary ammonium head group (independent of the solution pH) [15,16], but different hydrophobic alkyl chain lengths. The cationically charged quaternary ammonium headgroup interacts electrostatically with the charged groups of proteins, while the alkyl chains have hydrophobic interaction with the hydrophobic parts of the proteins. Protein fractionation patterns change in different FA*i*C-BPS owing to the different balances of electrostatic interactions and hydrophobic effects. The combination of these two forces, mostly governs protein fractionation pattern in proteomics practices. Here, we present novel FA*i*C-BPS and their application in proteomics workflow to address the challenges in characterization of low abundance and/or hydrophobic proteins. The electrostatic interaction and hydrophobic effect were studied by changing the type

of QUAT amphiphiles in FA*i*C-BPS. In addition, the effect of protein charge on fractionation pattern in the FA*i*C-BPS was studied. We also compare these FA*i*C-BPS in terms of transmembrane α-helical sequence coverage, and fractionation pattern of different protein groups between the two phases.

3.2 EXPERIMENTAL DESIGN

3.2.1 MATERIALS, CHEMICALS, AND REAGENTS

Tetrabutylammonium bromide (TBAB), Tetrapropylammonium bromide (T.Propyl.AB), tetrapentylammonium bromide (T.Pentyl.AB), and tetrahexylammonium bromide (THAB) were purchased from ACROS Organics. Tetraoctylammonium bromide (TOAB) was provided from Chem-Impex International and TCI America, respectively. 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Chemical, USA. Tris HCl and Tris base were purchased from Sigma-Aldrich. Dithiothreitol (DTT), Iodoacetamide (IAA) and Sequencing grade trypsin were purchased from Fisher BioReagents, Alfa Aesar™, and Promega Corporation, respectively. LC-MS acetonitrile (ACN), isopropanol (IPA) and deionized water were provided from Fisher Chemical, USA.

3.2.2 FORMATION OF DIFFERENT SUPRAMOLECULAR BIPHASIC SYSTEMS FOR PROTEIN EXTRACTION AND FRACTIONATION

Yeast cell lysate was prepared as explained in detail in a recently publish paper from this lab. 400 µg proteins (50 µL from the cell lysate with concentration of 8 µg/µL of protein) were subjected to phase separation by addition of the cell lysate to the solution of amphiphiles, followed by addition of HFIP and centrifugation. The concentration of amphiphiles in the original solution (prior to the centrifugation and phase separation) was kept constant at 50 mM, and percentage of HFIP was 8% (V/V). The pH of the solution was adjusted to the required value by addition of required amounts of 2.5 M Tris buffer with pH=8.5. Details about the original solutions to make

FA*i*C-BPS for protein fractionation are presented in Table 3-1, where all volumes are based on total volume of 1 mL in the initial solution. It is worth noting that all amphiphiles are water soluble except THAB and TOAB; these two amphiphiles were initially solubilized in HFIP and then added to the solution to induced two-phase system.

Table 3-1. List of different Fluoroalcohol induced Coacervate-Biphasic Systems (FA*i*C-BPS) used for protein extraction, fractionation, and enrichment (volumes are based on total volume of 1 mL)

System	μL of cell lysate (8 μg protein/μL)	Amphiphile	2.5 M Tris buffer, pH=8.5 (μL)	DI water (μL)	HFIP (μL)
T. Propyl AB_HFIP, pH≈3	50	50 mM (111 μL from 450 mM T. Propyl AB)	0	770	80
TBAB_HFIP, pH≈3	50	50 mM (111 μL from 450 mM TBAB)	0	759	80
T. Pentyl AB_HFIP, pH≈3	50	50 mM (250 μL from 200 mM T. Pentyl AB)	0	420	80
THAB_HFIP, pH≈3	50	50 mM (80 μL from 625 mM THAB in HFIP)	0	870	0
TOAB_HFIP, pH≈3	50	50 mM (80 μL from 625 mM TOAB in HFIP)	0	870	0
TBAB_HFIP_pH=5.5	50	50 mM (111 μL from 450 mM TBAB)	36	723	80
TBAB_HFIP_pH=8.4	50	50 mM (111 μL from 450 mM TBAB)	400	359	80

After the formation of two phases and fractionation of the yeast proteins, the aqueous and supramolecular phases were separated. The amphiphiles were removed using FASP and the proteins in each phase were digested on filter using trypsin. It is noteworthy that HFIP must be evaporated under nitrogen flow prior to loading the samples to the FASP filter because HFIP dissolves the cellulose filter. Dried samples were then dissolved the solution of 5 M urea and 2 M thiourea in 70% IPA, and then loaded on to FASP filters. Details about the cell lysate, protein fractionation and solubilization in FA*i*C-BPS, amphiphile removal, digestion and other sample preparation steps can be found in a recently published paper from this group.

3.2.3 MEASUREMENT OF RELATIVE HYDROPHOBICITY OF QUATS SUPRAMOLECULAR PHASES

Partition coefficient of homologous series of alkylphenones, including acetophenone, propiophenone, and butyrophenone between the coacervate and aqueous phases were measured by liquid chromatography. A RPLC C18 column (Thermo Scientific, part number: 28105-154630) with dimensions of 150×4 mm, and particle size of 5 µm (Spherical, Fully Porous) was used for separation of alkylphenones. Mobile phase A water, mobile phase B was ACN, and the gradient was 35% to 80% B in the first 4 minutes followed by 35% B from 4 to 7 minutes to recondition the column for the next injection. The coacervate phase was diluted 500 time before injection to prevent detector saturation and the absorbance was recorded at 254 nm. Related chromatograms and figures can be found in Appendix 3-3.

3.3 RESULTS AND DISCUSSIONS

3.3.1 IDENTIFICATION IMPROVEMENT OF LOW ABUNDANCE PROTEINS

The number of identified proteins in each of the FA*i*C-BPS samples were compared to the control system, 8M urea solution. In Figure 3-1, the Venn diagram of identified proteins in the two phases of each FA*i*C-BPS are shown, and the number of identified proteins in each system and their identification improvement compared to the urea control are listed.
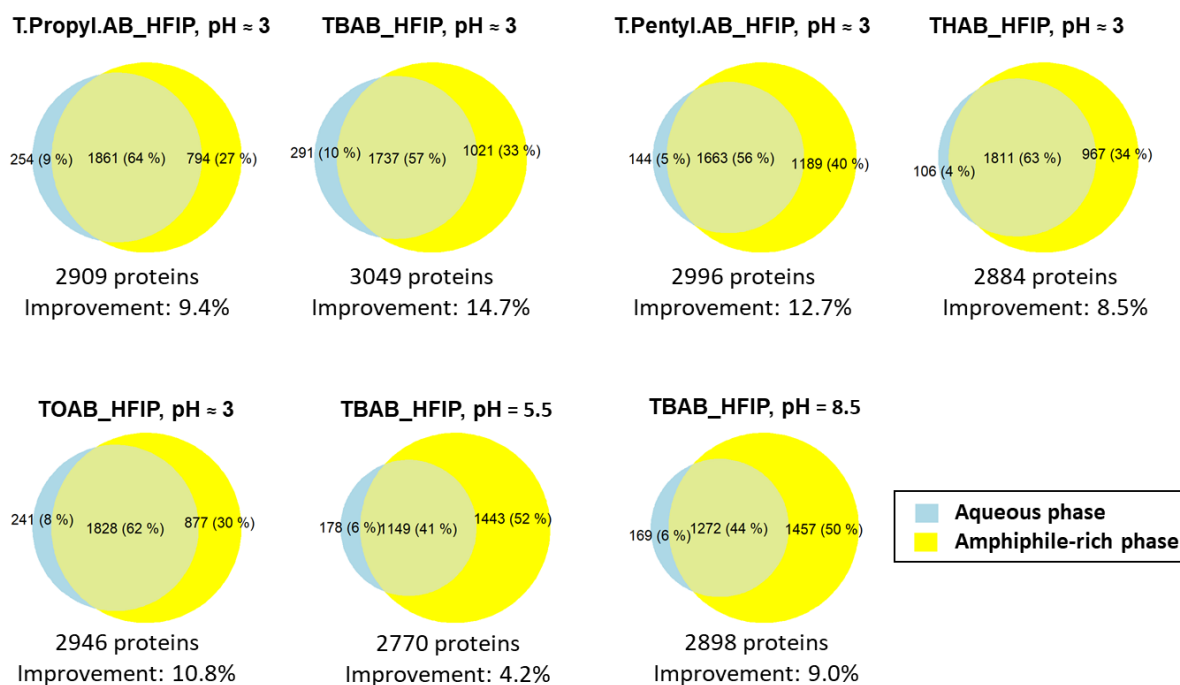
Figure 3-1. Venn diagram of identified proteins in different FA*i*C-BPS, and the identification improvement in each system compared to the urea control.

All the FA*i*C-BPS showed protein identification improvement compared to the control. Interestingly, further analysis revealed that in all these FA*i*C-BPS, most of the improvement is due to the identification of larger number of low abundance proteins (see section 3.3.2). Figure 3-2 compares the identification improvement of proteins in different abundance ranges using different FA*i*C-BPS. The main differences between systems with different chain lengths and pH values are different level of fractionation between two phases and coverage of low abundance proteins that will be discussed in the following sections.
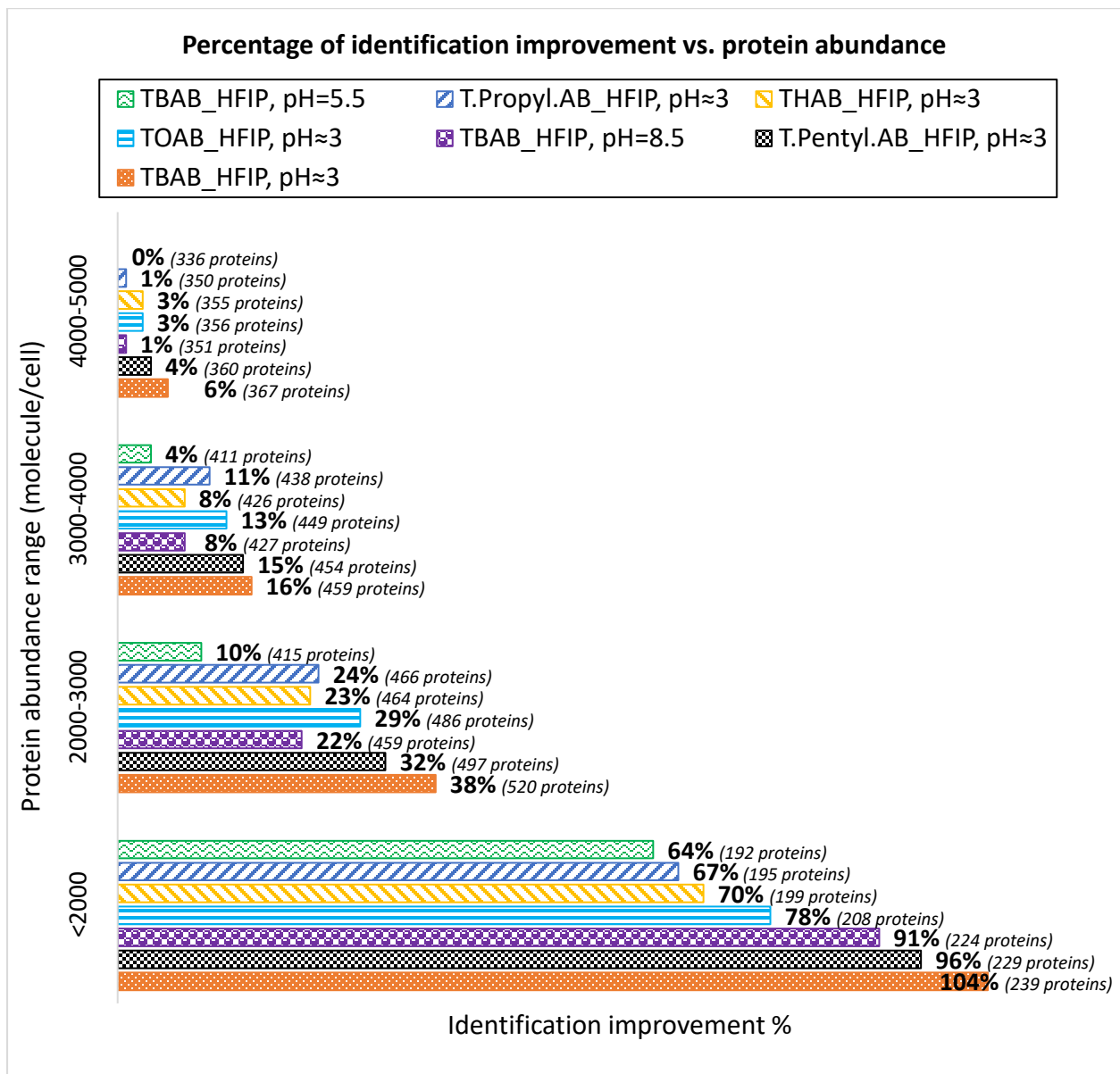
Figure 3-2. Identification improvement vs. abundance of proteins in different FA*i*C-BPS

3.3.2 SMALLER OVERLAP BETWEEN IDENTIFIED PROTEINS IN DIFFERENT FA*I*C-BPS AT LOWER ABUNDANCE RANGES

The identified proteins in four different FA*i*C-BPS, including TBAB_HFIP at pH≈3, T. Pentyl.AB_HFIP at pH≈3, TBAB_HFIP at pH=8.5, and TOAB_HFIP at pH≈3 was compared to each other and the results are shown in Figure 3-3(a). At lower abundance ranges, the overlap between the proteins in these FA*i*C-BPS is small; at the abundance range of less than 2000

(molecule/cell) only 19% of the proteins are common between these four systems. As the abundance of proteins increases, this percentage increases too, and at abundance range of 5000-7000 (molecule/cell), 88% of proteins are identified in all of above-mentioned systems. Therefore, at lower abundance rages, pooling the data from different systems would result in identification of larger number of unique low-abundance proteins. Figure 3-3(b) shows the Venn diagrams of proteins in the aqueous and amphiphile-rich phases of these four systems. Results shows that at lower abundance ranges, there is more fractionation of proteins between the two phases. As the abundance increases, the fractionation between the two phases decreases. Better identification improvement of FA*i*C-BPS at lower abundance ranges can be attributed to this fractionation patter.

Some of the most frequently used solubilizing reagents for proteins are urea, SDS, SDC, and SC, and each one has its own advantages and drawbacks. Urea as chaotropic compounds disrupts hydrophobic interactions and hydrogen bonds both between and within proteins.[17] SDS, as an anionic detergent with a long alkyl group, performs particularly well in solubilization of membrane proteins due to denaturing globular structure of proteins.[18] Unfortunately, small percentages of SDS inhibit trypsin activity dramatically, therefore needs several washing steps during sample preparation. SDC and SC, as common bile salts in proteomics studies, do not have the problem of decreasing trypsin activity, even up to 10%, and they can be easily removed before MS analysis by acidification.[19] Results of this study showed that by taking the advantage of FAiC-BPS, the identification of low-abundance proteins shows considerable improvement, regardless of the type of solubilizing reagent as the control.

To make a comparison between FA*i*C-BPS of this study and commonly used tryptic digestion approaches, four different controls with common solubilizing reagents were compared to the above-mentioned four FA*i*C-BPS of this study. A list of proteins, resulted from pooling the list of identified proteins in 4 controls with different solubilizing reagents, including urea, SDS, SDC,

was considered as the comprehensive control. Figure 3-3(c) compares the identification improvement of pooled FA*i*C-BPS versus the pooled control (the four FA*i*C-BPS and the four controls). Results show that pooled FA*i*C-BPS show better identification coverage compared to the pooled controls, especially at lower abundances. In addition to the fractionation of proteins between the two phases of FA*i*C-BPS, the enrichment of low abundance proteins in one of the phases would help to have better identification coverage for low-abundance proteins compared to different types of common solubilizing reagents. Moreover, the presence of fluoroalcohols would help to decrease tertiary structures that would results to improved digestion of proteins.
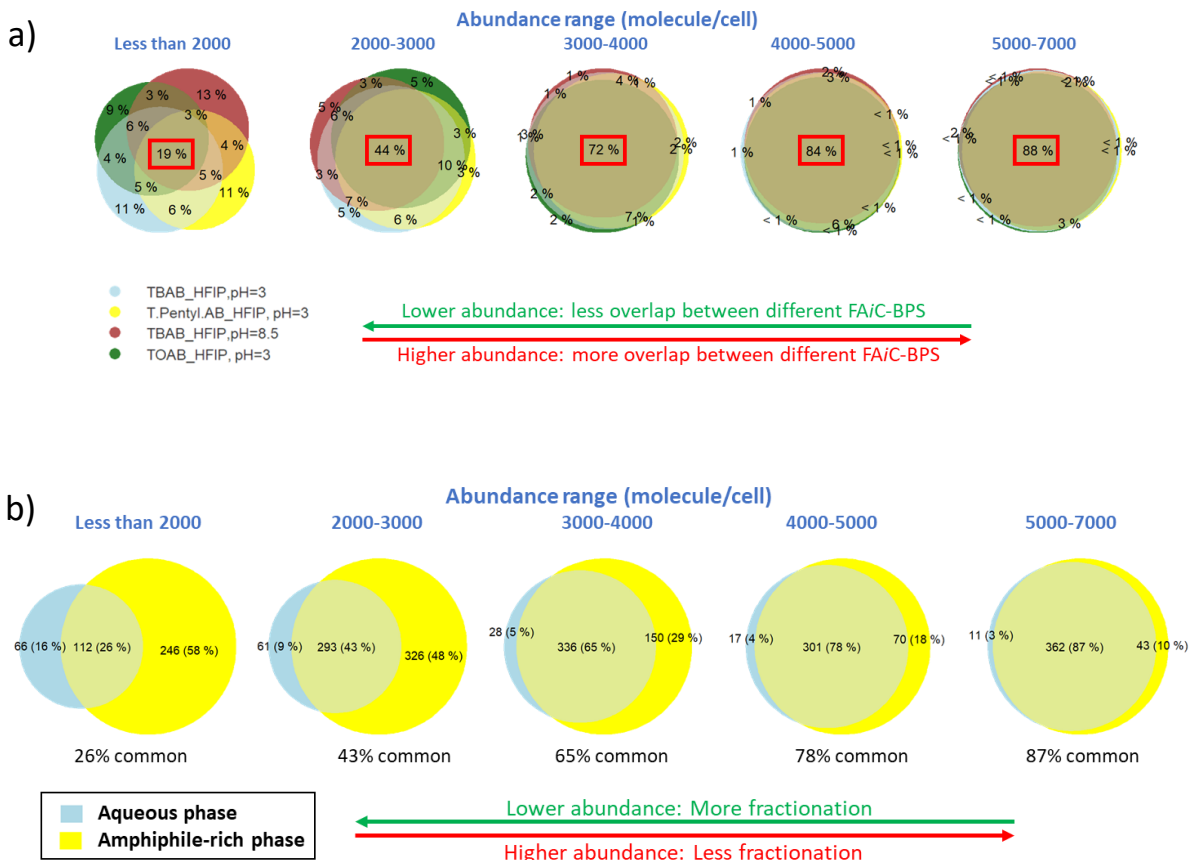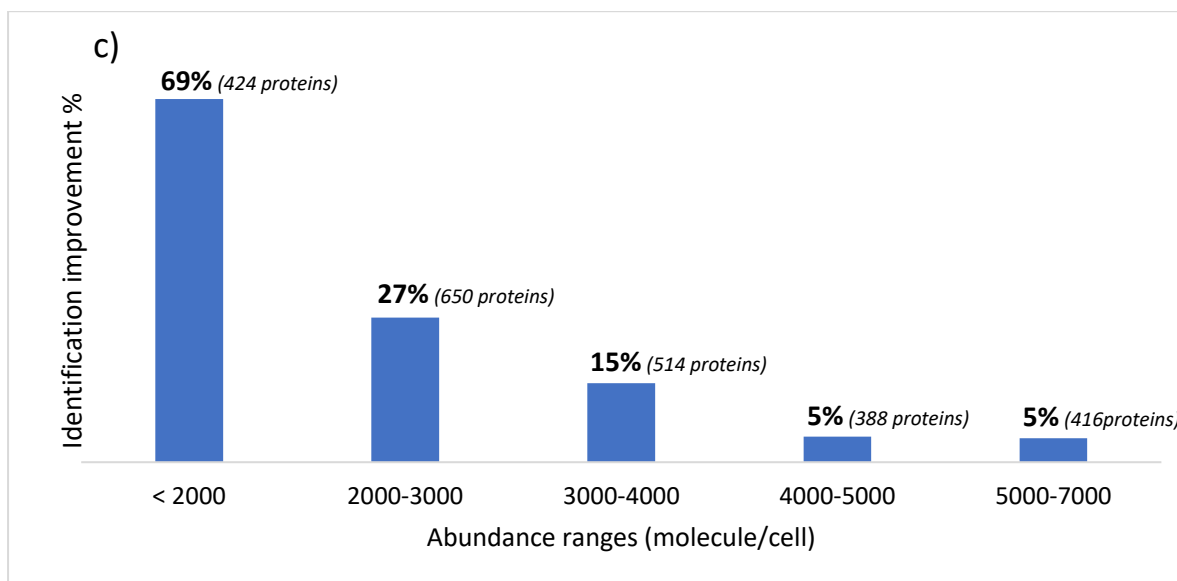
Figure 3-3. a) Venn diagrams that compare identified proteins in four different FA*i*C-BPS, including TBAB_HFIP at pH≈3, T. Pentyl.AB_HFIP at pH≈3, TBAB_HFIP at pH=8.5, and TOAB_HFIP at pH≈3; b) Venn diagrams that compare the proteins in aqueous and amphiphile-rich phases in the pooled list of FA*i*C-BPS of Figure 3-a; and c) percentage of identification improvement in the list of pooled FA*i*C-BPS at different abundance ranges compared to the pooled controls.
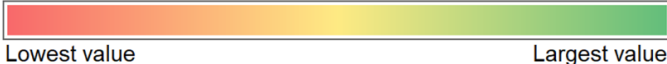
3.3.3 HIGHER FRACTIONATION BETWEEN THE PHASES IN LOW ABUNDANCE RANGES

In each FA*i*C-BPS and in any abundance range, some proteins are identified uniquely in either the aqueous phase or the organic phase, while some protein identified in both phases. Tables 3-2 compares the percentage of uniquely identified proteins (identified in either aqueous or organic phases) compared to the total number of proteins identified in an abundance range, and percentages are color-coded. In all FA*i*C-BPS, the uniquely identified proteins have the largest value for low-abundance proteins with abundance of less than 2000 molecule per cell. As the protein abundance in the cell increases, generally the percentage of uniquely identified proteins between two phases decreases. This means that at lower abundance ranges, larger number of proteins are enriched and identified in either the aqueous phases or organic phases. Further studies revealed that majority of these low-abundance proteins are enriched in the supramolecular/coacervate phases that have small volumes compared to the initial volume of solution. At the same time, amphiphile-rich

49

phases pose great solubility power due to the high concentration of amphiphiles and HFIP. The great solubilization power of amphiphile-rich phases, as well as their small volumes, makes these phases an ideal media for concomitant extraction, enrichment and solubilization of low abundance proteins that would result in enhanced identification of low abundance proteins.

Table 3-2. Percentage of uniquely identified proteins (identified in either aqueous or organic phases) at different abundance ranges.

FA*i*C-BPS	Abundance (molecule per cell)								
	<2000	2000-3000	3000-4000	4000-5000	5000-6000	6000-7000	7000-8000	8000-9000	9000-10000
T.Propyl.AB_HFIP, pH≈3	87%	71%	54%	33%	27%	19%	13%	19%	18%
TBAB_HFIP, pH≈3	87%	79%	59%	45%	34%	28%	29%	21%	23%
T.Pentyl.AB_HFIP, pH≈3	90%	81%	65%	50%	36%	30%	21%	19%	18%
THAB_HFIP, pH≈3	85%	79%	56%	38%	28%	16%	11%	14%	14%
TOAB_HFIP, pH≈3	84%	72%	56%	39%	30%	20%	18%	14%	21%
TBAB_HFIP, pH=5.5	92%	86%	79%	68%	62%	58%	55%	51%	49%
TBAB_HFIP, pH=8.5	88%	85%	76%	68%	59%	51%	42%	47%	41%
DTAB_HFIP, pH=5.5	74%	68%	60%	52%	46%	38%	40%	33%	38%

Lowest value   Largest value

3.3.4 CONTROLLING FRACTIONATION PATTERNS OF PROTEINS BASED ON PI BY MODULATING THE PH

When we look at specific groups of proteins with similar characteristics, we observe that affinity of specific protein groups toward the two phases changes by modulating the conditions of FA*i*C-BPS. One of the parameters to adjust is pH, which primarily affects the electrostatic interactions and consequently proteins distribution patterns based on isoelectric points. Previously published results from this groups shows that in the TBAB_HFIP system, basic proteins have a greater affinity toward the aqueous phase, and acidic proteins have larger affinity towards the coacervate phase. This phenomenon happens mostly because of the electrostatic interaction between proteins and the positively charged TBAB coacervate phases due to the high concentration of the QUATS in this phase. Since the system is very acidic; the proteins with higher pI values are positively charged, therefore they have little or no interaction with the coacervate phase and are repelled towards the aqueous phase. On the other hand, proteins with lower

isoelectric points are negatively charged, therefore they are attracted by the positive charge of QUATS in the coacervate phase. A good example of basic proteins is "mitochondrial ribosome" proteins, and generally they have large isoelectric points (negatively charged at acidic and neutral media).[20] In yeast, more than 94% of the mitochondrial ribosome proteins have isoelectric points of larger than 9. When FA*i*C-BPS are generated by addition of HFIP to the amphiphile solution of QUATS, the pH of aqueous phase is acidic (pH≈3 at 8% HFIP and 50 mM QUATS). Therefore, mitochondrial ribosome proteins with large isoelectric points are generally positively charged at low pHs, so they are repelled by the permanent positive charge of QUATS in the coacervate phase, as depicted in Figure 3-4.

By modulating the pH of the aqueous phase in the FA*i*C-BPS, the proteins distribution patterns can be controlled based on their pI. As shown in Figure 3-4, in acidic pH values (~3), mitochondrial ribosome proteins show more affinity towards the aqueous phase, but in higher pH values (more than 5.5) their affinity switches towards coacervate phase. Additionally, Figure 5 depicts the ditribution of isoelectric points of the proteins that are uniquely identified in the aqueous or coacervate phases of the TBAB_HFIP system at three different pHs. The resutls shows that as the pH increases, the affinity of basic proteins towards aqueous phase decreases. As shown in Figure 3-5, at pH of about 3, the median of isoelectric points of the proteins in the aqueous phase is 9.7, while the median in the cocarvet phase is 6.2. By increasing the pH of system, the median in the aqueous phase decreases from 9.7 at pH about 3, to 7.2 at pH=5.5, and 6.3 at pH=8.5. On the other hand, the variations in the median of isoelectric points of the proteins in the coacervate phases is almost negligible, from 6.2 at pH about 3, to 6.9 at pH=8.5. This shows that modulating the pH mostly affects the affinity of proteins towars the aqueous phase, while the hydrophobic effect plays a more dominant role in the interaction with the coacervate phases.
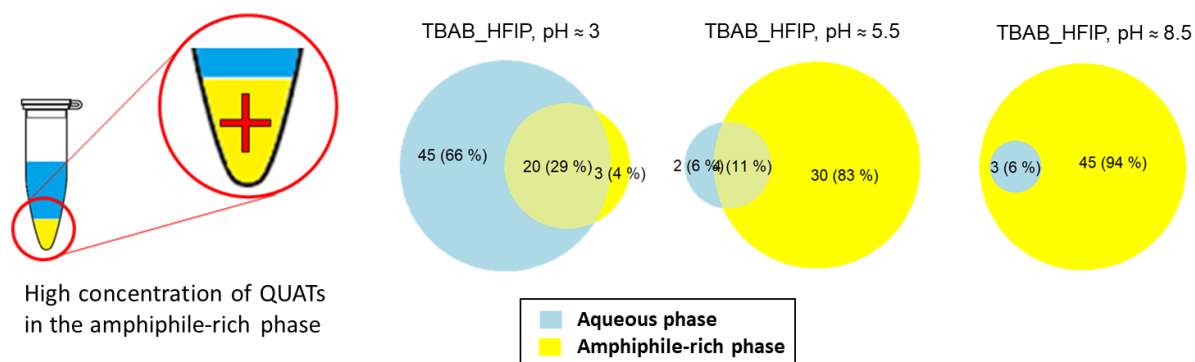
Figure 3-4. Switching affinity of mitochondrial ribosome proteins (with large isoelectric point of higher than 9) from the aqueous phases at low pHs to the coacervate phases at higher pHs
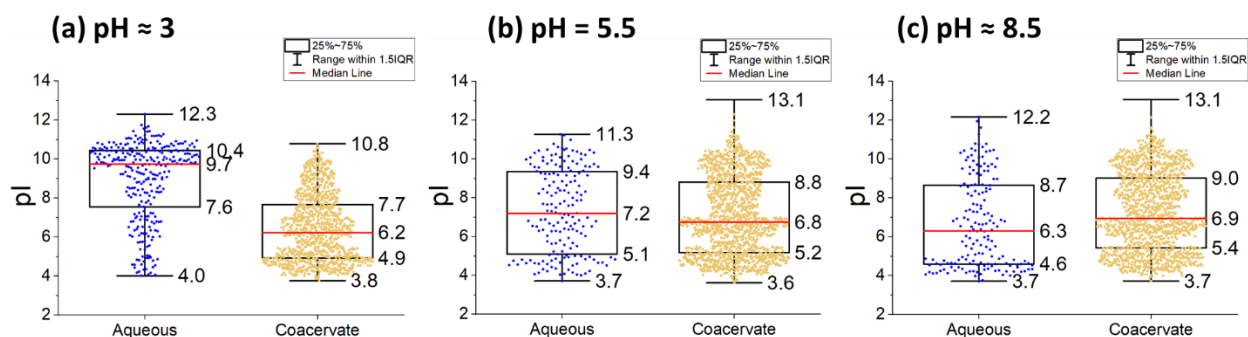
Figure 3-5. Distribution of isoelectric points of the proteins that are uniquely identified in the aqueous and coacervate phases of the TBAB_HFIP biphasic system at (a): pH about 3, (b): pH=5.5, and (c): pH=8.5

3.3.5 THE EFFECT OF CHAIN LENGTH OF ALKYL GROUPS OF QUATS ON ELECTROSTATIC INTERACTIONS

The electrostatic interaction between the positively charged head groups in the QUATS and proteins' charged groups would decrease as the QUATS alkyl chain lengths increase owing to the larger steric hindrance. Results showed that the aqueous phase of FA*i*C-BPS that have QUATS with shorter alkyl chains show more affinity towards basic proteins. Similarly, the coacervate phase of systems that have QUATS with shorter alkyl chain are more selective toward acidic proteins. As the alkyl chain length in the QUATS increases, the electrostatic interaction between the proteins and positively charged ammonium group decreases. Table 3-3 lists the percentage of

uniquely identified proteins in the aqueous phases of different FAiC-BPS in three different isoelectric point ranges of pI < 7, 7 ≤ pI < 9, and 9 ≥ pI, where percentages in each column are color coded and darker colors show larger numbers. Note the distinctly opposite trends in the number of unique proteins with pI<7 and those with pI>9 in systems with different chain lengths.

As shown in Table 3-3, more than 60% of the uniquely identified proteins in the aqueous phase of T.Propyl.AB_HFIP and TBAB_HFIP systems have isoelectric point of larger than 9. This percentage decreases systematically as the chain length increases. An opposite trend is observed for proteins with pI<7. The box charts in Figure 3-7 illustrates the distribution of pI of the uniquely identified proteins in both phases of FAiC-BPS at pH of 3. In T.Propyl.AB-HFIP and TBAB_HFIP systems, more than 75% of the protein in the aqueous phases have pI values of larger than 8 and 7.5, respectively. As the carbon number of alkyl group in the QUAT increases to pentyl, hexyl, and octyl groups, this value decreases to 6.7, 6.2, and 6, respectively. This corroborates that at a constant pH, as the alkyl group in QUATS increases, the electrostatic interaction between the proteins and the positive charge of QUATS in the organic phases decreases. However, this relationship is not linear due the presence of other types of interactions, especially hydrophobic effect. Figures 3-6 and 3-7 show that the electrostatic interaction between proteins and amphiphiles increases by decreasing the chain lengths of QUATS. Figure 3-6 depicts that in the systems that have QUATS with shorter chain lengths, mitochondrial ribosome proteins generally show more affinity towards the aqueous phase, which shows larger electrostatic interactions.

Results of this study show that this electrostatic interaction is stronger when the alkyl groups in the QUATS are shorter. As shown in Figure 3-6, in the FAiC-BPS with QUATS that have short alkyl chains, such as T.Propyl.AB and TBAB, majority of the mitochondrial ribosome proteins are uniquely identified only in the aqueous phase. In the T.Propyl.AB_HFIP and TBAB_HFIP system, 50% and 66% of the identified mitochondrial ribosome proteins in the aqueous phases are unique

to this phase (not identified in the organic phases), respectively. As the number of carbons in the alkyl group increases, this percentage decreases due to the larger steric hindrance. In the T.Pentyl.AB_HFIP, THAB_HFIP, and TOAB_HFIP systems, the percentage of uniquely identified mitochondrial ribosome proteins in the aqueous phases is 22%, 21%, and 11%, respectively. However, it is noteworthy that electrostatic interaction is not the only driving force in the FA*i*C-BPS, in the recently published results we have shown that hydrophobic effect is the other governing force in FA*i*C-BPS in addition to the electrostatic interaction.

Table 3-3. Percentage of proteins in different isoelectric point ranges in the aqueous phase of FA*i*C-BPS systems with different amphiphiles chain lengths.

QUAT	pI < 7	7 ≤ pI < 9	9 ≥ pI
T.Propyl.AB	20%	17%	63%
TBAB	21%	17%	62%
T.Pentyl.AB	29%	16%	55%
THAB	37%	17%	46%
TOAB	34%	20%	46%
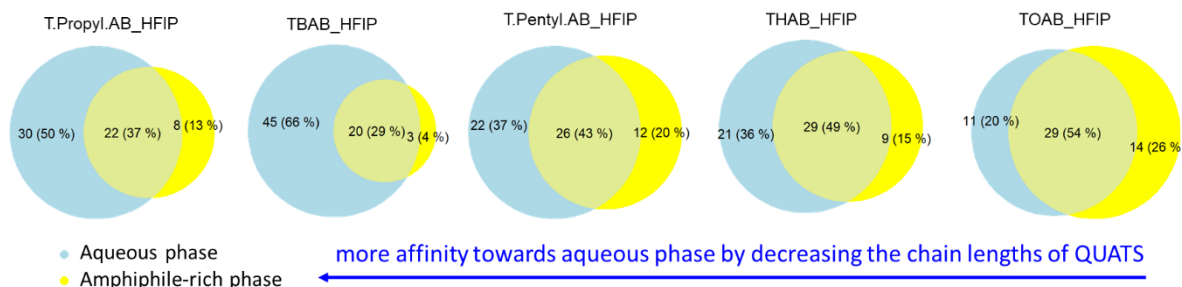
Figure 3-6. Larger affinity of mitochondrial ribosome proteins towards the aqueous phases as the chain lengths of QUATS decreases.
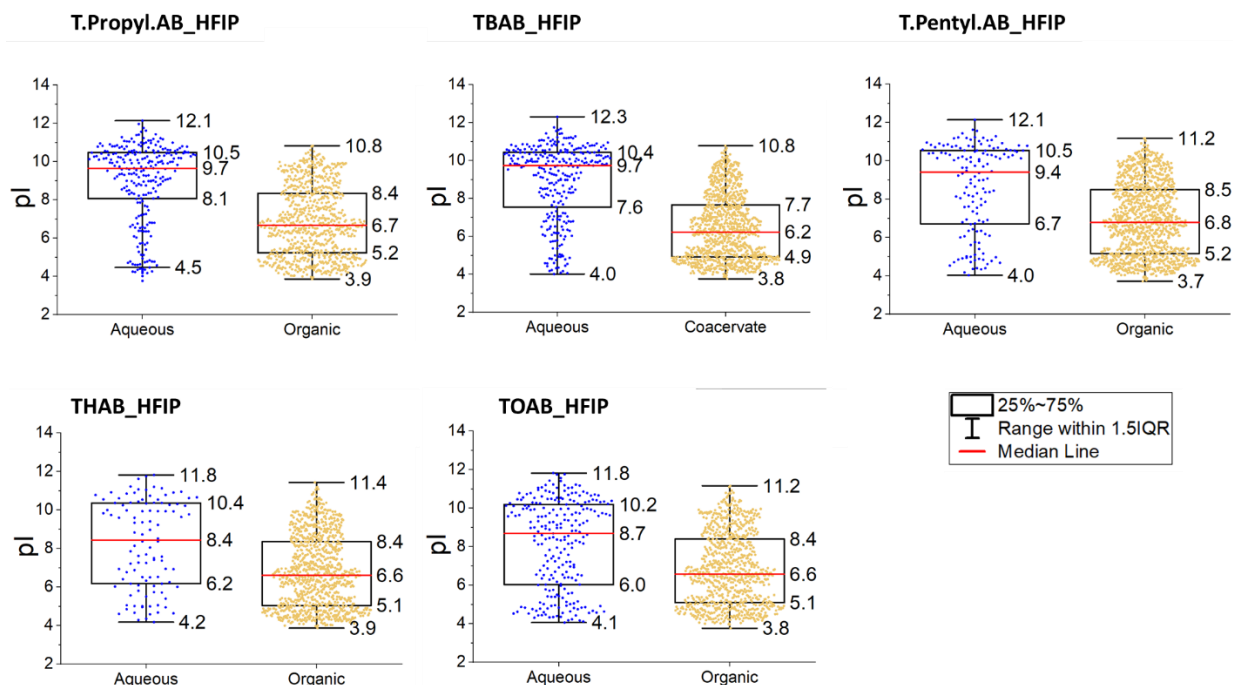
Figure 3-7. The distribution pattern of isoelectric points of the uniquely identified proteins in the two-phases of QUAT_HFIP induced biphasic systems at pH≈3, where QUATS have different alkyl chain lengths.

3.3.6 HYDROPHOBIC EFFECT ON PROTEINS DISTRIBUTION PATTERNS

In recent publications from this group, we have shown that generally more hydrophobic proteins have larger affinity towards coacervate phases.[5,6] As shown in the Appendix 3-2, the integral membrane protein, known as generally hydrophobic protein, show great affinity towards amphiphile-rich phases. This happens due to hydrophobic effect between proteins and coacervate phases, and the ability of amphiphile-rich phases to solubilize hydrophobic proteins. To compare the hydrophobic effect of different FA*i*C-BPS, their methylene selectivities were measured.

Methylene selectivity is in effect the partition coefficient of a methylene group from the aqueous phase into the coacervate phase and can be used as a relative measure of the coacervate phases hydrophobicity.[21] Theoretically, hydrophobicity has direct relationship with the QUATS chain lengths.[22] Methylene selectivity of each FA*i*C-BPS was determined from the slope of logarithm of partition coefficient of homologous series of alkylphenones, including acetophenone,

propiophenone, and butyrophenone versus the carbon number in the side chain of the alkylphenones. Comparing methylene selectivity of different FAiC-BPS (Table 3-4) shows that the coacervate phase of T.Propyl.AB_HFIP system is slightly smaller than other coacervates, however, there is no notable change in the hydrophobicity of coacervate phases after the number of carbons in alkyl groups of QUATS increases from C_4 to C_6. In other words, the hydrophobicity of the coacervate phases in FAiC-BPS does not significantly change by increasing the chain length of the QUATS alkyl groups. Additionally, as shown in Appendix 3-3, the distribution pattern of GRAVY values of the proteins in the two-phases of different FAiC-BPS do not show any tangible difference, which are in accordance with the results of methylene selectivity of different FAiC-BPS. In other words, more hydrophobic proteins with larger GRAVY value are extracted by the coacervate phases, however, neither changing the alky groups in QUATS nor changing the pH of system tangibly affect the affinity of hydrophobic proteins toward coacervate phases. In Figure 3-8, the distribution pattern of protein based on their GRAVY values and hydrophobicities is illustrated for two select system, and more details can be found in Appendix 3-3.

Table 3-4. Methylene selectivity in different FAiC-BPS.

FAiC-BPS	Number of carbons in the chain of QUATS	Methylene selectivity
T. Propyl AB_HFIP, pH≈3	3	0.30
TBAB_HFIP, pH≈3	4	0.36
T. Pentyl AB_HFIP, pH≈3	5	0.36
THAB_HFIP, pH≈3	6	0.37
TOAB_HFIP, pH≈3	8	0.37
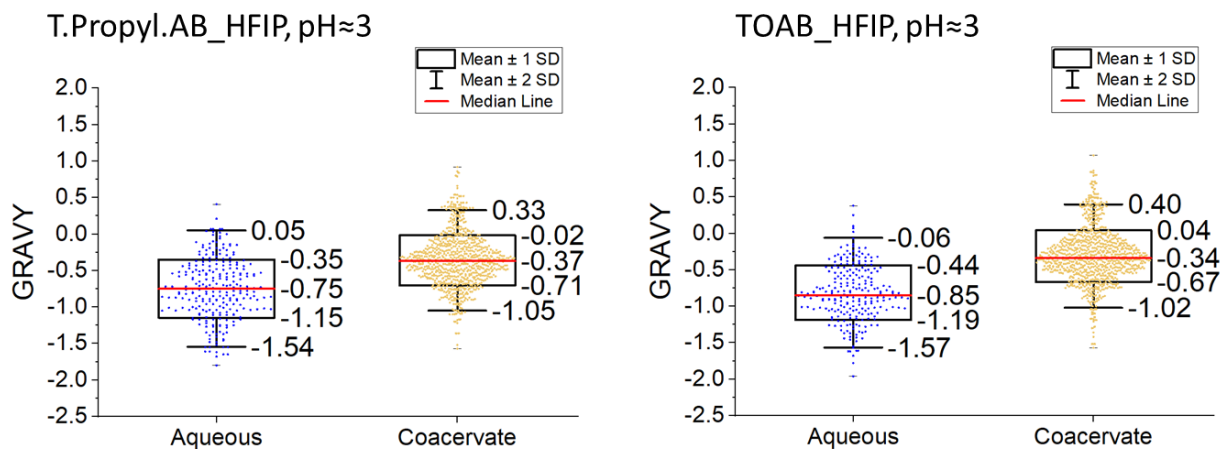
Figure 3-8. The distribution pattern of GRAVY values of the uniquely identified proteins in the two-phases of FA*i*C-BPS.

3.3.7 IMPROVED SEQUENCE COVERAGE OF TRANSMEMBRANE ALPHA-HELICES

Transmembrane proteins are consisting of three main domains, inner domain, outer domain and the transmembrane part that connects the first two parts. The transmembrane domain has alpha helical structure with a hydrophobic outer shell and traverses the lipid bilayer of the cell membrane.[23,24] In bottom-up proteomics approaches for identification of transmembrane proteins, proteins may be identified by their hydrophilic proteolytic peptides, but hydrophobic regions of membrane proteins are not identified in MS analysis[25] due to the low solubilization of α-helices in aqueous media. Better solubilization and coverage of α-helical transmembrane proteins can be used as an indication of efficiency of sample preparation techniques. To compare the efficiency of FA*i*C-BPS with the control, two measurements were taken into consideration, first, number of identified α-helical transmembrane proteins, and second, their sequence coverage.

Hidden Markov Model was used to make a database for α-helical transmembrane proteins (can be found at https://services.healthtech.dtu.dk/service.php?TMHMM-2.0).[26–28] We prepared the database of the sequence of α-helical segments of transmembrane proteins in yeast, and then

compared the identified sequences of our samples in mass spectrometry data with the sequences in the database. We developed an R program to measure the sequence coverage of α-helical segments of transmembrane proteins. The script of the R program and the instructions for using the script can be found in the Supplementary Information of the recently published article from this group.

Several studies have shown that fluoroalcohols, such as HFIP and trifluoroethanol (TFE), stabilize α-helical structures.[29–32] In addition, under low pH conditions, *e.g.* pH of 2, rapid loss of tertiary structure and increase in α-helices have been reported.[33] In FA*i*C-BPS, we take advantage of high concentrations of HFIP in the amphiphile-rich phases which can help to concomitantly stabilize and solubilize α-helical integral membrane proteins. In addition, the aqueous phases in this study contain about 7 to 7.5% (V/V) HFIP, and they have low pH values of about 3 due to the presence of HFIP, except in cases where the pH was buffered at higher values. The presence of HFIP can facilitate solubilization and identification of α-helical integral membrane proteins.

In the urea control sample, 436 integral membrane proteins (IMP) were detected, but only 14 of these IMP had detcted alpha-helical transmembrane peptides, and other proteins were detected by the peptides outside the membrane bilayer. As shown in Figure 9-a, in most FA*i*C-BPS, the number of IMP with detected alpha-helical transmembrane peptides are between 2 to 3 times larger than the control. For example, in TBAB_HFIP system, T.Pentyl.AB system, and THAB_HFIP system, the number of IMP with detected alpha-helical trasmembrane peptides are 44, 39, and 41, respectively. Interestingly, at pH≈3, the aqueous phases improvement in terms of detection of alpha-helical transmembrane peptides. Just in the aqueous phases of TBAB_HFIP and THAB_HFIP systems at pH≈3, respectively 29 and 25 protein had detected alpha-helical transmembranes peptides. This happens probably because of the better stability and solubilization α -helical structures due to the presence of about 7% (V/V) HFIP. As can be seen in Fig. 3-9b, no

α-helical transmembrane is identified in the aqueous phase of the TBAB_HFIP system at higher

pH values of 5.5 and 8.5 that contain buffer, while at pH 3, alpha helical peptides were identified

for 29 IMP.

In addition to the number of identified transmembrane α-helices, their sequence coverage is

also another parameter to assess the efficiency of FA*i*C-BPS in solubilization and identification of

α-helical transmembranes. Box charts of Figures 3-9-c and 3-9-d compare the sequence coverage

of the α-helical transmembranes in different FA*i*C-BPS, in the aqueous and amphiphile-rich

phases, respectively. Here, the sequence coverage is defined as the percentage of identified

aminoacids in the α-helical transmembrane segment by mass spectrometry data. The aqueous

phases of the TBAB_HFIP systems had the alpha-helical transmembrane sequence coverage, with

median of 44% coverage, compared to 9% in the control. The median of sequence coverage in the

aqueous phases of the T.Pentyl.AB_HFIP, T.Propyl.AB_HFIP, and THAB_HFIP were 35%, 26%,

and 23%, respectively. In it noteworthy to mention that the 100% coverage of transmembrane

alpha-helices was observed mostly in the aqueous phases. The sequence coverage in the aqueous

phases was generally higher at low pH. As opposed to the aqueous phases that showed better

sequence coverage at low pH values around 3, the amphiphile-rich phases showed better sequence

coverage of alpha-helical transmembranes at higher pH values of 5.5 and 8.5 in the aqueous

phases. It is noteworthy to mention that a previous study from this laboratory showed that using

mixed amphiphiles of a zwitterionic surfactant, 3- (N, N-Dimethyl myristyl ammonia) propane

sulfonate (DMMAPS), and a QUATS results in higher transmembrane α-helix coverage in FA*i*C-

BPS.

a) Identified trans membrane proteins in different FA*i*C-BPS at pH ≈ 3, different alkyl groups in QUATs

T.Propyl.AB_HFIP, pH ≈ 3   TBAB_HFIP, pH ≈ 3   T.Pentyl.AB_HFIP, pH ≈ 3

5 (21 %) 2 (8 %) 17 (71 %)   24 (55 %) 5 (11 %) 15 (34 %)   7 (18 %) 7 (18 %) 25 (64 %)

THAB_HFIP, pH ≈ 3   TOAB_HFIP, pH ≈ 3

17 (41 %) 8 (20 %) 16 (39 %)   2 (8 %) 4 (17 %) 18 (75 %)

b) Identified trans membrane proteins in FA*i*C-BPS at higher pHs

TBAB_HFIP, pH=5.5   TBAB_HFIP, pH=8.5

27 (100%)   33 (100%)

c) Sequence coverage of α-helical transmembranes in the aqueous phases

pH ≈ 3, no buffer added
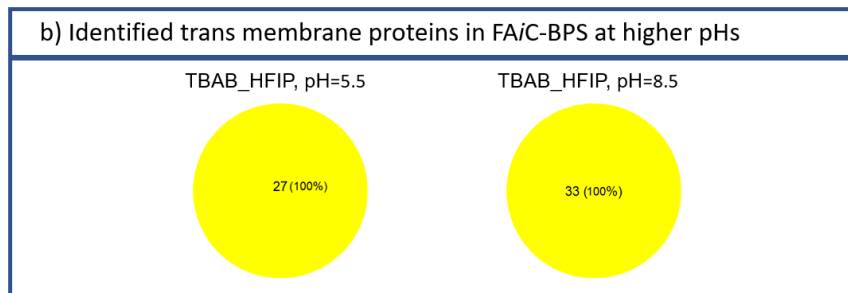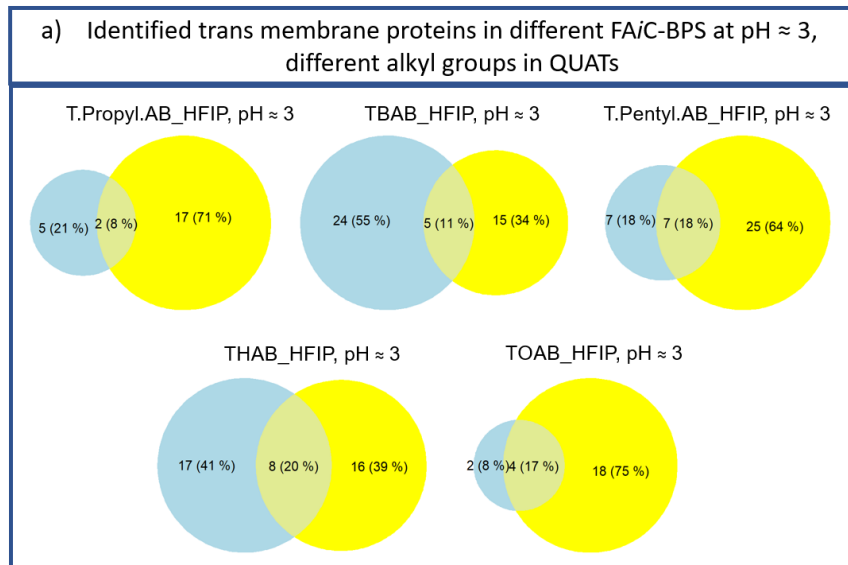
Legend:
25%~75%
Range within 1.5IQR
Median Line
Outliers
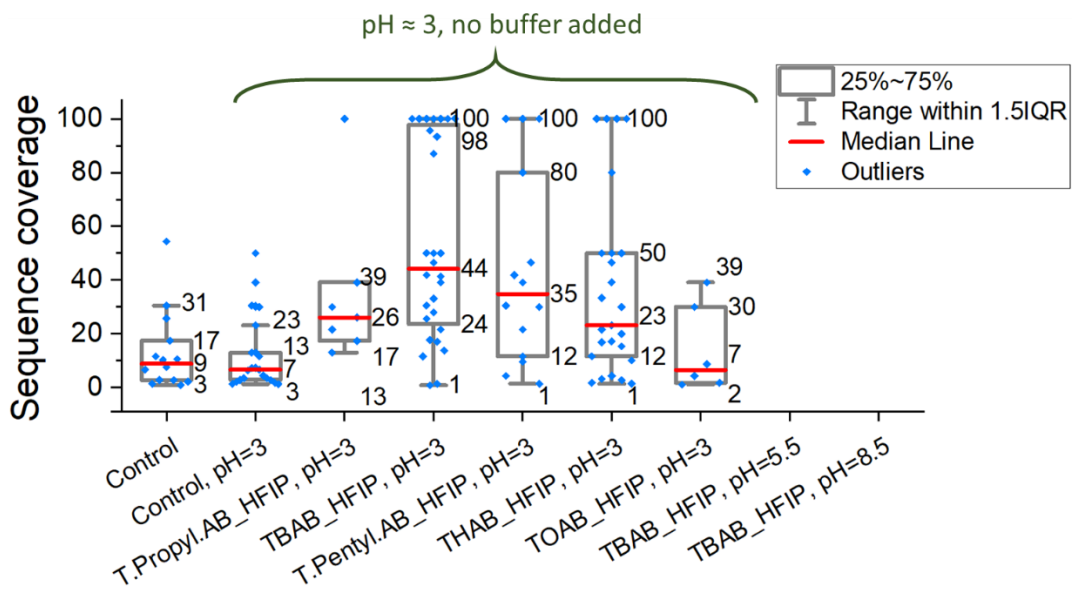
d) Sequence coverage of α-helical transmembranes in the amphiphile-rich phases
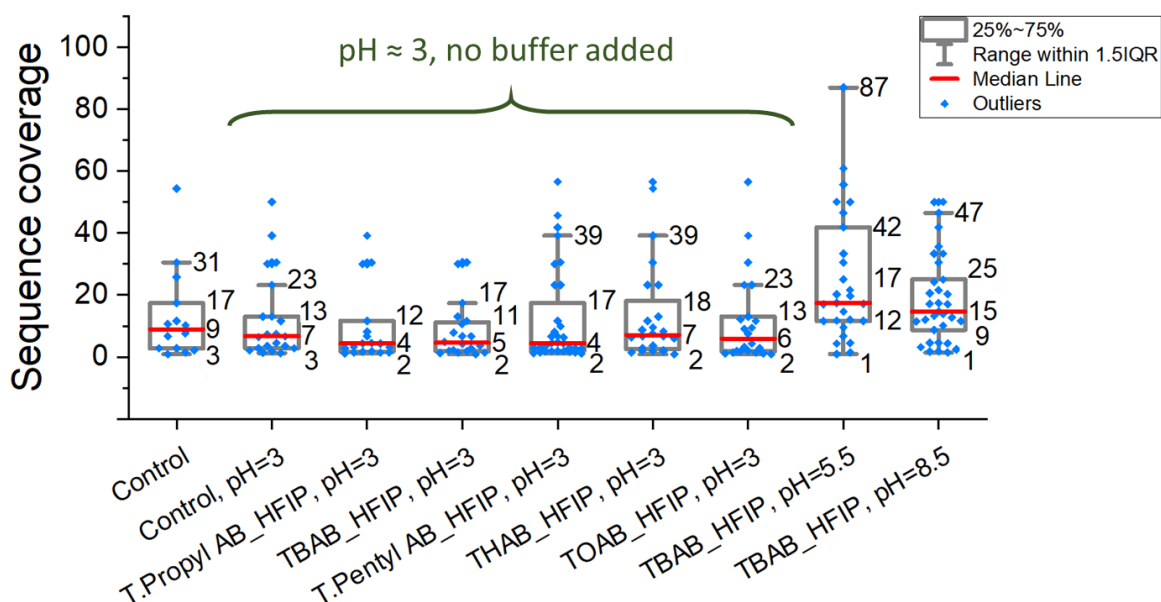
Figure 3-9. a) the number of integral membrane proteins with identified α-helical peptide segments in the two phases of different FA*i*C-BPS at pH≈3, b) the number of integral membrane proteins with identified α-helical segments in the two phases of different FA*i*C-BPS at pH=5.5 and pH=8.5, c) sequence coverage of the α-helical transmembranes in the aqueous phases of different FA*i*C-BPS, and d) sequence coverage of the α-helical transmembranes in the amphiphile-rich (coacervate) phases of different FA*i*C-BPS

3.4 CONCLUSIONS

By using FA*i*C-BPS in bottom-up proteomics, proteins were fractionated between the coacervate and aqueous phases based on their physicochemical properties, such as isoelectric points and hydrophilic/hydrophobic balance. Results showed that at lower abundance ranges, more fractionation happens in the FA*i*C-BPS, and low-abundance proteins are enriched in one of the phases. This enrichment results in improved identification of the low-abundance proteins. The trend of identification improvement versus abundance range shows that in all FA*i*C-BPS of this study, higher identification improvement happens at lower abundance ranges.

In addition to the hydrophobic effects in the coacervate phases due to the high concentration of QUATS and HFIP, the other driving force in the FA*i*C-BPS is electrostatic interaction between

the proteins and charged QUATS in the coacervate phases. By changing the pH of systems, protein net charge would change that would impact the electrostatic interaction, and subsequently proteins fractionation patterns. At pH of around 3, the majority of proteins at higher isoelectric points are positively charged and are extracted into the aqueous phase because of repulsion between the positively charged proteins and the positively charged QUATS in the coacervate phases. As the pH of system is modulated to higher values by addition of buffers, the tendency of basic proteins towards the aqueous phases decreases. Additionally, the results showed that QUATS with larger alkyl groups show less electrostatic interaction with proteins due to the higher steric hinderance.

With regards to alpha-helical transmembrane identification, generally most of the FA*i*C-BPS showed significantly better performance, in terms of both the number of transmembrane proteins with detected alpha-helical trasmembrane peptides and coverage of alpha-helical peptides. At acidic pH values (~ 3), the aqueous phases in the FA*i*C-BPS generally showed better alpha-helical transmembrane coverage. The aqueous phase of TBAB_HFIP system at pH of 3 showed significantly better transmembrane α-helix coverage as compared to the control and other FA*i*C-BPS, with up to 100% α-helical transmembranes coverage for specific proteins. At higher pH values of 5.5 and 8.5, only the coacervate phases showed α-helical transmembranes.

3.5 REFERENCES

(1)   Distler, A. M.; Kerner, J.; Peterman, S. M.; Hoppel, C. L. A Targeted Proteomic Approach for the Analysis of Rat Liver Mitochondrial Outer Membrane Proteins with Extensive Sequence Coverage. *Anal. Biochem.* **2006**, *356* (1), 18–29. https://doi.org/10.1016/j.ab.2006.03.053.

(2)   Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.; Chen, L.; Zhang, Y.; Wen, B.; He, F.; Xu, P. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *J. Proteome Res.* **2016**, *15* (11), 3988–3997. https://doi.org/10.1021/acs.jproteome.6b00390.

(3)   Zhang, Y.; Lin, Z.; Tan, Y.; Bu, F.; Hao, P.; Zhang, K.; Yang, H.; Liu, S.; Ren, Y. Exploration of Missing Proteins by a Combination Approach to Enrich the Low-Abundance Hydrophobic Proteins from Four Cancer Cell Lines. *J. Proteome Res.* **2020**, *19* (1), 401–408. https://doi.org/10.1021/acs.jproteome.9b00590.

(4)   Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography−Tandem Mass Spectrometry. *J. Proteome Res.* **2002**, *1* (4), 351–360. https://doi.org/10.1021/pr0255248.

(5)   Koolivand, A.; Azizi, M.; O'Brien, A.; Khaledi, M. G. Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies. *J. Proteome Res.* **2019**, *18* (4), 1595–1606. https://doi.org/10.1021/acs.jproteome.8b00857.

(6)   Koolivand, A.; Clayton, S.; Rion, H.; Oloumi, A.; O'Brien, A.; Khaledi, M. G. Fluoroalcohol - Induced Coacervates for Selective Enrichment and Extraction of Hydrophobic Proteins. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **2018**, *1083*, 180–188. https://doi.org/10.1016/j.jchromb.2018.03.004.

(7)   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.

(8)   Caballo, C.; Sicilia, M. D.; Rubio, S. Chapter 5 - Supramolecular Solvents for Green Chemistry. In *The Application of Green Solvents in Separation Processes*; Pena-Pereira, F., Tobiszewski, M., Eds.; Elsevier, 2017; pp 111–137. https://doi.org/10.1016/B978-0-12-805297-6.00005-X.

(9)   Jenkins, S. I.; Collins, C. M.; Khaledi, M. G. Perfluorinated Alcohols Induce Complex Coacervation in Mixed Surfactants. *Langmuir* **2016**, *32* (10), 2321–2330. https://doi.org/10.1021/acs.langmuir.5b04701.

(10)   Khaledi, M. G.; Jenkins, S. I.; Liang, S. Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles. *Langmuir* **2013**, *29* (8), 2458–2464. https://doi.org/10.1021/la303035h.

(11)   Nejati, M. M.; Khaledi, M. G. Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis. *Langmuir* **2015**, *31* (20), 5580–5589. https://doi.org/10.1021/acs.langmuir.5b00444.

(12)   Spruijt, E.; Sprakel, J.; Stuart, M. A. C.; Gucht, J. van der. Interfacial Tension between a Complex Coacervate Phase and Its Coexisting Aqueous Phase. *Soft Matter* **2010**, *6* (1), 172–178. https://doi.org/10.1039/B911541B.

(13)   Wang, Z.; Zhao, F.; Li, D. Determination of Solubilization of Phenol at Coacervate Phase of Cloud Point Extraction. *Colloids Surf. Physicochem. Eng. Asp.* **2003**, *216* (1), 207–214. https://doi.org/10.1016/S0927-7757(02)00560-5.

(14)   Khaledi, M. G.; Jenkins, S. I.; Liang, S. Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles. *Langmuir* **2013**, *29* (8), 2458–2464. https://doi.org/10.1021/la303035h.

(15)   Boyacı, E.; Pawliszyn, J. Micelle Assisted Thin-Film Solid Phase Microextraction: A New Approach for Determination of Quaternary Ammonium Compounds in Environmental Samples. *Anal. Chem.* **2014**, *86* (18), 8916–8921. https://doi.org/10.1021/ac5015673.

(16)   Rhodes, D.; Hanson, A. D. Quaternary Ammonium and Tertiary Sulfonium Compounds in Higher Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1993**, *44* (1), 357–384. https://doi.org/10.1146/annurev.pp.44.060193.002041.

(17)   Natarajan, S.; Xu, C.; Caperna, T. J.; Garrett, W. M. Comparison of Protein Solubilization Methods Suitable for Proteomic Analysis of Soybean Seed Proteins. *Anal. Biochem.* **2005**, *342* (2), 214–220. https://doi.org/10.1016/j.ab.2005.04.046.

(18)   Otzen, D. E. Protein Unfolding in Detergents: Effect of Micelle Structure, Ionic Strength, PH, and Temperature. *Biophys. J.* **2002**, *83* (4), 2219–2230. https://doi.org/10.1016/S0006-3495(02)73982-9.

(19)   Lin, Y.; Zhou, J.; Bi, D.; Chen, P.; Wang, X.; Liang, S. Sodium-Deoxycholate-Assisted Tryptic Digestion and Identification of Proteolytically Resistant Proteins. *Anal. Biochem.* **2008**, *377* (2), 259–266. https://doi.org/10.1016/j.ab.2008.03.009.

(20)   O'Brien, T. W. Properties of Human Mitochondrial Ribosomes. *IUBMB Life* **2003**, *55* (9), 505–513. https://doi.org/10.1080/15216540310001626610.

(21)   Bhagwat, V.; Bereznitski, Y.; Buszewski, B.; Jaroniec, M. Comparative Characterization of Octyl Bonded Phases Using Methylene Selectivity Data. *J. Liq. Chromatogr. Relat. Technol.* **1998**, *21* (7), 923–939. https://doi.org/10.1080/10826079808005859.

(22)   Tian, Z.; Meng, X.; Luo, Y.; Cao, S.; Zhao, G. A Novel Quaternary Ammonium Salts Derived from α-Amino Acids with Large Steric Hindrance Group and Its Application in Asymmetric Mannich Reaction. *Tetrahedron* **2020**, *76* (39), 131484. https://doi.org/10.1016/j.tet.2020.131484.

(23)   Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Membrane Proteins. *Mol. Biol. Cell 4th Ed.* **2002**.

(24)   Zhang, S.-Q.; Kulp, D. W.; Schramm, C. A.; Mravic, M.; Samish, I.; DeGrado, W. F. The Membrane- and Soluble-Protein Helix-Helix Interactome: Similar Geometry via Different Interactions. *Struct. Lond. Engl. 1993* **2015**, *23* (3), 527–541. https://doi.org/10.1016/j.str.2015.01.009.

(25)   Xiong, Y.; Chalmers, M. J.; Gao, F. P.; Cross, T. A.; Marshall, A. G. Identification of Mycobacterium Tuberculosis H37Rv Integral Membrane Proteins by One-Dimensional Gel Electrophoresis and Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry. *J. Proteome Res.* **2005**, *4* (3), 855–861. https://doi.org/10.1021/pr0500049.

(26)   Services https://services.healthtech.dtu.dk (accessed Dec 13, 2020).

(27)   Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes11Edited by F. Cohen. *J. Mol. Biol.* **2001**, *305* (3), 567–580. https://doi.org/10.1006/jmbi.2000.4315.

(28)   Käll, L.; Krogh, A.; Sonnhammer, E. L. L. Advantages of Combined Transmembrane Topology and Signal Peptide Prediction—the Phobius Web Server. *Nucleic Acids Res.* **2007**, *35* (suppl_2), W429–W432. https://doi.org/10.1093/nar/gkm256.

(29)   Hong, D.-P.; Hoshino, M.; Kuboi, R.; Goto, Y. Clustering of Fluorine-Substituted Alcohols as a Factor Responsible for Their Marked Effects on Proteins and Peptides. *J. Am. Chem. Soc.* **1999**, *121* (37), 8427–8433. https://doi.org/10.1021/ja990833t.

(30)   Roccatano, D.; Colombo, G.; Fioroni, M.; Mark, A. E. Mechanism by Which 2,2,2-Trifluoroethanol/Water Mixtures Stabilize Secondary-Structure Formation in Peptides: A Molecular Dynamics Study. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (19), 12179–12184. https://doi.org/10.1073/pnas.182199699.

(31)   Hirota, N.; Mizuno, K.; Goto, Y. Cooperative Alpha-Helix Formation of Beta-Lactoglobulin and Melittin Induced by Hexafluoroisopropanol. *Protein Sci. Publ. Protein Soc.* **1997**, *6* (2), 416–421. https://doi.org/10.1002/pro.5560060218.

(32)   Cammers-Goodwin, A.; Allen, T. J.; Oslick, S. L.; McClure, K. F.; Lee, J. H.; Kemp, D. S. Mechanism of Stabilization of Helical Conformations of Polypeptides by Water Containing Trifluoroethanol. *J. Am. Chem. Soc.* **1996**, *118* (13), 3082–3090. https://doi.org/10.1021/ja952900z.

(33)   Roccatano, D.; Fioroni, M.; Zacharias, M.; Colombo, G. Effect of Hexafluoroisopropanol Alcohol on the Structure of Melittin: A Molecular Dynamics Simulation Study. *Protein Sci. Publ. Protein Soc.* **2005**, *14* (10), 2582–2589. https://doi.org/10.1110/ps.051426605.

(34)   Fatima, S.; Ahmad, B.; Khan, R. H. Native-like Tertiary Structure in the Mucor Miehei Lipase Molten Globule State Obtained at Low PH. *IUBMB Life* **2007**, *59* (3), 179–186. https://doi.org/10.1080/15216540701335716.

(35)   Ku, T.; Lu, P.; Chan, C.; Wang, T.; Lai, S.; Lyu, P.; Hsiao, N. Predicting Melting Temperature Directly from Protein Sequences. *Comput. Biol. Chem.* **2009**, *33* (6), 445–450. https://doi.org/10.1016/j.compbiolchem.2009.10.002.

(36)   Bey, H.; Gtari, W.; Aschi, A.; Othman, T. Structure and Properties of Native and Unfolded Lysing Enzyme from T. Harzianum: Chemical and PH Denaturation. *Int. J. Biol. Macromol.* **2016**, *92*, 860–866. https://doi.org/10.1016/j.ijbiomac.2016.08.001.

CHAPTER 4

OPTIMIZATION OF FLUOROALCAHOL MEDIATED SUPRAMOLECULAR BIPHASIC SYSTEMS TO ENHANCE COVERAGE OF LOW ABUNDANCE AND HYDROPHOBIC PROTEINS: EFFECT OF AMPHIPHILES AND COACERVATORS

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

ABSTRACT

This article presents new classes of Supramolecular Biphasic Systems (S-BPS) and studies their application in enhanced identification and coverage of low-abundance and membrane proteins. Each S-BPS was composed of an amphiphile-rich phase with high concentration of fluoroalcohols and amphiphiles, plus an aqueous phase on the top. We took the advantage of S-BPS in bottom-up proteomics of Saccharomyces cerevisiae yeast as the proteins sample, and results were compared to routinely used solubilizing agents, such as urea, sodium dodecyl sulfate (SDS), and sodium deoxycholate (SDC). High solubilizing power of these S-BPS resulted in enhanced sequence coverage of alpha helical transmembranes, and identification improvement for hydrophobic proteins (*e.g.*, integral membrane proteins). All different S-BPS showed considerable improvement in identification of low abundance proteins, and interestingly, as the abundance of

proteins in the cell decreases, identification improvement showed more enhancement. In the biphasic system that decyltrimethylammonium bromide was used as amphiphile, identification of low-abundance proteins (less than 2000 molecule/cell) showed 150% improvement. Additionally, the effect of different parameters such as pH, type for surfactant, and fluoroalcohols with different hydrophobicities and hydrogen donor properties were studied. At pH=8.5, using cetrimonium bromide (CTAB) as amphiphile resulted in considerable sequence coverage enhancement in transmembrane alpha-helixes, and using tetraethylammonium bromide (TEAB) as amphiphile resulted in about 20% improvement in identification of hydrophobic integral membrane proteins.

KEYWORDS: Coacervate, Fluoroalcohols, Fractionation, Hydrophobic Proteins, Proteomics, Low Abundance Proteins.

4.1 INTRODUCTION

Membrane proteins have variety of vital roles for cell survival, such as transferring signals and nutrients between the internal cell components and the external ambient, enzymatic activities, adhesion to molecules, and involving immune responses [1–4]. Membrane proteins are target of more than 50% of therapeutic drugs [4–10]. Unhappily, membrane proteins generally have quite high hydrophobicity, low solubility in aqueous media, and they have propensity to form aggregates; therefore, their characterization is challenging [11]. Likewise, characterization of low abundance proteins poses technical challenges, in LC-MS/MS analysis they are generally underrepresented because proteins at higher concentrations cast a shadow over them [12]. In proteomics, the key role of effective sample preparation technique to achieve improved characterization of low abundance and membrane proteins cannot be taken for granted.

Recently, we have discovered and showed effectiveness of using new types of supramolecular biphasic systems (S-BPS) in sample preparation for bottom-up proteomics, these S-BPS were mediated by addition of fluoroalcohols to aqueous solution of amphiphiles [13–15]. Supramolecular biphasic system is a term to address water-immiscible supramolecular aggregates that form a second phase in aqueous media through a self-assembly process [16,17]. One of the main classes of these S-BPS is Fluoroalcohol-Induced Coacervate Biphasic Systems (FAiC-BPS) that was initially introduced by this laboratory [18]. Coacervation is known as the process self-assembly of amphiphiles that form a separate amphiphile-rich phase in aqueous media, this separate phase is called coacervate phase [19–22]. Addition of fluoroalcohols (they work as coacervators), such as hexafluoroisopropanol (HFIP), hexafluoro-2-methyl-isopropanol (HFMIP), and trifluoroethanol (TFE), to the aqueous solution of variety of amphiphiles induces formation of FAiC-BPS under a wide range of concentrations and pHs [21,22].

FAiC-BPS have unique properties that cannot be found in other coacervation systems. First, fluoroalcohols and amphiphiles enrich in the coacervate phase of FAiC-BPS, therefore the coacervate phases in these systems have a great solubilization power for hydrophobic compounds like integral membrane proteins. Second, the volume of coacervate phases of FAiC-BPS is generally a small percentage of the total volume of system, therefore using these systems for extraction purposes will simultaneously result in enrichment of compounds in the coacervate phase by up to about three orders of magnitude [13]. This characteristic is quite helpful in facilitating detection of trace elements, such as low-abundance proteins in proteomics applications. Third, we can take the advantage of FAiC-BPS for fractionation of complex mixtures; in the case of proteomics sample preparation, more hydrophobic proteins would be extracted into the coacervate phase while more hydrophilic proteins would remain in the aqueous phase. This feature is helpful in simply reducing the complexity of proteomes in early steps of sample treatment. Finally, FAiC-

BPS offer remarkable selectivity ranges that are comparable to those in solid phase extraction. This feature looks more striking when considering that FA*i*C-BPS is a kind of liquid-liquid extraction methods that offers large sample capacity with no sample loss; the issues of sample loss are routinely observed in solid phase extraction due to protein adsorption on solid phases.

Our lab showed the first implementation of FA*i*C-BPS in proteomics applications, we examined HFIP-induced coacervates by using three categories of amphiphiles with long alkyl groups: sodium dodecyl sulfate (SDS) as anionic surfactant, cetyltrimethylammonium bromide (CTAB) as cationic surfactant, and dimethylmyristylammoniopropane sulfonate (DMMAPS) as zwitterionic surfactant [13]. Results of sample preparation by taking the advantage of FA*i*C-BPS showed increased coverage compared to the control experiment, where the control was the routinely-used tryptic digestion method by using urea for proteome solubilization [13]. This increased coverage was particularly more notable for low abundance and hydrophobic proteins, such as membrane proteins. Afterwards, we investigated coacervate formation of tetrabutylammonium bromide (TBAB), induced by HFIP or TFE [14]. Opposed to surfactants with long alkyl chains, Quaternary Ammonium Salts (QUATS) with small alkyl chains, such as TBAB, do not form micelles by self-aggregation in aqueous media [14]. Consequently, long chain surfactants denature proteins and the interact with them via hydrophobic effect, on the other hand, the electrostatic interaction is more notable in the systems that contain QUATS with short-length chains such as TBAB.

The selectivity patterns in different FA*i*C-BPS appears to originate from a balance between hydrophobic effects and electrostatic interactions, that would be different from one system to another. Selectivity of coacervate phases of FA*i*C-BPS towards proteins can be modulated by judicious alteration of amphiphiles (cationic, anionic, or zwitterionic), such as changing the polar headgroups or altering chain length of hydrophobic tails of amphiphiles. Additionally, selectivity

can be controlled by changing the pH of aqueous phase that would result in alteration in charge of proteins and ionic strength of the aqueous phases. Finally, changing the fluoroalcohol (or fluoroacids), would result in having different polarities and hydrogen bonding properties of coacervator that can change the selectivity of phases in FA*i*C-BPS. The focus of this study is to investigate new FA*i*C-BPS systems induced with different coacervators and different amphiphiles at different pH values for proteomics applications. These new FA*i*C-BPS are compared mostly in terms of identification improvement for low abundance proteins and hydrophobic proteins, such as membrane proteins, and the sequence coverage of alpha-helical transmembranes.

4.2 MATERIALS AND METHODS

4.2.1 MATERIALS, CHEMICALS, AND REAGENTS

Tetrabutylammonium bromide (TBAB) was purchased from ACROS Organics. 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Chemical, USA. Hexafluoro-2-methylisopropanol (HFMIP) was purchased from Matrix Scientific. Tris HCl and Tris base were purchased from Sigma-Aldrich. Cetyl Trimethyl Ammonium Bromide (CTAB) was obtained from Amresco Inc. Dithiothreitol (DTT), Iodoacetamide (IAA) and Sequencing grade trypsin were purchased from Fisher BioReagents, Alfa Aesar™, and Promega Corporation, respectively. Dodecyltrimethylammonium Bromide (DTAB) was purchased from TCI America. LC-MS acetonitrile (ACN), isopropanol (IPA) and deionized water were provided from Fisher Chemical, USA. Centrifugal filter with pore size of 10 K and volume of 0.5 mL were provided from Amicon Ultra.

4.2.2 FORMATION OF DIFFERENT FA*i*C-BPS

Table 4-1 lists the ratio of constituents of different FA*i*C-BPS of this study. The pH of FA*i*C-BPS was adjusted by addition of Tris buffer (2.5 M, pH=8.5) to attain the required pH. For all samples, initially 400 µg proteins (in the case of this study: 50 µL × 8 µg/µL = 400 µg) from the

yeast cell lysate was added to the solution of amphiphiles to solubilize them, then required amount of Tris buffer (as mentioned in Table 4-1) was added to adjust pH, afterward coacervator was added and samples were mixed thoroughly to get to equilibrium state, finally samples were centrifugation at 10,000×g to form FA*i*C-BPS. It is noteworthy to mention that the details of cell culture and cell lysate preparation can be found in another recently published article from this lab.

Table 4-1. Detailed instructions of forming different Fluoroalcohol induced Coacervate-Biphasic Systems (FA*i*C-BPS). Volumes are based on total volume of 1 mL, and concentration of amphiphiles were kept constant at 50 mM.

FA*i*C-BPS	µL of cell lysate (protein concentration of 8 µg/µL)	Amphiphile	2.5 M Tris buffer, pH=8.5 (µL)	DI water (µL)	Coacervator (µL)
TEAB_HFIP, pH=5.5	50	50 mM (100 µL from 500 mM TEAB)	36	734	80 µL HFIP
TEAB_HFIP, pH=8.5	50	50 mM (100 µL from 500 mM TEAB)	400	370	80 µL HFIP
CTAB_HFIP, pH=5.5	50	50 mM (500 µL from 100 mM CTAB)	36	334	80 µL HFIP
CTAB_HFIP, pH=8.5	50	50 mM (500 µL from 100 mM CTAB)	370	0	80 µL HFIP
DTAB_HFIP, pH=5.5	50	50 mM (500 µL from 100 mM DTAB)	36	334	80 µL HFIP
TBAB_HFMIP, pH=3	50	50 mM (111 µL from 450 mM TBAB)	0	759	80 µL HFMIP
TBAB_HFMIP, pH=8.5	50	50 mM (111 µL from 450 mM TBAB)	400	359	80 µL HFMIP
TBAB_TFE, pH=8.5	50	50 mM (111 µL from 450 mM TBAB)	400	289	150 µL TFE

After formation of FA*i*C-BPS, the two phases were separated and subsequently prepared for sample clean-up prior to tryptic digestion. Filter Assisted Sample Preparation (FASP) method was used for the purpose of sample clean-up and subsequent on-filter digestion. It worth to mention that FASP filters are made of cellulose that would be dissolved by the coacervators, therefore, coacervators must be evaporated under nitrogen flow before loading the samples to the FASP filters. After evaporating the coacervators, protein samples were dissolved in the solution of and 2 M thiourea and 5 M urea in 70% isopropanol prior to loading on FASP filters. Finally, samples

were cleaned-up to wash surfactants, and proteins were tryptically digested on filter according to the procedure mentioned in a recently published paper by this group.

4.3 RESULTS AND DISCUSSIONS

4.3.1 IMPROVED IDENTIFICATION OF PROTEINS

The number of identified proteins in of the FAiC-BPS is listed in Table 4-2. These numbers are compared to the control, where in the control, 8 M urea was used for protein solubilization as the most common approach in sample preparation for proteomics. It is noteworthy to mention that we also tried other common solubilizing reagents for the control, such as sodium dodecyl sulfate (SDS), sodium deoxycholate (DSC), and sodium cholate (SC); however, the control with 8M urea resulted in identification of more proteins than other controls and it was selected as the reference control for this study. Figure 4-1 shows the distribution of proteins in the two phases of each of the FAiC-BPS and their identification improvement compared to the urea control.

Table 4-2. Comparing the number of identified proteins in different FAiC-BPS.

FAiC-BPS	Number of proteins in the amphiphile-rich phase	Number of proteins in the aqueous phase	Number of identified proteins after two-phase formation	Identification improvement
TEAB_HFIP, pH=8.5	2914	1354	3051	14.7%
TEAB_HFIP, pH=5.5	2918	1616	3013	13.3%
CTAB_HFIP, pH=8.5	2786	1688	3011	13.2%
CTAB_HFIP, pH=5.5	2722	1708	2989	12.4%
DTAB_HFIP, pH=5.5	2735	1868	2986	12.3%
TBAB_TFE, pH=8.5	2697	1307	2961	11.4%
TBAB_HFMIP, pH=3	2816	1934	2952	11.0%
TBAB_HFMIP, pH=8.5	2675	1917	2828	6.4%
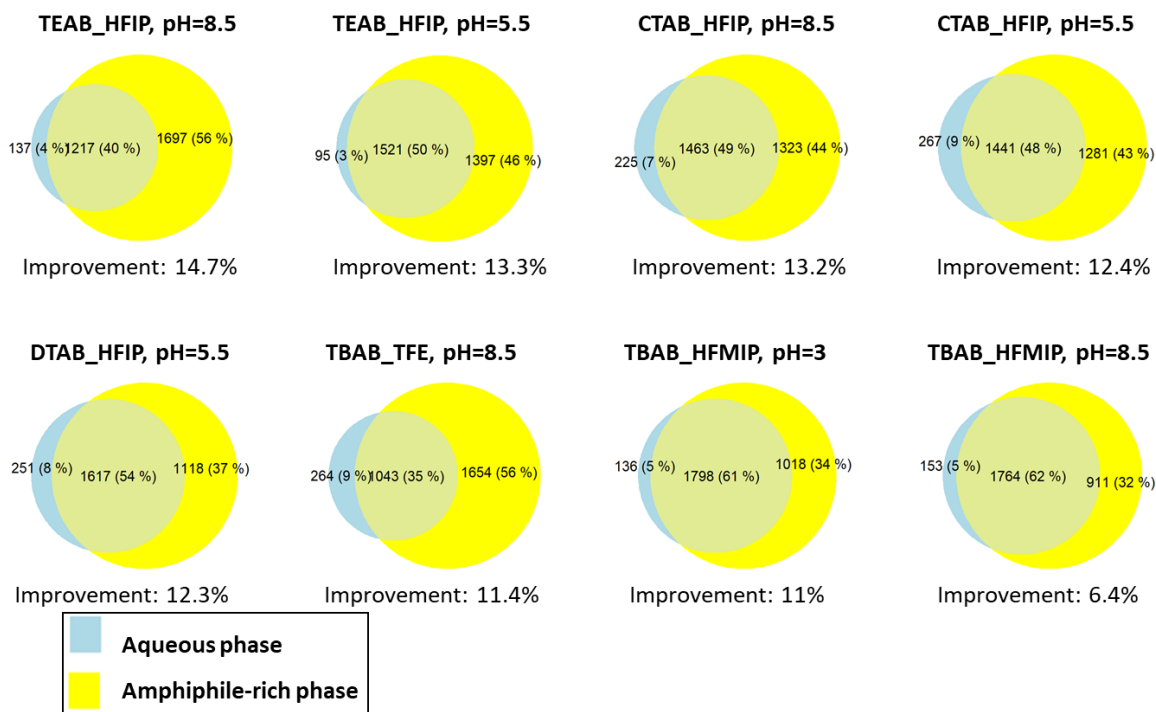
Figure 4-1. Venn diagram of identified proteins the two phases of different FA*i*C-BPS, and protein identification improvement of each FA*i*C-BPS compared to the urea control.

All the FA*i*C-BPS showed improved identification compared to the control. The largest identification improvement was observed for the systems that had TEAB as amphiphile and HFIP as the coacervator. These results are comparable to those that were recently published by this group, at which tetrabutylammonium bromide (TBAB) was used a QUATS with longer chain lengths. In the previously published results we have shown that in the FA*i*C-BPS at which QUATS are used as amphiphiles, electrostatic effect is a dominant interaction between the proteins and amphiphile-rich phases. Therefore, longer alkyl chain groups would cause greater steric hinderance. Comparison between the results of these two studies shows that at pH values of 5.5 and 8.5, TEAB as a QUATS with short chain length (ethyl) shows greater improvement than TBAB with longer chain length (butyl), probably due to less steric hinderance and greater electrostatic interaction between proteins and amphiphile-rich phase. At pH values of 5.5 and 8.5, TEAB_HFIP system shows 14.7% and 13.3% improvement, while the TBAB_HFIP system shows

4.2% and 9.0% improvement, respectively. It is noteworthy mentioning that the electrostatic interaction is not the only driving force in the FAiC-BPS, hydrophobic effect is the other interaction in these systems, and combination of these two forced mostly governs the interaction between proteins and the two phases of FAiC-BPS.

The FAiC-BPS originally have acidic pH values (around 3) due to the presence of fluoroalcohols, and pH can be adjusted to higher values by addition of buffers. It is noteworthy to mention that previous results showed that at pH value of 3, the identification improvement is greater than pH values of 5.5 and 8.5 mainly due to the effect of buffers at higher pHs. For example, the TBAB_HFIP system had shown improvement of 14.7% at pH=3, compared to 4.2% at pH=5.5 and 9.0% at pH=8.5. Unfortunately, when it comes to TEAB with shorter chain lengths, FAiC-BPS cannot be induced at pH=3 due to the slight hydrophobic effects. As the pH is increased, more HFIP would be deprotonated and the FAiC-BPS would start to form at pH>5 due to having deprotonated and anionic HFIP that can interaction with the positively charged amphiphile through electrostatic interactions.

4.3.2 SINIFICANT IDENTIFICATION IMPROVEMENT AT LOW ABUNDANCE RANGES

Interestingly, in all these FAiC-BPS, the greatest identification improvements were observed for low-abundance proteins. Figure 4-2 compares the identification improvement of proteins in different FAiC-BPS at different abundance ranges, and more details about the number of identified proteins in each abundance range can be found in Appendix 4-1. For proteins at very low abundance range of below 2000 molecules per cell, DTAB_HFIP system at pH=5.5 showed 150% improvement, which is slightly higher than other systems. In the FAiC-BPS that HFMIP is used as the coacervator, the identification improvement is lower than other systems, especially at pH=8.5, which is in accordance with the results of another study from this group that showed at

pH=3, identification improvement is greater than pH values of 5.5 and 8.5 owing to the buffer effect at higher pHs.
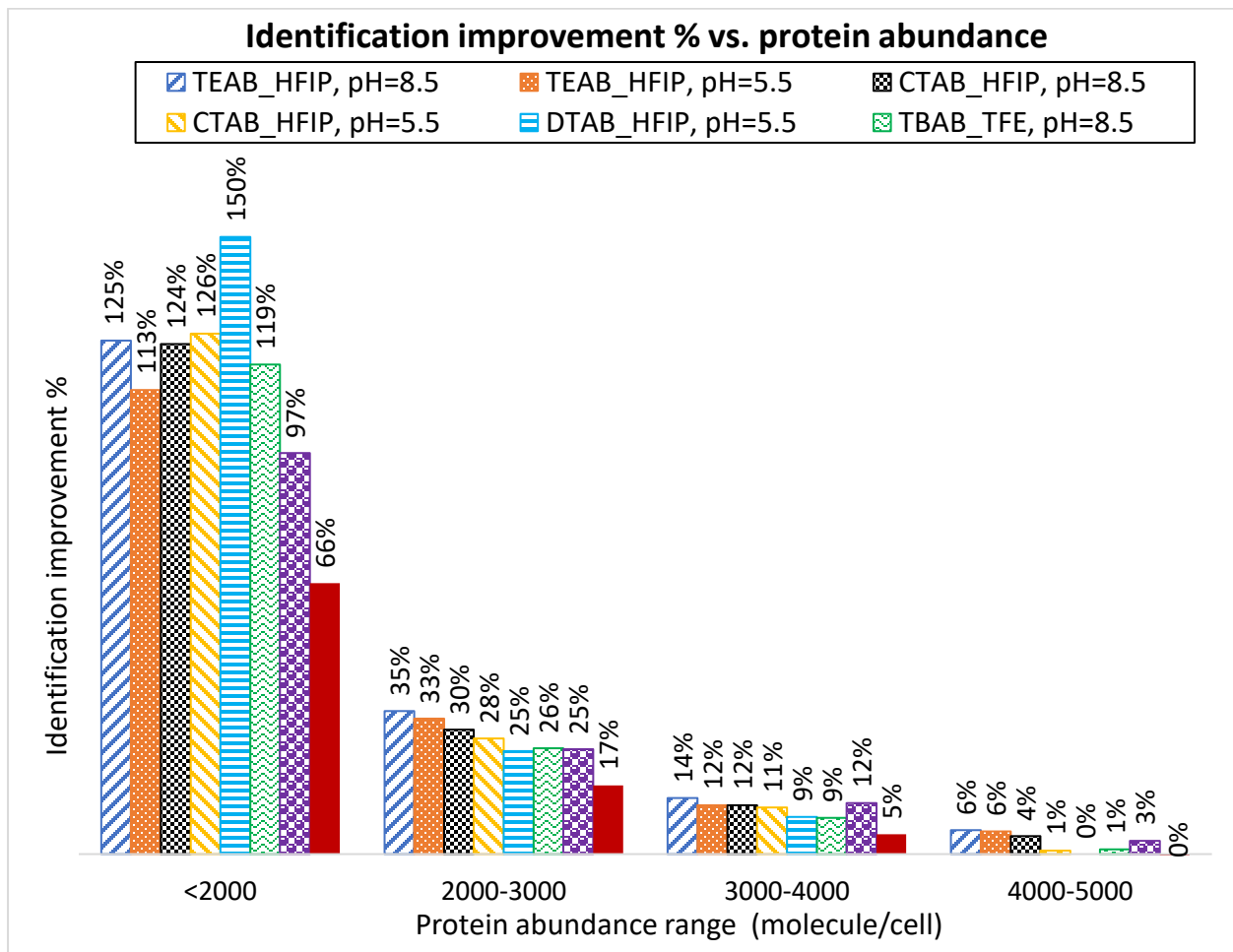
Figure 4-2. percentage of identification improvement in different FA*i*C-BPS vs. protein abundance ranges in the cell.

The main differences between systems with different amphiphiles, coacervators, and pH values are different levels of fractionation between the two phases and coverage of low abundance proteins that will be discussed in the following sections in detail.

4.3.3 SUPERIORITY OF FA*i*C-BPS IN TERMS OF LOW-ABUNDANCE PROTEIN IDENTIFICATION COMPARED TO COMMON SOLUBILIZING REAGENTS IN PROTEOMICS

The interactions between proteins and amphiphiles depend on the intrinsic properties of both, especially the charge of hydrophilic head group of the surfactant and the alkyl chain lengths[26]. Some of the most frequently used solubilizing reagents for proteins are urea, SDS, SDC, and SC, and each one has its own advantages and drawbacks. Urea as chaotropic compounds disrupts hydrophobic interactions and hydrogen bonds both between and within proteins [27]. SDS, as an anionic detergent with a long alkyl group, performs particularly well in solubilization of membrane proteins due to denaturing globular structure of proteins, however denatured proteins maintain a great degree of ordered structures, albeit non-native [26]. Unfortunately, small percentages of SDS inhibit trypsin activity dramatically, therefore needs several washing steps during sample preparation [28–32]. SDC and SC, as common bile salts in proteomics studies, do not have the problem of decreasing trypsin activity, even up to 10%, and they can be easily removed before MS analysis by acidification.[33] Results of this study showed that by taking the advantage of FA*i*C-BPS, the identification of low-abundance proteins shows considerable improvement, regardless of the type of solubilizing reagent as the control.

To make a fair comparison between the ability of FA*i*C-BPS and common sample preparation approaches, four different controls with common solubilizing agents were compared to four of the FA*i*C-BPS of this study. A list of proteins, resulted from pooling-up the list of identified proteins in 4 controls with different solubilizing agents, including urea, SDS, SDC, was considered as the comprehensive control. Similarly, a combined list of identified proteins in 4 FA*i*C-BPS, including DTAB_HFIP at pH=5.5, TEAB_HFIP at pH=8.5, CTAB_HFIP at pH=5.5, and TBAB_TFE at pH=8.5, was considered as pooled-up FA*i*C-BPS. Figure 4-3 compares the identification

improvement of these pooled-up FA*i*C-BPS versus the comprehensive control. In addition to the fractionation of proteins between the two phases of FA*i*C-BPS, the enrichment of low abundance proteins in one of the phases would help to have better identification coverage for low-abundance proteins compared to different types of common solubilizing reagents. However, utilization of different amphiphiles and different coacervators in FA*i*C-BPS would result in different interactions between proteins and the FA*i*C-BPS that causes various fractionation patterns and different identification improvement of low abundance proteins.
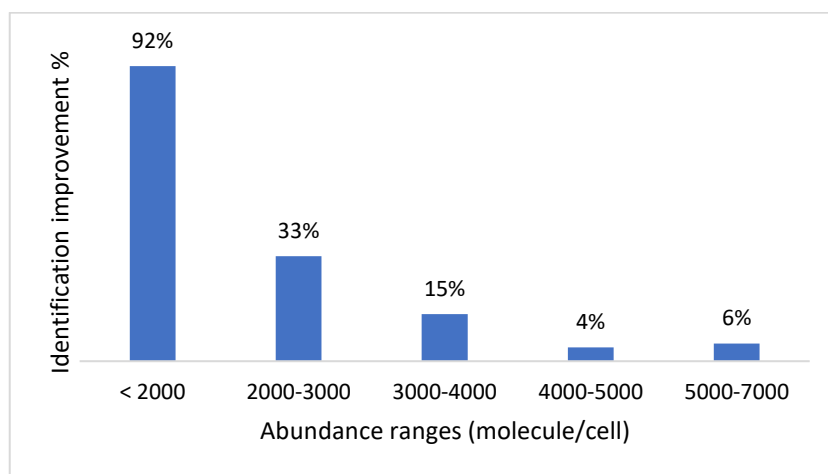
Figure 4-3. protein identification improvement in pooled-up list of 4 FA*i*C-BPS (DTAB_HFIP at pH=5.5, TEAB_HFIP at pH=8.5, CTAB_HFIP at pH=5.5, and TBAB_TFE at pH=8.5) versus pooled-up list of 4 controls with common solubilizing reagents including urea, SDS, SDC, and SC

4.3.4 MORE FRACTIONATION AND ENRICHMENT AT LOWER ABUNDANCE RANGES

The identified proitens in four different FAiC-BPS, including DTAB_HFIP at pH=5.5, TEAB_HFIP at pH=8.5, CTAB_HFIP at pH=5.5, and TBAB_TFE at pH=8.5 were compared to each other and the results are shown in Figure 4-4-a. At lower abundance ranges, the overlap between the proteins in these FAiC-BPS is small; at the abundance range of less than 2000 (molecule/cell) only 22% of the proteins are common between these four systems. As the abundance of proteins increases, this percentage increases too, and at abundance range of 5000-

77

7000 (molecule/cell), 90% of proteins are identified in all of above-mentioned systems. Therefore, at lower abundance rages, pooling the data from different systems would result in identification of larger number of unique low-abundance proteins. Figure 4-3-b shows the Venn diagrams of proteins in the aqueous and amphiphile-rich phases of these four systems. Results shows that at lower abundance ranges, there is more fractionation of proteins between the two phases. As the abundance increases, the fractionation between the two phases decreases. Better identification improvement of FAiC-BPS at lower abundance ranges can be attributed to this fractionation patter. Table 4-3 lists the percentage of proteins that are identified in either aqueous or amphiphile-rich phases at different abundance ranges. These percentages are color-coded, green colors mean more fractionation between the phases and enrichment of proteins in one phase, whereas red colors mean less fractionation. In all FAiC-BPS, more fractionation and enrichment were observed for proteins at lower abundances.
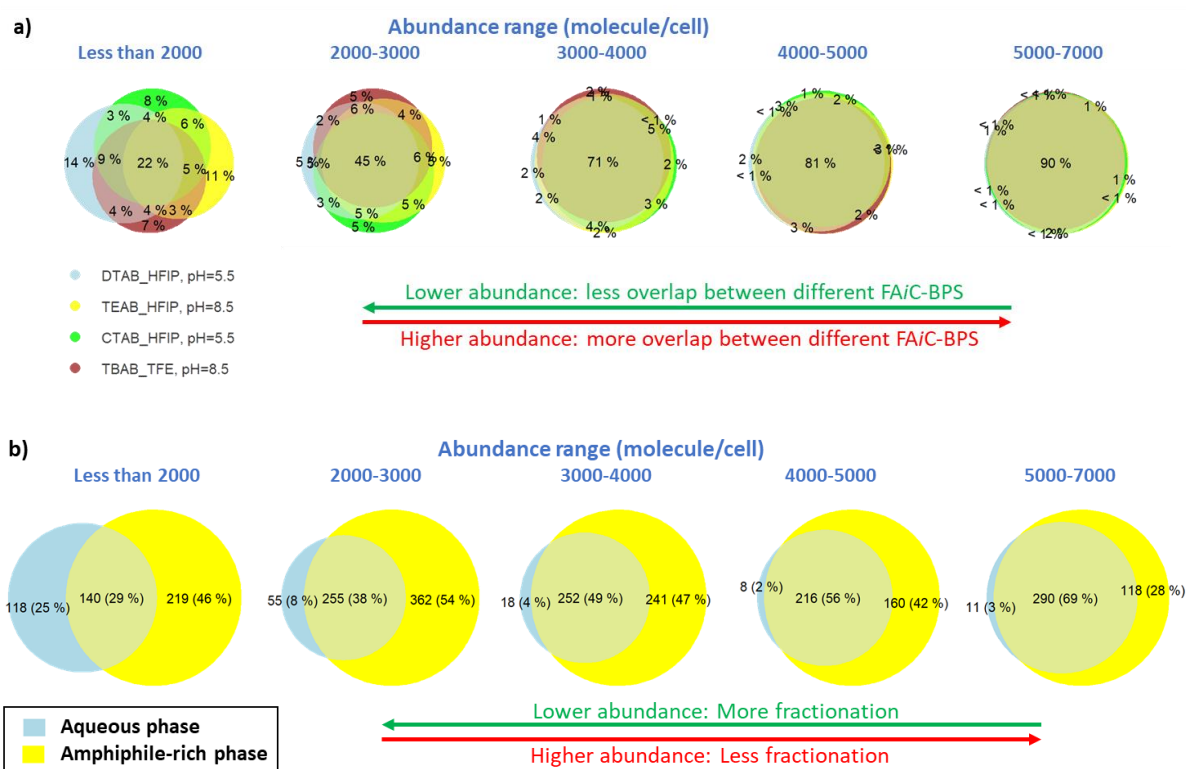
Figure 4-4. a) Venn diagrams that compare identified proteins in four different FAiC-BPS, including DTAB_HFIP at pH=5.5, TEAB_HFIP at pH=8.5, CTAB_HFIP at pH=5.5, and TBAB_TFE at pH=8.5;

b) Venn diagrams that compare the proteins in aqueous and amphiphile-rich phases in the pooled list of FAiC-BPS of Figure 3-a

Table 4-3. percentage of proteins that are identified in either aqueous or amphiphile-rich phases at different abundance ranges. More greenish colors represent more fractionation between the phases and enrichment of proteins in one phase, whereas more reddish colors represent less fractionation.

FA*i*C-BPS	Abundance range (molecule per cell)								
	<2 K	2-3 K	3-4 K	4-5 K	5-6 K	6-7 K	7-8 K	8-9 K	9-10 K
TEAB_HFIP, pH=8.5	90%	86%	78%	71%	62%	50%	50%	51%	49%
TEAB_HFIP, pH=5.5	90%	83%	70%	57%	43%	35%	31%	28%	29%
CTAB_HFIP, pH=8.5	84%	76%	69%	58%	55%	46%	44%	40%	40%
CTAB_HFIP, pH=5.5	86%	77%	70%	59%	58%	44%	39%	37%	44%
DTAB_HFIP, pH=5.5	74%	68%	60%	52%	46%	38%	40%	33%	38%
TBAB_TFE, pH=8.5	90%	78%	74%	69%	64%	62%	70%	59%	63%
TBAB_HFMIP, pH=3	88%	77%	55%	41%	32%	19%	16%	11%	18%
TBAB_HFMIP, pH=8.5	85%	75%	55%	34%	38%	25%	19%	22%	17%

Less fractionation   More fractionation and enrichment

In the TBAB_HFMIP, at both pH values of 3 and 8.5, smallest fractionations were observed compared to other systems (more reddish color at majority of abundance ranges). Consequently, the smallest protein identification improvement was observed in these two systems at which HFMIP was used as the coacervator. Previously we have shown that in the FAiC-BPS at which QUATS are used as amphiphiles, the main driving forces are electrostatic interactions and hydrophobic effects. A good balance between these two interactions would result in better protein fractionation and enrichment. When a very hydrophobic coacervator such as HFMIP is used, the hydrophobic effect outweighs the electrostatic interaction and proteins do not fractionate based on their isoelectric point.

In contrast, the TBAB_TFE system at pH=8.5, shows more fractionation compared to other FA*i*C-BPS of this study (more greenish color at majority of abundance ranges), however, this system did not show the largest identification improvement, probably due to smaller hydrophobicity of TFE compared to HFIP. A plausible justification for having less identification improvement while

having more fractionation would be the ability of TFE to solubilize proteins and break tertiary structures compared to HFIP. It worth mentioning that TFE does not induce two-phase system with TBAB at acidic pH values and in absence of salts. In this case, only deprotonated and ionized TFE at basic pHs would induce the two-phase system. Therefore, we cannot take the advantage of reduced tertiary structures at acidic pH values of about 2-3.

4.3.5 THE EFFECT OF COACERVATOR ON ELECTROSTATIC INTRACTIONS AND HYDROPHOBIC EFFECTS

To study the role of coacervator in fractionation pattern of proteins, TFE, HFIP, and HFMIP were used as three different coacervators in the same condition of 50 mM TBAB as amphiphile. Between HFIP and HFMIP, it has been shown that HFIP and HFMIP have a similar polar surface area, whereas HFMIP has substantially larger overall surface area than HFIP. Hence, HFMIP is more hydrophobic due to having a larger nonpolar area [34]. Between HFIP and TFE, it has been observed that HFIP, with six fluorinated atoms, performs as a stronger α-helical inducer compared to TFE.[35] HFIP with pK_a of 9.4 is more acidic than TFE with pK_a of 12.4.[34] Because of the presence of two –CF3 groups, HFIP shows greater H-bond donor characteristics than TFE. Therefore, HFIP should potentially show a better performance than TFE in perturbing interactions in proteins that are H-bonding or hydrophobic. It is noteworthy to mention that previous studies have shown that the equimolal mixture of TFE and HFIP is a stronger alpha-helical stabilizer than any of those.[35]

As shown in Figure 4-5(a), the results of a concurrent study from this group showed that electrostatic interaction have a dominant role in the TBAB-HFIP system, and by changing the pH in the aqueous phase of this system, alterations in electrostatic interactions will be happened which affects the distribution pattern of proteins based on their isoelectric points. In the HFIP-TBAB system at pH=3, the isoelectric points of the proteins in the aqueous phase had median value of about 3.5 greater than those in the amphiphile-rich phase; by changing the pH from 3 to 8.5, the

median value in the aqueous phase changes to be about 0.5 smaller than those in the amphiphile-rich phase. As shown in Figure 4-5(b), the results of this study show that by using a more hydrophobic coacervator, such as HFMIP, the hydrophobic interactions would outweigh the electrostatic interactions, and consequently, altering the pH of solution does not affect the fractionation of proteins based on their pI value and charges.

Figure 4-5(c) compares the TBAB_HFMIP and TBAB_TFE systems (at pH=8.5) in terms of the distribution pattern of hydrophobicities and GRAVY values of proteins in the two phases. Interestingly, HFMIP is more hydrophobic than TFE, but TFE shows better fractionation in terms of hydrophobicity. In the TBAB_HFMIP system, proteins in the amphiphile-rich phase had only slightly larger hydrophobicity, with median GRAVY value of -0.40 compared to -0.64 in the aqueous phase (0.24 difference). In the TBAB_TFE system, the difference between GRAVY values of proteins in the two phases is more notable, with median of -0.83 in the aqueous phase versus -0.34 in the amphiphile-rich phase (0.49 difference). One plausible justification would be the balance between electrostatic interactions and hydrophobic interactions. Previously published results from this group [cite] indicates that in the FA*i*C-BPS at which QUATS with short chain lengths, such as TBAB, are used as amphiphiles, electrostatic interaction shows a dominant role in fractionation of proteins due to the net positive charge in the amphiphile-rich phases. In the TBAB_TFE at pH=8.5, the coacervator (TFE) has much slighter hydrophobicity than HFIMIP. Therefore, in this system we can take the advantage of having both hydrophobic and electrostatic interactions that would cause better fractionation based on hydrophobicity. The distribution of pI and GRAVY values of proteins in different FAiC-BPS can be found in Appendix 4-2.
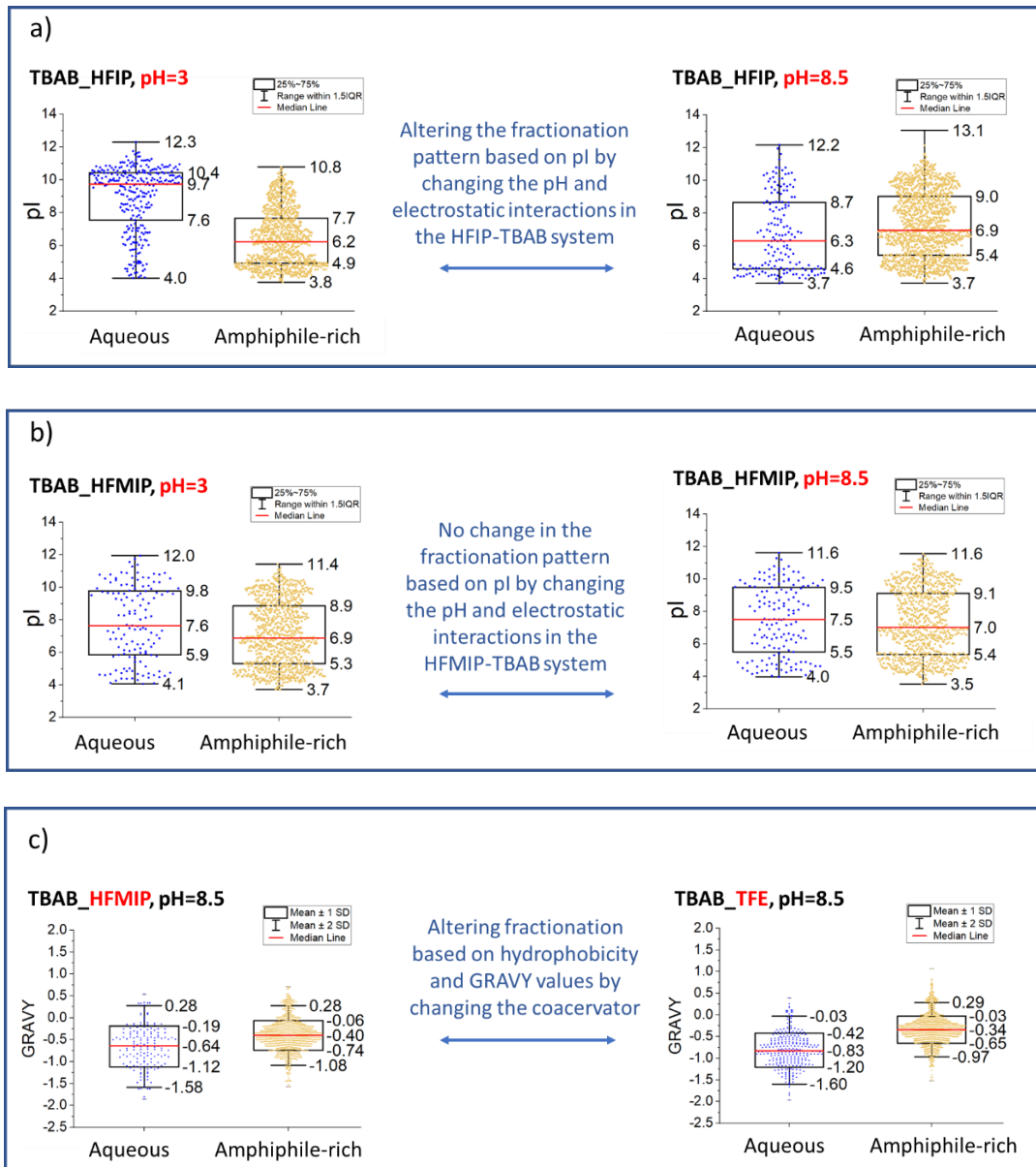
Figure 4-5. a) fractionation pattern of proteins based on their isoelectric points in the TBAB-HFIP system at pH values of 3 and 8.5 , b) fractionation pattern of proteins based their isoelectric points in the TBAB-HFMIP system at pH values of 3 and 8.5, c) fractionation pattern of proteins based on hydrophobicity in TBAB_HFMIP and TBAB-TFE systems at pH=8.5.

4.3.6 IMPROVED SEQUENCE COVERGE OF TRANSMEMBRANE ALPHA-HELICES

The part of transmembrane proteins that passes through the lipid bilayer of the cell is called transmembrane domain. This segment usually forms an alpha-helical structure or beta-sheet barrel shape, and among these two, the alpha-helical structure is the most dominant one. The outer side

of alpha-helical structure is hydrophobic that facilitates interaction between the protein and hydrophobic moiety of lipid bilayer [36,37]. In bottom-up proteomics, transmembrane proteins are mostly identified by their hydrophilic proteolytic peptides, and hydrophobic transmembrane segments generally remain uncharacterized and get lost in MS spectra because of the low solubilization of alpha-helices in aqueous media [38]. Better sequence coverage of α-helical parts of transmembrane proteins can be used as an indicator to assess efficiency of protein solubilization and sample preparation techniques. To evaluate the efficiency of FA*i*C-BPS, two measurements were considered, first, number of transmembrane proteins that have identified α-helical segments, and second, their sequence coverage.

We used Hidden Markov Model to make a database for alpha-helical transmembrane proteins. It is noteworthy to mention that details about Hidden Markov Model can be reached out at https://services.healthtech.dtu.dk/service.php?TMHMM-2.0.[39–41] We prepared the database of the sequence of α-helical segments of transmembrane proteins in yeast, and then the identified sequences of our samples in mass spec data were compared with this database. An in-house developed R program was used to measure the sequence coverage of alpha-helical segments of transmembrane proteins. Script of R program and instruction of using the script can be found in the Supplementary Information of the recently published article from this group.

Numerous researches have shown that in the presence of fluoroalcohols, such as HFIP, HFMIP, and TFE, alpha-helical structures are more stabilized.[42–45] In FA*i*C-BPS, high concentration of fluoroalcohols in the amphiphile-rich phases would facilitate concomitantly stabilization and solubilization of alpha-helices in transmembrane proteins. The number of integral membrane proteins with identified alpha-helical segments are presented in Table 4-4, and their sequence coverage in amphiphile-rich and aqueous phases are shown in Figure 4-6 (a and b), respectively. Results of both TEAB_HFIP and CTAB_HFIP systems show a notable improvement by increasing

the pH from 5.5 to 8.5. In the TEAB_HFIP system, at pH=5.5, only 18 alpha-helical transmembranes were identified with sequence coverage median of only 6%, rising the pH to 8.5 increased the number of alpha-helical transmembranes to 49 and improved the sequence coverage to median of 20%. By altering the amphiphile from TEAB as a QUATS with very short chain lengths (ethyl groups) to CTAB with one long hydrophobic chain (cetyl group), a significant improvement in the sequence coverage was observed, especially at pH=8.5 with median sequence coverage of 37%. It worth mentioning that previous results from this lab showed that addition of the zwitterionic surfactant 3-(N, N-Dimethyl myristyl ammonia) propane sulfonate (DMMAPS) to the solution of QUATS results in higher transmembrane α-helix coverage in FAiC-BPS. Combination of CTAB and DMMAPS could probably result in more improvement in sequence coverage of transmembrane alpha-helixes.

Several studies have reported loss of tertiary structures and stabilization of alpha-helices at pH values of 2 to 3 [23]. In a concurrently submitted manuscript, we have shown that at pH=3, in the aqueous phases of FA*i*C-BPS at which HFIP is used as coacervator and TBAB is used as amphiphile, sequence coverage of alpha-helical transmembranes improves significantly. However, by changing the coacervator to HFMIP and keeping other parameters constant, no noticeable change was observed compared to the control. Nevertheless, at pH=3, the TBAB_HFMIP systems shows better alpha-helical transmembrane coverage compared to pH=8.5 due to loss in tertiary structure at acidic pHs.
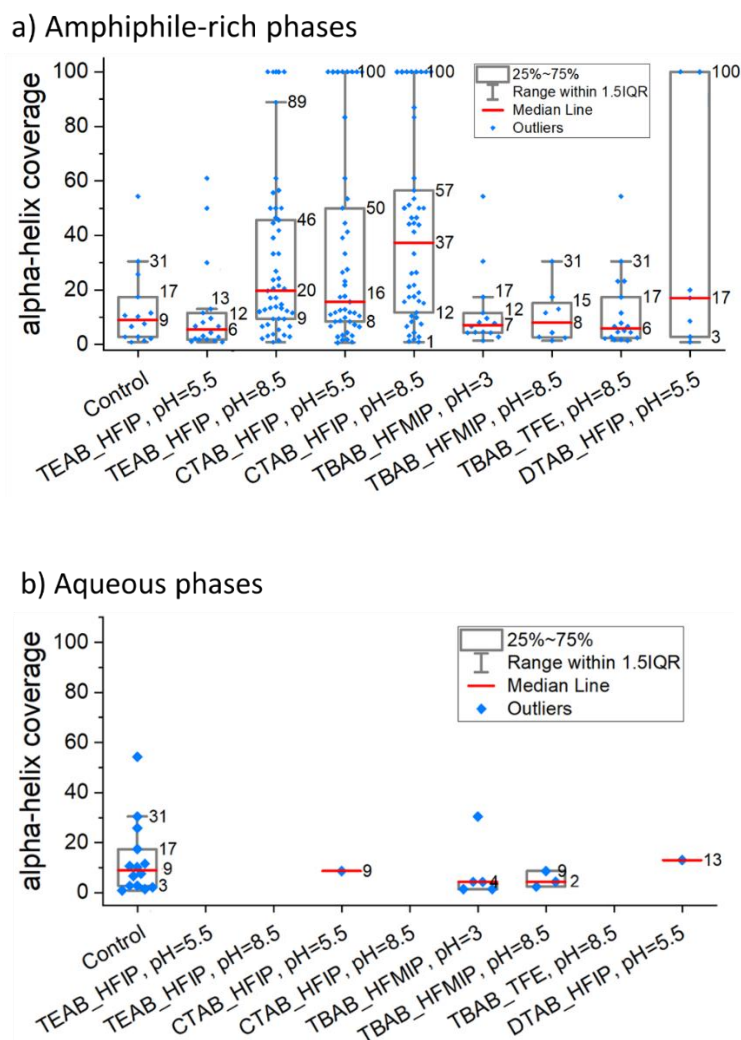
Figure 4-6. a) sequence coverage of the α-helical transmembranes in the amphiphile-rich phases of different FA*i*C-BPS, and b) sequence coverage of the α-helical transmembranes in the aqueous phases of different FA*i*C-BPS.

Table 4-4. Number of integral membrane proteins with identified α-helical transmembrane segments in FAiC-BPS and control.

FA*i*C-BPS	Amphiphile-rich phase	Aqueous phase
TEAB_HFIP, pH=5.5	18	0
TEAB_HFIP, pH=8.5	49	0
CTAB_HFIP, pH=5.5	45	1
CTAB_HFIP, pH=8.5	46	0
DTAB_HFIP, pH=5.5	7	1
TBAB_HFMIP, pH=3	14	5
TBAB_HFMIP, pH=8.5	8	3
TBAB_TFE, pH=8.5	18	0
Control	14	

4.3.6 SOLUBILIZATION AND IDENTIFICATION OF HYDROPHOBIC MEMBRANE AND INTEGRAL MEMBRANE PROTEINS

As shown in Table 4-5 by taking the advantage of FA*i*C_BPS, identification of membrane proteins and integral membrane proteins were improved due to their concomitant solubilization and enrichment in the amphiphile-rich phases with high concentration of fluoroalcohols. Nevertheless, different amphiphiles and different coacervators showed different performances in this regard. In terms of coacervators, comparison between the FA*i*C_BPS of this study and those from another concurrently filed manuscript shows that at pH=8.5, HFIP works better than TFE and HFMIP. One plausible justification is that TFE is not as hydrophobic as HFIP to solubilize hydrophobic integral membrane proteins, and in the case of HFMIP, it does not offer good fractionation and enrichment. In the case of amphiphiles, comparison between CTAB and DTAB shows that CTAB with a long hydrophobic chain performs better than DTAB in terms of hydrophobic integral membranes, even though DTAB led to identification of higher number of total proteins. Between different QUATS, at a same pH, TEAB with short tails performs better than other TBAB for internal membranes, probably due to less steric hinderance in electrostatic interaction. It is noteworthy to mention that at acidic pH values of around 3, proteins get more unfolded due to loss in tertiary structures that would help characterization of integral membranes, however, at this pH we cannot form FA*i*C-BPS by addition of HFIP to aqueous solution of TEAB, DTAB, or CTAB. Similarly, TFE as coacervator does not induce biphasic systems at low pHs without addition of salts. Appendix 4-3 presents the fractionation patterns of proteins belonging to different gene ontologies, such as membrane proteins and integral membrane proteins, in all FA*i*C-BPS od this study.

Table 4-5. Number of identified membrane proteins and integral membrane proteins, and their identification improvement compared to the control.

FA*i*C-BPS	Integral membrane proteins, Identification improvement %	Membrane proteins, Identification improvement %
TEAB_HFIP, pH=5.5	514, 18%	1040, 15%
TEAB_HFIP, pH=8.5	518, 19%	1037, 15%
CTAB_HFIP, pH=5.5	511, 17%	1033, 14%
CTAB_HFIP, pH=8.5	509, 17%	1031, 14%
DTAB_HFIP, pH=5.5	482, 11%	1009, 12%
TBAB_HFMIP, pH=3	497, 14%	1013, 12%
TBAB_HFMIP, pH=8.5	473, 8%	981, 9%
TBAB_TFE, pH=8.5	474, 9%	993, 10%
Control	436, N/A	903, N/A

4.4 CONCLUSIONS

In FA*i*C-BPS proteins fractionate between the amphiphile-rich aqueous phases, at the same time, amphiphile-rich phases provide a good solubilization power due to the high concentration of fluoroalcohols and amphiphiles; additionally, amphiphile-rich phases constitute a small fraction of total volume of system which leads to enrichment of proteins. These unique characteristics of FA*i*C-BPS make them favorable media for proteomics sample preparation, and by taking their advantages, proteins identification was improved. Interestingly, majority of this identification improvement was for low abundance proteins. Especially, identification of hydrophobic proteins such as integral membrane proteins was improved in FA*i*C-BPS and better sequence coverage of alpha-helical transmembranes was achieved in most systems. Finally, changing components and conditions of FA*i*C-BPS, including type of amphiphile, coacervator, and pH would result in changing fractionation patterns of proteins between the two phases, and some amphiphiles/coacervators showed more advantages in terms of low abundance or integral membrane proteins.

2.5 REFERENCES

(1)   Vuckovic, D.; Dagley, L. F.; Purcell, A. W.; Emili, A. Membrane Proteomics by High Performance Liquid Chromatography-Tandem Mass Spectrometry: Analytical Approaches and Challenges. *Proteomics* **2013**, *13* (3–4), 404–423. https://doi.org/10.1002/pmic.201200340.

(2)   Edel, K. H.; Marchadier, E.; Brownlee, C.; Kudla, J.; Hetherington, A. M. The Evolution of Calcium-Based Signalling in Plants. *Curr. Biol.* **2017**, *27* (13), R667–R679. https://doi.org/10.1016/j.cub.2017.05.020.

(3)   Yadeta, K. A.; Elmore, J. M.; Creer, A. Y.; Feng, B.; Franco, J. Y.; Rufian, J. S.; He, P.; Phinney, B.; Coaker, G. A Cysteine-Rich Protein Kinase Associates with a Membrane Immune Complex and the Cysteine Residues Are Required for Cell Death. *Plant Physiol.* **2017**, *173* (1), 771–787. https://doi.org/10.1104/pp.16.01404.

(4)   Almeida, J. G.; Preto, A. J.; Koukos, P. I.; Bonvin, A. M. J. J.; Moreira, I. S. Membrane Proteins Structures: A Review on Computational Modeling Tools. *Biochim. Biophys. Acta BBA - Biomembr.* **2017**, *1859* (10), 2021–2039. https://doi.org/10.1016/j.bbamem.2017.07.008.

(5)   Zhai, Y.; Su, J.; Ran, W.; Zhang, P.; Yin, Q.; Zhang, Z.; Yu, H.; Li, Y. Preparation and Application of Cell Membrane-Camouflaged Nanoparticles for Cancer Therapy. *Theranostics* **2017**, *7* (10), 2575–2592. https://doi.org/10.7150/thno.20118.

(6)   Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A Comprehensive Map of Molecular Drug Targets. *Nat. Rev. Drug Discov.* **2017**, *16* (1), 19–34. https://doi.org/10.1038/nrd.2016.230.

(7)   Petschnigg, J.; Moe, O. W.; Stagljar, I. Using Yeast as a Model to Study Membrane Proteins. *Curr. Opin. Nephrol. Hypertens.* **2011**, *20* (4), 425–432. https://doi.org/10.1097/MNH.0b013e3283478611.

(8)   Rucevic, M.; Hixson, D.; Josic, D. Mammalian Plasma Membrane Proteins as Potential Biomarkers and Drug Targets. *ELECTROPHORESIS* **2011**, *32* (13), 1549–1564. https://doi.org/10.1002/elps.201100212.

(9)   Arinaminpathy, Y.; Khurana, E.; Engelman, D. M.; Gerstein, M. B. Computational Analysis of Membrane Proteins: The Largest Class of Drug Targets. *Drug Discov. Today* **2009**, *14* (23–24), 1130–1135. https://doi.org/10.1016/j.drudis.2009.08.006.

(10)   Tan, S.; Tan, H. T.; Chung, M. C. M. Membrane Proteins and Membrane Proteomics. *Proteomics* **2008**, *8* (19), 3924–3932. https://doi.org/10.1002/pmic.200800597.

(11)   Rawlings, A. E. Membrane Proteins: Always an Insoluble Problem? *Biochem. Soc. Trans.* **2016**, *44* (3), 790–795. https://doi.org/10.1042/BST20160025.

(12)   Mesquita, R. O.; de Almeida Soares, E.; de Barros, E. G.; Loureiro, M. E. Method Optimization for Proteomic Analysis of Soybean Leaf: Improvements in Identification of New and Low-Abundance Proteins. *Genet. Mol. Biol.* **2012**, *35* (1 Suppl), 353–361. https://doi.org/10.1590/S1415-47572012000200017.

(13) McCord, J. P.; Muddiman, D. C.; Khaledi, M. G. Perfluorinated Alcohol Induced Coacervates as Extraction Media for Proteomic Analysis. *J. Chromatogr. A* **2017**, *1523*, 293–299. https://doi.org/10.1016/j.chroma.2017.06.025.

(14) Koolivand, A.; Clayton, S.; Rion, H.; Oloumi, A.; O'Brien, A.; Khaledi, M. G. Fluoroalcohol - Induced Coacervates for Selective Enrichment and Extraction of Hydrophobic Proteins. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **2018**, *1083*, 180–188. https://doi.org/10.1016/j.jchromb.2018.03.004.

(15) Koolivand, A.; Azizi, M.; O'Brien, A.; Khaledi, M. G. Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies. *J. Proteome Res.* **2019**, *18* (4), 1595–1606. https://doi.org/10.1021/acs.jproteome.8b00857.

(16) Ballesteros-Gómez, A.; Rubio, S.; Pérez-Bendito, D. Potential of Supramolecular Solvents for the Extraction of Contaminants in Liquid Foods. *J. Chromatogr. A* **2009**, *1216* (3), 530–539. https://doi.org/10.1016/j.chroma.2008.06.029.

(17) 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.

(18) Khaledi, M. G.; Jenkins, S. I.; Liang, S. Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles. *Langmuir* **2013**, *29* (8), 2458–2464. https://doi.org/10.1021/la303035h.

(19) Wang, Y.; Kimura, K.; Huang, Q.; Dubin, P. L.; Jaeger, W. Effects of Salt on Polyelectrolyte−Micelle Coacervation. *Macromolecules* **1999**, *32* (21), 7128–7134. https://doi.org/10.1021/ma990972v.

(20) Kayitmazer, A. B. Thermodynamics of Complex Coacervation. *Adv. Colloid Interface Sci.* **2017**, *239*, 169–177. https://doi.org/10.1016/j.cis.2016.07.006.

(21) Nejati, M. M.; Khaledi, M. G. Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis. *Langmuir* **2015**, *31* (20), 5580–5589. https://doi.org/10.1021/acs.langmuir.5b00444.

(22) Jenkins, S. I.; Collins, C. M.; Khaledi, M. G. Perfluorinated Alcohols Induce Complex Coacervation in Mixed Surfactants. *Langmuir* **2016**, *32* (10), 2321–2330. https://doi.org/10.1021/acs.langmuir.5b04701.

(23) Roccatano, D.; Fioroni, M.; Zacharias, M.; Colombo, G. Effect of Hexafluoroisopropanol Alcohol on the Structure of Melittin: A Molecular Dynamics Simulation Study. *Protein Sci. Publ. Protein Soc.* **2005**, *14* (10), 2582–2589. https://doi.org/10.1110/ps.051426605.

(24) Fatima, S.; Ahmad, B.; Khan, R. H. Native-like Tertiary Structure in the Mucor Miehei Lipase Molten Globule State Obtained at Low PH. *IUBMB Life* **2007**, *59* (3), 179–186. https://doi.org/10.1080/15216540701335716.

(25)   Ku, T.; Lu, P.; Chan, C.; Wang, T.; Lai, S.; Lyu, P.; Hsiao, N. Predicting Melting Temperature Directly from Protein Sequences. *Comput. Biol. Chem.* **2009**, *33* (6), 445–450. https://doi.org/10.1016/j.compbiolchem.2009.10.002.

(26)   Otzen, D. E. Protein Unfolding in Detergents: Effect of Micelle Structure, Ionic Strength, PH, and Temperature. *Biophys. J.* **2002**, *83* (4), 2219–2230. https://doi.org/10.1016/S0006-3495(02)73982-9.

(27)   Natarajan, S.; Xu, C.; Caperna, T. J.; Garrett, W. M. Comparison of Protein Solubilization Methods Suitable for Proteomic Analysis of Soybean Seed Proteins. *Anal. Biochem.* **2005**, *342* (2), 214–220. https://doi.org/10.1016/j.ab.2005.04.046.

(28)   Moore, S. M.; Hess, S. M.; Jorgenson, J. W. Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics. *J. Proteome Res.* **2016**, *15* (4), 1243–1252. https://doi.org/10.1021/acs.jproteome.5b01122.

(29)   Lin, Y.; Liu, H.; Liu, Z.; Liu, Y.; He, Q.; Chen, P.; Wang, X.; Liang, S. Development and Evaluation of an Entirely Solution-Based Combinative Sample Preparation Method for Membrane Proteomics. *Anal. Biochem.* **2013**, *432* (1), 41–48. https://doi.org/10.1016/j.ab.2012.09.023.

(30)   Lu, B.; McClatchy, D. B.; Kim, J. Y.; Yates, J. R. Strategies for Shotgun Identification of Integral Membrane Proteins by Tandem Mass Spectrometry. *Proteomics* **2008**, *8* (19), 3947–3955. https://doi.org/10.1002/pmic.200800120.

(31)   Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A. Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.* **2010**, *9* (6), 2863–2870. https://doi.org/10.1021/pr900949p.

(32)   Andersen, P.; Heron, I. Simultaneous Electroelution of Whole SDS-Polyacrylamide Gels for the Direct Cellular Analysis of Complex Protein Mixtures. *J. Immunol. Methods* **1993**, *161* (1), 29–39. https://doi.org/10.1016/0022-1759(93)90195-d.

(33)   Lin, Y.; Zhou, J.; Bi, D.; Chen, P.; Wang, X.; Liang, S. Sodium-Deoxycholate-Assisted Tryptic Digestion and Identification of Proteolytically Resistant Proteins. *Anal. Biochem.* **2008**, *377* (2), 259–266. https://doi.org/10.1016/j.ab.2008.03.009.

(34)   Basiri, B.; van Hattum, H.; van Dongen, W. D.; Murph, M. M.; Bartlett, M. G. The Role of Fluorinated Alcohols as Mobile Phase Modifiers for LC-MS Analysis of Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 190–199. https://doi.org/10.1021/jasms.8b05367.

(35) Jha, N. S.; Judy, E.; Kishore, N. 1,1,1,3,3,3-Hexafluoroisopropanol and 2,2,2-Trifluoroethanol Act More Effectively on Protein in Combination than Individually: Thermodynamic Aspects. *J. Chem. Thermodyn.* **2018**, *121*, 39–48. https://doi.org/10.1016/j.jct.2018.02.011.

(36)   Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Membrane Proteins. *Mol. Biol. Cell 4th Ed.* **2002**.

(37)   Zhang, S.-Q.; Kulp, D. W.; Schramm, C. A.; Mravic, M.; Samish, I.; DeGrado, W. F. The Membrane- and Soluble-Protein Helix-Helix Interactome: Similar Geometry via Different

Interactions. *Struct. Lond. Engl. 1993* **2015**, *23* (3), 527–541. https://doi.org/10.1016/j.str.2015.01.009.

(38)   Xiong, Y.; Chalmers, M. J.; Gao, F. P.; Cross, T. A.; Marshall, A. G. Identification of Mycobacterium Tuberculosis H37Rv Integral Membrane Proteins by One-Dimensional Gel Electrophoresis and Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry. *J. Proteome Res.* **2005**, *4* (3), 855–861. https://doi.org/10.1021/pr0500049.

(39)   Services https://services.healthtech.dtu.dk (accessed Dec 13, 2020).

(40)   Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes11Edited by F. Cohen. *J. Mol. Biol.* **2001**, *305* (3), 567–580. https://doi.org/10.1006/jmbi.2000.4315.

(41)   Käll, L.; Krogh, A.; Sonnhammer, E. L. L. Advantages of Combined Transmembrane Topology and Signal Peptide Prediction—the Phobius Web Server. *Nucleic Acids Res.* **2007**, *35* (suppl_2), W429–W432. https://doi.org/10.1093/nar/gkm256.

(42)   Hong, D.-P.; Hoshino, M.; Kuboi, R.; Goto, Y. Clustering of Fluorine-Substituted Alcohols as a Factor Responsible for Their Marked Effects on Proteins and Peptides. *J. Am. Chem. Soc.* **1999**, *121* (37), 8427–8433. https://doi.org/10.1021/ja990833t.

(43)   Roccatano, D.; Colombo, G.; Fioroni, M.; Mark, A. E. Mechanism by Which 2,2,2-Trifluoroethanol/Water Mixtures Stabilize Secondary-Structure Formation in Peptides: A Molecular Dynamics Study. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (19), 12179–12184. https://doi.org/10.1073/pnas.182199699.

(44)   Hirota, N.; Mizuno, K.; Goto, Y. Cooperative Alpha-Helix Formation of Beta-Lactoglobulin and Melittin Induced by Hexafluoroisopropanol. *Protein Sci. Publ. Protein Soc.* **1997**, *6* (2), 416–421. https://doi.org/10.1002/pro.5560060218.

(45)   Cammers-Goodwin, A.; Allen, T. J.; Oslick, S. L.; McClure, K. F.; Lee, J. H.; Kemp, D. S. Mechanism of Stabilization of Helical Conformations of Polypeptides by Water Containing Trifluoroethanol. *J. Am. Chem. Soc.* **1996**, *118* (13), 3082–3090. https://doi.org/10.1021/ja952900z.

CHAPTER 5

ENHANCING LIPIDOME COVERAGE BY FRACTIONATION OF COMPLEX
LIPID MIXTURES IN THE NOVEL DCM-HFIP-WATER MULTIPHASE
SYSTEM AS A COMPLEMENTARY METHOD FOR LC-MS- BASED
LIPIDOMICS

ABSTRACT

Lipid extracts from biological samples are very complex mixtures and their characterization
has been remained challenging. In LC-MS-based lipidomics, low abundance lipids are usually
underrepresented under the shadow of high abundance lipids. Additionally, coelution of different
lipids due the complexity of mixtures would result in losing lipidome coverage. Here, we introduce
a novel multiphase system composed of a top aqueous phase, a middle phase which is rich in
dichloromethane (DCM), and a bottom phase which is rich in hexafluoroisopropanol (HFIP).
Lipids are extracted and got enriched in different phases based on their physicochemical
properties. Combination of this simple fractionation step with LC-MS-based lipidomics would
result in significant improvements in lipidome coverage; in this case lipid coverage was improved
by more than 150%. Generally, lipids get fractionated and enriched in different phases based on
their polar headgroups and hydrophobic chains. The most hydrophobic lipids get enriched in the
DCM-rich phase, relatively less hydrophobic lipids get enriched in the HFIP-rich phase, and the

most hydrophilic lipids gets enriched in the aqueous phase. A simple comparison between the lipids that are uniquely identified in each phase reveals that phosphatidylinositols (PIs), with water-soluble myo-inositol headgroups, are mostly extracted by the HFIP-rich phase. About 27% of the unique lipids in the HFIP-rich phase were PIs, while only less than 1% of the unique lipids in the hydrophobic DCM-rich phase were PIs. On the other hand, phosphatidylglycerols (PGs) get enriched in the DCM-rich phase. About 21% of the unique lipids in the HFIP-rich phase were PGs, while only about 3% of the unique lipids in the hydrophobic DCM-rich phase were PGs; interestingly, the PGs in the DCM-rich phase had significantly longer nonpolar acyl tails than those in the HFIP-rich phase. Lyso-phospholipids and Fatty acid ester of hydroxyl fatty acids (FAHFAs) are other classes of lipids that get enriched in the HFIP-rich phase; while ceramides (Cer), and hexosylceramides (HexCer) are examples of lipid classes that get enriched in the DCM-rich phase. Addition of this simple fractionation step to the standard LC-MS-based lipidomics workflow would considerably enhance the lipidome coverage.

KEYWORDS: Chromatography-Mass Spectrometry (LC-MS), Fractionation and Enrichment, Lipidomics, Lipidome Coverage, Liquid.

5.1 INTRODUCTION

Lipids are the main component of cellular membranes and they have different key functions in the cell. Some examples of their functions are energy storage,[1] immune regulation and self-defense,[2] signaling molecules,[2] modulating protein function,[3] and they are substrates for post-translational protein-lipid modification.[4] Lipids have a great structural diversity, the compositional diversity of membrane lipids are not only different at various organisms and cell types,[5–7] but also affected by metabolic pathways.[8] Lipid levels are governed by a complex metabolic pathways and alterations of membrane lipid homeostasis are connected to many diseases.[8] For example changes

in phospholipids composition have been reported in human colorectal cancer cells.[9] Other examples are changes in hexosylceramides (HexCer) levels in brain aging,[10] and alterations of in the level of phosphatidylcholine in Alzheimer's disease.[11] Therefore, measuring such alterations in lipid composition widens our knowledge about the metabolic disease states. This shows the crucial importance of lipid characterization in biological samples.

Phospholipids, represent the most widespread lipid class in cell membranes.[12] The main class of phospholipids are amphiphilic diacyl-glycerophospholipids, composed of two hydrophobic nonpolar fatty acyl chains attached the glycerol group, and a polar phosphate head group is attached to the third carbon of the glycerol.[13–15] Diacyl-glycerophospholipids are classified based on their polar head groups to phosphatidic acid (PA), phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylinositol (PI), phosphatidylserine (PS). Lysophospholipids are other classes of phospholipids that are intermediates in the metabolism of lipids. Because they are synthesized from the hydrolysis of an acyl group from the sn-1 position of diacyl-glycerophospholipids, they are called lysophospholipids.[16] They are classified to subcategories of LysoPA, LysoPC, LysoPE, LysoPG, LysoPI, and LysoPS. In addition, numerous studies have shown the presence of ether-phospholipids, such as EtherPC and EtherPE, in membranes of different species.[17] The difference between diacyl-glycerophospholipids and ether-phospholipids is that in the former one, the two acyl chains are linked to the glycerol by ester bonds, while in the ether-phospholipids, the acyl chain in position sn-1 is bound to the glycerol backbone by either an alkyl- or an alkenyl-bond.[18,19] Ceramides (Cer), composed of sphingosine and a fatty acyl,[20] are the other major class of lipids in the cell membranes.[21] Additionally, hexosylceramides (HexCer), are other class of ceramides which contain either a glucosyl or a galactosyl sugar moiety.[10] Fatty acid ester of hydroxyl fatty

acid (FAHFA),[22–26] non-bilayer lipid monogalactosyldiacylglycerols (MGDG),[27–29] and sphingomyelins (SM) are other main classes of lipids.[30]

Diversity of lipids is not limited to different lipid classes; different hydrophobic backbones such as saturated, monounsaturated, and polyunsaturated acyl groups, with different carbon numbers at both sn-1 and sn-2 positions also affect lipids' diversity. This would result in very complex mixtures of lipid extracts from biological samples that makes lipid characterization very challenging. Thanks to the recent advances in liquid chromatography mass spectrometry (LCMS) techniques and advanced new instrumentation, lipidomics has been greatly developed, however, many challenges are still unresolved. For example, in LCMS techniques, low abundance lipids get lost under the shadow of high abundance lipids. Additionally, coelution of different lipids in LC due the complexity of mixtures would result in losing identification power. Prefractionation of lipids in multi-phase systems would results in fewer complex samples and concentrating specific lipids classes with similar physicochemical properties in each phase. Combining this simple and novel technique with LCMS would result in significant improvements in lipidomics. Previously, we have shown that using this technique in proteomics leads to significant improvement in identification of low abundance proteins.[31–33] The same terminology can be used to develop lipidomics techniques. Here we introduce the novel three-phase system consists of dichloromethane (DCM), Hexafluoroisopropanol (HFIP), and water as a complementary method for lipidomics studies. The three-phases system consists of the least hydrophobic aqueous phase on the top, a the most hydrophobic DCM-rich phase in the middle, and HFIP-rich phase at the bottom with slighter hydrophobicity. Due to the different hydrophobicity of phases, lipid can be fractionated between the phase base on their hydrophobicity. At the same, they can fractionate base on the lipid classes due to the different molecular interactions between the polar headgroups

of lipids and components of phases. This simple fractionation step, combined with LCMS lipidomics approaches, will result in significant lipid identification enhancement.

5.2 MATERIALS AND METHODS

When 22% (V/V) DCM, 18% HFIP, 60% water are mixed thoroughly and centrifuged, the result is a three-phase system, where the top phase is aqueous, middle phase is DCM-rich, and bottom phase is HFIP-rich (additional information about the compositional analysis of the phases can be found in section 5.2.1.). This system was used for lipid fractionation. Total yeast lipid extract was provided from Avanti® Lipids. Lipids were initially dissolved in chloroform with the concentration of 25 µg/µL. 10 mg lipid was taken (400 µL×25 µg/µL), chloroform was evaporated under nitrogen flow, then dissolved in 440 µL DCM, then 360 µL HFIP, 1200 µL water, and 2 µL formic acid were added to bring the volumetric ratios to 22% (V/V) DCM, 18% HFIP, 60% water and 0.1% formic acid (in total volume of 2 mL). The mixture was centrifuged at 5,000×g for 10 minutes, and the three phases were separated, evaporated, and finally dissolved in IPA. Final concentration of the lipids that were extracted by DCM-rich and HFIP-rich phases were adjusted to 25 µg/µL as explained in section 5.2.2. The whole process workflow is schematically illustrated in Figure 5-1. Finally, 400 µg of lipid from DCM-rich and HFIP-rich phases were subjected to LC-MS/MS analysis as explained in detail in section 5.2.3., and the results were analyzed as explained in section 5.2.4. It worth to note that the total amount of lipid in the aqueous phase (top phase) was far less than 400 µg, therefore all the lipid from the aqueous phase was used for a single injection.

After data processing, the lipids that were reproducibility identified in at least two out of the three biological replicates were selected as the reproducible lipids in the final list for further analysis. The results were compared to the control, where the control is the same lipid extract

without subjecting to phase separation before the LCMS analysis. Control was also analyzed three times and the lipids that were reproducibly identified in at least two replicates were considered as the control.
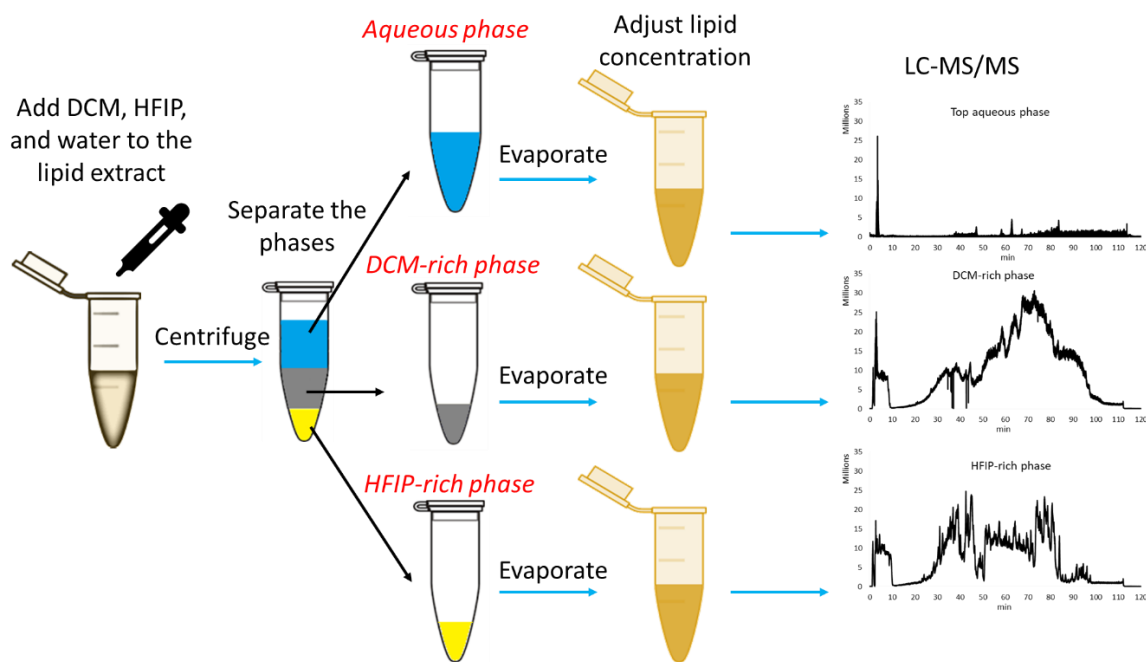
Figure 5-1. Schematic of the workflow

To measure the polarity of phases relative to each other, partition coefficient of homologous series of alkylphenones in DCM-rich and HFIP-rich phases were measured. Relatively nonpolar compounds, such as alkylphenones, can be fractionated and partitioned between different phases with different polarities. In this case, acetophenone, propiophenone, butyrophenone, and valerophenone were used as homologous series of alkylphenones. Partition coefficient of these compound between the DCM-rich and HFIP-rich phases by LC separation. 100 µmol of each alkylphenone (11.65 µL acetophenone, 13.28 µL propiophenone, 14.51 µL butyrophenone, and 16.42 µL valerophenone) were added to the system composed of 22% (V/V) DCM, 18% HFIP,

97

and 60% water (in total volume of 10mL). After reaching the equilibrium, phases were separated, then the DCM-rich and HFIP-rich phases were analyzed by HPLC as explained in section 2.2.

5.2.1 COMPOSITIONAL ANALYSIS OF PHASES BY GC-FID

GC-FID was used for compositional analysis of the phases. Column was SLB-IL60i ionic liquid column with inert treatment, where the temperature gradient was 100 to 140 °C with ramp of 10 °C/min and no hold time. Injection and detector were both at 250 °C, where injection volume was 1 µL, with split ratio of 100:1. Carrier gas: Helium, with flow rate of 1 mL/min at constant flow mode. Chromatograms are presented in Appendix 5-1 and results are discussed in section 5.3.1 (Compositional analysis of phases)

5.2.2 LC ANALYSIS FOR QUANTIFICATION OF HOMOLOGOUS SERIES OF ALKYLPHENONES

A RPLC C18 column (Thermo Scientific, part number: 28105-154630) with dimensions of 150×4 mm, and particle size of 5 µm (Spherical, Fully Porous) was used for separation of alkylphenones, where mobile phases A and B, were respectively water and ACN, and the gradient was 20-40% B in 20 minutes, followed by 40-50% B in 7 minutes, and 3 minutes of 20% B for reconditioning the column for the next injection. Chromatograms are presented in Appendix 5-2 and results are discussed in section 5.3.2.

5.2.3 LC-MS/MS ANALYSIS FOR LIPID SEPARATION, IDENTIFICATION, AND QUANTIFICATION

Shimadzu LCMS-9030 Q-TOF was used for separation and identification lipids. Lipids were separated on a C18 reverse phase liquid chromatography (RPLC) column (ACE Excel superC18, EXL-1711-1002U) with 1.7 µm particle size, 10 cm length, 2.1 mm ID. Mobile phase A was 40% (V/V) acetonitrile (ACN) in 60% water with 10 mM ammonium acetate and Mobile phase B was 0.1% (V/V) water in 99.9% isopropanol (IPA) with 10 mM ammonium acetate. Separation

gradient was 2 hours at the flow rate of 0.15 mL/min and room temperature, where gradient method was 0% B for 2 minutes, then ramp of 0% to 50% B from 2 to 30 minutes, followed by another ramp of 50% to 98% B from 30 to 100 minutes, then at 98% B for 10 minutes to make sure that all hydrophobic lipids are eluted, and finally reconditioning the column at 0% B for 10 min. For the control and the lipids extracted by HFIP-rich and DCM-rich phases, 400 µg ($16\,\mu L \times 25\,\mu g/\mu L$) lipid was injected. For the lipids extracted by the top aqueous phase, the whole lipid was injected due to the small amount of lipids extracted by this phase. For MS1 and MS2, m/z ranges were selected as 300-2000 and 150-1500, respectively. MS conditions were selected as follows: nebulizing gas flow rate was 3.0 L/min, heating gas flow was 10 L/min, interface temperature was set at 300 ˚C.

For quantification of the total amount of lipids in each phase and adjusting their concentrations before LCMS, 4 µL of sample was directly injected (without LC), where MS1 m/z range was 300-2000, with no MS2. The peek tubing had inner diameter (ID) of 0.005", flow rate was 0.1 mL/min, and eluent was isopropanol. For calibration line, the same conditions were used where known concentrations of lipid were injected.

It worth to note that the rinsing solution should be hydrophobic enough to clean the very hydrophobic lipid from the needle and injection parts. In this case, 25% IPA, 25% ACN, 25% methanol, and 25% water was used as the rinsing solution. Additionally, IPA was injected between any two lipid injections.

5.2.4 DATA PROCESSING

Files from Shimadzu QTOF (.lcd) were converted to centroid .mzML files, then centroid .mzML files were converted to .abf files with Reifysc Analysis Base File Convertor software. MS-DIAL software was used for identification of lipids. Searching parameters were as follows:

ionization type: soft ionization, method type: conventional LCMS or data dependent MS/MS, mass range begin: 0, mass range end: 2000 Da, MS1 tolerance: 0.01 Da, MS2 tolerance: 0.05 Da, Max charge: 2, smoothing level: 3 scans, minimum peak width: 5 scans, minimum peak height: 3000 amplitude, mass slice width: 0.1 Da, together with alignment: yes, identification score cut off: 85%, solvent adduct ion type: CH3COONH4, collision type: CID, adduct ion: [M-H]$^-$. The lipids which were identified at least in two replicates out of three were counted as the biological reproducible ones for further data analysis.

5.3 RESULTS AND DISCUSSIONS

For several years, various two-phase systems such as chloroform_methanol_water[34–36], and methyl-tert-butyl ether (MTBE)_methanol_water[37,38] have been used for extraction of lipids from biological tissues. The lipids extracted by these methods have a wide range of physicochemical properties, and this great structural diversity and chemical complexity among lipids causes the profiling of the complete lipidome of biological samples to be so challenging.[39,40] Here we show that fractionation of lipids based on their physicochemical properties would greatly facilitate their identification.

Like the chloroform_methanol mixture, the DCM_HFIP mixture is hydrophobic enough to solubilize and extract the lipids, but they have very distinct physicochemical properties compared to the widely used chloroform-methanol. Fluorinated groups on HFIP make it highly polar and strong hydrogen-bond-donor solvent,[41–43] it is fully miscible in water while having almost a large density. In addition, chloroform and DCM are very similar in chemical structure, however, as shown in Table 5-1, DCM has more relative polarity than chloroform, smaller density than chloroform and HFIP, and more water solubility than chloroform. Combination of all these distinct physicochemical properties make the DCM-HFIP mixture a suitable media to form a unique three-phase system in aqueous media at certain concentrations. To the best of our knowledge, the three-

phase system of DCM_HFIP_water has not been reported previously, and we use the unique properties of this system to improve LCMS-based lipidomics studies. As shown in Figure 5-2, a simple fractionation of lipids in this novel multiphase system prior to the LCMS, resulted in identification of several lipids that were not identified in the control sample under the same LCMS conditions. This significant improvement is due to the concentration of lipids with similar physicochemical properties in different phases that facilitates their identification by LCMS. In Figure 5-2 (a), the number of identified lipids after fractionation in the HFIP-DCM-water system is compared to the number of identified lipids in the control. Figure 5-2 (b) compares the number of identified lipids in each of the phases and the control, and numbers specified by circles around them are the lipids that were not identified in the control.

Table 5-1. Physicochemical properties of chloroform, DCM, and HFIP. [44]

Compound	Boiling point (°C)	Density (g/mL)	Solubility in H_2O (g/100g)	Relative polarity
Chloroform	61.2	1.498	0.8	0.259
DCM	39.8	1.326	1.32	0.309
HFIP	58.2	1.6	Totally miscible	0.969
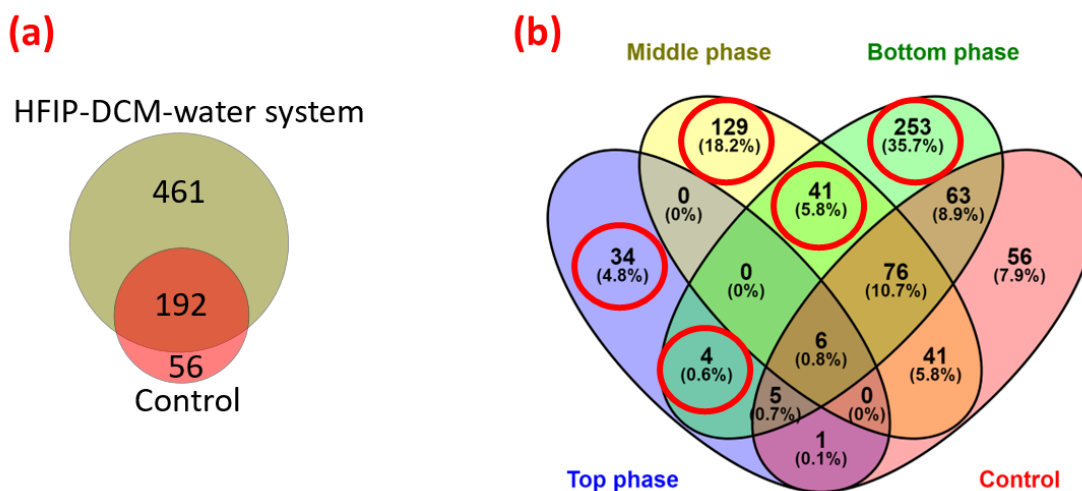
Figure 5-2. a), the number of identified lipids after fractionation in the HFIP-DCM-water multiphase system vs. the control under identical LCMS conditions. (b), the number of identified lipids in each of the phases of the HFIP-DCM-water multiphase system vs. the control under identical LCMS conditions; red

circles denote the number of lipids that are uniquely identified in each of the phases, but not in the control.

5.3.1 COMPOSITIONAL ANALYSIS OF PHASES

Composition of phases were measured by GC-FID and calibration lines for DCM and HFIP are presented in Appendix 5-1. As shown in Figure 5-3, the mixture of DCM, HFIP, and water forms a three-phase system at specific conditions. It worth to note that the blue color in Figure 5-3 is coming from the dye (methylene blue at very low concentration) which was added to the mixture to visually differentiate the phases. In all of the conditions shown in Figure 5-3, initial percentage of water is kept constant at 60% (V/V), while the ratio of DCM to HFIP is gradually increasing to show the trend from two-phase to three-phase and again to two-phase system. Initially, at water:DCM:HFIP volumetric ratios of 60:10:30, a two-phase system is formed where the top aqueous phase has composition of about 13% HFIP and 1.5% DCM, while the organic phase has composition of 54% HFIP and 17% DCM. As we continue to increase the ratio of DCM/HFIP, at water:DCM:HFIP volumetric ratios of 60:26:14, a third phase starts to appear. From this point to water:DCM:HFIP volumetric ratios of 60:14:26, a three-phase condition exists where the composition of top aqueous phase remains unchanged at around 13% HFIP and 1% DCM. This composition makes the top aqueous phase a suitable media for extraction of the lipids with the lowest relative lipophilicity. During the whole three-phase range, the middle phase is DCM-rich, and the HFIP percentage remains almost constant at about 20% HFIP, and percentage of DCM changes in the range of 69 to 78%. At this range, the bottom phase is HFIP-rich, and the HFIP percentage remains nearly unchanged at around 53% HFIP while the DCM percentage varies from 21 to 17%. For lipidomics application, the initial condition of water:DCM:HFIP volumetric ratios of 60:22:18 was chosen for lipid fractionation due to the following reasons, first, the difference in composition of DCM and HFIP between the middle and bottom phases is larger and

therefore it shows more hydrophobicity difference between the phases, second, the volume of both

phases is large enough for sample handling.
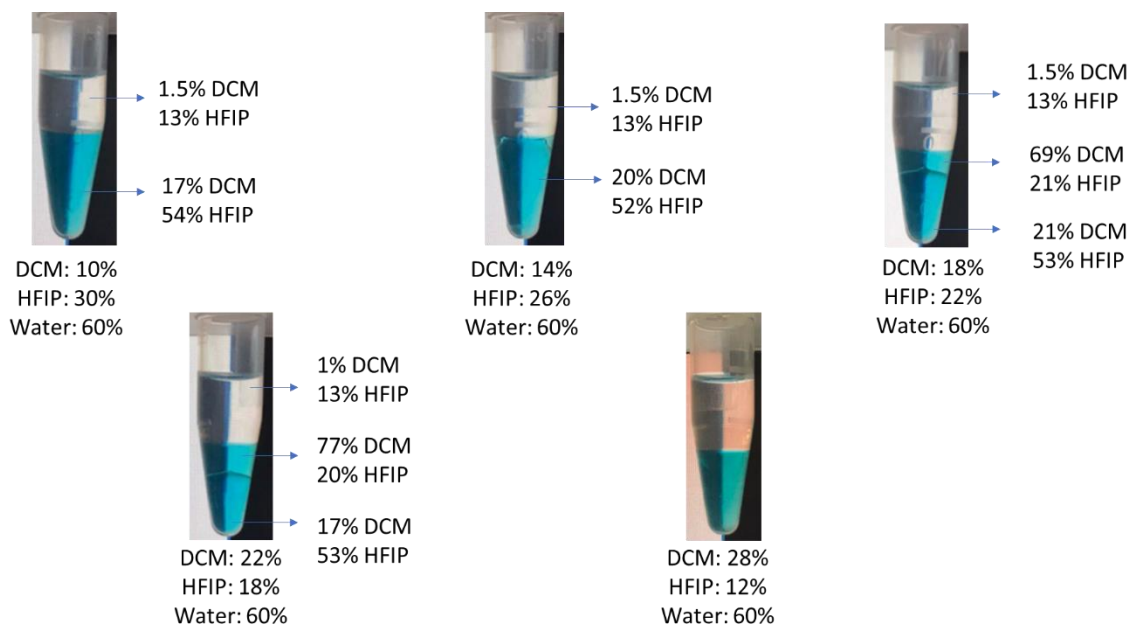
Figure 5-3. Compositional analysis of phases (V/V%) by gas chromatography (GC) in the range the three-phase system exists.

5.3.2 METHYLENE SELECTIVITY AND RELATIVE HYDROPHOBICITY OF DCM-RICH AND HFIP-RICH PHASES

Alkylphenones are quite non-polar compounds that can be partitioned between HFIP-rich and

DCM-rich phases. When alkylphenones are added to this system, as the alkyl group's chain lengths

increases, the ratio of concentration of alkylphenone in the DCM-rich phase to the HFIP-rich phase

increases. Therefore, changes in partition coefficient of homologous series of alkylphenones show

the relative hydrophobicity of phases. Acetophenone, propiophenone, butyrophenone, and

valerophenone were added to the system, then phases were separated and partition coefficients (P)

of these compounds between the DCM-rich phase and HFIP-rich phases were measured

(chromatograms are available in Appendix 5-2). As shown in the Figure 5-4 the small slope of

Log P vs. carbon number in the alkyl group shows that DCM-rich and HFIP-rich phases have

different hydrophobicities, however, this difference between hydrophobicity of phases is small and both phases have enough lipophilicity to dissolve lipids. As a result, less hydrophobic lipids can be extracted and enriched in the HFIP-rich phase, and more hydrophobic lipid can be enriched in the HFIP-rich phase.
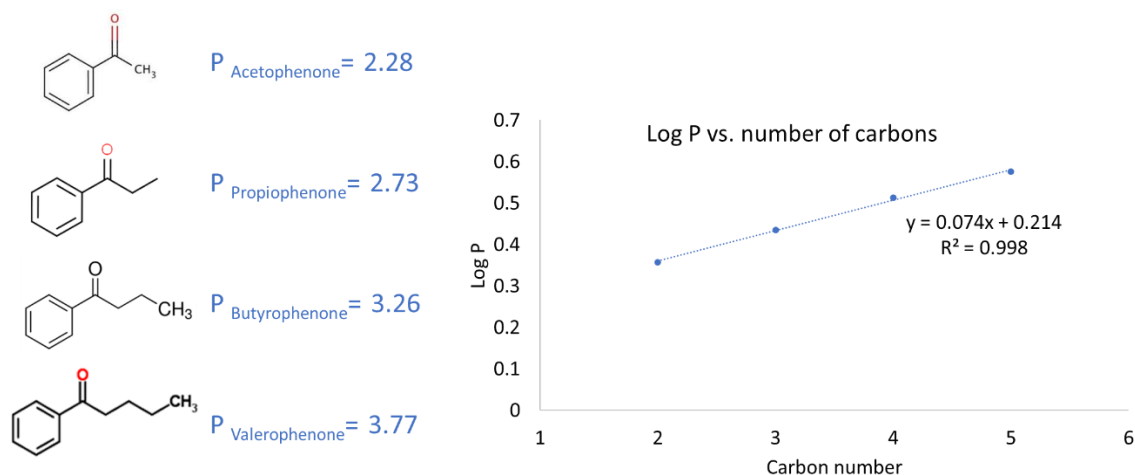
Figure 5- 4. Methylene selectivity, relative hydrophobicity of DCM-rich and HFIP-rich phases by

5.3.3 FRACTIONATION AND ENRICHMENT OF DIFFERENT CLASSES OF LIPIDS

Table 5-2 shows the relative percentages of lipids belonging to different classes in the control. The values are based on the peak area of individual identified lipids. Each percentage is calculated by dividing "the summation of peak areas of lipids in each class" over the "summation of peak areas of all identified lipids", which is shown in Equation 5-1.

$$\% \ of \ lipids \ in \ a \ specific \ class = \frac{\sum peak \ area \ of \ lipids \ in \ a \ specific \ class}{peak \ areas \ of \ all \ identified \ lipids} \qquad \text{(Equation 5-1)}$$

Table 5-2. Relative percentage of lipids in different classes in the control.

Lipid class	%	Lipid class	%
Cer	14.9	PA	8.6
EtherPC	3.7	PC	9.7
EtherPE	0.5	PE	1.7

FAHFA	0.9	**PG**	11.2
HexCer	13.7	**PI**	16.0
Lyso	10.2	**PS**	5.4
MGDG	2.5	**SM**	1.1

After fractionation of lipids in the DCM/HFIP/water system, each class of lipids generally gets enriched in either DCM-rich phase of HFIP-rich phase. Figure 5-5 compares the percentage of uniquely identified lipids in these two phases, where percentages are calculated based on their peak intensity. Some classes of lipids, such as FAHFA, PI, PE, and Lyso are enriched in the HFIP-rich phase, on the other hand, some other classes of lipids such as Cer, HexCer, MGDG, and PG are enriched in the DCM-rich phase. This enrichment of specific lipid classes in specific phases based on their physicochemical properties is the basis for identification improvement.

PIs are phospholipids comprising a water-soluble head group (myo-inositol),[45] and they have the most hydrophilic headgroup among all phospholipids. This water-soluble head group in addition to the hydrogen bond donor groups in HFIP that can interact with inositol headgroup, justify the enrichment of PIs in the HFIP-rich phase. As another example, Lyso groups have only one hydrophobic chain and generally they have less hydrophobicity among all phospholipids, which justifies their enrichment in the less hydrophobic HDIP-rich phase.

However, the nature of head group is not the only parameter that governs lipid fractionation patterns. For instance, PGs are mostly extracted by the most hydrophobic DCM-rich phase, however, the glycerol headgroup in the PG is not the most hydrophobic headgroup among phospholipids (*e.g.,* choline headgroup in PCs is more hydrophobic than the glycerol headgroup in PGs). If we compare PGs that are identified in the DCM-rich phase with those PGs that are identified in the HFIP-rich phase (shown in Figure 5-6), we observe a distinct difference between the average of number of carbons in their hydrophobic acyl tails. Different factors such as head

group, number of carbons and number of double bonds in the chains govern the distribution patterns.
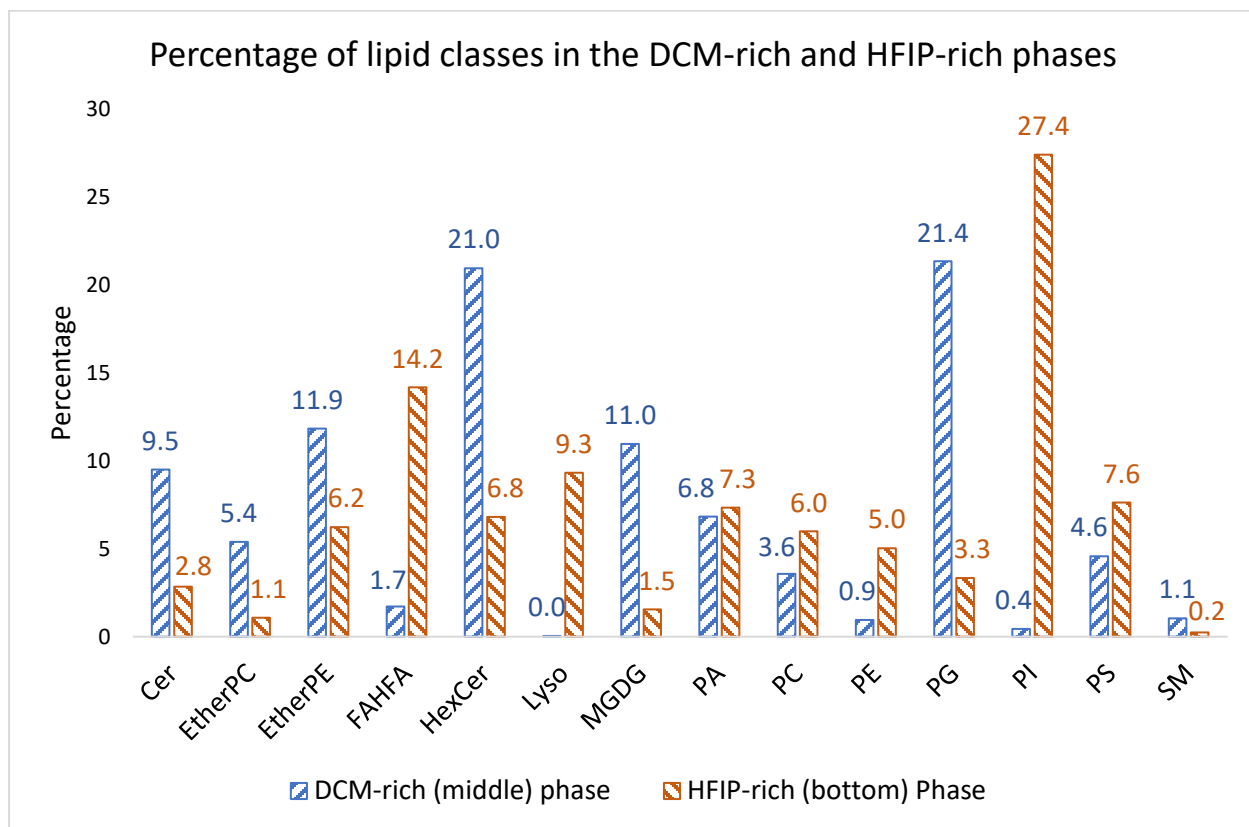
Figure 5-5. Percentage of lipid classes in the DCM-rich and HFIP-rich phases.
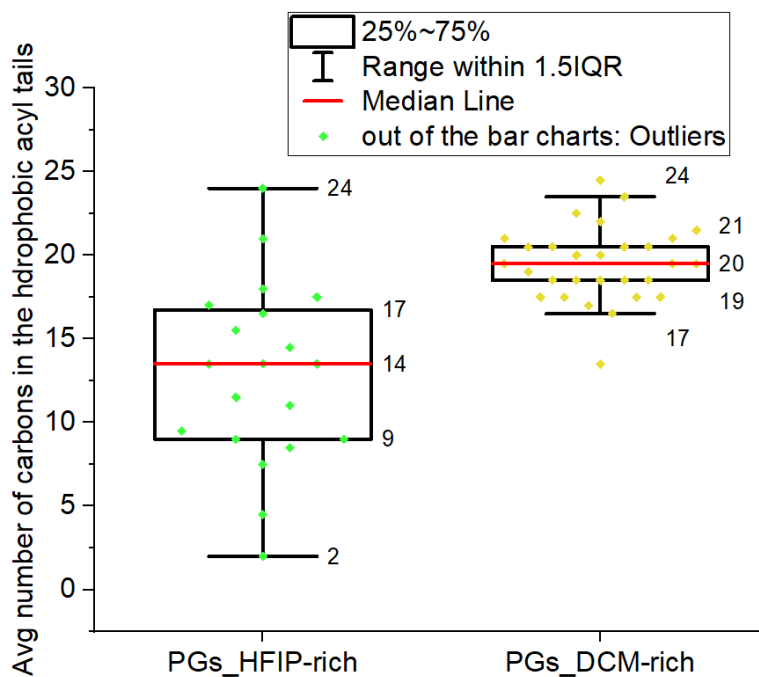
Figure 5-6. Comparison between the average number of carbons in the acyl chains of the PGs that are identified in HFIP-rich and DCM-rich phases.

5.3.4 ENRICHMENT OF LIPIDS BASED ON THEIR LIPOPHILICITY AND FATTY ACYL CHAIN

The relative hydrophobicity of lipids can be addressed with various indicators, such as their retention time on RPLC, their partition coefficient in octanol/water system (Log P), and their solubility in water (Log S). Figure 5-7 shows the distribution (box chart) of retention time of the lipids that are uniquely identified in each of the phases. Majority of the lipids that are selectively extracted by the aqueous top phase have small retention times of less than 3.7 minutes. This reveals that the aqueous phase top phase which is composed of around 13% HFIP and less than 2% DCM, selectively extracts and enriches the least hydrophobic lipids. Similarly, the lipids in the HFIP-rich phase generally have less retention times on RPLC compared to the lipids in the DCM-rich phase.
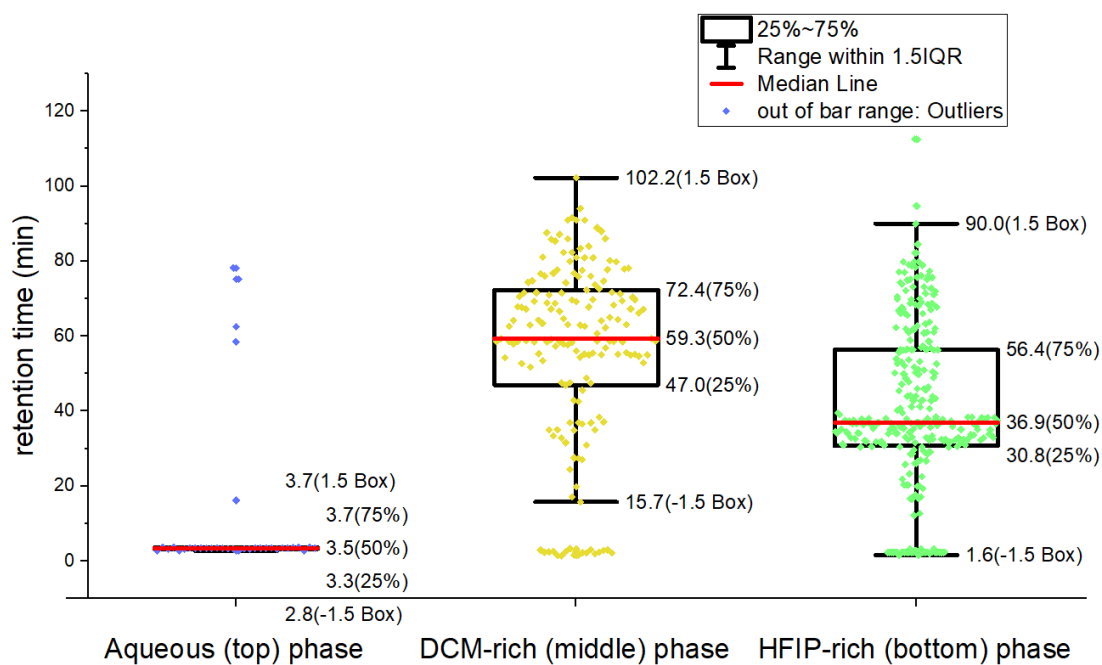
Figure 5-7. Distribution of retention time of the lipids that are uniquely identified in each phase.

These results are in correspondence with the RPLC chromatogram. Figure 5-8 shows the RPLC chromatogram of the lipids that are extracted by the top aqueous, DCM-rich, and HFIP-rich phases. It worth to note that the top aqueous phase extracts only few percentages of lipids due to its hydrophilic nature, that is why this phase is excluded in most parts of discussion and the focus is on the other two phases that extract majority of lipids.

In the first 50 minutes of the gradient, the intensities of peaks are higher in the HFIP-rich phase which shows higher concentration of relatively less lipophilic lipids in this phase. Between 70 to 90 minutes, DCM-rich phase shows slightly higher intensities, and after 90 minutes, the intensity of lipids in the DCM-rich phase are considerably greater than the HFIP-rich phase. This trend shows the fractionation and enrichment of lipids between HFIP-rich and DCM-rich phase.
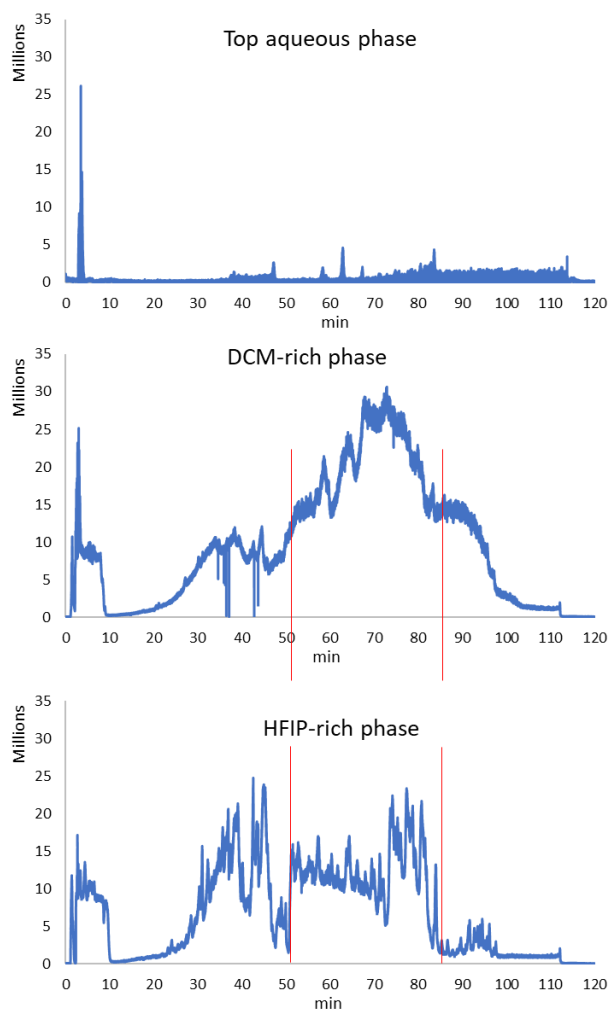
Figure 5-8. Chromatograms of lipid extracted by the least hydrophobic top aqueous phase, the most hydrophobic DCM-rich middle phase, and HFIP-rich bottom phase with intermediate hydrophobicity.

The database for partition coefficient of lipids in octanol/water system (Log P) can be found in http://www.vcclab.org/lab/alogps/ [46] where the values are predicted on the basis of neural network that provides reliable estimations of lipophilicity,[47,48] where lipid structures are represented by Simplified Molecular Line Entry Specification (SMILES)[49,50] format. It worth to note that using SMILES format to represent lipid structures is very accurate in terms of bond connectivity, valence and chirality, but the SMILES format does not include the orientation of the

structure.[51] Lipidomics scientists currently represent lipid structures based on their individual preferences, and hence a given lipid structure may be represented in quite different ways in different databases.[51] To brief, consistent method for lipid structure-drawing is not available currently. However, Log P is a powerful measurement method to predict hydrophobicity of lipids.

Comparison between the Log P of the lipids which are uniquely identified in each phase are provided in Figure 5-9. The medians of Log P values in DCM-rich and HFIP-rich phases are 8.4 and 6.3, respectively, which shows more than 2 orders of magnitude difference on the octanol/water partition coefficient scale.
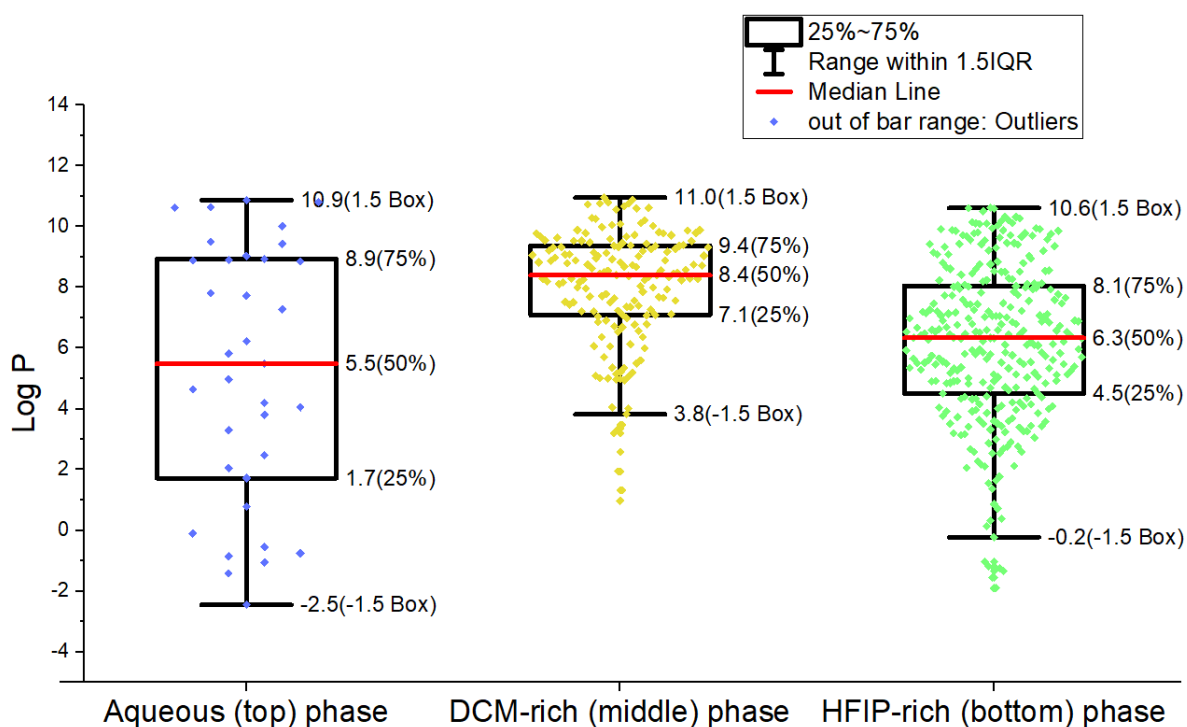
Figure 5-9. Distribution of Log P (partition coefficient in octanol/water system) of the lipids which are uniquely identified in each phase.

Another method to represent the lipids' hydrophobicity is aqueous solubility (Log S) and values can be found in http://www.vcclab.org/lab/alogps/,[46] and the distribution of Log S of unique lipids in each phase are depicted in Figure 5-10. Like Log P, these values are also calculated based

on SMILES format for molecular structure. Generally, the median of water solubility of uniquely identified lipids in the HFIP-rich phase is one order of magnitude larger than the lipids in the DCM-rich phase. Combination of RPLC, Log P, and Log S, show that lipids are generally extracted to the phases based on their lipophilicity.
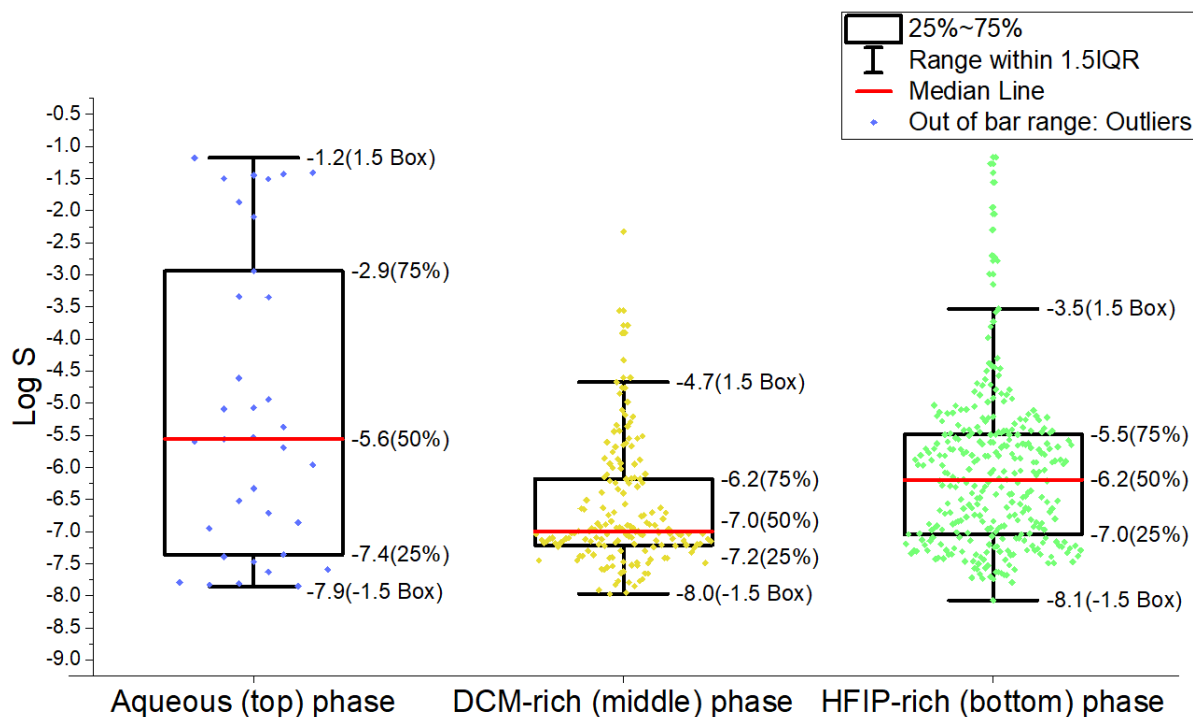
Figure 5-10. Distribution of Log S (aqueous solubility) of the lipids which are uniquely identified in each phase.

5.3.5 CHARACTERISTICS OF PHOSPHOLIPIDS IN EACH PHASE

Glycerophospholipids are the most abundant classes of lipids within the cell membranes.[52,53] Glycerophospholipids are consisted of a glycerol backbone, one or two fatty acyl chains that are linked at the sn1 and sn2 sites of the glycerol, and a polar head containing phosphate group.[11,14,15] The hydrophobic fatty acyl chains usually contain one or more Cis double bonds (C=C) that results in kinking in the tail which affects the 3D structure and fluidity of the bilayer.[54] The length of the fatty acyl chains also affect the fluidity of the bilayer.[55] Bilayers with shorter fatty acid tails have

less tendency to tightly pack together, therefore shorter chains increases the fluidity of the bilayer.[56] Therefore, physicochemical properties of glycerophospholipids describe the physicochemical properties of membrane lipid bilayers. The two main factors that affect the glycerophospholipids' properties, fatty acyl chain lengths and number of C=C bonds, are studied for the glycerophospholipids that are uniquely identified in each of the phases and results are presented in Figures 5-11 and 5-12. For the unique lipids extracted by each phase, Figures 5-11(a and b) respectively compare the distribution of the minimum number of carbons among the two fatty acyl chains, and the average number of carbons in the two chains.

Figure 5-12 reveals the following two characteristics that distinct the glycerophospholipids of the aqueous top phase; first, majority of them have at least one very short fatty acyl chain (C2 or C3), and second, they generally have multiple C=C double bonds in chain if the fatty acyl tail is long. Similarly, about 25% of the lipids in the less hydrophobic HFIP-rich (bottom) phase have at least one short chain length (C2 or C3). Figure 5-11(b) vividly depicts that more than 75% of the glycerophospholipids in the less hydrophobic HFIP-rich (bottom) phase have less than 17 carbons in their fatty acyl chains by average. On the other hand, more than 75% of the glycerophospholipids in the more hydrophobic DCM-rich (middle) phase averagely have more than 18 carbons in their fatty acyl chains. Therefore, it can be concluded that the number of carbons in the fatty acyl group of glycerophospholipids plays a crucial role in selectivity of phases towards phospholipids. However, fatty acyl's length is not the only factor that affects the hydrophobicity of phospholipids, and number of C=C double bonds affects hydrophobicity too.
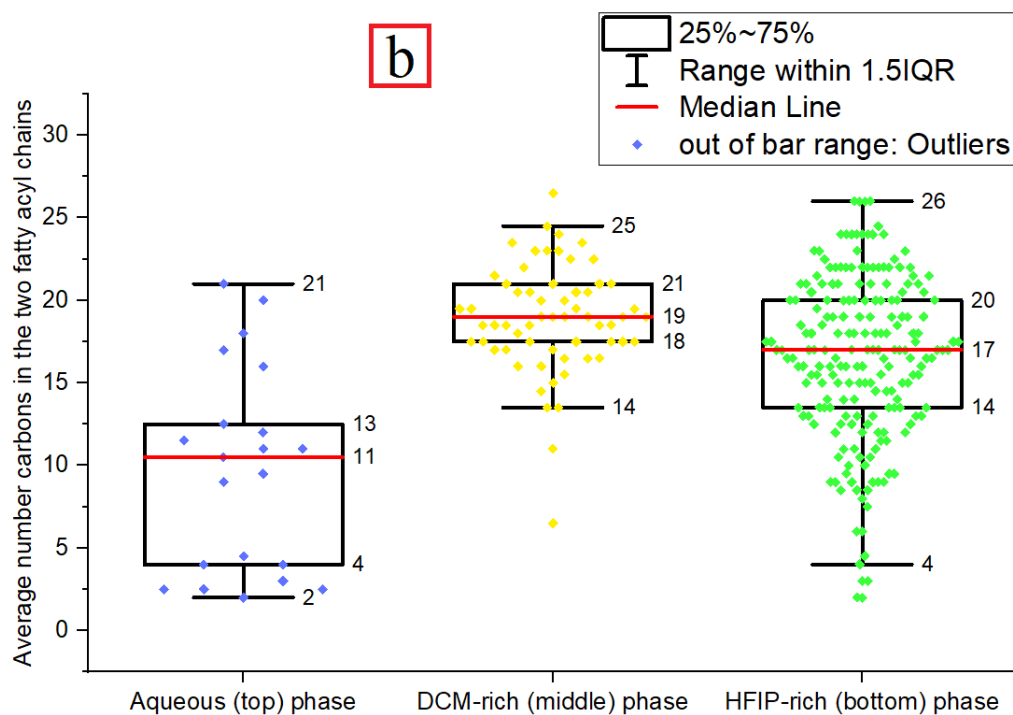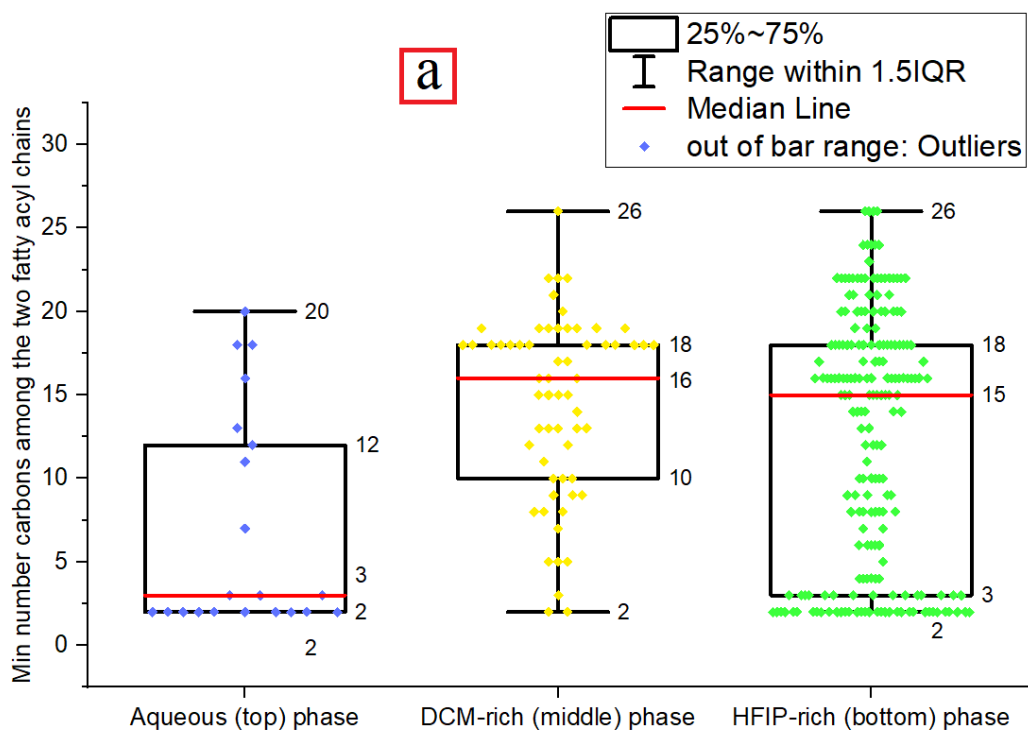
Figure 5-11. a) distribution of the minimum number of carbons among the two fatty acyl chains, and b) distribution of the average number of carbons in the two fatty acyl chains.

"Lipophilicity Index", being defined as $\frac{number\ of\ carbons\ on\ the\ fatty\ acyl\ chain}{number\ of\ C=C\ double\ bonds+1}$ can indicate the

relative hydrophobicity of fatty acyl group. Comparison between the Lipophilicity Index of the phospholipids in each phase is depicted in Figure 5-12, the trends show that generally as the hydrophobicity of phases increases, the Lipophilicity Index of the acyl chains increases too. However, Lipophilicity Index is not the only parameter that governs the distribution pattern of lipids. Overall, two major parameters control the distribution pattern, one is the headgroup, and the other one is lipophilicity of hydrophobic tails.
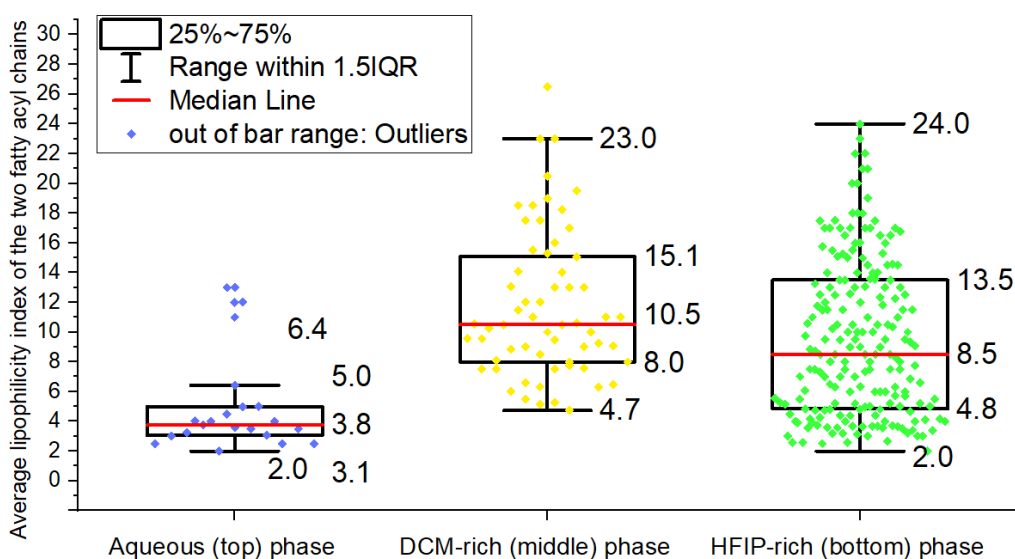
Figure 5-12. distribution of the "Lipophilicity Index" of the acyl chain groups of the lipids in each phase.

5.4 CONCLUSIONS

A wide range of hydrophobicity can be found in different phases of the DCM-HFIP-water multiphase system. Additionally, due to the different in chemical nature of phases, different chemical interactions between the phases and various classes of lipids would form. Combination of the hydrophobic effect and chemical interactions between lipids and phases would results in enrichment of lipids in different phases based on their physicochemical properties. This would

result in enhanced lipidome coverage. Very hydrophilic lipids with small retention times of about 3 minutes in the RP-LC were uniquely identified in the aqueous phase. More than 75% of the phospholipids in this phase had at least one short chain with 2-3 carbons in the acyl group. The HFIP-phase has a moderate hydrophobicity, between the aqueous phase and DCM-rich phase, and lipids with moderate hydrophobicity, such as PIs and Lyso phospholipids, get enriched in the HFIP-rich phase. About 50% of the phospholipids in the HFIP-rich phase had at least one short chain with 2-3 carbons in the acyl group. Finally, the most hydrophobic lipids with large hydrophobic tails generally get enriched in the DCM-rich as the most hydrophobic phase. In RPLC, the median of the retention time of lipids in the aqueous phase, HFIP-rich phase, and DCM-rich phases were 3.5, 37, and 59 minutes, respectively. This trend shows the enrichment and concentration of lipids in different phases based on the hydrophobic effect. Utilization of this simple step in lipidomics workflow would significantly improve lipidome coverage; in this case about 150% improvement was observed, even though the lipid extract in this case was extracted by chloroform/methanol system. Lipid extractions using a single solvent mixture is expected to be less efficient than two separate extraction steps, for example, combining methanol/chloroform extraction with methyl-tert-butyl ether (MTBE) extraction, will result in a more comprehensive lipid extract that has extracted lipids with a wider range of lipophilicity. Utilization of more comprehensive extraction methods and combining the extracted lipid mixture with this novel lipid fractionation technique with subsequent analyses under different LC separation gradients for different phases would result in a broader and more comprehensive information about the lipidome of biological samples. In addition, selection of various column combinations, for instance C4 for the more polar fractions in the aqueous top phase, C8 for the relatively less polar HFIP-rich phase, and C18 for more non-polar fractions in the DCM-rich phase would result in better separation of lipids, and consequently, more comprehensive lipidomics information. However, in this study, the

same LC column and gradient was used for different phases to compare the phases with the control

under identical conditions.

5.5 REFERENCES

(1)   Regulation of Energy Metabolism by Long-Chain Fatty Acids. *Prog. Lipid Res.* **2014**, *53*, 124–144. https://doi.org/10.1016/j.plipres.2013.12.001.

(2)   Shimizu, T. Lipid Mediators in Health and Disease: Enzymes and Receptors as Therapeutic Targets for the Regulation of Immunity and Inflammation. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 123–150. https://doi.org/10.1146/annurev.pharmtox.011008.145616.

(3)   Saliba, A.-E.; Vonkova, I.; Gavin, A.-C. The Systematic Analysis of Protein-Lipid Interactions Comes of Age. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (12), 753–761. https://doi.org/10.1038/nrm4080.

(4)   Resh, M. D. Fatty Acylation of Proteins: The Long and the Short of It. *Prog. Lipid Res.* **2016**, *63*, 120–131. https://doi.org/10.1016/j.plipres.2016.05.002.

(5)   Yamashita, A.; Hayashi, Y.; Nemoto-Sasaki, Y.; Ito, M.; Oka, S.; Tanikawa, T.; Waku, K.; Sugiura, T. Acyltransferases and Transacylases That Determine the Fatty Acid Composition of Glycerolipids and the Metabolism of Bioactive Lipid Mediators in Mammalian Cells and Model Organisms. *Prog. Lipid Res.* **2014**, *53*, 18–81. https://doi.org/10.1016/j.plipres.2013.10.001.

(6)   Harayama, T.; Eto, M.; Shindou, H.; Kita, Y.; Otsubo, E.; Hishikawa, D.; Ishii, S.; Sakimura, K.; Mishina, M.; Shimizu, T. Lysophospholipid Acyltransferases Mediate Phosphatidylcholine Diversification to Achieve the Physical Properties Required in Vivo. *Cell Metab.* **2014**, *20* (2), 295–305. https://doi.org/10.1016/j.cmet.2014.05.019.

(7)   Grösch, S.; Schiffmann, S.; Geisslinger, G. Chain Length-Specific Properties of Ceramides. *Prog. Lipid Res.* **2012**, *51* (1), 50–62. https://doi.org/10.1016/j.plipres.2011.11.001.

(8)   Harayama, T.; Riezman, H. Understanding the Diversity of Membrane Lipid Composition. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (5), 281–296. https://doi.org/10.1038/nrm.2017.138.

(9)   Dobrzyńska, I.; Szachowicz-Petelska, B.; Sulkowski, S.; Figaszewski, Z. Changes in Electric Charge and Phospholipids Composition in Human Colorectal Cancer Cells. *Mol. Cell. Biochem.* **2005**, *276* (1), 113–119. https://doi.org/10.1007/s11010-005-3557-3.

(10) Vozella, V.; Basit, A.; Misto, A.; Piomelli, D. Age-Dependent Changes in Nervonic Acid-Containing Sphingolipids in Mouse Hippocampus. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **2017**, *1862* (12), 1502–1511. https://doi.org/10.1016/j.bbalip.2017.08.008.

(11) Whiley, L.; Sen, A.; Heaton, J.; Proitsi, P.; García-Gómez, D.; Leung, R.; Smith, N.; Thambisetty, M.; Kloszewska, I.; Mecocci, P.; Soininen, H.; Tsolaki, M.; Vellas, B.; Lovestone, S.; Legido-Quigley, C. Evidence of Altered Phosphatidylcholine Metabolism in Alzheimer's Disease. *Neurobiol. Aging* **2014**, *35* (2), 271–278. https://doi.org/10.1016/j.neurobiolaging.2013.08.001.

(12) Yin, F.-W.; Zhou, D.-Y.; Zhao, Q.; Liu, Z.-Y.; Hu, X.-P.; Liu, Y.-F.; Song, L.; Zhou, X.; Qin, L.; Zhu, B.-W.; Shahidi, F. Identification of Glycerophospholipid Molecular Species of Mussel (Mytilus Edulis) Lipids by High-Performance Liquid Chromatography-Electrospray Ionization-Tandem Mass Spectrometry. *Food Chem.* **2016**, *213*, 344–351. https://doi.org/10.1016/j.foodchem.2016.06.094.

(13) Montealegre, C.; Verardo, V.; Luisa Marina, M.; Caboni, M. F. Analysis of Glycerophospho- and Sphingolipids by CE. *Electrophoresis* **2014**, *35* (6), 779–792. https://doi.org/10.1002/elps.201300534.

(14) Phoenix, D. A.; Harris, F.; Mura, M.; Dennison, S. R. The Increasing Role of Phosphatidylethanolamine as a Lipid Receptor in the Action of Host Defence Peptides. *Prog. Lipid Res.* **2015**, *59*, 26–37. https://doi.org/10.1016/j.plipres.2015.02.003.

(15) Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R. H.; Shimizu, T.; Spener, F.; Meer, G. van; Wakelam, M. J. O.; Dennis, E. A. Update of the LIPID MAPS Comprehensive Classification System for Lipids. *J. Lipid Res.* **2009**, *50* (Supplement), S9–S14. https://doi.org/10.1194/jlr.R800095-JLR200.

(16) Croset, M.; Brossard, N.; Polette, A.; Lagarde, M. Characterization of Plasma Unsaturated Lysophosphatidylcholines in Human and Rat. *Biochem. J.* **2000**, *345* (Pt 1), 61–67.

(17) Diagne, A.; Fauvel, J.; Record, M.; Chap, H.; Douste-Blazy, L. Studies on Ether Phospholipids: II. Comparative Composition of Various Tissues from Human, Rat and Guinea Pig. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* **1984**, *793* (2), 221–231. https://doi.org/10.1016/0005-2760(84)90324-2.

(18) da Silva, T. F.; Sousa, V. F.; Malheiro, A. R.; Brites, P. The Importance of Ether-Phospholipids: A View from the Perspective of Mouse Models. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **2012**, *1822* (9), 1501–1508. https://doi.org/10.1016/j.bbadis.2012.05.014.

(19) Paltauf, F. Ether Lipids in Biomembranes. *Chem. Phys. Lipids* **1994**, *74* (2), 101–139. https://doi.org/10.1016/0009-3084(94)90054-X.

(20) Davis, D.; Kannan, M.; Wattenberg, B. Orm/ORMDL Proteins: Gate Guardians and Master Regulators. *Adv. Biol. Regul.* **2018**, *70*, 3–18. https://doi.org/10.1016/j.jbior.2018.08.002.

(21) van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane Lipids: Where They Are and How They Behave. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (2), 112–124. https://doi.org/10.1038/nrm2330.

(22) Ma, Y.; Kind, T.; Vaniya, A.; Gennity, I.; Fahrmann, J. F.; Fiehn, O. An in Silico MS/MS Library for Automatic Annotation of Novel FAHFA Lipids. *J. Cheminformatics* **2015**, *7* (1), 53. https://doi.org/10.1186/s13321-015-0104-4.

(23) Tan, D.; Ertunc, M. E.; Konduri, S.; Zhang, J.; Pinto, A. M.; Chu, Q.; Kahn, B. B.; Siegel, D.; Saghatelian, A. Discovery of FAHFA-Containing Triacylglycerols and Their Metabolic Regulation. *J. Am. Chem. Soc.* **2019**, *141* (22), 8798–8806. https://doi.org/10.1021/jacs.9b00045.

(24) Buenger, E. W.; Reid, G. E. Shedding Light on Isomeric FAHFA Lipid Structures Using 213 Nm Ultraviolet Photodissociation Mass Spectrometry. *Eur. J. Mass Spectrom.* **2020**, *26* (5), 311–323. https://doi.org/10.1177/1469066720960341.

(25) Smith, U.; Kahn, B. B. Adipose Tissue Regulates Insulin Sensitivity: Role of Adipogenesis, de Novo Lipogenesis and Novel Lipids. *J. Intern. Med.* **2016**, *280* (5), 465–475. https://doi.org/10.1111/joim.12540.

(26) Ding, J.; Kind, T.; Zhu, Q.-F.; Wang, Y.; Yan, J.-W.; Fiehn, O.; Feng, Y.-Q. In-Silico-Generated Library for Sensitive Detection of 2-Dimethylaminoethylamine Derivatized FAHFA Lipids Using High-Resolution Tandem Mass Spectrometry. *Anal. Chem.* **2020**, *92* (8), 5960–5968. https://doi.org/10.1021/acs.analchem.0c00172.

(27) Seiwert, D.; Witt, H.; Janshoff, A.; Paulsen, H. The Non-Bilayer Lipid MGDG Stabilizes the Major Light-Harvesting Complex (LHCII) against Unfolding. *Sci. Rep.* **2017**, *7* (1), 5158. https://doi.org/10.1038/s41598-017-05328-7.

(28) Garab, G.; Ughy, B.; Goss, R. Role of MGDG and Non-Bilayer Lipid Phases in the Structure and Dynamics of Chloroplast Thylakoid Membranes. In *Lipids in Plant and Algae Development*; Nakamura, Y., Li-Beisson, Y., Eds.; Subcellular Biochemistry; Springer International Publishing: Cham, 2016; pp 127–157. https://doi.org/10.1007/978-3-319-25979-6_6.

(29) Seiwert, D.; Witt, H.; Ritz, S.; Janshoff, A.; Paulsen, H. The Nonbilayer Lipid MGDG and the Major Light-Harvesting Complex (LHCII) Promote Membrane Stacking in Supported Lipid Bilayers. *Biochemistry* **2018**, *57* (15), 2278–2288. https://doi.org/10.1021/acs.biochem.8b00118.

(30) Oresti, G. M.; Reyes, J. G.; Luquez, J. M.; Osses, N.; Furland, N. E.; Aveldaño, M. I. Differentiation-Related Changes in Lipid Classes with Long-Chain and Very Long-Chain Polyenoic Fatty Acids in Rat Spermatogenic Cells. *J. Lipid Res.* **2010**, *51* (10), 2909–2921. https://doi.org/10.1194/jlr.M006429.

(31) Koolivand, A.; Azizi, M.; O'Brien, A.; Khaledi, M. G. Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies. *J. Proteome Res.* **2019**, *18* (4), 1595–1606. https://doi.org/10.1021/acs.jproteome.8b00857.

(32) McCord, J. P.; Muddiman, D. C.; Khaledi, M. G. Perfluorinated Alcohol Induced Coacervates as Extraction Media for Proteomic Analysis. *Push. Boundaries Chromatogr. Electrophor.* **2017**, *1523* (Supplement C), 293–299. https://doi.org/10.1016/j.chroma.2017.06.025.

(33) Koolivand, A.; Clayton, S.; Rion, H.; Oloumi, A.; O'Brien, A.; Khaledi, M. G. Fluoroalcohol – Induced Coacervates for Selective Enrichment and Extraction of Hydrophobic Proteins. *J. Chromatogr. B* **2018**, *1083*, 180–188. https://doi.org/10.1016/j.jchromb.2018.03.004.

(34) Phillips, K. M.; Tarragó-Trani, M. T.; Grove, T. M.; Grün, I.; Lugogo, R.; Harris, R. F.; Stewart, K. K. Simplified Gravimetric Determination of Total Fat in Food Composites after Chloroform-Methanol Extraction. *J. Am. Oil Chem. Soc.* **1997**, *74* (2), 137–142. https://doi.org/10.1007/s11746-997-0158-1.

(35) Daugherty, C. E.; Lento, H. G.; Collaborators; Adams, M. L.; Beckert, E. W.; Bender, M. L.; Berman, S.; Chow, L.; Davis, C.; Gedang, D.; Howe, K.; Murphy, M. J.; Porcuna, M.; Sabolish, G.; Shen, C.-S. J.; Smith, N. M.; Tessaro, A. Chloroform-Methanol Extraction Method for Determination of Fat in Foods: Collaborative Study. *J. AOAC Int.* **1983**, *66* (4), 927–932. https://doi.org/10.1093/jaoac/66.4.927.

(36) Folch, J.; Lees, M.; Stanley, G. H. S. A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. *J. Biol. Chem.* **1957**, *226* (1), 497–509.

(37) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. Lipid Extraction by Methyl-Tert-Butyl Ether for High-Throughput Lipidomics. *J. Lipid Res.* **2008**, *49* (5), 1137–1146. https://doi.org/10.1194/jlr.D700041-JLR200.

(38) Chen, S.; Hoene, M.; Li, J.; Li, Y.; Zhao, X.; Häring, H.-U.; Schleicher, E. D.; Weigert, C.; Xu, G.; Lehmann, R. Simultaneous Extraction of Metabolome and Lipidome with Methyl Tert-Butyl Ether from a Single Small Tissue Sample for Ultra-High Performance Liquid Chromatography/Mass Spectrometry. *J. Chromatogr. A* **2013**, *1298*, 9–16. https://doi.org/10.1016/j.chroma.2013.05.019.

(39) Navarro-Reig, M.; Jaumot, J.; Tauler, R. An Untargeted Lipidomic Strategy Combining Comprehensive Two-Dimensional Liquid Chromatography and Chemometric Analysis. *J. Chromatogr. A* **2018**, *1568*, 80–90. https://doi.org/10.1016/j.chroma.2018.07.017.

(40) Sethi, S.; Brietzke, E. Recent Advances in Lipidomics: Analytical and Clinical Perspectives. *Prostaglandins Other Lipid Mediat.* **2017**, *128–129*, 8–16. https://doi.org/10.1016/j.prostaglandins.2016.12.002.

(41) Jenkins, S. I.; Collins, C. M.; Khaledi, M. G. Perfluorinated Alcohols Induce Complex Coacervation in Mixed Surfactants. *Langmuir* **2016**, *32* (10), 2321–2330. https://doi.org/10.1021/acs.langmuir.5b04701.

(42) Nejati, M. M.; Khaledi, M. G. Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis. *Langmuir* **2015**, *31* (20), 5580–5589. https://doi.org/10.1021/acs.langmuir.5b00444.

(43) Weisner, N.; Khaledi, M. G. Organic Synthesis in Fluoroalcohol-Water Two-Phase Systems. *Green Chem.* **2016**, *18* (3), 681–685. https://doi.org/10.1039/C5GC01463H.

(44) Solvents and Solvent Effects in Organic Chemistry, 4th, Updated and Enlarged Edition | Wiley https://www.wiley.com/en-us/Solvents+and+Solvent+Effects+in+Organic+Chemistry%2C+4th%2C+Updated+and+Enlarged+Edition-p-9783527324736 (accessed Oct 25, 2020).

(45) Maffucci, T.; Falasca, M. Chapter Five - Analysis, Regulation, and Roles of Endosomal Phosphoinositides. In *Methods in Enzymology*; Conn, P. M., Ed.; Endosome Signaling Part B; Academic Press, 2014; Vol. 535, pp 75–91. https://doi.org/10.1016/B978-0-12-397925-4.00005-5.

(46) On-line Lipophilicity/Aqueous Solubility Calculation Software http://www.vcclab.org/lab/alogps/ (accessed Oct 16, 2020).

(47) Tetko, I. V.; Tanchuk, V. Yu.; Villa, A. E. P. Prediction of N-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (5), 1407–1421. https://doi.org/10.1021/ci010368v.

(48) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull. (Tokyo)* **1992**, *40* (1), 127–130. https://doi.org/10.1248/cpb.40.127.

(49) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36.

(50) Fahy, E.; Cotter, D.; Sud, M.; Subramaniam, S. Lipid Classification, Structures and Tools. *Biochim. Biophys. Acta* **2011**, *1811* (11), 637–647. https://doi.org/10.1016/j.bbalip.2011.06.009.

(51) Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S. LIPID MAPS Online Tools for Lipid Research. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W606–W612. https://doi.org/10.1093/nar/gkm324.

(52) Han, X.; Gross, R. W. Shotgun Lipidomics: Electrospray Ionization Mass Spectrometric Analysis and Quantitation of Cellular Lipidomes Directly from Crude Extracts of Biological Samples. *Mass Spectrom. Rev.* **2005**, *24* (3), 367–412. https://doi.org/10.1002/mas.20023.

(53) Nicolaou, A.; Masoodi, M.; Mir, A. Lipidomic Analysis of Prostanoids by Liquid Chromatography–Electrospray Tandem Mass Spectrometry. In *Lipidomics: Volume 1: Methods and Protocols*; Armstrong, D., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2009; pp 271–286. https://doi.org/10.1007/978-1-60761-322-0_14.

(54) Brenner, R. R. Effect of Unsaturated Acids on Membrane Structure and Enzyme Kinetics. *Prog. Lipid Res.* **1984**, *23* (2), 69–96. https://doi.org/10.1016/0163-7827(84)90008-0.

(55) Poger, D.; Caron, B.; Mark, A. E. Effect of Methyl-Branched Fatty Acids on the Structure of Lipid Bilayers. *J. Phys. Chem. B* **2014**, *118* (48), 13838–13848. https://doi.org/10.1021/jp503910r.

(56) Sohlenkamp, C.; Geiger, O. Bacterial Membrane Lipids: Diversity in Structures and Pathways. *FEMS Microbiol. Rev.* **2016**, *40* (1), 133–159. https://doi.org/10.1093/femsre/fuv008.

CHAPTER 6

SUMMERY AND PERSPECTIVES

By using FA*i*C-BPS in bottom-up proteomics, proteins were fractionated between the coacervate and aqueous phases based on their physicochemical properties, such as isoelectric points and hydrophobicity. At lower abundance ranges, low-abundance proteins are generally enriched in one of the phases, and more fractionation happens in the FA*i*C-BPS. This enrichment results in enhanced identification of the low-abundance proteins. The trend of identification improvement versus abundance range shows that in all FA*i*C-BPS of this study, higher identification improvement happens at lower abundance ranges. Especially, identification of hydrophobic proteins such as IMP was improved, and better sequence coverage of alpha-helical transmembrane peptides was achieved in most FA*i*C-BPS. In addition to the hydrophobic effects in the coacervate phases due to the high concentration of amphiphiles and fluoroalcohols, the other driving force in the FA*i*C-BPS is electrostatic interaction between the proteins and charged amphiphiles in the coacervate phases. By changing the pH of systems, protein net charge would change that would impact the electrostatic interaction, and subsequently proteins fractionation patterns.

Similar to proteomics, lipid identification was enhanced by taking the advantage of a simple fractionation step prior to the LC-MS analysis. Lipid were fractionated in the water-DCM-HFIP multiphase system due to having a wide range of hydrophobicities in different phases.

Combination of the hydrophobic effect and chemical interactions between lipids and phases results in enrichment of lipids in different phases based on their physicochemical properties. This fractionation and enrichment resulted in enhanced lipidome coverage in LCMS-based lipidomics.

APPENDIX 2-1

SAMPLE PREPARATION WORKFLOW

SOLUTIONS

- 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)

- 450 mM TBAB: add 7.253 g tetrabutylammonium bromide in a 50 mL volumetric flask and bring the volume to 50 mL by adding water.

FASP PROTOCOL FOR COACERVATION WITH 8% HFIP AT 50 MM TBAB:

A) COACERVATION

1- Take 400 µg protein (for example, if the concentration of cell lysate is about 8 mg/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
- 111 µL 450 mM TBAB
- DI water: 1000 µL – 111 µL (Vol. TBAB Sol.) – 80 µL (Vol. of HFIP) – Vol. of cell lysate (µL)

- 80 μL HFIP

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

Note: measure the protein concentration in each phase before loading the sample to the filters. You will need the amount of protein in each phase when you want to add trypsin with the ratio of 1:25.

Note: For protein concentration measurement, because HFIP has interference with Bradford Assay, evaporate the HFIP prior to the protein measurement (measure the protein concentration in step B-3 for the coacervate phase and at the end of step C-1 for the aqueous phase).

B) FOR THE COACERVATE PHASE:

1- Dry the coacervate with nitrogen gas for about 1 min to get rid of HFIP (not completely dried, adjust the flow of nitrogen to prevent drip).

2- 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.

3- Add 450 μL 70% IPA to the coacervate, then add 76 mg thiourea to dissolve it. Vortex 30 sec, then sonicate 5 min at room temperature (not in ice, because it does not dissolve at low temperatures), and load the dissolved coacervate to the to the pre-conditioned FASP filters. Take a small amount of this solution for protein concentration measurement.

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

5- 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)

Note: 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.

6- 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)

7- 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

124

C) FOR THE AQUEOUS:

1- Put the Aqueous phase in concentrator to evaporate HFIP for about 1.5 hours, until the volume reaches to about 500 µL. At this point, majority of HFIP is evaporated (BP: 58 °C) and you can measure protein concentration.

2- 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.

3- 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:

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

2- 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.

3- Cool the sample to room temperature.

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

5- 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).

6- 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.

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

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

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

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

11- 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

(in the case of our study, for the aqueous: (250/25)*2= 20 µL trypsin; and for coacervate: (150/25)*2= 12 µL trypsin. These values may vary in other studies because the cell lysate is different.)

12- 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.

13- Transfer the filters to new collection tubes.

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

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

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

17- Acidify the sample with TFA to bring the pH below 2 (about 2-3 µL TFA is enough, DO NOT over-acidify)

18- Desalt the samples.

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 ACN, 1 mL 0.1 TFA in 75% ACN, 1 mL 0.1 TFA in 50% ACN, 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% 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.

APPENDIX 3-1

IDENTIFICATION IMPROVEMENT VS. ABUNDANCE OF PROTEINS IN
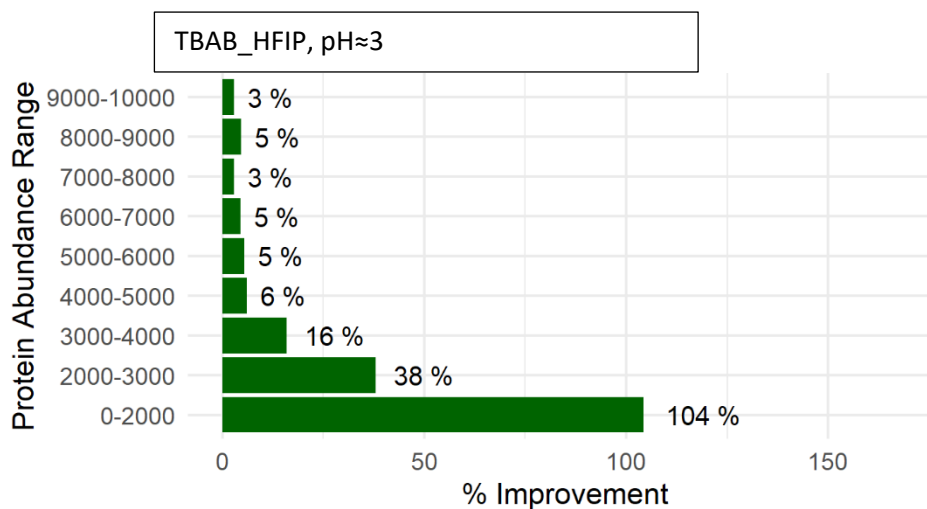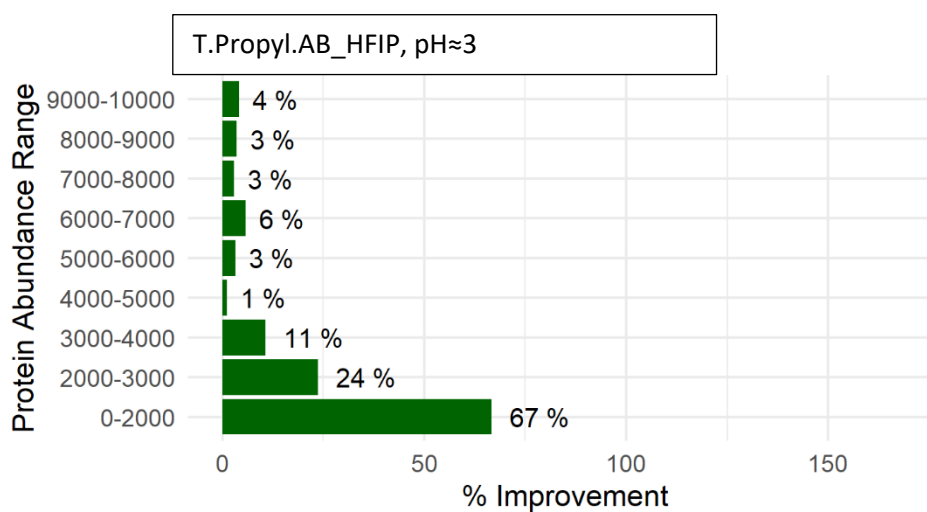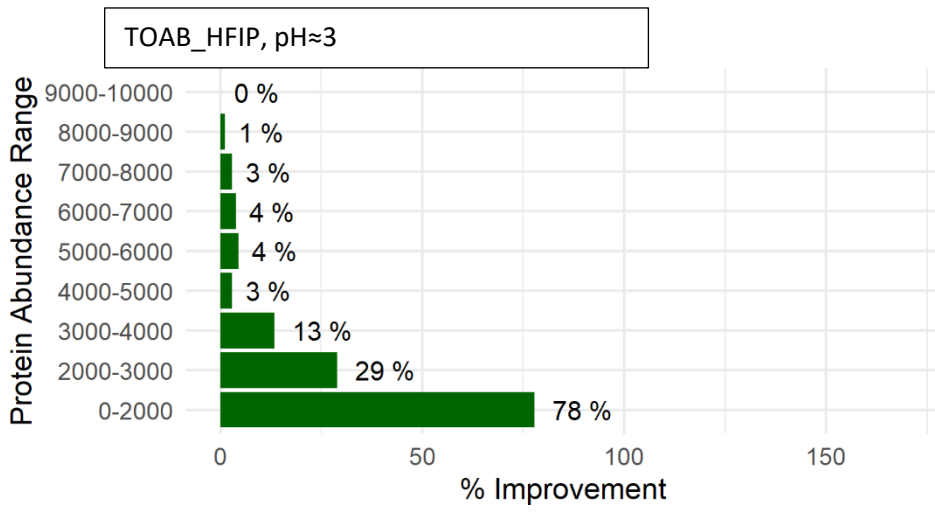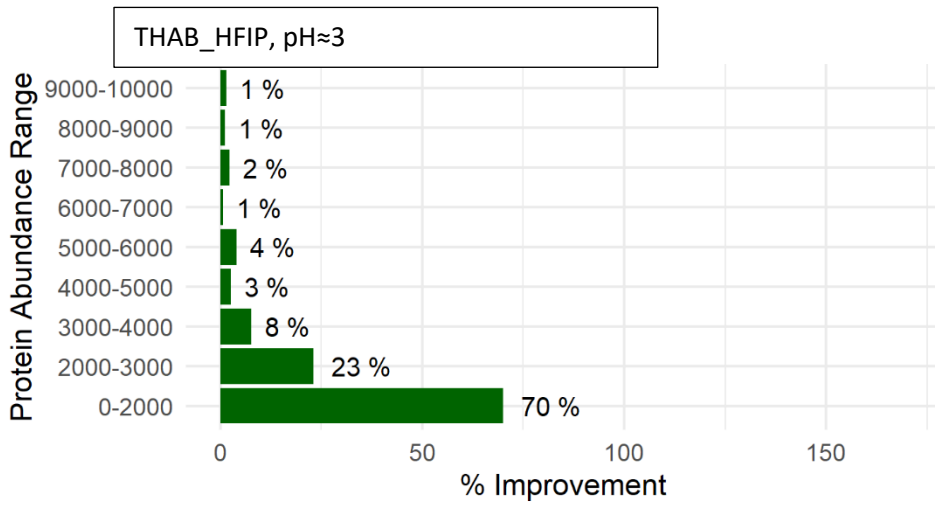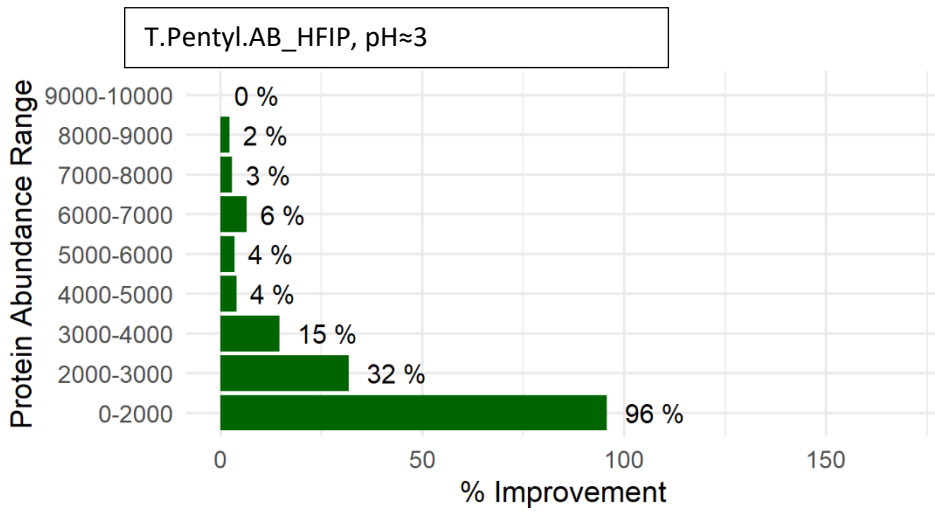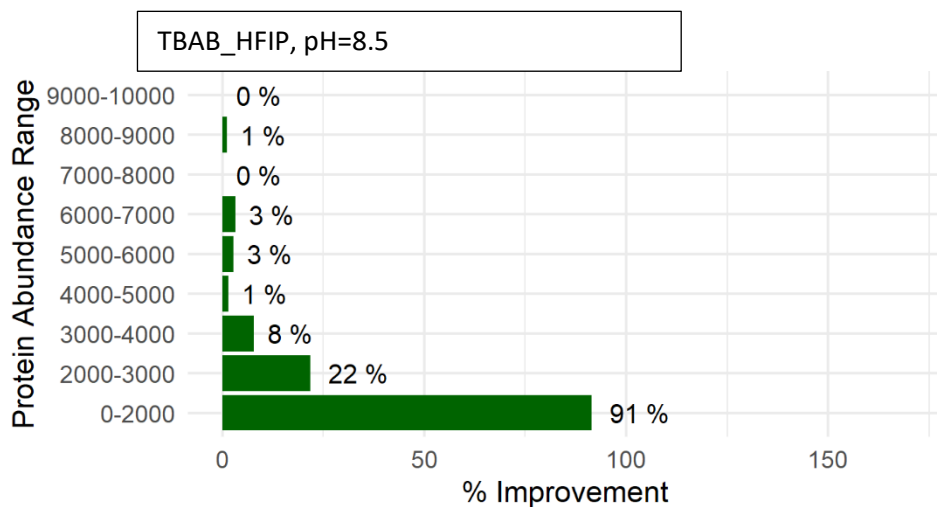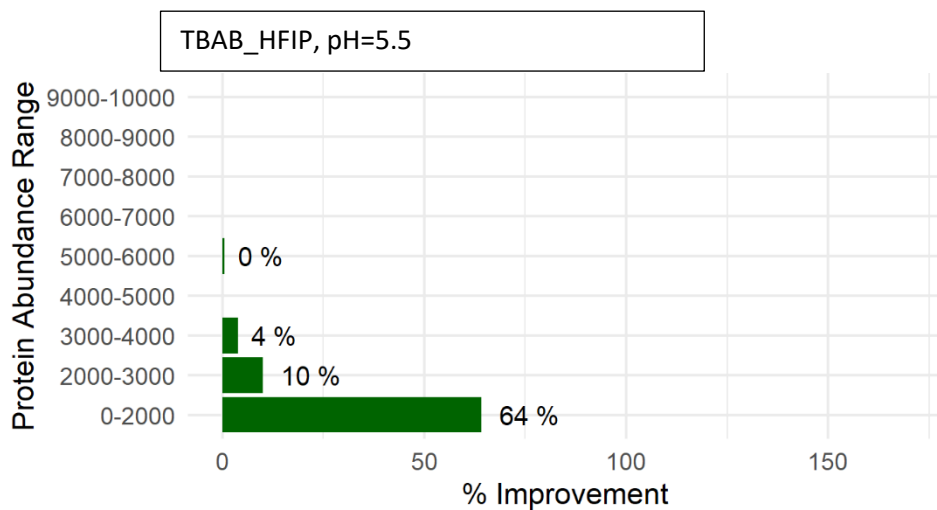
DIFFERENT FA*i*C-BPS

APPENDIX 3-1

IDENTIFICATION IMPROVEMENT VS. ABUNDANCE OF PROTEINS IN

DIFFERENT FA*i*C-BPS

T.Pentyl.AB_HFIP, pH≈3
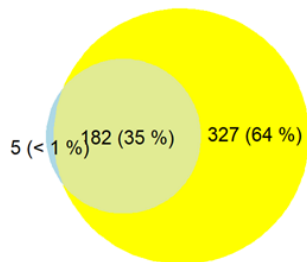

THAB_HFIP, pH≈3
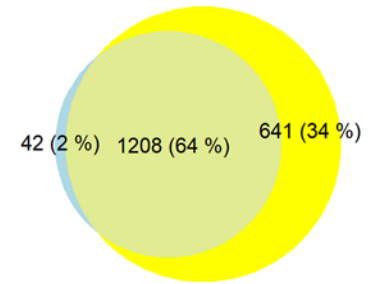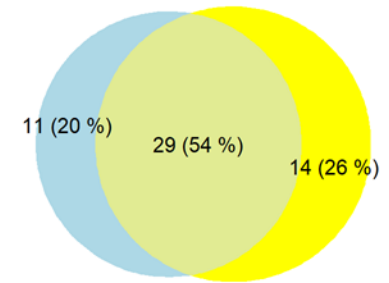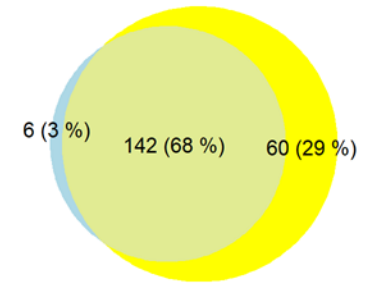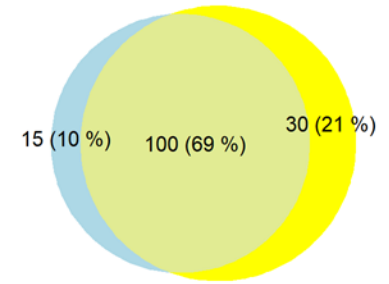
TOAB_HFIP, pH≈3
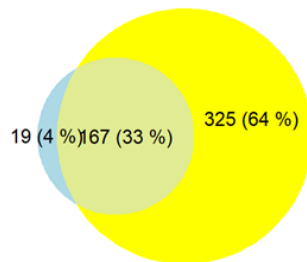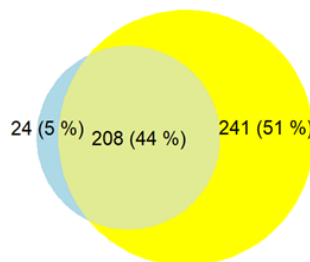
TBAB_HFIP, pH=5.5

TBAB_HFIP, pH=8.5

Integral membrane proteins

T.Propyl.AB_HFIP, pH≈3

29 (6 %)   259 (54 %)   195 (40 %)

- Aqueous Phase
- Organic Phase

TBAB _HFIP, pH≈3

50 (10 %)   236 (46 %)   228 (44 %)

- Aqueous Phase
- Organic Phase

T.Pentyl.AB _HFIP, pH≈3

10 (2 %)   200 (39 %)   307 (59 %)

- Aqueous Phase
- Organic Phase

THAB _HFIP, pH≈3

10 (2 %)   272 (55 %)   209 (43 %)

- Aqueous Phase
- Organic Phase

TOAB _HFIP, pH≈3

16 (3 %)   218 (44 %)   257 (52 %)

- Aqueous Phase
- Organic Phase

TBAB _HFIP, pH=5.5

19 (4 %)   111 (24 %)   333 (72 %)

- Aqueous Phase
- Organic Phase

TBAB _HFIP, pH=8.5

18 (4 %)   131 (26 %)   348 (70 %)

- Aqueous Phase
- Organic Phase

- **Aqueous phase**
- **Amphiphile-rich phase**

Membrane proteins

T.Propyl.AB_HFIP, pH≈3

64 (6 %) 601 (61 %) 327 (33 %)

TBAB _HFIP, pH≈3

89 (9 %) 565 (54 %) 384 (37 %)

T.Pentyl.AB _HFIP, pH≈3

34 (3 %) 509 (49 %) 495 (48 %)

THAB _HFIP, pH≈3

22 (2 %) 596 (61 %) 359 (37 %)

TOAB _HFIP, pH≈3

53 (5 %) 557 (56 %) 391 (39 %)

TBAB _HFIP, pH=5.5

44 (5 %) 353 (37 %) 551 (58 %)

TBAB _HFIP, pH=8.5

46 (5 %) 384 (38 %) 576 (57 %)

Aqueous phase

Amphiphile-rich phase

133

Ubiquitinylated lyside residue proteins

T.Propyl.AB_HFIP, pH≈3

59 (4 %)   1060 (75 %)   290 (21 %)

TBAB _HFIP, pH≈3

48 (3 %)   991 (69 %)   402 (28 %)

T.Pentyl.AB _HFIP, pH≈3

27 (2 %)   942 (65 %)   472 (33 %)

THAB _HFIP, pH≈3

361 (26 %)

20 (1 %)   1019 (73 %)

TOAB _HFIP, pH≈3

58 (4 %)   1038 (73 %)   324 (23 %)

TBAB _HFIP, pH=5.5

54 (4 %)   762 (55 %)   575 (41 %)

TBAB _HFIP, pH=8.5

56 (4 %)   840 (59 %)   532 (37 %)

■ Aqueous phase

■ Amphiphile-rich phase

Phosphorylated residue proteins

T.Propyl.AB_HFIP, pH≈3

137 (7 %)   1289 (68 %)   460 (24 %)

TBAB _HFIP, pH≈3

132 (7 %)   1201 (60 %)   653 (33 %)

T.Pentyl.AB _HFIP, pH≈3

66 (3 %)   1135 (58 %)   758 (39 %)

THAB _HFIP, pH≈3

42 (2 %)   1208 (64 %)   641 (34 %)

TOAB _HFIP, pH≈3

153 (8 %)   1292 (66 %)   515 (26 %)

TBAB _HFIP, pH=5.5

126 (7 %)   912 (49 %)   839 (45 %)

TBAB _HFIP, pH=8.5

126 (6 %)   1003 (51 %)   826 (42 %)

- Aqueous phase
- Amphiphile-rich phase

Mitochondrial ribosome proteins

T.Propyl.AB_HFIP, pH≈3

30 (50 %) 22 (37 %) 8 (13 %)

TBAB _HFIP, pH≈3

45 (66 %) 20 (29 %) 3 (4 %)

T.Pentyl.AB _HFIP, pH≈3

22 (37 %) 26 (43 %) 12 (20 %)

THAB _HFIP, pH≈3

11 (20 %) 29 (54 %) 14 (26 %)

TOAB _HFIP, pH≈3

21 (36 %) 29 (49 %) 9 (15 %)

TBAB _HFIP, pH=5.5

2 (6 %) 4 (11 %) 30 (83 %)

TBAB _HFIP, pH=8.5

3 (6 %) 45 (94 %)

- ■ **Aqueous phase**
- ■ **Amphiphile-rich phase**

Mitochondrial Membrane

T.Propyl.AB_HFIP, pH≈3

21 (10 %) 128 (59 %) 68 (31 %)

TBAB _HFIP, pH≈3

32 (14 %) 131 (58 %) 63 (28 %)

T.Pentyl.AB _HFIP, pH≈3

10 (5 %) 117 (53 %) 94 (43 %)

THAB _HFIP, pH≈3

6 (3 %) 142 (68 %) 60 (29 %)

TOAB _HFIP, pH≈3

17 (8 %) 132 (62 %) 65 (30 %)

TBAB _HFIP, pH=5.5

6 (3 %) 60 (30 %) 135 (67 %)

TBAB _HFIP, pH=8.5

8 (4 %) 59 (29 %) 136 (67 %)

Aqueous phase

Amphiphile-rich phase

Mitochondrial Matrix

T.Propyl.AB_HFIP, pH≈3

37 (24 %)   91 (59 %)   26 (17 %)

TBAB _HFIP, pH≈3

55 (33 %)   91 (55 %)   19 (12 %)

T.Pentyl.AB _HFIP, pH≈3

27 (18 %)   93 (62 %)   30 (20 %)

THAB _HFIP, pH≈3

15 (10 %)   100 (69 %)   30 (21 %)

TOAB _HFIP, pH≈3

26 (17 %)   100 (67 %)   23 (15 %)

TBAB _HFIP, pH=5.5

4 (3 %)   52 (43 %)   66 (54 %)

TBAB _HFIP, pH=8.5

1 (< 1 %)   49 (35 %)   89 (64 %)

| | Aqueous phase |
| | Amphiphile-rich phase |

Golgi Apparatus proteins

T.Propyl.AB_HFIP, pH≈3

5 (3 %)   118 (61 %)   70 (36 %)

TBAB _HFIP, pH≈3

8 (4 %)   113 (54 %)   89 (42 %)

T.Pentyl.AB _HFIP, pH≈3

4 (2 %)   103 (49 %)   102 (49 %)

THAB _HFIP, pH≈3

2 (< 1 %)   121 (60 %)   79 (39 %)

TOAB _HFIP, pH≈3

5 (2 %)   113 (54 %)   91 (44 %)

TBAB _HFIP, pH=5.5

5 (3 %)   67 (34 %)   125 (63 %)

TBAB _HFIP, pH=8.5

5 (2 %)   79 (39 %)   119 (59 %)

Aqueous phase

Amphiphile-rich phase

139

Endosome proteins

T.Propyl.AB_HFIP, pH≈3

1 (< 1 %)   59 (57 %)   43 (42 %)

TBAB _HFIP, pH≈3

7 (6 %)   53 (46 %)   54 (47 %)

T.Pentyl.AB _HFIP, pH≈3

4 (3 %)   52 (44 %)   63 (53 %)

THAB _HFIP, pH≈3

5 (4 %)   51 (46 %)   56 (50 %)

TOAB _HFIP, pH≈3

6 (5 %)   68 (62 %)   36 (33 %)

TBAB _HFIP, pH=5.5

7 (6 %)   44 (40 %)   59 (54 %)

TBAB _HFIP, pH=8.5

12 (10 %)   51 (42 %)   57 (48 %)

Aqueous phase

Amphiphile-rich phase

Chromosome proteins

T.Propyl.AB_HFIP, pH≈3

23 (9 %) 141 (56 %) 86 (34 %)

TBAB _HFIP, pH≈3

18 (7 %) 112 (42 %) 134 (51 %)

T.Pentyl.AB _HFIP, pH≈3

7 (3 %) 113 (44 %) 135 (53 %)

THAB _HFIP, pH≈3

7 (3 %) 112 (45 %) 131 (52 %)

TOAB _HFIP, pH≈3

29 (11 %) 138 (53 %) 94 (36 %)

TBAB _HFIP, pH=5.5

25 (10 %) 78 (32 %) 143 (58 %)

TBAB _HFIP, pH=8.5

24 (9 %) 94 (37 %) 139 (54 %)

Aqueous phase

Amphiphile-rich phase

141

APPENDIX 3-3

(a): chromatogram of separation of alkylphenones in the aqueous phase of T.propyl.AB_HFIP biphasic system where peaks 1, 2 and 3 correspond with acetophenone, propiophenone, and butyrophenone, respectively; (b): chromatogram of separation of alkylphenones in the coacervate phase of T.propyl.AB_HFIP biphasic system; (c)-(f): logarithm of partition coefficient of homologous series of alkylphenones in FAiC-BPS vs. carbon numbers in alkyl groups of QUATS of FAiC-BPS in (c): T.Propyl.AB_HFIP BPS, (d): TBAB_HFIP BPS, (e): T.Pentyl.AB_HFIP BPS, and (f): THAB_HFIP BPS.
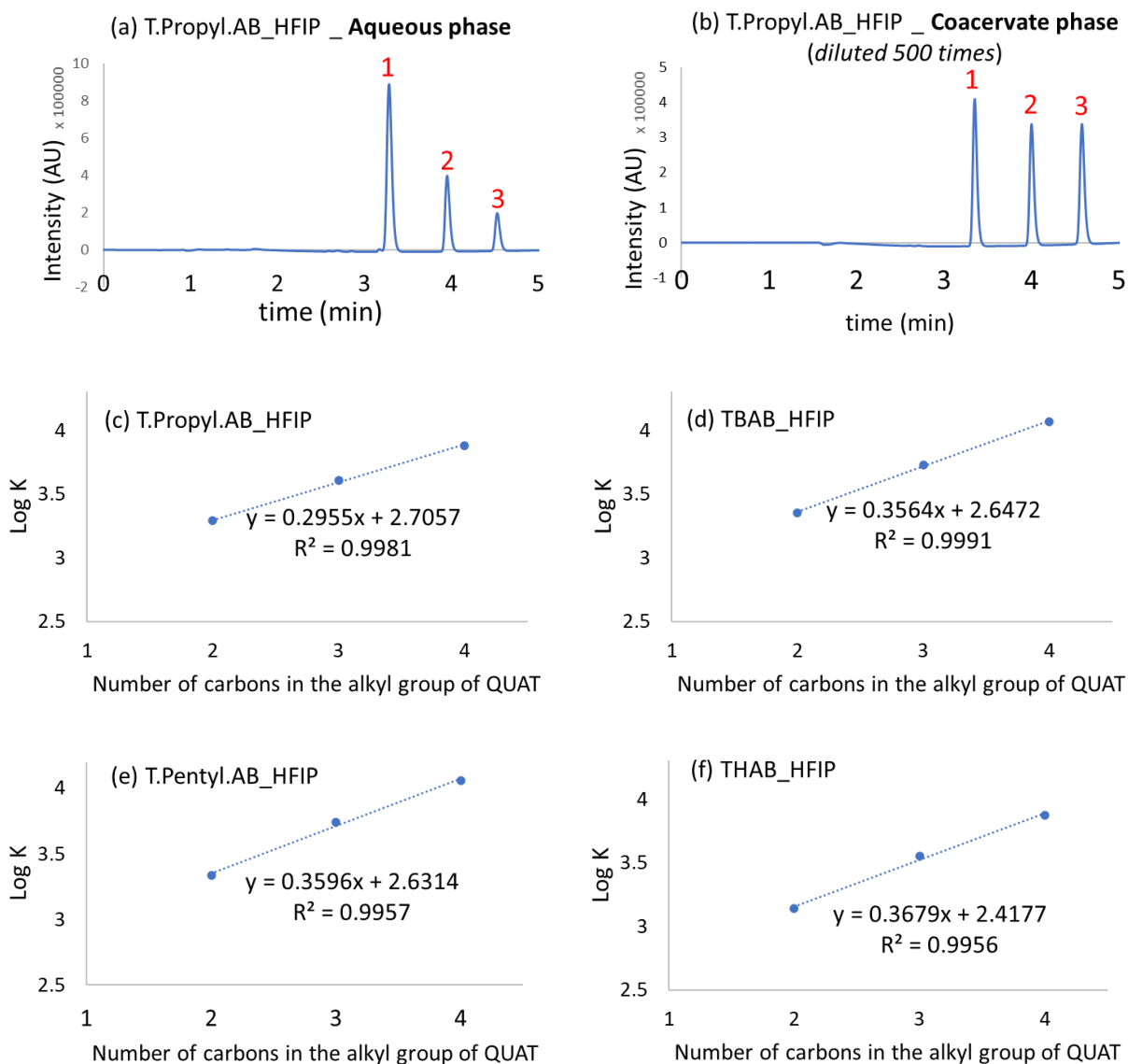
The distribution pattern of GRAVY values of the uniquely identified proteins in the two-phases of different FA*i*C-BPS:
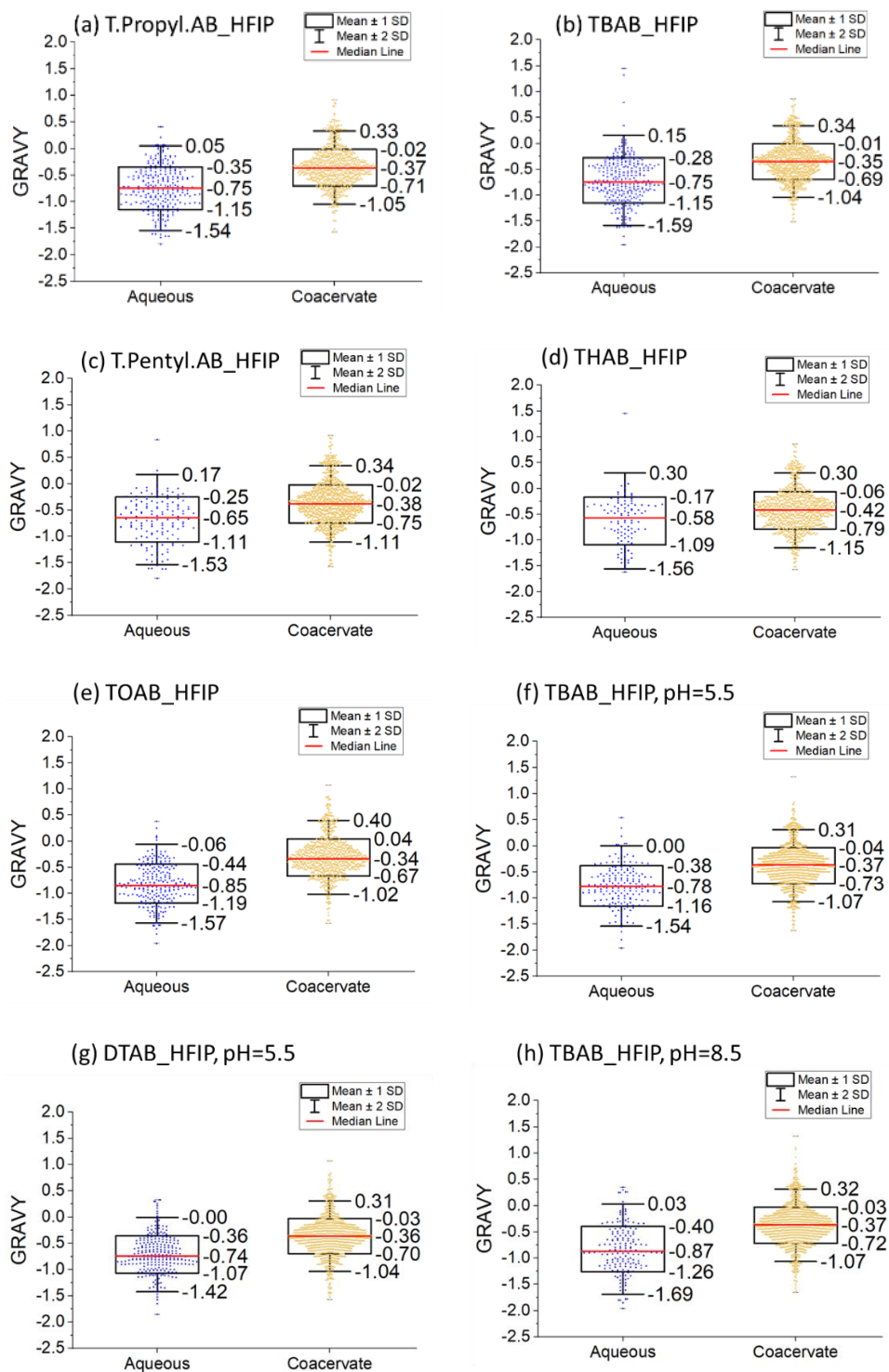
APPENDIX 4-1

TEAB_HFIP, pH=5.5

TEAB_HFIP, pH=8.5

144

CTAB_HFIP, pH=5.5

CTAB_HFIP, pH=8.5

DTAB_HFIP, pH=5.5

TBAB_HFMIP, pH=3
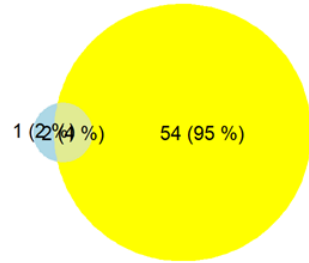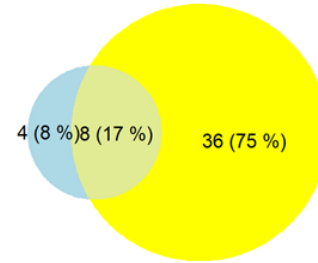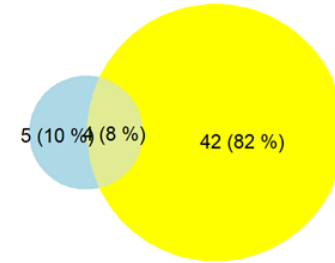

146
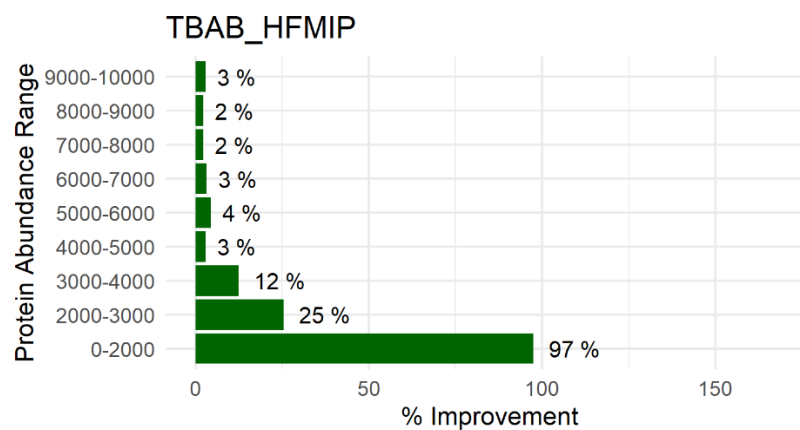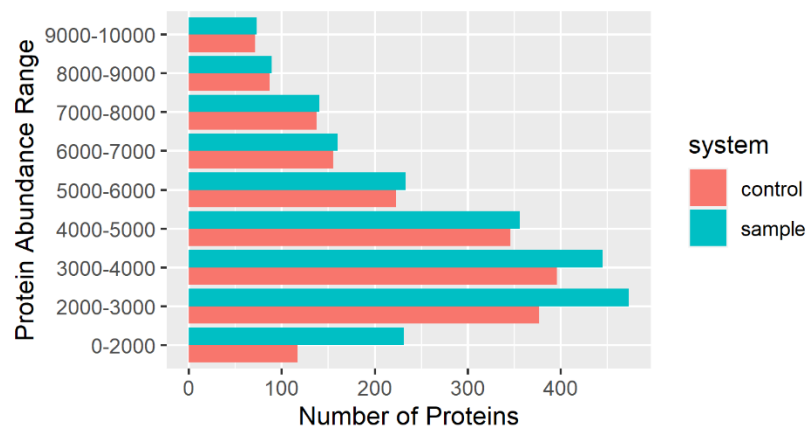
TBAB_HFMIP, pH=8.5

TBAB_TFE, pH=8.5
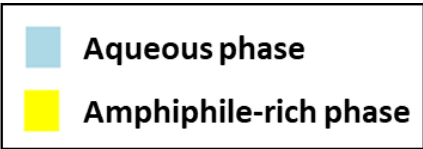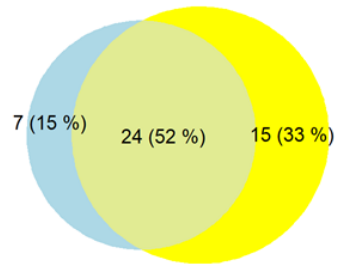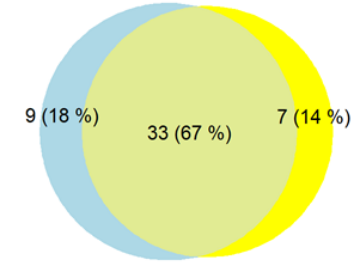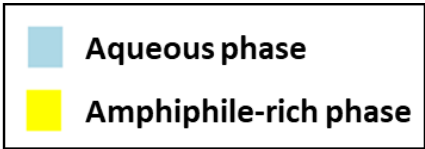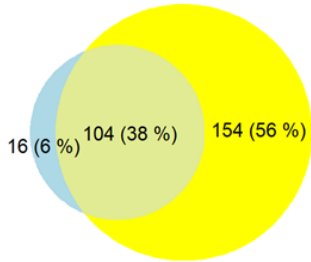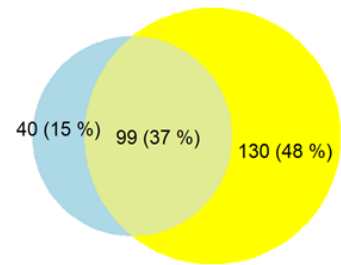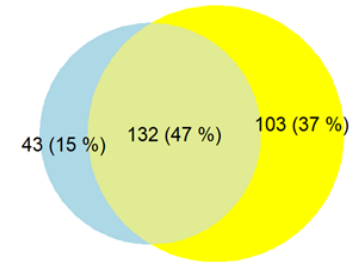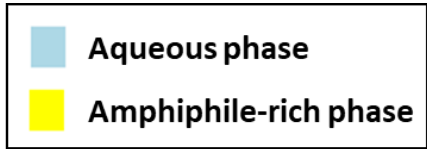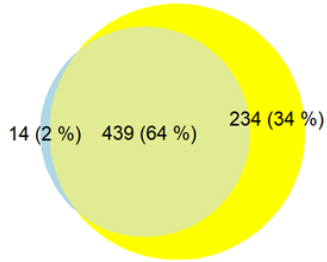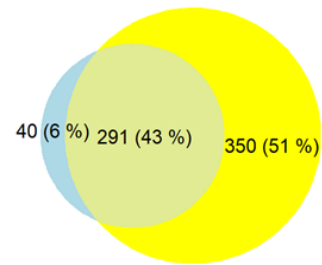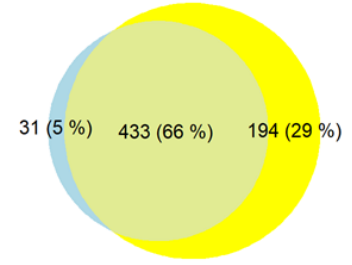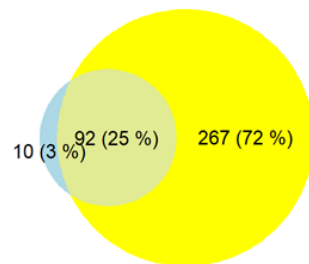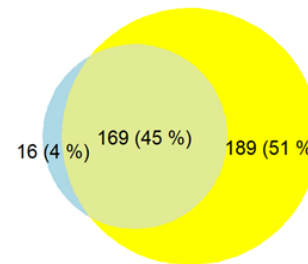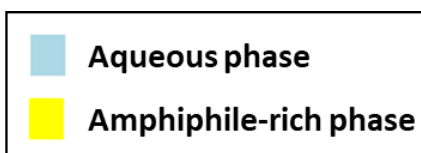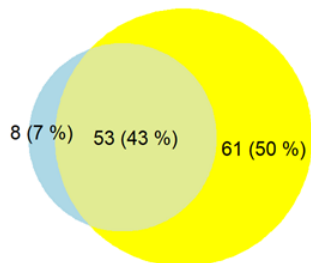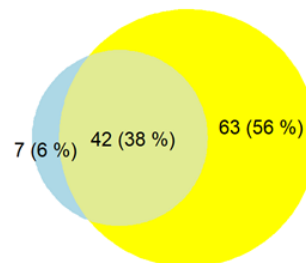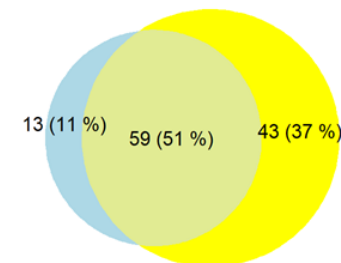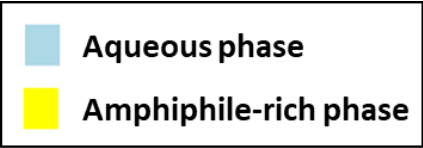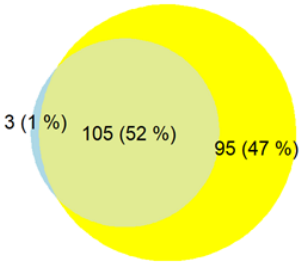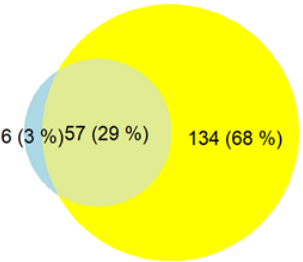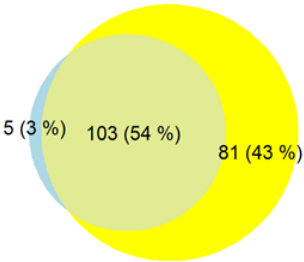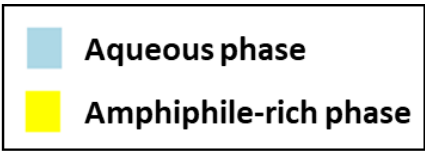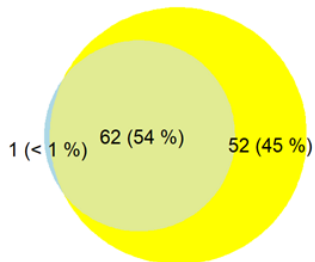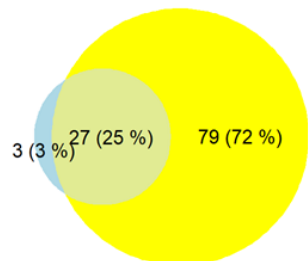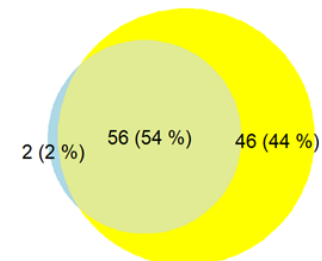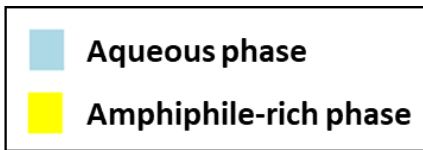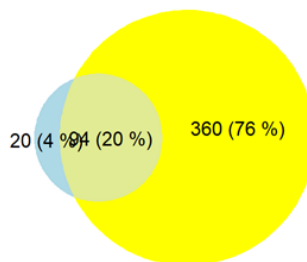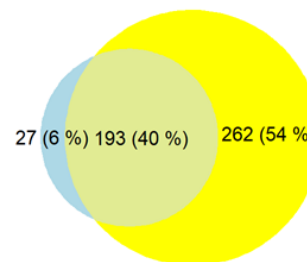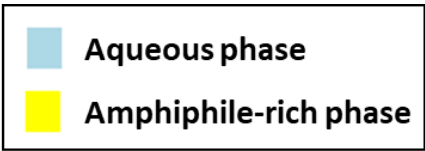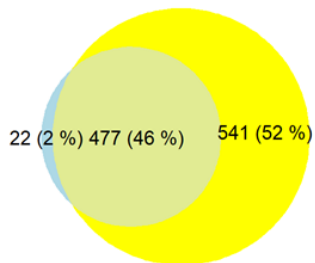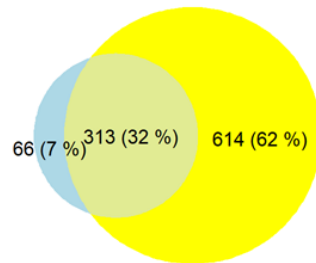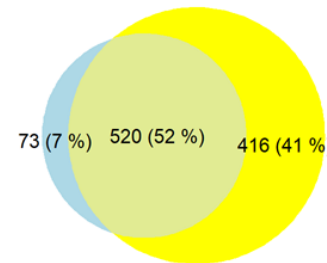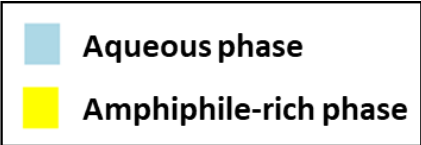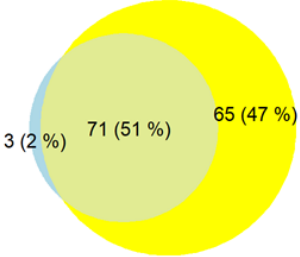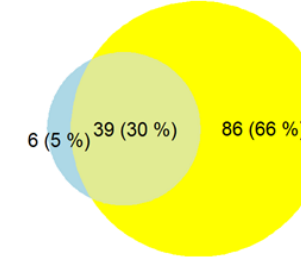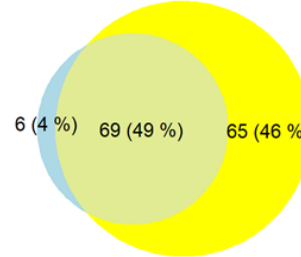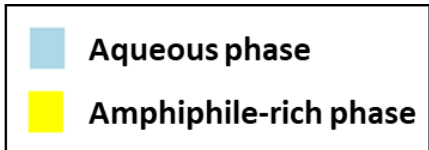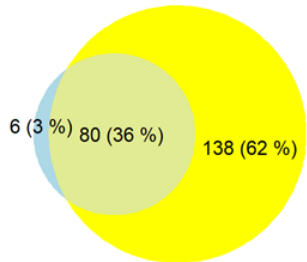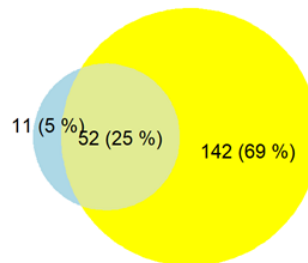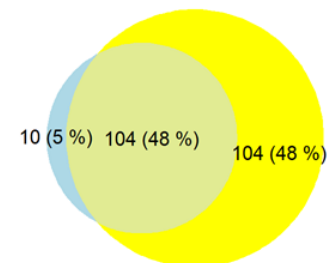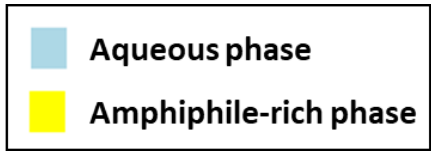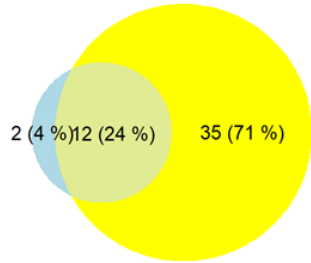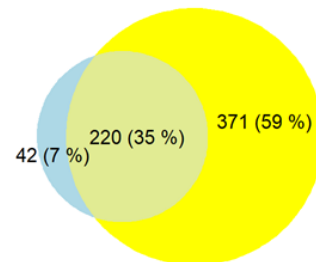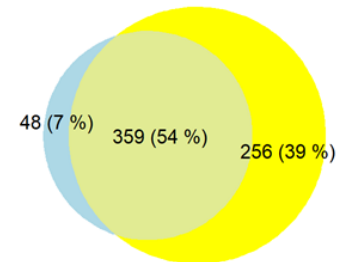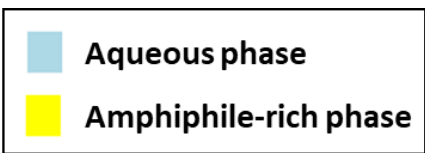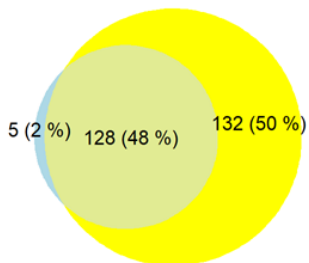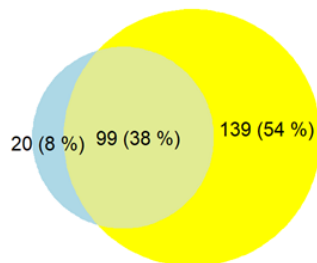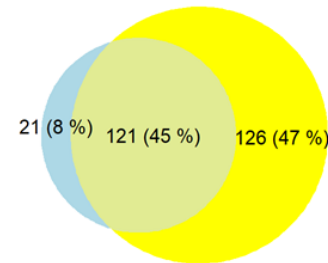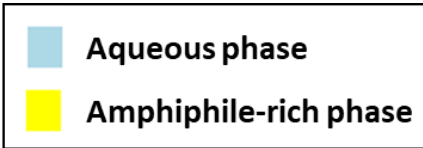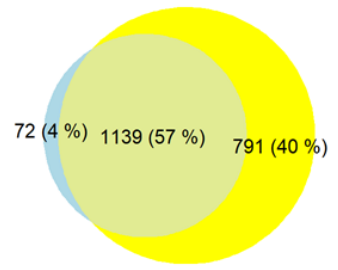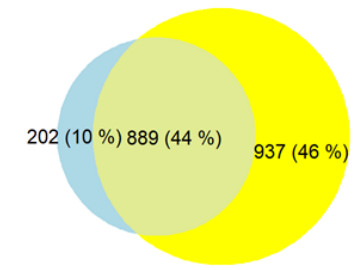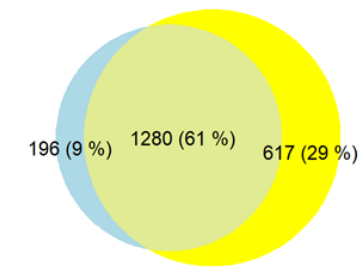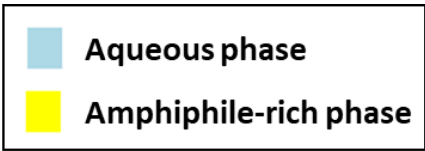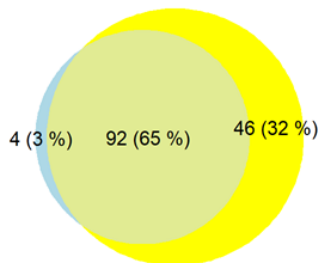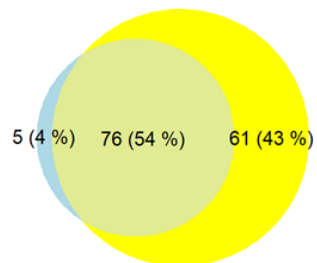
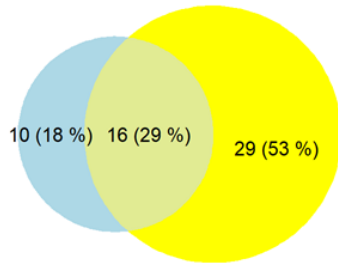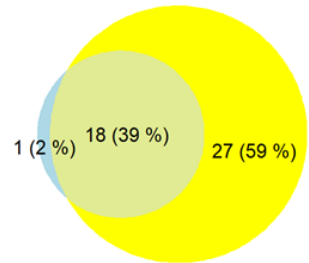TEAB_HFIP, pH=5.5
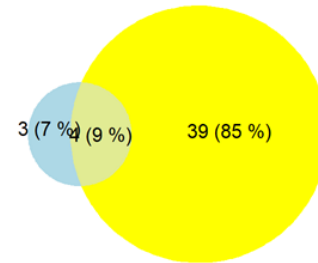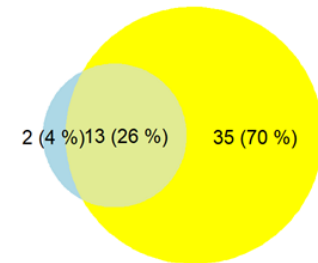
TEAB_HFIP, pH=8.5

CTAB_HFIP, pH=5.5

CTAB_HFIP, pH=8.5
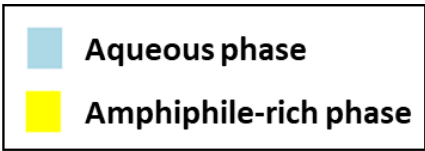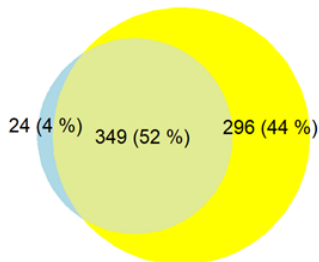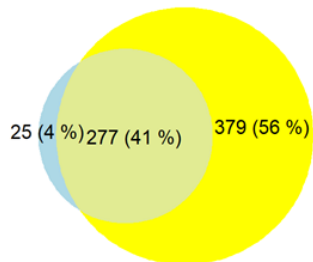
DTAB_HFIP, pH=5.5

TBAB_TFE, pH=8.5

TBAB_HFMIP, pH=3

TBAB_HFMIP, pH=8.5

Total identified proteins

| | Aqueous phase |
| Amphiphile-rich phase |

TEAB_HFIP, pH=5.5

95 (3 %) 1521 (50 %) 1397 (46 %)

TEAB_HFIP, pH=8.5

137 (4 %) 1217 (40 %) 1697 (56 %)

CTAB_HFIP, pH=5.5

267 (9 %) 1441 (48 %) 1281 (43 %)

CTAB_HFIP, pH=8.5

225 (7 %) 1463 (49 %) 1323 (44 %)

TBAB_HFMIP, pH≈3

136 (5 %) 1798 (61 %) 1018 (34 %)

TBAB_HFMIP, pH=8.5

153 (5 %) 1764 (62 %) 911 (32 %)

TBAB_TFE, pH=8.5

264 (9 %) 1043 (35 %) 1654 (56 %)

DTAB_HFIP, pH=5.5

251 (8 %) 1617 (54 %) 1118 (37 %)

Catalytic complex proteins

Aqueous phase
Amphiphile-rich phase

TEAB_HFIP, pH=5.5

17 (4 %) 223 (49 %) 216 (47 %)

TEAB_HFIP, pH=8.5

13 (3 %) 157 (34 %) 287 (63 %)

CTAB_HFIP, pH=5.5

40 (9 %) 206 (47 %) 189 (43 %)

CTAB_HFIP, pH=8.5

35 (8 %) 215 (49 %) 192 (43 %)

TBAB_HFMIP, pH≈3

19 (4 %) 268 (61 %) 154 (35 %)

TBAB_HFMIP, pH=8.5

21 (5 %) 261 (64 %) 124 (31 %)

TBAB_TFE, pH=8.5

41 (9 %) 149 (33 %) 258 (58 %)

DTAB_HFIP, pH=5.5

36 (8 %) 231 (51 %) 182 (41 %)

Cell wall proteins

TEAB_HFIP, pH=5.5

1 (2 %) 35 (81 %) 7 (16 %)

TEAB_HFIP, pH=8.5

2 (5 %) 33 (77 %) 8 (19 %)

CTAB_HFIP, pH=5.5

7 (15 %) 34 (71 %) 7 (15 %)

CTAB_HFIP, pH=8.5

6 (12 %) 34 (71 %) 8 (17 %)

TBAB_HFMIP, pH≈3

3 (7 %) 34 (77 %) 7 (16 %)

TBAB_HFMIP, pH=8.5

4 (9 %) 33 (73 %) 8 (18 %)

TBAB_TFE, pH=8.5

7 (15 %) 24 (52 %) 15 (33 %)

DTAB_HFIP, pH=5.5

9 (18 %) 33 (67 %) 7 (14 %)

Chromosome proteins

Legend:
- Aqueous phase
- Amphiphile-rich phase

TEAB_HFIP, pH=5.5
16 (6 %) 104 (38 %) 154 (56 %)

TEAB_HFIP, pH=8.5
13 (5 %) 97 (35 %) 164 (60 %)

CTAB_HFIP, pH=5.5
40 (15 %) 117 (44 %) 109 (41 %)

CTAB_HFIP, pH=8.5
26 (10 %) 116 (44 %) 121 (46 %)

TBAB_HFMIP, pH≈3
9 (4 %) 130 (52 %) 111 (44 %)

TBAB_HFMIP, pH=8.5
13 (6 %) 140 (60 %) 81 (35 %)

TBAB_TFE, pH=8.5
40 (15 %) 99 (37 %) 130 (48 %)

DTAB_HFIP, pH=5.5
43 (15 %) 132 (47 %) 103 (37 %)

153

Cytosol proteins

Legend: Aqueous phase (blue), Amphiphile-rich phase (yellow)

TEAB_HFIP, pH=5.5
14 (2 %)   439 (64 %)   234 (34 %)

TEAB_HFIP, pH=8.5
18 (3 %)   361 (52 %)   316 (45 %)

CTAB_HFIP, pH=5.5
43 (6 %)   420 (61 %)   231 (33 %)

CTAB_HFIP, pH=8.5
27 (4 %)   420 (61 %)   243 (35 %)

TBAB_HFMIP, pH≈3
27 (4 %)   481 (71 %)   172 (25 %)

TBAB_HFMIP, pH=8.5
27 (4 %)   477 (73 %)   145 (22 %)

TBAB_TFE, pH=8.5
40 (6 %)   291 (43 %)   350 (51 %)

DTAB_HFIP, pH=5.5
31 (5 %)   433 (66 %)   194 (29 %)

Endoplasmic reticulum proteins

Legend: Aqueous phase (blue), Amphiphile-rich phase (yellow)

TEAB_HFIP, pH=5.5 — 3 (< 1 %), 165 (43 %), 219 (57 %)

TEAB_HFIP, pH=8.5 — 6 (2 %), 116 (30 %), 267 (69 %)

CTAB_HFIP, pH=5.5 — 15 (4 %), 151 (39 %), 221 (57 %)

CTAB_HFIP, pH=8.5 — 15 (4 %), 159 (42 %), 204 (54 %)

TBAB_HFMIP, pH≈3 — 6 (2 %), 206 (54 %), 170 (45 %)

TBAB_HFMIP, pH=8.5 — 12 (3 %), 191 (52 %), 162 (44 %)

TBAB_TFE, pH=8.5 — 10 (3 %), 92 (25 %), 267 (72 %)

DTAB_HFIP, pH=5.5 — 16 (4 %), 169 (45 %), 189 (51 %)

155

Endosome proteins

Legend:
- Aqueous phase (light blue)
- Amphiphile-rich phase (yellow)

TEAB_HFIP, pH=5.5
8 (7 %)　53 (43 %)　61 (50 %)

TEAB_HFIP, pH=8.5
8 (7 %)　46 (39 %)　65 (55 %)

CTAB_HFIP, pH=5.5
11 (10 %)　55 (48 %)　49 (43 %)

CTAB_HFIP, pH=8.5
11 (10 %)　58 (51 %)　44 (39 %)

TBAB_HFMIP, pH≈3
4 (3 %)　62 (53 %)　50 (43 %)

TBAB_HFMIP, pH=8.5
11 (10 %)　60 (55 %)　39 (35 %)

TBAB_TFE, pH=8.5
7 (6 %)　42 (38 %)　63 (56 %)

DTAB_HFIP, pH=5.5
13 (11 %)　59 (51 %)　43 (37 %)

Golgi apparatus proteins

Aqueous phase
Amphiphile-rich phase

TEAB_HFIP, pH=5.5

3 (1 %) 105 (52 %) 95 (47 %)

TEAB_HFIP, pH=8.5

5 (2 %) 73 (35 %) 130 (62 %)

CTAB_HFIP, pH=5.5

8 (4 %) 91 (44 %) 106 (52 %)

CTAB_HFIP, pH=8.5

9 (4 %) 91 (46 %) 100 (50 %)

TBAB_HFMIP, pH≈3

7 (3 %) 116 (56 %) 86 (41 %)

TBAB_HFMIP, pH=8.5

10 (5 %) 117 (59 %) 71 (36 %)

TBAB_TFE, pH=8.5

6 (3 %) 57 (29 %) 134 (68 %)

DTAB_HFIP, pH=5.5

5 (3 %) 103 (54 %) 81 (43 %)

Golgi membrane proteins

Aqueous phase
Amphiphile-rich phase

TEAB_HFIP, pH=5.5

1 (< 1 %)   62 (54 %)   52 (45 %)

TEAB_HFIP, pH=8.5

3 (3 %)   41 (35 %)   72 (62 %)

CTAB_HFIP, pH=5.5

3 (3 %)   48 (44 %)   58 (53 %)

CTAB_HFIP, pH=8.5

2 (2 %)   51 (47 %)   55 (51 %)

TBAB_HFMIP, pH≈3

3 (3 %)   71 (62 %)   40 (35 %)

TBAB_HFMIP, pH=8.5

3 (3 %)   71 (65 %)   36 (33 %)

TBAB_TFE, pH=8.5

3 (3 %)   27 (25 %)   79 (72 %)

DTAB_HFIP, pH=5.5

2 (2 %)   56 (54 %)   46 (44 %)

158

Integral membrane proteins

Legend:
- Aqueous phase (blue)
- Amphiphile-rich phase (yellow)

TEAB_HFIP, pH=5.5
5 (< 1 %) 182 (35 %) 327 (64 %)

TEAB_HFIP, pH=8.5
11 (2 %) 137 (26 %) 370 (71 %)

CTAB_HFIP, pH=5.5
19 (4 %) 167 (33 %) 325 (64 %)

CTAB_HFIP, pH=8.5
28 (6 %) 169 (33 %) 312 (61 %)

TBAB_HFMIP, pH≈3
13 (3 %) 211 (42 %) 273 (55 %)

TBAB_HFMIP, pH=8.5
24 (5 %) 208 (44 %) 241 (51 %)

TBAB_TFE, pH=8.5
20 (4 %) 94 (20 %) 360 (76 %)

DTAB_HFIP, pH=5.5
27 (6 %) 193 (40 %) 262 (54 %)

159

Membrane proteins

Legend: Aqueous phase (blue), Amphiphile-rich phase (yellow)

TEAB_HFIP, pH=5.5 — 22 (2 %), 477 (46 %), 541 (52 %)

TEAB_HFIP, pH=8.5 — 34 (3 %), 378 (36 %), 625 (60 %)

CTAB_HFIP, pH=5.5 — 62 (6 %), 475 (46 %), 496 (48 %)

CTAB_HFIP, pH=8.5 — 68 (7 %), 474 (46 %), 489 (47 %)

TBAB_HFMIP, pH≈3 — 30 (3 %), 542 (54 %), 441 (44 %)

TBAB_HFMIP, pH=8.5 — 49 (5 %), 548 (56 %), 384 (39 %)

TBAB_TFE, pH=8.5 — 66 (7 %), 313 (32 %), 614 (62 %)

DTAB_HFIP, pH=5.5 — 73 (7 %), 520 (52 %), 416 (41 %)

160

Mitochondrial matrix proteins

Aqueous phase

Amphiphile-rich phase

TEAB_HFIP, pH=5.5

3 (2 %) 71 (51 %) 65 (47 %)

TEAB_HFIP, pH=8.5

4 (3 %) 47 (32 %) 97 (66 %)

CTAB_HFIP, pH=5.5

8 (6 %) 56 (41 %) 72 (53 %)

CTAB_HFIP, pH=8.5

8 (6 %) 53 (38 %) 79 (56 %)

TBAB_HFMIP, pH≈3

13 (9 %) 79 (53 %) 56 (38 %)

TBAB_HFMIP, pH=8.5

4 (3 %) 81 (60 %) 49 (37 %)

TBAB_TFE, pH=8.5

6 (5 %) 39 (30 %) 86 (66 %)

DTAB_HFIP, pH=5.5

6 (4 %) 69 (49 %) 65 (46 %)

Mitochondrial membrane proteins

Legend:
- Aqueous phase
- Amphiphile-rich phase

TEAB_HFIP, pH=5.5 — 6 (3 %), 80 (36 %), 138 (62 %)

TEAB_HFIP, pH=8.5 — 5 (2 %), 64 (28 %), 157 (69 %)

CTAB_HFIP, pH=5.5 — 8 (4 %), 87 (40 %), 120 (56 %)

CTAB_HFIP, pH=8.5 — 10 (4 %), 87 (39 %), 128 (57 %)

TBAB_HFMIP, pH≈3 — 3 (1 %), 115 (54 %), 94 (44 %)

TBAB_HFMIP, pH=8.5 — 8 (4 %), 102 (49 %), 99 (47 %)

TBAB_TFE, pH=8.5 — 11 (5 %), 52 (25 %), 142 (69 %)

DTAB_HFIP, pH=5.5 — 10 (5 %), 104 (48 %), 104 (48 %)

162

Mitochondrial ribosome proteins

Aqueous phase
Amphiphile-rich phase

TEAB_HFIP, pH=5.5

2 (4 %) 12 (24 %) 35 (71 %)

TEAB_HFIP, pH=8.5

1 (2 %) 4 (4 %) 54 (95 %)

CTAB_HFIP, pH=5.5

4 (8 %) 8 (17 %) 36 (75 %)

CTAB_HFIP, pH=8.5

5 (10 %) 4 (8 %) 42 (82 %)

TBAB_HFMIP, pH≈3

10 (18 %) 16 (29 %) 29 (53 %)

TBAB_HFMIP, pH=8.5

1 (2 %) 18 (39 %) 27 (59 %)

TBAB_TFE, pH=8.5

3 (7 %) 4 (9 %) 39 (85 %)

DTAB_HFIP, pH=5.5

2 (4 %) 13 (26 %) 35 (70 %)

163

Mitochondrion proteins

Aqueous phase

Amphiphile-rich phase

TEAB_HFIP, pH=5.5

24 (4 %) 349 (52 %) 296 (44 %)

TEAB_HFIP, pH=8.5

25 (4 %) 277 (41 %) 379 (56 %)

CTAB_HFIP, pH=5.5

42 (6 %) 319 (48 %) 297 (45 %)

CTAB_HFIP, pH=8.5

36 (5 %) 323 (48 %) 320 (47 %)

TBAB_HFMIP, pH≈3

27 (4 %) 409 (62 %) 221 (34 %)

TBAB_HFMIP, pH=8.5

22 (4 %) 395 (63 %) 211 (34 %)

TBAB_TFE, pH=8.5

42 (7 %) 220 (35 %) 371 (59 %)

DTAB_HFIP, pH=5.5

48 (7 %) 359 (54 %) 256 (39 %)

Nucleoplasm proteins

Aqueous phase
Amphiphile-rich phase

TEAB_HFIP, pH=5.5

5 (2 %)   128 (48 %)   132 (50 %)

TEAB_HFIP, pH=8.5

13 (5 %)   82 (29 %)   183 (66 %)

CTAB_HFIP, pH=5.5

24 (9 %)   102 (39 %)   138 (52 %)

CTAB_HFIP, pH=8.5

21 (8 %)   112 (43 %)   128 (49 %)

TBAB_HFMIP, pH≈3

9 (3 %)   144 (55 %)   108 (41 %)

TBAB_HFMIP, pH=8.5

9 (4 %)   145 (58 %)   94 (38 %)

TBAB_TFE, pH=8.5

20 (8 %)   99 (38 %)   139 (54 %)

DTAB_HFIP, pH=5.5

21 (8 %)   121 (45 %)   126 (47 %)

Phosphorylated residue proteins

Aqueous phase
Amphiphile-rich phase

TEAB_HFIP, pH=5.5

72 (4 %)   1139 (57 %)   791 (40 %)

TEAB_HFIP, pH=8.5

99 (5 %)   965 (48 %)   965 (48 %)

CTAB_HFIP, pH=5.5

205 (10 %)   1157 (57 %)   661 (33 %)

CTAB_HFIP, pH=8.5

167 (8 %)   1162 (57 %)   693 (34 %)

TBAB_HFMIP, pH≈3

71 (4 %)   1288 (65 %)   623 (31 %)

TBAB_HFMIP, pH=8.5

105 (5 %)   1307 (68 %)   517 (27 %)

TBAB_TFE, pH=8.5

202 (10 %)   889 (44 %)   937 (46 %)

DTAB_HFIP, pH=5.5

196 (9 %)   1280 (61 %)   617 (29 %)

166

Ribosome proteins

Legend:
- Aqueous phase
- Amphiphile-rich phase

TEAB_HFIP, pH=5.5
4 (3 %)   92 (65 %)   46 (32 %)

TEAB_HFIP, pH=8.5
2 (1 %)   75 (50 %)   74 (49 %)

CTAB_HFIP, pH=5.5
7 (5 %)   91 (63 %)   46 (32 %)

CTAB_HFIP, pH=8.5
5 (3 %)   90 (62 %)   49 (34 %)

TBAB_HFMIP, pH≈3
15 (10 %)   99 (67 %)   34 (23 %)

TBAB_HFMIP, pH=8.5
2 (1 %)   102 (73 %)   35 (25 %)

TBAB_TFE, pH=8.5
5 (4 %)   76 (54 %)   61 (43 %)

DTAB_HFIP, pH=5.5
3 (2 %)   85 (59 %)   55 (38 %)

Ubiquitinylated lyside residue proteins

Legend:
- Aqueous phase (light blue)
- Amphiphile-rich phase (yellow)

TEAB_HFIP, pH=5.5
27 (2 %)   944 (65 %)   485 (33 %)

TEAB_HFIP, pH=8.5
47 (3 %)   792 (54 %)   625 (43 %)

CTAB_HFIP, pH=5.5
84 (6 %)   913 (62 %)   471 (32 %)

CTAB_HFIP, pH=8.5
72 (5 %)   906 (62 %)   482 (33 %)

TBAB_HFMIP, pH≈3
30 (2 %)   1044 (73 %)   358 (25 %)

TBAB_HFMIP, pH=8.5
50 (4 %)   1043 (73 %)   334 (23 %)

TBAB_TFE, pH=8.5
69 (5 %)   658 (46 %)   716 (50 %)

DTAB_HFIP, pH=5.5
66 (5 %)   937 (65 %)   441 (31 %)

Vacuole proteins

Legend:
- Aqueous phase
- Amphiphile-rich phase

TEAB_HFIP, pH=5.5
9 (4 %)   108 (47 %)   111 (49 %)

TEAB_HFIP, pH=8.5
10 (5 %)   93 (42 %)   118 (53 %)

CTAB_HFIP, pH=5.5
19 (8 %)   98 (42 %)   116 (50 %)

CTAB_HFIP, pH=8.5
21 (9 %)   103 (46 %)   100 (45 %)

TBAB_HFMIP, pH≈3
9 (4 %)   114 (51 %)   99 (45 %)

TBAB_HFMIP, pH=8.5
17 (8 %)   121 (57 %)   75 (35 %)

TBAB_TFE, pH=8.5
25 (11 %)   58 (26 %)   142 (63 %)

DTAB_HFIP, pH=5.5
19 (9 %)   106 (49 %)   91 (42 %)

169

Vesicle proteins

TEAB_HFIP, pH=5.5

8 (4 %) 99 (47 %) 102 (49 %)

TEAB_HFIP, pH=8.5

9 (4 %) 81 (39 %) 117 (57 %)

CTAB_HFIP, pH=5.5

15 (7 %) 93 (46 %) 94 (47 %)

CTAB_HFIP, pH=8.5

14 (7 %) 103 (52 %) 83 (42 %)

TBAB_HFMIP, pH≈3

8 (4 %) 116 (56 %) 85 (41 %)

TBAB_HFMIP, pH=8.5

14 (7 %) 114 (58 %) 69 (35 %)

TBAB_TFE, pH=8.5

7 (4 %) 71 (36 %) 117 (60 %)

DTAB_HFIP, pH=5.5

18 (9 %) 104 (53 %) 76 (38 %)

CALIBRATION LINES AND RESULTS OF SEPARATION AND

QUANTIFICATION OF DCM AND HFIP

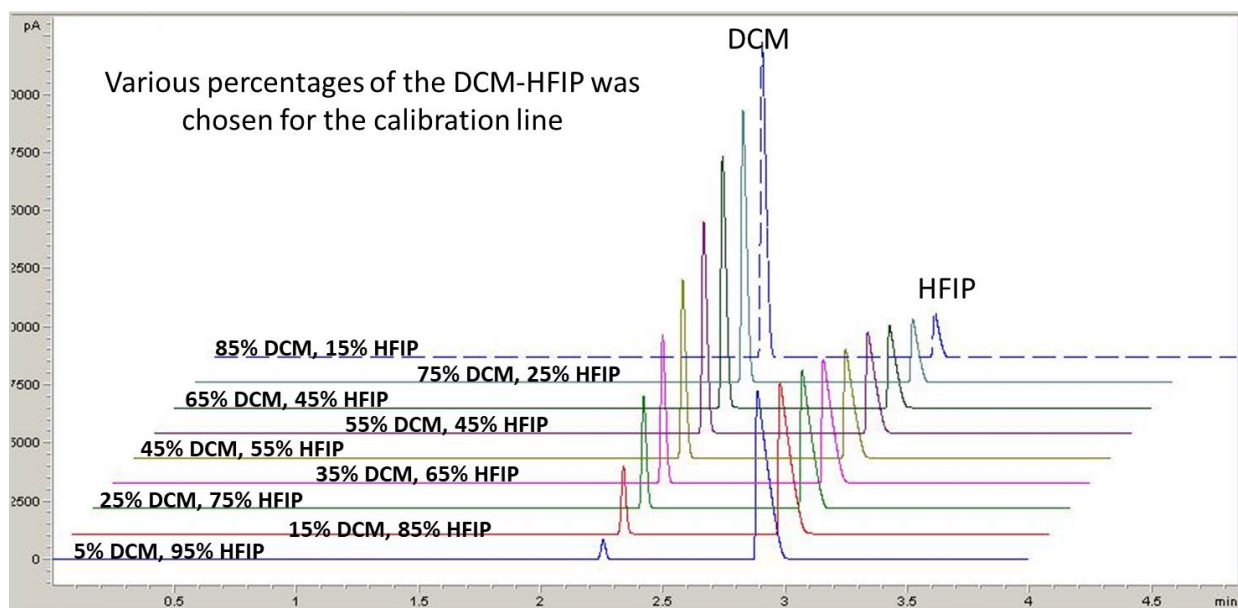<u>Quantification of phases by CG-FID</u>:

Type of column: SLB-IL60i, ionic liquid column with inert treatment

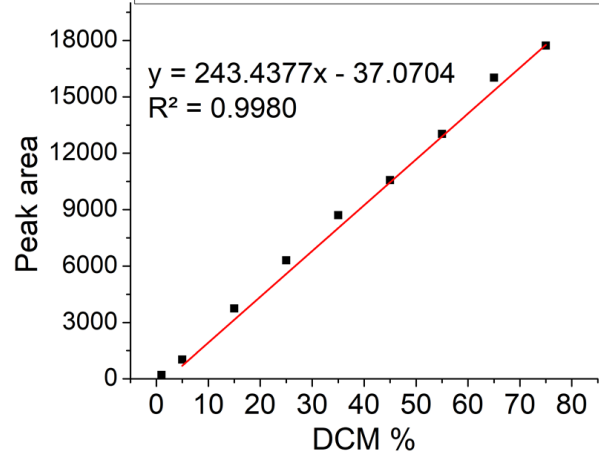Temperature gradient: 100 to 140 °C with ramp of 10 °C/min (no hold time)

Injection and detector both at 250 °C

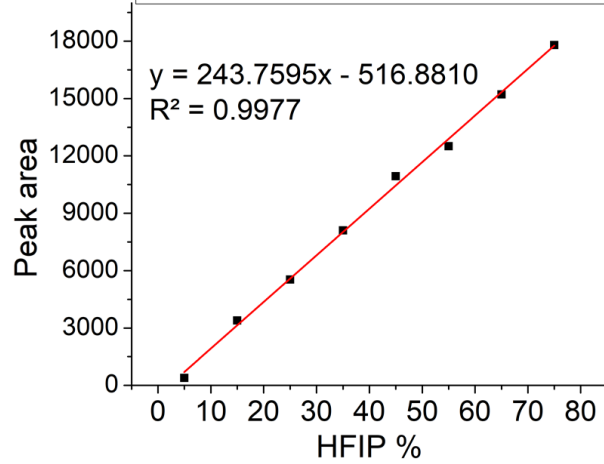Injection volume: 1 µL, with split ratio of 100:1

Carrier gas: Helium, with flow rate of 1 mL/min, constant flow mode

Calibration line for DCM

y = 243.4377x - 37.0704
R² = 0.9980

Calibration line for HFIP

y = 243.7595x - 516.8810
R² = 0.9977
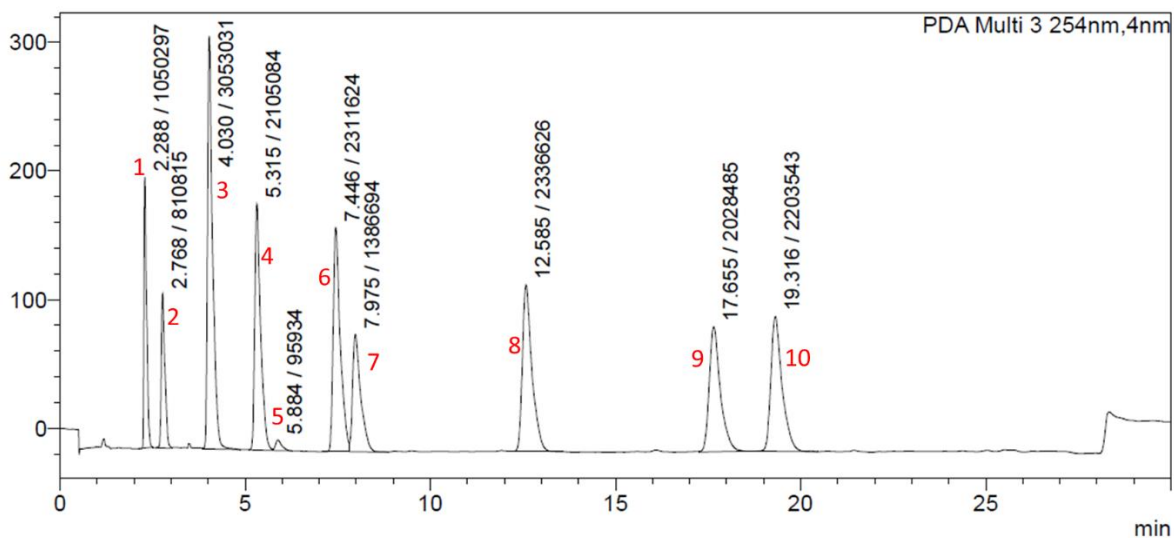
1) phenol, 2) 4-fluorophenol, 3) Acetophenone, 4) nitrobenzene, 5) benzene, 6) propiophenone, 7) 4-iodophenol, 8) butyrophenone, 9) naphthalene, 10) valerophenone. Mobile phases are A water and acetonitrile, C18 stationary phase.

BIOGRAPHICAL INFORMATION

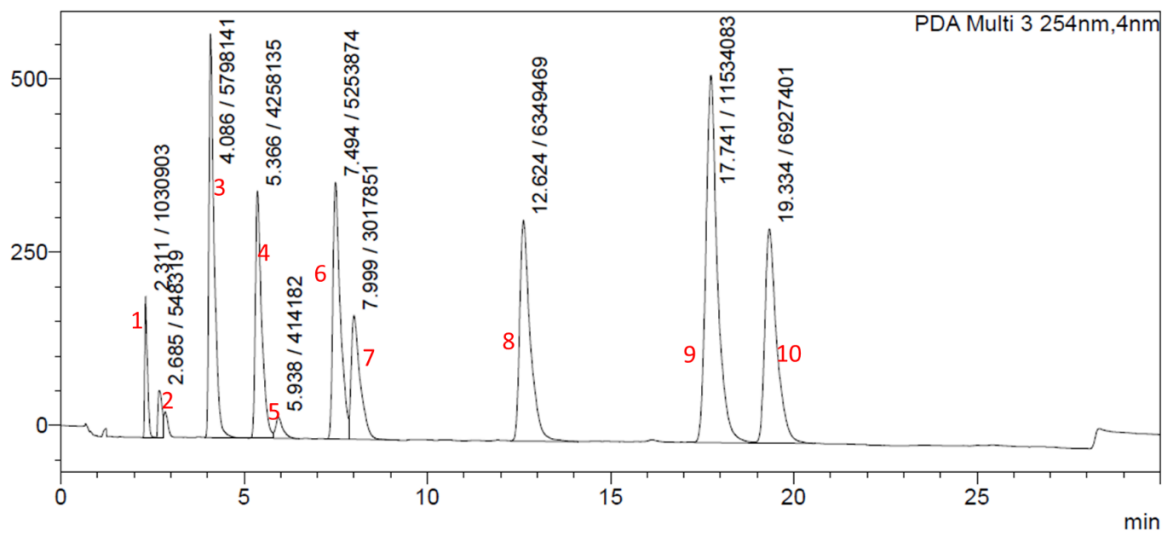Mohammadmehdi Azizi was born and raised in Shiraz, Iran. He obtained his bachelor's and master's degree in Chemical Engineering at Shiraz University, Iran. He then chose to further his education by joining Dr. Morteza Khaledi's group in 2015 at the University of Texas at Arlington. His research was focused on the development of fluoroalcohol induced supramolecular biphasic systems, and their applications in proteomics and lipidomics.