
A

Presented to
the faculty of the School of Engineering and Applied Science
University of Virginia

in partial fulfillment
of the requirements for the degree

by

APPROVAL SHEET

This

is submitted in partial fulfillment of the requirements
for the degree of

Author:

Advisor:

Advisor:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

A handwritten signature in black ink that reads "Jennifer L. West". The signature is written in a cursive style with a large initial 'J' and 'W'.

Jennifer L. West, School of Engineering and Applied Science

Abstract

Per- and polyfluoroalkyl substances (PFAS) are a large group of synthetic fluorinated chemicals with surface active and water-repellent properties that have been considered “forever chemicals” due to their rapid emergence as environmental contaminants and resistance to biological degradation. PFAS have been developed for use and production of everyday items like stains, oil and water-resistant textiles non-stick cookware and, a majority of aqueous-film forming foams used in fire suppression. The combination of wide-spread use in industrial processes and consumer products with the chemicals’ extended biological half-lives leads to accumulation of PFAS in the environment and subsequently people. Exposure and accumulation of PFAS chemicals, specifically perfluorooctanoic acid (PFOA) and perfluorooctane sulfonic acid (PFOS), has been linked to a multitude of health effects, and a majority of human exposure is linked to ingestion through water and food grown in contaminated soils. However, remediation of these contaminants across a wide range of conditions remains difficult as their strong carbon-fluorine bonds result in limited reactivity and resistance to degradation. Even the detection of PFAS poses a unique challenge and often requires expensive and timely analysis techniques unsuitable for on-site diagnostics. As the extent of the PFAS problem continues to be revealed, the overall goal of this work is to begin addressing the need for quick, on-site detection technologies as well as feasible methods for remediation of soils using biological methods.

For detection strategies, we focused on the creation of biosensors for PFAS detection by utilizing human liver fatty acid binding protein (hLFABP) as a scaffold. Chapter 2 focuses on the development of a purified protein-based sensor capable of detecting several PFAS through a rationally incorporated fluorophore (acrylodan) while Chapter 3 shows the application of this

biosensor on environmental samples and comparison to analytic methods. Chapter 4 illustrates the development of a separate biosensor based on the incorporation of circularly permuted green fluorescent protein (cpGFP) into a split hLFABP motif that is promising for genetically encoded whole cell sensing. Chapter 5 addresses remediation of soil by studying phytoaccumulation and distribution of PFAS into industrial hemp (*Cannabis sativa*). Collectively, this dissertation lays important groundwork for the use of synthetic biology and other biologically inspired techniques to begin overcoming the challenges in PFAS detection and remediation.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Bryan Berger, for welcoming me into his lab five years ago. Your endless creativity and passion have inspired me and allowed me to grow immeasurably as a researcher, and I'm so glad to have had the opportunity to learn about research, careers, and life in general from you. Thank you to my committee members, Dr. Geoffrey Geise, Dr. Roseanne Ford, Dr. Lisa Colosi Peterson, and Dr. Nick Vecchiarello for your insight and guidance.

Completing my PhD would not have been possible without my immense support system. To my family, thank you for your faith in my ability, your understanding and patience during my years of putting school first, and the pride you've shown regardless of my accomplishments. I also want to thank the most incredible friends I've made over the last five years that have not only made my time in Charlottesville fun but have helped and guided me in lab and in life. I came to UVA because it felt like home, and I wasn't disappointed. Thank you to my lab members old and new that taught me more than they even realize, and a special thank you to my undergraduate, Victoria Kartseva, for working so tirelessly and enthusiastically on our mutual success.

Table of Contents

Chapter 1: Introduction	1
1.1 Per- and polyfluoroalkyl substances.....	1
1.2 Detection of PFAS.....	6
1.3 Biosensors	8
1.4 Remediation of PFAS.....	10
1.5 Research Objectives	13
1.6 References	14
Chapter 2: Engineering human liver fatty acid binnding protein for detection of per- and polyfluoroalkyl substances	31
2.1 Abstract.....	31
2.2 Introduction	32
2.3 Materials and Methods.....	34
2.4 Results and Discussion.....	40
2.5 References	49
Chapter 3: Field performance of acrylodan labeled human liver fatty acid binding protein biosensor	55
3.1 Abstract.....	55
3.2 Introduction	56
3.3 Materials and Methods.....	57
3.4 Results and Discussion.....	59
3.5 References	66
Chapter 4: A genetically encoded biosensor for direct detection of perfluorooctanoic acid.....	71
4.1 Abstract.....	71
4.2 Introduction	72
4.3 Materials and Methods.....	75
4.4 Results and Discussion.....	79

4.5 Conclusions and Future Directions	87
4.6 Supplemental Information.....	93
4.7 References	95
Chapter 5:Uptake and distribution of perfluoroalkyl substances in industrial hemp (<i>Cannabis sativa</i>)	103
5.1 Introduction	103
5.2 Preliminary Studies	105
5.3 Ongoing work and future directions.....	107
5.4 References	111
Chapter 6: Conclusions	116
Appendix A: Towards improving sensitivity of hLFABP baed PFAS detection systems	118
A.1 Introduction	118
A.2 Materials and methods.....	119
A.3 Results and discussion.....	124
A.4 Future Work and Considerations	130
A.5 Supplemental information	131
A.6 References	135
Appendix B: Understanding how transmembrane domains regulate interactions between human BST-2 and the SARS-CoV-2 accessory protein ORF7a	138
B.1 Abstract.....	138
B.2 Introduction	139
B.3 Experimental materials and methods.....	143
B.4 Results and discussion	146
B.5 Supplemental information	156
B.6 References	186

List of Figures

Figure 1.1: Overview of main biosensor components: analyte, biorecognition unit (bioreceptor), and transducer that produces a signal. Created with Biorender.	9
Figure 2.1: hLFABP in complex with two oleic acid molecules. The residues chosen for mutation (L28, F50, F63) are highlighted in red. Figure created using PyMOL (PDB: 2LKK)	40
Figure 2.2: (A) Weighted average center of mass of fluorescence spectra for F50W hLFABP (10 μ M) after titration with PFOA. (B) Binding of 1,8-ANS to wild type (WT) and F50W hLFABP (1 μ M) characterized using relative fluorescence intensity as a function of fluorophore (1,8-ANS) concentration. The data was fit to a one-site binding model using Prism. The represented points are mean values \pm SE with n=3. (C) Fluorescence decrease, quantified as the % of initial fluorescence for PFOA. Data were fit to a dose-response model using Prism. The represented points are mean values \pm SE with n=3.....	42
Figure 2.3: (A) SDS-PAGE gel of His-trap purified [Lanes] 1: whole cell lysate, 2: column flowthrough, 3-6: purified protein:: F50W hLFABP containing both monomer (~19kDa) and formed dimer (~38kDa) bands. F50W/C69S hLFABP containing only monomer band (~19 kDa) (B) CD Spectra for unlabeled (-) and acrylodan labeled (-) C69S/F50C hLFABP.....	44
Figure 2.4: (A-C) Weighted average center of mass of fluorescence spectra for acrylodan conjugated C69S/F50C hLFABP (1 μ M) after titration with PFOA (A), PFOS (B), and PFHxS (C) in PBS as well as in creek water. Blue shift in spectra upon increased PFAS concentration corresponds to binding and is visualized as a decrease in spectra center of mass (X_{cm}). :: (D) Binding of PFAS to acrylodan conjugated C69S/F50C hLFABP (1 μ M) fitted to a one site binding model with the fractional occupancy represented as the magnitude of change in center of mass ($ \Delta X_{cm} $). (E) Fluorescence spectra of acrylodan conjugated C69S/F50C hLFABP (1 μ M) with 46 (-) and without (-) PFOA in PBS.	46
Figure 2.5: Weighted average center of mass (X_{cm}) of fluorescence spectra for acrylodan conjugated C69S/F50C hLFABP (1 μ M) after titration with PFOA and PFOS in PBS containing SDS (1 μ M).....	48
Figure 3.1: Illustration of collection sites near (A) Chapman pit and (B) Malabean Lake. The left panel depicts the area around Loring Airforce Base with an approximation of the Aroostook National Wildlife Refuge overlaid in white.	61

Figure 3.2: (LEFT) Sensor calibration curves for varying PFOA concentrations spiked in water. Data from these spiked standards were fit to a four-parameter log-dose response model, and water sample concentrations were calculated using the parameters obtained from the non-linear regression for each day the analysis was performed. (RIGHT) Comparison of predicted and verified concentrations of PFOA and total PFAS in parts per billion (ppb) for surface water samples. The calculated concentrations are mean values \pm SE with n=3. 63

Figure 4.1: cpGFP.hLFABP construct design overview. (A) linear representation of construct with linkers L1 (GSG) L2 (GGTGGG) and L3 (GG) (B) schematic of circularly permuted GFP.hLFABP at the protein level depicting binding of PFAS (C) overlay of apo form (orange) and holo form (blue) of wild type hLFABP complexed with palmitic acid (PBD IDs 3STN and 3STK respectively) ⁵⁰. Location of receptor splitting was partly based on perturbations at residues 56 and 57. Figure created using Biorender.com. 81

Figure 4.2: SDS-PAGE analysis of cpGFP.hLFABP purification via Ni-NTA resin. Lane M: protein marker, Lanes 1-6: purification fractions containing increasing concentrations of imidazole (50, 75, 100, 125, 200, and 500mM respectively). cpGFP.hLFABP is shown as a band of approximately 46.6 kDa. 82

Figure 4.3: Representative absorption spectra for cpGFP.hLFABP (2 μ M). Spectral changes upon increased PFOA addition show a decreased and increased trend in intensity at 395nm and 485nm respectively. Curves are smoothed for qualitative analysis. 83

Figure 4.4: Representative fluorescence emission spectra for cpGFP.hLFABP (1 μ M) after excitation at 395nm (LEFT) and 485nm (RIGHT). Spectral changes upon PFOA titration shows a decreased and increased overall intensity for excitation at 395nm and 485nm respectively. All curves are smoothed for qualitative analysis. 83

Figure 4.5: Binding of PFOA to cpGFP.hLFABP (1 μ M) after titration in PBS (black circles), creek water (blue squares), and PBS containing 1 μ M SDS (green triangles). Data are fitted to a one site binding model with the fractional occupancy represented as the fractional occupancy represented as the fractional change in fluorescence intensity at 510nm after excitation at 485nm. The represented points are mean values \pm SEM with n=3. Bottom table displays calculated model constants and limits of detection in all systems. 84

Figure 4.6: Fluorescence intensity change at 510nm after excitation at 485nm of *E. coli* expressing cpGFP.hLFABP upon PFOA titration. Data are normalized by cell concentration (OD600) and fitted to a dose response model. The represented points are mean values \pm SEM with n=3. 86

Figure 4.7: Fluorescence intensity change at 510nm after excitation at 485nm of *E. coli* expressing cpGFP.hLFABP with (blue) and without (black) FadL expression upon PFOA titration. Data are normalized by cell concentration (OD600) and fitted to a dose response model. The represented points are mean values \pm SEM with n=3. 89

Figure 4.8: Anti myc-tag Western Blot of whole cell lysates for *E. coli* expressing (1) cpGFP.hLFABP-AIDAc fusion and (2) AIDAc protein without passenger protein used as a control. Arrow points to band representing full length cpGFP.hLFABP-AIDAc fusion (~90 kDa)..... 92

Figure 5.1: Hemp plants grown in high dosed (10ppm) soil after 7 days of growth in comparison to control plants..... 106

Figure 5.2: PFAS concentration ($\mu\text{g kg}^{-1}$ wet weight) of hemp grown in cumulatively spiked soils at either low (LEFT) or high (RIGHT) concentrations..... 107

Figure A.1: Sequence alignment of Wild Type hLFABP with full-length mutants from library screening. Blue highlighted residues show variation from Wild Type at residues I41, I52, F63, V92, T93, S100, and T102. Mutant #48 also has variation at residue K21. MSAs generated with SnapGene. 125

Figure A.2: *E. coli* AS19 library transformation plated on spectinomycin containing LB agar after incubation with 250 ppm (LEFT) or 500 ppm (RIGHT). 126

Figure A.3: SDS-PAGE analysis of Wild Type (WT) hLFABP and full-length mutant expression solubility. Bands representing protein variants are depicted at ~19kDa. 127

Figure A.4: Binding of PFOA shown through displacement of 1,8-ANS. Fluorescence decrease is quantified as the % of initial fluorescence. Data were fit to a dose-response model. 128

Figure A.S1: Representation of hLFABP gene library containing NNK codons for randomization. Using the wild type hLFABP gene as a template, 4 segments will be amplified out using primers containing NNK nucleotides. Once sections are amplified out using NNK containing primers, they will be “stitched” back together and incorporated into expression vector using compatible overhang sections created using *BsaI* restriction sites. **Figure made using Biorender** 132

Figure A.S2: Final OD₆₀₀ of *E. coli* after 8 hours of growth in varying PFOA concentrations. Values are normalized by growth in no PFOA. (A) AS19 and BL21 strains grown in either regular

or 0.4M NaCl containing LB (B) AS19 grown in 0.4M NaCl containing LB expressing WT hLFABP cytosolically compared to no protein expression..... 133

Figure B.1: (A) AraTM and DN-AraTM assay. In the AraTM assay, pAraTMwt expresses the full-length AraC fused to a target sequence under control of the PTrc promoter. This plasmid is co-expressed with the pAraGFP plasmid that contains an eGFP reporter gene under control of the PBAD promoter. In the DN-AraTM assay, addition of the pAraTMDN plasmid expresses the truncated AraC (AraC*) fused to a competitor sequence under control of the PTrc promoter. Co-expression of both constructs with pARAGFP in *E. coli* leads to a distribution of interactions between AraC and AraC* fusions, and the relative affinity for homo- versus heterooligomerization can be independently quantified in terms of eGFP expression; reduced GFP expression is indicative of preferential heterooligomerization, whereas increased GFP expression is indicative of preferential homooligomerization. (B) Truncated amino acid sequences of SARS-Cov-2 ORF7a and BST-2 domains used for the current work. (C) Key residue contacting pairs of both WT and I28S heterodimerization revealed from MD simulations..... 147

Figure B.2: AraTM homodimerization assay for varying BST-2 domains..... 148
Ratios of GFP fluorescence intensity vs. cell concentration (OD₆₀₀) for each construct are compared with the background (bkgd). In the background sample, *E. coli* was transformed with the empty pTrcRSF scaffold plasmid that does not express the MBP-AraC-construct. (Upper) Representation of full-length human BST-2 protein and residue numbers used for all BST-2 domain chimeras (Accession Number Q10589). (N=5, **** represents p<0.0001)..... 148

Figure B.3: DN-AraTM dimerization assay for BST-2 SNPs with SARS-CoV-2 ORF7a..... 150
Ratios of GFP fluorescence intensity vs. cell concentration (OD₆₀₀) for each construct are compared with the background (denoted as – in both the AraC and AraC* containing construct table). In the background sample, *E. coli* was transformed with the empty pTrcRSF scaffold for pAraTMwt and pTrc99a scaffold for pAraTMDN plasmids that do not express the MBP-AraC or MBP-AraC* constructs. (N=5, ***, **, and * represent p<0.001, p<0.01, and p<0.05 respectively) 150

Figure B.4: Results from AA-MD for ORF7a dimerization with WT BST-2..... 151

(A-D): Contact maps of four groupings calculated by using the top 5000 configurations from the center of GMVAE clustering groups. The color bars represent the percentage of contacting distance less than 9 Å. (E-H): Snapshots of dimer groups made by the superimposed top 50 configurations

from the center of the group. Nonpolar, polar, acidic, and basic residues are colored white, green, red, and blue, respectively..... 151

Figure B.5: Results from AA-MD for ORF7a dimerization with mutant I28S BST-2..... 152

(A-E): Contact maps of five groupings calculated by using the top 5000 configurations from the center of GMVAE clustering groups. The color bars represent the percentage of contacting distance less than 9 Å. **(F-J):** Snapshots of dimer groups made by the superimposed top 50 configurations from the center of the group. Nonpolar, polar, acidic, and basic residues are colored white, green, red, and blue, respectively..... 152

Figure B.6: I28S mutation effects. 153

The I28S SNP in BST-2 alters the local molecular interactions and water contacts per residues. The configurations demonstrate that **(A)** S²⁸ forms an extra sidechain-backbone hydrogen bond with L²⁴, which does not occur for **(B)** I²⁸. **(C)** The water contacts for each BST-2 residue computed water contacting residues within 3 Å over the WT Group 3 and I28S Group 4 which the individual membership possibility of heterodimeric configuration is more than 0.75 for its own group.... 153

Figure B.7: HEK293T cells expressing BST-2 variants exhibit changes in glycosylation patterns when cotransfected with SARS-CoV-2 ORF7a. **(Top)** Western blot of cells transfected with wild-type as well as I28S and I28T BST-2 exhibit an increase in intensity for lower molecular weight band upon coexpression with ORF7a. **(Bottom)** Densitometry analysis of protein bands normalized to GAPDH internal controls. (n=3) This lower molecular weight band represents unglycosylated BST-2 which suggests ORF7a expression leads to decreased levels of glycosylated BST-2 in all variants. 155

Figure B.S1: The procedure of the MD simulations and the initial placements of dimers..... 158

Figure B.S2: Results from CG-MD for the ORF7a dimerization with the WT BST-2. The top three panels are contact maps for three CG runs **(A-C)**. The color bars are in angstrom (Å), and the maps are averages over the last 2.5 μs of simulation. End snapshots from the 5 μs CG-MD runs are shown in the bottom panels **(D-F)**, with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively. The phosphate atoms are colored gold..... 161

Figure B.S3: Results from CG-MD for the ORF7a dimerization with the I28S mutated BST-2. The top three panels are contact maps for three CG runs **(A-C)**. The color bars are in angstrom (Å), and the maps are averages over the last 2.5 μs of simulation. End snapshots from the 5 μs CG-MD runs

are shown in the bottom panels **(D-F)**, with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively. The phosphate atoms are colored gold. 161

Figure B.S4: Total loss, cross-entropy loss, and reconstruction loss of the train **(LEFT)** and the validation set **(RIGHT)** for the ORF7a/WT BST-2 dimerization. 163

Figure B.S5: Total loss, cross-entropy loss, and reconstruction loss of the train **(LEFT)** and the validation set **(RIGHT)** for the ORF7a/I28S BST-2 dimerization. 163

Figure B.S6: The reconstruction loss with the latent dimension for the ORF7a/WT BST-2 **(A)** and the ORF7a/I28S BST-2 **(B)**. 163

Figure B.S7: TSNE visualization of 5D latent space colored according to their cluster assignment probabilities. Only points with more than 0.75 membership probability are shown. 164

Figure B.S8: Contact maps of the WT clusters form GMVAE with a membership possibility over 0.75. The color bars are in percentage of contacting distance less than 9 Å. 166

Figure B.S9: Contact maps of the I28S clusters form GMVAE with a membership possibility over 0.75. The color bars are in percentage of contacting distance less than 9 Å. 167

Table B.S4: Occupancy percentage of contact map for the WT clusters. 168

Figure B.S10: AA dimer snapshots of the WT clusters made by the supper position of the top 50 configurations from the center of each cluster, with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively. 169

Figure B.S11: AA dimer snapshots of the I28S clusters made by the supper position of the top 50 configurations from the center of each cluster, with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively. 169

Figure B.S13: Distribution of crossing angle **(A, C)** and helices distance **(B, D)** of the WT and I28S groups using the top 5000 configurations from the center of GMVAE clustering groups. 173

Figure BS.14: Molecular structure of the helix-packing motif with detailed side chain contacts in the WT Group 3 Oxygen, nitrogen, and carbon atoms colored red, blue, and gray, respectively and BST-2 and ORF7a colored orange and pink in QuickSurf style. The labeled residues have contacts with the other peptide chain within 3 Å. 178

Figure BS.15: Molecular structure of the helix-packing motif with detailed side chain contacts in the I28S Group 4. Oxygen, nitrogen, and carbon atoms colored red, blue, and gray, respectively and I28S BST-2 and ORF7a colored orange and pink in QuickSurf style. The labeled residues have contacts with the other peptide chain within 3 Å. 178

Figure BS.16: TM helix-helix interface of the WT groups.....	179
Figure BS17: TM helix-helix interface of the I28S groups.	180
Figure BS.18: Interaction energies for amino acid residues	181
(A) BST-2 and (B) ORF7a computed over the WT Group 3 and I28S Group 4. The individual membership possibility of heterodimeric configuration is more than 0.75 for its own group. The mutated residue is marked with an “*”.....	181
Figure B.S21: Salt bridges formed in WT Group 3 and I28S Group 4.....	184
Salt bridges are formed by (A) E ¹²¹ (ORF7a) and R ¹⁹ (WT BST-2), (B) E ¹²¹ (ORF7a) and K ²¹ (WT BST-2) and (C) E ¹²¹ (ORF7a) and R ¹⁹ (I28S BST-2). Nonpolar, polar, acidic, and basic residues in secondary structure are colored white, green, red, and blue, respectively. Oxygen, nitrogen, carbon and hydrogen atoms in highlight molecular structures are colored red, blue, cyan, and white, respectively.	184
Figure B.S22: Contacting complexes of ORF7a/BST-2 heterodimerization at the EC/TM interface	184
In WT packing, (A) a salt bridge formed by E91 (ORF7a) and K47 (BST-2) or (B) a hydrogen bond between Q94a (ORF7a) and K47 (BST-2) promotes an amino- π interaction between Q94a (ORF7a) and F44 (BST-2) and π - π stacking between Y97 (ORF7a) and P44 (BST-2). In I28S packing, (C) an ORF7a intra-hydrogen bond formed by Q94a and Y97 in the extended helix domain promotes the same residues to form amino- π stacking and a π - π stacking interactions.	184

List of Tables

Table 2.1: Binding parameters of PFOA for WT and F50W hLFABP as well as PFOA, PFOS, and PFHxS for C69S, C69S/F50W, and C69S/F50C hLFABP. These values were determined using displacement of the fluorophore 1,8-ANS. Error is reported as propagated SEM.	43
Table 3.1: Results of surface water samples from LC/MS-MS. Verified concentrations are in ng/L with LOQ defined as limit of quantification.	62
Table A.1: Binding parameters for PFOA-protein interactions. Values were determined using displacement of the fluorophore 1,8-ANS (ANS)	129
Table A.S1: Oligos used for library generation and cloning	131
Table A.S2: Colony Forming Units of AS19 <i>E. coli</i> plated on high salt LB media containing 0.4M NaCl. Cells were transformed with spectinomycin resistant plasmid and allowed to recover in 0.4M NaCl containing LB for 2 hours prior to plating.	134
Table B.S1: Hyperparameters for both dimerization systems.	162
Table B.S2: Data distribution of the WT cluster form GMVAE clustering.	164
Table B.S3: Data distribution of the I28S cluster form GMVAE clustering.	165
Table B.S5: Grouping of the GMVAE identified dimer clusters.	170
Table B.S6: Occupancy percentage of contact map for the I28S clusters.	170
Table B.S7: Length, distance, and crossing angle of helices. The mean and standard error have been provided.	173
Table B.S8: Length, distance, and crossing angle of helices for the WT clusters. The mean and standard deviation error were calculated over 5000 sampling of individual clusters.	174
Table B.S9: Length, distance, and crossing angle of helices for the I28S clusters. The mean and standard deviation error were calculated over 5000 sampling of individual clusters.	174
Table B.S10: Summary of WT Helix-Helix C α Contacting Residues Pairs.	176
Table B.S11: Summary of I28S Helix-Helix C α Contacting Residues Pairs.	177

Chapter 1

Introduction

1.1 Per- and polyfluoroalkyl substances

Per and poly-fluoroalkyl substances (PFAS) are a group of man-made chemicals that have become an increasing concern. Entitled “forever chemicals” by news and media outlets, these fluorochemicals are pervasive environmental contaminants without a robust natural degradation pathway.^{1,2} First manufactured in the 1940s, PFAS have been used in a variety of industrial and commercial products including fire-fighting foams, synthetic fabrics, medical devices, food packaging, and cosmetic goods.³⁻⁶ PFAS consist of chains of highly fluorinated carbon atoms bound to polar head groups which are typically carboxylic acids, tertiary amines, or sulfide moieties.^{1,7} Over 3000 different types of PFAS have been industrially manufactured with two major categories consisting of perfluorocarboxylic acids (PFCAs) and perfluorosulfonic acids (PFSAs).¹ Due to their unique chemical composition, PFAS are water and/or oil resistant which is valuable in a number of commercial applications. As a result, PFAS are abundant in everyday items including stain and water-resistant fabrics, non-stick cookware, and even hygienic products like dental floss.^{6,8-11} Furthermore, their amphiphilic properties confer high solubility under aqueous conditions, causing these chemicals to be a prevalent and mobile set of environmental contaminants.¹²⁻¹⁴

While a universally accepted definition still lacking, a broad interpretation has been recently proposed as: “fluorinated substances that contain at least one fully fluorinated methyl or methylene carbon atom (without any H/Cl/Br/I atom attached to it), i.e., with a few noted exceptions (represented by a carbon atom instead having H/Cl/Br/I atoms attached), any chemical with at least a perfluorinated methyl group ($-CF_3$) or a perfluorinated methylene group ($-CF_2-$) is a PFAS”.² These chemicals are classified through the length of their fluorinated carbons chain (C4-C17) and grouped into either polymeric or non-polymeric molecule categories.¹ Polymeric PFAS include fluoropolymers like polytetrafluoroethylene (PTFE) and polyvinylidene fluoride (PVDF), side-chain fluorinated polymers like fluorinated acrylate polymers, and perfluoropolyethers. These chemicals often break down in the environment into non-polymeric PFAS which are classified into groups representing perfluoroalkyl and polyfluoroalkyl substances. Polyfluoroalkyl PFAS contain at least one partially fluorinated carbon atom bound to oxygen or hydrogen, while perfluoroalkyl substances include chemicals with carbon chains completely fluorinated aside from the terminal end. This terminal moiety is a polar function group such as carboxylates, sulfonates, or phosphates, and the combination of saturated fluoro-carbon chains and ionic head group confer amphiphilicity. Perfluoroalkyl PFAS can then be further subdivided into groups among which perfluoroalkyl acids (PFAAs) contain some of the most common and extensively studied PFAS like perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA).¹⁵

1.1.1 PFAS health effects and human exposure pathways

Despite only having been manufactured for less than a century, most residents of industrialized countries have been exposed to PFAS.¹⁶⁻¹⁸ Among a representative sample of the U.S. population, 95% of human serum analyses yielded a positive result for PFAS.^{19,20} With a half-life of several

years and the inability to naturally degrade, PFAS can accumulate in human tissues through long-term exposure, even with a source containing relatively low concentrations.^{5,21} While new toxicological effects are continuing to be discovered, long-term exposure to PFAS, specifically medium chain perfluoroalkyl acids like PFOA, have been linked to numerous health problems including increased cholesterol levels²²⁻²⁴, various cancers²⁵⁻²⁷, and reproductive issues²⁸. Recent studies have suggested high levels of PFOA are correlated with reduced vaccine response,^{29,30} and chronic exposure to PFAS could render a person more susceptible to pathogens or viral infections.³¹ PFAS have been shown to transfer from mother to fetus,³² and prenatal exposure to PFAS has been linked to low birth weight,³³ neurobehavioral issues,^{34,35} as well as childhood adiposity^{36,37}. Many reviews summarizing toxicity and health information have been published.^{5,28,38-40} Furthermore, biomonitoring studies in a variety of species have shown that upon accumulation, the highest concentrations of PFOA are found in the blood plasma and liver.^{41,42} This has since been further elucidated as PFOA is shown to bind to relevant proteins including liver fatty acid binding protein (LFABP) and serum albumin.^{43,44} Human exposure to PFAS has been contributed to multiple pathways including ingestion of contaminated drinking water and food including crops grown in contaminated soils and biosolids as well as general dermal adsorption.⁴⁵

1.1.2 Environmental distribution and accumulation of PFAS

Environmental release of PFAS occurs through a variety of sources. Many PFAS are directly released in the form of chemical intermediates, waste, and final products from industrial production of fluoropolymers, building materials, and other consumer commodities like textiles, food packaging, and cosmetics. Major sources of PFAS contamination, however, are contributed

to the professional use of aqueous firefighting foams (AFFFs)^{46,47}, industrial and municipal wastewater treatment plants (WWTPs)^{48,49}, landfills⁵⁰, and even plants specifically created for the recycling and incineration of PFAS-containing products⁵¹. By virtue of their chemical properties and global use, PFAS have been found in almost every region on Earth in a variety of environmental matrices, living organisms, and even humans. While resistant to biotic and abiotic degradation, some polyfluorinated forms of PFAS, including the long chain fluorotelomer-based compounds used in these AFFFs, undergo partial biotransformation which results in shorter chain PFAS like PFAAs which have even greater biologic impacts.^{52,53,7} PFOA and PFOS, in particular, are often directly released into the environment as industrial impurities but are also formed through the degradation and even atmospheric oxidation of fluorotelomer alcohols and other long chain PFAS.

Atmospheric dispersion of PFAS has a nontrivial impact on subsequent transport and deposition into other environmental medias like soil and water.⁵⁴⁻⁵⁷ Volatile PFAS, like fluorotelomers, can partition into the atmosphere partially contributing to these chemicals' environmental ubiquities. Studies have found various volatile PFAS in numerous places including indoor air in schools and residences⁵⁸ and in the atmosphere above several oceans and seas^{6,59}. Even nonvolatile PFAS are highly mobile in the air due to their high adsorption onto organic particulates in the atmosphere,⁶⁰ as many varieties have been found to reach polar regions like Antarctica as well as Tibetan highlands^{61,62}.

Some PFAS, including ionic PFAAs with short to medium carbon chains ($C \leq 8$) are extremely soluble in water and partition at air-water interfaces⁶³ while longer chain chemicals tend to be

found in water-sediment fractions⁶⁴. Therefore, bodies of water, including drinking water, are consistently found to contain PFAS and act as major vehicles for further contamination. In fact, exposure to PFAS through drinking water is a global human-health concern; studies conducted by the United States Geological Survey indicate at least one PFAS substance is detected in 45% of United States drinking water.⁶⁵ In response to rising concerns, the United States Environmental Protection Agency proposed in March of 2023 to change national drinking water limits to 4 parts per trillion for PFOA and PFOS.⁶⁶

Atmospheric deposition of PFAS leads to soil contamination even in areas where direct human activity is absent with PFOA and PFOS being common at low levels.^{62,67,68} However, soil is primarily polluted through water run-off, direct contamination, or application of contaminated biosolids and irrigation water. High concentrations of PFAS are found near manufacturing sites and areas where AFFFs have been consistently utilized. Studies have shown concentrations on the order of multiple milligrams per kilogram of soil after acute or chronic use of AFFFs⁶⁹⁻⁷¹ as well as manufacturing facilities⁷².

Soil, specifically in areas of agricultural use, are also commonly contaminated with PFAS through the use of recycled water and sludge materials from WWTPs.^{73,74} Discharge of contaminated effluents from treatment facilities is a large contribution of PFAS in the environment.^{13,48,75} In fact, WWTPs often have higher outputs of PFAAs as polyfluorinated precursors are oxidized to recalcitrant PFAS by the diverse community of microorganisms vital in water treatment processes.^{48,49,76} Since municipalities are not capable of removing PFAS, accumulation is often seen in the sludge produced.^{75,77} This sludge is a byproduct formed after chemical and/or biological

treatment of waste solids after separation. After treatment, biosolids are often added to landfills, applied to agricultural cropland, or used at reclamations sites due to the numerous benefits it has as a sustainable fertilizer source with nitrogen and nutrient boosting properties; In 2018, the National Biosolids Data Project estimated about 40% (~2.3 million dry metric tons) of biosolids were recycled and applied to agricultural soils in the United States while another 45% was either landfilled or incinerated.⁷⁸ However, adsorption of PFAS into these materials has been seen to cause contamination in the U.S. in the realm of 1000-6000 µg/kg dry weight with PFOA, PFOS, and PFDA as major contributors.^{73,79}

Water and soil contamination, regardless of source, contribute to the entry to PFAS into the terrestrial food chain. While studies are limited, it has been shown that livestock exposed to contaminated water, soil, air, or food accumulate the chemicals which can be transferred to offspring as well as commercial products like milk and eggs.⁸⁰⁻⁸⁵ There are many studies that show the uptake of PFAS into edible crops grown in contaminated soils as well as hydroponic systems. In fact, PFAS have been found to accumulate in a large variety of food crops: cereals like wheat corn and oats⁸⁶⁻⁸⁸, tomato and other fruits^{86,89}, and a wide array of vegetables including carrots celery radishes lettuce and cucumbers^{86,89-91}.

1.2 Detection of PFAS

With a rise in evidence of PFAS accumulation and toxicity comes a wave of regulatory changes and calls for action that highlight the necessity of quick, relatively easy ways to detect chemicals like PFOA.^{66,92} However, this challenge has proven non-trivial given the diversity of the chemicals as well as their limited reactivity and vast concentrations ranges. Currently, standard PFAS

detection relies on chromatography techniques coupled with tandem mass spectroscopy. These methods are highly precise with detection limits in the range of 1 ng/L for aqueous samples (EPA Methods 533, 537, and 537.1).⁹³⁻⁹⁵ These methods also require highly sophisticated equipment and training, and therefore are not readily available for widespread, rapid use in testing foods, commercial products or water samples.^{95,96}

Increasing health concerns and new regulations in response to these concerns have led to development of new PFAS sensors capable of detecting compounds in drinking water. Most PFAS technologies rely on complexation with organic dyes, allowing for optical detection. Specifically, complexation of PFAS with a cationic dye has been used to develop an app-based test kit for PFOS.⁹⁷ While this technology is portable and offers detection on the order of parts per billion (ppb), prior clean-up steps in the form of liquid or solid phase extraction are required as interference of inorganic ions like chloride inhibit detection.⁹⁷ Additionally, immunoassays take advantage of receptor-ligand interactions, e.g. peroxisome proliferator-activated receptor alpha response elements were modified with gold nanoparticles to detect PFOS at ppt levels using optical density changes.⁹⁸ While promising, immunoassays can be challenging to implement in field applications due to the time, infrastructure and reagents required for multiple washing and incubation steps prior to detection, as well as a need for pre-treatment to remove potentially interfering compounds in complex mixtures. While still constrained to bench level research, the most successful technologies revolve around the use of molecularly imprinted polymers (MIPS) to capture and detect PFAS down to the ppt levels determined in health advisories set by the EPA.^{66,99-101} Nonetheless, these methods often require extensive sample preparation, and/or are

limited to drinking water samples making PFAS detection impractical for other applications and inaccessible for most communities.

1.3 Biosensors

In contrast to sensors with synthetic scaffolds such as nanoparticles and polymers, biosensors, provide alternative detection strategies by utilizing biological recognition elements as receptors. Biosensors have seen rapid development due to their abilities to identify wide ranges of analytes and their applicability to numerous fields including environmental monitoring where matrix complexity are a challenge.^{102,103} Overall, biosensors are commonly classified by the most important components, bioreceptors and transducers.¹⁰⁴ Bioreceptors biochemically recognize elements from analyzed samples while transducers convert the outcome into a quantified signal; this signal can be electrical, thermal, or optical.¹⁰⁵ A general overview of biosensors and is outlined in **Figure 1.1**.

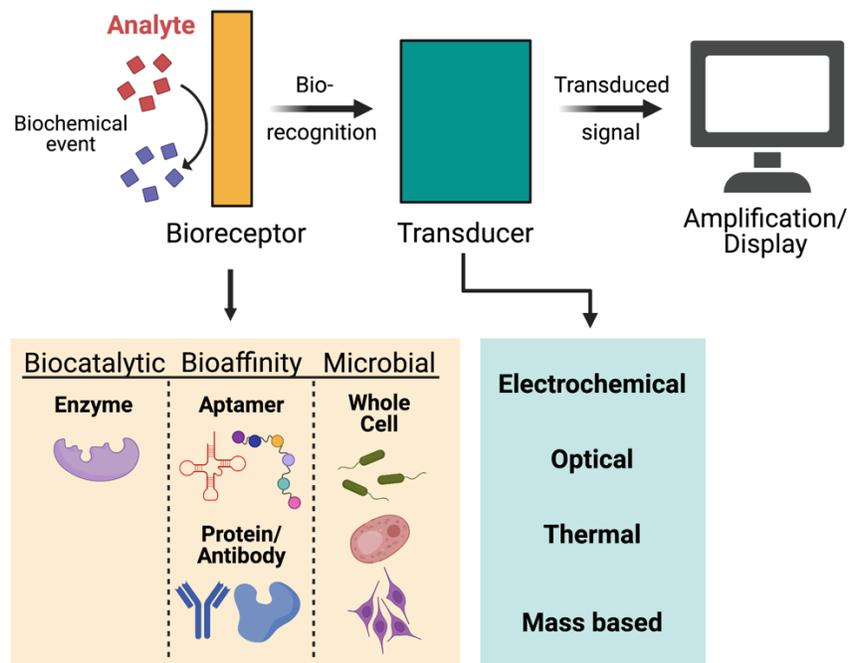


Figure 1.1: Overview of main biosensor components: analyte, biorecognition unit (bioreceptor), and transducer that produces a signal. Created with Biorender.

Successful bioreceptors are comprised of biological molecules including proteins, nucleic acids, antigens, cells, tissues and are commonly grouped as biocatalytic, bioaffinity, or microbial based¹⁰⁴. Biocatalytic systems use enzymes as sensing elements, and their signals are created through enzymatic reactions. Bioaffinity based sensors utilize specific binding interactions between analytes and biomolecules like antibodies, nucleic acid or peptide aptamers, and binding proteins. Microbial sensors are whole cell systems utilizing microbes like yeasts, bacteria, and algae for detection. Cells are capable of producing recognition elements without the need for purification or extraction and exhibiting biochemical responses upon analyte interaction.^{106–108} The vast array of design elements outlined in **Figure 1.1** highlights the multidisciplinary nature of biosensor design. After bioreceptor and analyte interaction optimization, biosensor design focuses on increasing signal transduction and amplification processes in which the incorporation of

nanoparticles and microfluidic devices have become areas of increasing interest.^{109–111} However, one of the main benefits of biosensor technology revolves around the tunability of biomolecules. Through the use of genetic engineering and molecular biology techniques, bioreceptors with improved sensitivity and selectivity can be identified or produced for a variety of applications. Specifically, the work in this dissertation focuses on the development of novel PFAS bioreceptors with incorporated fluorescent optical signal methods for detection.

1.4 Remediation of PFAS

Resistance to biodegradation, low volatility, and high electronegativity from fluorine have created a significant challenge when it comes to PFAS remediation.¹¹² In fact, most conventional water treatment options including flocculation, sedimentation, as well as anaerobic biological treatment are ineffective in treating PFAS¹¹³; therefore current treatment methods focus on sequestration of PFAS using adsorption onto matrices like granulated activated carbon (GAC) and ion exchange resins or filtration through reverse osmosis and nanofiltration.¹¹⁴ While these techniques are effective in removal of PFAS from water, some have high operation costs (filtration) while others have limited capability in trapping shorter chain PFAS molecules (adsorption). Furthermore, all these remediation techniques result in secondary waste from matrix regeneration or single use resins and filters.^{115–117} Currently, this waste is dealt with by incineration which mineralizes PFAS incompletely, leaving behind short chain intermediate species byproducts.^{117,118}

Emerging technologies for breakdown of PFAS molecules are often based on advanced redox reactions that require high chemical and/or energy requirements. Electrochemical oxidation utilizing metal oxide anodes have been proven effective as breakdown occurs through

decarboxylation via direct electron transfer. Specifically, boron doped diamond anodes have been shown to breakdown many PFAS molecules including fluorotelomers as well as shorter chain PFCAs like PFOA and PFOS.^{119,120} However, electrochemical oxidation becomes less effective for shorter chain PFAS and these intermediates persist even after residence times of 10 hours. In addition, electrochemical oxidation also creates other toxic by-products like perchlorate, hydrogen fluoride, chlorine gas, and organic halides.^{119,120}

While ongoing work is being done to investigate effective breakdown techniques and mechanisms, emerging technologies focusing on degradation are only feasible for highly concentrated amounts of PFAS making them unfit full-scale application.¹²¹ Therefore, the development of large scale, economically viable solutions for remediation and/or concentration of PFAS is vital. Furthermore, PFAS' absorption tendencies, heterogeneous nature of environmental matrices, and sheer volume of material make the remediation of contaminated soils an even bigger challenge. Currently, two broad approaches, mobilization and immobilization, may prove to be economically viable approaches for soil remediation. Immobilization redistributes contaminants and reduces mobility through sorption onto various materials akin to the water sequestration described above. Activated carbon,^{122–125} ionic polymer,^{126,127} and organic matter^{128–130} based amendments are potentially valuable for PFAS immobilization. While these techniques can reduce PFAS leaching significantly, the actual removal of these contaminants from soil, or mobilization, is necessary to remediate soil for societal use. Soil flushing and washing are *in situ* processes in which solutions are used with the goal of extracting contaminants. However, this often utilizes harsh chemicals like organic or inorganic acids and bases or harsh solvents and requires the use of copious amounts of water that then must be further decontaminated.¹³¹ Due to their proven ability to accumulate in a multitude

of plants, phytoremediation could be an effective strategy of sequestering PFAS from various contaminated water and soil systems. Phytoremediation is a technique that has been used for decades in soil remediation of heavy metals.^{132,133} Furthermore, many studies have shown the application and potential of this technique for the clean-up of a variety of pollutants including radionuclides¹³⁴, herbicides¹³⁵, and hydrocarbons¹³⁶. This concept is based around the use of plants and their related microenvironment and soil amendments to remove degrade or stabilize pollutants and can be divided into several categories: phytoextraction/phytoaccumulation, phytodegradation, rhizofiltration, phytostabilization, and phytovolatilization.¹³⁷ As PFAS are resistant to biodegradation, phytoaccumulation is a feasible strategy for sequestration of these chemicals.

1.5 Research Objectives

The overall goal of this thesis is to address the need for quick, on-site detection technologies for per- and polyfluoroalkyl substances (PFAS) as well as feasible methods of remediating agricultural and other lands through the use of biological methods and inspiration.

Chapter 2 discusses the development of a biosensor based on human liver fatty acid binding protein (hLFABP) that can detect several PFAS molecules through the solvatochromic fluorophore, acrylodan. **Chapter 3** goes on to demonstrate the application of this Ac-hLFABP sensor on real world environmental samples through comparison with the gold standard method, LC-MS/MS. **Chapter 4** illustrates the development of a separate biosensor based on incorporation of circularly permuted green fluorescent protein (cpGFP) into a split hLFABP motif that is promising for the incorporation into a genetically encoded whole cell sensing system. **Chapter 5** goes on to address remediation of soil from PFAS. Explicitly this chapter discusses the feasibility and study of phytoaccumulation and distribution of PFAS into industrial hemp (*Cannabis sativa*), a well known plant for phytoremediation potential.

1.6 References

- (1) Buck, R. C.; Franklin, J.; Berger, U.; Conder, J. M.; Cousins, I. T.; de Voogt, P.; Jensen, A. A.; Kannan, K.; Mabury, S. A.; van Leeuwen, S. P. Perfluoroalkyl and Polyfluoroalkyl Substances in the Environment: Terminology, Classification, and Origins. *Integr Environ Assess Manag* **2011**, 7 (4), 513–541. <https://doi.org/10.1002/ieam.258>.
- (2) Wang, Z.; Buser, A. M.; Cousins, I. T.; Demattio, S.; Drost, W.; Johansson, O.; Ohno, K.; Patlewicz, G.; Richard, A. M.; Walker, G. W.; White, G. S.; Leinala, E. A New OECD Definition for Per- and Polyfluoroalkyl Substances. *Environ. Sci. Technol.* **2021**, 55 (23), 15575–15578. <https://doi.org/10.1021/acs.est.1c06896>.
- (3) D’Agostino, L. A.; Mabury, S. A. Aerobic Biodegradation of 2 Fluorotelomer Sulfonamide-Based Aqueous Film-Forming Foam Components Produces Perfluoroalkyl Carboxylates. *Environ Toxicol Chem* **2017**, 36 (8), 2012–2021. <https://doi.org/10.1002/etc.3750>.
- (4) Dauchy, X.; Boiteux, V.; Bach, C.; Rosin, C.; Munoz, J.-F. Per- and Polyfluoroalkyl Substances in Firefighting Foam Concentrates and Water Samples Collected near Sites Impacted by the Use of These Foams. *Chemosphere* **2017**, 183, 53–61. <https://doi.org/10.1016/j.chemosphere.2017.05.056>.
- (5) Fenton, S. E.; Ducatman, A.; Boobis, A.; DeWitt, J. C.; Lau, C.; Ng, C.; Smith, J. S.; Roberts, S. M. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environmental Toxicology and Chemistry* **2021**, 40 (3), 606–630. <https://doi.org/10.1002/etc.4890>.
- (6) Wang, Z.; DeWitt, J. C.; Higgins, C. P.; Cousins, I. T. A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environ. Sci. Technol.* **2017**, 51 (5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>.
- (7) Harding-Marjanovic, K. C.; Houtz, E. F.; Yi, S.; Field, J. A.; Sedlak, D. L.; Alvarez-Cohen, L. Aerobic Biotransformation of Fluorotelomer Thioether Amido Sulfonate (Lodyne) in AFFF-Amended Microcosms. *Environ. Sci. Technol.* **2015**, 49 (13), 7666–7674. <https://doi.org/10.1021/acs.est.5b01219>.
- (8) Favreau, P.; Poncioni-Rothlisberger, C.; Place, B. J.; Bouchex-Bellomie, H.; Weber, A.; Tremp, J.; Field, J. A.; Kohler, M. Multianalyte Profiling of Per- and Polyfluoroalkyl Substances (PFASs) in Liquid Commercial Products. *Chemosphere* **2017**, 171, 491–501. <https://doi.org/10.1016/j.chemosphere.2016.11.127>.

- (9) Kotthoff, M.; Müller, J.; Jürling, H.; Schlummer, M.; Fiedler, D. Perfluoroalkyl and Polyfluoroalkyl Substances in Consumer Products. *Environ Sci Pollut Res Int* **2015**, *22* (19), 14546–14559. <https://doi.org/10.1007/s11356-015-4202-7>.
- (10) Robel, A. E.; Marshall, K.; Dickinson, M.; Lunderberg, D.; Butt, C.; Peaslee, G.; Stapleton, H. M.; Field, J. A. Closing the Mass Balance on Fluorine on Papers and Textiles. *Environ. Sci. Technol.* **2017**, *51* (16), 9022–9032. <https://doi.org/10.1021/acs.est.7b02080>.
- (11) Schaidler, L. A.; Balan, S. A.; Blum, A.; Andrews, D. Q.; Strynar, M. J.; Dickinson, M. E.; Lunderberg, D. M.; Lang, J. R.; Peaslee, G. F. Fluorinated Compounds in U.S. Fast Food Packaging. *Environ. Sci. Technol. Lett.* **2017**, *4* (3), 105–111. <https://doi.org/10.1021/acs.estlett.6b00435>.
- (12) Ahrens, L.; Bundschuh, M. Fate and Effects of Poly- and Perfluoroalkyl Substances in the Aquatic Environment: A Review. *Environ Toxicol Chem* **2014**, *33* (9), 1921–1929. <https://doi.org/10.1002/etc.2663>.
- (13) Arvaniti, O. S.; Stasinakis, A. S. Review on the Occurrence, Fate and Removal of Perfluorinated Compounds during Wastewater Treatment. *Science of The Total Environment* **2015**, *524–525*, 81–92. <https://doi.org/10.1016/j.scitotenv.2015.04.023>.
- (14) Brunn, H.; Arnold, G.; Körner, W.; Rippen, G.; Steinhäuser, K. G.; Valentin, I. PFAS: Forever Chemicals—Persistent, Bioaccumulative and Mobile. Reviewing the Status and the Need for Their Phase out and Remediation of Contaminated Sites. *Environ Sci Eur* **2023**, *35* (1), 20. <https://doi.org/10.1186/s12302-023-00721-8>.
- (15) Cousins, I. T.; DeWitt, J. C.; Glüge, J.; Goldenman, G.; Herzke, D.; Lohmann, R.; Miller, M.; Ng, C. A.; Scheringer, M.; Vierke, L.; Wang, Z. Strategies for Grouping Per- and Polyfluoroalkyl Substances (PFAS) to Protect Human and Environmental Health. *Environ. Sci.: Processes Impacts* **2020**, *22* (7), 1444–1460. <https://doi.org/10.1039/D0EM00147C>.
- (16) Kato, K.; Kalathil, A. A.; Patel, A. M.; Ye, X.; Calafat, A. M. Per- and Polyfluoroalkyl Substances and Fluorinated Alternatives in Urine and Serum by on-Line Solid Phase Extraction–Liquid Chromatography–Tandem Mass Spectrometry. *Chemosphere* **2018**, *209*, 338–345. <https://doi.org/10.1016/j.chemosphere.2018.06.085>.
- (17) Kato, K.; Wong, L.-Y.; Jia, L. T.; Kuklenyik, Z.; Calafat, A. M. Trends in Exposure to Polyfluoroalkyl Chemicals in the U.S. Population: 1999–2008. *Environ. Sci. Technol.* **2011**, *45* (19), 8037–8045. <https://doi.org/10.1021/es1043613>.

- (18) Olsen, G. W.; Mair, D. C.; Lange, C. C.; Harrington, L. M.; Church, T. R.; Goldberg, C. L.; Herron, R. M.; Hanna, H.; Nobiletti, J. B.; Rios, J. A.; Reagen, W. K.; Ley, C. A. Per- and Polyfluoroalkyl Substances (PFAS) in American Red Cross Adult Blood Donors, 2000–2015. *Environmental Research* **2017**, *157*, 87–95. <https://doi.org/10.1016/j.envres.2017.05.013>.
- (19) Calafat, A. M.; Kuklenyik, Z.; Caudill, S. P.; Reidy, J. A.; Needham, L. L. Perfluorochemicals in Pooled Serum Samples from United States Residents in 2001 and 2002. *Environ. Sci. Technol.* **2006**, *40* (7), 2128–2134. <https://doi.org/10.1021/es0517973>.
- (20) Calafat, A. M.; Kuklenyik, Z.; Reidy, J. A.; Caudill, S. P.; Tully, J. S.; Needham, L. L. Serum Concentrations of 11 Polyfluoroalkyl Compounds in the U.S. Population: Data from the National Health and Nutrition Examination Survey (NHANES) 1999–2000. *Environ. Sci. Technol.* **2007**, *41* (7), 2237–2242. <https://doi.org/10.1021/es062686m>.
- (21) Post, G. B.; Cohn, P. D.; Cooper, K. R. Perfluorooctanoic Acid (PFOA), an Emerging Drinking Water Contaminant: A Critical Review of Recent Literature. *Environmental Research* **2012**, *116*, 93–117. <https://doi.org/10.1016/j.envres.2012.03.007>.
- (22) Lin, P.-I. D.; Cardenas, A.; Hauser, R.; Gold, D. R.; Kleinman, K. P.; Hivert, M.-F.; Fleisch, A. F.; Calafat, A. M.; Webster, T. F.; Horton, E. S.; Oken, E. Per- and Polyfluoroalkyl Substances and Blood Lipid Levels in Pre-Diabetic Adults—Longitudinal Analysis of the Diabetes Prevention Program Outcomes Study. *Environment International* **2019**, *129*, 343–353. <https://doi.org/10.1016/j.envint.2019.05.027>.
- (23) Steenland, K.; Fletcher, T.; Savitz, D. A. Epidemiologic Evidence on the Health Effects of Perfluorooctanoic Acid (PFOA). *Environ Health Perspect* **2010**, *118* (8), 1100–1108. <https://doi.org/10.1289/ehp.0901827>.
- (24) Steenland, K.; Tinker, S.; Frisbee, S.; Ducatman, A.; Vaccarino, V. Association of Perfluorooctanoic Acid and Perfluorooctane Sulfonate With Serum Lipids Among Adults Living Near a Chemical Plant. *American Journal of Epidemiology* **2009**, *170* (10), 1268–1278. <https://doi.org/10.1093/aje/kwp279>.
- (25) Imir, O. B.; Kaminsky, A. Z.; Zuo, Q.-Y.; Liu, Y.-J.; Singh, R.; Spinella, M. J.; Irudayaraj, J.; Hu, W.-Y.; Prins, G. S.; Madak Erdogan, Z. Per- and Polyfluoroalkyl Substance Exposure Combined with High-Fat Diet Supports Prostate Cancer Progression. *Nutrients* **2021**, *13* (11), 3902. <https://doi.org/10.3390/nu13113902>.

- (26) Law, H.; Armstrong, B.; D'este, C.; Hosking, R.; Smurthwaite, K.; Trevenar, S.; Lucas, R.; Lazarevic, N.; Kirk, M.; Korda, R. Relative Rates of Cancers and Deaths in Australian Communities with PFAS Environmental Contamination Associated with Firefighting Foams: A Cohort Study Using Linked Data. *Cancer Epidemiology* **2023**, *82*, 102296. <https://doi.org/10.1016/j.canep.2022.102296>.
- (27) Messmer, M. F.; Salloway, J.; Shara, N.; Locwin, B.; Harvey, M. W.; Traviss, N. Risk of Cancer in a Community Exposed to Per- and Poly-Fluoroalkyl Substances. *Environ Health Insights* **2022**, *16*, 117863022210767. <https://doi.org/10.1177/11786302221076707>.
- (28) Chambers, W. S.; Hopkins, J. G.; Richards, S. M. A Review of Per- and Polyfluorinated Alkyl Substance Impairment of Reproduction. *Front. Toxicol.* **2021**, *3*, 732436. <https://doi.org/10.3389/tox.2021.732436>.
- (29) Abraham, K.; Mielke, H.; Fromme, H.; Völkel, W.; Menzel, J.; Peiser, M.; Zepp, F.; Willich, S. N.; Weikert, C. Internal Exposure to Perfluoroalkyl Substances (PFASs) and Biological Markers in 101 Healthy 1-Year-Old Children: Associations between Levels of Perfluorooctanoic Acid (PFOA) and Vaccine Response. *Arch Toxicol* **2020**, *94* (6), 2131–2147. <https://doi.org/10.1007/s00204-020-02715-4>.
- (30) Looker, C.; Luster, M. I.; Calafat, A. M.; Johnson, V. J.; Burleson, G. R.; Burleson, F. G.; Fletcher, T. Influenza Vaccine Response in Adults Exposed to Perfluorooctanoate and Perfluorooctanesulfonate. *Toxicol Sci* **2014**, *138* (1), 76–88. <https://doi.org/10.1093/toxsci/kft269>.
- (31) Bulka, C. M.; Avula, V.; Fry, R. C. Associations of Exposure to Perfluoroalkyl Substances Individually and in Mixtures with Persistent Infections: Recent Findings from NHANES 1999–2016. *Environmental Pollution* **2021**, *275*, 116619. <https://doi.org/10.1016/j.envpol.2021.116619>.
- (32) Mamsen, L. S.; Björvang, R. D.; Mucs, D.; Vinnars, M.-T.; Papadogiannakis, N.; Lindh, C. H.; Andersen, C. Y.; Damdimopoulou, P. Concentrations of Perfluoroalkyl Substances (PFASs) in Human Embryonic and Fetal Organs from First, Second, and Third Trimester Pregnancies. *Environment International* **2019**, *124*, 482–492. <https://doi.org/10.1016/j.envint.2019.01.010>.

- (33) Washino, N.; Saijo, Y.; Sasaki, S.; Kato, S.; Ban, S.; Konishi, K.; Ito, R.; Nakata, A.; Iwasaki, Y.; Saito, K.; Nakazawa, H.; Kishi, R. Correlations between Prenatal Exposure to Perfluorinated Chemicals and Reduced Fetal Growth. *Environ Health Perspect* **2009**, *117* (4), 660–667. <https://doi.org/10.1289/ehp.11681>.
- (34) Hoffman, K.; Webster, T. F.; Weisskopf, M. G.; Weinberg, J.; Vieira, V. M. Exposure to Polyfluoroalkyl Chemicals and Attention Deficit/Hyperactivity Disorder in U.S. Children 12–15 Years of Age. *Environ Health Perspect* **2010**, *118* (12), 1762–1767. <https://doi.org/10.1289/ehp.1001898>.
- (35) Vuong, A. M.; Yolton, K.; Webster, G. M.; Sjödin, A.; Calafat, A. M.; Braun, J. M.; Dietrich, K. N.; Lanphear, B. P.; Chen, A. Prenatal Polybrominated Diphenyl Ether and Perfluoroalkyl Substance Exposures and Executive Function in School-Age Children. *Environ Res* **2016**, *147*, 556–564. <https://doi.org/10.1016/j.envres.2016.01.008>.
- (36) Mora, A. M.; Oken, E.; Rifas-Shiman, S. L.; Webster, T. F.; Gillman, M. W.; Calafat, A. M.; Ye, X.; Sagiv, S. K. Prenatal Exposure to Perfluoroalkyl Substances and Adiposity in Early and Mid-Childhood. *Environ Health Perspect* **2017**, *125* (3), 467–473. <https://doi.org/10.1289/EHP246>.
- (37) Mora, A. M.; Fleisch, A. F.; Rifas-Shiman, S. L.; Baidal, J. W.; Pardo, L.; Webster, T. F.; Calafat, A. M.; Ye, X.; Oken, E.; Sagiv, S. K. Early Life Exposure to Per- and Polyfluoroalkyl Substances and Mid-Childhood Lipid and Alanine Aminotransferase Levels. *Environ Int* **2018**, *111*, 1–13. <https://doi.org/10.1016/j.envint.2017.11.008>.
- (38) Boyd, R. I.; Ahmad, S.; Singh, R.; Fazal, Z.; Prins, G. S.; Madak Erdogan, Z.; Irudayaraj, J.; Spinella, M. J. Toward a Mechanistic Understanding of Poly- and Perfluoroalkylated Substances and Cancer. *Cancers* **2022**, *14* (12), 2919. <https://doi.org/10.3390/cancers14122919>.
- (39) Stanifer, J. W.; Stapleton, H. M.; Souma, T.; Wittmer, A.; Zhao, X.; Boulware, L. E. Perfluorinated Chemicals as Emerging Environmental Threats to Kidney Health: A Scoping Review. *CJASN* **2018**, *13* (10), 1479–1492. <https://doi.org/10.2215/CJN.04670418>.
- (40) Steenland, K.; Winquist, A. PFAS and Cancer, a Scoping Review of the Epidemiologic Evidence. *Environmental Research* **2021**, *194*, 110690. <https://doi.org/10.1016/j.envres.2020.110690>.

- (41) Houde, M.; De Silva, A. O.; Muir, D. C. G.; Letcher, R. J. Monitoring of Perfluorinated Compounds in Aquatic Biota: An Updated Review: PFCs in Aquatic Biota. *Environ. Sci. Technol.* **2011**, *45* (19), 7962–7973. <https://doi.org/10.1021/es104326w>.
- (42) Ng, C. A.; Hungerbuehler, K. Exploring the Use of Molecular Docking to Identify Bioaccumulative Perfluorinated Alkyl Acids (PFAAs). *Environ. Sci. Technol.* **2015**, *49* (20), 12306–12314. <https://doi.org/10.1021/acs.est.5b03000>.
- (43) Bischel, H. N.; MacManus-Spencer, L. A.; Luthy, R. G. Noncovalent Interactions of Long-Chain Perfluoroalkyl Acids with Serum Albumin. *Environ. Sci. Technol.* **2010**, *44* (13), 5263–5269. <https://doi.org/10.1021/es101334s>.
- (44) Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol.* **2013**, *47* (19), 11293–11301. <https://doi.org/10.1021/es4026722>.
- (45) De Silva, A. O.; Armitage, J. M.; Bruton, T. A.; Dassuncao, C.; Heiger-Bernays, W.; Hu, X. C.; Kärrman, A.; Kelly, B.; Ng, C.; Robuck, A.; Sun, M.; Webster, T. F.; Sunderland, E. M. PFAS Exposure Pathways for Humans and Wildlife: A Synthesis of Current Knowledge and Key Gaps in Understanding. *Environ Toxicol Chem* **2021**, *40* (3), 631–657. <https://doi.org/10.1002/etc.4935>.
- (46) Backe, W. J.; Day, T. C.; Field, J. A. Zwitterionic, Cationic, and Anionic Fluorinated Chemicals in Aqueous Film Forming Foam Formulations and Groundwater from U.S. Military Bases by Nonaqueous Large-Volume Injection HPLC-MS/MS. *Environ. Sci. Technol.* **2013**, *47* (10), 5226–5234. <https://doi.org/10.1021/es3034999>.
- (47) Korucu, M. K.; Gedik, K.; Weber, R.; Karademir, A.; Kurt-Karakus, P. B. Inventory Development for Perfluorooctane Sulfonic Acid (PFOS) in Turkey: Challenges to Control Chemicals in Articles and Products. *Environ Sci Pollut Res* **2015**, *22* (19), 14537–14545. <https://doi.org/10.1007/s11356-014-3924-2>.
- (48) Eriksson, U.; Haglund, P.; Kärrman, A. Contribution of Precursor Compounds to the Release of Per- and Polyfluoroalkyl Substances (PFASs) from Waste Water Treatment Plants (WWTPs). *Journal of Environmental Sciences* **2017**, *61*, 80–90. <https://doi.org/10.1016/j.jes.2017.05.004>.
- (49) Xiao, F.; Halbach, T. R.; Simcik, M. F.; Gulliver, J. S. Input Characterization of Perfluoroalkyl Substances in Wastewater Treatment Plants: Source Discrimination by

- Exploratory Data Analysis. *Water Research* **2012**, *46* (9), 3101–3109. <https://doi.org/10.1016/j.watres.2012.03.027>.
- (50) Weber, R.; Watson, A.; Forter, M.; Oliaei, F. Review Article: Persistent Organic Pollutants and Landfills - a Review of Past Experiences and Future Challenges. *Waste Manag Res* **2011**, *29* (1), 107–121. <https://doi.org/10.1177/0734242X10390730>.
- (51) Aleksandrov, K.; Gehrmann, H.-J.; Hauser, M.; Mätzing, H.; Pigeon, D.; Stapf, D.; Wexler, M. Waste Incineration of Polytetrafluoroethylene (PTFE) to Evaluate Potential Formation of per- and Poly-Fluorinated Alkyl Substances (PFAS) in Flue Gas. *Chemosphere* **2019**, *226*, 898–906. <https://doi.org/10.1016/j.chemosphere.2019.03.191>.
- (52) Houtz, E. F.; Higgins, C. P.; Field, J. A.; Sedlak, D. L. Persistence of Perfluoroalkyl Acid Precursors in AFFF-Impacted Groundwater and Soil. *Environ. Sci. Technol.* **2013**, *47* (15), 8187–8195. <https://doi.org/10.1021/es4018877>.
- (53) Butt, C. M.; Muir, D. C. G.; Mabury, S. A. Biotransformation Pathways of Fluorotelomer-Based Polyfluoroalkyl Substances: A Review. *Environmental Toxicology and Chemistry* **2014**, *33* (2), 243–267. <https://doi.org/10.1002/etc.2407>.
- (54) Casas, G.; Martínez-Varela, A.; Roscales, J. L.; Vila-Costa, M.; Dachs, J.; Jiménez, B. Enrichment of Perfluoroalkyl Substances in the Sea-Surface Microlayer and Sea-Spray Aerosols in the Southern Ocean. *Environmental Pollution* **2020**, *267*, 115512. <https://doi.org/10.1016/j.envpol.2020.115512>.
- (55) Pike, K. A.; Edmiston, P. L.; Morrison, J. J.; Faust, J. A. Correlation Analysis of Perfluoroalkyl Substances in Regional U.S. Precipitation Events. *Water Research* **2021**, *190*, 116685. <https://doi.org/10.1016/j.watres.2020.116685>.
- (56) Schroeder, T.; Bond, D.; Foley, J. PFAS Soil and Groundwater Contamination via Industrial Airborne Emission and Land Deposition in SW Vermont and Eastern New York State, USA. *Environ. Sci.: Processes Impacts* **2021**, *23* (2), 291–301. <https://doi.org/10.1039/DOEM00427H>.
- (57) Vierke, L.; Ahrens, L.; Shoeib, M.; Palm, W.-U.; Webster, E. M.; Ellis, D. A.; Ebinghaus, R.; Harner, T. In Situ Air–Water and Particle–Water Partitioning of Perfluorocarboxylic Acids, Perfluorosulfonic Acids and Perfluorooctyl Sulfonamide at a Wastewater Treatment Plant. *Chemosphere* **2013**, *92* (8), 941–948. <https://doi.org/10.1016/j.chemosphere.2013.02.067>.

- (58) Fromme, H.; Dreyer, A.; Dietrich, S.; Fembacher, L.; Lahrz, T.; Völkel, W. Neutral Polyfluorinated Compounds in Indoor Air in Germany – The LUPE 4 Study. *Chemosphere* **2015**, *139*, 572–578. <https://doi.org/10.1016/j.chemosphere.2015.07.024>.
- (59) Lai, S.; Song, J.; Song, T.; Huang, Z.; Zhang, Y.; Zhao, Y.; Liu, G.; Zheng, J.; Mi, W.; Tang, J.; Zou, S.; Ebinghaus, R.; Xie, Z. Neutral Polyfluoroalkyl Substances in the Atmosphere over the Northern South China Sea. *Environmental Pollution* **2016**, *214*, 449–455. <https://doi.org/10.1016/j.envpol.2016.04.047>.
- (60) Ge, H.; Yamazaki, E.; Yamashita, N.; Taniyasu, S.; Ogata, A.; Furuuchi, M. Particle Size Specific Distribution of Perfluoro Alkyl Substances in Atmospheric Particulate Matter in Asian Cities. *Environ. Sci.: Processes Impacts* **2017**, *19* (4), 549–560. <https://doi.org/10.1039/C6EM00564K>.
- (61) Joerss, H.; Xie, Z.; Wagner, C. C.; Von Appen, W.-J.; Sunderland, E. M.; Ebinghaus, R. Transport of Legacy Perfluoroalkyl Substances and the Replacement Compound HFPO-DA through the Atlantic Gateway to the Arctic Ocean—Is the Arctic a Sink or a Source? *Environ. Sci. Technol.* **2020**, *54* (16), 9958–9967. <https://doi.org/10.1021/acs.est.0c00228>.
- (62) Rankin, K.; Mabury, S. A.; Jenkins, T. M.; Washington, J. W. A North American and Global Survey of Perfluoroalkyl Substances in Surface Soils: Distribution Patterns and Mode of Occurrence. *Chemosphere* **2016**, *161*, 333–341. <https://doi.org/10.1016/j.chemosphere.2016.06.109>.
- (63) Schaefer, C. E.; Culina, V.; Nguyen, D.; Field, J. Uptake of Poly- and Perfluoroalkyl Substances at the Air–Water Interface. *Environ. Sci. Technol.* **2019**, *53* (21), 12442–12448. <https://doi.org/10.1021/acs.est.9b04008>.
- (64) Langberg, H. A.; Breedveld, G. D.; Slinde, G. Aa.; Grønning, H. M.; Høisæter, Å.; Jartun, M.; Rundberget, T.; Jenssen, B. M.; Hale, S. E. Fluorinated Precursor Compounds in Sediments as a Source of Perfluorinated Alkyl Acids (PFAA) to Biota. *Environ. Sci. Technol.* **2020**, *54* (20), 13077–13089. <https://doi.org/10.1021/acs.est.0c04587>.
- (65) Smalling, K. L.; Romanok, K. M.; Bradley, P. M.; Morriss, M. C.; Gray, J. L.; Kanagy, L. K.; Gordon, S. E.; Williams, B. M.; Breitmeyer, S. E.; Jones, D. K.; DeCicco, L. A.; Eagles-Smith, C. A.; Wagner, T. Per- and Polyfluoroalkyl Substances (PFAS) in United States Tapwater: Comparison of Underserved Private-Well and Public-Supply Exposures and

- Associated Health Implications. *Environment International* **2023**, *178*, 108033. <https://doi.org/10.1016/j.envint.2023.108033>.
- (66) United States Environmental Protection Agency. *PFAS National Primary Drinking Water Regulation Rulemaking*; Document Citation: 88 FR 18638; Proposed Rule 2023–05471; 2023; pp 18638–18754. <https://www.federalregister.gov/documents/2023/03/29/2023-05471/pfas-national-primary-drinking-water-regulation-rulemaking#addresses>.
- (67) Barber, J. L.; Berger, U.; Chaemfa, C.; Huber, S.; Jahnke, A.; Temme, C.; Jones, K. C. Analysis of Per- and Polyfluorinated Alkyl Substances in Air Samples from Northwest Europe. *J. Environ. Monit.* **2007**, *9* (6), 530. <https://doi.org/10.1039/b701417a>.
- (68) Beser, M. I.; Pardo, O.; Beltrán, J.; Yusà, V. Determination of Per- and Polyfluorinated Substances in Airborne Particulate Matter by Microwave-Assisted Extraction and Liquid Chromatography–Tandem Mass Spectrometry. *Journal of Chromatography A* **2011**, *1218* (30), 4847–4855. <https://doi.org/10.1016/j.chroma.2011.04.082>.
- (69) Bräunig, J.; Baduel, C.; Barnes, C. M.; Mueller, J. F. Leaching and Bioavailability of Selected Perfluoroalkyl Acids (PFAAs) from Soil Contaminated by Firefighting Activities. *Science of The Total Environment* **2019**, *646*, 471–479. <https://doi.org/10.1016/j.scitotenv.2018.07.231>.
- (70) Mejia-Avenidaño, S.; Munoz, G.; Vo Duy, S.; Desrosiers, M.; Benoît, P.; Sauvé, S.; Liu, J. Novel Fluoroalkylated Surfactants in Soils Following Firefighting Foam Deployment During the Lac-Mégantic Railway Accident. *Environ. Sci. Technol.* **2017**, *51* (15), 8313–8323. <https://doi.org/10.1021/acs.est.7b02028>.
- (71) Milley, S. A.; Koch, I.; Fortin, P.; Archer, J.; Reynolds, D.; Weber, K. P. Estimating the Number of Airports Potentially Contaminated with Perfluoroalkyl and Polyfluoroalkyl Substances from Aqueous Film Forming Foam: A Canadian Example. *Journal of Environmental Management* **2018**, *222*, 122–131. <https://doi.org/10.1016/j.jenvman.2018.05.028>.
- (72) Liu, Z.; Lu, Y.; Song, X.; Jones, K.; Sweetman, A. J.; Johnson, A. C.; Zhang, M.; Lu, X.; Su, C. Multiple Crop Bioaccumulation and Human Exposure of Perfluoroalkyl Substances around a Mega Fluorochemical Industrial Park, China: Implication for Planting Optimization and Food Safety. *Environment International* **2019**, *127*, 671–684. <https://doi.org/10.1016/j.envint.2019.04.008>.

- (73) Sepulvado, J. G.; Blaine, A. C.; Hundal, L. S.; Higgins, C. P. Occurrence and Fate of Perfluorochemicals in Soil Following the Land Application of Municipal Biosolids. *Environ. Sci. Technol.* **2011**, *45* (19), 8106–8112. <https://doi.org/10.1021/es103903d>.
- (74) Zareitalabad, P.; Siemens, J.; Hamer, M.; Amelung, W. Perfluorooctanoic Acid (PFOA) and Perfluorooctanesulfonic Acid (PFOS) in Surface Waters, Sediments, Soils and Wastewater – A Review on Concentrations and Distribution Coefficients. *Chemosphere* **2013**, *91* (6), 725–732. <https://doi.org/10.1016/j.chemosphere.2013.02.024>.
- (75) Lenka, S. P.; Kah, M.; Padhye, L. P. A Review of the Occurrence, Transformation, and Removal of Poly- and Perfluoroalkyl Substances (PFAS) in Wastewater Treatment Plants. *Water Research* **2021**, *199*, 117187. <https://doi.org/10.1016/j.watres.2021.117187>.
- (76) Llorca, M.; Farré, M.; Picó, Y.; Müller, J.; Knepper, T. P.; Barceló, D. Analysis of Perfluoroalkyl Substances in Waters from Germany and Spain. *Science of The Total Environment* **2012**, *431*, 139–150. <https://doi.org/10.1016/j.scitotenv.2012.05.011>.
- (77) Coggan, T. L.; Moodie, D.; Kolobaric, A.; Szabo, D.; Shimeta, J.; Crosbie, N. D.; Lee, E.; Fernandes, M.; Clarke, B. O. An Investigation into Per- and Polyfluoroalkyl Substances (PFAS) in Nineteen Australian Wastewater Treatment Plants (WWTPs). *Heliyon* **2019**, *5* (8), e02316. <https://doi.org/10.1016/j.heliyon.2019.e02316>.
- (78) Beecher, Ned; Juliana Beecher; Janine Burke-Wells; Maile Lono-Batura; Nora Goldstein; Greg Kester; Bill Toffey. National Biosolids Data Project: Biosolids Management in the U.S., 2022. <https://www.biosolidsdata.org>, (accessed 2023-07-22).
- (79) Washington, J. W.; Yoo, H.; Ellington, J. J.; Jenkins, T. M.; Libelo, E. L. Concentrations, Distribution, and Persistence of Perfluoroalkylates in Sludge-Applied Soils near Decatur, Alabama, USA. *Environ. Sci. Technol.* **2010**, *44* (22), 8390–8396. <https://doi.org/10.1021/es1003846>.
- (80) Brambilla, G.; D’Hollander, W.; Oliaei, F.; Stahl, T.; Weber, R. Pathways and Factors for Food Safety and Food Security at PFOS Contaminated Sites within a Problem Based Learning Approach. *Chemosphere* **2015**, *129*, 192–202. <https://doi.org/10.1016/j.chemosphere.2014.09.050>.
- (81) Death, C.; Bell, C.; Champness, D.; Milne, C.; Reichman, S.; Hagen, T. Per- and Polyfluoroalkyl Substances (PFAS) in Livestock and Game Species: A Review. *Science of The Total Environment* **2021**, *774*, 144795. <https://doi.org/10.1016/j.scitotenv.2020.144795>.

- (82) Fernandes, A. R.; Lake, I. R.; Dowding, A.; Rose, M.; Jones, N. R.; Petch, R.; Smith, F.; Panton, S. The Potential of Recycled Materials Used in Agriculture to Contaminate Food through Uptake by Livestock. *Science of The Total Environment* **2019**, *667*, 359–370. <https://doi.org/10.1016/j.scitotenv.2019.02.211>.
- (83) Guruge, K. S.; Manage, P. M.; Yamanaka, N.; Miyazaki, S.; Taniyasu, S.; Yamashita, N. Species-Specific Concentrations of Perfluoroalkyl Contaminants in Farm and Pet Animals in Japan. *Chemosphere* **2008**, *73* (1), S210–S215. <https://doi.org/10.1016/j.chemosphere.2006.12.105>.
- (84) Hlouskova, V.; Hradkova, P.; Poustka, J.; Brambilla, G.; De Filippis, S. P.; D'Hollander, W.; Bervoets, L.; Herzke, D.; Huber, S.; De Voogt, P.; Pulkrabova, J. Occurrence of Perfluoroalkyl Substances (PFASs) in Various Food Items of Animal Origin Collected in Four European Countries. *Food Additives & Contaminants: Part A* **2013**, *30* (11), 1918–1932. <https://doi.org/10.1080/19440049.2013.837585>.
- (85) Van Asselt, E. D.; Kowalczyk, J.; Van Eijkeren, J. C. H.; Zeilmaker, M. J.; Ehlers, S.; Fürst, P.; Lahrssen-Wiederholt, M.; Van Der Fels-Klerx, H. J. Transfer of Perfluorooctane Sulfonic Acid (PFOS) from Contaminated Feed to Dairy Milk. *Food Chemistry* **2013**, *141* (2), 1489–1495. <https://doi.org/10.1016/j.foodchem.2013.04.035>.
- (86) Blaine, A. C.; Rich, C. D.; Hundal, L. S.; Lau, C.; Mills, M. A.; Harris, K. M.; Higgins, C. P. Uptake of Perfluoroalkyl Acids into Edible Crops via Land Applied Biosolids: Field and Greenhouse Studies. *Environ. Sci. Technol.* **2013**, *47* (24), 14062–14069. <https://doi.org/10.1021/es403094q>.
- (87) Krippner, J.; Falk, S.; Brunn, H.; Georgii, S.; Schubert, S.; Stahl, T. Accumulation Potentials of Perfluoroalkyl Carboxylic Acids (PFCAs) and Perfluoroalkyl Sulfonic Acids (PFASs) in Maize (*Zea Mays*). *J. Agric. Food Chem.* **2015**, *63* (14), 3646–3653. <https://doi.org/10.1021/acs.jafc.5b00012>.
- (88) Stahl, T.; Heyn, J.; Thiele, H.; Hüther, J.; Failing, K.; Georgii, S.; Brunn, H. Carryover of Perfluorooctanoic Acid (PFOA) and Perfluorooctane Sulfonate (PFOS) from Soil to Plants. *Arch Environ Contam Toxicol* **2009**, *57* (2), 289–298. <https://doi.org/10.1007/s00244-008-9272-9>.
- (89) Blaine, A. C.; Rich, C. D.; Sedlacko, E. M.; Hyland, K. C.; Stushnoff, C.; Dickenson, E. R. V.; Higgins, C. P. Perfluoroalkyl Acid Uptake in Lettuce (*Lactuca Sativa*) and Strawberry

- (*Fragaria Ananassa*) Irrigated with Reclaimed Water. *Environ. Sci. Technol.* **2014**, *48* (24), 14361–14368. <https://doi.org/10.1021/es504150h>.
- (90) Bizkarguenaga, E.; Zabaleta, I.; Prieto, A.; Fernández, L. A.; Zuloaga, O. Uptake of 8:2 Perfluoroalkyl Phosphate Diester and Its Degradation Products by Carrot and Lettuce from Compost-Amended Soil. *Chemosphere* **2016**, *152*, 309–317. <https://doi.org/10.1016/j.chemosphere.2016.02.130>.
- (91) Blaine, A. C.; Rich, C. D.; Sedlacko, E. M.; Hundal, L. S.; Kumar, K.; Lau, C.; Mills, M. A.; Harris, K. M.; Higgins, C. P. Perfluoroalkyl Acid Distribution in Various Plant Compartments of Edible Crops Grown in Biosolids-Amended Soils. *Environ. Sci. Technol.* **2014**, *48* (14), 7858–7865. <https://doi.org/10.1021/es500016s>.
- (92) United States Environmental Protection Agency. EPA's Per- and Polyfluoroalkyl Substances (PFAS) Action Plan EPA 823R18004, 2019.
- (93) J.A. Shoemaker; Grimmett, P.; Boutin, B. Determination of Selected Perfluorinated Alkyl Acids in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS), 2008.
- (94) Rosenblum, Laura, W., Steven. METHOD 533: DETERMINATION OF PER- AND POLYFLUOROALKYL SUBSTANCES IN DRINKING WATER BY ISOTOPE DILUTION ANION EXCHANGE SOLID PHASE EXTRACTION AND LIQUID CHROMATOGRAPHY/TANDEM MASS SPECTROMETRY, 2019.
- (95) Shoemaker, J., D. T. Method 537.1 Determination of Selected Per- and Polyfluorinated Alkyl Substances in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS), 2020.
- (96) Hu, X.; Liu, Y.; Yang, R.; Jiang, Y.; Meng, M.; Liu, Z.; Ni, L.; Wu, W.; Liu, H. Adsorption for Perfluorooctanoic Acid with Graphitic-Phase Carbon Nitride and Its HPLC Fluorescence Determination. *The Canadian Journal of Chemical Engineering* **2020**, *98* (1), 394–403. <https://doi.org/10.1002/cjce.23625>.
- (97) Fang, C.; Zhang, X.; Dong, Z.; Wang, L.; Megharaj, M.; Naidu, R. Smartphone App-Based/Portable Sensor for the Detection of Fluoro-Surfactant PFOA. *Chemosphere* **2018**, *191*, 381–388. <https://doi.org/10.1016/j.chemosphere.2017.10.057>.

- (98) Xia, W.; Wan, Y.-J.; Wang, X.; Li, Y.-Y.; Yang, W.-J.; Wang, C.-X.; Xu, S. Sensitive Bioassay for Detection of PPAR α Potentially Hazardous Ligands with Gold Nanoparticle Probe. *J Hazard Mater* **2011**, *192* (3), 1148–1154. <https://doi.org/10.1016/j.jhazmat.2011.06.023>.
- (99) Feng, H.; Wang, N.; Tran, T. T.; Yuan, L.; Li, J.; Cai, Q. Surface Molecular Imprinting on Dye-(NH₂)-SiO₂ NPs for Specific Recognition and Direct Fluorescent Quantification of Perfluorooctane Sulfonate. *Sensors and Actuators B: Chemical* **2014**, *195*, 266–273. <https://doi.org/10.1016/j.snb.2014.01.036>.
- (100) Jiao, Z.; Li, J.; Mo, L.; Liang, J.; Fan, H. A Molecularly Imprinted Chitosan Doped with Carbon Quantum Dots for Fluorometric Determination of Perfluorooctane Sulfonate. *Microchim Acta* **2018**, *185* (10), 473. <https://doi.org/10.1007/s00604-018-2996-y>.
- (101) Pitruzzella, R.; Arcadio, F.; Perri, C.; Del Prete, D.; Porto, G.; Zeni, L.; Cennamo, N. Ultra-Low Detection of Perfluorooctanoic Acid Using a Novel Plasmonic Sensing Approach Combined with Molecularly Imprinted Polymers. *Chemosensors* **2023**, *11* (4), 211. <https://doi.org/10.3390/chemosensors11040211>.
- (102) Mehrotra, P. Biosensors and Their Applications – A Review. *J Oral Biol Craniofac Res* **2016**, *6* (2), 153–159. <https://doi.org/10.1016/j.jobcr.2015.12.002>.
- (103) Justino, C.; Duarte, A.; Rocha-Santos, T. Recent Progress in Biosensors for Environmental Monitoring: A Review. *Sensors* **2017**, *17* (12), 2918. <https://doi.org/10.3390/s17122918>.
- (104) Bhalla, N.; Jolly, P.; Formisano, N.; Estrela, P. Introduction to Biosensors. *Essays in Biochemistry* **2016**, *60* (1), 1–8. <https://doi.org/10.1042/EBC20150001>.
- (105) Naresh, Varnakavi.; Lee, N. A Review on Biosensors and Recent Development of Nanostructured Materials-Enabled Biosensors. *Sensors* **2021**, *21* (4), 1109. <https://doi.org/10.3390/s21041109>.
- (106) Kylilis, N.; Riangrunroj, P.; Lai, H.-E.; Salema, V.; Fernández, L. Á.; Stan, G.-B. V.; Freemont, P. S.; Polizzi, K. M. Whole-Cell Biosensor with Tunable Limit of Detection Enables Low-Cost Agglutination Assays for Medical Diagnostic Applications. *ACS Sens.* **2019**, *4* (2), 370–378. <https://doi.org/10.1021/acssensors.8b01163>.
- (107) Belkin, S. Microbial Whole-Cell Sensing Systems of Environmental Pollutants. *Current Opinion in Microbiology* **2003**, *6* (3), 206–212. [https://doi.org/10.1016/S1369-5274\(03\)00059-6](https://doi.org/10.1016/S1369-5274(03)00059-6).

- (108) Harms, H.; Wells, M. C.; Van Der Meer, J. R. Whole-Cell Living Biosensors—Are They Ready for Environmental Application? *Appl Microbiol Biotechnol* **2006**, *70* (3), 273–280. <https://doi.org/10.1007/s00253-006-0319-4>.
- (109) S. Kucherenko, I.; O. Soldatkin, O.; Yu. Kucherenko, D.; V. Soldatkina, O.; V. Dzyadevych, S. Advances in Nanomaterial Application in Enzyme-Based Electrochemical Biosensors: A Review. *Nanoscale Advances* **2019**, *1* (12), 4560–4577. <https://doi.org/10.1039/C9NA00491B>.
- (110) Kumar, H.; Kumari, N.; Sharma, R. Nanocomposites (Conducting Polymer and Nanoparticles) Based Electrochemical Biosensor for the Detection of Environment Pollutant: Its Issues and Challenges. *Environmental Impact Assessment Review* **2020**, *85*, 106438. <https://doi.org/10.1016/j.eiar.2020.106438>.
- (111) Bahadır, E. B.; Sezgentürk, M. K. Applications of Commercial Biosensors in Clinical, Food, Environmental, and Biothreat/Biowarfare Analyses. *Analytical Biochemistry* **2015**, *478*, 107–120. <https://doi.org/10.1016/j.ab.2015.03.011>.
- (112) Ross, I.; McDonough, J.; Miles, J.; Storch, P.; Thelakkat Kochunarayanan, P.; Kalve, E.; Hurst, J.; S. Dasgupta, S.; Burdick, J. A Review of Emerging Technologies for Remediation of PFASs. *Remediation* **2018**, *28* (2), 101–126. <https://doi.org/10.1002/rem.21553>.
- (113) Kwon, B. G.; Lim, H.-J.; Na, S.-H.; Choi, B.-I.; Shin, D.-S.; Chung, S.-Y. Biodegradation of Perfluorooctanesulfonate (PFOS) as an Emerging Contaminant. *Chemosphere* **2014**, *109*, 221–225. <https://doi.org/10.1016/j.chemosphere.2014.01.072>.
- (114) Meegoda, J. N.; Kewalramani, J. A.; Li, B.; Marsh, R. W. A Review of the Applications, Environmental Release, and Remediation Technologies of Per- and Polyfluoroalkyl Substances. *International Journal of Environmental Research and Public Health* **2020**, *17* (21), 8117. <https://doi.org/10.3390/ijerph17218117>.
- (115) Horst, J.; McDonough, J.; Ross, I.; Dickson, M.; Miles, J.; Hurst, J.; Storch, P. Water Treatment Technologies for PFAS: The Next Generation. *Groundwater Monitoring & Remediation* **2018**, *38* (2), 13–23. <https://doi.org/10.1111/gwmmr.12281>.
- (116) Darlington, R.; Barth, E.; McKernan, J. The Challenges of PFAS Remediation. *Mil Eng* **2018**, *110* (712), 58–60.
- (117) Watanabe, N.; Takata, M.; Takemine, S.; Yamamoto, K. Thermal Mineralization Behavior of PFOA, PFHxA, and PFOS during Reactivation of Granular Activated Carbon (GAC) in

- Nitrogen Atmosphere. *Environ Sci Pollut Res* **2018**, 25 (8), 7200–7205. <https://doi.org/10.1007/s11356-015-5353-2>.
- (118) Yamada, T.; Taylor, P. H.; Buck, R. C.; Kaiser, M. A.; Giraud, R. J. Thermal Degradation of Fluorotelomer Treated Articles and Related Materials. *Chemosphere* **2005**, 61 (7), 974–984. <https://doi.org/10.1016/j.chemosphere.2005.03.025>.
- (119) Gomez-Ruiz, B.; Gómez-Lavín, S.; Diban, N.; Boiteux, V.; Colin, A.; Dauchy, X.; Urriaga, A. Efficient Electrochemical Degradation of Poly- and Perfluoroalkyl Substances (PFASs) from the Effluents of an Industrial Wastewater Treatment Plant. *Chemical Engineering Journal* **2017**, 322, 196–204. <https://doi.org/10.1016/j.cej.2017.04.040>.
- (120) Trautmann, A. M.; Schell, H.; Schmidt, K. R.; Mangold, K.-M.; Tiehm, A. Electrochemical Degradation of Perfluoroalkyl and Polyfluoroalkyl Substances (PFASs) in Groundwater. *Water Science and Technology* **2015**, 71 (10), 1569–1575. <https://doi.org/10.2166/wst.2015.143>.
- (121) Nzeribe, B. N.; Crimi, M.; Mededovic Thagard, S.; Holsen, T. M. Physico-Chemical Processes for the Treatment of Per- And Polyfluoroalkyl Substances (PFAS): A Review. *Critical Reviews in Environmental Science and Technology* **2019**, 49 (10), 866–915. <https://doi.org/10.1080/10643389.2018.1542916>.
- (122) Hale, S. E.; Arp, H. P. H.; Slinde, G. A.; Wade, E. J.; Bjørseth, K.; Breedveld, G. D.; Straith, B. F.; Moe, K. G.; Jartun, M.; Høisæter, Å. Sorbent Amendment as a Remediation Strategy to Reduce PFAS Mobility and Leaching in a Contaminated Sandy Soil from a Norwegian Firefighting Training Facility. *Chemosphere* **2017**, 171, 9–18. <https://doi.org/10.1016/j.chemosphere.2016.12.057>.
- (123) Söregård, M.; Lindh, A.-S.; Ahrens, L. Thermal Desorption as a High Removal Remediation Technique for Soils Contaminated with Per- and Polyfluoroalkyl Substances (PFASs). *PLoS ONE* **2020**, 15 (6), e0234476. <https://doi.org/10.1371/journal.pone.0234476>.
- (124) Kupryianchyk, D.; Hale, S. E.; Breedveld, G. D.; Cornelissen, G. Treatment of Sites Contaminated with Perfluorinated Compounds Using Biochar Amendment. *Chemosphere* **2016**, 142, 35–40. <https://doi.org/10.1016/j.chemosphere.2015.04.085>.
- (125) McGregor, R. *In Situ* Treatment of PFAS-Impacted Groundwater Using Colloidal Activated Carbon. *Remediation* **2018**, 28 (3), 33–41. <https://doi.org/10.1002/rem.21558>.

- (126) Aly, Y. H.; McInnis, D. P.; Lombardo, S. M.; Arnold, W. A.; Pennell, K. D.; Hatton, J.; Simcik, M. F. Enhanced Adsorption of Perfluoro Alkyl Substances for *in Situ* Remediation. *Environ. Sci.: Water Res. Technol.* **2019**, *5* (11), 1867–1875. <https://doi.org/10.1039/C9EW00426B>.
- (127) Anderson, E. L.; Mousavi, M. P. S.; Aly, Y. H.; Chen, X. V.; Simcik, M. F.; Bühlmann, P. Remediation of Perfluorooctylsulfonate Contamination by *in Situ* Sequestration: Direct Monitoring of PFOS Binding to Polyquaternium Polymers. *ACS Omega* **2019**, *4* (1), 1068–1076. <https://doi.org/10.1021/acsomega.8b03275>.
- (128) Zhou, Q.; Deng, S.; Yu, Q.; Zhang, Q.; Yu, G.; Huang, J.; He, H. Sorption of Perfluorooctane Sulfonate on Organo-Montmorillonites. *Chemosphere* **2010**, *78* (6), 688–694. <https://doi.org/10.1016/j.chemosphere.2009.12.005>.
- (129) Qian, J.; Shen, M.; Wang, P.; Wang, C.; Hou, J.; Ao, Y.; Liu, J.; Li, K. Adsorption of Perfluorooctane Sulfonate on Soils: Effects of Soil Characteristics and Phosphate Competition. *Chemosphere* **2017**, *168*, 1383–1388. <https://doi.org/10.1016/j.chemosphere.2016.11.114>.
- (130) Wei, C.; Song, X.; Wang, Q.; Hu, Z. Sorption Kinetics, Isotherms and Mechanisms of PFOS on Soils with Different Physicochemical Properties. *Ecotoxicology and Environmental Safety* **2017**, *142*, 40–50. <https://doi.org/10.1016/j.ecoenv.2017.03.040>.
- (131) Prasad, M. N. V.; Nakbanpote, W.; Sebastian, A.; Panitlertumpai, N.; Phadermrod, C. Phytomanagement of Padaeng Zinc Mine Waste, Mae Sot District, Tak Province, Thailand. In *Soil Remediation and Plants*; Elsevier, 2015; pp 661–687. <https://doi.org/10.1016/B978-0-12-799937-1.00023-1>.
- (132) Dushenkov, Viatcheslav.; Kumar, P. B. A. Nanda.; Motto, Harry.; Raskin, Ilya. Rhizofiltration: The Use of Plants to Remove Heavy Metals from Aqueous Streams. *Environ. Sci. Technol.* **1995**, *29* (5), 1239–1245. <https://doi.org/10.1021/es00005a015>.
- (133) Cunningham, S. D.; Ow, D. W. Promises and Prospects of Phytoremediation. *Plant Physiol.* **1996**, *110* (3), 715–719. <https://doi.org/10.1104/pp.110.3.715>.
- (134) Černe, M.; Smodiš, B.; Štok, M. Uptake of Radionuclides by a Common Reed (*Phragmites Australis* (Cav.) Trin. Ex Steud.) Grown in the Vicinity of the Former Uranium Mine at Žirovski Vrh. *Nuclear Engineering and Design* **2011**, *241* (4), 1282–1286. <https://doi.org/10.1016/j.nucengdes.2010.04.003>.

- (135) Huang, H.; Yu, N.; Wang, L.; Gupta, D. K.; He, Z.; Wang, K.; Zhu, Z.; Yan, X.; Li, T.; Yang, X. The Phytoremediation Potential of Bioenergy Crop *Ricinus Communis* for DDTs and Cadmium Co-Contaminated Soil. *Bioresource Technology* **2011**, *102* (23), 11034–11038. <https://doi.org/10.1016/j.biortech.2011.09.067>.
- (136) Hou, J.; Liu, W.; Wang, B.; Wang, Q.; Luo, Y.; Franks, A. E. PGPR Enhanced Phytoremediation of Petroleum Contaminated Soil and Rhizosphere Microbial Community Response. *Chemosphere* **2015**, *138*, 592–598. <https://doi.org/10.1016/j.chemosphere.2015.07.025>.
- (137) Pandey, V. C.; Bajpai, O. Phytoremediation. In *Phytomanagement of Polluted Sites*; Elsevier, 2019; pp 1–49. <https://doi.org/10.1016/B978-0-12-813912-7.00001-6>.

Chapter 2

Engineering human liver fatty acid binding protein for detection of per- and polyfluoroalkyl substances

This chapter has been adapted from the following publication: Mann, M.M., Tang, J.D., Berger, B.W. “Engineering human liver fatty acid binding protein for detection of poly- and perfluoroalkyl substances.” *Biotechnology and Bioengineering* **2022**

2.1 Abstract

Per- and polyfluoroalkyl substances (PFAS) are a large group of synthetic fluorinated chemicals with surface active and water-repellent properties. The combination of wide-spread use in numerous consumer and industrial products and extended biological half-lives arising from strong carbon-fluorine bonds has led to significant accumulation of PFAS in humans. As most human interaction with PFAS comes from ingestion, it is important to be able to detect PFAS in drinking water as well as in agricultural water. Here we present an approach to designing a fluorescence-based biosensor for the rapid detection of PFAS based on human liver fatty acid binding protein (hLFABP). Introduction of solvatochromic fluorophores within the ligand binding pocket (F50) allowed for intrinsic detection of PFOA, PFOS, and PFHxS via blue-shifts in fluorescence emission spectra. Initially, a single tryptophan mutation (F50W) was found to be able to detect PFOA with a LOD of 2.8ppm. We improved the sensitivity of the biosensor by exchanging tryptophan for the thiol reactive fluorophore, acrylodan. The acrylodan conjugated C69S/F50C

hLFABP variant is capable of detecting PFOA, PFOS, and PFHxS in PBS with LODs of 112ppb, 345 ppb, and 1.09 ppm respectively. The protein-based sensor is also capable of detecting these contaminants at similar ranges in spiked environmental water samples, including samples containing an interfering anionic surfactant SDS. Overall, this work demonstrates engineered hLFABP is a useful platform for detection of PFAS in environmental water samples and highlights its ease of use and versatility in field applications.

2.2 Introduction

Per- and polyfluoroalkyl substances (PFAS) have become increasingly notorious environmental contaminants and a growing public health concern. This large group of chemicals contain hydrophobic, highly fluorinated carbon chains (C₆-C₁₀) and terminal hydrophilic headgroups. These headgroups are typically comprised of carboxylic acids or sulfates, which impart unique interfacial properties such as the ability to repel oils and water. As a result, PFAS are abundant in everyday items including stain and water-resistant fabrics, non-stick cookware, and even hygienic products like dental floss¹⁻⁵. Additionally, aqueous film-forming foams (AFFFs) containing PFAS have been used for decades at airports as well as various manufacturing facilities where highly flammable fuel or solvents are stored⁶.

Prolonged exposure to PFAS has been linked to several negative health effects,⁷ including increased cholesterol levels,^{8,9} liver and kidney disease^{10,11} as well as impairment of the immune system. Recent studies have suggested high levels of PFOA are correlated with reduced vaccine response,^{12,13} and chronic exposure to PFAS could render a person more susceptible to pathogens or viral infections¹⁴. PFAS has been shown to transfer from mother to fetus,¹⁵ and prenatal

exposure to PFAS has been linked to low birth weight,¹⁶ neurobehavioral issues,^{17,18} as well as childhood adiposity^{19,20}.

The current standard for detection of PFAS is via liquid and gas chromatography coupled with mass spectrometry.²¹ This method is highly sensitive, capable of detecting ppt levels depending on the type of sample matrix being tested. These methods also require highly sophisticated equipment and training, and therefore is not readily available for widespread, rapid use in testing foods, commercial products or water samples.^{21,22} Given the diversity of PFAS contaminated materials, the range of concentrations necessary for testing, and desire for portable and flexible testing, there is an increasing need to develop rapid and robust methodologies for PFAS detection.

In contrast to synthetic scaffolds such as nanoparticles or polymers, engineered proteins can be tuned for selectivity and affinity for a given ligand from complex mixtures (e.g., environmental samples) and can incorporate fluorescence and other highly sensitive optical detection methods. One such potential protein scaffold is the human liver fatty acid binding protein (hLFABP), a cytosolic protein expressed in the liver, kidneys, and intestines²³ whose expression has been shown to be upregulated in response to PFOA exposure.²⁴ LFABP contains a large hydrophobic β barrel binding region composed of antiparallel β sheets with two α helices covering one end creating a helix-turn-helix (HTH) motif.^{25,26} This well-defined β barrel motif is typical of other fatty acid binding proteins, but LFABP is unique in its ability to bind two cognate ligands.^{25,26} PFAS, particularly perfluoroalkyl acids, share a structural similarity to the natural ligands of hLFABP. In an attempt to understand toxicity and bioaccumulation, others have shown that PFAS can bind to the protein in a similar manner as native fatty acid ligands.²⁷ Also, the native forms of both rat and

human LFABP binds PFOA and PFOS as well as short-chain PFHxS with moderate affinity and micromolar dissociation constants (K_d).^{24,28}

In this study, we describe a strategy to reengineer human LFABP (hLFABP) as a scaffold for PFOA, PFOS, and PFHxS detection via the introduction of solvatochromic fluorophores into the protein's ligand binding pocket. Using a structure-guided approach, we identified key mutations in the inner binding pocket that are sensitive to PFAS ligand binding as a function of chain length and demonstrated binding using intrinsic tryptophan fluorescence. By modifying one such residue (Phe50) with the solvatochromic fluorophore, 6-acryloyl-2-dimethylaminonaphthalene [acrylodan], we significantly increased the limit of detection to ppb levels. Furthermore, we demonstrate robustness to and specificity of PFAS ligand binding in surface water samples, and in samples spiked with the anionic surfactant, sodium dodecyl sulfate (SDS), which can compete with PFAS binding. Collectively, our results indicate that hLFABP can be reengineered to significantly improve detection for PFAS ligands and suggests possible improvements for further detection of this challenging and important class of environmental toxins.

2.3 Materials and Methods

2.3.1 Molecular biology

An *E. coli* codon optimized form of human liver fatty acid binding protein (NCBI 2168) was synthesized and subcloned into the pET-28a(+) vector using BamHI/XhoI restriction sites (GenScript). The gene includes N- and C-terminal hexahistidine tags. Unless otherwise stated, all other molecular biology procedures for PCR amplification, plasmid preparation, cell transformation and subcloning were performed according to standard methods supplied by

manufacturers. For site-directed mutagenesis, primers were designed using PrimerX, and mutations were introduced using the QuikChange II site-directed mutagenesis kit (Agilent). Mutated sequences were verified by DNA sequencing (GeneWiz). For DNA maintenance, *E. coli* strain DH5 α was used and for protein expression *E. coli* strain BL21 (DE3) was used.

2.3.2 Protein expression and purification

Transformed *E. coli* BL21 cells were grown to saturation in 10 mL of LB containing kanamycin (50 μ g/mL) overnight at 37°C with 200 rpm agitation. The next day, cells were pelleted by centrifugation (3,000 g), and resuspended in fresh media containing kanamycin. The suspension was grown to an OD of 0.7 at 37°C. Protein expression was then induced via addition of 1mM isopropyl β -D-1-thiogalactopyranoside (IPTG), and cultures transferred to 20°C. Cells were harvested after an 18-hour growth by centrifugation (12,000 g) and resuspended in lysis buffer (50mM Tris-Cl, 100mM NaCl, 5% v/v glycerol) before lysis via sonication. The lysis mixture was clarified by centrifugation (12,000 g), and the insoluble fraction discarded. The supernatant containing soluble protein was then separated by affinity chromatography on a 1mL HisTrap HP column using an AKTA Pure system (GE Healthcare). The column was equilibrated with 50 mM Tris-Cl buffer (pH 8) containing 10 mM imidazole, and protein was separated using a stepwise elution of imidazole up to 500 mM. The fractions collected were analyzed by SDS-PAGE to determine purity. Pure hLFABP containing fractions were pooled and dialyzed against 50 mM Tris-Cl buffer (pH 8) or phosphate buffered saline (PBS) buffer (pH 7.5).

Delipidation of the protein was performed at 37°C using hydroxyalkoxypropyl-dextran (Sigma Aldrich), a Lipidex-5000 equivalent resin. Resin was equilibrated with buffer for 1 hour at 37°C

before sample was applied and incubated for 2 hours at 37°C with gentle agitation. Protein samples were collected after centrifugation and filtration through a 0.2 micrometer filter. Concentration of all protein samples was measured using a Pierce BCA protein assay kit.

2.3.3 Acrylodan Labelling

Acrylodan labelling reactions containing 20 mM protein, 50 mM acrylodan and 8 M urea in 5 mL of PBS (pH 7.5) were incubated at room temperature for 10 hours with gentle mixing. Removal of unreacted acrylodan and refolding of denatured protein was performed on-column with a gradient from 8 M to 0 M urea in 50 mM Tris-HCl buffer (pH 8) containing 10 mM imidazole. The protein was then eluted using 500 mM imidazole containing buffer, pooled, and dialyzed into PBS buffer overnight. The degree of labelling was determined by measuring the protein concentration using a BCA protein assay, as well as acrylodan concentration measured by absorption at 370 nm (extinction coefficient $16,400 \text{ M}^{-1}\text{cm}^{-1}$).²⁹ Degree of labelling for C69S/F50C hLFABP was 1.0.

2.3.4 Fluorescence displacement assays

Binding of the fluorophore 1-anilinonaphthalene-8-sulfonic acid (ANS) to hLFABP was performed in a similar manner as described previously in order to obtain max fluorescence intensity for the complex.³⁰ Briefly, hLFABP (1 μM) in either 50 mM Tris buffer containing 1 mM BME or PBS buffer (pH 7.5) was titrated with ANS (0-50 μM). After an equilibration time of 5 minutes, fluorescence emission spectra between 420 and 600 nm as well as end point intensity at 470 nm were collected with excitation at 400 nm. The data were corrected for background free protein and ANS fluorescence. The maximum fluorescence upon saturation of fluorophore (F_{max}) as well as the dissociation constant ($K_{\text{d, ANS}}$) were determined by fitting the fluorescence intensity

(F) and corresponding concentration of ANS ([ANS]) to a one-site binding model using nonlinear regression:

$$F = F_{\max} * \frac{[\text{ANS}]}{(K_{d,\text{ANS}} + [\text{ANS}])} \quad (1)$$

This $K_{d,\text{ANS}}$ value is equivalent to the concentration of free fluorophore in the system at half the maximum fluorescence, and was used to calculate the inhibition constant (K_i) of PFAS for the proteins.

The binding of PFAS to hLFABP variants was measured by displacement of bound ANS. hLFABP (2 μM) was first equilibrated with ANS (100 μM) in excess for 5 minutes. PFAS was then titrated into hLFABP-ANS samples resulting in a final protein concentration of 1 μM . After a 5-minute incubation, fluorescence emissions spectra were collected between 420-600 nm with excitation at 400 nm. Measurements were corrected using blanks containing the ANS-hLFABP complex as well as protein in buffer only. The displacement of ANS was characterized as the percent loss in fluorescence and was calculated as the decrease in area under the emission spectra for various PFAS concentrations. The concentration of PFAS necessary to displace half of the bound ANS and inhibit fluorescence by 50% (IC_{50}) was found by fitting displacement data to a sigmoidal dose-response curve where F_{\min} and F_{\max} represent the maximum and minimum ANS displacement by PFAS respectively:

$$\% \text{ Initial Fluorescence} = F_{\min} + \frac{(F_{\max} - F_{\min})}{(1 + 10^{(\log(\text{IC}_{50}) - \log([\text{PFOA}] * \text{HS}))})} \quad (2)$$

The inhibition constants (K_i) were then found by relating the value to the half maximal inhibitory constants (IC50) and the dissociation constant for ANS in the absence of PFAS ($K_{d,ANS}$).

$$K_i = \frac{IC50}{\frac{[ANS]_{total}}{K_{d,ANS}} + 1} \quad (3)$$

2.3.5 Circular dichroism

Human LFABP variant C69S/F50C unlabeled and labelled with acrylodan were diluted to 10 μ M in 50mM sodium phosphate buffer (pH 7.5). Circular Dichroism spectra were obtained in a quartz cuvette (0.1cm) at 20C using a Jasco J-1500 CD spectrometer from 250 to 190 nm. Reported Spectra are averages of 4 scans and are expressed in terms of mean residue ellipticity (MRE) (deg cm² dmol⁻¹). MRE values are based on a mean residue weight of 112 and 113 for unlabeled and acrylodan labelled C69S/F50C hLFABP respectively. The content of structural elements were estimated using the web based server BeStSel.³¹

2.3.6 Equilibrium fluorescence titration assays

All fluorescence measurements for wild type hLFABP and variants were performed using a Synergy Neo2 Hybrid Multi-Mode Microplate Reader (Biotek) at room temperature and under steady state conditions. Initial PFOA binding assays were performed by titrating up to 1000 μ M PFOA into wild type hLFABP and tryptophan containing variants (L28W, F50W, F18W, F63W) in 50 mM Tris-Cl buffer (pH 8). After a 5-minute incubation, fluorescence spectra were recorded for a wavelength range of 300-400 nm after excitation at 280 nm.

For the acrylodan labelled C69S/F50C hLFABP variant, PFAS binding assays were performed by titrating PFAS into 1 μM protein in either PBS buffer (pH 7.5) or creek water. To determine sensor ability in systems containing other anionic surfactants, the assay was also performed in PBS buffer (pH 7.5) with sodium dodecyl sulfate (SDS) at a final concentration of 1 μM. Samples were allowed to equilibrate for 5 minutes before fluorescence spectra were recorded over 420-600 nm after excitation at 395 nm.

To quantify averaged shifts in fluorescence spectra after ligand binding, the center of mass was calculated for each curve using the following equation: For tryptophan fluorescence, the peak center of mass was calculated over a wavelength (I) range of 300-400 nm while a range of 420-600 nm was used for acrylodan.

$$x_{cm} = \frac{\sum_{i=\lambda_1}^{\lambda_2} I_i x_i}{\sum_{i=\lambda_1}^{\lambda_2} I_i} \quad (4)$$

2.3.7 Limit of Detection Calculations

The limits of detection (LOD) were calculated for each non-linear system similarly to existing literature³². The base equation (Eq. 5) determines LOD, the lowest analyte concentration that can be distinguished from blanks, by utilizing the limit of blank (LOB) which is the highest apparent analyte concentration to be found in blank replicates. The LOB is first determined using Eq. 6 in which the 1.645 constant is multiplied to the standard deviations such that no more than 5% of values fall out of range.³²

$$LOD = LOB + 1.645(SD_{low\ concentration\ sample}) \quad (5)$$

$$LOB = mean_{blank} + 1.645(SD_{blank}) \quad (6)$$

2.4 Results and Discussion

Tryptophan fluorescence is extremely sensitive to environmental polarity and is often used as an intrinsic probe for the study of protein conformational changes and ligand binding.^{33,34} Human LFABP contains no natural tryptophan residues. Prior work has shown that replacing a leucine residue in the outer binding pocket with tryptophan (L28W) creates a mutant sensitive to natural fatty acid ligand binding.³⁵ Therefore, single tryptophan mutations within the inner and outer ligand binding pockets were introduced as probes for PFAS binding, and residues chosen for mutation are shown in **Figure 2.1**.³⁶

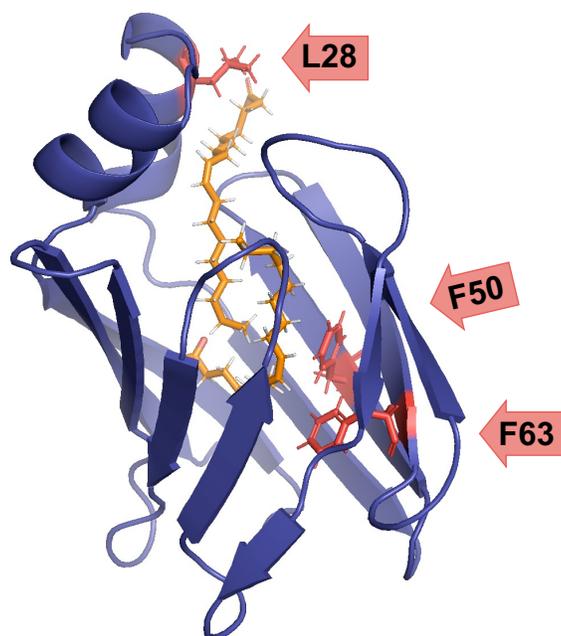


Figure 2.1: hLFABP in complex with two oleic acid molecules. The residues chosen for mutation (L28, F50, F63) are highlighted in red. Figure created using PyMOL (PDB: 2LKK)

Several tryptophan containing hLFABP variants were created, including L28W as well as conservative phenylalanine to tryptophan mutations along the inner binding pocket (F50W and F63W). These mutations were chosen based on proximity to natural ligands as shown in **Figure 2.1**. While neither L28W and F63W showed changes in emission nor fluorescence intensity upon PFOA binding, one mutant (F50W) exhibited blue-shifted emission spectra upon titration of PFOA as shown in **Figure 2.2A**. This suggests that upon PFOA binding, aqueous solvent is displaced within the inner binding pocket and the polarity of the microenvironment surrounding F50W is decreased. This reduction in polarity decreases solvent relaxation of the fluorophore after excitation, leading to lower energy emission and a blue-shifted spectrum.^{37,38} Published molecular dynamics studies of PFOA binding to wild type hLFABP indicates that a single PFOA ligand docks in a “head-out” mode, allowing the fluorinated tail to be stabilized by hydrophobic residues within the binding pocket, including F50.^{24,27,39} Thus, at equilibrium PFOA may not sufficiently interact with the 28th residue located in the more solvated outer binding region nor the 63rd residue, located slightly deeper within the inner binding pocket than the 8-carbon ligand is able to interact with. This is consistent with the lack of tryptophan mutant sensitivity to PFOA binding for the L28W and F63W mutants. As for the binding sensitive F50W hLFABP mutant, the shifted spectra data was fit to a dose response model (**Figure 2.2A**), and a limit of detection (LOD) of PFOA for the tryptophan was calculated to be 6.7 μ M (2.8 ppm).

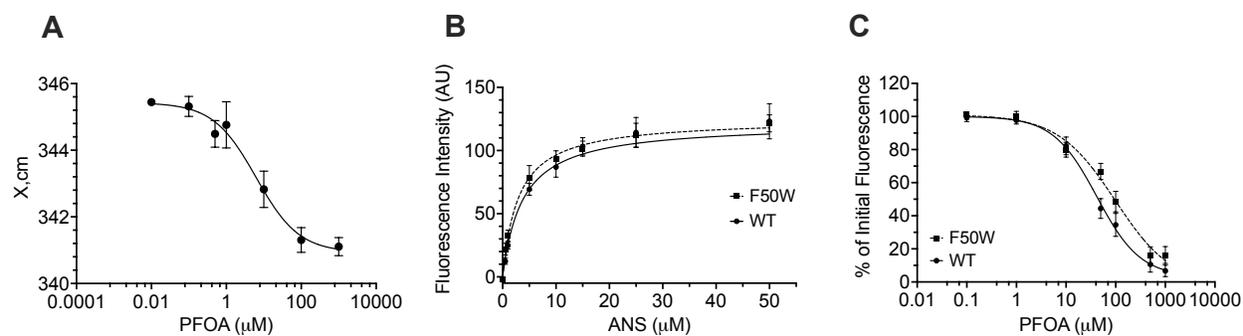


Figure 2.2: (A) Weighted average center of mass of fluorescence spectra for F50W hLFABP (10 μ M) after titration with PFOA. (B) Binding of 1,8-ANS to wild type (WT) and F50W hLFABP (1 μ M) characterized using relative fluorescence intensity as a function of fluorophore (1,8-ANS) concentration. The data was fit to a one-site binding model using Prism. The represented points are mean values \pm SE with n=3. (C) Fluorescence decrease, quantified as the % of initial fluorescence for PFOA. Data were fit to a dose-response model using Prism. The represented points are mean values \pm SE with n=3.

Affinity for PFAS ligands to WT and mutant hLFABP was measured by ANS displacement. First, binding affinity of the fluorescent dye ANS to wild type (WT) and mutant hLFABP was assessed by measuring equilibrium fluorescence intensities for 1 μ M protein samples titrated with increasing concentrations of ANS. Data were fit using nonlinear regression to determine the dissociation constants ($K_{d, \text{ANS}}$). **Figure 2.2B** presents fitted binding curves for WT and F50W, demonstrating no change in ANS affinity. Second, displacement of bound ANS was used to assess the binding of PFAS to hLFABP variants. Protein was first equilibrated with excess ANS, and the fluorescence spectra after PFAS titration was measured. As unlabeled ligand (PFAS) binds to the protein, ANS is displaced from hLFABP, and fluorescence is quenched. Thus, the displacement of the ANS fluorophore was characterized as the percent loss in fluorescence. These values were then plotted as a function of PFAS concentration and fit to a sigmoidal dose-response model to obtain IC50 values. **Figure 2.2C** demonstrates PFOA binding to WT and F50W hLFABP. The IC50 values obtained in this way were then used to calculate an inhibition constant for PFAS ($K_{i, \text{PFOA}}$). All calculated inhibition constants, shown in **Table 2.1**, are in the micromolar range, and it can be

inferred that the addition of all mutations does not negatively affect the binding of PFAS molecules.

Table 2.1: Binding parameters of PFOA for WT and F50W hLFABP as well as PFOA, PFOS, and PFHxS for C69S, C69S/F50W, and C69S/F50C hLFABP. These values were determined using displacement of the fluorophore 1,8-ANS. Error is reported as propagated SEM.

	1,8 ANS	PFOA		PFOS		PFHxS	
	K_d (μ M)	$\log(IC50)$	K_i (μ M)	$\log(IC50)$	K_i (μ M)	$\log(IC50)$	K_i (μ M)
WT	3.76 ± 0.61	1.62 ± 0.09	2.92 ± 0.60	--	--	--	--
F50W	2.97 ± 0.55	2.00 ± 0.26	5.66 ± 3.40	--	--	--	--
C69S	8.78 ± 3.28	1.60 ± 0.05	5.88 ± 0.75	1.00 ± 0.03	1.49 ± 0.12	1.59 ± 0.05	5.77 ± 0.68
C69S/F50W	4.89 ± 1.70	2.18 ± 0.15	13.34 ± 4.68	1.60 ± 0.03	3.51 ± 0.25	1.89 ± 0.21	6.96 ± 3.37
C69S/F50C	5.86 ± 1.80	1.95 ± 0.15	9.29 ± 3.14	1.40 ± 0.04	2.65 ± 0.26	1.97 ± 0.27	9.90 ± 6.16

To increase the sensitivity of the detection system, tryptophan was replaced with another solvatochromic fluorophore, 6-Acryloyl-2-Dimethylaminonaphthalene [Acrylodan]. Acrylodan is a prodan derivative that has been used to study protein conformational changes and ligand binding due to its high sensitivity to polarity.⁴⁰⁻⁴³ Conjugation of this probe to intestinal fatty acid binding protein has also been used to detect free fatty acid levels as well as membrane partitioning of fatty acids.^{44,45}

The native cysteine at position 69 was first replaced with serine to ensure site specific thiol-fluorophore conjugation. This variant (C69S/F50W) resulted in inhibited dimerization as shown in **Figure 2.3A**. For labeling, the inner binding pocket-sensitive position for ligand binding was replaced with cysteine (F50C), resulting in a single cysteine containing hLFABP variant (C69S/F50C). Proper refolding of C69S/F50C hLFABP after urea denaturation during acrylodan labelling was confirmed by comparing CD spectra before and after fluorophore conjugation

(Figure 2.3B). The spectra for unlabeled and labeled proteins are essentially identical, with both samples containing 9.3% α -helix and 30.6% β -sheet as determined using the web-based server, BeStSel. BeStSel allows for characterization of eight different elements of protein structure to provide a more detailed description of protein secondary structure: regular α -helix, distorted α -helix, left twisted β -strand, relaxed β -strand, right-twisted β -strand, parallel β -strand, turn and other³¹. Unlabeled and labeled C69S/F50C hLFABP showed identical compositions for all eight with 3%, 6.3%, 6.0%, 14.2%, 10.4%, 0.5%, 13.9%, and 45.6% respectively, further confirming no change in protein structure occurs upon labeling.

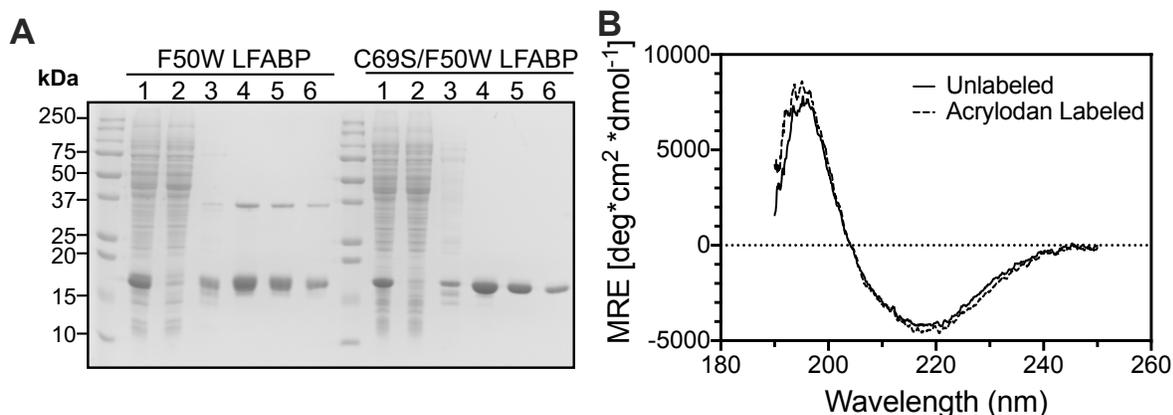


Figure 2.3: (A) SDS-PAGE gel of His-trap purified [Lanes] 1: whole cell lysate, 2: column flowthrough, 3-6: purified protein:: F50W hLFABP containing both monomer (~19kDa) and formed dimer (~38kDa) bands. F50W/C69S hLFABP containing only monomer band (~19 kDa) (B) CD Spectra for unlabeled (—) and acrylodan labeled (---) C69S/F50C hLFABP.

After acrylodan conjugation, the protein was titrated with three PFAS molecules (PFOA, PFOS, and PFHxS) and equilibrium fluorescence spectra were measured in PBS buffer (pH 7.5) as well as in water samples taken from the Dell and Meadow Creek on the grounds of University of Virginia. For all three ligands, an increase in fluorescence as well as a blue shift in spectra were observed in both media systems suggesting this hLFABP variant is able to detect PFAS binding.

Fitted data of the calculated spectra center of mass are shown in **Figure 2.4A-C**, with a larger dose-response shift in fluorescence upon titration of PFAS ligand to the tryptophan-containing protein (F50W; **Figure 2.2A**). Binding of 10 μM of the longer-chain PFAS molecules (PFOA and PFOS) to the acrylodan-based detection system resulted in a shifted center of mass of 10 nm in both PBS and surface water. This is more than a 4-fold increase in signal as compared to PFOA binding to the tryptophan-based system, indicating improved sensitivity of detection using acrylodan. The estimated detection limit of PFOA in PBS was significantly improved with the substitution of acrylodan for tryptophan, with an LOD of 112 ppb and 10-fold reduction in protein needed for detection as compared to the F50W mutant. The calculated LOD for PFOS and PFHxS are 345 ppb, and 1.09 ppm respectively. Thus, use of acrylodan enhances LOD by an order of magnitude as compared to tryptophan and enables detection directly from solution with a LOD within the reported range for other fluorescence and optical-based detection methods for PFAS.

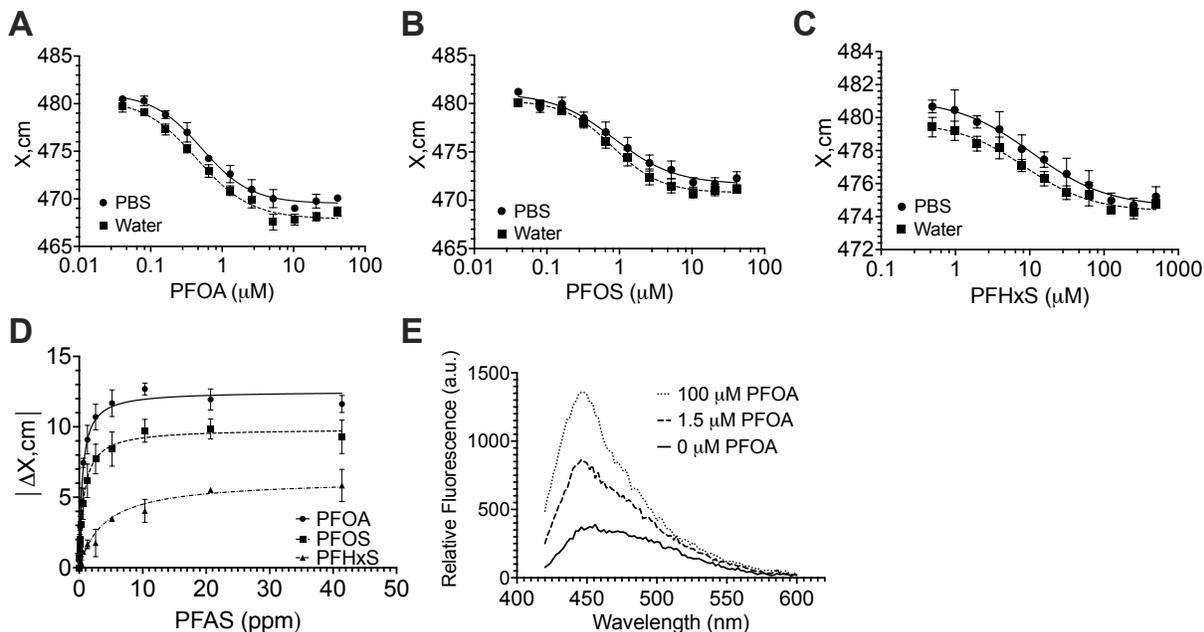


Figure 2.4: (A-C) Weighted average center of mass of fluorescence spectra for acrylodan conjugated C69S/F50C hLFABP (1 μM) after titration with PFOA (A), PFOS (B), and PFHxS (C) in PBS as well as in creek water. Blue shift in spectra upon increased PFAS concentration corresponds to binding and is visualized as a decrease in spectra center of mass (X_{cm}). \therefore (D) Binding of PFAS to acrylodan conjugated C69S/F50C hLFABP (1 μM) fitted to a one site binding model with the fractional occupancy represented as the magnitude of change in center of mass ($|\Delta X_{cm}|$). (E) Fluorescence spectra of acrylodan conjugated C69S/F50C hLFABP (1 μM) with (--) and without (-) PFOA in PBS.

Binding affinity for the acrylodan conjugated protein was assessed by using non-linear regression to fit the data to both a one site and two site binding model. To correct for the negative response upon binding that is associated with a decrease in spectra center of mass, the binding occupancy was represented as the magnitude of the shift ($|\Delta X_{cm}|$). The binding data conformed well only to the one site model (Figure 2.4D) which is consistent with previous studies suggesting PFOA is only capable of interacting with one binding site on the wild type hLFABP.²⁷ Dissociation constants for PFOA, PFOS, and PFHxS were obtained from the one site model and are $1.18 \pm 0.15 \mu\text{M}$, $1.73 \pm 0.37 \mu\text{M}$, and $9.48 \pm 2.20 \mu\text{M}$ respectively. Binding affinity of PFHxS to the acrylodan labelled hLFABP variant is slightly weaker than the longer chain PFAS molecules PFOA and

PFOS. This is consistent with binding affinities for hydrocarbon-chain containing fatty acids ligands, where binding affinity has been shown to increase with carbon chain length for C4-C11 molecules as their hydrophobic tails are able to fully extend in hLFABP's inner binding pocket.³⁹ Thus, it is possible that the shorter-chain PFHxS, when bound to the inner binding pocket, provides less available contact area for specific interactions and stabilization. This is also consistent with the higher LODs the acrylodan labelled sensor has for PFHxS. It is speculated that the shorter chain PFAS interacts less with the 50th residue resulting in decreased change in polarity as compared to the 8 carbon PFOA and PFOS.

The specificity of the acrylodan-conjugated hLFABP variant was further tested by introducing the anionic surfactant, sodium dodecyl sulfate (SDS), as a competitor for binding. **Figure 2.5** shows the change in spectral center of mass for the protein sensor (1 μ M) in PBS containing SDS (1 μ M) upon titration of both PFOA and PFOS. The LODs for PFOA and PFOS were calculated to be 112 ppb and 165 ppb respectively. While the magnitude of the fluorescence shift is decreased compared to system without SDS, the LOD values are comparable to those containing no additional surfactant. Thus, the engineered hLFABP sensor is capable of detecting both PFOA and PFOS at ppb levels in complex media containing anionic surfactants such as SDS that can act as inhibitors of binding.

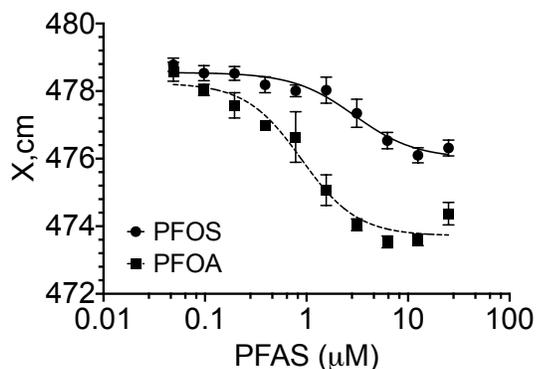


Figure 2.5: Weighted average center of mass (X, cm) of fluorescence spectra for acrylodan conjugated C69S/F50C hLFABP ($1 \mu\text{M}$) after titration with PFOA and PFOS in PBS containing SDS ($1 \mu\text{M}$).

Our approach is a useful initial step in development of protein-based biosensors to detect PFAS. Unlike PPAR α and other mammalian receptors that bind PFAS, FABP is a robust scaffold that can be readily expressed and purified recombinantly in high yield versus extracted from mammalian tissues. Furthermore, direct detection from environmental samples is possible and detection is not impaired by the presence of interfering ligands such as SDS, unlike other approaches that require pre-treatment to enrich PFAS or remove interfering substances. We also demonstrate the ability to modify the protein to amplify signal and thus LOD, which we anticipate will enable further engineering approaches to improve affinity and LOD, as well as a complement to other ELISA- and QD-based biochemical PFAS detection strategies. Given that protein-based sensors have been developed as green, effective alternatives for detection of many environmental contaminants including heavy metals, pharmaceuticals, and pesticides, LFABP could eventually serve as a useful scaffold for improved direct detection of PFAS in important solid and liquid matrices.^{46,47}

2.5 References

- (1) Wang, Z.; DeWitt, J. C.; Higgins, C. P.; Cousins, I. T. A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environ. Sci. Technol.* **2017**, *51* (5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>.
- (2) Favreau, P.; Poncioni-Rothlisberger, C.; Place, B. J.; Bouchex-Bellomie, H.; Weber, A.; Tremp, J.; Field, J. A.; Kohler, M. Multianalyte Profiling of Per- and Polyfluoroalkyl Substances (PFASs) in Liquid Commercial Products. *Chemosphere* **2017**, *171*, 491–501. <https://doi.org/10.1016/j.chemosphere.2016.11.127>.
- (3) Robel, A. E.; Marshall, K.; Dickinson, M.; Lunderberg, D.; Butt, C.; Peaslee, G.; Stapleton, H. M.; Field, J. A. Closing the Mass Balance on Fluorine on Papers and Textiles. *Environ. Sci. Technol.* **2017**, *51* (16), 9022–9032. <https://doi.org/10.1021/acs.est.7b02080>.
- (4) Schaidler, L. A.; Balan, S. A.; Blum, A.; Andrews, D. Q.; Strynar, M. J.; Dickinson, M. E.; Lunderberg, D. M.; Lang, J. R.; Peaslee, G. F. Fluorinated Compounds in U.S. Fast Food Packaging. *Environ. Sci. Technol. Lett.* **2017**, *4* (3), 105–111. <https://doi.org/10.1021/acs.estlett.6b00435>.
- (5) Kotthoff, M.; Müller, J.; Jürling, H.; Schlummer, M.; Fiedler, D. Perfluoroalkyl and Polyfluoroalkyl Substances in Consumer Products. *Environ Sci Pollut Res Int* **2015**, *22* (19), 14546–14559. <https://doi.org/10.1007/s11356-015-4202-7>.
- (6) Dauchy, X.; Boiteux, V.; Bach, C.; Rosin, C.; Munoz, J.-F. Per- and Polyfluoroalkyl Substances in Firefighting Foam Concentrates and Water Samples Collected near Sites Impacted by the Use of These Foams. *Chemosphere* **2017**, *183*, 53–61. <https://doi.org/10.1016/j.chemosphere.2017.05.056>.
- (7) Fenton, S. E.; Ducatman, A.; Boobis, A.; DeWitt, J. C.; Lau, C.; Ng, C.; Smith, J. S.; Roberts, S. M. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environmental Toxicology and Chemistry* **2021**, *40* (3), 606–630. <https://doi.org/10.1002/etc.4890>.
- (8) Steenland, K.; Tinker, S.; Frisbee, S.; Ducatman, A.; Vaccarino, V. Association of Perfluorooctanoic Acid and Perfluorooctane Sulfonate With Serum Lipids Among Adults Living Near a Chemical Plant. *American Journal of Epidemiology* **2009**, *170* (10), 1268–1278. <https://doi.org/10.1093/aje/kwp279>.

- (9) Steenland, K.; Fletcher, T.; Savitz, D. A. Epidemiologic Evidence on the Health Effects of Perfluorooctanoic Acid (PFOA). *Environ Health Perspect* **2010**, *118* (8), 1100–1108. <https://doi.org/10.1289/ehp.0901827>.
- (10) Armstrong, L. E.; Guo, G. L. Understanding Environmental Contaminants' Direct Effects on Non-Alcoholic Fatty Liver Disease Progression. *Curr Envir Health Rpt* **2019**, *6* (3), 95–104. <https://doi.org/10.1007/s40572-019-00231-x>.
- (11) Stanifer, J. W.; Stapleton, H. M.; Souma, T.; Wittmer, A.; Zhao, X.; Boulware, L. E. Perfluorinated Chemicals as Emerging Environmental Threats to Kidney Health: A Scoping Review. *CJASN* **2018**, *13* (10), 1479–1492. <https://doi.org/10.2215/CJN.04670418>.
- (12) Looker, C.; Luster, M. I.; Calafat, A. M.; Johnson, V. J.; Burleson, G. R.; Burleson, F. G.; Fletcher, T. Influenza Vaccine Response in Adults Exposed to Perfluorooctanoate and Perfluorooctanesulfonate. *Toxicol Sci* **2014**, *138* (1), 76–88. <https://doi.org/10.1093/toxsci/kft269>.
- (13) Abraham, K.; Mielke, H.; Fromme, H.; Völkel, W.; Menzel, J.; Peiser, M.; Zepp, F.; Willich, S. N.; Weikert, C. Internal Exposure to Perfluoroalkyl Substances (PFASs) and Biological Markers in 101 Healthy 1-Year-Old Children: Associations between Levels of Perfluorooctanoic Acid (PFOA) and Vaccine Response. *Arch Toxicol* **2020**, *94* (6), 2131–2147. <https://doi.org/10.1007/s00204-020-02715-4>.
- (14) Bulka, C. M.; Avula, V.; Fry, R. C. Associations of Exposure to Perfluoroalkyl Substances Individually and in Mixtures with Persistent Infections: Recent Findings from NHANES 1999–2016. *Environmental Pollution* **2021**, *275*, 116619. <https://doi.org/10.1016/j.envpol.2021.116619>.
- (15) Mamsen, L. S.; Björvang, R. D.; Mucs, D.; Vinnars, M.-T.; Papadogiannakis, N.; Lindh, C. H.; Andersen, C. Y.; Damdimopoulou, P. Concentrations of Perfluoroalkyl Substances (PFASs) in Human Embryonic and Fetal Organs from First, Second, and Third Trimester Pregnancies. *Environment International* **2019**, *124*, 482–492. <https://doi.org/10.1016/j.envint.2019.01.010>.
- (16) Washino, N.; Saijo, Y.; Sasaki, S.; Kato, S.; Ban, S.; Konishi, K.; Ito, R.; Nakata, A.; Iwasaki, Y.; Saito, K.; Nakazawa, H.; Kishi, R. Correlations between Prenatal Exposure to

- Perfluorinated Chemicals and Reduced Fetal Growth. *Environ Health Perspect* **2009**, *117* (4), 660–667. <https://doi.org/10.1289/ehp.11681>.
- (17) Hoffman, K.; Webster, T. F.; Weisskopf, M. G.; Weinberg, J.; Vieira, V. M. Exposure to Polyfluoroalkyl Chemicals and Attention Deficit/Hyperactivity Disorder in U.S. Children 12–15 Years of Age. *Environ Health Perspect* **2010**, *118* (12), 1762–1767. <https://doi.org/10.1289/ehp.1001898>.
- (18) Vuong, A. M.; Yolton, K.; Webster, G. M.; Sjödin, A.; Calafat, A. M.; Braun, J. M.; Dietrich, K. N.; Lanphear, B. P.; Chen, A. Prenatal Polybrominated Diphenyl Ether and Perfluoroalkyl Substance Exposures and Executive Function in School-Age Children. *Environ Res* **2016**, *147*, 556–564. <https://doi.org/10.1016/j.envres.2016.01.008>.
- (19) Mora, A. M.; Oken, E.; Rifas-Shiman, S. L.; Webster, T. F.; Gillman, M. W.; Calafat, A. M.; Ye, X.; Sagiv, S. K. Prenatal Exposure to Perfluoroalkyl Substances and Adiposity in Early and Mid-Childhood. *Environ Health Perspect* **2017**, *125* (3), 467–473. <https://doi.org/10.1289/EHP246>.
- (20) Mora, A. M.; Fleisch, A. F.; Rifas-Shiman, S. L.; Baidal, J. W.; Pardo, L.; Webster, T. F.; Calafat, A. M.; Ye, X.; Oken, E.; Sagiv, S. K. Early Life Exposure to Per- and Polyfluoroalkyl Substances and Mid-Childhood Lipid and Alanine Aminotransferase Levels. *Environ Int* **2018**, *111*, 1–13. <https://doi.org/10.1016/j.envint.2017.11.008>.
- (21) Shoemaker, J., D. T. Method 537.1 Determination of Selected Per- and Polyfluorinated Alkyl Substances in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS), 2020.
- (22) Hu, X.; Liu, Y.; Yang, R.; Jiang, Y.; Meng, M.; Liu, Z.; Ni, L.; Wu, W.; Liu, H. Adsorption for Perfluorooctanoic Acid with Graphitic-Phase Carbon Nitride and Its HPLC Fluorescence Determination. *The Canadian Journal of Chemical Engineering* **2020**, *98* (1), 394–403. <https://doi.org/10.1002/cjce.23625>.
- (23) Atshaves, B. P.; Martin, G. G.; Hostetler, H. A.; McIntosh, A. L.; Kier, A. B.; Schroeder, F. Liver Fatty Acid Binding Protein and Obesity. *J Nutr Biochem* **2010**, *21* (11), 1015–1032. <https://doi.org/10.1016/j.jnutbio.2010.01.005>.
- (24) Sheng, N.; Cui, R.; Wang, J.; Guo, Y.; Wang, J.; Dai, J. Cytotoxicity of Novel Fluorinated Alternatives to Long-Chain Perfluoroalkyl Substances to Human Liver Cell Line and Their

- Binding Capacity to Human Liver Fatty Acid Binding Protein. *Arch Toxicol* **2018**, *92* (1), 359–369. <https://doi.org/10.1007/s00204-017-2055-1>.
- (25) Thompson, J.; Winter, N.; Terwey, D.; Bratt, J.; Banaszak, L. The Crystal Structure of the Liver Fatty Acid-Binding Protein: A COMPLEX WITH TWO BOUND OLEATES. *J. Biol. Chem.* **1997**, *272* (11), 7140–7150. <https://doi.org/10.1074/jbc.272.11.7140>.
- (26) Sharma, A.; Sharma, A. Fatty Acid Induced Remodeling within the Human Liver Fatty Acid-Binding Protein. *J. Biol. Chem.* **2011**, *286* (36), 31924–31928. <https://doi.org/10.1074/jbc.M111.270165>.
- (27) Sheng, N.; Li, J.; Liu, H.; Zhang, A.; Dai, J. Interaction of Perfluoroalkyl Acids with Human Liver Fatty Acid-Binding Protein. *Arch Toxicol* **2016**, *90* (1), 217–227. <https://doi.org/10.1007/s00204-014-1391-7>.
- (28) Cheng, W.; Ng, C. A. Predicting Relative Protein Affinity of Novel Per- and Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach. *Environ. Sci. Technol.* **2018**, *52* (14), 7972–7980. <https://doi.org/10.1021/acs.est.8b01268>.
- (29) Prendergast, F. G.; Meyer, M.; Carlson, G. L.; Iida, S.; Potter, J. D. Synthesis, Spectral Properties, and Use of 6-Acryloyl-2-Dimethylaminonaphthalene (Acrylodan). A Thiol-Selective, Polarity-Sensitive Fluorescent Probe. *J Biol Chem* **1983**, *258* (12), 7541–7544.
- (30) Velkov, T.; Chuang, S.; Wielens, J.; Sakellaris, H.; Charman, W. N.; Porter, C. J. H.; Scanlon, M. J. The Interaction of Lipophilic Drugs with Intestinal Fatty Acid-Binding Protein. *J. Biol. Chem.* **2005**, *280* (18), 17769–17776. <https://doi.org/10.1074/jbc.M410193200>.
- (31) Micsonai, A.; Wien, F.; Bulyáki, É.; Kun, J.; Moussong, É.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. BeStSel: A Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra. *Nucleic Acids Res* **2018**, *46* (Web Server issue), W315–W322. <https://doi.org/10.1093/nar/gky497>.
- (32) Armbruster, D. A.; Pry, T. Limit of Blank, Limit of Detection and Limit of Quantitation. *Clin Biochem Rev* 2008, *29* Suppl 1 (Suppl 1), S49-52.
- (33) Markiewicz, B. N.; Mukherjee, D.; Troxler, T.; Gai, F. Utility of 5-Cyanotryptophan Fluorescence as a Sensitive Probe of Protein Hydration. *J Phys Chem B* **2016**, *120* (5), 936–944. <https://doi.org/10.1021/acs.jpcc.5b12233>.

- (34) Royer, C. A. Probing Protein Folding and Conformational Transitions with Fluorescence. *Chem. Rev.* **2006**, *106* (5), 1769–1784. <https://doi.org/10.1021/cr0404390>.
- (35) Hagan, R. M.; Worner-Gibbs, J.; Wilton, D. C. Tryptophan Insertion Mutagenesis of Liver Fatty Acid-Binding Protein: L28W MUTANT PROVIDES IMPORTANT INSIGHTS INTO LIGAND BINDING AND PHYSIOLOGICAL FUNCTION. *J. Biol. Chem.* **2005**, *280* (3), 1782–1789. <https://doi.org/10.1074/jbc.M407131200>.
- (36) Cai, J.; Lücke, C.; Chen, Z.; Qiao, Y.; Klimtchuk, E.; Hamilton, J. A. Solution Structure and Backbone Dynamics of Human Liver Fatty Acid Binding Protein: Fatty Acid Binding Revisited. *Biophysical Journal* **2012**, *102* (11), 2585–2594. <https://doi.org/10.1016/j.bpj.2012.04.039>.
- (37) Lakowicz, J. R.; Masters, B. R. Principles of Fluorescence Spectroscopy, Third Edition. *J. Biomed. Opt.* **2008**, *13* (2), 029901. <https://doi.org/10.1117/1.2904580>.
- (38) Draganski, A. R.; Friedman, J. M.; Ludescher, R. D. Solvent-Slaved Dynamic Processes Observed by Tryptophan Phosphorescence of Human Serum Albumin. *Biophysical Journal* **2017**, *112* (5), 881–891. <https://doi.org/10.1016/j.bpj.2016.12.048>.
- (39) Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol.* **2013**, *47* (19), 11293–11301. <https://doi.org/10.1021/es4026722>.
- (40) Gilardi, Gianfranco.; Zhou, L. Qing.; Hibbert, Linda.; Cass, A. E. G. Engineering the Maltose Binding Protein for Reagentless Fluorescence Sensing. *Anal. Chem.* **1994**, *66* (21), 3840–3847. <https://doi.org/10.1021/ac00093a047>.
- (41) Hibbs, R. E.; Talley, T. T.; Taylor, P. Acrylodan-Conjugated Cysteine Side Chains Reveal Conformational State and Ligand Site Locations of the Acetylcholine-Binding Protein*. *Journal of Biological Chemistry* **2004**, *279* (27), 28483–28491. <https://doi.org/10.1074/jbc.M403713200>.
- (42) Park, H. Y.; Kim, S. A.; Korlach, J.; Rhoades, E.; Kwok, L. W.; Zipfel, W. R.; Waxham, M. N.; Webb, W. W.; Pollack, L. Conformational Changes of Calmodulin upon Ca²⁺ Binding Studied with a Microfluidic Mixer. *Proceedings of the National Academy of Sciences* **2008**, *105* (2), 542–547. <https://doi.org/10.1073/pnas.0710810105>.
- (43) Hickey, J. M.; Sahni, N.; Chaudhuri, R.; D'Souza, A.; Metters, A.; Joshi, S. B.; Russell Middaugh, C.; Volkin, D. B. Effect of Acrylodan Conjugation and Forced Oxidation on the

- Structural Integrity, Conformational Stability, and Binding Activity of a Glucose Binding Protein SM4 Used in a Prototype Continuous Glucose Monitor. *Protein Sci* **2017**, *26* (3), 527–535. <https://doi.org/10.1002/pro.3102>.
- (44) Richieri, G. V.; Ogata, R. T.; Kleinfeld, A. M. A Fluorescently Labeled Intestinal Fatty Acid Binding Protein. Interactions with Fatty Acids and Its Use in Monitoring Free Fatty Acids. *J Biol Chem* **1992**, *267* (33), 23495–23501.
- (45) Simard, J. R.; Kamp, F.; Hamilton, J. A. Acrylodan-Labeled Intestinal Fatty Acid-Binding Protein to Measure Concentrations of Unbound Fatty Acids. *Methods in Membrane Lipids* **2007**, 27–43. https://doi.org/10.1007/978-1-59745-519-0_3.
- (46) Rodriguez-Mozaz, S.; Reder, S.; Lopez de Alda, M.; Gauglitz, G.; Barceló, D. Simultaneous Multi-Analyte Determination of Estrone, Isoproturon and Atrazine in Natural Waters by the RIver ANALyser (RIANA), an Optical Immunosensor. *Biosensors and Bioelectronics* **2004**, *19* (7), 633–640. [https://doi.org/10.1016/S0956-5663\(03\)00255-0](https://doi.org/10.1016/S0956-5663(03)00255-0).
- (47) Thavarajah, W.; Verosloff, M. S.; Jung, J. K.; Alam, K. K.; Miller, J. D.; Jewett, M. C.; Young, S. L.; Lucks, J. B. A Primer on Emerging Field-Deployable Synthetic Biology Tools for Global Water Quality Monitoring. *npj Clean Water* **2020**, *3* (1), 1–10. <https://doi.org/10.1038/s41545-020-0064-8>.

Chapter 3

Field performance of acrylodan labeled human liver fatty acid binding protein biosensor

3.1 Abstract

Per- and polyfluoroalkyl substances (PFAS) are a large set of emerging contaminants pervasive in the environment due to amphiphilic properties and strong carbon-fluorine bonds resistant to biodegradation. With an ever-increasing prevalence, the need for precise detection of these chemicals at low levels in drinking water is clear. However, ground and surface water as well as soil and other biosolids have become reservoirs for PFAS at extremely high levels. In fact, PFAS concentrations at part per billion and part per million levels are found in environmental samples taken near high contamination sites including industrial facilities and military bases. In this work, we demonstrate the application of a biosensor based on human liver fatty acid binding protein to detect perfluorooctanoic acid (PFOA) in surface water samples taken near Loring Airforce Base. We show this sensor can detect the high levels of PFOA found in the samples quickly and easily without the use of extensive sample pre-treatment or analytical methods. Therefore, we hope the future of this technology will better assess PFAS detection needs for a multitude of end point users.

3.2 Introduction

PFAS (per-/polyfluoroalkyl substances) are a group of environmental and toxicological contaminants of increasing concern. Since their development in the 1940s, over 3000 variants of PFAS have been manufactured and have entered the global market for use in industrial and consumer applications.^{1,2} While their fluorinated carbon chains and polar head moieties impart oil and water repellency ideal for industrial surfactants as well as consumer goods³⁻⁵, the unparalleled chemical stability of environmental end products like perfluoroalkyl acids (PFAAs) allow for extreme environmental accumulation.^{4,6} Accumulation of these shorter chain PFAS like perfluorooctanoic acid (PFOA) and perfluorooctane sulfonic acid (PFOS) has become increasingly prevalent in humans^{5,7,8} and has been linked to a variety of health effects including endocrine dysfunction, thyroid, pancreatic, and liver diseases⁸⁻¹¹ as well as a variety of reproductive issues and cancers^{8,12-15}.

In recent years many new regulations have been developed and proposed to ensure strict limits on PFAS in drinking water.¹⁶⁻¹⁸ Therefore, detection of these contaminants at the extremely low (part per trillion) levels relevant for drinking water has become a growing priority for researchers. However, it is important to note that human exposure to these chemicals occurs through multiple avenues including food ingestion as well as general contact.^{1,4} Specifically, it is becoming increasingly relevant to be able to quickly and easily detect the presence of PFAS in the soil and groundwater around sites with large amounts of these types of contaminants.¹⁹⁻²¹

In this study we aim to show the feasibility of a protein-based biosensor to detect PFAS in groundwater samples collected in Aroostook County, Maine. Loring Air Force base is located in

the town of Limestone within Aroostook County and is an EPA Superfunds site with PFAS contamination. Adjacent to this site are lands belonging to the Mi'kmaq Nation as well as the Aroostook National Wildlife Refuge²²; prior studies have investigated use of phytoremediation to remove PFAS from soils at Loring AFB, and recent testing has detected PFAS in drinking water in Aroostook County. Our prior work developing and benchmarking an acrylodan [6-acryloyl-2-dimethylaminonaphthalene] labeled biosensor based on human liver fatty acid binding protein (hLFABP) enabled detection of PFOA and PFOS in spiked creek water samples with limits of detection in the hundreds of parts per billion.²³ In this study we demonstrate this sensor can detect PFAS compounds such as PFOA in contaminated water samples collected adjacent to Loring AFB and are in agreement with independent LC/MS-MS testing for PFAS using method EPA 1633.²⁴ These results provide an important initial validation of our biosensor to augment LC/MS-MS testing and demonstrates the use of engineered biosensors for rapid detection of environmental contaminants.

3.3 Materials and Methods

3.3.1 Sample collection

Samples were collected in June onsite in Aroostook County, ME. Unless otherwise stated, water samples collected were shipped to University of Virginia and Cyclopure for analysis.

3.3.2 LC/MS-MS quantification

Surface water samples were first filtered using a 0.2 μ m filter to remove particulates and diluted as needed in sterile, ultrapure water. PFAS levels in filtered, diluted samples were measured using

the Water Test Kit Pro (Cyclopure), which provides measured concentration of 55 PFAS compounds utilizing EPA methods 533, 537 and 1633.

3.3.3 Sensor production

Acrylodan labeled hLFABP (Ac-hLFABP) was produced as previously described.²³ Briefly, the single cysteine containing hLFABP F50C/C69S mutant was expressed via the pET28a(+) vector in *E. coli* BL21 (DE3). The protein was then purified using nickel affinity chromatography, dialyzed against PBS (pH 7.6), and thrombin cleaved using Thrombin CleanCleave Kit (Sigma Aldrich) to remove the N-terminal histidine tag. After cleavage, the protein was conjugated with acrylodan [6-acryloyl-2-dimethylaminonaphthalene] at a 2.5:1 fluorophore to protein ratio in denaturing conditions (8M urea). Removal of unreacted acrylodan and refolding of the denatured protein was performed on-column with a step-down gradient to 0M urea. Properly folded Ac-hLFABP was removed using 500mM imidazole, and once again dialyzed into PBS (pH 7.6). Protein concentration was found using a BCA protein assay (Pierce), and acrylodan concentration was measured by absorption at 370 nm (extinction coefficient $16,400 \text{ M}^{-1}\text{cm}^{-1}$)²⁵. The degree of labeling for acrylodan conjugated C69S/F50C hLFABP was 1.0.

3.3.4 Sample analysis

Fluorescence measurements were performed using a Synergy Neo2 Hybrid Multi-Mode Microplate Reader (Biotek) at room temperature and under steady state conditions. Calibration curves and sample data were collected as previously described.²³ Calibration curves were generated by titrating PFOA in water into Ac-hLFABP in PBS (pH 7.6) to a final micromolar ratio of 100:1 ligand to sensor respectively. Samples and standards were allowed to equilibrate for 5

minutes before fluorescence spectra were recorded over 400-600 nm after excitation at 360 nm. Signal was quantified as the shift in fluorescence spectra calculated as the change in the average center of mass (Δx_{cm}) from 400-460nm (Eq. 1). Calibration curve data was fit to a four-parameter log dose response model (Eq. 2), and sample PFOA concentrations were calculated using the obtained fit parameters: Hillslope coefficient (HS), minimum and maximum signal ($\Delta X_{cm_{min}}$ & $\Delta X_{cm_{max}}$), and half maximal effective concentration (EC50).

$$x_{cm} = \frac{\sum_{i=1}^{12} I_i x_i}{\sum_{i=1}^{12} I_i} \quad (1)$$

$$\Delta X_{cm} = \Delta X_{cm_{min}} + \frac{[PFOA]^{HS} * (\Delta X_{cm_{max}} - \Delta X_{cm_{min}})}{\Delta X_{cm}^{HS} + EC50^{HS}} \quad (2)$$

3.4 Results and Discussion

The developed protein-based biosensor utilizes human liver fatty acid binding protein (hLFABP) as a PFAS-binding scaffold.²³ In the body, hLFABP is upregulated in response to PFOA exposure and has also been shown to bind several PFAS variants with moderate affinity due to their structural similarity to endogenous fatty acids.²⁶⁻²⁸ By introducing a thiol conjugated solvatochromic fluorophore (acrylodan) at the 50th residue position within the ligand binding pocket, the hLFABP mutant F50C/C69S is capable of binding several PFAS compounds including PFOA while producing a dose dependent blue-shift in acrylodan emission spectra. This shift in emission spectra to higher energy wavelengths occurs due to a polarity change in the acrylodan microenvironment upon ligand binding. The spectral shift is then quantified as the change in peak center of mass (ΔX_{cm}).

Aqueous fire-fighting foams (AFFFs), such as those found at Loring AFB, can consist of over 50 different PFAS compounds ranging from C2-C12 in length as well as many fluorotelomers that act as precursors to PFAAs such as PFOA and PFOS.^{6,29,30} Wide use of AFFF chemicals at military bases during training, emergency responses, and equipment maintenance make these sites high risk for PFAS contamination.^{29,30} In fact, AFFF fluorotelomers as well as perfluorocarboxylates and perfluorosulfonates of varying chain sizes have been found at multiple U.S. military sites, sometimes even reaching part per million levels in groundwater,^{19,20,31–33} and are known to have extremely high transport potential.^{29,34,35}

Like many other military sites, Loring AFB was a site of known contamination with many hazardous toxins including PFAS, and in the years following the base's deactivation, it remains a EPA Superfunds cleanup site.^{24,36} Recently, concerns have been raised about persistent contamination based on high levels of PFAS found in drinking water in Aroostook county as well as in agricultural soils on land belonging to the Mi'kmaq; one potential source of this contamination is from overuse of firefighting foams that contain PFAS.³⁷ Over 70 different types of PFAS have been detected in soil samples around the Loring AFB, with sulfonic and carboxylic acids being the primary contaminants at concentrations of up to 150 ppb in soil.³⁶ A recent test of area schools in Aroostook County show unsafe levels of both lead and PFAS in drinking water.^{38,39}

To demonstrate our biosensor's utility in application settings, we measured total signal for PFAS binding in samples collected adjacent to the Loring AFB as well as in major waterways such as the Aroostook River. **Figure 3.1** illustrates the sites where samples were collected.



Figure 3.1: Illustration of collection sites near (A) Chapman pit and (B) Malabeam Lake. The left panel depicts the area around Loring Airforce Base with an approximation of the Aroostook National Wildlife Refuge overlaid in white.

Based on LC/MS-MS analysis, all samples contained primarily PFOA and PFHxS, accounting for 87-98% of the total PFAS levels measured across all samples collected; PFHxS was commonly used in AFFF, and PFOA is one of the most abundant PFAS compounds found in the environment, and often is a breakdown product formed from higher molecular weight PFAS compounds. The results of the LC/MS-MS data are summarized in *Table 3.1*. Our previous work showed the biosensor is most sensitive to PFOA at <1 ppm levels unlike PFHxS at greater than >1 ppm levels. Therefore, we prepared standards containing solely PFOA to calibrate the biosensor, determine apparent signal from PFAS binding in environmental samples, and compare to LC/MS-MS results.

Table 3.1: Results of surface water samples from LC/MS-MS. Verified concentrations are in ng/L with LOQ defined as limit of quantification.

	Chapman Pit (A)	Malabeam Lake (B)
PFBA	<LOQ	<LOQ
PFPeA	112.50	<LOQ
PFHxA	183.75	47.73
PFHpA	1417.50	202.27
PFOA	50477.50	78840.91
PFNA	<LOQ	<LOQ
PFDA	86.25	<LOQ
HFPO-DA (GenX)	<LOQ	<LOQ
PFBS	268.75	831.82
PFHxS	581.25	47920.45
PFOS	881.25	1379.55
ΣPFAS	54008.75	129222.73

The right panel of **Figure 3.2** shows the calibration curves of the assay replicates utilized for sample PFOA quantification. These data were fit to four-parameter log-dose response models where Hill slope coefficient (HS), minimum and maximum signal (ΔXcm_{min} & ΔXcm_{max}), and half maximal effective concentration (EC50) constants were derived. Four separate sets of derived constants were used for quantifying PFOA concentration in the water samples rather than an average amongst the replicates to gauge concentrations more accurately in lower PFOA containing samples and reduce day to day error.

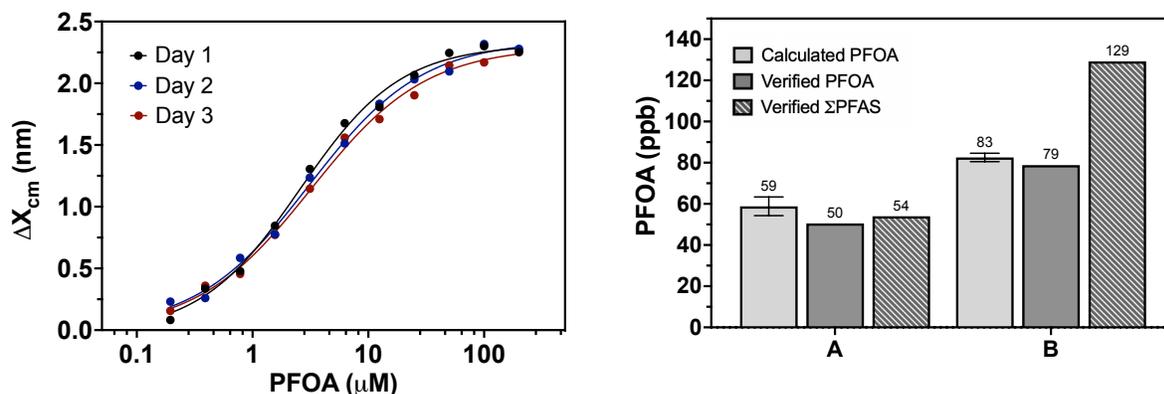


Figure 3.2: (LEFT) Sensor calibration curves for varying PFOA concentrations spiked in water. Data from these spiked standards were fit to a four-parameter log-dose response model, and water sample concentrations were calculated using the parameters obtained from the non-linear regression for each day the analysis was performed. (RIGHT) Comparison of predicted and verified concentrations of PFOA and total PFAS in parts per billion (ppb) for surface water samples. The calculated concentrations are mean values \pm SE with $n=3$.

The results of LC/MS-MS indicated samples A and B contained 50 and 79 ppb PFOA respectively. Based on our calibration curve using PFOA, from our biosensor measurements we obtained values of 59 ± 5 ppb and 83 ± 2 ppb for these samples; this provides further evidence that the biosensor is robust to environmental testing and can be used for direct determination of PFOA. PFOA was the main contaminant found in sample A which was taken from Chapman Pit. Along with PFOA, sample B, taken from Malabeam Lake, also contained 48 ppb PFHxS, a 6-carbon sulfonated perfluoroalkyl acid. The ability of the biosensor to detect PFOA in this sample with relative accuracy but not PFHxS indicates a selectivity for the 8-carbon chain carboxylic acid contaminant. This is consistent with prior work that suggests shorter chain PFAS like PFHxS, when bound to the inner binding pocket of hLFABP, provides less available contact area for specific interactions and stabilization.^{23,40}

It is important to note challenges with sample variability; for surface water, constituents including organic matter and other co-contaminants can interfere with PFAS detection. While our testing on field samples with relatively limited sample pre-processing generated results consistent with independent LC/MS-MS detection, further application of this biosensor requires analysis of how other pollutants or compounds might interfere with PFAS detection. Our prior work demonstrated that representative surfactants such as SDS which bind hLFABP do not interfere with PFAS detection; however, as a Superfund site, the Loring base was identified as having soil and groundwater containing volatile organic compounds (VOCs), polychlorinated biphenyls (PCBs), polyaromatic hydrocarbons (PAHs), pesticides, metals, and other petroleum related compounds.²⁴ Even among AFFFs, formulations include various hydrocarbon surfactants, polymers, and organic solvents amongst the fluorinated chemicals.⁴¹⁻⁴³ Because hLFABP is known to bind aliphatic fatty acids, it is likely that aliphatic, hydrocarbon-based contaminants could also be detected using this biosensor despite low hydrocarbon solubility in water. Our current efforts are focused on determining what hydrocarbons and other pollutants may be present in field samples and how they impact biosensor sensitivity, as well as methods to multiplex co-detection of multiple pollutants.

Overall, this study provides evidence that an engineered, protein-based biosensor can be used to detect PFAS in field samples using fluorescence readout with minimal pre-processing. With recent studies indicating widespread and significant PFAS levels in soils, water, agricultural products, food, consumer products and the human body, a multi-faceted approach to PFAS detection using multiple approaches is needed. As illustrated in this work, biosensors have the advantage of straightforward implementation with reduced sample pre-processing; as such, they can serve as a useful initial screen for contamination to augment more sophisticated, detailed detections methods

such as LC/MS-MS. As more studies show that PFAS levels in surface water and soils can reach levels greater than parts per trillion (ppt), it is also important to develop testing methods able to give reliable results in this concentration range that are cost-effective and readily available. As mentioned above, our current efforts are focused on understanding how other pollutants or environmental compounds impact detection, as well as extending detection to other compounds often found where PFAS contamination is present.

3.5 References

- (1) Wang, Z.; DeWitt, J. C.; Higgins, C. P.; Cousins, I. T. A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environ. Sci. Technol.* **2017**, *51* (5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>.
- (2) Wickham, G. M.; Shriver, T. E. Emerging Contaminants, Coerced Ignorance and Environmental Health Concerns: The Case of Per- and Polyfluoroalkyl Substances (PFAS). *Sociol Health Illn* **2021**, *43* (3), 764–778. <https://doi.org/10.1111/1467-9566.13253>.
- (3) Harding-Marjanovic, K. C.; Houtz, E. F.; Yi, S.; Field, J. A.; Sedlak, D. L.; Alvarez-Cohen, L. Aerobic Biotransformation of Fluorotelomer Thioether Amido Sulfonate (Lodyne) in AFFF-Amended Microcosms. *Environ. Sci. Technol.* **2015**, *49* (13), 7666–7674. <https://doi.org/10.1021/acs.est.5b01219>.
- (4) Brunn, H.; Arnold, G.; Körner, W.; Rippen, G.; Steinhäuser, K. G.; Valentin, I. PFAS: Forever Chemicals—Persistent, Bioaccumulative and Mobile. Reviewing the Status and the Need for Their Phase out and Remediation of Contaminated Sites. *Environ Sci Eur* **2023**, *35* (1), 20. <https://doi.org/10.1186/s12302-023-00721-8>.
- (5) Jian, J.-M.; Chen, D.; Han, F.-J.; Guo, Y.; Zeng, L.; Lu, X.; Wang, F. A Short Review on Human Exposure to and Tissue Distribution of Per- and Polyfluoroalkyl Substances (PFASs). *Science of The Total Environment* **2018**, *636*, 1058–1069. <https://doi.org/10.1016/j.scitotenv.2018.04.380>.
- (6) Houtz, E. F.; Higgins, C. P.; Field, J. A.; Sedlak, D. L. Persistence of Perfluoroalkyl Acid Precursors in AFFF-Impacted Groundwater and Soil. *Environ. Sci. Technol.* **2013**, *47* (15), 8187–8195. <https://doi.org/10.1021/es4018877>.
- (7) Calafat, A. M.; Kuklennyik, Z.; Reidy, J. A.; Caudill, S. P.; Tully, J. S.; Needham, L. L. Serum Concentrations of 11 Polyfluoroalkyl Compounds in the U.S. Population: Data from the National Health and Nutrition Examination Survey (NHANES) 1999–2000. *Environ. Sci. Technol.* **2007**, *41* (7), 2237–2242. <https://doi.org/10.1021/es062686m>.
- (8) Fenton, S. E.; Ducatman, A.; Boobis, A.; DeWitt, J. C.; Lau, C.; Ng, C.; Smith, J. S.; Roberts, S. M. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environmental Toxicology and Chemistry* **2021**, *40* (3), 606–630. <https://doi.org/10.1002/etc.4890>.

- (9) Lee, J. E.; Choi, K. Perfluoroalkyl Substances Exposure and Thyroid Hormones in Humans: Epidemiological Observations and Implications. *Ann Pediatr Endocrinol Metab* **2017**, *22* (1), 6. <https://doi.org/10.6065/apem.2017.22.1.6>.
- (10) Guillette, T. C.; McCord, J.; Guillette, M.; Polera, M. E.; Rachels, K. T.; Morgeson, C.; Kotlarz, N.; Knappe, D. R. U.; Reading, B. J.; Strynar, M.; Belcher, S. M. Elevated Levels of Per- and Polyfluoroalkyl Substances in Cape Fear River Striped Bass (*Morone Saxatilis*) Are Associated with Biomarkers of Altered Immune and Liver Function. *Environment International* **2020**, *136*, 105358. <https://doi.org/10.1016/j.envint.2019.105358>.
- (11) Das, K. P.; Grey, B. E.; Rosen, M. B.; Wood, C. R.; Tatum-Gibbs, K. R.; Zehr, R. D.; Strynar, M. J.; Lindstrom, A. B.; Lau, C. Developmental Toxicity of Perfluorononanoic Acid in Mice. *Reproductive Toxicology* **2015**, *51*, 133–144. <https://doi.org/10.1016/j.reprotox.2014.12.012>.
- (12) Blake, B. E.; Fenton, S. E. Early Life Exposure to Per- and Polyfluoroalkyl Substances (PFAS) and Latent Health Outcomes: A Review Including the Placenta as a Target Tissue and Possible Driver of Peri- and Postnatal Effects. *Toxicology* **2020**, *443*, 152565. <https://doi.org/10.1016/j.tox.2020.152565>.
- (13) Fei, C.; McLaughlin, J. K.; Lipworth, L.; Olsen, J. Maternal Levels of Perfluorinated Chemicals and Subfecundity. *Human Reproduction* **2009**, *24* (5), 1200–1205. <https://doi.org/10.1093/humrep/den490>.
- (14) VanNoy, B. N.; Lam, J.; Zota, A. R. Breastfeeding as a Predictor of Serum Concentrations of Per- and Polyfluorinated Alkyl Substances in Reproductive-Aged Women and Young Children: A Rapid Systematic Review. *Curr Envir Health Rpt* **2018**, *5* (2), 213–224. <https://doi.org/10.1007/s40572-018-0194-z>.
- (15) Šabović, I.; Cosci, I.; De Toni, L.; Ferramosca, A.; Stornaiuolo, M.; Di Nisio, A.; Dall’Acqua, S.; Garolla, A.; Foresta, C. Perfluoro-Octanoic Acid Impairs Sperm Motility through the Alteration of Plasma Membrane. *J Endocrinol Invest* **2020**, *43* (5), 641–652. <https://doi.org/10.1007/s40618-019-01152-0>.
- (16) United States Environmental Protection Agency. Drinking Water Health Advisory for Perfluorooctane Sulfonate (PFOS) EPA Document Number 822-R-16-004, 2016.
- (17) United States Environmental Protection Agency. Drinking Water Health Advisory for Perfluorooctanoic Acid (PFOA) EPA Document Number: 822-R-16-005, 2016.

- (18) United States Environmental Protection Agency. *PFAS National Primary Drinking Water Regulation Rulemaking*; Document Citation: 88 FR 18638; Proposed Rule 2023–05471; 2023; pp 18638–18754. <https://www.federalregister.gov/documents/2023/03/29/2023-05471/pfas-national-primary-drinking-water-regulation-rulemaking#addresses>.
- (19) Brusseau, M. L.; Anderson, R. H.; Guo, B. PFAS Concentrations in Soils: Background Levels versus Contaminated Sites. *Science of The Total Environment* **2020**, *740*, 140017. <https://doi.org/10.1016/j.scitotenv.2020.140017>.
- (20) Dauchy, X.; Boiteux, V.; Bach, C.; Rosin, C.; Munoz, J.-F. Per- and Polyfluoroalkyl Substances in Firefighting Foam Concentrates and Water Samples Collected near Sites Impacted by the Use of These Foams. *Chemosphere* **2017**, *183*, 53–61. <https://doi.org/10.1016/j.chemosphere.2017.05.056>.
- (21) Dauchy, X.; Boiteux, V.; Colin, A.; Hémar, J.; Bach, C.; Rosin, C.; Munoz, J.-F. Deep Seepage of Per- and Polyfluoroalkyl Substances through the Soil of a Firefighter Training Site and Subsequent Groundwater Contamination. *Chemosphere* **2019**, *214*, 729–737. <https://doi.org/10.1016/j.chemosphere.2018.10.003>.
- (22) United States Airforce. Former Loring Airforce Base, Maine. *Official United States Air Force Website*. December 2012. www.afcec.af.mil/About-Us/Fact-Sheets/Display/Article/466134/former-loring-air-force-base-maine/.
- (23) Mann, M. M.; Tang, J. D.; Berger, B. W. Engineering Human Liver Fatty Acid Binding Protein for Detection of Poly- and Perfluoroalkyl Substances. *Biotech & Bioengineering* **2022**, *119* (2), 513–522. <https://doi.org/10.1002/bit.27981>.
- (24) United States Environmental Protection Agency. *Loring Air Force Base Limestone, ME; Announcements and Key Topics*. www.epa.gov/superfund/loring (accessed 2023-06-06).
- (25) Prendergast, F. G.; Meyer, M.; Carlson, G. L.; Iida, S.; Potter, J. D. Synthesis, Spectral Properties, and Use of 6-Acryloyl-2-Dimethylaminonaphthalene (Acrylodan). A Thiol-Selective, Polarity-Sensitive Fluorescent Probe. *J Biol Chem* **1983**, *258* (12), 7541–7544.
- (26) Sheng, N.; Li, J.; Liu, H.; Zhang, A.; Dai, J. Interaction of Perfluoroalkyl Acids with Human Liver Fatty Acid-Binding Protein. *Arch Toxicol* **2016**, *90* (1), 217–227. <https://doi.org/10.1007/s00204-014-1391-7>.

- (27) Ku, C.-Y.; Liu, Y.-H.; Lin, H.-Y.; Lu, S.-C.; Lin, J.-Y. Liver Fatty Acid-Binding Protein (L-FABP) Promotes Cellular Angiogenesis and Migration in Hepatocellular Carcinoma. *Oncotarget* **2016**, *7* (14), 18229–18246. <https://doi.org/10.18632/oncotarget.7571>.
- (28) Smathers, R. L.; Petersen, D. R. The Human Fatty Acid-Binding Protein Family: Evolutionary Divergences and Functions. *Hum Genomics* **2011**, *5* (3), 170. <https://doi.org/10.1186/1479-7364-5-3-170>.
- (29) Anderson, R. H.; Long, G. C.; Porter, R. C.; Anderson, J. K. Occurrence of Select Perfluoroalkyl Substances at U.S. Air Force Aqueous Film-Forming Foam Release Sites Other than Fire-Training Areas: Field-Validation of Critical Fate and Transport Properties. *Chemosphere* **2016**, *150*, 678–685. <https://doi.org/10.1016/j.chemosphere.2016.01.014>.
- (30) Place, B. J.; Field, J. A. Identification of Novel Fluorochemicals in Aqueous Film-Forming Foams Used by the US Military. *Environ. Sci. Technol.* **2012**, *46* (13), 7120–7127. <https://doi.org/10.1021/es301465n>.
- (31) Nickerson, A.; Rodowa, A. E.; Adamson, D. T.; Field, J. A.; Kulkarni, P. R.; Kornuc, J. J.; Higgins, C. P. Spatial Trends of Anionic, Zwitterionic, and Cationic PFASs at an AFFF-Impacted Site. *Environ. Sci. Technol.* **2021**, *55* (1), 313–323. <https://doi.org/10.1021/acs.est.0c04473>.
- (32) McGuire, M. E.; Schaefer, C.; Richards, T.; Backe, W. J.; Field, J. A.; Houtz, E.; Sedlak, D. L.; Guelfo, J. L.; Wunsch, A.; Higgins, C. P. Evidence of Remediation-Induced Alteration of Subsurface Poly- and Perfluoroalkyl Substance Distribution at a Former Firefighter Training Area. *Environ. Sci. Technol.* **2014**, *48* (12), 6644–6652. <https://doi.org/10.1021/es5006187>.
- (33) Moody, C. A.; Field, J. A. Determination of Perfluorocarboxylates in Groundwater Impacted by Fire-Fighting Activity. *Environ. Sci. Technol.* **1999**, *33* (16), 2800–2806. <https://doi.org/10.1021/es981355+>.
- (34) Guelfo, J. L.; Higgins, C. P. Subsurface Transport Potential of Perfluoroalkyl Acids at Aqueous Film-Forming Foam (AFFF)-Impacted Sites. *Environ. Sci. Technol.* **2013**, *47* (9), 4164–4171. <https://doi.org/10.1021/es3048043>.
- (35) Reinikainen, J.; Perkola, N.; Äystö, L.; Sorvari, J. The Occurrence, Distribution, and Risks of PFAS at AFFF-Impacted Sites in Finland. *Science of The Total Environment* **2022**, *829*, 154237. <https://doi.org/10.1016/j.scitotenv.2022.154237>.

- (36) Nason, S. L.; Koelmel, J.; Zuverza-Mena, N.; Stanley, C.; Tamez, C.; Bowden, J. A.; Godri Pollitt, K. J. Software Comparison for Nontargeted Analysis of PFAS in AFFF-Contaminated Soil. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (4), 840–846. <https://doi.org/10.1021/jasms.0c00261>.
- (37) Nason, S. L.; Stanley, C. J.; PeterPaul, C. E.; Blumenthal, M. F.; Zuverza-Mena, N.; Silliboy, R. J. A Community Based PFAS Phytoremediation Project at the Former Loring Airforce Base. *iScience* **2021**, *24* (7), 102777. <https://doi.org/10.1016/j.isci.2021.102777>.
- (38) *Testing for Lead in K-12 School Drinking Water*; Confirmed Results; Maine CDC Drinking Water Program, 2023. <https://www.maine.gov/dhhs/mecdc/environmental-health/dwp/pws/testingLeadSchoolWater.shtml>.
- (39) *PFAS in Public Water Systems*; Confirmed Results; Maine CDC Drinking Water Program, 2022. <https://www.maine.gov/dhhs/mecdc/environmental-health/dwp/pws/pfas.shtml>.
- (40) Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol.* **2013**, *47* (19), 11293–11301. <https://doi.org/10.1021/es4026722>.
- (41) Field, J. A.; Schultz, M.; Barofsky, D. Identifying Hydrocarbon and Fluorocarbon Surfactants in Specialty Chemical Formulations of Environmental Interest by Fast Atom Bombardment/Mass Spectrometry. *Chimia* **2003**, *57* (9), 556. <https://doi.org/10.2533/000942903777679028>.
- (42) Moe, M. K.; Huber, S.; Svenson, J.; Hagenaaers, A.; Pabon, M.; Trümper, M.; Berger, U.; Knapen, D.; Herzke, D. The Structure of the Fire Fighting Foam Surfactant Forafac®1157 and Its Biological and Photolytic Transformation Products. *Chemosphere* **2012**, *89* (7), 869–875. <https://doi.org/10.1016/j.chemosphere.2012.05.012>.
- (43) García, R. A.; Chiaia-Hernández, A. C.; Lara-Martin, P. A.; Loos, M.; Hollender, J.; Oetjen, K.; Higgins, C. P.; Field, J. A. Suspect Screening of Hydrocarbon Surfactants in AFFFs and AFFF-Contaminated Groundwater by High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2019**, *53* (14), 8068–8077. <https://doi.org/10.1021/acs.est.9b01895>.

Chapter 4

A genetically encoded biosensor for direct detection of perfluorooctanoic acid

4.1 Abstract

Determination of per- and polyfluoroalkyl substances (PFAS) in drinking water at the low levels set by regulatory officials has been a major focus for sensor developing researchers. However, it is becoming more apparent that detection of these contaminants at part per billion and even part per million levels is extremely environmentally relevant and necessary. Here, a fluorescent biosensor for the rapid detection of PFOA was engineered based on human liver fatty acid binding protein (hLFABP). By conjugating circularly permuted green fluorescent protein (cp.GFP) to a split hLFABP construct, the biosensor was able to detect perfluorooctanoic acid PFOA in PBS as well as spiked environmental water samples with LODs of 235 and 335 ppb respectively. Furthermore, the *E. coli* cells cytosolically expressing the protein-based sensor were demonstrated to quickly detect PFOA, demonstrating feasibility of whole-cell sensing. Overall, this work demonstrates a platform technology utilizing a circularly permuted GFP and split hLFABP conjugate as a label free optical sensing strategy for PFOA.

4.2 Introduction

With a rise in evidence of PFAS accumulation and toxicity comes a wave of regulatory changes and calls for action that highlight the necessity of quick, relatively easy ways to detect chemicals like PFOA.^{1,2} However, this challenge has proven non-trivial given the diversity of the chemicals as well as their limited reactivity and vast concentrations ranges. Currently, standard PFAS detection relies on chromatography techniques coupled with tandem mass spectroscopy. These methods are highly precise with detection limits in the range of 1 ng/L (1 ppt) for aqueous samples (EPA Methods 533, 537, and 537.1).³⁻⁵ Increasing health concerns and new regulations in response to these concerns have led to development of new PFAS sensors capable of detecting compounds in drinking water. While still constrained to bench level research, the most successful technologies revolve around the use of molecularly imprinted polymers (MIPS) to capture and detect PFAS down to the ppt levels determined in health advisories set by the EPA.^{2,6-8} Nonetheless, these methods often require extensive sample preparation, and/or are limited to drinking water samples making PFAS detection impractical for other applications and inaccessible for most communities.

Human exposure to PFAS has been contributed to multiple pathways including ingestion of contaminated water and food including crops grown in contaminated soils and biosolids as well as general dermal adsorption.⁹ Therefore, as PFAS contamination is found to be more and more ubiquitous in the environment, the necessity of detection at these higher concentrations and in a vast variety of matrices becomes clearer. In fact, several states in the U.S. report PFOA and PFOS concentrations in the part per million (ppm) in non-drinking water and various soil sources.^{10,11} High levels of PFAS pollution are common near manufacturing facilities, sites storing PFAS related waste, and areas utilizing fluorinated AFFFs like airports.¹² In fact, limited analyses of

ground water and sediment samples in locations where PFAS containing AFFs were used, have demonstrated that a wide variety of PFAS chemicals can persist in the environment at high concentrations, several decades after release.¹³⁻¹⁷

Agricultural lands have become reservoirs for PFAS as these chemicals are emitted directly into the environment or brought in through irrigation waters and soil amendments like treated sewage sludge and soil conditioners.¹⁸⁻²⁰ Therefore, this contamination is reflected in food crops. Studies have shown plants are capable of taking up and accumulating PFAS with preference to medium and short chain chemicals like PFOA and PFOS.^{21,22} Specifically, for highly contaminated areas, this has been shown to range from mg/kg to mg/kg dry weight levels in a variety of crops²³⁻²⁵, which is orders of magnitude above the limits set for drinking water. With this in mind, there is a clear need for easy and rapid PFAS detection in a multitude of matrices and in a wide range of concentrations with minimal pre-processing. In order to grow necessary food crops and livestock, and allow safe use of outdoor recreation areas, people must be able to easily determine contamination.

Biosensors have often been used for detection of pathogens and contaminants in agricultural products and environmental samples as they offer the advantage of tunability in terms of sensitivity and selectivity as well as the possibility for minimal sample pre-treatment.²⁶⁻²⁹ Despite not having been assessed for real world feasibility, several biosensor platforms have been developed for PFAS detection. Some of these technologies are considered “whole cell biosensors”, created around bacteria with engineered biological promoters that induce fluorescent protein expression upon PFAS interaction.^{30,31} However, these have yet to be optimized for quick read-outs, often taking

24-48 hours. Biosensors utilizing individual binding proteins and antibodies as PFAS receptors rather than whole cell systems, have also been shown to detect PFOA and/or PFOS with various transducers including fiber optics.^{32,33} We previously designed an acrylodan based fluorescent sensor for detection of several PFAS in water based on human liver fatty acid binding protein (hLFABP).³⁴ While promising, to utilize the robustness and ease of whole-cell biosensors for direct detection in a variety of sample formats, we must first develop a genetically encoded system or protein capable of intrinsic detection of PFAS.

In this study, we introduce a biosensing scaffold capable of detecting PFOA in aqueous solutions based on hLFABP and utilizing circularly permuted green fluorescent protein (GFP) that can be further optimized for whole cell detection. This fusion protein construct shown in **Figure 6.1**, exhibits increased intrinsic fluorescence upon PFOA binding *in vitro* with a LOD of 236 ppb, a level well within concentration ranges seen in highly contaminated sites. This is also achieved with minimal protein expression and purification steps and no secondary, post purification modifications. We also demonstrate the feasibility of this construct to be utilized *in vivo* through cytosolic *E. coli* expression. Our results provide a promising detection platform for use in non-aqueous and heterogeneous media detection as whole cell sensors offer robustness, ease of use, application flexibility as compared to sensors based on purified proteins.³⁵

4.3 Materials and Methods

4.3.1 Molecular Biology

For DNA maintenance, *E. coli* strain DH5a was used and *E. coli* strain BL21 (DE3) was used for protein expression. Unless otherwise stated, all molecular biology procedures for PCR amplification, plasmid preparation, cell transformation and subcloning were performed according to standard methods supplied by manufacturers. The gene encoding *E. coli* codon optimized human liver fatty acid binding protein (hLFABP) (NCBI 2168) was previously subcloned into pET-28a(+) using *Bam*HI/*Xho*I restriction sites³⁴ and the cp.GFP.PPYF gene fragment was amplified from EcMBP165-cpGFP.PPYF.pRSET (Addgene plasmid #33372)³⁶. The cp.GFP.PPYF fragment as well as the destination vector, hLFABP-pET28a(+) were amplified using primer sets FragmentCP.GFP and VectorCP.GFP respectively. The insert and linear vector were then simultaneously digested and ligated via Golden Gate assembly utilizing *Paq*CI and T4 DNA Ligase (New England Biolabs). All ligated sequences were verified by Sanger Sequencing (Europhins Genomics).

FragmentCP.GFP_F: 5' TATCACCTGCACTAaggcagcggcagctacaacgtcttcat 3'

FragmentCP.GFP_R: 5' TATCACCTGCACTAacccccgtaaagtgtactccagcttg 3'

VectorCP.GFP_F: 5' TATCACCTGCACTAagggtaaagtgatccaaaacgaattaccgttg 3'

Vector.CP.GFP_R: 5' TATCACCTGCACTAtgccgctgcccgcgtaat 3'

4.3.2 Protein expression and purification

The recombinant protein was produced after transformation into *E. coli* BL21 (DE3). Saturated cell solutions were pelleted by centrifugation (3,000 g), resuspended in fresh LB media containing

kanamycin (50 µg/mL), and grown at 37°C to an OD600 of 0.6. Protein expression was then induced via addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cultures transferred to 20°C for 18 hours. Harvested cells were then pelleted by centrifugation (10,000 g) and resuspended in lysis buffer (50mM Tris-Cl, 100mM NaCl, 5% v/v glycerol, 1mM phenylmethylsulfonyl fluoride) before lysis via sonication. The clarified supernatant was then purified by Ni²⁺-affinity chromatography using Chelating Sepharose Fast Flow (Cytiva). The 1 mL column was equilibrated with 50 mM Tris-Cl buffer (pH 8) containing 10 mM imidazole, and protein was separated using a stepwise elution of imidazole up to 500 mM. The fractions collected were analyzed by SDS-PAGE to determine purity. Pure fractions were then pooled and dialyzed against phosphate buffered saline (PBS) pH 7.6. Concentration of all protein samples was measured using a Pierce BCA protein assay kit.

4.3.3 Spectroscopy

All absorbance and fluorescence measurements for cpGFP.hLFABP were performed using a Synergy Neo2 Hybrid Multi-Mode Microplate Reader (Biotek) at room temperature and under steady state conditions.

4.3.4 *In vitro* assays

PFOA binding assays *in vitro* were performed by titrating PFOA into protein (1 µM final concentration) in either PBS buffer (pH 7.6) or creek water taken from the Dell and Meadow Creek on the grounds of University of Virginia. To determine sensor ability in systems containing other anionic surfactants, the assay was also performed in PBS buffer (pH 7.5) with sodium dodecyl sulfate (SDS) at a final concentration of 1 µM. Samples were allowed to equilibrate for 5 minutes

before fluorescence spectra were recorded over 500-600 nm after excitation at 395 and 485 nm. For more quantitative data, fluorescence intensity endpoint reads at 510 nm after excitation at 485 nm were also collected. The dissociation constant (K_d) was determined by fitting the fractional change in fluorescence intensity (F) at 510nm and corresponding PFOA concentrations to a one-site binding model using nonlinear regression after correcting for protein in buffer only fluorescence:

$$\frac{\Delta F}{F_0} = \frac{\Delta F}{F_{0\max}} * \frac{[\text{PFOA}]}{(K_{d,\text{PFOA}} + [\text{PFOA}])} \quad (1)$$

4.3.5 *In vivo* assays

For *in vivo* assays, the induced cells were harvested and resuspended in PBS (pH 7.6) to cut down on background media fluorescence. After titration with PFOA, cells were allowed to equilibrate for 30 minutes at room temperature. After 30 seconds of orbital shaking, the fluorescence spectra as well as endpoint data were collected as described above. After fluorescence data was normalized by OD600, the fractional change in fluorescence intensity at 510nm was plotted against PFOA concentration. However, to account for more nonspecific interactions, the data were fit to a log-dose response model where a half maximal effective concentration (EC50) was obtained:

$$\frac{\Delta F}{F_0} = \frac{\Delta F}{F_{0\min}} + \frac{\frac{\Delta F}{F_{0\max}} - \frac{\Delta F}{F_{0\min}}}{\left(1 + \frac{EC50^{HS}}{[\text{PFOA}]}\right)} \quad (2)$$

4.3.6 Limit of Detection Calculations

The limits of detection (LOD) were calculated for each non-linear system similarly to existing literature.³⁷ The base equation (**Eq. 3**) determines LOD, represented by the deviation in concentration (x_D), by multiplying the standard deviation of blank samples, (s_{y0}), with the coefficient for a Student's t distribution, (t)³⁸.

$$x_D = t * s_{y0} \quad (3)$$

While this is often the simplest method for LOD determination, it does not accurately consider response deviation, nor does it measure standard deviation of calibration measurements. For *in vitro* assays, we see a first order binding dependence of signal on PFOA concentration shown in **Eq. 4** after rearrangement and simplification of **Eq. 1**. Concentration of PFOA is now represented as (x) while the change in fluorescence response is now (y) with max response as (B). Therefore, to properly obtain LOD, we must calculate the contribution of all terms in **Eq. 4** in the deviation of x .

$$x = \frac{B * K_d}{(B - y)} \quad (4)$$

$$s_x^2 = \left(\frac{\partial x}{\partial y}\right)^2 s_y^2 + \left(\frac{\partial x}{\partial K_d}\right)^2 s_{K_d}^2 + \left(\frac{\partial x}{\partial B}\right)^2 s_B^2 \quad (5)$$

Standard deviation of x , (s_x) was calculated based on **Eq. 5** with (s_y), (s_{K_d}), and (s_B) representing error in the calibration curve from the response signal, dissociation constant, and

maximum signal respectively. The calculated deviation obtained at a specific point, (s_{y_0}), from **Eq. 5** was substituted into Eq. 3 to calculate LOD. As our system is based on the change in fluorescence, measurement deviation from a low concentration standard was used for s_{y_0} rather than blank samples. The final equation (**Eq. 6**) was then used for calculation of LOD for in vitro assays with the confidence factor $t=3$ as it corresponds to the confidence level of 95%.

$$x_D = t * \left[\left(\frac{K_d * B}{(B - y_0)^2} \right)^2 s_{y_0}^2 + \left(\frac{-K_d * y_0}{(B - y_0)^2} \right)^2 s_B^2 + \left(\frac{y_0}{(B - y_0)} \right)^2 s_{Kd}^2 \right]^{\frac{1}{2}} \quad (6)$$

This process was done similarly for the LOD calculation of the cell-based assays. However, **Eq. 2** was used as the basis instead of **Eq. 1** in order to properly model the non-specific effects that come with the complexity of using whole cells instead of purified protein for binding assays.

4.4 Results and Discussion

Fluorescent proteins (FPs) contain optical properties that are extremely dependent on the microenvironment surrounding their chromophores.^{39,40} This chromophore sensitivity has thus led to the wide use of FPs like GFP as sensing tools since small changes to protonation equilibrium are transduced easily. This can be through direct interaction of analytes and chromophores like with FP based pH and ion sensors⁴¹ or through the addition of a separate binding unit where ligand induced conformation change leads FRET-based activity or even allosteric based fluorescence changes.⁴²⁻⁴⁴ While the addition of ligand-binding receptors to FPs is conducive for binding events that lead to large changes in receptor conformation, the use of circularly permuted GFP has been proven to be capable of transducing binding events for proteins with a wide range of

conformational flexibilities.⁴⁵ The circular permutation process involves fusing the natural GFP termini which forms new termini that can then be fused to insert a receptor of interest closer to the chromophore.⁴⁶

To create a sensor capable of detecting PFOA, circularly permuted GFP (cpGFP) and human liver fatty acid binding protein were utilized as optical signal and recognition units respectively. The chosen receptor, hLFABP has not only been shown to bind PFOA with moderate affinity^{47,48} but has also been used previously as a scaffold for a PFAS biosensor not suitable for genetic encoding.³⁴ However, hLFABP is not known to have extensive conformational changes upon binding of endogenous fatty acids,^{49,50} and the little structural information of PFOA binding shows only minimal changes in alpha helical composition.^{48,51} Therefore, it is imperative that the cpGFP construct is fused near the PFOA binding region of hLFABP without interrupting residues directly responsible for ligand interactions. The residues S56 and K57 in the loop region shown in **Figure 4.1** were chosen for domain splitting.

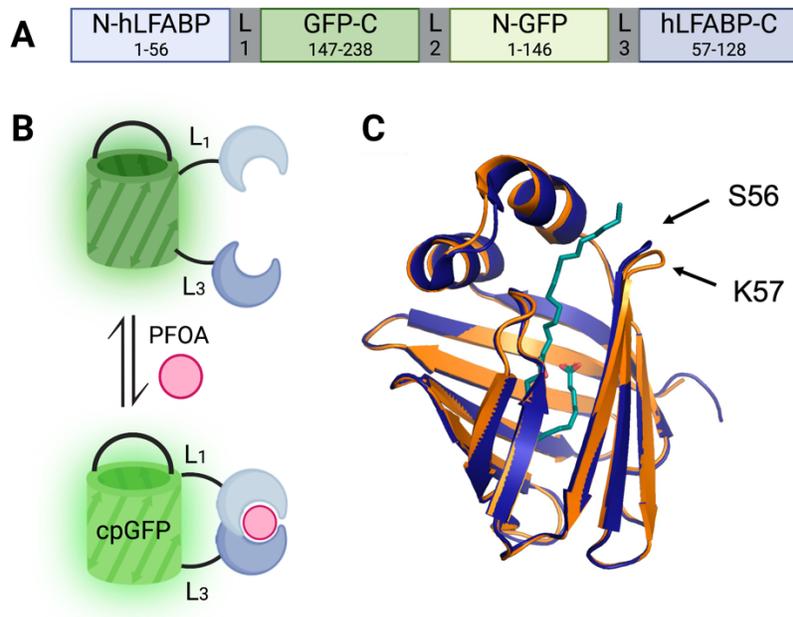


Figure 4.1: cpGFP.hLFABP construct design overview. (A) linear representation of construct with linkers L1 (GSG) L2 (GGTGG) and L3 (GG) (B) schematic of circularly permuted GFP.hLFABP at the protein level depicting binding of PFAS (C) overlay of apo form (orange) and holo form (blue) of wild type hLFABP complexed with palmitic acid (PDB IDs 3STN and 3STK respectively)⁵⁰. Location of receptor splitting was partly based on perturbations at residues 56 and 57. Figure created using Biorender.com.

While located in a flexible region away from residues taking part in electrostatic interactions^{47,48}, S56 and K57 are in a region where modest change in structure occurs upon binding of palmitic acid as shown in **Figure 4.1** as the overlay of the apo and holo form of hLFABP. (PDB 3STN and 3STK respectively).⁵⁰ Furthermore, previous work in our lab has shown insertion of the fluorophore, acrylodan, into a reasonably close residue (F50) of hLFABP is able to probe PFOA binding.³⁴ **Figure 4.1** also shows a linear map of the construct as well as a graphical overview of the sensor function. As ligand is bound by the split hLFABP, the change in microenvironment around cpGFP's chromophore will elicit a change in fluorescence. Specifically, this change is seen as an increase in fluorescence after excitation.

The designed construct was subsequently cloned as described above, expressed in *E. coli* BL21 (DE3), and purified via immobilized metal affinity chromatography. Purified fractions containing the single band of cpGFP.hLFABP at ~46.5 kDa (**Figure 4.2**) were then dialyzed and used for *in vitro* analysis.

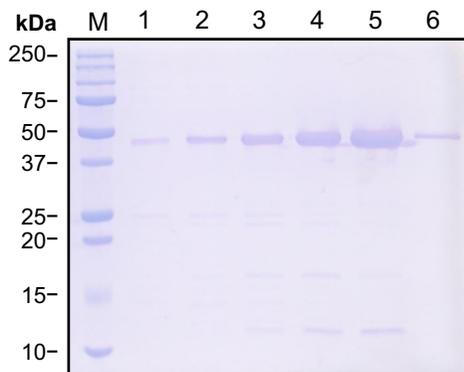


Figure 4.2: SDS-PAGE analysis of cpGFP.hLFABP purification via Ni-NTA resin. Lane M: protein marker, Lanes 1-6: purification fractions containing increasing concentrations of imidazole (50, 75, 100, 125, 200, and 500mM respectively). cpGFP.hLFABP is shown as a band of approximately 46.6 kDa.

Enhanced green fluorescent protein (EGFP), the original basis of this cpGFP signal unit, as well as our cpGFP.hLFABP fusion have absorbance peaks (**Figure 4.3**) corresponding with the protonated and deprotonated forms at ~395nm and ~490nm respectively.⁵² Upon titration of PFOA, the absorbance at 395nm increased and decreased at 485nm. This indicates shifts in the equilibrium state of the chromophore from protonated to deprotonated states upon binding of PFOA to the split hLFABP.

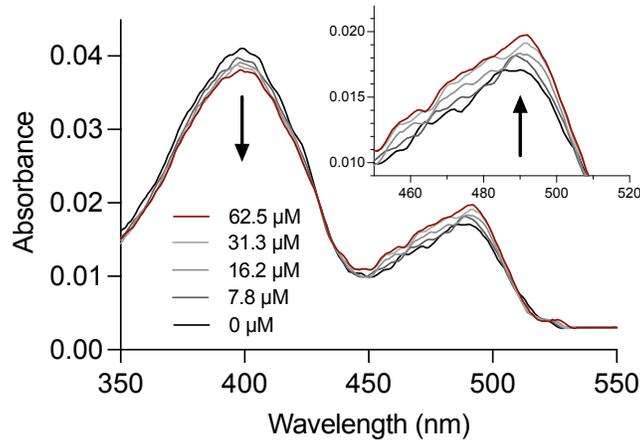


Figure 4.3: Representative absorption spectra for cpGFP.hLFABP (2 μ M). Spectral changes upon increased PFOA addition show a decreased and increased trend in intensity at 395nm and 485nm respectively. Curves are smoothed for qualitative analysis.

Fluorescence response at these two wavelengths was assessed upon addition of PFOA. **Figure 4.4** shows that while emission spectra changes occur at both wavelengths, excitation at 485nm exhibits a much more exaggerated response in overall fluorescence intensity change.

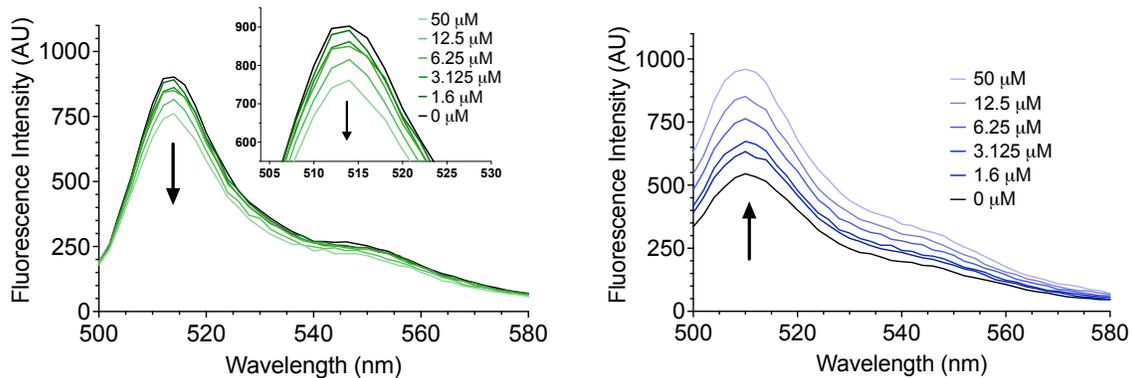
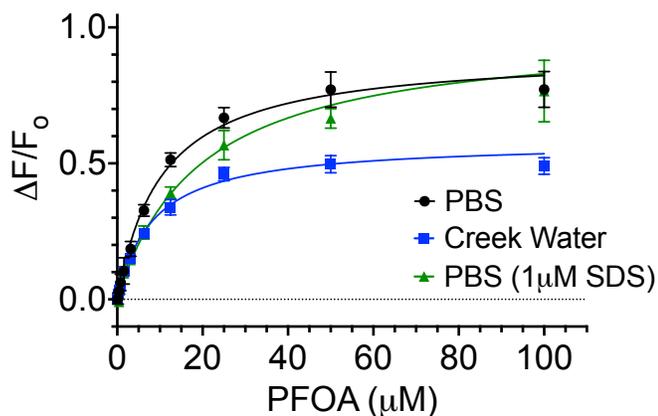


Figure 4.4: Representative fluorescence emission spectra for cpGFP.hLFABP (1 μ M) after excitation at 395nm (LEFT) and 485nm (RIGHT). Spectral changes upon PFOA titration shows a decreased and increased overall intensity for excitation at 395nm and 485nm respectively. All curves are smoothed for qualitative analysis.

Therefore, to quantify binding of PFOA to cpGFP.hLFABP, endpoint intensity at 510nm after excitation at 485nm was used. **Figure 4.5** shows binding of PFOA to cpGFP.hLFABP in PBS (pH 7.6) as a fractional change in fluorescence. This data is shown fitted to a one site binding model based on the best fit to experimental data. This is consistent with studies suggesting PFOA binds in a 1:1 stoichiometry to WT hLFABP despite cpGFP.hLFABP utilizing a split hLFABP domain.^{34,48,51} The calculated dissociation constant K_d was determined to be 11.9 ± 1.6 mM which is also consistent with previous studies characterizing binding of PFOA to WT hLFABP and variants.^{34,48,51,53}



	PBS	Creek Water	PBS + SDS
K_d (μM)	10.4 ± 1.4	8.2 ± 0.8	19.4 ± 2.3
$(\Delta F/F_0)_{\text{max}}$	0.91 ± 0.04	0.58 ± 0.02	0.99 ± 0.04
LOD (μM)	0.57 ± 0.2	0.8 ± 0.3	1.13 ± 0.7
LOD (ppb)	236	330	467

Figure 4.5: Binding of PFOA to cpGFP.hLFABP ($1 \mu\text{M}$) after titration in PBS (black circles), creek water (blue squares), and PBS containing $1 \mu\text{M}$ SDS (green triangles). Data are fitted to a one site binding model with the fractional occupancy represented as the fractional change in fluorescence intensity at 510nm after excitation at 485nm. The represented points are mean values \pm SEM with $n=3$. Bottom table displays calculated model constants and limits of detection in all systems.

To assess feasibility and selectivity of the sensor in more realistic application systems, these experiments were also done with spiked water samples taken from the Dell and Meadow Creek on

the grounds of University of Virginia as well as in buffer containing the anionic surfactant, sodium dodecyl sulfate (SDS), as a competitor for binding (**Figure 4.5**). While K_d values obtained in all buffer systems were similar as shown in **Figure 4.5**, the maximum response, and limits of detection (LODs) varied. While maximum response, $(\Delta F/F_o)_{\max}$, is reduced in creek water-based assays, the overall binding affinity and limits of detection remain comparable to PBS based data with LODs in the hundreds of ppb. While this limit is pushed toward the part per million for buffer containing SDS, the ability to detect PFOA at this level in a system with a known hLFABP binder is promising for application of samples that contain other anionic surfactant co-contaminants.

One of the most valuable aspects of genetically encoded sensors is the ability to be used *in vivo*. Specifically, in environmental detection, utilizing whole bacterial cells introduces a robustness to a sensor as compared to proteins alone in terms of tolerance to physical changes like pH and temperature as cell membranes act as a barrier from harsh environmental conditions.⁵⁴ Furthermore, whole cell-based systems are usually more amenable to immobilization-based implementation than proteins⁵⁵ and offer advantages such as portability and the possibility of detection in complex matrices with minimal sample preparation.⁵⁶

Therefore, as a first pass at feasibility of whole cell detection, *E. coli* cells expressing cpGFP.hLFABP cytosolically were incubated with PFOA, and changes in fluorescence monitored. The increase in fluorescence upon PFOA titration is shown in **Figure 4.6**. To account for complications the bacterial interactions bring, the data was fit to a log-dose response model rather than a one site binding model. While only eliciting a maximal response of around 10%, the sensor expressing cells are capable of detecting PFOA in PBS (pH 7.6) with a LOD of 2.4 ppm.

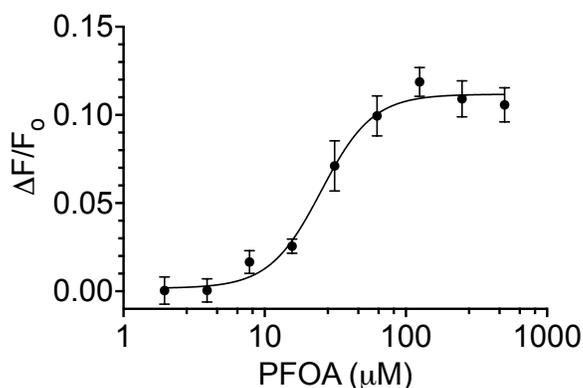


Figure 4.6: Fluorescence intensity change at 510nm after excitation at 485nm of *E. coli* expressing cpGFP.hLFABP upon PFOA titration. Data are normalized by cell concentration (OD600) and fitted to a dose response model. The represented points are mean values \pm SEM with n=3.

Overall, in this study we demonstrate the design and initial applicability of an intrinsic fluorescent PFOA sensor based on conjugation of cpGFP to a split hLFABP construct. The purified fusion protein exhibited dose-dependent changes in absorption spectra as well as fluorescence intensity that demonstrate saturation behavior. This shows that despite small changes in structural conformation, splitting hLFABP allows PFOA induced changes large enough for transduction by cpGFP without detrimental changes to binding as compared to wild type hLFABP.

Direct detection of PFOA was achieved in buffer as well spiked water samples with LODs within relative environmental concentration ranges without sample pre-treatment. We also demonstrate the feasibility of utilizing this construct in whole bacterial cells through cytosolic expression to detect PFOA. This is extremely promising as minimal optimization was done in terms of individual GFP and hLFABP moieties. As has been shown prior, just optimizing the linkers between cpGFP and receptor proteins can improve optical signal by more than 200%.³⁶

4.5 Conclusions and Future Directions

As further evidence continues to elucidate the wide-reaching contamination of PFAS chemicals, the need for detection strategies grows. Specifically, detection in a wide variety of sample types and concentrations are necessary. As mentioned above, biosolids and irrigation waters have become reservoirs as well as transporters of PFAS to crops, livestock, and people. In the last 10 years, PFAS have been found at detectable levels not only in food crops but also livestock, including milk from dairies with groundwater and hay contamination.⁵⁷ Therefore, to address the need for easy detection in all sample types, our lab looks towards utilizing synthetic biology. At the bench scale, biosensor feasibility studies are usually conducted in liquid culture. However, genetically encoded biosensors have the capability for a multitude of applications including immobilization for the design of test strips, and microfluidic devices.^{35,55} We also know that with biological tool kits, researchers have the capability to modify proteins and enzymes to enhance sensitivity and selectivity drastically.^{42,58} Furthermore, genetically encoded biosensors can be designed with multiple functions. By introducing new proteins and enzymes or creating genetic circuits, whole cell sensors seem to have unlimited capabilities, including pre-treatment simplification.^{26,35} Therefore, by demonstrating cpGFP.hLFABP as a promising platform for intrinsic fluorescent-based detection, we hope to enable further synthetic biology-based approaches for PFAS detection. In this section I will discuss strategies to improve the whole cell-based utilization of the cpGFP.hLFABP biosensor as well as outline at a high level several methods for improving sensitivity and specificity of the biosensor.

Whole cell biosensors most commonly use simplistic gene circuits for analyte detection which can be transcriptionally or translationally based. As a simplistic overview, transcriptionally based

systems rely on specific interactions of analyte species with sensing domains (oftentimes transcription factors) that then drive the transcription and translation of reporter proteins.⁵⁶ Translational sensing mechanisms often use regulatory RNA molecules for sensing where aptamer domains are linked to regulatory domains.⁵⁹ These “riboswitches” regulate expression of reporter genes through specific binding. Unlike these categories of sensors, the discussed cpGFP.hLFABP construct does not rely on the regulation of reporter gene expression. Instead, the sensing mechanism and signal output are fused allowing for coupled activation and reporting. While this creates a simpler and more robust system than classic genetic circuit-based sensing, response to extracellular contaminants still poses a challenge for all whole-cell biosensors as the physical cell walls and membranes act as physical barriers. Approaches to circumvent this issue revolve around increasing the physical interaction of protein and analyte through increased uptake of environmental contaminants into the cell or expression of the protein biosensor on the cell surface.

One such method would be through the use of expression host systems with more permeable membranes to increase analyte uptake. Specifically, future work should look at the use of *E. coli* AS19. As further discussed in **Appendix A**, AS19 have hyperpermeable outer membranes due to defects in the lipopolysaccharides of the outer membrane.⁶⁰ Aside from changing the overall outer membrane integrity, it is possible to adjust membrane permeability through the deletion or amplification of genes encoding membrane proteins. Membrane proteins account for about 20-30% of synthesized proteins in all organisms and are essential to a variety of cellular functions including selective passage of solutes and key metabolites across cell membranes.⁵⁴ In *E. coli* there are two main proteins required for transport of exogenous fatty acids; FadL and FadD which facilitate fatty acid uptake through the outer and inner membranes respectively.^{61,62} Since PFAS

like PFOA have been shown to bind to fatty acid binding proteins with moderate affinity, it is believed that overexpression of the outer membrane transport protein FadL will facilitate uptake of PFAS molecules and increase interaction of these analytes with our protein biosensor. It has even been previously shown that increased FadL expression in *E. coli* leads to increased uptake of palmitic acid.⁶³

Preliminary studies were conducted on the effect of FadL overexpression on biosensor sensitivity for whole cell PFOA detection, and experimental details are outlined in the Supplemental Information section of this chapter. The native *E. coli* gene *fadL* was cloned into the pRSET plasmid and the outer membrane protein was co-expressed through a T7 promoter system with our cpGFP.hLFABP construct. Upon direct comparison of PFOA sensing in **Figure 4.7**, it appears that greater sensitivity was not achieved.

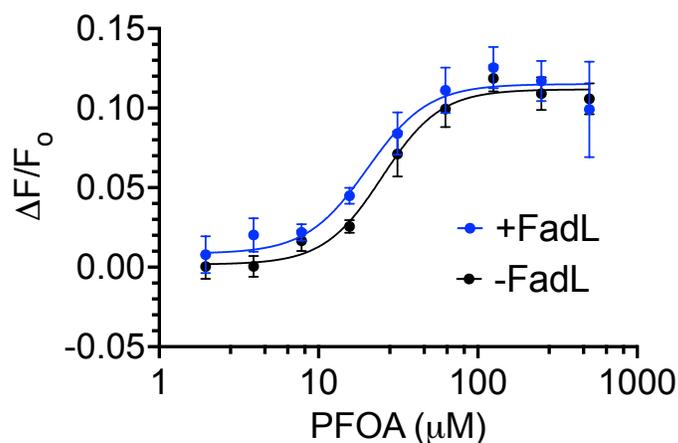


Figure 4.7: Fluorescence intensity change at 510nm after excitation at 485nm of *E. coli* expressing cpGFP.hLFABP with (blue) and without (black) FadL expression upon PFOA titration. Data are normalized by cell concentration (OD600) and fitted to a dose response model. The represented points are mean values \pm SEM with n=3.

The transport of fatty acids through the β -barrel protein is thought to involve lateral diffusion of the substrate as well as conformational change of FadL that results in the formation of the exit channel.⁶⁴ Since PFOA (8 carbons) is smaller than average long chain fatty acids (12-18 carbons) it is possible the interaction of PFOA with FadL does not sufficiently activate the exit formation resulting in low uptake and transport. However, the feasibility of this method cannot be ruled out as membrane proteins are notoriously difficult to express recombinantly. Specifically, overproduction of membrane proteins often leads to incredibly low yields and can overwhelm the hosts folding machinery or lead to host toxicity.⁶⁵ The challenge of low yields is extrapolated through the co-expression of cpGFP.LFABP as cellular machinery and energy are expended in the formation and folding of this extremely soluble and easily expressed protein. Therefore, further work should investigate optimization of both FadL expression as well as co-expression of the membrane protein and biosensor.

Along with increasing analyte uptake, the method of cell surface display could improve biosensor sensitivity by removing the physical cell barrier to maximize sensor-PFOA interaction and decrease noise from intracellular expression. Expression of recombinant proteins on a cells surface is used in a multitude of applications, most commonly in the development and screening of protein and antibody libraires. This technique has even been utilized for biosensor applications in several prokaryote and eukaryote hosts including phages, yeast, and bacteria.⁶⁶ Outer membrane proteins and autotransporters are frequently used as protein carriers for Gram-negative bacteria like *E. coli*.⁶⁶ Therefore, preliminary work with methods outlined in the Supplemental Information section of this chapter investigates the ability of surface display of cpGFP.hLFABP on *E. coli* BL21 through a system based on the autotransporter protein AIDA-I. Autotransporter proteins function

in cell adhesion and protein hydrolysis, and their structure promotes translocation of N-terminal linked passenger proteins across membranes through their porin-like β -barrel structures.^{67,68} As they do not require other proteins for translocation, they are capable of transporting larger proteins than common outer membrane proteins used for bacterial surface display like OmpF.^{69,70} Autotransporter proteins like AIDA-I are also extremely efficient and can exhibit more than 250,000 passenger proteins on the surface of a single bacterial cell.⁷¹ However, as with all membrane proteins, many factors play into their expression efficiency. The protein surface display system consists of several elements; an AIDA specific signal sequence for localization, passenger protein (cpGFP.hLFABP), linker and myc antibody tag regions, and the translocation unit (AIDAc).

Upon expression of the biosensor-AIDAc fusion, however, cells were no longer visibly fluorescent compared to cytosolically expressed cpGFP.hLFABP. While it is possible that the efficiency of cytosolic expression outmatches that of surface display, the western blot in **Figure 4.8** indicates improper surface expression.

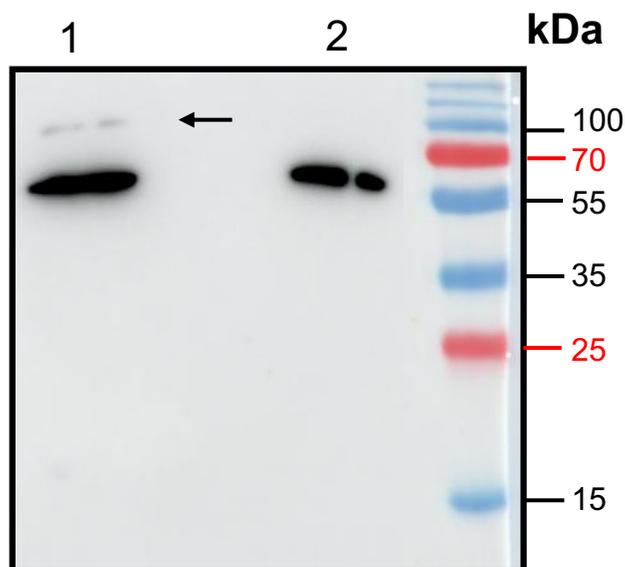


Figure 4.8: Anti myc-tag Western Blot of whole cell lysates for *E. coli* expressing (1) cpGFP.hLFABP-AIDAc fusion and (2) AIDAc protein without passenger protein used as a control. Arrow points to band representing full length cpGFP.hLFABP-AIDAc fusion (~90 kDa)

Specifically, there is clear indication of the presence of the AIDAc protein at ~60 kDa in both cpGFP.hLFABP.pAIDA as well as the pAIDA control in which no passenger was attached. While the full-size fusion protein (~90 kDa) is seen, the band is faint indicating possible cleavage at the TEV protease site or other degradation. Therefore, more work should be done to develop the surface display of this biosensor. The use of different *E. coli* strains to reduce cleavage of surface expressed proteins⁶⁵ or the removal of proteolytic regions are natural starting points.

There are many methods for optimizing and designing whole cell biosensors. One of the most appealing aspects of protein-based sensing is the ability to increase specificity and selectivity of analytes to the recognition protein. This is often done through the use of directed evolution to mutate proteins and screen for those with increased function. The design and implementation of a library for increased hLFABP binding affinity to PFOA is discussed in **Appendix A**. However, the

fluorescent nature of the circularly permuted, split hLFABP fusion is an ideal candidate for future library screening as increases in fluorescence upon binding allows the utilization of optical screening methods. Aside from protein engineering, there are many other areas for optimization of biosensors including choice of host, expression system and location, as well as implementation and integration into an end-use system. In this section, I hope to outline a starting point for future optimization of our genetically encoded platform system utilizing cpGFP.hLFABP for PFOA detection.

4.6 Supplemental Information

Simultaneous expression of cpGFP.hLFABP and FadL for in vivo PFOA detection

The gene encoding the FadL protein was PCR amplified from *E. coli* after genomic DNA extraction (Monarch Genomic DNA Purification Kit) using primers FadLfrag_F and FadLfrag_R. The insert was then ligated into the pRSET vector via golden gate cloning with *paqCI* and T4 DNA Ligase (New England Biolabs) after backbone amplification with primers pRSET_F and pRSET_R. After transformation into *E. coli* BL21, protein expression and assays were conducted as previously described in this chapter with the exception of antibiotic marker; pRSET includes resistance to ampicillin so cells were grown in LB media containing (50 µg/mL) ampicillin.

FadLfrag_F: 5' *tattcacctgcactacacccatgagccagaaaaccctgttacaag* 3'

FadLfrag_R: 5' *tattcacctgcactatctcagaacgcgtagttaaagttagtaccga* 3'

pRSET_F: 5' *tattcacctgcactagagaattcgaagcttgatccggct* 3'

pRSET_R: 5' *tattcacctgcactaggtgggatccttategtcatcg* 3'

Bacterial surface display of cpGFP.hLFABP

The gene encoding cpGFP.hLFABP was PCR amplified out of cpGFP.hLFABP-pET28a(+) using primers cpGFPLFABP_SD_F and cpGFPLFABP_SD_R to add *SphI* and *SacI* sites onto 5' and 3' ends respectively. After amplification of the pAIDA1 vector (Addgene plasmid #79180)⁶⁶ using primers pAIDA_F and pAIDA_R to add the corresponding *SphI* site both the insert and backbone DNA were doubly digested with *SphI/SacI*, ethanol precipitated, resuspended in nuclease free water, and ligated overnight at 16°C using T4 DNA Ligase (New England Biolabs). After transformation into *E. coli* BL21, protein expression was conducted as previously described in this chapter with the exception of antibiotic marker; pAIDA1 includes resistance to chloramphenicol so cells were grown in LB media containing the antibiotic at 25 µg/mL. After induction, cells were pelleted via centrifugation (3,000 g), resuspended in a 20mM Tris based lysis buffer containing 5mM EDTA and 0.5% SDS, and heated at 90°C for 10 minutes. Lysed whole-cell samples were run on 16% Tris-tricine SDS-PAGE under reducing conditions and transferred to nitrocellulose membrane using standard protocols for Western Blot analysis with HRP-conjugated rabbit anti-myc tag antibody (Abcam).

cpGFPLFABP_SD_F: 5' *cagegtaccgcatgcatgagcttcagcggca* 3'

cp.GFPLFABP_SD_R: 5' *cagtacagagctcaatacgtttgctgatacgc* 3'

pAIDA_F: 5' *acagegtaccgagctcgaaa* 3'

pAIDA_R: 5' *cagttctggcatgccggaccctggaacagcg* 3'

4.7 References

- (1) United States Environmental Protection Agency. EPA's Per- and Polyfluoroalkyl Substances (PFAS) Action Plan EPA 823R18004, 2019.
- (2) United States Environmental Protection Agency. *PFAS National Primary Drinking Water Regulation Rulemaking*; Document Citation: 88 FR 18638; Proposed Rule 2023–05471; 2023; pp 18638–18754. <https://www.federalregister.gov/documents/2023/03/29/2023-05471/pfas-national-primary-drinking-water-regulation-rulemaking#addresses>.
- (3) J.A. Shoemaker; Grimmett, P.; Boutin, B. Determination of Selected Perfluorinated Alkyl Acids in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS), 2008.
- (4) Rosenblum, Laura, W., Steven. METHOD 533: DETERMINATION OF PER- AND POLYFLUOROALKYL SUBSTANCES IN DRINKING WATER BY ISOTOPE DILUTION ANION EXCHANGE SOLID PHASE EXTRACTION AND LIQUID CHROMATOGRAPHY/TANDEM MASS SPECTROMETRY, 2019.
- (5) Shoemaker, J., D. T. Method 537.1 Determination of Selected Per- and Polyfluorinated Alkyl Substances in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS), 2020.
- (6) Feng, H.; Wang, N.; Tran, T.; Yuan, L.; Li, J.; Cai, Q. Surface Molecular Imprinting on Dye-(NH₂)-SiO₂ NPs for Specific Recognition and Direct Fluorescent Quantification of Perfluorooctane Sulfonate. *Sensors and Actuators B: Chemical* **2014**, *195*, 266–273. <https://doi.org/10.1016/j.snb.2014.01.036>.
- (7) Jiao, Z.; Li, J.; Mo, L.; Liang, J.; Fan, H. A Molecularly Imprinted Chitosan Doped with Carbon Quantum Dots for Fluorometric Determination of Perfluorooctane Sulfonate. *Microchim Acta* **2018**, *185* (10), 473. <https://doi.org/10.1007/s00604-018-2996-y>.
- (8) Pitruzzella, R.; Arcadio, F.; Perri, C.; Del Prete, D.; Porto, G.; Zeni, L.; Cennamo, N. Ultra-Low Detection of Perfluorooctanoic Acid Using a Novel Plasmonic Sensing Approach Combined with Molecularly Imprinted Polymers. *Chemosensors* **2023**, *11* (4), 211. <https://doi.org/10.3390/chemosensors11040211>.
- (9) De Silva, A. O.; Armitage, J. M.; Bruton, T. A.; Dassuncao, C.; Heiger-Bernays, W.; Hu, X. C.; Kärrman, A.; Kelly, B.; Ng, C.; Robuck, A.; Sun, M.; Webster, T. F.; Sunderland, E. M. PFAS Exposure Pathways for Humans and Wildlife: A Synthesis of Current Knowledge

- and Key Gaps in Understanding. *Environ Toxicol Chem* **2021**, *40* (3), 631–657. <https://doi.org/10.1002/etc.4935>.
- (10) Andrews, D. Q.; Hayes, J.; Stoiber, T.; Brewer, B.; Campbell, C.; Naidenko, O. V. Identification of Point Source Dischargers of Per- and Polyfluoroalkyl Substances in the United States. *AWWA Water Science* **2021**, *3* (5). <https://doi.org/10.1002/aws2.1252>.
- (11) Carlson, G. L.; Tupper, S. Ski Wax Use Contributes to Environmental Contamination by Per- and Polyfluoroalkyl Substances. *Chemosphere* **2020**, *261*, 128078. <https://doi.org/10.1016/j.chemosphere.2020.128078>.
- (12) Salvatore, D.; Mok, K.; Garrett, K. K.; Poudrier, G.; Brown, P.; Birnbaum, L. S.; Goldenman, G.; Miller, M. F.; Patton, S.; Poehlein, M.; Varshavsky, J.; Corder, A. Presumptive Contamination: A New Approach to PFAS Contamination Based on Likely Sources. *Environ. Sci. Technol. Lett.* **2022**, *9* (11), 983–990. <https://doi.org/10.1021/acs.estlett.2c00502>.
- (13) Dauchy, X.; Boiteux, V.; Bach, C.; Rosin, C.; Munoz, J.-F. Per- and Polyfluoroalkyl Substances in Firefighting Foam Concentrates and Water Samples Collected near Sites Impacted by the Use of These Foams. *Chemosphere* **2017**, *183*, 53–61. <https://doi.org/10.1016/j.chemosphere.2017.05.056>.
- (14) Dauchy, X.; Boiteux, V.; Colin, A.; Hémar, J.; Bach, C.; Rosin, C.; Munoz, J.-F. Deep Seepage of Per- and Polyfluoroalkyl Substances through the Soil of a Firefighter Training Site and Subsequent Groundwater Contamination. *Chemosphere* **2019**, *214*, 729–737. <https://doi.org/10.1016/j.chemosphere.2018.10.003>.
- (15) Guelfo, J. L.; Higgins, C. P. Subsurface Transport Potential of Perfluoroalkyl Acids at Aqueous Film-Forming Foam (AFFF)-Impacted Sites. *Environ. Sci. Technol.* **2013**, *47* (9), 4164–4171. <https://doi.org/10.1021/es3048043>.
- (16) Houtz, E. F.; Higgins, C. P.; Field, J. A.; Sedlak, D. L. Persistence of Perfluoroalkyl Acid Precursors in AFFF-Impacted Groundwater and Soil. *Environ. Sci. Technol.* **2013**, *47* (15), 8187–8195. <https://doi.org/10.1021/es4018877>.
- (17) Reinikainen, J.; Perkola, N.; Äystö, L.; Sorvari, J. The Occurrence, Distribution, and Risks of PFAS at AFFF-Impacted Sites in Finland. *Science of The Total Environment* **2022**, *829*, 154237. <https://doi.org/10.1016/j.scitotenv.2022.154237>.

- (18) Brunn, H.; Arnold, G.; Körner, W.; Rippen, G.; Steinhäuser, K. G.; Valentin, I. PFAS: Forever Chemicals—Persistent, Bioaccumulative and Mobile. Reviewing the Status and the Need for Their Phase out and Remediation of Contaminated Sites. *Environ Sci Eur* **2023**, *35* (1), 20. <https://doi.org/10.1186/s12302-023-00721-8>.
- (19) Ghisi, R.; Vamerali, T.; Manzetti, S. Accumulation of Perfluorinated Alkyl Substances (PFAS) in Agricultural Plants: A Review. *Environmental Research* **2019**, *169*, 326–341. <https://doi.org/10.1016/j.envres.2018.10.023>.
- (20) Stahl, T.; Gassmann, M.; Falk, S.; Brunn, H. Concentrations and Distribution Patterns of Perfluoroalkyl Acids in Sewage Sludge and in Biowaste in Hesse, Germany. *J. Agric. Food Chem.* **2018**, *66* (39), 10147–10153. <https://doi.org/10.1021/acs.jafc.8b03063>.
- (21) Pérez, F.; Llorca, M.; Köck-Schulmeyer, M.; Škrbić, B.; Oliveira, L. S.; Da Boit Martinello, K.; Al-Dhabi, N. A.; Antić, I.; Farré, M.; Barceló, D. Assessment of Perfluoroalkyl Substances in Food Items at Global Scale. *Environmental Research* **2014**, *135*, 181–189. <https://doi.org/10.1016/j.envres.2014.08.004>.
- (22) Wang, W.; Rhodes, G.; Ge, J.; Yu, X.; Li, H. Uptake and Accumulation of Per- and Polyfluoroalkyl Substances in Plants. *Chemosphere* **2020**, *261*, 127584. <https://doi.org/10.1016/j.chemosphere.2020.127584>.
- (23) Blaine, A. C.; Rich, C. D.; Sedlacko, E. M.; Hyland, K. C.; Stushnoff, C.; Dickenson, E. R. V.; Higgins, C. P. Perfluoroalkyl Acid Uptake in Lettuce (*Lactuca Sativa*) and Strawberry (*Fragaria Ananassa*) Irrigated with Reclaimed Water. *Environ. Sci. Technol.* **2014**, *48* (24), 14361–14368. <https://doi.org/10.1021/es504150h>.
- (24) Liu, Z.; Lu, Y.; Song, X.; Jones, K.; Sweetman, A. J.; Johnson, A. C.; Zhang, M.; Lu, X.; Su, C. Multiple Crop Bioaccumulation and Human Exposure of Perfluoroalkyl Substances around a Mega Fluorochemical Industrial Park, China: Implication for Planting Optimization and Food Safety. *Environment International* **2019**, *127*, 671–684. <https://doi.org/10.1016/j.envint.2019.04.008>.
- (25) Zhu, H.; Kannan, K. Distribution and Partitioning of Perfluoroalkyl Carboxylic Acids in Surface Soil, Plants, and Earthworms at a Contaminated Site. *Science of The Total Environment* **2019**, *647*, 954–961. <https://doi.org/10.1016/j.scitotenv.2018.08.051>.

- (26) Karadurmus, L.; Kaya, S. I.; Ozkan, S. A. Recent Advances of Enzyme Biosensors for Pesticide Detection in Foods. *Food Measure* **2021**, *15* (5), 4582–4595. <https://doi.org/10.1007/s11694-021-01032-3>.
- (27) Mehrotra, P. Biosensors and Their Applications – A Review. *J Oral Biol Craniofac Res* **2016**, *6* (2), 153–159. <https://doi.org/10.1016/j.jobcr.2015.12.002>.
- (28) Nigam, V. K.; Shukla, P. Enzyme Based Biosensors for Detection of Environmental Pollutants-A Review. *Journal of Microbiology and Biotechnology* **2015**, *25* (11), 1773–1781. <https://doi.org/10.4014/jmb.1504.04010>.
- (29) Sahu, S.; Roy, R.; Anand, R. Harnessing the Potential of Biological Recognition Elements for Water Pollution Monitoring. *ACS Sens.* **2022**, *7* (3), 704–715. <https://doi.org/10.1021/acssensors.1c02579>.
- (30) Sunantha, G.; Vasudevan, N. A Method for Detecting Perfluorooctanoic Acid and Perfluorooctane Sulfonate in Water Samples Using Genetically Engineered Bacterial Biosensor. *Science of The Total Environment* **2021**, *759*, 143544. <https://doi.org/10.1016/j.scitotenv.2020.143544>.
- (31) Young, N. A.; Lambert, R. L.; Buch, A. M.; Dahl, C. L.; Harris, J. D.; Barnhart, M. D.; Sitko, J. C.; Jordan Steel, J. A Synthetic Biology Approach Using Engineered Bacteria to Detect Perfluoroalkyl Substance (PFAS) Contamination in Water. *Mil Med* **2021**, *186* (Suppl 1), 801–807. <https://doi.org/10.1093/milmed/usaa367>.
- (32) Cennamo, N.; Zeni, L.; Tortora, P.; Regonesi, M. E.; Giusti, A.; Staiano, M.; D’Auria, S.; Varriale, A. A High Sensitivity Biosensor to Detect the Presence of Perfluorinated Compounds in Environment. *Talanta* **2018**, *178*, 955–961. <https://doi.org/10.1016/j.talanta.2017.10.034>.
- (33) Moro, G.; Chiavaioli, F.; Liberi, S.; Zubiato, P.; Del Villar, I.; Angelini, A.; De Wael, K.; Baldini, F.; Moretto, L. M.; Giannetti, A. Nanocoated Fiber Label-Free Biosensing for Perfluorooctanoic Acid Detection by Lossy Mode Resonance. *Results in Optics* **2021**, *5*, 100123. <https://doi.org/10.1016/j.rio.2021.100123>.
- (34) Mann, M. M.; Tang, J. D.; Berger, B. W. Engineering Human Liver Fatty Acid Binding Protein for Detection of Poly- and Perfluoroalkyl Substances. *Biotech & Bioengineering* **2022**, *119* (2), 513–522. <https://doi.org/10.1002/bit.27981>.

- (35) Gavrilaş, S.; Ursachi, C. Ştefan; Perța-Crişan, S.; Munteanu, F.-D. Recent Trends in Biosensors for Environmental Quality Monitoring. *Sensors* **2022**, *22* (4), 1513. <https://doi.org/10.3390/s22041513>.
- (36) Marvin, J. S.; Schreiter, E. R.; Echevarría, I. M.; Looger, L. L. A Genetically Encoded, High-signal-to-noise Maltose Sensor. *Proteins* **2011**, *79* (11), 3025–3036. <https://doi.org/10.1002/prot.23118>.
- (37) Van Hao, P.; Xuan, C. T.; Thanh, P. D.; Thuat, N.-T.; Hai, N. H.; Tuan, M. A. Detection Analysis Limit of Nonlinear Characteristics of DNA Sensors with the Surface Modified by Polypyrrole Nanowires and Gold Nanoparticles. *Journal of Science: Advanced Materials and Devices* **2018**, *3* (2), 129–138. <https://doi.org/10.1016/j.jsamd.2018.04.002>.
- (38) Rodbard, D. Statistical Estimation of the Minimal Detectable Concentration (“Sensitivity”) for Radioligand Assays. *Analytical Biochemistry* **1978**, *90* (1), 1–12. [https://doi.org/10.1016/0003-2697\(78\)90002-7](https://doi.org/10.1016/0003-2697(78)90002-7).
- (39) Oltrogge, L. M.; Wang, Q.; Boxer, S. G. Ground-State Proton Transfer Kinetics in Green Fluorescent Protein. *Biochemistry* **2014**, *53* (37), 5947–5957. <https://doi.org/10.1021/bi500147n>.
- (40) Tsien, R. Y. THE GREEN FLUORESCENT PROTEIN. *Annu. Rev. Biochem.* **1998**, *67* (1), 509–544. <https://doi.org/10.1146/annurev.biochem.67.1.509>.
- (41) Hess, S. T.; Heikal, A. A.; Webb, W. W. Fluorescence Photoconversion Kinetics in Novel Green Fluorescent Protein pH Sensors (pHluorins). *J. Phys. Chem. B* **2004**, *108* (28), 10138–10148. <https://doi.org/10.1021/jp0362077>.
- (42) Doi, N.; Yanagawa, H. Design of Generic Biosensors Based on Green Fluorescent Proteins with Allosteric Sites by Directed Evolution. *FEBS Letters* **1999**, *453* (3), 305–307. [https://doi.org/10.1016/S0014-5793\(99\)00732-2](https://doi.org/10.1016/S0014-5793(99)00732-2).
- (43) Hires, S. A.; Tian, L.; Looger, L. L. Reporting Neural Activity with Genetically Encoded Calcium Indicators. *Brain Cell Bio* **2008**, *36* (1–4), 69–86. <https://doi.org/10.1007/s11068-008-9029-4>.
- (44) Zhou, X.; Herbst-Robinson, K. J.; Zhang, J. Visualizing Dynamic Activities of Signaling Enzymes Using Genetically Encodable FRET-Based Biosensors. In *Methods in Enzymology*; Elsevier, 2012; Vol. 504, pp 317–340. <https://doi.org/10.1016/B978-0-12-391857-4.00016-1>.

- (45) Sakaguchi, R.; Endoh, T.; Yamamoto, S.; Tainaka, K.; Sugimoto, K.; Fujieda, N.; Kiyonaka, S.; Mori, Y.; Morii, T. A Single Circularly Permuted GFP Sensor for Inositol-1,3,4,5-Tetrakisphosphate Based on a Split PH Domain. *Bioorganic & Medicinal Chemistry* **2009**, *17* (20), 7381–7386. <https://doi.org/10.1016/j.bmc.2009.08.015>.
- (46) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. Circular Permutation and Receptor Insertion within Green Fluorescent Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (20), 11241–11246. <https://doi.org/10.1073/pnas.96.20.11241>.
- (47) Cheng, W.; Ng, C. A. Predicting Relative Protein Affinity of Novel Per- and Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach. *Environ. Sci. Technol.* **2018**, *52* (14), 7972–7980. <https://doi.org/10.1021/acs.est.8b01268>.
- (48) Sheng, N.; Li, J.; Liu, H.; Zhang, A.; Dai, J. Interaction of Perfluoroalkyl Acids with Human Liver Fatty Acid-Binding Protein. *Arch Toxicol* **2016**, *90* (1), 217–227. <https://doi.org/10.1007/s00204-014-1391-7>.
- (49) Cai, J.; Lücke, C.; Chen, Z.; Qiao, Y.; Klimtchuk, E.; Hamilton, J. A. Solution Structure and Backbone Dynamics of Human Liver Fatty Acid Binding Protein: Fatty Acid Binding Revisited. *Biophysical Journal* **2012**, *102* (11), 2585–2594. <https://doi.org/10.1016/j.bpj.2012.04.039>.
- (50) Sharma, A.; Sharma, A. Fatty Acid Induced Remodeling within the Human Liver Fatty Acid-Binding Protein. *J. Biol. Chem.* **2011**, *286* (36), 31924–31928. <https://doi.org/10.1074/jbc.M111.270165>.
- (51) Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol.* **2013**, *47* (19), 11293–11301. <https://doi.org/10.1021/es4026722>.
- (52) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-Optimized Mutants of the Green Fluorescent Protein (GFP). *Gene* **1996**, *173* (1), 33–38. [https://doi.org/10.1016/0378-1119\(95\)00685-0](https://doi.org/10.1016/0378-1119(95)00685-0).
- (53) Sheng, N.; Cui, R.; Wang, J.; Guo, Y.; Wang, J.; Dai, J. Cytotoxicity of Novel Fluorinated Alternatives to Long-Chain Perfluoroalkyl Substances to Human Liver Cell Line and Their Binding Capacity to Human Liver Fatty Acid Binding Protein. *Arch Toxicol* **2018**, *92* (1), 359–369. <https://doi.org/10.1007/s00204-017-2055-1>.

- (54) Koebnik, R.; Locher, K. P.; Van Gelder, P. Structure and Function of Bacterial Outer Membrane Proteins: Barrels in a Nutshell. *Mol Microbiol* **2000**, *37* (2), 239–253. <https://doi.org/10.1046/j.1365-2958.2000.01983.x>.
- (55) Su, L.; Jia, W.; Hou, C.; Lei, Y. Microbial Biosensors: A Review. *Biosensors and Bioelectronics* **2011**, *26* (5), 1788–1799. <https://doi.org/10.1016/j.bios.2010.09.005>.
- (56) Lei, Y.; Chen, W.; Mulchandani, A. Microbial Biosensors. *Analytica Chimica Acta* **2006**, *568* (1–2), 200–210. <https://doi.org/10.1016/j.aca.2005.11.065>.
- (57) Death, C.; Bell, C.; Champness, D.; Milne, C.; Reichman, S.; Hagen, T. Per- and Polyfluoroalkyl Substances (PFAS) in Livestock and Game Species: A Review. *Science of The Total Environment* **2021**, *774*, 144795. <https://doi.org/10.1016/j.scitotenv.2020.144795>.
- (58) Jonsson, A.; Dogan, J.; Herne, N.; Abrahmsén, L.; Nygren, P.-A. Engineering of a Femtomolar Affinity Binding Protein to Human Serum Albumin. *Protein engineering, design & selection: PEDS* **2008**, *21* (8), 515–527. <https://doi.org/10.1093/protein/gzn028>.
- (59) Chen, S.; Chen, X.; Su, H.; Guo, M.; Liu, H. Advances in Synthetic-Biology-Based Whole-Cell Biosensors: Principles, Genetic Modules, and Applications in Food Safety. *IJMS* **2023**, *24* (9), 7989. <https://doi.org/10.3390/ijms24097989>.
- (60) Zorzópulos, J.; Long, S. D. D.; Chapman, V.; Kozloff, L. Evidence for a Net-like Organization of Lipopolysaccharide Particles in the Escherichia Coli Outer Membrane. *FEMS microbiology letters* **1989**. <https://doi.org/10.1111/J.1574-6968.1989.TB03545.X>.
- (61) Black, P. N.; DiRusso, C. C. Transmembrane Movement of Exogenous Long-Chain Fatty Acids: Proteins, Enzymes, and Vectorial Esterification. *Microbiol Mol Biol Rev* **2003**, *67* (3), 454–472. <https://doi.org/10.1128/MMBR.67.3.454-472.2003>.
- (62) Black, P. N. The fadL Gene Product of Escherichia Coli Is an Outer Membrane Protein Required for Uptake of Long-Chain Fatty Acids and Involved in Sensitivity to Bacteriophage T2. *J Bacteriol* **1988**, *170* (6), 2850–2854. <https://doi.org/10.1128/jb.170.6.2850-2854.1988>.
- (63) Bae, J. H.; Park, B. G.; Jung, E.; Lee, P.-G.; Kim, B.-G. fadD Deletion and fadL Overexpression in Escherichia Coli Increase Hydroxy Long-Chain Fatty Acid Productivity. *Appl Microbiol Biotechnol* **2014**, *98* (21), 8917–8925. <https://doi.org/10.1007/s00253-014-5974-2>.

- (64) Lepore, B. W.; Indic, M.; Pham, H.; Hearn, E. M.; Patel, D. R.; Van Den Berg, B. Ligand-Gated Diffusion across the Bacterial Outer Membrane. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (25), 10121–10126. <https://doi.org/10.1073/pnas.1018532108>.
- (65) Wagner, S.; Baars, L.; Ytterberg, A. J.; Klussmeier, A.; Wagner, C. S.; Nord, O.; Nygren, P.-Å.; Van Wijk, K. J.; De Gier, J.-W. Consequences of Membrane Protein Overexpression in Escherichia Coli. *Molecular & Cellular Proteomics* **2007**, *6* (9), 1527–1550. <https://doi.org/10.1074/mcp.M600431-MCP200>.
- (66) Park, M. Surface Display Technology for Biosensor Applications: A Review. *Sensors* **2020**, *20* (10), 2775. <https://doi.org/10.3390/s20102775>.
- (67) Charbonneau, M.-È.; Mourez, M. The Escherichia Coli AIDA-I Autotransporter Undergoes Cytoplasmic Glycosylation Independently of Export. *Research in Microbiology* **2008**, *159* (7–8), 537–544. <https://doi.org/10.1016/j.resmic.2008.06.009>.
- (68) Laarmann, S.; Schmidt, M. A. The Escherichia Coli AIDA Autotransporter Adhesin Recognizes an Integral Membrane Glycoprotein as Receptor. *Microbiology* **2003**, *149* (7), 1871–1882. <https://doi.org/10.1099/mic.0.26264-0>.
- (69) Gustavsson, M.; Bäcklund, E.; Larsson, G. Optimisation of Surface Expression Using the AIDA Autotransporter. *Microb Cell Fact* **2011**, *10* (1), 72. <https://doi.org/10.1186/1475-2859-10-72>.
- (70) Chen, T.; Wang, K.; Chi, X.; Zhou, L.; Li, J.; Liu, L.; Zheng, Q.; Wang, Y.; Yu, H.; Gu, Y.; Zhang, J.; Li, S.; Xia, N. Construction of a Bacterial Surface Display System Based on Outer Membrane Protein F. *Microb Cell Fact* **2019**, *18* (1), 70. <https://doi.org/10.1186/s12934-019-1120-2>.
- (71) Lee, M. J.; Kim, P. Recombinant Protein Expression System in Corynebacterium Glutamicum and Its Application. *Front. Microbiol.* **2018**, *9*, 2523. <https://doi.org/10.3389/fmicb.2018.02523>.
- (72) Jarmander, J.; Gustavsson, M.; Do, T.-H.; Samuelson, P.; Larsson, G. A Dual Tag System for Facilitated Detection of Surface Expressed Proteins in Escherichia Coli. *Microb Cell Fact* **2012**, *11* (1), 118. <https://doi.org/10.1186/1475-2859-11-118>.

Chapter 5

Uptake and distribution of perfluoroalkyl substances in industrial hemp (*Cannabis sativa*)

5.1 Introduction

Per- and polyfluoroalkyl substances (PFAS) are considered “forever chemicals” due to their rapid emergence as environmental contaminants and their resistance to biological degradation.¹ This large group of amphiphilic molecules are abundant in the products we use every day, such as foams in fire extinguishers, stain and water-resistant fabrics, and non-stick cooking items.² The risk posed by these man-made chemicals is enhanced by their ability to stick around, accumulating in the water, soil, and even the human body. Long term exposure and buildup of PFAS can lead to adverse health effects like neurotoxicity and certain types of cancers.³ For most people, exposure to PFAS occurs through consumption of contaminated drinking water as well as food exposed to the chemicals during production or packaging. Specifically, soil and groundwater can act as reservoirs and vehicles for PFAS rendering a large amount of land useless for agricultural, recreational, and other purposes.^{4,5}

In the face of PFAS and their inability to biodegrade, phytoextraction or phytoaccumulation could be an efficient option for remediation of heavily contaminated regions. In fact, recent studies have shown the ability of a variety of plant species to take up PFAS including crops like lettuce^{6,7},

tomato^{8,9} and grains¹⁰⁻¹² as well as a multitude of tree, reed, and bushes species¹³⁻¹⁵. While this is deemed an incredible issue in the realm of agricultural crop plants meant for consumption, it opens the door to potential phytoremediation of these chemicals.

Uptake of contaminants into plants is predominately achieved through active transport, and efficiency of PFAS movement past the root systems is mainly governed by sorption of the chemicals to root-solids¹⁶⁻¹⁸. Previous work has elucidated several factors that affect the uptake of PFAS in plants including chain length, and functional group. Hydrophobicity and lipophobicity increase with increasing chain length of PFAS. Therefore, mobility and overall uptake of PFAS decreases as carbon number increases, and long chain PFAAs are more likely to be found in root systems than above ground foliage.^{6,19-21}

Ideal candidates for phytoremediation species are hyperaccumulators with high biomass production that are known to have effective evapotranspiration rates meaning quick absorption of water and nutrients from soil and efficient transport into plant shoots^{22,23}. One such widely studied hyperaccumulator is industrial hemp (*Cannabis sativa*). Hemp is an annual herbaceous plant with short maturation periods, a 4 to 8 month life cycle, and fast growth rate that lead to high biomass yields²⁴. Industrial hemp also develops dense root systems capable of penetrating up to 2 meters into soil increasing surface area and therefore accumulation potential^{25,26}. As a whole, phytoremediation has been proven to be a cost and labor effective, sustainable, and simple method for environmental contamination removal. Hemp, however, has low nutrient requirements and a high tolerance to toxic contaminants allowing it to survive in sites like landfills and industrial mining areas²⁷. While hemp has been extensively proven to remediate soil from contamination of

various heavy metals²⁸⁻³¹, it has also been used to detoxify soils of polycyclic aromatic hydrocarbons, petrol hydrocarbons, and other contaminants of emerging concern^{32,33}. In this and ongoing work, we look to assess the potential of industrial hemp for PFAS phytoextraction by determining concentrations and plant distribution of several PFAS: PFOA, PFOS, and PFHxS.

5.2 Preliminary Studies

5.2.1 Materials and Methods

Experiments were conducted in the greenhouse at the University of Virginia where temperatures reached approximately 35 ± 5 °C during the day and 25 ± 5 °C at night with $70\% \pm 20\%$ relative humidity. Hemp plants were grown for 4 weeks in growing medium consisting of PRO-MIX high porosity mycorrhizae soil. After this period, soil dehydration was achieved through water withholding for 8 hours. Plants were then grown for seven days with daily watering of water containing either high doses (10ppm), low doses (10ppb), or no PFAS; PFAS studied include perfluorooctanoic acid (PFOA), perfluorooctane sulfonic acid (PFOS), and perfluorohexane sulfonic acid (PFHxS) at 10 ppm. On Day 1 plants were given 300mLs of water and 200mLs every day after for seven days; a total of 1.7 liters was given to each plant with a total of 17 μ g and 17mg of PFAS given to low and high dose samples respectively. After seven days of growth, four biological replicates in each treatment group were harvested and roots were washed thoroughly in deionized water before extraction and analysis. Samples were sent to Pace Analytical for tissue matrix extraction and quantitative analysis of 36 perfluorinated compounds using method DOD QSM 5.3.

5.2.2 Results and Discussion

After seven days of growth in soil watered with extremely high concentrations of PFAS in a cumulative manner, the plants still appeared to be unaffected; all sample groups grew and flowered normally as compared to control plants receiving only water as shown in **Figure 5.1**. This attests to the extreme robustness of hemp in the face of environmental toxins and aligns well with their use for soil cleanup of a variety of toxins.

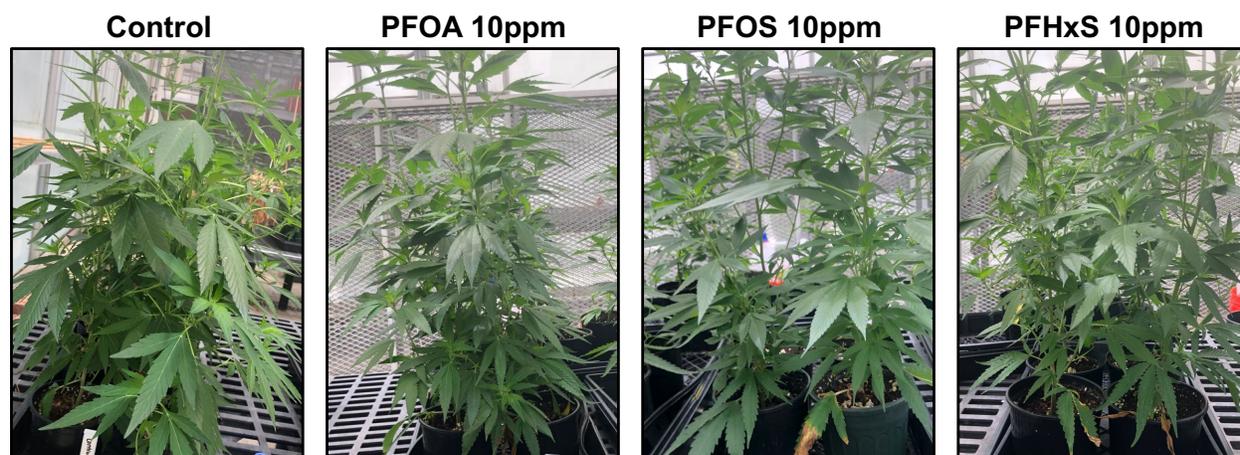


Figure 5.1: Hemp plants grown in high dosed (10ppm) soil after 7 days of growth in comparison to control plants.

Not only the plants appear resistant to PFAS toxicity at extremely high levels, but sample analysis shows promising uptake in both high and low dose plant pools (**Figure 5.2**). Here we see μg per kg wet weight uptake for both PFOA and PFOS in low dose plants and even mg per kg levels of all three studied PFAS at the high dose concentrations. At low concentrations, PFOA was taken up at a higher level than PFOS. This is attributed to the differing functional groups (carboxylate and sulfonate for PFOA and PFOS respectively) as PFCAs are known to have better root uptake compared to PFSAs.^{12,16,18} It is interesting to note that high doses of PFOS accumulate similarly to the shorter chain PFHxS and carboxylic acid containing PFOA. This could be attributed to soil adsorption differences in the three compounds driving bioavailability. Sorption affinity and

reversibility decrease when hydrophobicity of PFAS increases, and the sulfonated chemicals absorb stronger to soil organic matter.^{17,34} Therefore, saturation of the soils in the high dose samples could decrease the bioavailability burden of PFOS through soil adsorption.

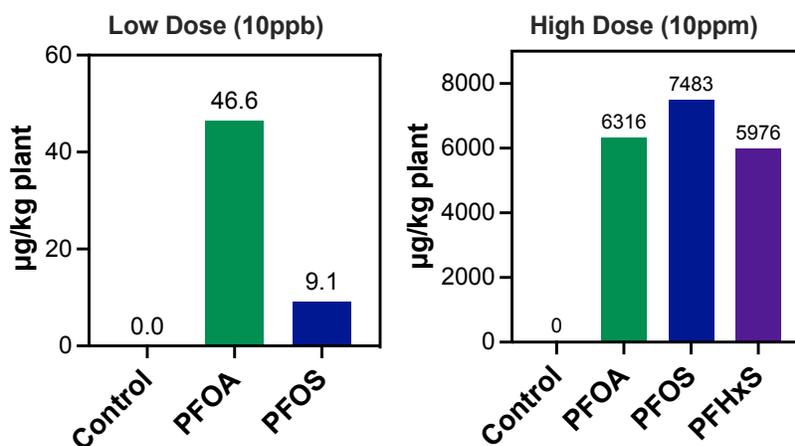


Figure 5.2: PFAS concentration ($\mu\text{g kg}^{-1}$ wet weight) of hemp grown in cumulatively spiked soils at either low (LEFT) or high (RIGHT) concentrations.

5.3 Ongoing work and future directions

5.3.1 Materials and Methods

Experiments were conducted in the greenhouse at the University of Virginia. Hemp plants were grown for seven weeks in pots with a layer PRO-MIX high porosity mycorrhizae soil topped with Turface soil aggregate under ~ 20 hours of light daily. Experiments were conducted in the greenhouse at the University of Virginia where temperatures reached approximately 28 ± 5 °C during the day and 18 ± 5 °C at night with $50\% \pm 10\%$ relative humidity. Flowering was induced through the switching of light cycle time to include 12 hours of light daily. However, it must be noted that along with artificial light, environmental sunlight exposure through windows extended daylight to around 13.5 hours and prolonged the flowering period. Soil dehydration was achieved

through water withholding for 8 hours prior to PFAS addition. Plant soils were spiked with concentrations of 1 mg kg⁻¹ of either perfluorooctanoic acid (PFOA), perfluorooctane sulfonic acid (PFOS), perfluorohexane sulfonic acid (PFHxS), or no PFAS and allowed to grow and flower over the course of 3 weeks. During this period, plants were irrigated with water containing 15-15-15 fertilizer at 200 ppm nitrogen to ensure proper hydration.

The above root sections of plants were harvested after 3 weeks of growth and flowering. Samples were separated into shoots, leaves, and flower sections, and 3 biological replicates for each condition were sent to CEM for solvent extraction and quantitative analysis via LC-MS/MS. Samples of growth media were also sent for extraction and analysis. Upon analysis completion, the bioconcentration factors for the three plant parts will be calculated using the following equation:

$$BCF = \frac{PFAS \text{ concentration in plant part (ng kg}^{-1}\text{)}}{PFAS \text{ concentration in soil (ng kg}^{-1}\text{)}}$$

5.3.2 *Expected outcomes and future work*

In this study, we will look at uptake of several perfluoroalkyl acids at environmentally relevant soil concentrations and determine distribution into various plant parts: leaves, flowers, and vasculature stems. By also focusing on hemp plants in the flowering phase, we hope to elucidate PFAS distribution and accumulation in a crucial and extensive phase of the plant's life cycle. The preliminary studies outlined above show uptake of µg kg⁻¹ levels of PFAS on a wet weight plant mass basis. These studies were performed at PFAS concentrations orders of magnitudes higher than most environmentally contaminated soils, with younger and smaller plant samples. Therefore,

for this study we expect to see ng kg^{-1} levels of total uptake with low $\mu\text{g kg}^{-1}$ concentrations feasible given the 3-week growth period. In alignment with our low concentration studies as well as work on accumulation in other plant species, we expect to observe lower concentrations of PFOS. The main mechanism of PFAS uptake in plants is through vasculature tissues. Therefore, it is intuitive that PFAS with higher water solubilities like PFOA and PFHxS will accumulate at high levels than that of PFOS, especially in the tissues farthest from root systems like flowers and leaves. Hemp flowers contain a high concentration of polyunsaturated fatty acids as well as other molecules like terpenes that contain hydrophobic characteristics.³⁵ It is expected that accumulation trends in flowers will prove to be interesting as PFAS molecules are amphiphilic with extremely hydrophobic carbon-fluorine chains making them a candidate for final PFAS destination in the plants.

There are many factors that affect PFAS uptake into plants species including soil content and the presence of microbial species and their interactions. Studies have shown that PFAS adsorb to a wide variety of matrices. In soil specifically, it has been found that increasing mineral concentrations like Ca^{2+} found in many fertilizers increases PFAS sorption to soils, while soil organic matter and soil organic carbon concentration and make up play a significant role in soil sorption.^{9,17,36} Microbial communities are important for plant growth optimization as they convert organic nutrients like nitrogen and phosphorus into available forms, boost plant immune responses, and promote production of an array of bioactive compounds.³⁷ Future work in the phytoaccumulation of PFAS into industrial hemp should look at the effects of soil and microbiome makeup and the addition supplements on PFAS uptake in order to estimate the practicality of using

hemp for remediation. This work would also help elucidate ways to increase the efficiency of uptake.

Overall, this work looks to assess feasibility of phytoaccumulation in industrial hemp for PFAS remediation. Sequestration of these contaminants from otherwise unusable lands can then be further coupled with degradation techniques that are only economically viable and scalable for low amounts of highly concentrated matrices.

5.4 References

- (1) United States Environmental Protection Agency. EPA's Per- and Polyfluoroalkyl Substances (PFAS) Action Plan EPA 823R18004, 2019.
- (2) Post, G. B.; Cohn, P. D.; Cooper, K. R. Perfluorooctanoic Acid (PFOA), an Emerging Drinking Water Contaminant: A Critical Review of Recent Literature. *Environmental Research* 2012, 116, 93–117. <https://doi.org/10.1016/j.envres.2012.03.007>.
- (3) Sunderland, E. M.; Hu, X. C.; Dassuncao, C.; Tokranov, A. K.; Wagner, C. C.; Allen, J. G. A Review of the Pathways of Human Exposure to Poly- and Perfluoroalkyl Substances (PFASs) and Present Understanding of Health Effects. *J Expo Sci Environ Epidemiol* 2019, 29 (2), 131–147. <https://doi.org/10.1038/s41370-018-0094-1>.
- (4) Costello, M. C. S.; Lee, L. S. Sources, Fate, and Plant Uptake in Agricultural Systems of Per- and Polyfluoroalkyl Substances. *Curr Pollution Rep* 2020. <https://doi.org/10.1007/s40726-020-00168-y>.
- (5) Brusseau, M. L.; Anderson, R. H.; Guo, B. PFAS Concentrations in Soils: Background Levels versus Contaminated Sites. *Science of The Total Environment* 2020, 740, 140017. <https://doi.org/10.1016/j.scitotenv.2020.140017>.
- (6) Felizeter, S.; McLachlan, M. S.; De Voogt, P. Uptake of Perfluorinated Alkyl Acids by Hydroponically Grown Lettuce (*Lactuca Sativa*). *Environ. Sci. Technol.* 2012, 46 (21), 11735–11743. <https://doi.org/10.1021/es302398u>.
- (7) Blaine, A. C.; Rich, C. D.; Sedlacko, E. M.; Hyland, K. C.; Stushnoff, C.; Dickenson, E. R. V.; Higgins, C. P. Perfluoroalkyl Acid Uptake in Lettuce (*Lactuca Sativa*) and Strawberry (*Fragaria Ananassa*) Irrigated with Reclaimed Water. *Environ. Sci. Technol.* 2014, 48 (24), 14361–14368. <https://doi.org/10.1021/es504150h>.
- (8) Blaine, A. C.; Rich, C. D.; Hundal, L. S.; Lau, C.; Mills, M. A.; Harris, K. M.; Higgins, C. P. Uptake of Perfluoroalkyl Acids into Edible Crops via Land Applied Biosolids: Field and Greenhouse Studies. *Environ. Sci. Technol.* 2013, 47 (24), 14062–14069. <https://doi.org/10.1021/es403094q>.
- (9) Blaine, A. C.; Rich, C. D.; Sedlacko, E. M.; Hundal, L. S.; Kumar, K.; Lau, C.; Mills, M. A.; Harris, K. M.; Higgins, C. P. Perfluoroalkyl Acid Distribution in Various Plant Compartments of Edible Crops Grown in Biosolids-Amended Soils. *Environ. Sci. Technol.* 2014, 48 (14), 7858–7865. <https://doi.org/10.1021/es500016s>.

- (10)Stahl, T.; Heyn, J.; Thiele, H.; Hüther, J.; Failing, K.; Georgii, S.; Brunn, H. Carryover of Perfluorooctanoic Acid (PFOA) and Perfluorooctane Sulfonate (PFOS) from Soil to Plants. *Arch Environ Contam Toxicol* 2009, 57 (2), 289–298. <https://doi.org/10.1007/s00244-008-9272-9>.
- (11)Krippner, J.; Falk, S.; Brunn, H.; Georgii, S.; Schubert, S.; Stahl, T. Accumulation Potentials of Perfluoroalkyl Carboxylic Acids (PFCAs) and Perfluoroalkyl Sulfonic Acids (PFSA) in Maize (*Zea Mays*). *J. Agric. Food Chem.* 2015, 63 (14), 3646–3653. <https://doi.org/10.1021/acs.jafc.5b00012>.
- (12)Ghisi, R.; Vamerli, T.; Manzetti, S. Accumulation of Perfluorinated Alkyl Substances (PFAS) in Agricultural Plants: A Review. *Environmental Research* 2019, 169, 326–341. <https://doi.org/10.1016/j.envres.2018.10.023>.
- (13)Gobelius, L.; Lewis, J.; Ahrens, L. Plant Uptake of Per- and Polyfluoroalkyl Substances at a Contaminated Fire Training Facility to Evaluate the Phytoremediation Potential of Various Plant Species. *Environ. Sci. Technol.* 2017, 51 (21), 12602–12610. <https://doi.org/10.1021/acs.est.7b02926>.
- (14)Liu, Z.; Lu, Y.; Song, X.; Jones, K.; Sweetman, A. J.; Johnson, A. C.; Zhang, M.; Lu, X.; Su, C. Multiple Crop Bioaccumulation and Human Exposure of Perfluoroalkyl Substances around a Mega Fluorochemical Industrial Park, China: Implication for Planting Optimization and Food Safety. *Environment International* 2019, 127, 671–684. <https://doi.org/10.1016/j.envint.2019.04.008>.
- (15)Liu, Y.; Hou, X.; Chen, W.; Kong, W.; Wang, D.; Liu, J.; Jiang, G. Occurrences of Perfluoroalkyl and Polyfluoroalkyl Substances in Tree Bark: Interspecies Variability Related to Chain Length. *Science of The Total Environment* 2019, 689, 1388–1395. <https://doi.org/10.1016/j.scitotenv.2019.06.454>.
- (16)Gredelj, A.; Nicoletto, C.; Polesello, S.; Ferrario, C.; Valsecchi, S.; Lava, R.; Barausse, A.; Zanon, F.; Palmeri, L.; Guidolin, L.; Bonato, M. Uptake and Translocation of Perfluoroalkyl Acids (PFAAs) in Hydroponically Grown Red Chicory (*Cichorium Intybus* L.): Growth and Developmental Toxicity, Comparison with Growth in Soil and Bioavailability Implications. *Science of The Total Environment* 2020, 720, 137333. <https://doi.org/10.1016/j.scitotenv.2020.137333>.

- (17) Milinovic, J.; Lacorte, S.; Vidal, M.; Rigol, A. Sorption Behaviour of Perfluoroalkyl Substances in Soils. *Science of The Total Environment* 2015, 511, 63–71. <https://doi.org/10.1016/j.scitotenv.2014.12.017>.
- (18) Felizeter, S.; McLachlan, M. S.; De Voogt, P. Root Uptake and Translocation of Perfluorinated Alkyl Acids by Three Hydroponically Grown Crops. *J. Agric. Food Chem.* 2014, 62 (15), 3334–3342. <https://doi.org/10.1021/jf500674j>.
- (19) Xie, L.-N.; Wang, X.-C.; Dong, X.-J.; Su, L.-Q.; Zhu, H.-J.; Wang, C.; Zhang, D.-P.; Liu, F.-Y.; Hou, S.-S.; Dong, B.; Shan, G.-Q.; Zhang, X.; Zhu, Y. Concentration, Spatial Distribution, and Health Risk Assessment of PFASs in Serum of Teenagers, Tap Water and Soil near a Chinese Fluorochemical Industrial Plant. *Environment International* 2021, 146, 106166. <https://doi.org/10.1016/j.envint.2020.106166>.
- (20) Krippner, J.; Brunn, H.; Falk, S.; Georgii, S.; Schubert, S.; Stahl, T. Effects of Chain Length and pH on the Uptake and Distribution of Perfluoroalkyl Substances in Maize (*Zea Mays*). *Chemosphere* 2014, 94, 85–90. <https://doi.org/10.1016/j.chemosphere.2013.09.018>.
- (21) Zhang, D. Q.; Wang, M.; He, Q.; Niu, X.; Liang, Y. Distribution of Perfluoroalkyl Substances (PFASs) in Aquatic Plant-Based Systems: From Soil Adsorption and Plant Uptake to Effects on Microbial Community. *Environmental Pollution* 2020, 257, 113575. <https://doi.org/10.1016/j.envpol.2019.113575>.
- (22) Rodriguez, L.; Lopez-Bellido, F. J.; Carnicer, A.; Recreo, F.; Tallos, A.; Monteagudo, J. M. Mercury Recovery from Soils by Phytoremediation. In *Environmental Chemistry*; Lichtfouse, E., Schwarzbauer, J., Robert, D., Eds.; Springer-Verlag: Berlin/Heidelberg, 2005; pp 197–204. https://doi.org/10.1007/3-540-26531-7_18.
- (23) Pandey, V. C.; Bajpai, O. Phytoremediation. In *Phytomanagement of Polluted Sites*; Elsevier, 2019; pp 1–49. <https://doi.org/10.1016/B978-0-12-813912-7.00001-6>.
- (24) Clarke, R. C.; Watson, D. P. Cannabis and Natural Cannabis Medicines. In *Marijuana and the Cannabinoids*; ElSohly, M. A., Ed.; Forensic Science And Medicine; Humana Press: Totowa, NJ, 2007; pp 1–15. https://doi.org/10.1007/978-1-59259-947-9_1.
- (25) Fike, J. Industrial Hemp: Renewed Opportunities for an Ancient Crop. *Critical Reviews in Plant Sciences* 2016, 35 (5–6), 406–424. <https://doi.org/10.1080/07352689.2016.1257842>.
- (26) Todde, G.; Carboni, G.; Marras, S.; Caria, M.; Sirca, C. Industrial Hemp (*Cannabis Sativa L.*) for Phytoremediation: Energy and Environmental Life Cycle Assessment of Using

- Contaminated Biomass as an Energy Resource. *Sustainable Energy Technologies and Assessments* 2022, 52, 102081. <https://doi.org/10.1016/j.seta.2022.102081>.
- (27) Mihoc, M.; Pop, G.; Alexa, E.; Radulov, I. Nutritive Quality of Romanian Hemp Varieties (*Cannabis Sativa* L.) with Special Focus on Oil and Metal Contents of Seeds. *Chemistry Central Journal* 2012, 6 (1), 122. <https://doi.org/10.1186/1752-153X-6-122>.
- (28) Pejic, B. M.; Vukcevic, M. M.; Pajic-Lijakovic, I. D.; Lausevic, M. D.; Kostic, M. M. Mathematical Modeling of Heavy Metal Ions (Cd^{2+} , Zn^{2+} and Pb^{2+}) Biosorption by Chemically Modified Short Hemp Fibers. *Chemical Engineering Journal* 2011, 172 (1), 354–360. <https://doi.org/10.1016/j.cej.2011.06.016>.
- (29) Linger, P.; Müssig, J.; Fischer, H.; Kobert, J. Industrial Hemp (*Cannabis Sativa* L.) Growing on Heavy Metal Contaminated Soil: Fibre Quality and Phytoremediation Potential. *Industrial Crops and Products* 2002, 16, 33–42. [https://doi.org/10.1016/S0926-6690\(02\)00005-5](https://doi.org/10.1016/S0926-6690(02)00005-5).
- (30) Tofan, L.; Teodosiu, C.; Paduraru, C.; Wenkert, R. Cobalt (II) Removal from Aqueous Solutions by Natural Hemp Fibers: Batch and Fixed-Bed Column Studies. *Applied Surface Science* 2013, 285, 33–39. <https://doi.org/10.1016/j.apsusc.2013.06.151>.
- (31) Stonehouse, G. C.; McCarron, B. J.; Guignardi, Z. S.; El Mehdawi, A. F.; Lima, L. W.; Fakra, S. C.; Pilon-Smits, E. A. H. Selenium Metabolism in Hemp (*Cannabis Sativa* L.)—Potential for Phytoremediation and Biofortification. *Environ. Sci. Technol.* 2020, 54 (7), 4221–4230. <https://doi.org/10.1021/acs.est.9b07747>.
- (32) Campbell, S.; Paquin, D.; Awaya, J. D.; Li, Q. X. Remediation of Benzo[a]Pyrene and Chrysene-Contaminated Soil with Industrial Hemp (*Cannabis Sativa*). *International Journal of Phytoremediation* 2002, 4 (2), 157–168. <https://doi.org/10.1080/15226510208500080>.
- (33) Liste, H.-H.; Prutz, I. Plant Performance, Dioxygenase-Expressing Rhizosphere Bacteria, and Biodegradation of Weathered Hydrocarbons in Contaminated Soil. *Chemosphere* 2006, 62 (9), 1411–1420. <https://doi.org/10.1016/j.chemosphere.2005.05.018>.
- (34) Campos Pereira, H.; Ullberg, M.; Kleja, D. B.; Gustafsson, J. P.; Ahrens, L. Sorption of Perfluoroalkyl Substances (PFASs) to an Organic Soil Horizon – Effect of Cation Composition and pH. *Chemosphere* 2018, 207, 183–191. <https://doi.org/10.1016/j.chemosphere.2018.05.012>.

- (35)Vuerich, M.; Ferfuaia, C.;Zuliani, F.; Piani, B.; Sepulcri, A.; Baldini, M. Yield and Quality of Essential Oils in Hemp Varieties in Different Environments. *Agronomy* **2019**, *9* (7), 356. <https://doi.org/10.3390/agronomy9070356>
- (36)Higgins, C. P.; Luthy, R. G. Sorption of Perfluorinated Surfactants on Sediments. *Environ. Sci. Technol.* 2006, *40* (23), 7251–7256. <https://doi.org/10.1021/es061000n>.
- (37)Koza, N.; Adedayo, A.; Babalola, O.; Kappo, A. Microorganisms in Plant Growth and Development: Roles in Abiotic Stress Tolerance and Secondary Metabolites Secretion. *Microorganisms* 2022, *10* (8), 1528. <https://doi.org/10.3390/microorganisms10081528>.

Chapter 6

Conclusions

The work presented in this thesis focused on designing platform technologies for the detection and remediation of per- and polyfluorinated alkyl substances (PFAS).

In Chapter 2, we focused on engineering human liver fatty acid binding protein (hLFABP) for PFAS detection. By introducing the solvatochromic fluorophore, acrylodan, within the ligand binding pocket (F50) we could detect several PFAS via blue-shifts in fluorescence emission spectra. The protein-based sensor was capable of detecting PFOA, PFOS, and PFHxS with LODs of 112 ppb, 345 ppb, and 1.09 ppm respectively and demonstrates hLFABP as a useful sensing platform. In Chapter 3, we demonstrate the application of this acrylodan labeled hLFABP biosensor to detect perfluorooctanoic acid (PFOA) in surface water samples taken near Loring Airforce Base. We show this sensor can detect the high levels of PFOA found in the environmental samples quickly and easily without the use of extensive sample pre-treatment or analytical methods.

In Chapter 4, we developed a separate fluorescent biosensor for the rapid detection of PFOA by conjugating circularly permuted green fluorescent protein (cp.GFP) to a split hLFABP construct. This biosensor was able to detect perfluorooctanoic acid PFOA in PBS as well as spiked environmental water samples with LODs of 235 and 335 ppb respectively. Furthermore, the *E. coli*

cells cytosolically expressing the protein-based sensor were demonstrated to quickly detect PFOA, showcasing the feasibility of whole-cell sensing.

Chapter 5 then focuses on remediation of PFAS from soils through phytoaccumulation. Our preliminary studies show industrial hemp (*Cannabis sativa*) is capable of surviving in media containing large amounts of PFAS and hyperaccumulates to $\mu\text{g kg}^{-1}$ and even mg kg^{-1} plant material depending on soil concentration. Ongoing work described in this chapter looks to characterize uptake of several PFAS by studying distribution into various plant parts to further assess phytoaccumulation potential.

Taken together, this dissertation develops a multitude of platform technologies for PFAS detection and remediation using biology-based approaches. By engineering protein-based detection strategies and studying phytoaccumulation, we hope to open the door for further PFAS biosensor design and implementation as well as feasible remediation strategies.

Appendix A

Towards improving sensitivity of hLFABP based PFAS detection systems

A.1 Introduction

Protein based sensors have been proven to be convenient and cost-effective green alternatives for detection of many environmental contaminants including heavy metals, pharmaceuticals, and pesticides.^{1,2} While, to our knowledge, the acrylodan labeled sensor engineered in **Chapter 2** is the first direct fluorescent biosensor designed for PFAS detection, much improvement can be made in order to decrease limits of detection. While this acrylodan labelled sensor was shown in **Chapter 3** to be extremely useful for detecting PFOA at the high levels seen in overly contaminated sites, we believe it is feasible to reach sub-ppb limits through optimization flexibility unique to biosensors. This can be achieved by focusing on two critical aspects of biosensor design outlined in **Chapter 1**: the receptor binding domain (hLFABP) and the signal output (fluorophore). Specifically, by improving binding affinity of the protein receptor to PFAS ligands, the specificity and sensitivity of the engineered biosensor can be increased. In a similar sense, limits of detection can also be decreased via the introduction of more sensitive fluorophores in place of the implemented acrylodan in **Chapter 2**.

By harnessing nature through directed evolution, researchers have been able to alter ligand specificity and affinity for various proteins and even improve binding protein affinity for ligands

by several orders of magnitude (nanomolar to femtomolar).³ Directed evolution is an *in vitro* tool capable of generating proteins with enhanced functions through creation of a genetic library via iterative mutations and subsequent screening of the created new protein variants for enhanced desired properties.

In this work, we designed a library of hLFABP mutants through mutagenesis saturation of several residues near the estimated binding site of PFOA to increase binding affinity. We also developed and utilized an *in vivo* library screen based on the survival of *E. coli* strain AS19 challenged by high concentrations of PFOA. Library mutant sequencing and characterization shows extreme bias towards shortened and aggregated forms of hLFABP. These results, while unexpected, shed light on the relationship of PFOA-protein interactions *in vivo* and offer starting platforms for the further optimization of hLFABP based sensors as well as a survival-based screening methods for directed evolution of a variety of proteins.

A.2 Materials and methods

A.2.1 General Molecular Biology

For DNA maintenance, *E. coli* strain DH5a was used. *E. coli* strain AS19 was used for the library screen, and *E. coli* strain BL21 (DE3) was used for protein expression for the intent of purification. Unless otherwise stated, all molecular biology procedures for PCR amplification, plasmid preparation, cell transformation and subcloning were performed according to standard methods supplied by manufacturers. The gene encoding *E. coli* codon optimized human liver fatty acid binding protein (hLFABP) (NCBI 2168) was previously subcloned into pTrcCDF,pelB. This vector (sequence in supplemental) was previously made in our lab to facilitate golden gate assembly

cloning as it contains no *BsaI* sites. The backbone also contains the *Pectobacterium carotovorum* PelB leader sequence for periplasmic expression of recombinant proteins.⁴ All oligos used for this work are listed in supplemental (**Table A.S1**), and all PCR was done using high fidelity Phusion polymerase with HF buffer (New England Biolabs).

A.2.2 Library Generation and cloning

Saturation mutagenesis was performed using hLFABP-pET28a(+) as a template to introduce NNK codon degeneracy at 7 positions simultaneously. The NNK codons were introduced into hLFABP using degenerate oligonucleotides containing the mixed bases as primers for PCR. Four sets of degenerate oligos were purchased from Integrated DNA Technologies that were used to encode for this randomization through PCR amplification of the gene in four separate fragments. *BsaI* restriction sites were incorporated on the 5' and 3' ends of each degenerate oligo primer. Primers pTrcCDF_LibF and pTrcCDF_LibR were used to amplify the destination vector pTrcCDF with *BsaI* sites on 5' and 3' end. The library fragments as well as amplified pTrcCDF were purified and concentrated using a Monarch DNA and PCR cleanup kit and eluted in sterile deionized water. The library constructs were then “stitched” together and simultaneously ligated into the pTrcCDF vector through Golden Gate Assembly. This construction method is outlined in **Figure A.S1** for clarity. Mutants chosen for further characterization were then PCR amplified out of the pTrcCDF vector using primers Lib_FullF and Lib_FullR and ligated into pET28a(+) using *BamHI* and *XhoI*. Sequences were verified through sanger sequencing (Europhins). Purified DNA for each mutant was then transformed via electroporation into *E. coli* BL21 (DE3) for protein expression. Trigger factor (TF) hLFABP mutant library fusions in pET28a(+) were created utilizing golden gate assembly with *PaqCI* digestion after amplification of trigger factor from TF-silicatein-pet28

(Addgene #205479) using primers TF_FragF and TF_FragR. hLFABP_pET28a(+) mutant vectors were amplified with primers TF_VectorF and TF_VectorR.

A.2.3 Library Transformation and survival screen implementation

After assembly, the library DNA (approximately 250ng) was ethanol precipitated, and gently mixed with 100 μ L of electrocompetent AS19 cells, incubated on ice, and transformed via electroporation. The cells were allowed to recover for 30 minutes in 0.4M NaCl containing LB before spectinomycin (50 μ g/mL) was added and protein expression was induced with 1mM isopropyl β -D-1-thiogalactopyranoside (IPTG). After 2 hours of induction, cells were further incubated for 2 hours after addition of either 250 or 500 parts per million (ppm) PFOA. Cells were then plated on LB agar plates containing spectinomycin (50 μ g/mL). Colonies were selected from the high PFOA (500ppm) containing agar plates and grown overnight for plasmid harvest, extraction, and sequencing via Sanger sequencing (Europhins). All other transformations for protein production were conducted via standard protocols for electroporation.

A.2.4 Protein production

E. coli BL21 cells transformed with hLFABP variants in pET28a(+) were grown to saturation in 10 mL of LB containing kanamycin (50 μ g/mL) overnight at 37°C. The next day, cells were pelleted by centrifugation (3,000 g), and resuspended in fresh media containing kanamycin. The suspension was grown to an OD of 0.6 at 37°C, and protein expression was induced via addition of 1mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Cultures were harvested after an 18-hour growth at 20°C, harvested via centrifugation (10,000 g), and resuspended in lysis buffer (50mM Tris-Cl, 100mM NaCl, 5% v/v glycerol, 0.5mM PMSF protease inhibitor) before lysis via

sonication. The lysis mixture was clarified by centrifugation (12,000 g), and the insoluble fraction was resuspended in 8M urea. Lysates and solubilized pellets were analyzed for recombinant protein via SDS-PAGE analysis. For soluble mutants, the supernatant containing soluble protein was purified by Ni²⁺-affinity chromatography using Chelating Sepharose Fast Flow (Cytiva). The 1 mL column was equilibrated with 50 mM Tris-Cl buffer (pH 8) containing 10 mM imidazole, and protein was separated using a stepwise elution of imidazole up to 500 mM. For insoluble mutants, the pellets were solubilized in 8M urea for 30 minutes. Purification was done similarly as described above with the addition of 8M urea in all buffers. The fractions collected were analyzed by SDS-PAGE to determine purity. Pure fractions were then pooled and dialyzed against phosphate buffered saline (PBS) pH 7.6. For insoluble mutants, dialysis was done in a step-down urea system. Buffers were replaced with PBS pH 7.6 with lower urea concentrations after 4 hours of equilibration until proteins aggregated. Concentration of all protein samples was measured using a Pierce BCA protein assay kit.

A.2.5 Fluorescent Displacement Assays

Characterization of PFOA binding to the hLFABP mutants was conducted via displacement of the fluorophore 1-anilinonaphthalene-8-sulfonic acid (ANS). These assays were performed as described previously in **Chapter 2**. Briefly, hLFABP (1 μ M) in either PBS buffer (pH 7.6) was titrated with ANS. After an equilibration time of 5 minutes, fluorescence intensity at 470 nm were collected after excitation at 400 nm. The maximum fluorescence upon saturation of fluorophore (F_{\max}) as well as the dissociation constant ($K_{d, \text{ANS}}$) were determined by fitting the fluorescence intensity (F) and corresponding concentration of ANS ($[\text{ANS}]$) to a one-site binding model using nonlinear regression:

$$F = F_{\max} * \frac{[ANS]}{(K_{d,ANS}+[ANS])} \quad (1)$$

The binding of PFAS to hLFABP variants was measured by displacement of bound ANS. hLFABP variants (2 μ M) were first equilibrated with ANS (100 μ M or 200 μ M for TF fusions) in excess, and PFAS was then titrated into protein-ANS samples resulting in a final protein concentration of 1 μ M. After a 5-minute incubation, fluorescence emissions spectra were collected between 420-600 nm with excitation at 400 nm. Measurements were corrected using blanks containing the ANS-protein complex as well as protein in buffer only.

The displacement of ANS was characterized as the percent loss in fluorescence and was calculated as the decrease in area under the emission spectra for various PFAS concentrations. The concentration of PFAS necessary to displace half of the bound ANS and inhibit fluorescence by 50% (IC50) was found by fitting displacement data to a sigmoidal dose-response curve where F_{\min} and F_{\max} represent the maximum and minimum ANS displacement by PFAS respectively:

$$\% \text{ Initial Fluorescence} = F_{\min} + \frac{(F_{\max}-F_{\min})}{(1+10^{((\log(IC50)-\log([PFOA]))*HS)})} \quad (2)$$

The inhibition constants (K_i) were then found by relating the value to the half maximal inhibitory constants (IC50) and the dissociation constant for ANS in the absence of PFAS ($K_{d,ANS}$).

$$K_i = \frac{IC50}{\frac{[ANS]_{total}}{K_{d,ANS}} + 1} \quad (3)$$

A.3 Results and discussion

A.3.1 Library Generation

There are many ways to create DNA libraries of varying diversity including completely random error prone PCR, DNA shuffling between genes with varying degrees of homology, and rational based chosen point mutations. As discussed previously, it is known that upon binding of fatty acid ligands as well as PFAS like PFOA and PFOS, hydro and fluorocarbon chains are stabilized by hydrophobic interactions with residues in lining binding pocket.^{5,6} Therefore, seven specific residues believed to be capable of interacting with PFAS ligands were randomized utilizing NNK codons. Currently, there are no solved crystal structures of hLFABP:PFAS interactions available. Therefore, these residues were chosen based on a solved crystal structure of hLFABP with bound oleic acid (PDB 2LKK) as well as residue interaction insight of hLFABP:PFAS binding events from published molecular dynamics studies.^{5,6} These seven residues chosen are shown highlighted in the sequence alignment of the wild type (WT) hLFABP with full-length mutants shown in **Figure A.1**.

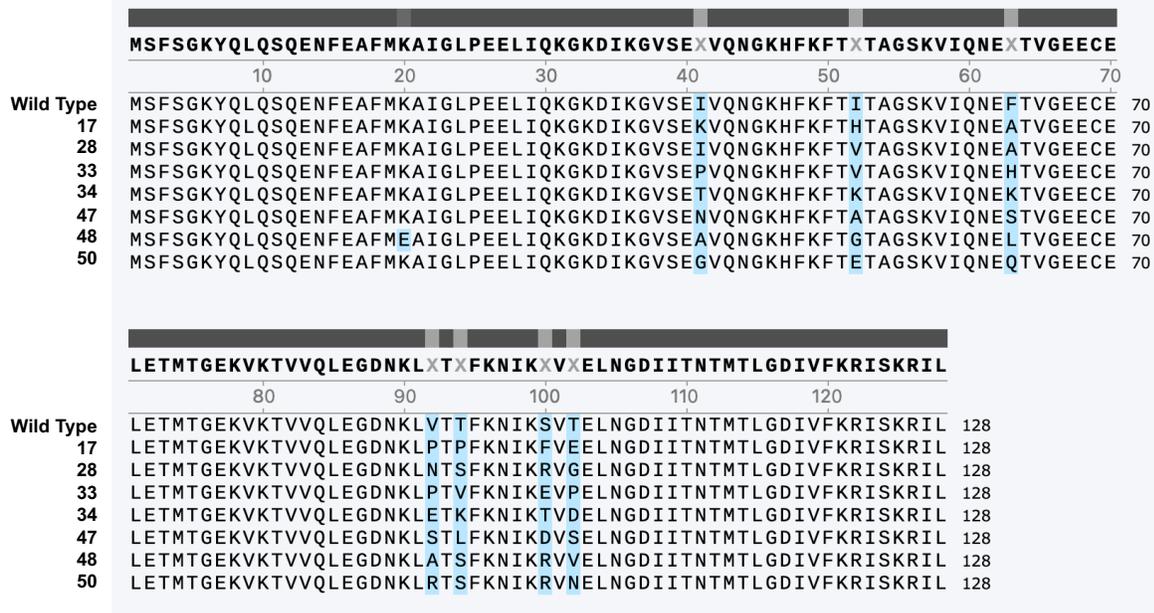


Figure A.1: Sequence alignment of Wild Type hLFABP with full-length mutants from library screening. Blue highlighted residues show variation from Wild Type at residues I41, I52, F63, V92, T93, S100, and T102. Mutant #48 also has variation at residue K21. MSAs generated with SnapGene.

A.3.2 Screen

After generation of a large library of potential hLFABP variants, an efficient screen is necessary to determine high affinity binding proteins for PFOA. Previous studies have demonstrated that the toxicity of PFOA to *E. coli* is largely due to oxidative stress, DNA damage, and membrane disruption.⁷ Along with increased membrane fluidity due to lipid peroxidation of polyunsaturated fatty acids and damage to DNA encoding membrane proteins regulating permeability, PFOA causes cytoplasm outflow in *E. coli* which points to overall disruption of cell membranes.^{8,9} This membrane disruption and toxicity may also be explained by the overall chemical properties of the contaminants. PFOS and PFOA have low polar surface areas, and readily partition into and disrupt phospholipid assemblies like the lipid bilayers composing cell membranes.¹⁰ Therefore, we developed a survival assay based on the notion that increased binding of PFOA to expressed

protein would in turn lower the concentration of toxic PFOA in solution. Thus, there would be increased survival of the bacterial host cells that express proteins with higher binding affinity.

As mentioned before, PFOA cell toxicity has been largely attributed to cell membrane disruption. Because of this, another *E. coli* strain with an already weakened cell membrane was also chosen to test growth inhibition. *E. coli* AS19 is a β -lactam antibiotic sensitive B strain derivative created through chemical mutagenesis for study of *E. coli* uptake of charged antibiotics.¹¹ Through later studies, it has been determined that AS19 hypersensitivity is due to defects in the lipopolysaccharides (LPS) of the outer membrane; specifically, weak attachment of LPS to the peptidoglycan layer.¹² It has also been previously shown that *E. coli* has greater sensitivity to the anionic surfactant sodium dodecyl sulfate (SDS) in high-osmolarity solutions.¹³ Therefore survival screenings were performed through incubation of AS19 expressing hLFABP mutants with PFOA in LB containing 0.4M NaCl. The transformed cells after screening were plated on LB agar and shown in **Figure A.2**.

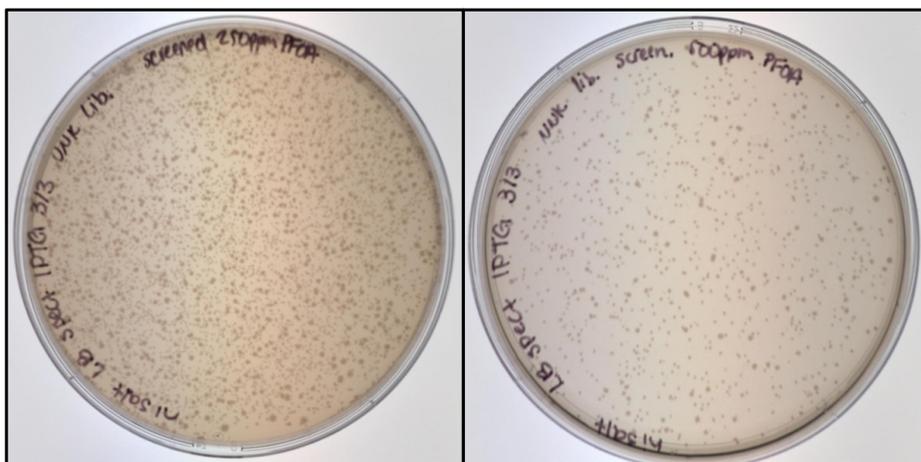


Figure A.2: *E. coli* AS19 library transformation plated on spectinomycin containing LB agar after incubation with 250 ppm (LEFT) or 500 ppm (RIGHT).

A.3.3 Mutant selection and characterization

DNA sequencing showed <25% of the AS19 library colonies harvested, contained hLFABP mutants of full length with ~50% being less than half the length of the wild-type protein. The genes of full length mutants (alignments in **Figure A.1**) were then cloned into pE28a(+) for recombinant overproduction in BL21. Following expression, cells were lysed, and soluble fractions were separated from insoluble pellets via centrifugation. After pellet solubilization in 8M urea, soluble and insoluble fractions were assessed via SDS-PAGE analysis. **Figure A.3** shows that while WT and mutant 17 are evident in the soluble cell lysate fractions, the other six hLFABP variants are only expressed in the insoluble pellet after lysis.

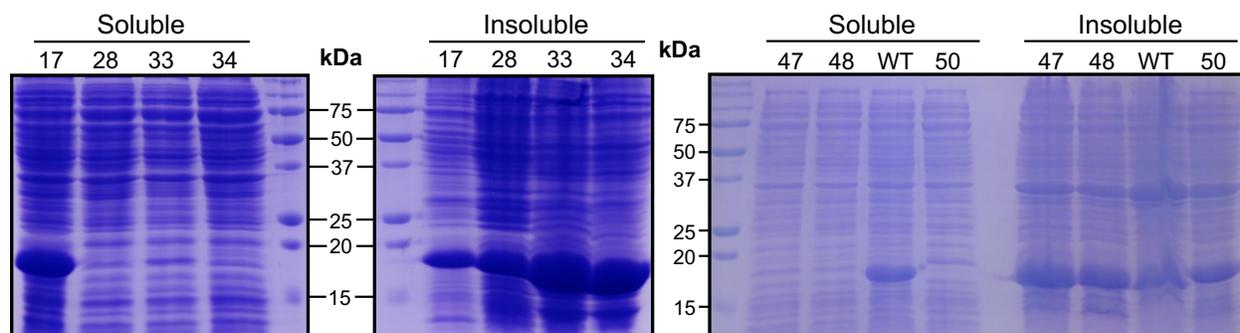


Figure A.3: SDS-PAGE analysis of Wild Type (WT) hLFABP and full-length mutant expression solubility. Bands representing protein variants are depicted at ~19kDa.

This indicates that aside from 17, all full-length hLFABP mutants were expressed as insoluble inclusion bodies. To compare binding of PFOA to hLFABP mutants, assays characterizing the displacement of the fluorophore 1,8-ANS were used to obtain inhibition constants (K_i) as discussed in detail in **Chapter 2**. Briefly, binding of PFOA to our proteins (as quantified by K_i) is correlated to direct binding of the fluorophore ($K_{d,ANS}$) as well as the concentration of PFOA necessary to

displace 50% of the fluorophore from the protein (IC₅₀); this is seen in **Eq. 3**. Characterization of mutant 17 was done directly and compared to wild type hLFABP. **Figure A.4A** shows the displacement of 1,8-ANS by PFOA quantified as the % of initial fluorescence and fit to Equation 2. We see similar inhibition constants in the micromolar range for both wild type hLFABP as well as mutant 17 (**Table A.1**) indicating similar binding capabilities.

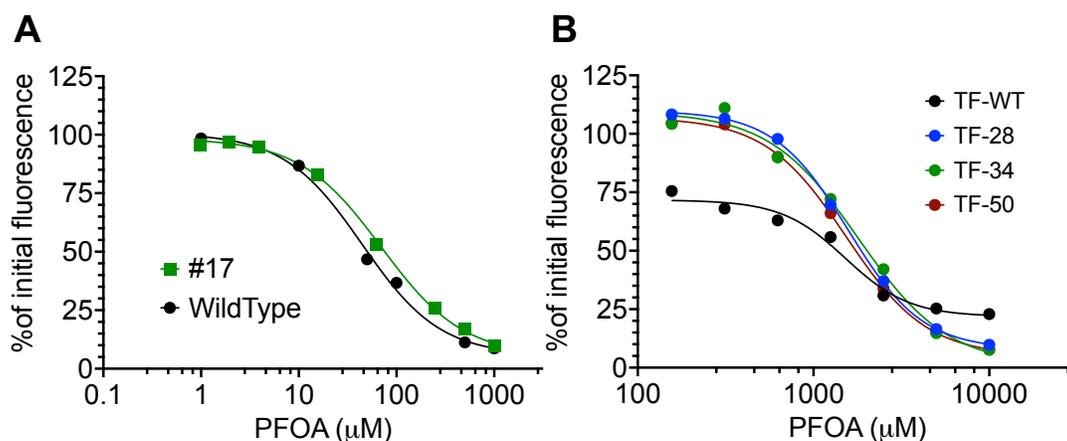


Figure A.4: Binding of PFOA shown through displacement of 1,8-ANS. Fluorescence decrease is quantified as the % of initial fluorescence. Data were fit to a dose-response model.

Several insoluble mutants were characterized after the addition of trigger factor (TF) as a solubility tag to the N-terminal of the proteins. This native *E. coli* chaperone protein is often used as a solubility enhancing tag for insoluble proteins.^{14,15} Direct comparisons of these mutant fusions with the TF-WT hLFABP fusion in **Figure A.4B** and **Table A.1**, shows decreased binding affinity for PFOA in all mutants studied (#28, #34, and #50).

Table A.1: Binding parameters for PFOA-protein interactions. Values were determined using displacement of the fluorophore 1,8-ANS (ANS)

	WT	17	TF-WT	TF-28	TF-34	TF-50
$K_{d,ANS}$ (μM)	4.35	2.30	21.99	53.26	101.89	92.91
IC_{50} (μM)	44.50	70.07	1519.84	1576.97	1804.10	1535.29
$K_{i,PFOA}$ (μM)	3.56	3.08	273.97	548.02	910.50	739.43

The majority of full length hLFABP library mutants were insoluble and showed no sign of improved binding when expressed fused to a solubility tag. While it is possible the addition of a bulky tag like trigger factor (~48kDa) hinders interaction with PFOA, the TF-WT control was utilized as a comparison. Therefore, it is possible that survival against PFOA for cells expressing these mutants is contingent on the aggregation and inclusion body expression. Inclusion bodies are amorphous aggregates of proteins that often arise during recombinant over expression of proteins in bacteria due to an unbalanced equilibrium between translation, protein folding, aggregation, and degradation.¹⁶ While there are many ways to try and mitigate the formation of inclusion bodies including regulating the rate of protein expression through changes in promoter, plasmid, copy number, and codon usage, this requires empirical optimization. Furthermore, these insoluble hLFABP mutants show little to no soluble expression indicating aggregation is not an artifact of flux incompatibilities between translation and folding.¹⁶ While not well understood, inclusion bodies formed upon protein expression have been shown to maintain biological activity after solubilization in chaotropes, purification, and proper refolding.¹⁷ Therefore, to assess function of refolded hLFABP mutants, step-down dialysis was performed on mutants purified in 8M urea. In this process all insoluble mutants re-aggregated during equilibration to 4M urea (data not shown). This further points toward the formation of these aggregates being inherent to protein structure rather than an anomaly is overproduction.

A.4 Future Work and Considerations

There are a multitude of avenues that can be explored based on the work displayed in this chapter. Overall, a library of hLFABP mutants was created via saturation mutagenesis of seven residues near the expected binding site of PFOA. This library was screened through a novel survival-based screen designed to select based on increased binding affinity for the contaminant. However, the mutant sequences show less than a quarter of variants with full-length sequences with no correlation in truncations, and characterized full-length mutants show limited to no solubility. Therefore, it is hypothesized that our library screen shows preference toward aggregated proteins expressed in inclusion bodies that could act as a physical barrier for non-specific binding of PFOA. It is also possible that aggregation or insolubility of mutant variants confers survival through allosteric binding of oligomerized hLFABP variants as PFOA binding is predominantly driven by hydrophobic interactions.^{5,6,18} Therefore, further study should focus on utilizing in cell study of protein-ligand interactions of the library mutants and/or implementation of a library screening that does not rely on cell survival.

Fluorescent protein fusions could be implemented for either of these strategies. Fluorescent proteins (FPs) contain optical properties that are dependent on their surrounding microenvironment and have often been utilized for in vivo imaging, biosensing, as well as cell sorting and library screening through flow cytometry.¹⁹⁻²¹ Specifically, we have discussed in **Chapter 4** that circularly permuted green fluorescent protein (cp.GFP) is capable of transducing PFOA binding when fused to a split hLFABP moiety. Therefore, a strategy to improve binding of PFAS to hLFABP and improve limits of detection for both the acrylodan labeled hLFABP and the genetically encoded

sensors is to screen the created library as a fluorescent protein fusion. Overall, this work lays a platform for directed evolution of hLFABP for PFAS binding as well as a survival-based assay using *E. coli* AS19 that holds promise for future library screening applications.

A.5 Supplemental information

Table A.S1: Oligos used for library generation and cloning

Oligo Name	Sequence (5' → 3')
hLFABP_Lib1F	ggctacggctccatgagcttcagcggcaagtaccag
hLFABP_Lib1R	ggctacggctcagtgcttgcggttctgaacKNNctcgetcagccctt
hLFABP_Lib2F	ggctacggctcgcacttcaaattaccNNKaccgcgggcagc
hLFABP_Lib2R	ggctacggctcatggtttccagctcgcattcctcgccaacgggKNNttcgttttg
hLFABP_Lib3F	ggctacggctcaaccatgaccggcgagaaggtagaaaaccgtg
hLFABP_Lib3R	ggctacggctcgtaatgttcttgaKNNgggKNNcagctgttatcacc
hLFABP_Lib4F	ggctacggctcgattaaaNNKgttNNKgagctgaacggtgacatc
hLFABP_Lib4R	ggcttaggtctccgggtggtggtggtggtg
pTrcCDF_LibF	ggcttaggtctcgaccgctgagcaataactagc
pTrcCDF_LibR	ggctacggctcgtcatggccatcgccggctgggcagcaggag
Lib_FullF	attctgacaggatccatgagcttcagc
Lib_FullR	aactcagctctcagctcgcagaatacgtttg
TF_VectorF	tatcacctgcactagcgggagggcagtgggcg
TF_VectorR	tatcacctgcactatggatccgcgacccatttg
TF_FragF	tatcacctgcactatccatgcaagtttcagttgaacc
TF_FragR	tatcacctgcactaccgctgctggttcacagc

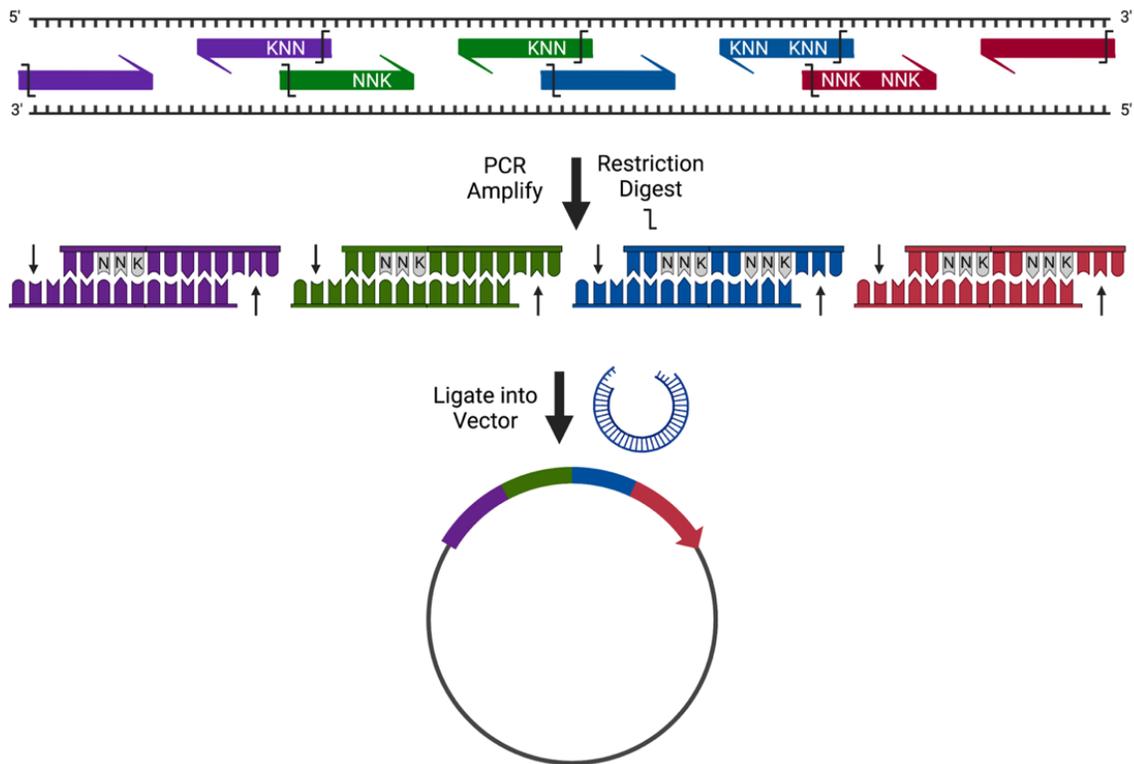


Figure A.S1: Representation of hLFABP gene library containing NNK codons for randomization. Using the wild type hLFABP gene as a template, 4 segments will be amplified out using primers containing NNK nucleotides. Once sections are amplified out using NNK containing primers, they will be “stitched” back together and incorporated into expression vector using compatible overhang sections created using *BsaI* restriction sites. **Figure made using Biorender**

Screen Development

Development of the survival-based screen is explained here. In an ideal survival-based screen you would expect to see (1) almost complete decimation of cells expressing no protein or protein variants with diminished binding capabilities and (2) slight increased survival in cells expressing the wild-type protein to ensure the binding event is promoting cell survival. To develop the screening method used in this chapter, we first looked at overall bacterial survival as well as survival of cells expressing WT hLFABP. To do this, cells were first grown to saturation overnight at 37°C. Cells (*E. coli* BL21 or AS19) were then harvested, washed in deionized water, and resuspended in either LB media or LB media with added sodium chloride (4M final concentration)

and varying concentrations of PFOA. Growth was then monitored over an 8-hour period. **Figure A.S2** shows final OD₆₀₀ data at 8 hours of growth normalized by the growth of the cultures containing no PFOA. In both media conditions (**Figure A.S2A**), AS19 shows clear increased sensitivity to PFOA compared to BL21. This sensitivity was further increased with addition of high concentrations of NaCl; nearly complete inhibition is seen at 1000ppm and 500ppm PFOA for the LB and LB+NaCl media, respectively.

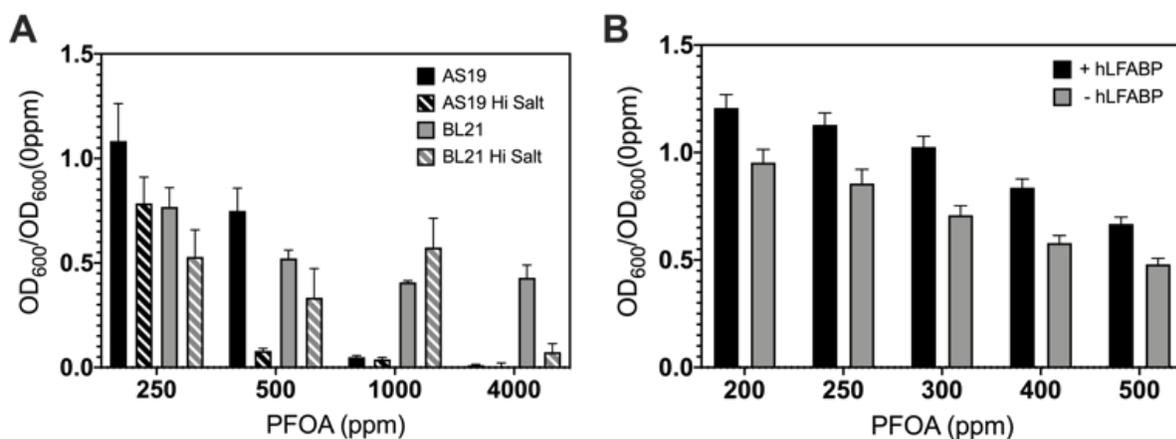


Figure A.S2: Final OD₆₀₀ of *E. coli* after 8 hours of growth in varying PFOA concentrations. Values are normalized by growth in no PFOA. (A) AS19 and BL21 strains grown in either regular or 0.4M NaCl containing LB (B) AS19 grown in 0.4M NaCl containing LB expressing WT hLFABP cytosolically compared to no protein expression.

Baseline survival of AS19 was further assessed upon cytosolic expression of wild type hLFABP. After 4 hours of induction, AS19 transformed with either an hLFABP containing or empty pTrcCDF were screened for PFOA survival in 0.4M NaCl containing LB. **Figure A.S2B** shows increased survival of AS19 expressing the protein at all PFOA concentrations.

Finally, utilization of a post-transformation, plate-based screen was assessed by plating AS19 on high salt LB-agar containing PFOA after post-transformation recovery in high salt LB. **Table 5.S2** shows the clear reduction in colony forming units from “Too Numerous to Count” on the control to complete inhibition by 1000ppm PFOA.

Table A.S2: Colony Forming Units of AS19 *E. coli* plated on high salt LB media containing 0.4M NaCl. Cells were transformed with spectinomycin resistant plasmid and allowed to recover in 0.4M NaCl containing LB for 2 hours prior to plating.

AS19 Post-Transformation Survival on PFOA				
PFOA (ppm)	0	250	500	1000
Colony Forming Units	TNTC	109	33	0

A.6 References

- (1) Rodriguez-Mozaz, S.; Reder, S.; Lopez de Alda, M.; Gauglitz, G.; Barceló, D. Simultaneous Multi-Analyte Determination of Estrone, Isoproturon and Atrazine in Natural Waters by the RIver ANALyser (RIANA), an Optical Immunosensor. *Biosensors and Bioelectronics* 2004, 19 (7), 633–640. [https://doi.org/10.1016/S0956-5663\(03\)00255-0](https://doi.org/10.1016/S0956-5663(03)00255-0).
- (2) Thavarajah, W.; Verosloff, M. S.; Jung, J. K.; Alam, K. K.; Miller, J. D.; Jewett, M. C.; Young, S. L.; Lucks, J. B. A Primer on Emerging Field-Deployable Synthetic Biology Tools for Global Water Quality Monitoring. *npj Clean Water* 2020, 3 (1), 1–10. <https://doi.org/10.1038/s41545-020-0064-8>.
- (3) Jonsson, A.; Dogan, J.; Herne, N.; Abrahmsén, L.; Nygren, P.-A. Engineering of a Femtomolar Affinity Binding Protein to Human Serum Albumin. *Protein engineering, design & selection: PEDS* 2008, 21 (8), 515–527. <https://doi.org/10.1093/protein/gzn028>.
- (4) Singh, P.; Sharma, L.; Kulothungan, S. R.; Adkar, B. V.; Prajapati, R. S.; Ali, P. S. S.; Krishnan, B.; Varadarajan, R. Effect of Signal Peptide on Stability and Folding of Escherichia Coli Thioredoxin. *PLoS ONE* 2013, 8 (5), e63442. <https://doi.org/10.1371/journal.pone.0063442>.
- (5) Sheng, N.; Li, J.; Liu, H.; Zhang, A.; Dai, J. Interaction of Perfluoroalkyl Acids with Human Liver Fatty Acid-Binding Protein. *Arch Toxicol* 2016, 90 (1), 217–227. <https://doi.org/10.1007/s00204-014-1391-7>.
- (6) Cheng, W.; Ng, C. A. Predicting Relative Protein Affinity of Novel Per- and Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach. *Environ. Sci. Technol.* 2018, 52 (14), 7972–7980. <https://doi.org/10.1021/acs.est.8b01268>.
- (7) Liu, G.; Zhang, S.; Yang, K.; Zhu, L.; Lin, D. Toxicity of Perfluorooctane Sulfonate and Perfluorooctanoic Acid to Escherichia Coli: Membrane Disruption, Oxidative Stress, and DNA Damage Induced Cell Inactivation and/or Death. *Environmental Pollution* 2016, 214, 806–815. <https://doi.org/10.1016/j.envpol.2016.04.089>.
- (8) Nobels, I.; Dardenne, F.; Coen, W. D.; Blust, R. Application of a Multiple Endpoint Bacterial Reporter Assay to Evaluate Toxicological Relevant Endpoints of Perfluorinated

- Compounds with Different Functional Groups and Varying Chain Length. *Toxicology in Vitro* 2010, 24 (6), 1768–1774. <https://doi.org/10.1016/j.tiv.2010.07.002>.
- (9) Cabisco, E.; Tamarit, J.; Ros, J. Oxidative Stress in Bacteria and Protein Damage by Reactive Oxygen Species. *Int. Microbiol.* 2000, 3 (1), 3–8.
 - (10) Oldham, E. D.; Xie, W.; Farnoud, A. M.; Fiegel, J.; Lehmler, H.-J. Disruption of Phosphatidylcholine Monolayers and Bilayers by Perfluorobutane Sulfonate. *J Phys Chem B* 2012, 116 (33), 9999–10007. <https://doi.org/10.1021/jp304412p>.
 - (11) Sekiguchi, M.; Iida, S. Mutants of Escherichia Coli Permeable to Actinomycin. *Proc Natl Acad Sci U S A* 1967, 58 (6), 2315–2320. <https://doi.org/10.1073/pnas.58.6.2315>.
 - (12) Zorzópulos, J.; Long, S. D. D.; Chapman, V.; Kozloff, L. Evidence for a Net-like Organization of Lipopolysaccharide Particles in the Escherichia Coli Outer Membrane. *FEMS microbiology letters* 1989. <https://doi.org/10.1111/J.1574-6968.1989.TB03545.X>.
 - (13) Rajagopal, S.; Eis, N.; Bhattacharya, M.; Nickerson, K. W. Membrane-Derived Oligosaccharides (MDOs) Are Essential for Sodium Dodecyl Sulfate Resistance in Escherichia Coli. *FEMS Microbiology Letters* 2003, 223 (1), 25–31. [https://doi.org/10.1016/S0378-1097\(03\)00323-9](https://doi.org/10.1016/S0378-1097(03)00323-9).
 - (14) Miranda, B. N. M.; Fotoran, W. L.; Canduri, F.; Souza, D. H. F.; Wunderlich, G.; Carrilho, E. Heterologous Expression of Homo Sapiens Alpha-Folate Receptors in E. Coli by Fusion with a Trigger Factor for Enhanced Solubilization. *Protein Expression and Purification* 2018, 142, 75–80. <https://doi.org/10.1016/j.pep.2017.10.006>.
 - (15) Sahdev, S.; Khattar, S. K.; Saini, K. S. Production of Active Eukaryotic Proteins through Bacterial Expression Systems: A Review of the Existing Biotechnology Strategies. *Mol Cell Biochem* 2007, 307 (1–2), 249–264. <https://doi.org/10.1007/s11010-007-9603-6>.
 - (16) Bhatwa, A.; Wang, W.; Hassan, Y. I.; Abraham, N.; Li, X.-Z.; Zhou, T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia Coli and Strategies to Address Them for Industrial Applications. *Front. Bioeng. Biotechnol.* 2021, 9, 630551. <https://doi.org/10.3389/fbioe.2021.630551>.
 - (17) Singh, A.; Upadhyay, V.; Upadhyay, A. K.; Singh, S. M.; Panda, A. K. Protein Recovery from Inclusion Bodies of Escherichia Coli Using Mild Solubilization Process. *Microb Cell Fact* 2015, 14 (1), 41. <https://doi.org/10.1186/s12934-015-0222-8>.

- (18) Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; Hamm, P.; Stote, R. H.; Eberhardt, J.; Chebaro, Y.; Dejaegere, A.; Cecchini, M.; Changeux, J.-P.; Bolhuis, P. G.; Vreede, J.; Faccioli, P.; Orioli, S.; Ravasio, R.; Yan, L.; Brito, C.; Wyart, M.; Gkeka, P.; Rivalta, I.; Palermo, G.; McCammon, J. A.; Panecka-Hofman, J.; Wade, R. C.; Di Pizio, A.; Niv, M. Y.; Nussinov, R.; Tsai, C.-J.; Jang, H.; Padhorny, D.; Kozakov, D.; McLeish, T. Allostery in Its Many Disguises: From Theory to Applications. *Structure* 2019, 27 (4), 566–578. <https://doi.org/10.1016/j.str.2019.01.003>.
- (19) Daugherty, P. S.; Iverson, B. L.; Georgiou, G. Flow Cytometric Screening of Cell-Based Libraries. *Journal of Immunological Methods* 2000, 243 (1–2), 211–227. [https://doi.org/10.1016/S0022-1759\(00\)00236-2](https://doi.org/10.1016/S0022-1759(00)00236-2).
- (20) Tsien, R. Y. THE GREEN FLUORESCENT PROTEIN. *Annu. Rev. Biochem.* 1998, 67 (1), 509–544. <https://doi.org/10.1146/annurev.biochem.67.1.509>.
- (21) Doi, N.; Yanagawa, H. Design of Generic Biosensors Based on Green Fluorescent Proteins with Allosteric Sites by Directed Evolution. *FEBS Letters* 1999, 453 (3), 305–307. [https://doi.org/10.1016/S0014-5793\(99\)00732-2](https://doi.org/10.1016/S0014-5793(99)00732-2).

Appendix B

Understanding how transmembrane domains regulate interactions between human BST-2 and the SARS-CoV-2 accessory protein ORF7a

This chapter has been adapted from the following publication: Mann, M. M.; Hsieh, M.-K.; Tang, J. D.; Hart, W. S.; Lazzara, M. J.; Klauda, J. B.; Berger, B. W. “Understanding How Transmembrane Domains Regulate Interactions between Human BST-2 and the SARS-CoV-2 Accessory Protein ORF7a” *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2023**

B.1 Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID, replicates at intracellular membranes. Bone marrow stromal antigen 2 (BST-2; tetherin) is an antiviral response protein that inhibits transport of viral particles after budding within infected cells. RNA viruses such as SARS-CoV-2 use various strategies to disable BST-2, including use of transmembrane ‘accessory’ proteins that interfere with BST-2 oligomerization. ORF7a is a small, transmembrane protein present in SARS-CoV-2 shown previously to alter BST-2 glycosylation and function. In this study, we investigated the structural basis for BST-2 ORF7a interactions, with a particular focus on transmembrane and juxtamembrane interactions. Our results indicate that transmembrane domains play an important role in BST-2 ORF7a interactions and mutations to the transmembrane domain of BST-2 can alter these interactions, particularly single-nucleotide

polymorphisms in BST-2 that result in mutations such as I28S. Using molecular dynamics simulations, we identified specific interfaces and interactions between BST-2 and ORF7a to develop a structural basis for the transmembrane interactions. Differences in glycosylation are observed for BST-2 transmembrane mutants interacting with ORF7a, consistent with the idea that transmembrane domains play a key role in their heterooligomerization. Overall, our results indicate that ORF7a transmembrane domain interactions play a key role along with extracellular and juxtamembrane domains in modulating BST-2 function.

B.2 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified to be the causative agent of a fatal respiratory illness that emerged in the city of Wuhan, China at the end of 2019.¹ Since then, the outbreak of COVID-19, the disease caused by SARS-CoV-2, has infected more than 36.5 million individuals and claimed over one million lives worldwide.² The pathogenesis of SARS-CoV-2 infection in humans has been well documented; SARS-CoV-2 enters the respiratory tract through the binding of the viral structural spike (S) protein to the angiotensin-converting enzyme 2 (ACE2) receptor that is present on the surface of host cells.³⁻⁷ Its genome organization is shared with other betacoronaviruses, and has a 79% and 50% genome sequence identity with SARS-CoV and Middle Eastern Respiratory Syndrome (MERS)-CoV, respectively.⁸ In addition to the functional open reading frames (ORFs) such as the replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M) and nucleocapsid (N), the SARS-CoV-2 genome also contains several putative ORFs encoding accessory and nonstructural proteins interspersed between structural genes that promote SARS-CoV-2 virulence and replication.⁹

One of these accessory proteins, ORF7a, is a type-I transmembrane protein comprised of a N-terminal signal peptide, an ectodomain, a transmembrane region, and a cytoplasmic di-lysine motif (KRKTE) for ER localization.¹⁰ This protein shares 85.2% identity and 95.9% sequence similarity with SARS-CoV ORF7a.¹¹ It was previously shown that SARS-CoV-2 ORF7a has the ability to antagonize the interferon (IFN-I) response by suppressing nuclear translocation of phosphorylated signal transducers and activators of transcription 1 (STAT1) during IFN-I signaling.¹² Similarly, *Cao et al.* reported that upon IFN- α stimulation, ORF7a suppressed STAT2 but not STAT1 phosphorylation in a dose-dependent manner,¹³ suggesting that ORF7a does not effectively block nuclear transport, which may explain the increase in sensitivity of SARS-CoV-2 to IFN pretreatment. This inhibition of STAT2 phosphorylation may be attributed to ORF7a polyubiquitination which subsequently enhances IFN-I antagonism. A recent study has shown that the SARS-CoV-2 ORF7a ectodomain binds to CD14⁺ monocytes in human peripheral blood with high efficiency, leading to decreased antigen-presenting ability and inducing a dramatic expression of proinflammatory cytokines by human immune cells.¹⁴ These cytokines, including IL-1 β , IL-6, IL-8, and TNF are dictated by NF- κ B, and linked to cytokine release syndrome, implicating the positive associations with severe disease outcome.¹⁵ A previous study demonstrated that ORF7a of SARS-CoV-2 can activate NF- κ B function and increases these proinflammatory cytokine expressions.¹⁶ Since IL-1 β is in part responsible for the cytokine storm by SARS-CoV and MERS-CoV¹⁷, ORF7a may play a significant role in the clinical severity of COVID-19, however, further studies are needed to characterize the molecular details and biological functions related to how ORF7a initiates viral pathogenesis.

Bone marrow stromal antigen 2 (BST-2; also known as CD317 or tetherin) is an IFN-inducible gene. As viral infections trigger expression of IFN, BST-2 expression is increased due to IFN responsive regulatory elements in BST-2's promoter region.^{18,19}

BST-2 is known for its ability to block the egress of enveloped viruses from infected cells. In the absence of an antagonist, BST-2 is incorporated into budding viral particles and homodimerizes with other resident BST-2 molecules present on the cell surface,²⁰ thereby 'tethering' the viral particle to the host plasma membrane and restricting cell-to-cell virus spread. BST-2's antiviral function was first identified for HIV-1,²¹ where it was initially thought to restrict HIV-1 virus release. Further studies revealed that BST-2 can also impair the release of various enveloped viruses belonging to *Retroviridae*: HIV type 2 (HIV-2) such as simian immunodeficiency viruses (SIVs),²² equine infectious anemia virus (EIAV),²³ feline immunodeficiency virus (FIV),²⁴ prototype foamy virus (PFV), and Mason-Pfizer monkey virus (MPMV).²⁵

The ability of BST-2 to impair the release of a broad spectrum of viruses would suggest that various viruses have evolved a way to antagonize the cellular restriction factor(s) of BST-2. For instance, the widely studied Vpu accessory protein of HIV-1 is known to physically interact with the transmembrane (TM) domain of BST-2.²⁶ Infection of cells by the virus results in the internalization of BST-2 from the plasma membrane through a clathrin-dependent endocytosis pathway compared to Δ Vpu viruses.²⁷ This suggests that Vpu may affect resupply or surface delivery of BST-2.²⁷ Vpu was also found to reduce total cellular levels of endogenous as well as exogenously expressed BST-2,²⁸ although the mechanism of cell surface BST-2 downregulation is unknown.²⁹ Interestingly, a recent analysis of the TM domains of human and rhesus BST-2

revealed a number of differences including deletions and non-synonymous mutations that affect Vpu sensitivity.³⁰⁻³² Previous studies indicated that a deletion of a GI amino acid motif present in human BST-2, but absent in non-human BST-2 variants,^{26,33} combined with mutation of T45I, resulted in complete loss of Vpu sensitivity.³¹ Similarly, a I48T mutation induced partial sensitivity of rhesus BST-2 to Vpu,³⁰ suggesting that the BST-2 TM domain contains the determinants responsible for the species-specific sensitivity to Vpu.

Biochemical analyses reveal that BST-2 is post-translationally modified by N-linked glycosylation, and forms stable cysteine-linked homodimers.³⁴ In a recent study, BST-2 was found to significantly block human coronavirus 229E progeny virus release,³⁵ indicating that BST-2 is also capable of inhibiting virus budding at intracellular membranes. Similarly, SARS-CoV ORF7a was demonstrated to preferentially bind to unglycosylated BST-2, suggesting that the blocking of glycosylation by ORF7a is directly responsible for the antagonism of BST-2.³⁶ Several studies have implicated BST-2 dimerization as essential for inhibition of HIV-1 release.^{20,37} A recent investigation into the role of BST-2 dimer formation in the release of viral particles showed the mutation of multiple cysteine residues (C53A, C63A, and C91A) prevented both dimer formation and BST-2 function.^{20,37} However, single and double substitutions had no effect, suggesting that promiscuous dimer formation is important for BST-2 anti-viral activity.^{20,37} Thus, the dimerization interface of BST-2 and SARS-CoV-2 may play a unique role in viral release and BST-2 antagonism.

The present study focuses on identifying specific interactions between BST-2 and SARS-CoV-2 ORF7a, with particular emphasis on the TM domain interactions between BST-2 and ORF7a.

Based on dimerization assays in cell membranes using the AraTM and DN-AraTM systems, we demonstrate the sensitivity of homodimerization to varying BST-2 domain truncations as well as point mutations within the TM domain. Similarly, we observed preferential heterodimeric interactions of the TM domains within ORF7a and BST-2 as well as the I28T BST-2. Our results in collaboration with the Lazzara lab at UVA also show that in expression of the full length constructs, SARS-CoV-2 ORF7a disrupts BST-2 glycosylation for all variants studied which has been previously linked to its viral restriction activity.³⁶ Molecular dynamics (MD) simulations completed by Jeff Klauda's group at the University of Maryland College Park revealed the BST-2 and ORF7a protein-protein interface as well as the details of the heterodimerizations. By surveying the heterodimeric configurations, the MD simulations provide molecular insights that the I28S BST-2 variant has enhanced its helix-helix hydrophobic interaction, compared to the WT, to support our observation from *Ara-TM and DN-AraTM* experiments.

B.3 Experimental materials and methods

B.3.1 Subcloning

Unless otherwise stated, all molecular biology procedures were performed according to standard methods supplied by manufacturers. For site-directed mutagenesis, primers were designed using PrimerX, and mutations were introduced using the QuikChange II site-directed mutagenesis kit (Agilent). Mutated sequences were verified by DNA sequencing (GeneWiz). *E. coli* strain DH5 α was used for DNA maintenance.

B.3.2 Plasmids

For bacterial assays, plasmids pAraTMwt and pAraTMDN were generated as previously described from pTrcRSF and pTrc99a scaffolds respectively.^{38,39} The reporter plasmid, pAraGFP was previously derived from plasmid pDS439.³⁸ The DNA sequences coding for the BST-2 domains of interest were cloned into the pAraTMwt (coding for AraC) plasmid, and the ORF7-a transmembrane domains of interest for both COVID and SARS were cloned into the pAraTMDN (coding for the inactive form of AraC, AraC*) plasmid. For mammalian cell expression, pCAGGS-mCherry as a gift from Phil Sharp (Addgene plasmid # 41583 ; <http://n2t.net/addgene:41583> ; RRID:Addgene_41583),⁴⁰ and pcDNA3.1(+) was purchased from Invitrogen. BST-2-FLAG was constructed where the FLAG epitope was fused to the N-terminal of human BST-2 (UniProtKB – Q10589) and amplified using forward primer (5'-AAACTTAAGCTTGGTACCGCCACCATGGATTACAAGGA-3') and reverse primer (5'-CTCTAGACTCGAGCGGCCG-3'). PCR products were digested with KpnI and NotI and cloned into pcDNA3.1(+). For the SARS-CoV-2 ORF7a-myc construct, forward primer (5'-ATCATTTTGGCAAAGAATTCTAGCGCC-3') and reverse primer (5'-GATCGAGATCTGAGTCCGGATTACAGA-3') were used to amplify the sequence. A c-myc tag was fused in-frame to the C-terminal of ORF7a to replace the stop codon at the 3' terminal end. The PCR product was then digested with EcoRI and BglII and cloned into pCAGGS-mCherry.

B.3.3 Ara-TM and DN-AraTM dimerization assays

Ara-TM and DN-AraTM dimerization assays were conducted as previously described.^{38,39} The constructs in the pAraTMwt plasmids and the reporter plasmid (pAraGFPCDF) were co-transformed with or without the pAraTMDN constructs for the Ara-TM homodimerization and

DN-AraTM heterodimerization assays respectively into the AraC-deficient *E. coli* strain SB1676 and streaked onto selective LB plates (100 µg/mL ampicillin, 50 µg/mL kanamycin, and 100 µg/mL spectinomycin). Colonies were picked for each construct and grown in 2mL of selective lysogeny broth (LB) for 12 hours at 37°C and 250 rpm. Cultures were then diluted into selective media with 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for protein induction and grown in a 2.0-mL-deep, 96-well PP plate (PlateOne) for additional 6 hours at 37°C and 250 rpm. We then transferred 200 µL of each culture to a black 96-well, clear bottom plate (Greiner). Absorbance measurements at 600 nm as well as GFP fluorescence emission measurements at 530 nm after excitation at 485 nm were collected using a Synergy Neo2 Hybrid Multi-Mode Microplate Reader (Biotek). The results are reported as the ratio of fluorescence emission at 530 nm to absorbance at 600 nm.

B.3.4 BST-2 and SARS-CoV-2 ORF7a protein co-transfection

HEK293T cells were plated in 6-well plates at a density of 4e5 cells per well and allowed to adhere overnight before transfection. Co-transfection was performed using Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. A total of 5 µg of DNA was used, with 2.5 µg of pcDNA3.1(+)-BST2-FLAG (or empty vector control) and 2.5 µg of pCAGGS-ORF7A-myc (or control). Cells were lysed 48 hours after transfection.

B.3.5 Lysis and Western Blotting

Cells were lysed in standard cell lysis buffer (Invitrogen) supplemented with protease and phosphatase inhibitors (Sigma-Aldrich, P8340, P5726, P0044). Lysates were clarified by centrifugation at 20,000×g for 10 min at 4°C, and total protein concentrations were determined by

micro-bicinchoninic assay (Pierce). Approximately 20 µg of denatured protein was loaded per lane onto 4-12% gradient polyacrylamide gels (Thermo Fisher Scientific). Samples were transferred to 0.2 µm nitrocellulose membranes (Bio-Rad). Membranes were probed with antibodies and imaged using a LICOR Odyssey CLx system or a Bio-Rad ChemiDoc. Antibodies used included FLAG (Cell Signaling Technology, #8146), GAPDH (Santa Cruz, sc-32233), and myc-HRP conjugate (Cell Signaling Technology, #9B11).

B.4 Results and discussion

B.4.1 AraTM and DN-AraTM dimerization assays

The Ara-C based transcriptional reporter assay, AraTM,³⁸ was used to study the effects of BST-2 domains (cytoplasmic, transmembrane, and extracellular) on dimerization. In this assay (**Figure B.1A**), varying domains of BST-2 are fused at the N-terminus to maltose binding protein (MBP) which allows for placement in the bacterial inner membrane. The C-terminus is fused to the AraC transcriptional factor which induces expression of eGFP through activation of the P_{BAD} promoter upon dimerization. eGFP fluorescence is quantified from culture and directly correlates with the extent of dimerization of the inserted BST-2 domains in the chimera.³⁸ Note that BST-2 is a Type II integral membrane protein, whereas ORF7a is a Type I integral membrane protein; constructs were cloned into AraTM assays to reflect the appropriate Type I and Type II configurations for each construct tested.

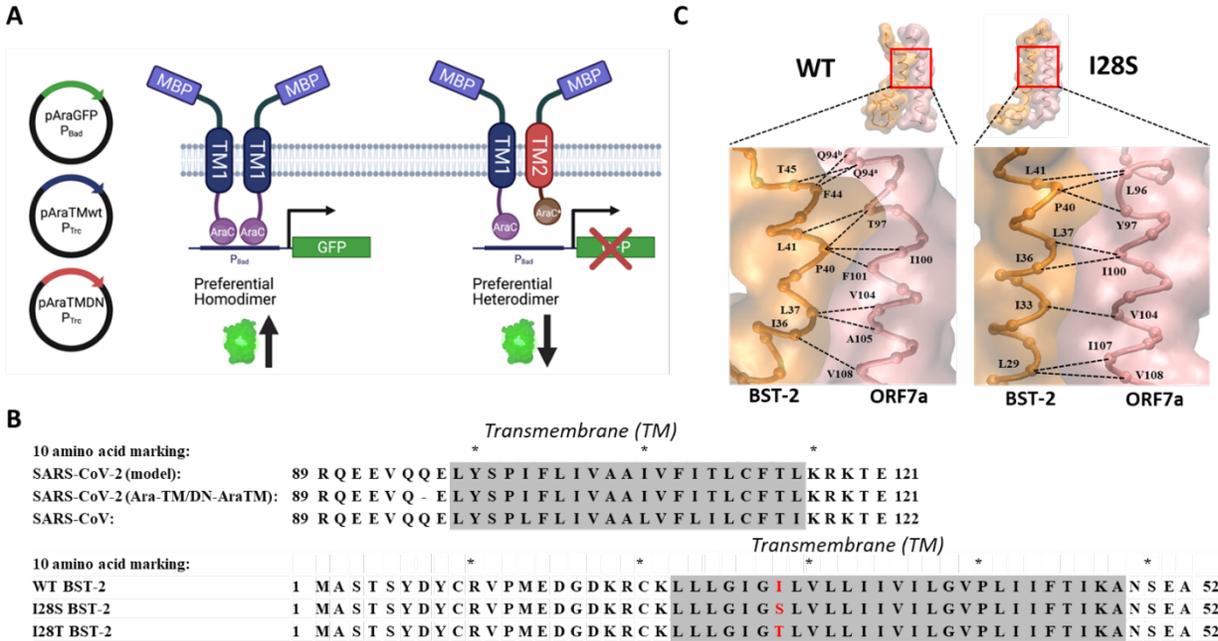


Figure B.1: (A) AraTM and DN-AraTM assay. In the AraTM assay, pAraTMwt expresses the full-length AraC fused to a target sequence under control of the PTrc promoter. This plasmid is co-expressed with the pAraGFP plasmid that contains an eGFP reporter gene under control of the PBAD promoter. In the DN-AraTM assay, addition of the pAraTMDN plasmid expresses the truncated AraC (AraC*) fused to a competitor sequence under control of the PTrc promoter. Co-expression of both constructs with pARAGFP in *E. coli* leads to a distribution of interactions between AraC and AraC* fusions, and the relative affinity for homo- versus heterooligomerization can be independently quantified in terms of eGFP expression; reduced GFP expression is indicative of preferential heterooligomerization, whereas increased GFP expression is indicative of preferential homooligomerization. (B) Truncated amino acid sequences of SARS-Cov-2 ORF7a and BST-2 domains used for the current work. (C) Key residue contacting pairs of both WT and I28S heterodimerization revealed from MD simulations.

The MBP-BST2-AraC constructs containing various forms of the BST-2 domains were transformed into the AraC-deficient *E. coli* strain SB1676 (Figure B.2) containing the engineered *araBAD::GFP* reporter. Interestingly, domains containing the first 52 amino acids, which include the cytosolic domain and transmembrane domain of BST-2, exhibit the highest dimerization in the AraTM assay. Moreover, addition of extracellular domain regions to the TM and juxtamembrane (JM) region of BST-2 (corresponding to amino acids 80-147) do not cause a substantial increase in observed homodimerization, indicating that the primary signal observed in the AraTM assay is

due to the TM and JM interactions. Previous crystal structures show that BST-2 forms a continuous α -helix through the cytoplasmic domain that dimerizes through interactions at the C-terminus, consistent with our observed results.^{41,42} While prior structures of ORF7a and BST-2 indicate that potential regions within the soluble domains are also responsible for oligomerization, our primary focus was on the role of TM and juxtamembrane interactions. Thus, the cytoplasmic tail and TM domain containing BST-2 chimera (BST2-7) was used in the proceeding heterodimer assays.

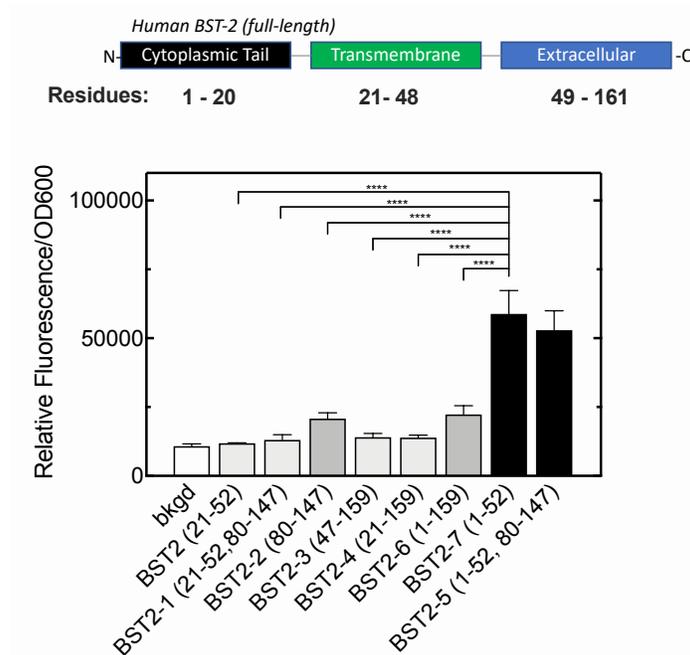


Figure B.2: AraTM homodimerization assay for varying BST-2 domains.

Ratios of GFP fluorescence intensity vs. cell concentration (OD600) for each construct are compared with the background (bkgd). In the background sample, *E. coli* was transformed with the empty pTrcRSF scaffold plasmid that does not express the MBP-AraC-construct. (Upper) Representation of full-length human BST-2 protein and residue numbers used for all BST-2 domain chimeras (Accession Number Q10589). (N=5, **** represents $p < 0.0001$)

As discussed previously, BST-2 exhibits evidence of positive, adaptive selection in response to exposure of viral pathogens, and therefore there are a wide range of identified variations in the TM sequence for human BST-2. To investigate the effects of identified sequence variants available in

dbSNPs⁴³, mutants of the cytoplasmic-TM domain containing BST-2 chimera (BST2-7) were created. To assess self-assembly and heterodimerization interactions of the cytoplasmic-TM domain containing wild type (WT) BST-2 chimera (BST2-7) as well as BST2-7 SNP mutants I28S and I28T with the TM domain of the ORF7a accessory protein from SARS-CoV-2 (C19), a dominant-negative AraC-based transcriptional reporter assay (DN-AraTM) was used in addition to the homodimer AraTM assay.³⁹ The DN-AraTM assay measures the relative affinity of two TM-containing domains to heterodimerize and enables relative comparisons to TM-containing domain homodimerization. This process is illustrated in **Figure B.1A**, where a wild-type AraC chimera containing the TM and JM truncated form of BST-2 (BST2-7) is co-expressed with an inactivated, dominant-negative form (AraC*) containing ORF7a or BST-2. When these chimeras are co-expressed from unique plasmids (pAraTMwt and pAraTMDN), the AraC*-containing ORF7a or BST-2 chimeras act as competitors to BST-2 self-association, thereby decreasing the GFP signal.

As expected, **Figure B.3** shows co-expression of WT BST-2 in both wild-type and dominant-negative AraC constructs causes a significant decrease in GFP signal, confirming the strong self-association seen in the AraTM assay (**Figure B.2**). While BST-2 WT and I28T exhibit this self-association, co-expression of these proteins with the AraC*-containing ORF7a also showed a significant reduction in GFP indicating dimerization competition with the accessory protein. However, the I28S mutant does not show self-association in the DN-AraTM assay. While heterodimer competition with ORF7a cannot be studied in this system upon this loss of self-association of the I28S variant, this lack of homodimerization upon a single nucleotide substitution in the transmembrane domain of BST-2 is interesting to note; the lack of self-association of BST-2 I28S mutant relative to BST-2 WT renders it less susceptible to heterodimerization. Specifically,

while promiscuous in mechanism, dimerization of BST-2 has been previously linked to viral response function.^{20,37} Collectively, these results are consistent with heterodimerization with the transmembrane domain of SARS-CoV-2 ORF7a being significant relative to BST-2 WT and I28T homodimerization.

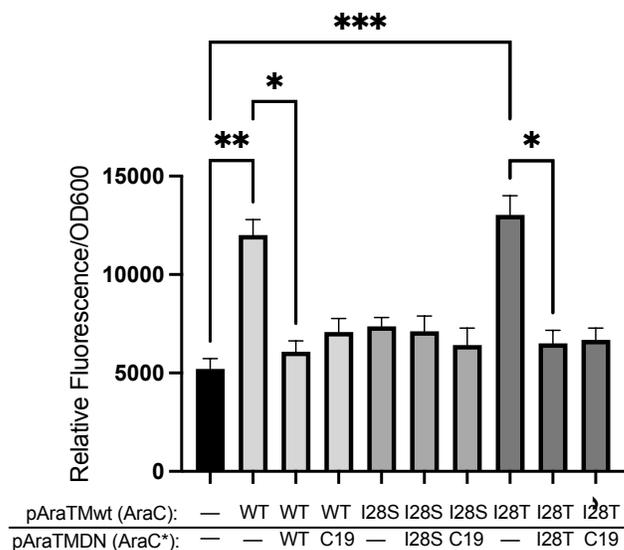


Figure B.3: DN-AraTM dimerization assay for BST-2 SNPs with SARS-CoV-2 ORF7a. Ratios of GFP fluorescence intensity vs. cell concentration (OD_{600}) for each construct are compared with the background (denoted as – in both the AraC and AraC* containing construct table). In the background sample, *E. coli* was transformed with the empty pTrcRSF scaffold for pAraTMwt and pTrc99a scaffold for pAraTMDN plasmids that do not express the MBP-AraC or MBP-AraC* constructs. (N=5, ***, **, and * represent $p < 0.001$, $p < 0.01$, and $p < 0.05$ respectively)

B.4.2 Summary of Molecular Dynamic (MD) simulation work

Further evidence to support the experimental observations of differences in ORF7a-BST-2 interactions was provided through molecular dynamics (MD) simulations. This work from Dr. Klauda’s lab and is discussed in more detail in the Supplemental Information section of this chapter. Briefly, a structural model for the protein-protein interface is presented using multi-scale MD simulations. Course grained MD provided conformations, and all-atom MD refined these

structures. The heterodimeric structures have been classified by multilayer clustering including GMVAE and contact map matching approaches, resulting in 4 and 5 configuration clusters for the WT and I28S heterodimers, respectively. **Figures B.4 & B.5** show contact map results of identified GMVAE clusters for the ORF7a/WT BST-2 and ORF7a/I28S BST-2 dimerization interactions respectively.

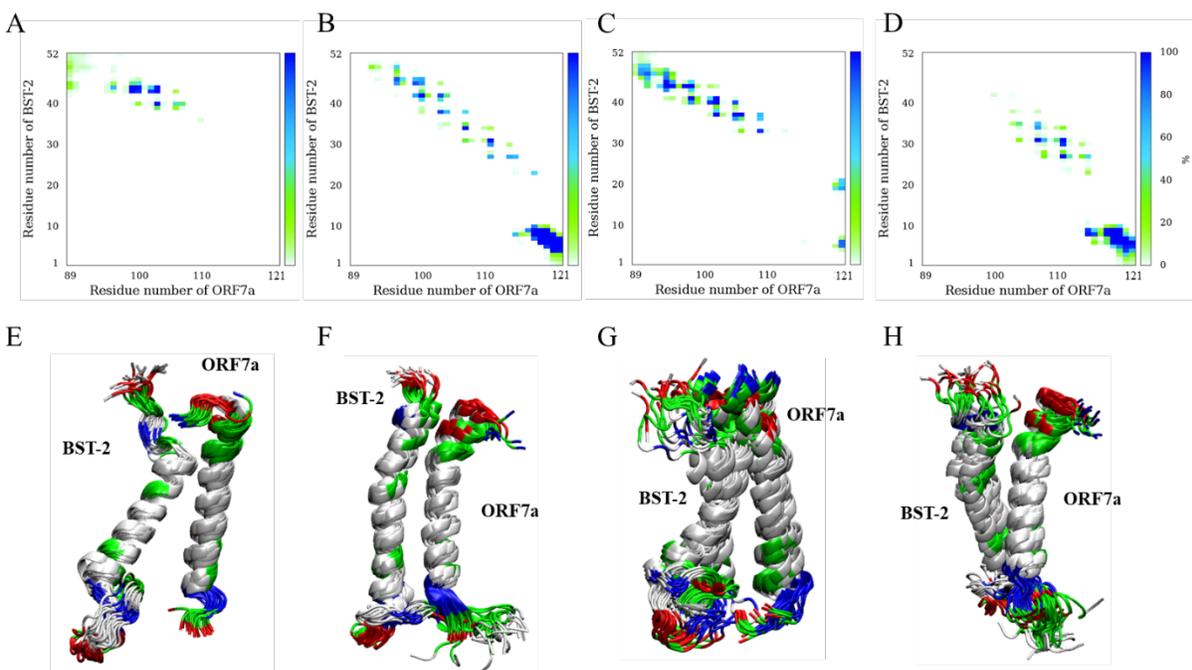


Figure B.4: Results from AA-MD for ORF7a dimerization with WT BST-2.

(A-D): Contact maps of four groupings calculated by using the top 5000 configurations from the center of GMVAE clustering groups. The color bars represent the percentage of contacting distance less than 9 Å. **(E-H):** Snapshots of dimer groups made by the superimposed top 50 configurations from the center of the group. Nonpolar, polar, acidic, and basic residues are colored white, green, red, and blue, respectively.

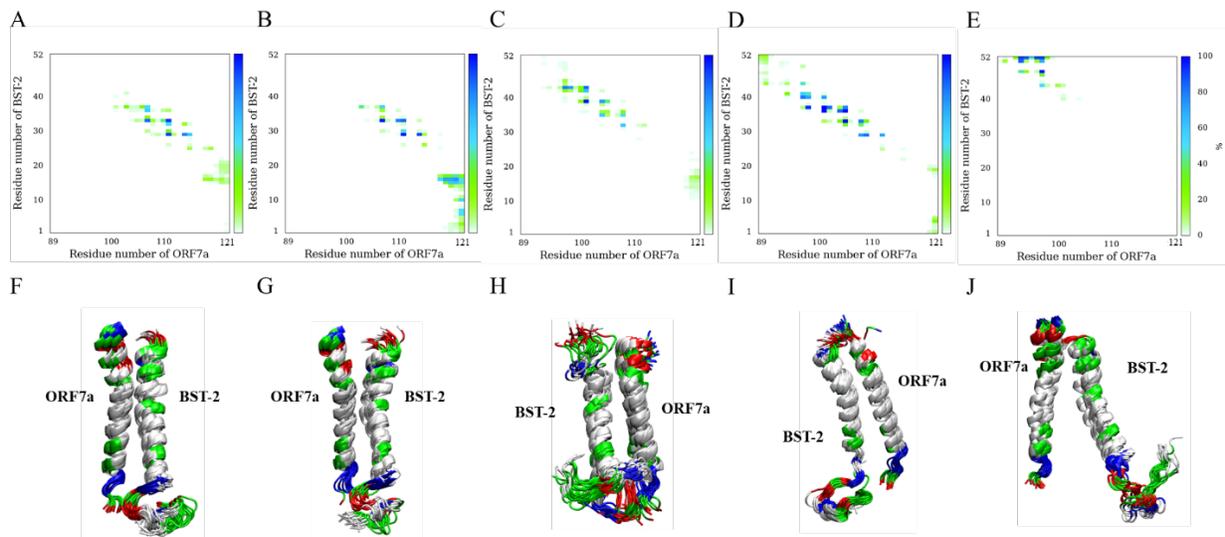


Figure B.5: Results from AA-MD for ORF7a dimerization with mutant I28S BST-2.

(A-E): Contact maps of five groupings calculated by using the top 5000 configurations from the center of GMVAE clustering groups. The color bars represent the percentage of contacting distance less than 9 Å. **(F-J):** Snapshots of dimer groups made by the superimposed top 50 configurations from the center of the group. Nonpolar, polar, acidic, and basic residues are colored white, green, red, and blue, respectively.

Among these clusters, residue contacts between the two peptides show variations in the domains and intensity. Although many strong residue pairs have been observed in the EC and CYTO domains, the tight helical packing structures revealed that coexisting interactions within the CYTO domain (the salt bridges) and at the TM/EC interface (the interacting complex) are likely the primary driving forces for stabilizing TM contacts in both heterodimers. The tight packing structures also revealed a longer TM helix-helix contact region for the I28S dimer compared to the WT. **Figure B.6** shows that the I28S mutation affects the backbone structure by forming an extra sidechain-backbone hydrogen bond (formation probability over 0.99) between S²⁸ and L²⁴ of BST-2 which does not occur in the WT BST-2.

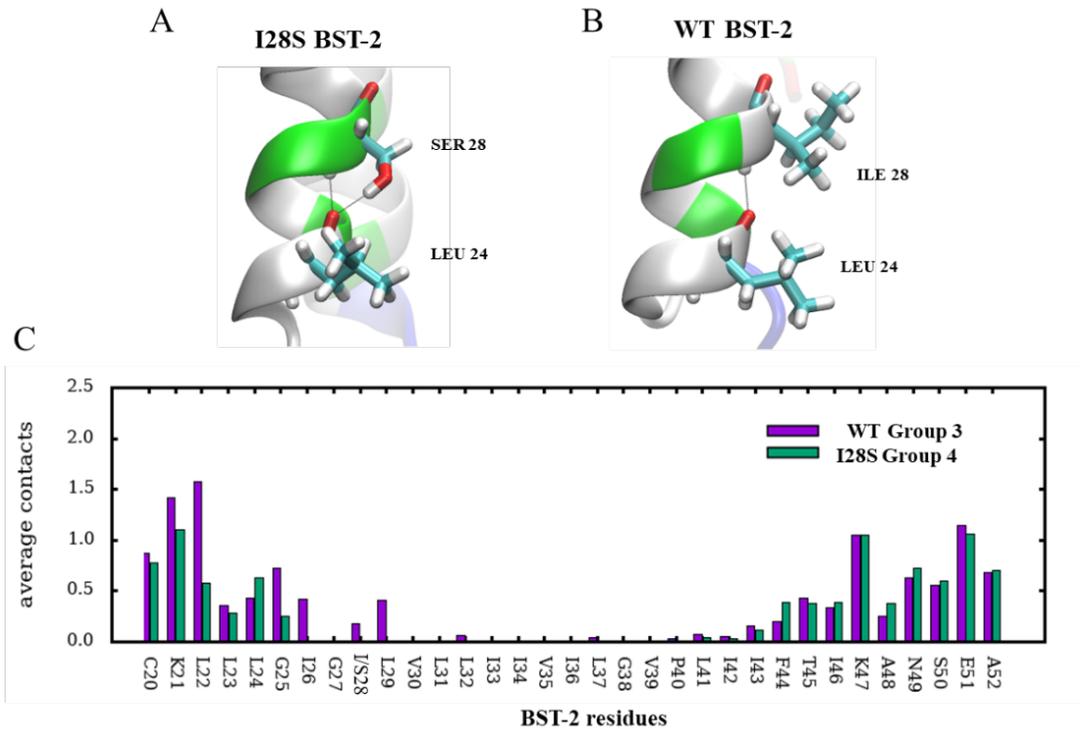


Figure B.6: I28S mutation effects.

The I28S SNP in BST-2 alters the local molecular interactions and water contacts per residues. The configurations demonstrate that **(A)** S²⁸ forms an extra sidechain-backbone hydrogen bond with L²⁴, which does not occur for **(B)** I²⁸. **(C)** The water contacts for each BST-2 residue computed water contacting residues within 3 Å over the WT Group 3 and I28S Group 4 which the individual membership possibility of heterodimeric configuration is more than 0.75 for its own group.

I28S appears to alter the rigidity of the helix structure, which affects the heterodimeric packing. Examination of contacting water within 3 Å of each BST-2 residue shows that contacting water stops at G25 from the cytoplasm in the I28S Group 4 while it goes further to L²⁹ in the WT Group 3. This indicates the extra hydrogen bond reduces the steric restriction which prevents water molecules from penetrating deeper into the hydrophobic region and prompts the tight helix-helix packing. Moreover, the increase in rigidity of the I28S BST-2 helix structure affects the whole domain. By creating better alignment of the backbone atoms, the extended L_{Helix} of I28S BST-2 about 3 residues (averaging over all groups) longer than WT BST-2 can promote hydrophobic

interactions between the two helices. These MD simulations provide molecular insights on how mutations at I28 can influence the dimerization profile with ORF7a and are explained in further detail in the Supplemental Information.

B.4.3 Coexpression of SARS-CoV-2 ORF7a and BST-2 in HEK293T cells

As explained in a previous section, heterodimerization of I28S BST-2 with ORF7a could not be assessed via the bacterial transmembrane DN-AraTM assay. However, the MD simulations presented here have not only elucidated the probability of heterodimerization but also extensively characterized these interaction mechanisms and compared them to wild type BST-2. Therefore, protein-protein interactions were further assessed using the full-length sequences of both BST-2 and its variants as well as SARS-CoV-2 ORF7a in HEK293T cells. Previously, it has been shown that SARS-CoV-2 ORF7a interferes with glycosylation of BST-2.³⁶ Thanks to collaboration with the Lazzara lab, we show that cells cotransfected with SARS-CoV-2 ORF7a also led to a higher abundance of lower molecular weight bands for BST-2 (**Figure B.7**). This increase in unglycosylated BST-2 also occurs in I28S and I28T variants despite the inhibition of self-association for the I28S transmembrane domain seen in our bacterial assays.

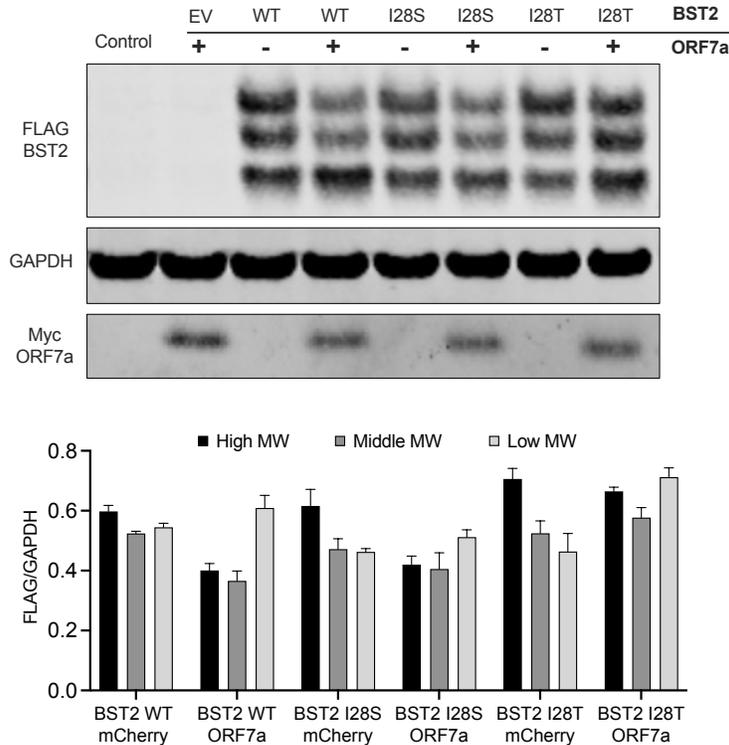


Figure B.7: HEK293T cells expressing BST-2 variants exhibit changes in glycosylation patterns when cotransfected with SARS-CoV-2 ORF7a. **(Top)** Western blot of cells transfected with wild-type as well as I28S and I28T BST-2 exhibit an increase in intensity for lower molecular weight band upon coexpression with ORF7a. **(Bottom)** Densitometry analysis of protein bands normalized to GAPDH internal controls. (n=3) This lower molecular weight band represents unglycosylated BST-2 which suggests ORF7a expression leads to decreased levels of glycosylated BST-2 in all variants.

In this work, we employed BST-2 mutational studies to show that oligomerization is regulated through specific TM residue interactions. As alluded to previously, BST-2 exhibits evidence of positive, adaptive selection in response to viral pathogen exposure. Therefore, there are a wide range of identified single nucleotide polymorphisms (SNPs) within the TM sequence for human BST-2. By analogy to HIV-1 Vpu, it is intriguing to speculate that these mutations may enhance or diminish the heterooligomerization of BST-2 with SARS-CoV-2 ORF7a, thereby enhancing or diminishing the immunomodulatory effect of ORF7a. Based on our bacterial *AraC* transcription-based assay, we observed that preferential heterodimeric interactions of the TM domains within

ORF7a and BST-2 occur but self-association of BST-2 may be affected by an I28 variant. Our AA structural model for protein-protein interface, however, shows heterodimerization of ORF7a with I28S BST-2. Furthermore, it reveals the helix-packing motif involves more residue contacting pairs in the I28S mutant that extend to the bottom TM domain compared to the WT which only exists at the top and middle TM domains. This primary change alters the hydrophobic interaction of the TM association promoting a strong helix-packing in the I28S heterodimeric configuration. We also show that HEK293T cells co-transfected with SARS-CoV-2 ORF7a and BST-2 display different glycosylation patterns. These patterns are consistent with the previously studied SARS-CoV ORF7a – WT BST-2 interactions.³⁶ As heterodimerization of SARS-CoV ORF7a with unglycosylated BST-2 is preferred³⁶, these results are consistent with a mechanism by which SARS-CoV-2 ORF7a could also be responsible for BST-2 antagonism, including both wild-type and BST-2 variants studied.

B.5 Supplemental information

B.5.1 MD Simulation methods

Homology Modeling

A model of the ORF7a (amino acids 89-121 of NCB reference sequence: NC_004718.3; (RQEEVQ^{94a}Q^{94b}ELYSPIFLIVAAIVFITLCFTLKRKTE) (**Figure B.1B**) was generated by *ab initio* modeling with ROBETTA.⁴⁴ This is a chimeric sequence where the N-terminus (EC domain) is from SARS-CoV while the remaining sequence is SARS-CoV-2. The only difference is in the two consecutive glutamines starting at the 94th residue, while only a single Glutamine at position in SARS-CoV-2. This mistake should have minimal influence on our structures obtained below.

Hence, we labeled the two consecutive Glutamine as Q^{94a} and Q^{94b}, for keeping consistency of the sequence number with ORF7a of SARS-Cov-2. A model of the BST-2 (amino acids 1-52 of NCB accession number BAD96844.1) (**Figure B.1B**) was generated using known crystal structures (4P6Z (X-ray; 3Å) (<https://www.rcsb.org/structure/4p6z>), 2LK9 (NMR) (<https://www.rcsb.org/structure/2LK9> BMRB: 17985), and 3MQC (X-ray; 2.8 Å) with MODELLER. The mutated BST-2 (I28S) was also generated by MODLLER after the WT BST-2 model.

Coarse-Grained Molecular Dynamics (CG-MD)

MD simulations were performed on the TM interface between ORF7a and BST-2 and the approaches is summarized in **Figure B.S1**. The wild-type structure of ORF7a was taken to form dimers with the wild-type and the I28S mutant structures of BST-2 to identify the effect of the mutation on dimerization. The initial structure was constructed using the CHARMM-GUI *Martini Bilayer Maker*⁴⁵⁻⁴⁷ with a separation of ~2 nm based on their TM helical axis. Three selected placements (**Figure B.S1**) were set for three replicas in both WT and I28S. The simulation box also includes a 15 Å layer of explicit nonpolarizable water as a buffer on the top and the bottom of the system, a bilayer of 100 lipids of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) per leaflet. MD simulations were conducted using GROMACS 2019.4^{48,49} with the Martini 2.2^{50,51} on Intel® Ivy Bridge E5-2680v2 processors at the Deepthought2 High-Performance Computing (HPC) cluster. First, the systems were energy minimized using 5000 steps of a steepest descent algorithm followed by equilibrium runs. Then, 5 μs production runs (unscaled time) were performed in the isothermal–isobaric ensemble (NPT) ensemble at 310K and 1 bar. The Berendsen method⁵² as applied with semi-isotropic scaling at a pressure of 1 bar for pressure coupling. The

potential-shift⁵³ was used for both electrostatics and van der Waals calculations with cutoff at 11 Å. A 25 fs time step was used, and trajectories were saved every 125ps. The last 2.5 μs of simulations (after equilibrium) was taken to perform distance calculations and create inter-residue contact maps using CHARMM scripts.

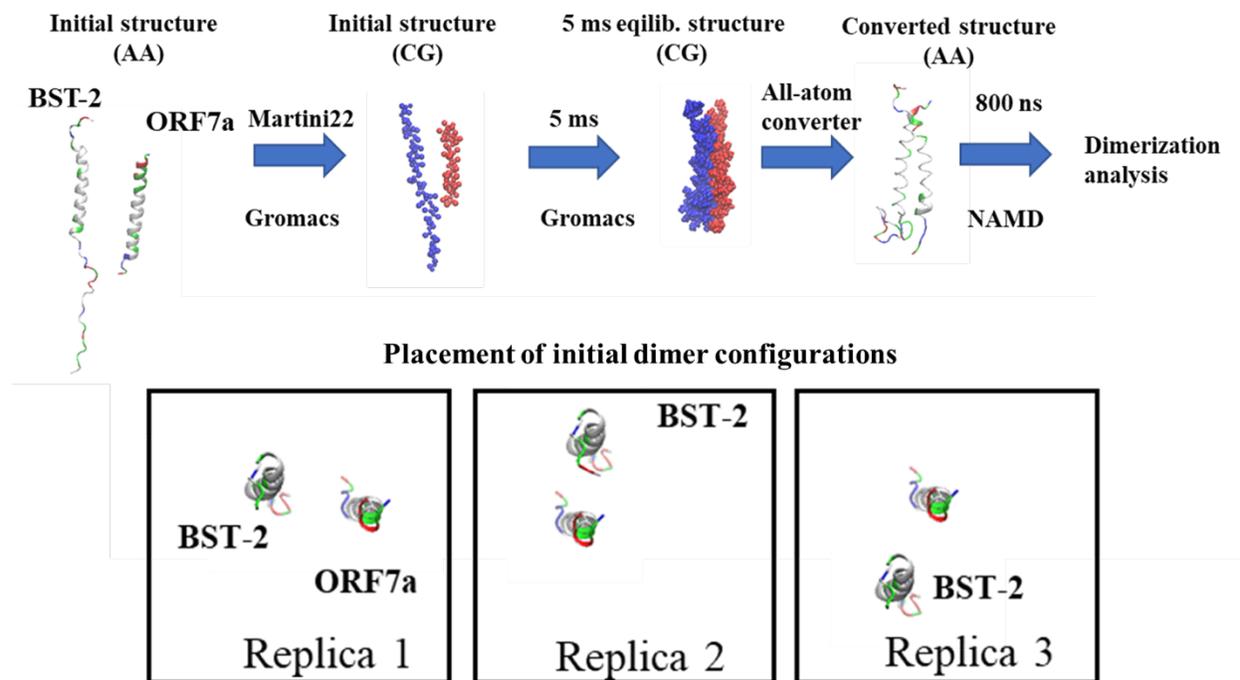


Figure B.S1: The procedure of the MD simulations and the initial placements of dimers.

All-Atom MD (AA-MD)

Each ORF7a/BST-2 dimer structures obtained from the last time step of CG simulations were taken as initial configurations of AA-MD simulations. CG to AA conversion of a whole simulation system including proteins, POPC lipids, and water, was conducted using the CHARMM-GUI *Martini to All-atom Converter*.⁵⁴ The NAMD 2.14^{55,56} simulation package was used with the CHARMM36 (C36) force field for protein and lipids^{57,58} and TIP3P^{59,60} for water, and NPT

ensemble at 310K and 1 atm, performed on NVIDIA[®] Tesla[®] K80 and P100 GPUs at the Bluecrab HPC cluster. The Lennard-Jones potential was used to describe van der Waals interactions, and a force-based switching function in the range of 10-12 Å was chosen.⁶¹ Langevin dynamics maintained the temperature, and the Nosé-Hoover Langevin piston algorithm was applied to maintain the pressure.^{62,63} Hydrogen atoms were constrained by using the RATTLE algorithm.⁶⁴ Simulations were run for 800 ns with a time step was 2 fs.

Clustering for AA configurations

A machine learning method, Gaussian mixture variational autoencoder (GMVAE),⁶⁵ that can perform an unsupervised clustering of protein conformations using the distance matrix of C_α as inputs was applied to identify the main dimer clusters. The GMVAE has shown its capacity for identifying configurations and used to study several proteins in their folding mechanism.⁶⁶ Here, we will demonstrate the potential application of the GMVAE for clustering helix-helix dimerization. The GMVAE clustering by the distance matrix of C_α between two peptide chains could sort the backbone configurations with structural similarity, which would not limit contacting configurations. All three replicas (dropping the first 200 ns each) were combined for GMVAE clustering. To further investigate the contacting of two peptides, we grouped the cluster with the occupancy of contact map to remove the clusters with weak contacts and combined the clusters with high similarity in their contact maps. These identified dimer clusters were further examined for their contacts between ORF7a and BST-2 to construct the protein-protein interface structural model for dimerization to provide the insight of molecular interaction at the atomic level.

B.5.2 Detailed results of MD simulation work

Coarse-grained molecular dynamics simulations

Further evidence to support the experimental observations of differences in ORF7a-BST-2 interactions was provided through coarse-grained molecular dynamics (CG MD) simulations. ORF7a-BST-2 heterodimers were separated initially in a POPC bilayer, and CG MD simulations with MARTINI force field were performed (**Figure B.S1**). Within 100 ns, the helical dimers associated from their initial separation and remained tightly packed for the 5 μ s simulation. The contact maps of the WT heterodimer, calculated by the distances of inter-residue backbone (BB) units between two monomers (**Figure B.S2A-C**) and the end snapshots (**Figure B.S2D-F**), show slight variations between replicas but consistently show interacting transmembrane (TM) and cytoplasmic (CYTO) domains. Similarly, the BB contact maps between the ORF7a and the I28S BST-2 (**Figure B.S3A-C**) as well as the end snapshots (**Figure B.S3D-F**) show a consistent interaction in the TM and CYTO domains. Based on these CG-MD simulations, some slight variations in the dimer interactions exist, but examples of complete dissociation are lacking. Thus, heterodimerization is consistently observed over multiple interactions of specific TM and CYTO interfaces of the ORF7a and BST-2 in both WT and I28S.

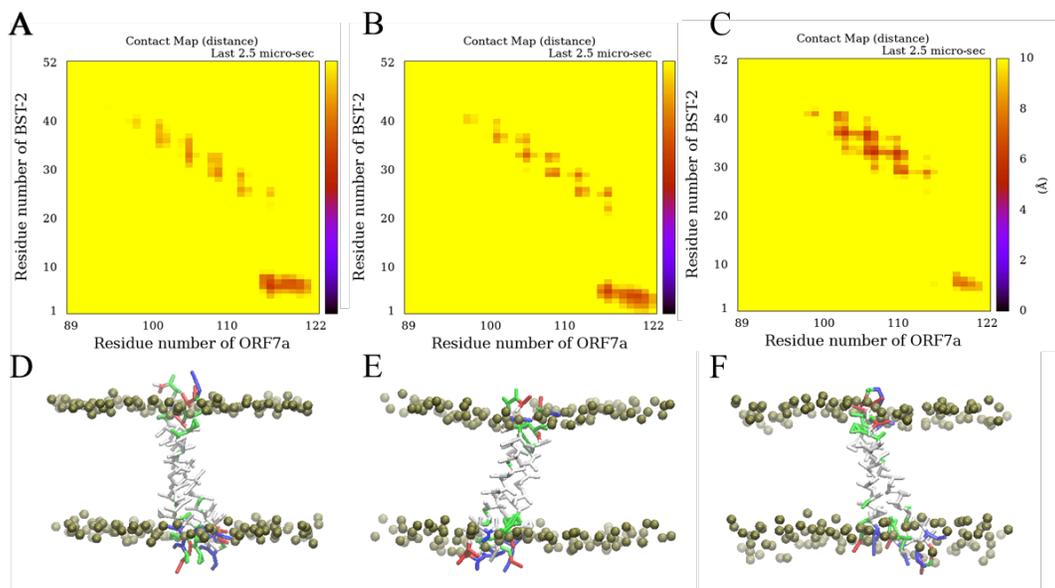


Figure B.S2: Results from CG-MD for the ORF7a dimerization with the WT BST-2. The top three panels are contact maps for three CG runs (A-C). The color bars are in angstrom (\AA), and the maps are averages over the last 2.5 μs of simulation. End snapshots from the 5 μs CG-MD runs are shown in the bottom panels (D-F), with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively. The phosphate atoms are colored gold.

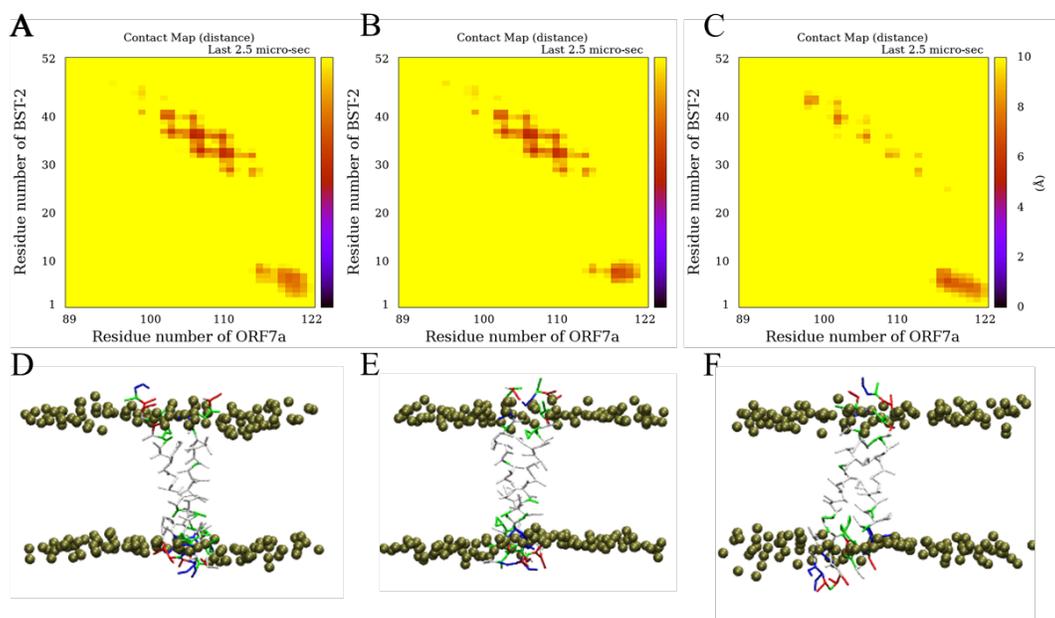


Figure B.S3: Results from CG-MD for the ORF7a dimerization with the I28S mutated BST-2. The top three panels are contact maps for three CG runs (A-C). The color bars are in angstrom (\AA), and the maps are averages over the last 2.5 μs of simulation. End snapshots from the 5 μs CG-MD runs are shown in the bottom panels (D-F), with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively. The phosphate atoms are colored gold.

All-atom MD and Clustering of Dimer Structures

Since the MARTINI FF used in this work is known to be in favor of residue affinity resulting in an overestimation of protein-protein contacts,⁶⁷ MD simulations with an all-atom representation were performed to refine the modeled CG-based structure.⁶⁸ The equilibrated AA trajectory data were prepared by dropping the first 200ns of each replica and then combining three heterodimeric replicas for both WT and the I28S sets. The approach to determining the hyperparameters for heterodimeric clustering is described in previous work.⁶⁶ The chosen hyperparameters for both dimers were optimal based on a cross-validated reconstruction loss, shown in **Table B.S1**.

Table B.S1: Hyperparameters for both dimerization systems.

systems	number of layers	number of neurons	latent dimension	number of clusters	Batch-size	Temperature	Kernel size	Learning rate	number of filters	pooling sizes
WT	3	64	5	9	2500	0.05	[3,3,3]	0.001	[64,64,32]	[2,2,1]
I28S	3	64	5	13	2500	0.05	[3,3,3]	0.001	[64,64,32]	[2,2,1]

During training, the data were split into a train/validation set with a fraction of 0.8 for the training set and 0.2 for the validation set. The total loss, cross-entropy loss, and reconstruction loss decreased during 100 epochs and remained at stable values for both dimerization data (**Figure B.S4-B.S5**). **Figure B.S6** shows the reconstruction loss as a function of the latent space dimension for both systems, indicating higher dimensions result in better reconstruction loss. The latent space dimension of 5 was chosen for minimizing the reconstruction loss of the validation set for both systems. The cluster size was determined using the threshold scheme that has membership possibility higher than 0.95 cut-off. A total of 9 and 13 clusters were identified for the dimerization of ORF7a with the WT and the I28S mutant BST-2, respectively.

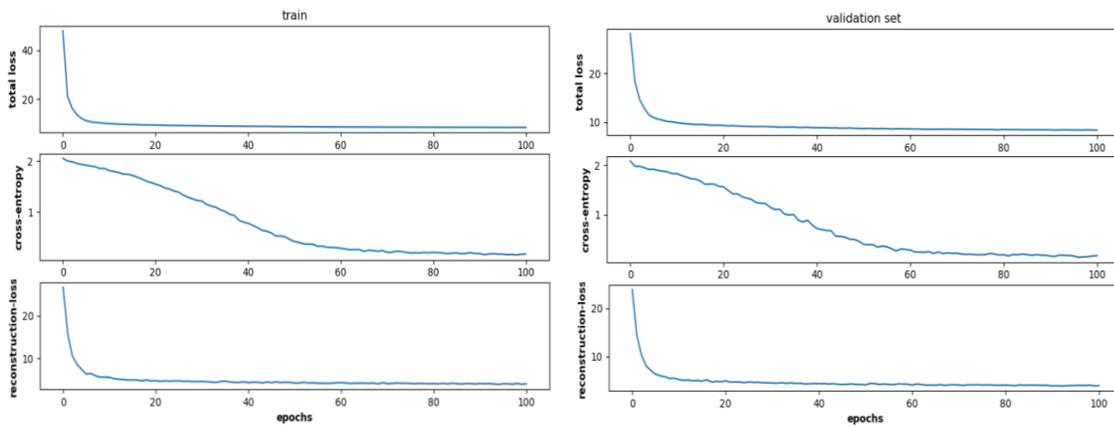


Figure B.S4: Total loss, cross-entropy loss, and reconstruction loss of the train (**LEFT**) and the validation set (**RIGHT**) for the ORF7a/WT BST-2 dimerization.

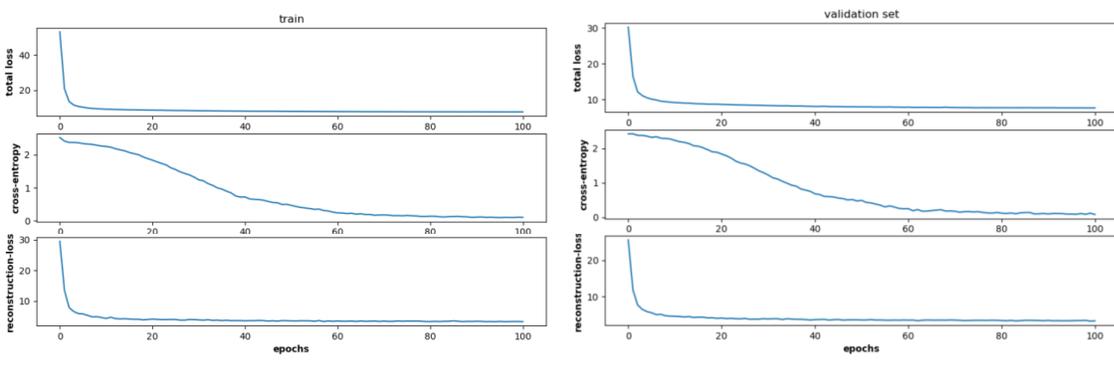


Figure B.S5: Total loss, cross-entropy loss, and reconstruction loss of the train (**LEFT**) and the validation set (**RIGHT**) for the ORF7a/I28S BST-2 dimerization.

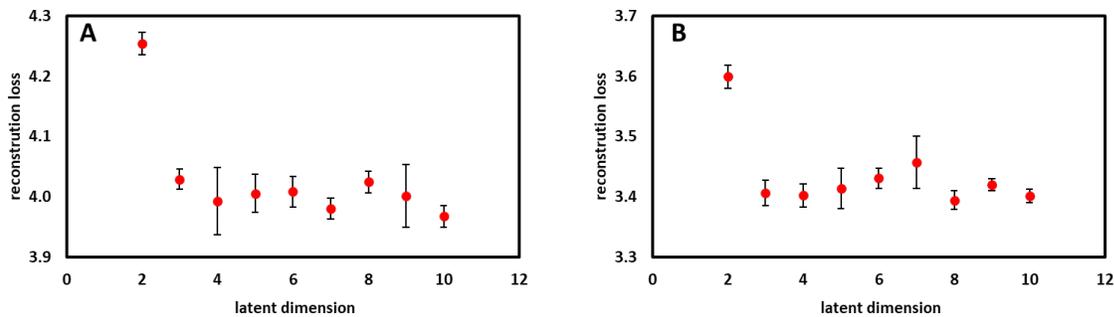


Figure B.S6: The reconstruction loss with the latent dimension for the ORF7a/WT BST-2 (**A**) and the ORF7a/I28S BST-2 (**B**).

The t-distributed stochastic neighborhood embedding (T-SNE)⁶⁹ was applied to visualize the 5D latent space by transforming the five-dimensional embedding into two dimensions (**Figure B.S7**), showing the clusters are highly separated on this landscape with membership possibility over 0.75 for both systems.

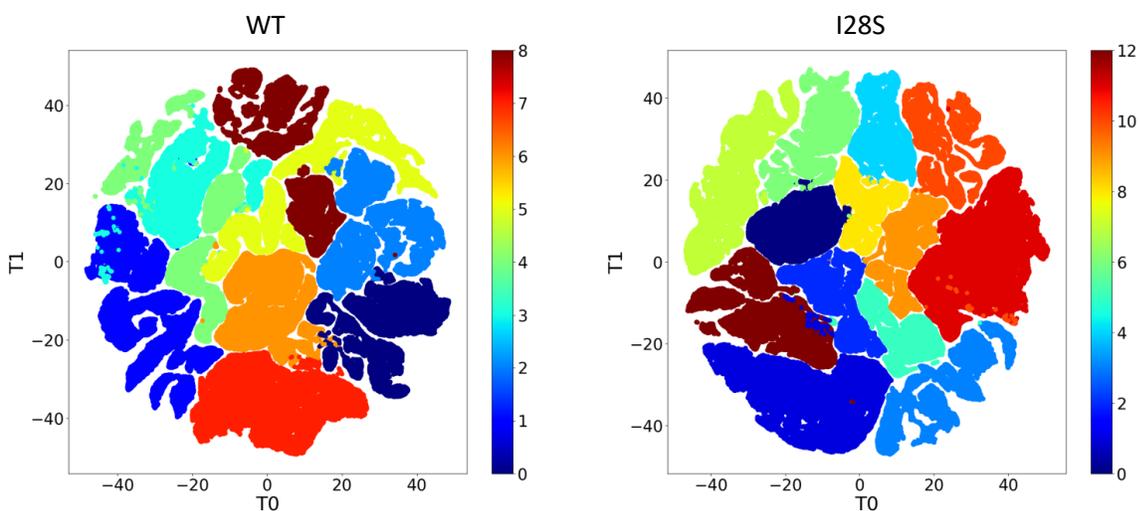


Figure B.S7: TSNE visualization of 5D latent space colored according to their cluster assignment probabilities. Only points with more than 0.75 membership probability are shown.

The distribution of data (**Table B.S2-B.S3**) shows the even fractions among the WT clusters from single or two resource replicas for individual clusters while higher fractions in two clusters than others for the I28S clusters only from single replica for each cluster. 91% and 96% of data points have been identified for the WT and I28S clusters, respectively, with the membership probability more than 0.75.

Table B.S2: Data distribution of the WT cluster form GMVAE clustering.

	% of total data	% of data from replica 1	% of data from replica 2	% of data from replica 3
Cluster 1	9.24	0	51	49
Cluster 2	12.33	100	0	0
Cluster 3	10.42	0	0	100
Cluster 4	8.28	100	0	0
Cluster 5	9.08	61	39	0

Cluster 6	10.65	46	0	54
Cluster 7	10.96	0	100	0
Cluster 8	10.61	0	100	0
Cluster 9	9.5	0	0	100

Table B.S3: Data distribution of the I28S cluster form GMVAE clustering.

	% of total data	% of data from replica 1	% of data from replica 2	% of data from replica 3
Cluster 1	6.71	0	0	100
Cluster 2	10.49	0	100	0
Cluster 3	6.83	100	0	0
Cluster 4	6.88	0	100	0
Cluster 5	5.04	0	0	100
Cluster 6	4.96	0	100	0
Cluster 7	6.91	0	0	100
Cluster 8	7.22	0	0	100
Cluster 9	4.95	0	0	100
Cluster 10	6.58	100	0	0
Cluster 11	6.35	100	0	0
Cluster 12	12.77	100	0	0
Cluster 13	9.89	0	100	0

This GMAVE clustering approach demonstrates its capability to classify the dimer configurations from AA MD dimerization simulations and sort data across different independent runs by using the C_{α} distances between two peptides only. Since the intra-peptide C_{α} distances and other detail sidechain structural information have not been included, the structural variation within a cluster were expected. These GMAVE clusters were taken to further build our structural model for the protein-protein interface, where we are looking for mainly TM helix-helix packing of the heterodimeric system. In examining the contact map of these clusters (**Figure B.S8-B.S9**), several have a similar profile of contacting pairs since this GMAVE was not trained using criteria of contacts between two peptides. Therefore, we further classified the GMVAE clusters into groups by using the contacting occupancy within a whole contact map.

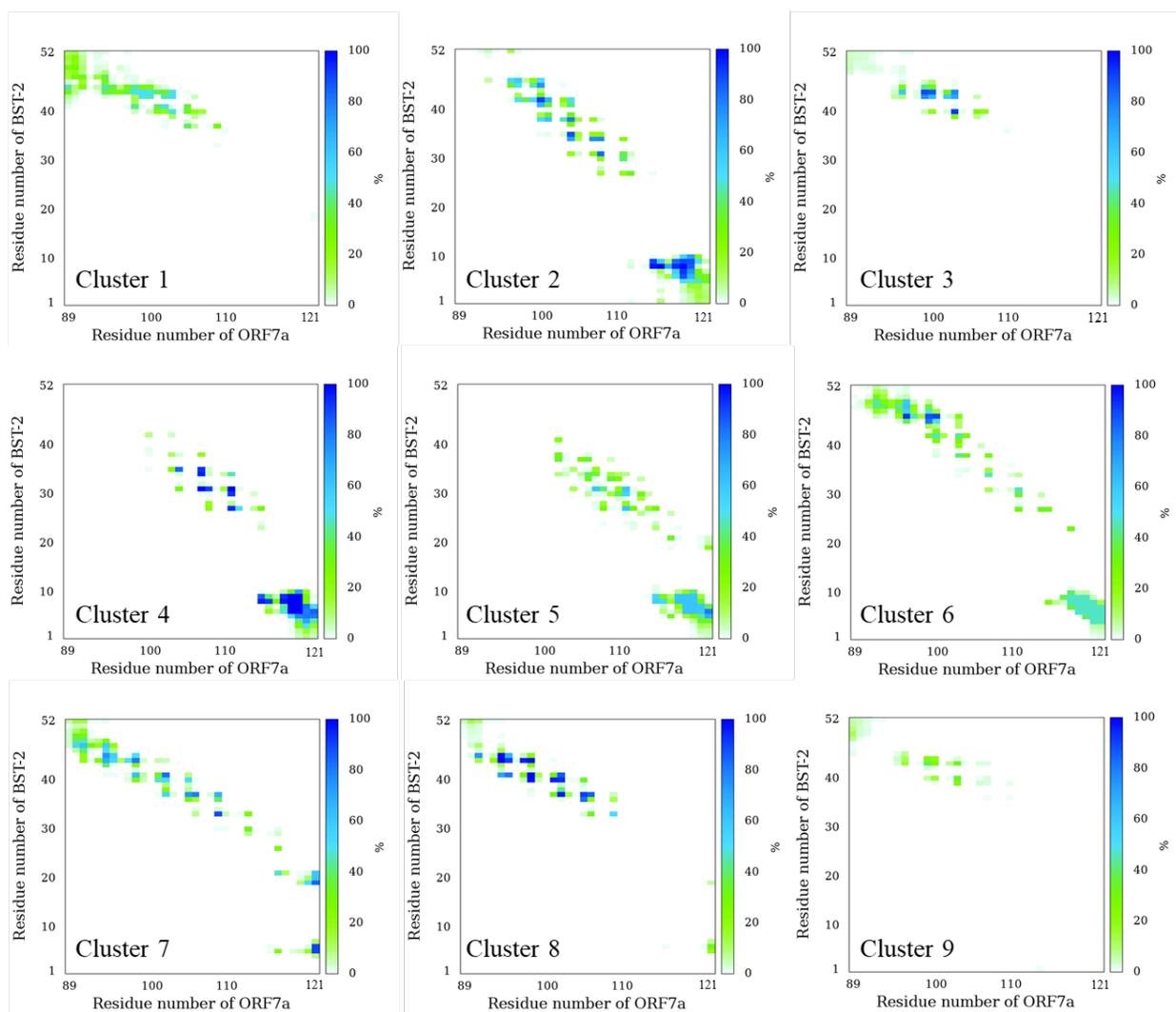


Figure B.S8: Contact maps of the WT clusters from GMVAE with a membership possibility over 0.75. The color bars are in percentage of contacting distance less than 9 Å.

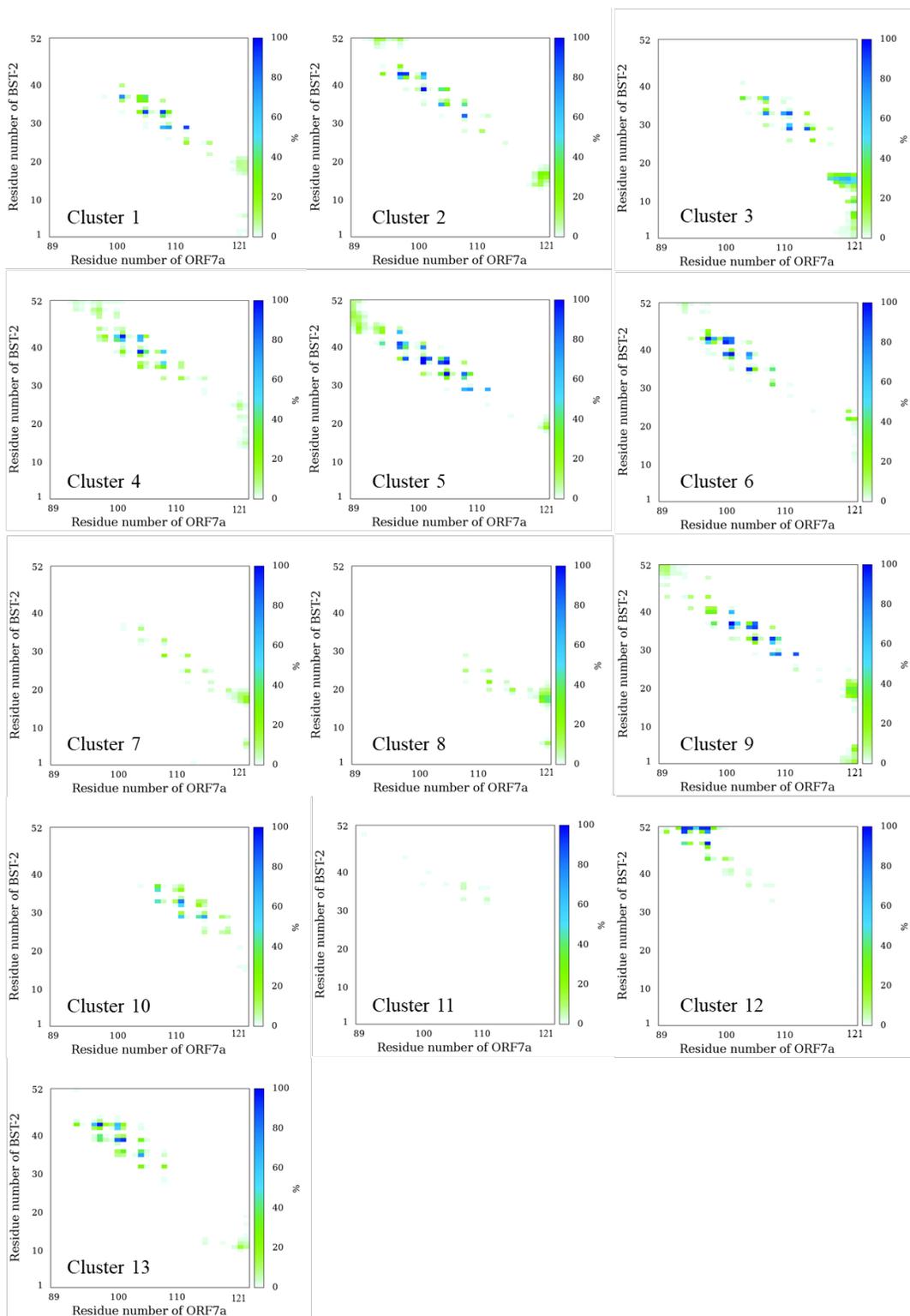


Figure B.S9: Contact maps of the I28S clusters from GMVAE with a membership possibility over 0.75. The color bars are in percentage of contacting distance less than 9 Å.

Contact Map

Figures B.4 and **B.S8** show the contact maps (frequency of contacting distance less than 9 Å) of the GMVAE identified clusters for the ORF7a/WT BST-2 dimerization, showing the contacting residue pairs occupy 0.21 ~ 2.3 % of the whole contact map (**Table B.S4**) for the identified GMVAE clusters. The contacting occupancies for each division were listed as well. These clusters were further grouped by the distribution of their regional contact occupancy. Specifically, the regional contact domains on BST-2 were defined by ranges of residue positions: EC domain (residue 49-52), top of the TM domain (residue 40-48), middle of the TM domain (residue 31-39), bottom of the TM domain (residue 22-30) and CYTO domain (residue 1-21).

Table B.S4: Occupancy percentage of contact map for the WT clusters.

	overall	EC	TM_top	TM_mid	TM_bot	IC
Cluster 1	1.058	0.11	0.892	0.056	0	0.001
Cluster 2	2.364	0.012	0.556	0.547	0.075	1.175
Cluster 3	0.573	0.026	0.531	0.017	0	0
Cluster 4	2.044	0	0.007	0.416	0.186	1.435
Cluster 5	1.488	0	0.013	0.339	0.23	0.905
Cluster 6	1.795	0.171	0.76	0.137	0.095	0.633
Cluster 7	1.581	0.082	0.816	0.309	0.042	0.331
Cluster 8	1.25	0.025	0.861	0.327	0	0.037
Cluster 9	0.205	0.029	0.156	0.02	0	0

Figures B.4 and **B.S10** also show the corresponding superpositions of the cluster snapshots which were generated from the top 50 secondary configurations of peptides referring to the center of individual clusters (sorting by argmax of RMSD) for visualizing the dimeric structures. The centers of structural configuration represent the average C_{α} coordinate over the whole cluster. **Table B.S5** summarizes the grouping classification based on domain contacts for dimer clusters of ORF7a/WT BST-2 as well as ORF7a/I28S BST-2. Note that the ORF7a/WT BST-2 cluster with an overall occupancy of < 0.3%, i.e., Cluster 9, was not classified due to minimal contacts of the dimer.

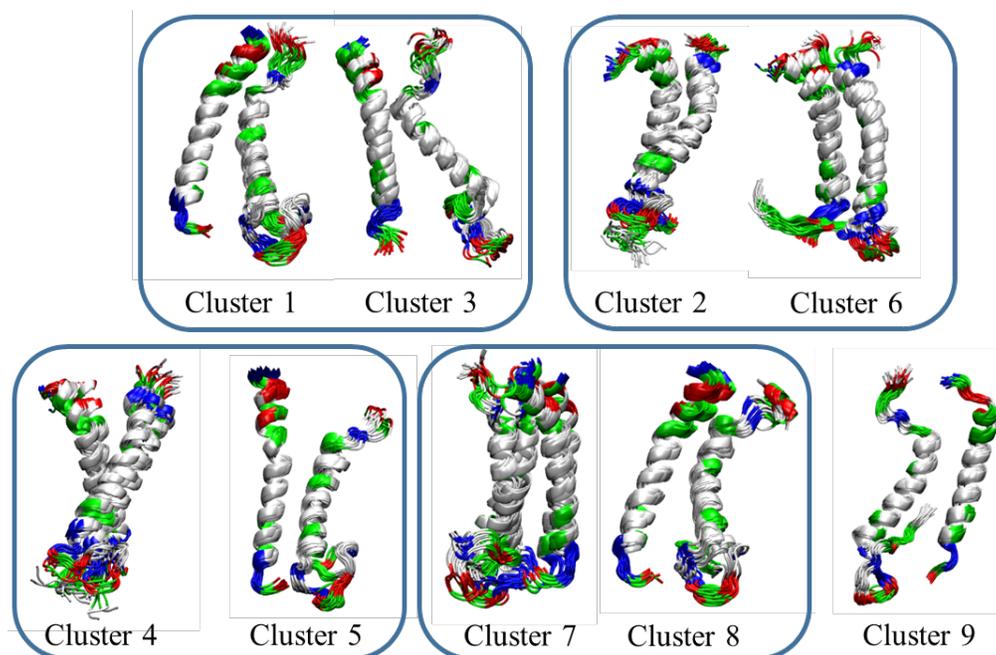


Figure B.S10: AA dimer snapshots of the WT clusters made by the superposition of the top 50 configurations from the center of each cluster, with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively.

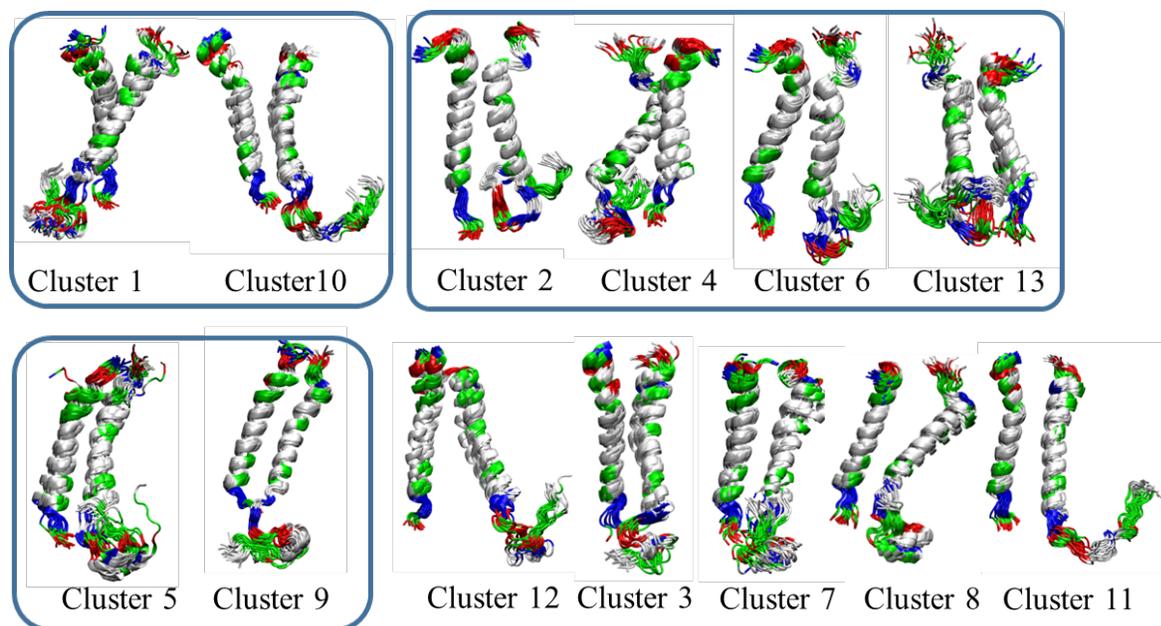


Figure B.S11: AA dimer snapshots of the I28S clusters made by the superposition of the top 50 configurations from the center of each cluster, with nonpolar, polar, acidic, and basic residues colored white, green, red, and blue, respectively.

Table B.S5: Grouping of the GMVAE identified dimer clusters

Group*	GMVAE clusters	Contacting domains	% of data
WT Group 1	Cluster 1, Cluster 3	top TM	19.66
WT Group 2	Cluster 2, Cluster 6	top and middle TM, CYTO	22.98
WT Group 3	Cluster 7, Cluster 8	top and middle TM	21.57
WT Group 4	Cluster 4, Cluster 5	middle and bottom TM, CYTO	17.36
I28S Group 1	Cluster 1, Cluster10	middle and bottom TM	13.29
I28S Group 2	Cluster 3	middle and bottom TM, CYTO	6.83
I28S Group 3	Cluster 2, Cluster 4, Cluster 6, Cluster 13	top and middle TM, CYTO	32.23
I28S Group 4	Cluster 5, Cluster 9	middle TM	9.99
I28S Group 5	Cluster 12	EC, top TM	12.77

*WT groups classify the GMVAE clusters from ORF7a/WT BST-2 dimerization while I28S groups classify the GMVAE clusters from ORF7a/I28S BST-2 dimerization.

The same grouping approach for ORF7a/I28S BST-2 were applied after the dimer configurations were clustered by GMAVE. **Figure B.5** and **B.S9** show the contact maps of the ORF7a/I28S BST-2 dimer clusters with the contacting residue pair occupies 0.01 ~ 1.1 % of whole contact map (**Table B.S6**) of the identified clusters. Note that Cluster 7, 8, and 11 were not included in the grouping due to a low occupancy of their contact maps (**Table B.S5**). Therefore, a total of 4 and 5 structural groups with various contacting behaviors were identified for WT and I28S BST-2 respectively. The variation in residue contacts indicates the heterodimeric dimerization of BST-2 and ORF7a has multiple states of association.

Table B.S6: Occupancy percentage of contact map for the I28S clusters.

	overall	EC	TM_top	TM_mid	TM_bot	IC
Cluster 1	0.483	0	0.005	0.292	0.154	0.033
Cluster 2	0.64	0.037	0.259	0.246	0.008	0.09
Cluster 3	0.849	0	0.001	0.23	0.172	0.446
Cluster 4	0.584	0.036	0.279	0.243	0.016	0.009
Cluster 5	1.148	0.029	0.365	0.612	0.12	0.022
Cluster 6	0.772	0.011	0.39	0.327	0.04	0.005
Cluster 7	0.112	0	0	0.015	0.035	0.062
Cluster 8	0.16	0	0	0	0.028	0.132
Cluster 9	0.984	0.038	0.111	0.52	0.15	0.165
Cluster 10	0.342	0	0	0.206	0.135	0.001
Cluster 11	0.011	0	0.001	0.01	0	0

Cluster 12	0.553	0.399	0.149	0.005	0	0
Cluster 13	0.578	0.001	0.274	0.279	0	0.025

Helix-Helix C_α Backbone Packing

To examine the helical packing of each group, a tight packing form (the top 5000 sampling from the center of each group) was analyzed and represent visually for individual groups. To further examine the interaction of residues, the whole configurations (with a membership probability of more than 0.75) will be included to obtain the energetic behavior of individual groups.

To construct the protein-protein interface structural model of these heterodimeric type I transmembrane proteins, the lateral association of helix-helix crossing angle (Ω), packing distance ($D_{\text{Helix-Helix}}$),⁷⁰ and packing motifs were examined using the C_α coordinates of individual clusters. Then, residue-residue interactions, including sidechain contacts, were further assessed by residue binding energetics. The above clustering of atomic configurations enhancing their regional C_α contacts were used to assess the lateral helix-helix packing. **Figures B.S12** and **B.S13** demonstrate the distribution of the Ω and $D_{\text{Helix-Helix}}$ for each cluster in which the top 5000 samples of each group (and cluster) were taken and the helices of each configuration were defined as the largest continuously helical domain (including 3-helix, α -helix, and 5-helix structures) by using DSSP-2.2.0.⁷¹ *Tables B.S7, B.S8, and B.S9* summarizes the size (L_{Helix} , number of residues) of the helices, Ω and $D_{\text{Helix-Helix}}$ for each heterodimeric cluster.

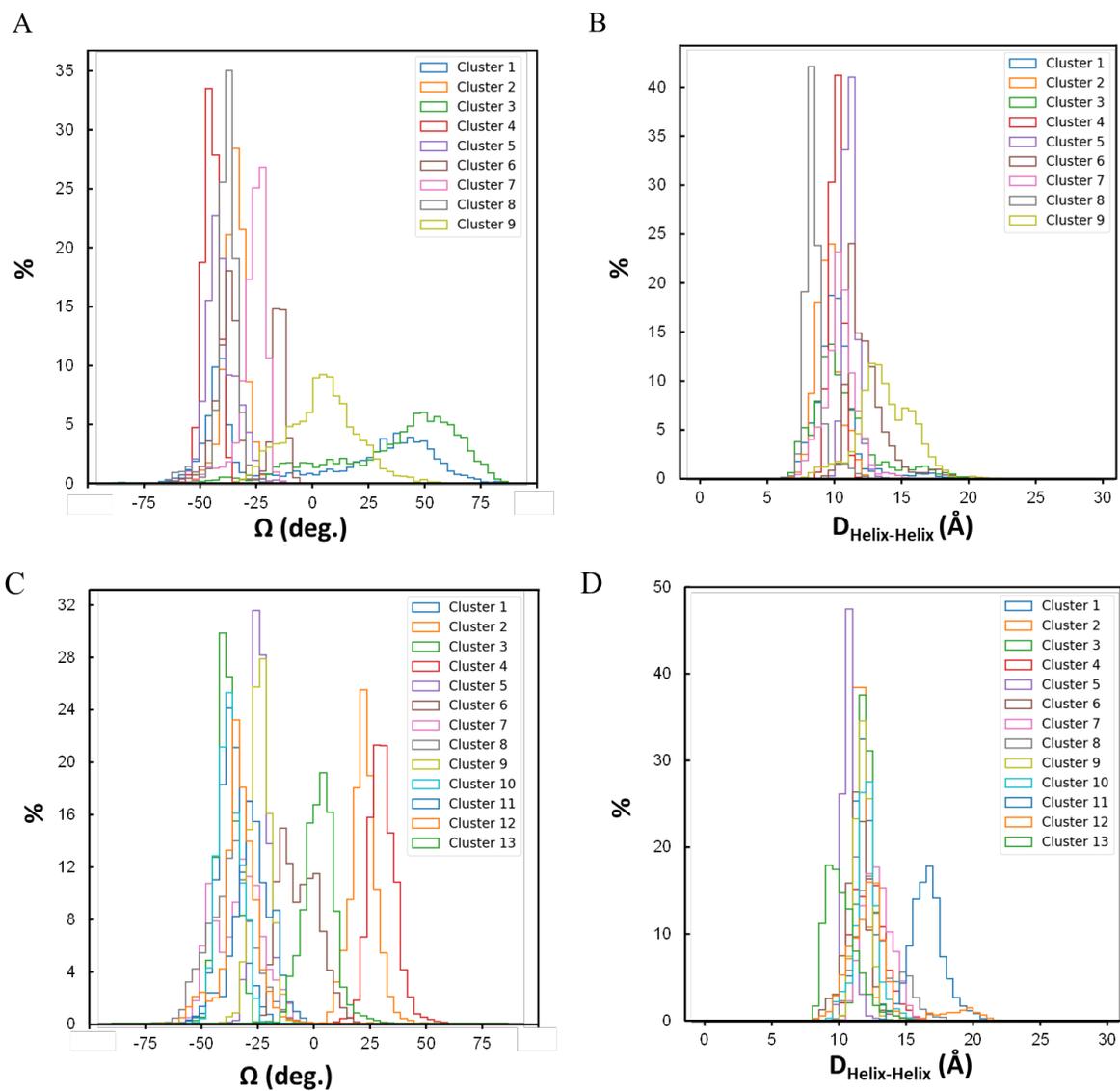


Figure B.S12: Distribution of crossing angle (A, C) and helices distance (B, D) of the WT and I28S clusters using the top 5000 configurations from the center of GMVAE clusters.

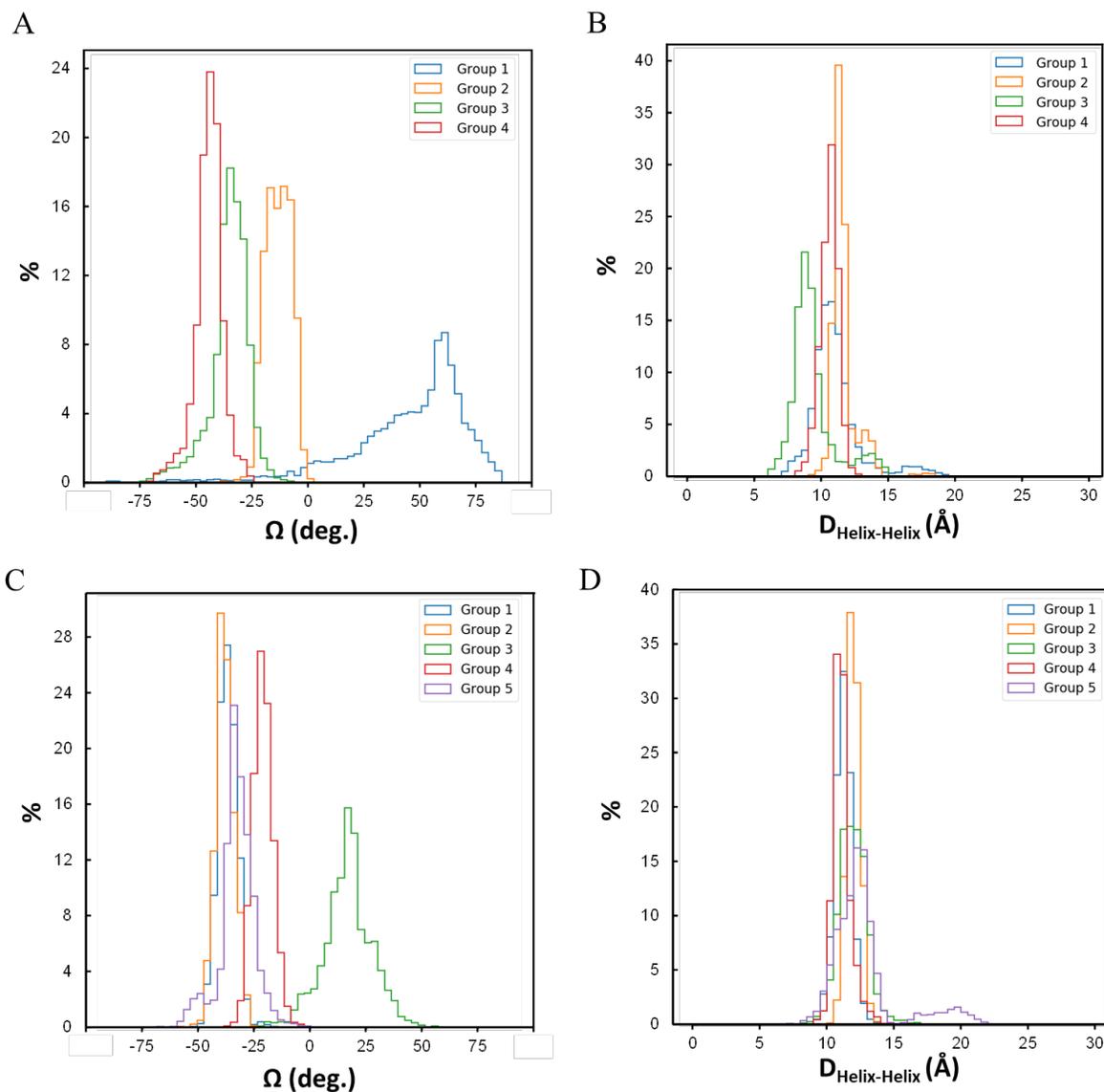


Figure B.S13: Distribution of crossing angle (A, C) and helices distance (B, D) of the WT and I28S groups using the top 5000 configurations from the center of GMVAE clustering groups.

Table B.S7: Length, distance, and crossing angle of helices. The mean and standard error have been provided.

	BST-2 L_{Helix}^*	ORF7a L_{helix}	$D_{\text{helix-helix}}$ (Å)	Ω (deg.)
WT Group 1	21.5 ± 0.02	26.43 ± 0.03	11.1 ± 0.03	44.85 ± 0.39
WT Group 2	26.37 ± 0.03	25.9 ± 0.02	11.67 ± 0.02	-13.07 ± 0.08
WT Group 3	20.57 ± 0.04	27.36 ± 0.04	9.38 ± 0.02	-35.01 ± 0.13
WT Group 4	25.21 ± 0.05	25.28 ± 0.03	10.56 ± 0.01	-44.04 ± 0.09
I28S Group 1	29.18 ± 0.04	27.51 ± 0.03	11.24 ± 0.01	-37.41 ± 0.06
I28S Group 2	29.89 ± 0.03	27.9 ± 0.03	11.98 ± 0.01	-38.5 ± 0.06
I28S Group 3	21.38 ± 0.02	24.36 ± 0.04	11.97 ± 0.02	16.49 ± 0.15
I28S Group 4	23.1 ± 0.02	23.82 ± 0.03	11.13 ± 0.01	-21.53 ± 0.06

I28S Group 5	28.8 ± 0.06	28.17 ± 0.03	12.71 ± 0.04	-33.05 ± 0.11
--------------	-----------------	------------------	------------------	-------------------

* L_{Helix} is the number of the residue in the DSSP identified biggest helix domain.

Table B.S8: Length, distance, and crossing angle of helices for the WT clusters. The mean and standard deviation error were calculated over 5000 sampling of individual clusters.

	BST-2 L_{Helix}	ORF7a L_{Helix}	$D_{\text{Helix-Helix}}$ (Å)	Ω (deg.)
Cluster 1	21.55 ± 0.02	26.6 ± 0.03	10.12 ± 0.02	-0.13 ± 0.58
Cluster 2	25.16 ± 0.05	24.25 ± 0.05	9.49 ± 0.01	-34.87 ± 0.07
Cluster 3	21.41 ± 0.02	24.43 ± 0.04	10.62 ± 0.03	36.21 ± 0.41
Cluster 4	25.78 ± 0.03	24.87 ± 0.03	10.1 ± 0.01	-45.65 ± 0.05
Cluster 5	26.37 ± 0.05	24.72 ± 0.04	11.15 ± 0.01	-40.79 ± 0.09
Cluster 6	23.5 ± 0.04	25.9 ± 0.03	12.31 ± 0.02	-29.8 ± 0.18
Cluster 7	21.81 ± 0.03	27.52 ± 0.04	10.4 ± 0.02	-26.51 ± 0.08
Cluster 8	22.4 ± 0.02	28.15 ± 0.02	8.45 ± 0.01	-38.59 ± 0.08
Cluster 9	22 ± 0.02	21.73 ± 0.03	13.76 ± 0.03	4.34 ± 0.22

Table B.S9: Length, distance, and crossing angle of helices for the I28S clusters. The mean and standard deviation error were calculated over 5000 sampling of individual clusters.

	BST-2 L_{Helix}	ORF7a L_{Helix}	$D_{\text{Helix-Helix}}$ (Å)	Ω (deg.)
Cluster 1	23.25 ± 0.03	26.04 ± 0.04	11.81 ± 0.01	-36.86 ± 0.08
Cluster 2	21.48 ± 0.01	24.41 ± 0.03	11.57 ± 0.01	22.55 ± 0.08
Cluster 3	29.89 ± 0.03	27.9 ± 0.03	11.98 ± 0.01	-38.5 ± 0.06
Cluster 4	21.08 ± 0.02	24.22 ± 0.04	12.32 ± 0.02	30.98 ± 0.08
Cluster 5	22.62 ± 0.01	24.34 ± 0.03	10.75 ± 0.01	-23.9 ± 0.05
Cluster 6	20.76 ± 0.01	25.33 ± 0.03	11.34 ± 0.01	-6.1 ± 0.12
Cluster 7	23.26 ± 0.02	25.69 ± 0.03	12.85 ± 0.02	-33.35 ± 0.15
Cluster 8	23.25 ± 0.02	26.22 ± 0.03	12.82 ± 0.02	-36.82 ± 0.14
Cluster 9	23.54 ± 0.02	24.68 ± 0.03	11.74 ± 0.01	-23.58 ± 0.06
Cluster 10	29.47 ± 0.02	26.84 ± 0.05	12.03 ± 0.01	-37.27 ± 0.07
Cluster 11	28.84 ± 0.04	27.99 ± 0.03	16.63 ± 0.02	-27.27 ± 0.12
Cluster 12	28.81 ± 0.06	28.17 ± 0.03	12.71 ± 0.04	-33.04 ± 0.11
Cluster 13	21.68 ± 0.03	25.51 ± 0.02	10.26 ± 0.02	2.72 ± 0.1

In general, a single helical domain per peptide was found, but a short second helix domain was found in I28S Group 4. A short β - β contact has been found in the CYTO domain in WT Group 2. A variation of Ω distributions indicate various orientations of both helices, and their associations could be either right-handed (a negative crossing angle) or left-handed (a positive crossing angle) crossing, though the right-handed crossing is the majority. **Figure B.S13B** shows a distinguishable

$D_{\text{Helix-Helix}}$ across WT groups while **Figure B.S13D** demonstrates the similarity of the $D_{\text{Helix-Helix}}$ distributions among the I28S groups. The results show helix packing variations between groups demonstrating what appears to be high entropic heterodimeric dimerization of BST-2 and ORF7a. It is worth noting that in Group 3 with the shortest $D_{\text{Helix-Helix}}$, BST-2 has a short L_{Helix} compared to other groups while the ORF7a L_{Helix} is longest among all WT groups, resulting in ~ 7 residue differences between the helices in the WT packing. Also, the average L_{Helix} of BST-2 over I28S groups is about 3 residues longer, compared to the WT BST-2 while the average L_{Helix} of ORF7a is the same in both heterodimers.

Binding Energetics and Heli-packing Motifs

Tables B.S10 & B.S11 list the residue pairing with a distance less than the cutoff (i.e., 9 Å) indicating that the WT heterodimerization is more likely driven by the interactions within the CYTO and EC domains, compared to the I28S heterodimerization. Examination of the contacts within the TM domain demonstrate that various pairing residues in the helix-packing region neither correlate to the length of the L_{Helix} nor the strength of interactions in the CYTO and EC domains, indicating the TM contacts result from a complex mechanism of protein-protein and protein-lipid interactions. The hydrophobic interaction of these TM contacts is the main contributor to helix-helix associations within the membrane. These contacts can vary from helix-helix packing that spans the TM region to packing that is limited to shorter ranges. Extended TM contact exists for the WT Group 2 and 3 (**Table B.S10**) and I28S Group 4 (**Table B.S11**), compared to more localized regions with other classified groups. WT Group 3 has a tighter packing, having a shorter distance average among the TM contacting pairs, compared to WT Group 2, as expected due to the shortest $D_{\text{Helix-Helix}}$ of WT Group 3.

Table B.S10: Summary of WT Helix-Helix C α Contacting Residues Pairs.

Group 1			Group 2			Group 3			Group 4		
BST-2	ORF7a	distance (Å)									
P 40	L 102	6.89 ± 0.01	G 27	C 113	8.78 ± 0.01	I 33	V 108	8.2 ± 0.01	G 27	F 114	8.98 ± 0.02
I 43	F 101	7.91 ± 0.02	G 27	F 114	8.86 ± 0.01	I 36	V 104	8.63 ± 0.02	G 27	I 110	7.46 ± 0.02
I 43	L 102	6.74 ± 0.02	G 27	I 110	8.67 ± 0.01	I 36	V 108	8.87 ± 0.01	V 30	I 110	8.16 ± 0.01
I 43	P 99	7.99 ± 0.02	V 30	I 110	8.68 ± 0.01	L 37	A 105	6.93 ± 0.01	L 31	A 106	7.33 ± 0.01
I 43	S 98	6.51 ± 0.01	L 31	I 110	7.78 ± 0.01	L 37	F 101	7.41 ± 0.01	L 31	I 107	8.04 ± 0.03
F 44	E 95	8.76 ± 0.02	I 34	A 106	7.72 ± 0.01	L 37	V 104	7.06 ± 0.02	L 31	I 110	7.33 ± 0.01
F 44	L 102	7.33 ± 0.01	G 38	A 106	8.98 ± 0.01	L 37	V 108	8.84 ± 0.01	I 34	A 106	8.37 ± 0.03
F 44	P 99	6.57 ± 0.02	G 38	I 103	8.86 ± 0.01	P 40	F 101	6.94 ± 0.02	V 35	A 106	8.68 ± 0.02
F 44	S 98	6.1 ± 0.01	G 38	L 102	8.27 ± 0.01	P 40	I 100	7.68 ± 0.02			
			I 42	L 102	8.16 ± 0.02	P 40	V 104	8.29 ± 0.01			
			I 42	P 99	6.32 ± 0.02	P 40	Y 97	8.1 ± 0.02			
			I 42	S 98	8.7 ± 0.02	L 41	F 101	7.52 ± 0.01			
			T 45	E 95	8.33 ± 0.03	L 41	Y 97	7.17 ± 0.02			
			T 45	P 99	7.49 ± 0.04	I 43	Q 94a	8.49 ± 0.02			
			T 45	S 98	7.59 ± 0.02	F 44	L 96	7.7 ± 0.02			
			I 46	E 95	7.73 ± 0.03	F 44	Q 94a	5.34 ± 0.01			
			I 46	P 99	8.51 ± 0.03	F 44	Q 94b	7.07 ± 0.02			
			I 46	S 98	8.88 ± 0.02	F 44	Y 97	6.18 ± 0.01			
						T 45	Q 94a	7.35 ± 0.02			
						T 45	Q 94b	8.3 ± 0.03			
						T 45	Y 97	8.76 ± 0.02			
						I 46	E 91	8.75 ± 0.03			
						I 46	Q 94a	8.08 ± 0.02			

Table B.S11: Summary of I28S Helix-Helix C α Contacting Residues Pairs.

Group 1			Group 2			Group 3			Group 4			Group 5		
BST-2	ORF7a	distance (Å)												
L 29	C 113	8.89 ± 0.03	L 29	C 113	8.41 ± 0.01	V 35	I 103	8.85 ± 0.02	L 29	I 107	8.7 ± 0.01	F 44	L 96	8.97 ± 0.02
L 29	F 114	8.87 ± 0.03	L 29	I 110	7.99 ± 0.01	V 39	I 100	7.63 ± 0.02	L 29	T 111	8.67 ± 0.01	A 48	L 96	7.27 ± 0.02
L 29	I 110	7.91 ± 0.03	V 30	I 110	8.94 ± 0.02	V 39	I 103	8.61 ± 0.02	L 29	V 108	8.45 ± 0.01	E 51	L 96	8.97 ± 0.02
I 33	A 106	8.02 ± 0.03	I 33	A 106	7.56 ± 0.02	I 42	I 100	8.78 ± 0.03	I 33	I 107	8.45 ± 0.02	E 51	V 93	7.69 ± 0.03
I 33	I 110	8.18 ± 0.03	I 33	F 119	8.51 ± 0.02	I 43	I 100	8.22 ± 0.02	I 33	V 104	6.56 ± 0.01	A 52	E 95	6.93 ± 0.05
			I 33	I 110	8.28 ± 0.02	I 43	L 96	8.04 ± 0.03	I 36	I 100	7.45 ± 0.01	A 52	L 96	6.17 ± 0.04
			L 37	A 106	8.96 ± 0.02	I 43	Y 97	8.44 ± 0.03	I 36	I 103	8.44 ± 0.01	A 52	Q 94a	8.44 ± 0.05
									I 36	V 104	7.92 ± 0.01	A 52	V 93	6.6 ± 0.05
									L 37	F 101	8.55 ± 0.01			
									L 37	I 100	6.77 ± 0.01			
									L 37	V 104	8.56 ± 0.01			
									L 37	Y 97	7.86 ± 0.02			
									P 40	I 100	8.75 ± 0.01			
									P 40	L 96	8.31 ± 0.02			
									P 40	Y 97	8.61 ± 0.02			
									L 41	L 96	8.94 ± 0.03			

According to the C α pairing, the helix-helix packing motif could be identified as the “knobs-into-holes” heptad⁶³ repeats within the closest helix packing configurations (**Figures 1C**), but it was not clear for other groups due to a short contacting region. The WT Group 3 helices associate through I³⁶L³⁷xxP⁴⁰L⁴¹xxF⁴⁴T⁴⁵ and V¹⁰⁸xxA¹⁰⁵V¹⁰⁴xxF¹⁰¹I¹⁰⁰xxY⁹⁷xxQ^{94b}Q^{94a} of BST-2 and ORF7a respectively and form a packing complex. On the other hand, the I28S Group 4 TM helices pack with heptad repeats through L²⁹xxxI³³xxI³⁶L³⁷xxP⁴⁰L⁴¹ and V¹⁰⁸I¹⁰⁷xxV¹⁰⁴xxxI¹⁰⁰xxY⁹⁷L⁹⁶ of BST-2 and ORF7a respectively, and form TM heterodimer structural complexes. Both involve stacking interactions among F⁴⁴, Y⁹⁷, and Q^{94a} at the EC/TM interface. **Figures B.S14-17** visualize the helix-helix contacting motif including side-chain contacts identified based on a contact distance of less than 3Å in the center configuration of individual clusters. The contacting pairs in

the WT Group 3 located across the top and middle of the TM domain while the contacting motif in the I28S Group 4 dimerization shifts closer to the cytoplasmic side across the middle and bottom of the TM domain, indicating the I28S mutation initiates helix-helix lateral association close to the mutated position (at the bottom of the TM domain).

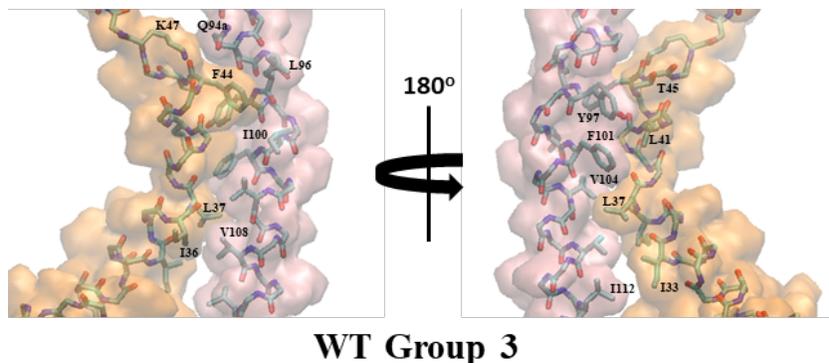


Figure BS.14: Molecular structure of the helix-packing motif with detailed side chain contacts in the WT Group 3 Oxygen, nitrogen, and carbon atoms colored red, blue, and gray, respectively and BST-2 and ORF7a colored orange and pink in QuickSurf style. The labeled residues have contacts with the other peptide chain within 3 Å.

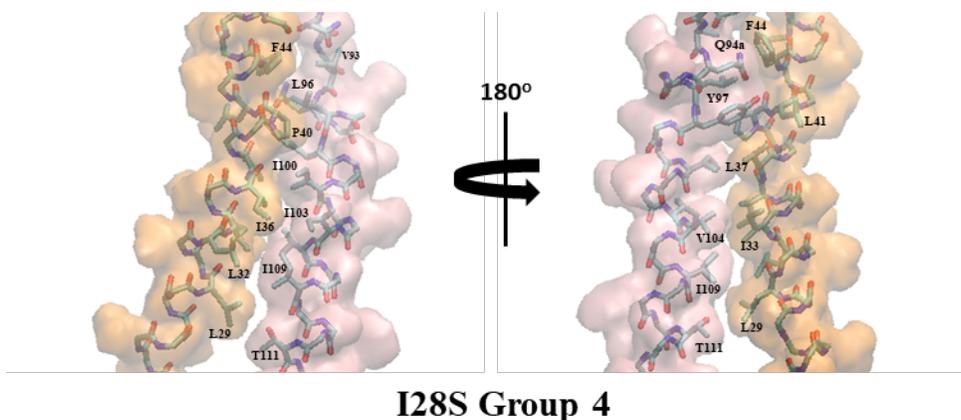
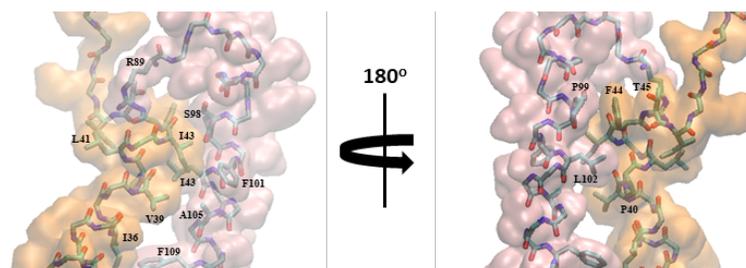
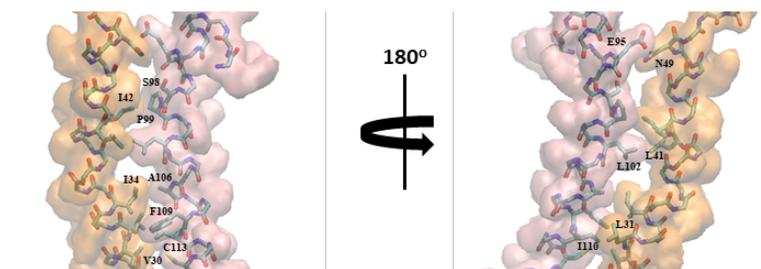


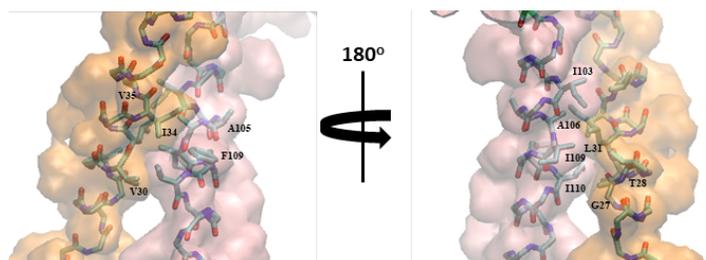
Figure BS.15: Molecular structure of the helix-packing motif with detailed side chain contacts in the I28S Group 4. Oxygen, nitrogen, and carbon atoms colored red, blue, and gray, respectively and I28S BST-2 and ORF7a colored orange and pink in QuickSurf style. The labeled residues have contacts with the other peptide chain within 3 Å.



WT Group 1

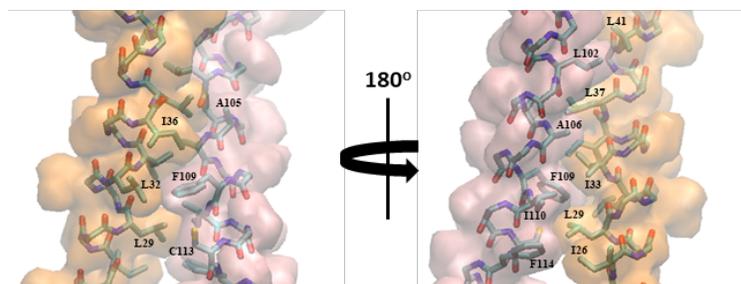


WT Group 2

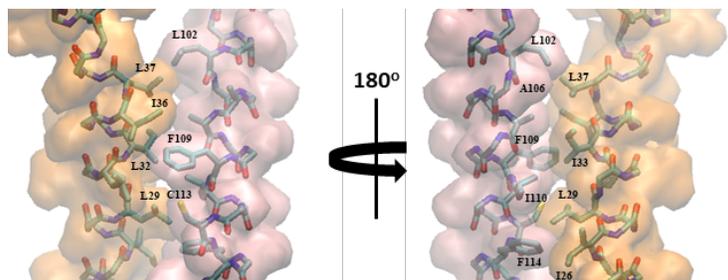


WT Group 4

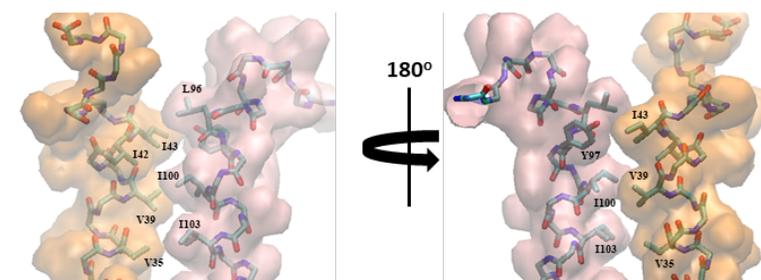
Figure BS.16: TM helix-helix interface of the WT groups.



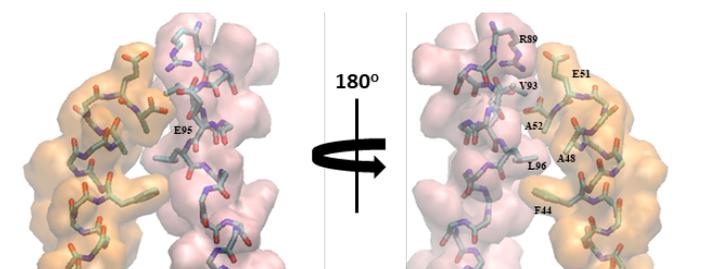
I28S Group 1



I28S Group 2



I28S Group 3



I28S Group 5

Figure BS17: TM helix-helix interface of the I28S groups.

Figures B.S18-20 show the energetics of dimerization contributions from individual residues, supporting the observation in the C_α pairing. The energetics also revealed the binding enthalpic contributions of mainly sidechain contacts quantitatively.

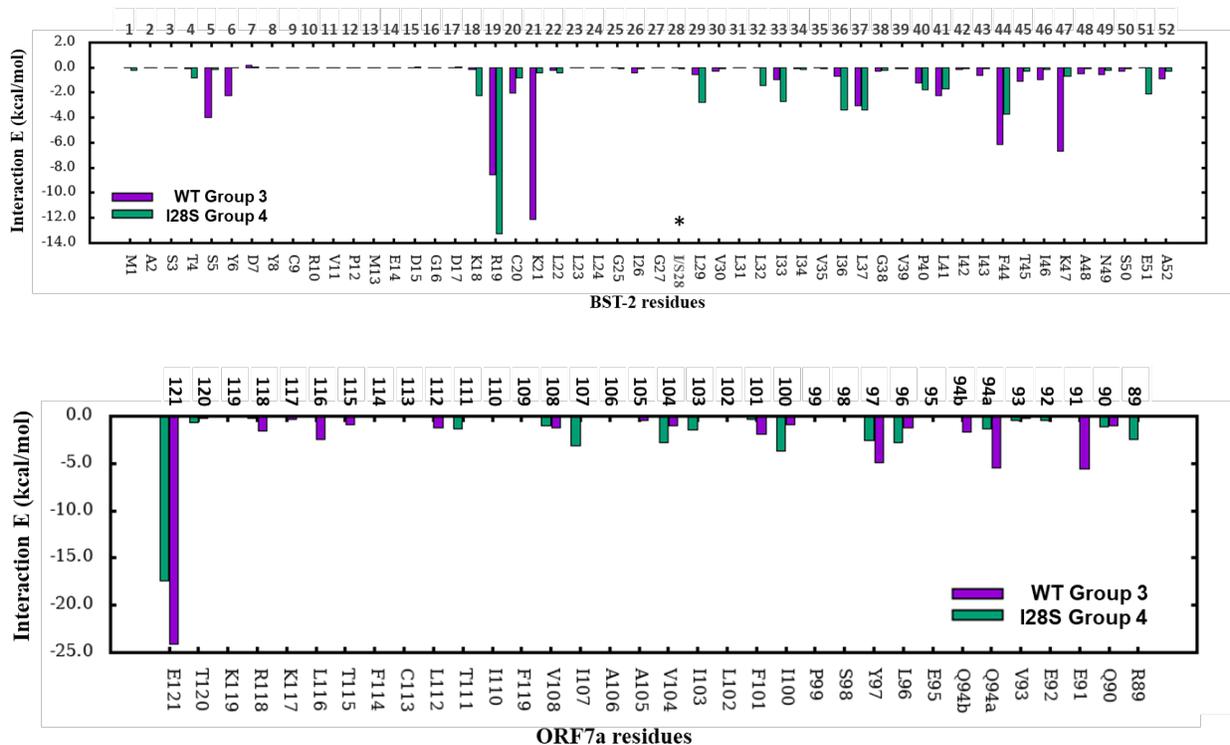


Figure BS.18: Interaction energies for amino acid residues (A) BST-2 and (B) ORF7a computed over the WT Group 3 and I28S Group 4. The individual membership possibility of heterodimeric configuration is more than 0.75 for its own group. The mutated residue is marked with an “*”.

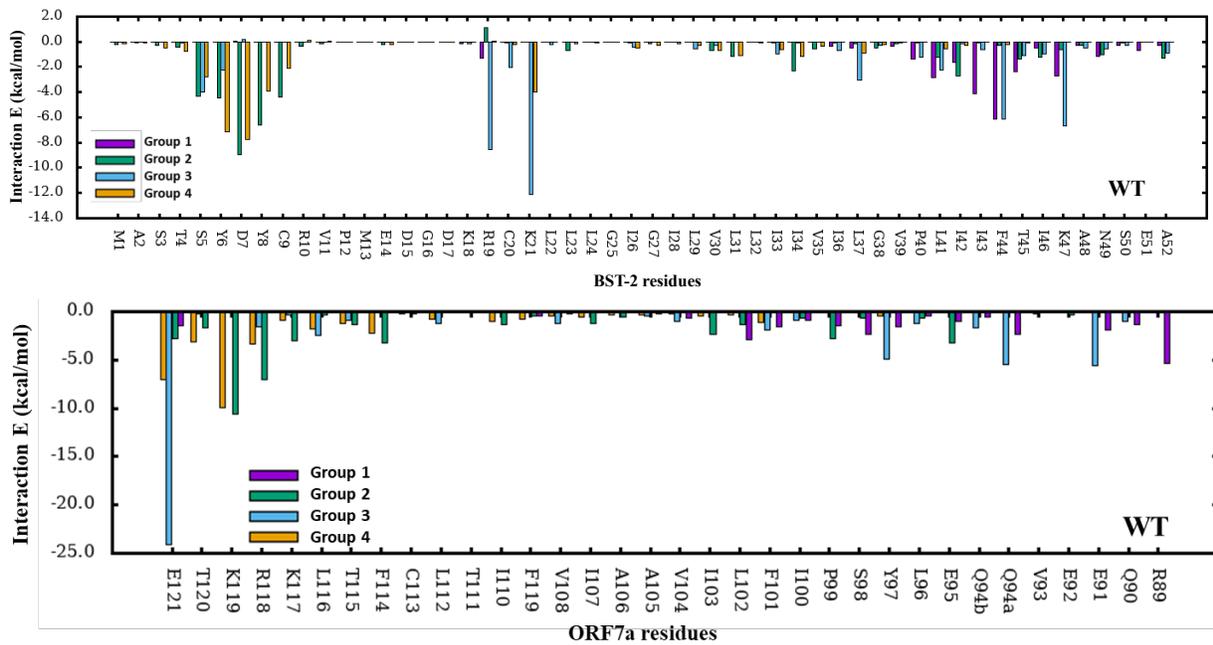


Figure B.S19: Interaction energy for each residue in (A) BST-2 and (B) ORF7a computed over the WT groups which the individual membership possibility of heterodimeric configuration is more 0.75 for its own group.

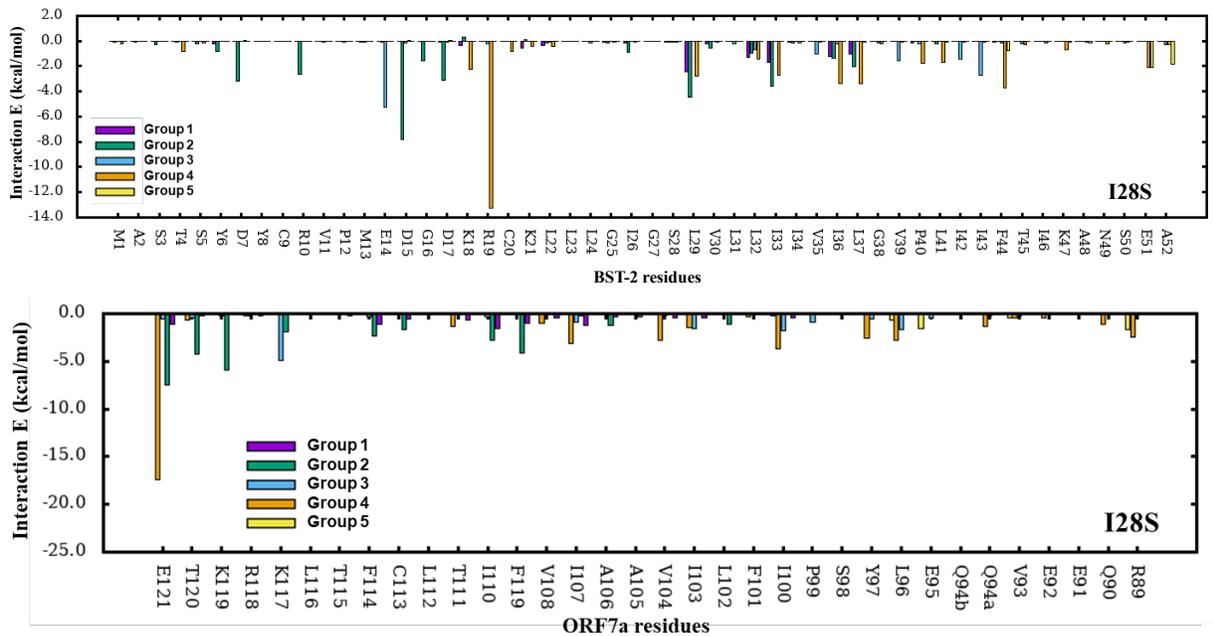


Figure B.S20: Interaction energy for each residue in (A) BST-2 and (B) ORF7a computed over the I28S groups which the individual membership possibility of heterodimeric configuration is more 0.75 for its own group.

These TM helix pairings within helix-packing motif are stabilized by van der Waals interactions of large side chain residues (valine, leucine, and isoleucine) which contribute to residue binding energetics. Moreover, energetically strong interacting residues have been found in the CYTO domain of loose packing groups. In WT Group 2 and 4 (**Figure B.S19**), energetic residues in the CYTO domains of WT BST-2 (S⁵, Y⁶, D⁷, Y⁸, and C⁹) interact with the CYTO residues of ORF7a (F¹¹⁴, T¹¹⁵, K¹¹⁷, R¹¹⁸, K¹¹⁹, and T¹²⁰). On the other hand, in I28S Group2 (**Figure B.S11**), the different interacting residues in the CYTO domain of I28S BST-2 (D⁷, R¹⁰, E¹⁴, D¹⁵, G¹⁶, and D¹⁷) interact with the same region of the ORF7a CYTO domain (K¹¹⁷, K¹¹⁹, T¹²⁰, and E¹²¹).

In the most extended packing conformations (WT Group 3 and I28S Group 4), E¹²¹ (ORF7a) in the CYTO domain forms a salt bridge with R¹⁹ (**Figure B.S21A**) or K²¹ (**Figure B.S21B**) in the WT with the formation probability of 0.83 and 0.76, respectively; E¹²¹ (ORF7a) also forms a salt bridge with R¹⁹ (**Figure B.S21C**) in the I28S mutant with the formation probability of 1.10. Moreover, the EC domain can dimerize with the formation of varied contacts. In WT Group 3, either a salt bridge formed by E⁹¹ (ORF7a) and K⁴⁷ (BST-2) (**Figure B.S22A**) with the formation probability of 0.21 or hydrogen bond between Q^{94a} (ORF7a) and K⁴⁷ (BST-2) (**Figure B.S22B**) with the formation probability of 0.12 promotes an amino- π interaction between Q^{94a} (ORF7a) and F⁴⁴ (BST-2) and a π - π stacking between Y⁹⁷ (ORF7a) and F⁴⁴ (BST-2) with the formation probability of 0.19 and 0.15, respectively. On the other hand, in I28S Group 4, the contacting complex is formed by Q^{94a}, L⁹⁶, Y⁹⁷ (ORF7a) and F⁴⁴ (BST-2) (**Figure B.S22C**) where Q^{94a} (ORF7a) and F⁴⁴ (BST-2) form an amino- π stacking with its formation probability of 0.22 as Q^{94a} and Y⁹⁷ form a backbone hydrogen bonding. L⁹⁶ (ORF7a) has a hydrophobic interaction in the complex which contributed to its residual energy (Figure 8).

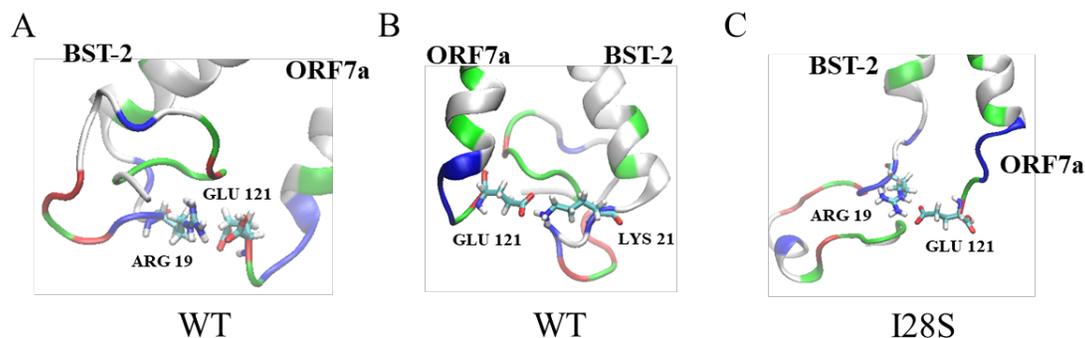


Figure B.S21: Salt bridges formed in WT Group 3 and I28S Group 4

Salt bridges are formed by (A) E¹²¹ (ORF7a) and R¹⁹ (WT BST-2), (B) E¹²¹ (ORF7a) and K²¹ (WT BST-2) and (C) E¹²¹ (ORF7a) and R¹⁹ (I28S BST-2). Nonpolar, polar, acidic, and basic residues in secondary structure are colored white, green, red, and blue, respectively. Oxygen, nitrogen, carbon and hydrogen atoms in highlight molecular structures are colored red, blue, cyan, and white, respectively.

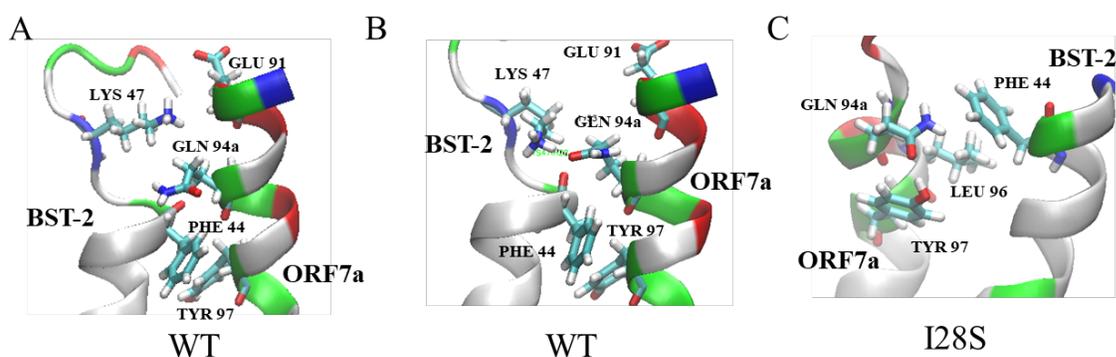


Figure B.S22: Contacting complexes of ORF7a/BST-2 heterodimerization at the EC/TM interface. In WT packing, (A) a salt bridge formed by E91 (ORF7a) and K47 (BST-2) or (B) a hydrogen bond between Q94a (ORF7a) and K47 (BST-2) promotes an amino- π interaction between Q94a (ORF7a) and F44 (BST-2) and π - π stacking between Y97 (ORF7a) and P44 (BST-2). In I28S packing, (C) an ORF7a intra-hydrogen bond formed by Q94a and Y97 in the extended helix domain promotes the same residues to form amino- π stacking and a π - π stacking interactions.

These CYTO and EC interactions involving polar/charged residues (i.e., glutamine, glutamic acid, lysine) can stabilize protein-protein binding by creating strong hydrogen bonding or salt-bridges.

Moreover, these residues present within a TM helix drive a highly stable helix-helix association

(I28S Group 4)⁶⁴ and form a stacking interaction in a helix-packing complex even enhancing the stability of both associations⁶⁵. It has also been observed that the polar residues contribute to drive stable heterodimerization, but they also interrupt the helix packing pattern. Interestingly, although dimerization profiles vary in the BST-2 WT and I28S mutant, it is consistently observed that the contacting region of the I28S dimerization extends deeper in the bilayer and away from polar residues of both peptides, indicating interactions within the hydrophobic region could essentially affect the helix-helix association of ORF7a and BST-2 heterodimer. According to the known structural and biofunctional features of ORF7a of SARS-Cov-2, K119, as a ubiquitination position⁶⁶, shows a strong interactive energy in the WT packing (Groups 2 and 4), compared to the I28S packing, indicating tight dimeric packing would reduce the activity of the K¹¹⁹ and may prohibit ubiquitination. Additionally, one of the identified key contact residues on ORF7a, A¹⁰⁵, has been reported as mutating to valine, resulting in a strong interaction among L¹⁰², I¹⁰³, V¹⁰⁴, and A¹⁰⁵ which is associated with increased severity and lethality of the infection in a group of Romanian patients⁶⁷. A¹⁰⁵ is listed as a key contacting residue (Table S8) in the WT packing (Group 3), matching the observation in the study of Romanian cases. On the other hand, several identified TM contact residues on BST-2 (Tables S8-S9): I³⁴, L³⁷, P⁴⁰, L⁴¹, and T⁴⁵ match those reported when this protein forms a heterodimer with the HIV-1 viral protein U.⁶⁸

B.6 References

- (1) Cai, M.; Xun, L. Organization and Regulation of Pentachlorophenol-Degrading Genes in *Sphingobium Chlorophenicum* ATCC 39723. *Journal of Bacteriology* **2002**, *184* (17), 4672–4680. <https://doi.org/10.1128/JB.184.17.4672-4680.2002>.
- (2) Wu, Z. Y.; McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *Jama-J Am Med Assoc Jama-J Am Med Assoc* **2020**, *323* (13), 1239–1242.
- (3) Zhou, P.; Yang, X. L.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin (Vol 579, Pg 270, 2020). *Nature Nature* **2020**, *588* (7836), E6–E6.
- (4) Shang, J.; Ye, G.; Shi, K.; Wan, Y. S.; Luo, C. M.; Aihara, H.; Geng, Q. B.; Auerbach, A.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature Nature* **2020**, *581* (7807), 221-+.
- (5) Chandrashekar, A.; Liu, J. Y.; Martinot, A. J.; McMahan, K.; Mercado, N. B.; Peter, L.; Tostanoski, L. H.; Yu, J. Y.; Maliga, Z.; Nekorchuk, M.; Busman-Sahay, K.; Terry, M.; Wrijil, L. M.; Ducat, S.; Martinez, D. R.; Atyeo, C.; Fischinger, S.; Burke, J. S.; Slein, M. D.; Pessaint, L.; Van Ry, A.; Greenhouse, J.; Taylor, T.; Blade, K.; Cook, A.; Finneyfrock, B.; Brown, R.; Teow, E.; Velasco, J.; Zahn, R.; Wegmann, F.; Abbink, P.; Bondzie, E. A.; Dagotto, G.; Gebre, M. S.; He, X.; Jacob-Dolan, C.; Kordana, N.; Li, Z. F.; Lifton, M. A.; Mahrokhian, S. H.; Maxfield, L. F.; Nityanandam, R.; Nkolola, J. P.; Schmidt, A. G.; Miller, A. D.; Baric, R. S.; Alter, G.; Sorger, P. K.; Estes, J. D.; Andersen, H.; Lewis, M. G.; Barouch, D. H. SARS-CoV-2 Infection Protects against Rechallenge in Rhesus Macaques. *Science Science* **2020**, *369* (6505), 812-+.
- (6) Shang, J.; Wan, Y. S.; Luo, C. M.; Ye, G.; Geng, Q. B.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *P Natl Acad Sci USA P Natl Acad Sci USA* **2020**, *117* (21), 11727–11734.

- (7) Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 Entry into Cells. *Nat Rev Mol Cell Bio Nat Rev Mol Cell Bio* **2021**.
- (8) Lu, R. J.; Zhao, X.; Li, J.; Niu, P. H.; Yang, B.; Wu, H. L.; Wang, W. L.; Song, H.; Huang, B. Y.; Zhu, N.; Bi, Y. H.; Ma, X. J.; Zhan, F. X.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z. H.; Zhou, W. M.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J. Y.; Xie, Z. H.; Ma, J. M.; Liu, W. J.; Wang, D. Y.; Xu, W. B.; Holmes, E. C.; Gao, G. F.; Wu, G. Z.; Chen, W. J.; Shi, W. F.; Tan, W. J. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *Lancet Lancet* **2020**, *395* (10224), 565–574.
- (9) Chan, J. F. W.; Kok, K. H.; Zhu, Z.; Chu, H.; To, K. K. W.; Yuan, S.; Yuen, K. Y. Genomic Characterization of the 2019 Novel Human-Pathogenic Coronavirus Isolated from a Patient with Atypical Pneumonia after Visiting Wuhan (Vol 23, Pg 2511, 2019). *Emerg Microbes Infec Emerg Microbes Infec* **2020**, *9* (1), 540–540.
- (10) Arya, R.; Kumari, S.; Pandey, B.; Mistry, H.; Bihani, S. C.; Das, A.; Prashar, V.; Gupta, G. D.; Panicker, L.; Kumar, M. Structural Insights into SARS-CoV-2 Proteins. *J Mol Biol J Mol Biol* **2021**, *433* (2).
- (11) Yoshimoto, F. K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J Protein J* **2020**, *39* (3), 198–216.
- (12) Lei, X. B.; Dong, X. J.; Ma, R. Y.; Wang, W. J.; Xiao, X.; Tian, Z. Q.; Wang, C. H.; Wang, Y.; Li, L.; Ren, L. L.; Guo, F.; Zhao, Z. D.; Zhou, Z.; Xiang, Z. C.; Wang, J. W. Activation and Evasion of Type I Interferon Responses by SARS-CoV-2. *Nat Commun Nat Commun* **2020**, *11* (1).
- (13) Cao, Z. G.; Xia, H. J.; Rajsbaum, R.; Xia, X. Z.; Wang, H. L.; Shi, P. Y. Ubiquitination of SARS-CoV-2 ORF7a Promotes Antagonism of Interferon Response. *Cell Mol Immunol Cell Mol Immunol* **2021**, *18* (3), 746–748.
- (14) Zhou, Z. L.; Huang, C. L.; Zhou, Z. C.; Huang, Z. X.; Su, L. L.; Kang, S. S.; Chen, X. X.; Chen, Q. Y.; He, S. H.; Rong, X.; Xiao, F.; Chen, J.; Chen, S. D. Structural Insight Reveals SARS-CoV-2 ORF7a as an Immunomodulating Factor for Human CD14(+) Monocytes. *Iscience Iscience* **2021**, *24* (3).

- (15) Lucas, C.; Wong, P.; Klein, J.; Castro, T. B. R.; Silva, J.; Sundaram, M.; Ellingson, M. K.; Mao, T.; Oh, J. E.; Israelow, B.; Takahashi, T.; Tokuyama, M.; Lu, P.; Venkataraman, A.; Park, A.; Mohanty, S.; Wang, H.; Wyllie, A. L.; Vogels, C. B. F.; Earnest, R.; Lapidus, S.; Ott, I. M.; Moore, A. J.; Muenker, M. C.; Fournier, J. B.; Campbell, M.; Odio, C. D.; Casanovas-Massana, A.; Yale IMPACT Team; Obaid, A.; Lu-Culligan, A.; Nelson, A.; Brito, A.; Nunez, A.; Martin, A.; Watkins, A.; Geng, B.; Kalinich, C.; Harden, C.; Todeasa, C.; Jensen, C.; Kim, D.; McDonald, D.; Shepard, D.; Courchaine, E.; White, E. B.; Song, E.; Silva, E.; Kudo, E.; DeIuliis, G.; Rahming, H.; Park, H.-J.; Matos, I.; Nouws, J.; Valdez, J.; Fauver, J.; Lim, J.; Rose, K.-A.; Anastasio, K.; Brower, K.; Glick, L.; Sharma, L.; Sewanan, L.; Knaggs, L.; Minasyan, M.; Batsu, M.; Petrone, M.; Kuang, M.; Nakahata, M.; Campbell, M.; Linehan, M.; Askenase, M. H.; Simonov, M.; Smolgovsky, M.; Sonnert, N.; Naushad, N.; Vijayakumar, P.; Martinello, R.; Datta, R.; Handoko, R.; Bermejo, S.; Prophet, S.; Bickerton, S.; Velazquez, S.; Alpert, T.; Rice, T.; Khoury-Hanold, W.; Peng, X.; Yang, Y.; Cao, Y.; Strong, Y.; Herbst, R.; Shaw, A. C.; Medzhitov, R.; Schulz, W. L.; Grubaugh, N. D.; Dela Cruz, C.; Farhadian, S.; Ko, A. I.; Omer, S. B.; Iwasaki, A. Longitudinal Analyses Reveal Immunological Misfiring in Severe COVID-19. *Nature* **2020**, *584* (7821), 463–469. <https://doi.org/10.1038/s41586-020-2588-y>.
- (16) Su, C.-M.; Wang, L.; Yoo, D. Activation of NF- κ B and Induction of Proinflammatory Cytokine Expressions Mediated by ORF7a Protein of SARS-CoV-2. *Sci Rep* **2021**, *11* (1), 13464. <https://doi.org/10.1038/s41598-021-92941-2>.
- (17) Cheung, C. Y.; Poon, L. L. M.; Ng, I. H. Y.; Luk, W.; Sia, S.-F.; Wu, M. H. S.; Chan, K.-H.; Yuen, K.-Y.; Gordon, S.; Guan, Y.; Peiris, J. S. M. Cytokine Responses in Severe Acute Respiratory Syndrome Coronavirus-Infected Macrophages In Vitro: Possible Relevance to Pathogenesis. *J Virol* **2005**, *79* (12), 7819–7826. <https://doi.org/10.1128/JVI.79.12.7819-7826.2005>.
- (18) Ohtomo, T.; Sugamata, Y.; Ozaki, Y.; Ono, K.; Yoshimura, Y.; Kawai, S.; Koishihara, Y.; Ozaki, S.; Kosaka, M.; Hirano, T.; Tsuchiya, M. Molecular Cloning and Characterization of a Surface Antigen Preferentially Overexpressed on Multiple Myeloma Cells. *Biochem Bioph Res Co Biochem Bioph Res Co* **1999**, *258* (3), 583–591.

- (19) Randall, R. E.; Goodbourn, S. Interferons and Viruses: An Interplay between Induction, Signalling, Antiviral Responses and Virus Countermeasures. *J Gen Virol J Gen Virol* **2008**, *89*, 1–47.
- (20) Andrew, A. J.; Miyagi, E.; Kao, S.; Strebel, K. The Formation of Cysteine-Linked Dimers of BST-2/Tetherin Is Important for Inhibition of HIV-1 Virus Release but Not for Sensitivity to Vpu. *Retrovirology Retrovirology* **2009**, *6*.
- (21) Terwilliger, E. F.; Cohen, E. A.; Lu, Y. C.; Sodroski, J. G.; Haseltine, W. A. Functional-Role of Human Immunodeficiency Virus Type-1 Vpu. *P Natl Acad Sci USA P Natl Acad Sci USA* **1989**, *86* (13), 5163–5167.
- (22) Lim, E. S.; Malik, H. S.; Emerman, M. Ancient Adaptive Evolution of Tetherin Shaped the Functions of Vpu and Nef in Human Immunodeficiency Virus and Primate Lentiviruses. *J Virol J Virol* **2010**, *84* (14), 7124–7134.
- (23) Yin, X.; Hu, Z.; Gu, Q. Y.; Wu, X. L.; Zheng, Y. H.; Wei, P.; Wang, X. J. Equine Tetherin Blocks Retrovirus Release and Its Activity Is Antagonized by Equine Infectious Anemia Virus Envelope Protein. *J Virol J Virol* **2014**, *88* (2), 1259–1270.
- (24) Dietrich, I.; McMonagle, E. L.; Petit, S. J.; Vijaykrishnan, S.; Logan, N.; Chan, C. N.; Towers, G. J.; Hosie, M. J.; Willett, B. J. Feline Tetherin Efficiently Restricts Release of Feline Immunodeficiency Virus but Not Spreading of Infection. *J Virol J Virol* **2011**, *85* (12), 5840–5852.
- (25) Bray, M.; Prasad, S.; Dubay, J. W.; Hunter, E.; Jeang, K. T.; Rekosh, D.; Hammarskjold, M. L. A Small Element from the Mason-Pfizer Monkey Virus Genome Makes Human-Immunodeficiency-Virus Type-1 Expression and Replication Rev-Independent. *P Natl Acad Sci USA P Natl Acad Sci USA* **1994**, *91* (4), 1256–1260.
- (26) McNatt, M. W.; Zang, T.; Hatzioannou, T.; Bartlett, M.; Ben Fofana, I.; Johnson, W. E.; Neil, S. J. D.; Bieniasz, P. D. Species-Specific Activity of HIV-1 Vpu and Positive Selection of Tetherin Transmembrane Domain Variants. *Plos Pathog Plos Pathog* **2009**, *5* (2).
- (27) Le Tortorec, A.; Neil, S. J. D. Antagonism to and Intracellular Sequestration of Human Tetherin by the Human Immunodeficiency Virus Type 2 Envelope Glycoprotein. *J Virol J Virol* **2009**, *83* (22), 11966–11978.
- (28) Mitchell, R. S.; Katsura, C.; Skasko, M. A.; Fitzpatrick, K.; Lau, D.; Ruiz, A.; Stephens, E. B.; Margottin-Goguet, F.; Benarous, R.; Guatelli, J. C. Vpu Antagonizes BST-2-Mediated

- Restriction of HIV-1 Release via Beta-TrCP and Endo-Lysosomal Trafficking. *Plos Pathog Plos Pathog* **2009**, 5 (5).
- (29) Douglas, J. L.; Viswanathan, K.; McCarroll, M. N.; Gustin, J. K.; Fruh, K.; Moses, A. V. Vpu Directs the Degradation of the Human Immunodeficiency Virus Restriction Factor BST-2/Tetherin via a Beta TrCP-Dependent Mechanism. *J Virol J Virol* **2009**, 83 (16), 7931–7947.
- (30) Yoshida, T.; Kao, S.; Strebel, K. Identification of Residues in the BST-2 TM Domain Important for Antagonism by HIV-1 Vpu Using a Gain-of-Function Approach. *Front. Microbio.* **2011**, 2. <https://doi.org/10.3389/fmicb.2011.00035>.
- (31) Gupta, D.; Sharma, P.; Singh, M.; Kumar, M.; Ethayathulla, A. S.; Kaur, P. Structural and Functional Insights into the Spike Protein Mutations of Emerging SARS-CoV-2 Variants. *Cell Mol Life Sci Cell Mol Life Sci* **2021**, 78 (24), 7967–7989.
- (32) Goffinet, C.; Allespach, I.; Homann, S.; Tervo, H. M.; Habermann, A.; Rupp, D.; Oberbremer, L.; Kern, C.; Tibroni, N.; Welsch, S.; Krijnse-Locker, J.; Banting, G.; Krausslich, H. G.; Fackler, O. T.; Keppler, O. T. HIV-1 Antagonism of CD317 Is Species Specific and Involves Vpu-Mediated Proteasomal Degradation of the Restriction Factor. *Cell Host Microbe Cell Host Microbe* **2009**, 5 (3), 285–297.
- (33) Rong, L.; Zhang, J.; Lu, J.; Pan, Q.; Lorgeoux, R.-P.; Aloysius, C.; Guo, F.; Liu, S.-L.; Wainberg, M. A.; Liang, C. The Transmembrane Domain of BST-2 Determines Its Sensitivity to Down-Modulation by Human Immunodeficiency Virus Type 1 Vpu. *J Virol* **2009**, 83 (15), 7536–7546. <https://doi.org/10.1128/JVI.00620-09>.
- (34) Ishikawa, J.; Kaisho, T.; Tomizawa, H.; Lee, B. O.; Kobune, Y.; Inazawa, J.; Oritani, K.; Itoh, M.; Ochi, T.; Ishihara, K.; Hirano, T. Molecular-Cloning and Chromosomal Mapping of a Bone-Marrow Stromal Cell-Surface Gene, Bst2, That May Be Involved in Pre-B-Cell Growth. *Genomics Genomics* **1995**, 26 (3), 527–534.
- (35) Wang, S. M.; Huang, K. J.; Wang, C. T. BST2/CD317 Counteracts Human Coronavirus 229E Productive Infection by Tethering Virions at the Cell Surface. *Virology Virology* **2014**, 449, 287–296.
- (36) Taylor, J. K.; Coleman, C. M.; Postel, S.; Sisk, J. M.; Bernbaum, J. G.; Venkataraman, T.; Sundberg, E. J.; Frieman, M. B. Severe Acute Respiratory Syndrome Coronavirus ORF7a Inhibits Bone Marrow Stromal Antigen 2 Virion Tethering through a Novel Mechanism of

- Glycosylation Interference. *Journal of Virology* **2015**, *89* (23), 11820–11833. <https://doi.org/10.1128/JVI.02274-15>.
- (37) Perez-Caballero, D.; Zang, T.; Ebrahimi, A.; McNatt, M. W.; Gregory, D. A.; Johnson, M. C.; Bieniasz, P. D. Tetherin Inhibits HIV-1 Release by Directly Tethering Virions to Cells. *Cell* **2009**, *139* (3), 499–511. <https://doi.org/10.1016/j.cell.2009.08.039>.
- (38) Su, P.-C.; Berger, B. W. Identifying Key Juxtamembrane Interactions in Cell Membranes Using AraC-Based Transcriptional Reporter Assay (AraTM). *J. Biol. Chem.* **2012**, *287* (37), 31515–31526. <https://doi.org/10.1074/jbc.M112.396895>.
- (39) Su, P.-C.; Berger, B. W. A Novel Assay for Assessing Juxtamembrane and Transmembrane Domain Interactions Important for Receptor Heterodimerization. *Journal of Molecular Biology* **2013**, *425* (22), 4652–4658. <https://doi.org/10.1016/j.jmb.2013.07.022>.
- (40) Gurtan, A. M.; Lu, V.; Bhutkar, A.; Sharp, P. A. In Vivo Structure-Function Analysis of Human Dicer Reveals Directional Processing of Precursor miRNAs. *Rna* **2012**, *18* (6), 1116–1122. <https://doi.org/10.1261/rna.032680.112>.
- (41) Yang, H.; Wang, J.; Jia, X.; McNatt, M. W.; Zang, T.; Pan, B.; Meng, W.; Wang, H.-W.; Bieniasz, P. D.; Xiong, Y. Structural Insight into the Mechanisms of Enveloped Virus Tethering by Tetherin. *Proceedings of the National Academy of Sciences* **2010**, *107* (43), 18428–18432. <https://doi.org/10.1073/pnas.1011485107>.
- (42) Deng, W.; Li, R. Juxtamembrane Contribution to Transmembrane Signaling: Transmembrane Signaling. *Biopolymers* **2015**, *104* (4), 317–322. <https://doi.org/10.1002/bip.22651>.
- (43) Sherry, S. T. dbSNP: The NCBI Database of Genetic Variation. *Nucleic Acids Research* **2001**, *29* (1), 308–311. <https://doi.org/10.1093/nar/29.1.308>.
- (44) Kim, D. E.; Chivian, D.; Baker, D. Protein Structure Prediction and Analysis Using the Robetta Server. *Nucleic Acids Research* **2004**, *32* (Web Server), W526–W531. <https://doi.org/10.1093/nar/gkh468>.
- (45) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865. <https://doi.org/10.1002/jcc.20945>.

- (46) Qi, Y.; Ingólfsson, H. I.; Cheng, X.; Lee, J.; Marrink, S. J.; Im, W. CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* **2015**, *11* (9), 4486–4494. <https://doi.org/10.1021/acs.jctc.5b00513>.
- (47) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoseck, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614. <https://doi.org/10.1002/jcc.21287>.
- (48) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>.
- (49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (50) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9* (1), 687–697. <https://doi.org/10.1021/ct300646g>.
- (51) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819–834. <https://doi.org/10.1021/ct700324x>.
- (52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690. <https://doi.org/10.1063/1.448118>.
- (53) van der Spoel, D.; van Maaren, P. J. The Origin of Layer Structure Artifacts in Simulations of Liquid Water. *J. Chem. Theory Comput.* **2006**, *2* (1), 1–11. <https://doi.org/10.1021/ct0502256>.

- (54) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10* (2), 676–690. <https://doi.org/10.1021/ct400617g>.
- (55) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 044130. <https://doi.org/10.1063/5.0014475>.
- (56) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802. <https://doi.org/10.1002/jcc.20289>.
- (57) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114* (23), 7830–7843. <https://doi.org/10.1021/jp101759q>.
- (58) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. <https://doi.org/10.1002/jcc.23354>.
- (59) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-Induced Forces between Two Hydrophilic Groups. *J. Phys. Chem.* **1994**, *98* (8), 2198–2202. <https://doi.org/10.1021/j100059a038>.
- (60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (61) Steinbach, P. J.; Brooks, B. R. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J. Comput. Chem.* **1994**, *15* (7), 667–683. <https://doi.org/10.1002/jcc.540150702>.

- (62) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *The Journal of Chemical Physics* **1995**, *103* (11), 4613–4621. <https://doi.org/10.1063/1.470648>.
- (63) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *The Journal of Chemical Physics* **1994**, *101* (5), 4177–4189. <https://doi.org/10.1063/1.467468>.
- (64) Andersen, H. C. Rattle: A “Velocity” Version of the Shake Algorithm for Molecular Dynamics Calculations. *Journal of Computational Physics* **1983**, *52* (1), 24–34. [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1).
- (65) Dilokthanakul, N.; Mediano, P. A. M.; Garnelo, M.; Lee, M. C. H.; Salimbeni, H.; Arulkumaran, K.; Shanahan, M. Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders. **2016**. <https://doi.org/10.48550/ARXIV.1611.02648>.
- (66) Ghorbani, M.; Prasad, S.; Klauda, J. B.; Brooks, B. R. Variational Embedding of Protein Folding Simulations Using Gaussian Mixture Variational Autoencoders. *J. Chem. Phys.* **2021**, *155* (19), 194108. <https://doi.org/10.1063/5.0069708>.
- (67) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive Aggregation of Membrane Proteins in the Martini Model. *PLoS ONE* **2017**, *12* (11), e0187936. <https://doi.org/10.1371/journal.pone.0187936>.
- (68) Barton, R.; Khakbaz, P.; Bera, I.; Klauda, J. B.; Iovine, M. K.; Berger, B. W. Interplay of Specific Trans- and Juxtamembrane Interfaces in Plexin A3 Dimerization and Signal Transduction. *Biochemistry* **2016**, *55* (35), 4928–4938. <https://doi.org/10.1021/acs.biochem.6b00517>.
- (69) Van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *Journal of Machine Learning Research* **2008**, *1*, 1–48.
- (70) Lee, J.; Im, W. Implementation and Application of Helix–Helix Distance and Crossing Angle Restraint Potentials. *J. Comput. Chem.* **2007**, *28* (3), 669–680. <https://doi.org/10.1002/jcc.20614>.
- (71) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577–2637. <https://doi.org/10.1002/bip.360221211>.