UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Genetic Engineering of Amphiphilic Intrinsically Disordered Proteins

Permalink

https://escholarship.org/uc/item/6gq7j04z

Author

Gleason, Jamie Marie

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Genetic Engineering of Amphiphilic Intrinsically Disordered Proteins

by

Jamie M Gleason

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Matthew B Francis, Chair
Professor Evan Miller
Professor Sabeeha Merchant

Fall 2022

Genetic Engineering of Amphiphilic Intrinsically Disordered Proteins

Copyright 2022
by
Jamie M Gleason

Abstract

Genetic Engineering of Amphiphilic Intrinsically Disordered Proteins

by

Jamie M Gleason

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Matthew B Francis, Chair

Due to the widespread use of surfactants, there have been renewed efforts to develop bio-derived and bio-degradable alternatives. However, the large-scale production of designer bio-surfactants for medical or environmental purposes remains a challenge regarding economic viability compared to their synthetic counterparts. This work shows that the previously studied and characterized intrinsically disordered protein (IDP) that was genetically modified to self-assemble into micelles (IDP-2Yx2A) has promise to be used as a vehicle for traditionally unusable hydrophobic drugs for *in vivo* delivery to mice. Additionally, this work created a scalable and versatile production method for these amphiphilic proteins and has allowed for the development of different iterations and mutant sequences that were easily attained. The protein sequence was separated and explored by its two opposing hydrophilic and hydrophobic domains. Variants with increased overall charges and systematically decreased hydrophilic domain lengths were produced and examined for their sizes, morphologies, and critical micelle concentrations (CMCs). Additionally, to explore the role of the hydrophobic domain on self-assembly properties, we developed a bioconjugation method to attach N-terminal cysteine peptides of varying hydrophobicities to the C-terminal tyrosine residue of the hydrophilic domain. This represents a facile method for creating a large number of hydrophobic library members to be screened as IDP-based bio-surfactants.

Dedicated to my mother, Laurie Gleason. Growing up with you as my role model is why I am here today.

Contents

Contents **ii**

List of Figures **v**

List of Tables **vii**

1 Insoluble Expression of Amphiphilic Intrinsically Disordered Proteins **1**
1.1 Abstract . . . 1
1.2 Introduction . . . 2
1.3 Results and Discussion . . . 4
1.3.1 Protein Expression . . . 6
1.3.2 Protein Purification . . . 6
1.4 Conclusion . . . 8
1.5 Materials and Methods . . . 9
1.5.1 Materials . . . 9
1.5.2 Instrumentation . . . 10
1.5.3 Methods of protein production . . . 11
1.6 Additional Figures . . . 14
1.7 Acknowledgements . . . 16
1.8 References . . . 16

2 *In vivo* Characterization of Intrinsically Disordered Proteins as Vehicles for Delivery to Glioblastoma Tumors **19**
2.1 Abstract . . . 19
2.2 Introduction . . . 20
2.3 Results and Discussion . . . 22
2.3.1 IDP-2Yx2A *in vitro* characterization . . . 22
2.3.2 Solubilizing SN38 . . . 24

2.3.3 SN38-IDP-2Yx2A *in vitro* studies . . . 26
2.3.4 Evaluating SN38-IDP-2Yx2A *in vivo* convection-enhanced delivery . . . 27
2.4 Conclusion . . . 30
2.5 Materials and Methods . . . 30
2.5.1 Reagents and Instrumentation . . . 30
2.5.2 Protein Expression and Purification . . . 32
2.5.3 IDP-2Yx2A Characterization . . . 33
2.5.4 Endotoxin Assay . . . 34
2.5.5 Dye Modification . . . 34
2.5.6 FRET Release Assay . . . 34
2.5.7 SN38 Release Study . . . 35
2.5.8 Cell Culture . . . 35
2.5.9 Cell Viability Assay . . . 35
2.5.10 Convection-enhanced Delivery (CED Infusion) . . . 35
2.5.11 *In vivo* Studies . . . 36
2.6 Additional Figures . . . 37
2.7 Acknowledgements . . . 40
2.8 References . . . 41

3 Genetic Engineering of an Intrinsically Disordered Protein's Hydrophilic Domain 46
3.1 Abstract . . . 46
3.2 Introduction . . . 47
3.3 Results and Discussion . . . 48
3.3.1 Exploring IDP length and charge . . . 48
3.3.2 Exploring hydrophilic-lipophilic balance . . . 54
3.4 Conclusion and Outlook . . . 56
3.5 Materials and Methods . . . 57
3.5.1 Materials . . . 57
3.5.2 Instrumentation . . . 58
3.5.3 Methods for protein production . . . 60
3.6 Additional Figures . . . 68
3.6.1 DLS analysis of IDP domain . . . 68
3.6.2 Disorder Profiles . . . 69
3.6.3 DLS analysis of IDP-Mutants . . . 69
3.6.4 Tinopal OB Solubilization . . . 73
3.7 Acknowledgements . . . 74
3.8 References . . . 74

4 Sequence Exploration of the Amphiphilic Intrinsically Disordered Protein's Hydrophobic Domain **77**
4.1 Abstract . . . 77
4.2 Introduction . . . 78
4.3 Results and Discussion . . . 80
4.4 Conclusion . . . 85
4.5 Materials and Methods . . . 85
4.5.1 Materials . . . 85
4.5.2 Instrumentation . . . 86
4.5.3 Methods for protein production . . . 88
4.6 Additional Figures . . . 95
4.7 Acknowledgements . . . 99
4.8 References . . . 99

List of Figures

1.1 Generation of self-assembling IDP micelles. 3
1.2 Structures for the Maltose Binding Protein (MBP) and Ketosteroid Isomerase (KSI) fusion partners. 5
1.3 Insoluble expression of IDP-2Yx2A proteins. 6
1.4 Schematic for insoluble epression and purification of KSI-IDP-2Yx2A. . . 7
1.5 KSI-IDP-2Yx2A detection during protein purification. 14
1.6 Yield of KSI-IDP-2Yx2A in LB vs. TB media. 15
1.7 KSI-IDP-2Yx2A detection in LB vs. TB media. 15

2.1 IDP micelles encapsulate hydrophobic cargo. 21
2.2 *In vitro* toxicity characterization of IDP-2Yx2A. 22
2.3 Hydrophobic cargo release kinetics of IDP-2Yx2A. 23
2.4 SN38 solubilization and formulated SN38-IDP-2Yx2A cell viability. . . . 25
2.5 *In vivo* toxicity studies for intracranial injection of SN38-IDP-2Yx2A. . . 27
2.6 Preliminary efficacy study of U87 tumor bearing mice. 28
2.7 Efficacy study of formulated SN38-IDP-2Yx2A for CED treatment of GBM in mice. 29
2.8 IDP-2Yx2A AF-488 NHS ester modification. 37
2.9 Analysis of AF-594 gly. 38
2.10 DLS measurements and SN38 release study 39

3.1 Generation of IDP variants with increased charges and decreasing lengths. 49
3.2 AFM analysis of IDP variants with decreasing lengths.AFM analysis of IDP variants with decreasing lengths. 51
3.3 Encapsulation of hydrophobic molecules in the micelle core. 53
3.4 Characterization of 0.25(Pro)-IDP-2Yx2A 55
3.5 Effect of increased temperature on IDP hydrophilic domain. 68
3.6 DLS of the averaged mean diameters of %Number of the 1.0-IDP hydrophilic control. 68
3.7 Intrinsically disordered profiles of IDP variants. 69

3.8 DLS of %Number 1.0-IDP-2Yx2A. 69
3.9 PDI, Z-average, and diffusion for all variants. 70
3.10 DLS of %Intensity IDP-2Yx2A mutants. 71
3.11 DLS of %Number for IDP-2Yx2A mutants. 71
3.12 DLS dilution of IDP variants. 72
3.13 Pyrene Assay for CMC determination of IDP-2Yx2A mutants. 72
3.14 Contact angle measurements of 1.0-IDP-2Yx2A and 0.5-IDP-2Yx2A mutants. 73
3.15 Tinopal OB®solubilization. 73

4.1 Bio-synthetic pathway for the production of pheomelanin from tyrosine. . 79
4.2 DLS analysis off %Number for the hydrophobic domain library. 82
4.3 Production of IDP-Y with a C-terminal tyrosine residue from an inclusion body expression system in *E. coli* 83
4.4 Organic solvent screen for megaTYR compatibility. 84
4.5 Mass Spectra of glutathione and dodecanethiol addition to IDP-Y. 84
4.6 Pyrene Assay for CMC determination of 1.0-IDP-2Yx2T variant. 95
4.8 C-terminal tyrosine residue with N-terminal cysteine peptides. 96
4.7 Analysis of fast and slow steps of phenomelanin type linkage formation. . 96
4.9 Stability of the 1.0-IDP-CIIIG conjugate in the presence of glutathione. . 97
4.10 Mass spectra for the organic solvent screen of IDP-Y oxidation with megaTYR. 98

List of Tables

1.1 Standard PCR Reaction Protocol 11
1.2 Standard Ligation Reaction Protocol 12

2.1 Drug loading and encapsulation efficiency of SN38. 39
2.2 Preliminary efficacy study: survival analysis. 40
2.3 BLI data for efficacy study. 40
2.4 Efficacy study: survival analysis. 40

3.1 Ionic surfactant HLB values 55
3.2 Standard PCR Reaction Protocol 65
3.3 Standard Ligation Reaction Protocol 65
3.4 CMCs derived from logistic fit of I1/I3 data. 73
3.5 CMCs derived from logistic fit of I1/ex data. 74

4.1 Hydrophobic library sequences. 80
4.2 Gibson Assembly Protocol 88

Acknowledgments

To all the amazing people in my life: I could not have done this without you.

Most importantly, I want to thank my mom, Laurie. Without you, I know I would not have been able to accomplish any of this. You have been supportive of everything I have ever done in my life, and it has given me the confidence to succeed at one of the most challenging experiences of my life. I cannot express how grateful I am for every concert, music lesson, sporting event, or dance show that you sacrificed your time to be there for me in support. I watched how hard you worked to support us growing up, and that has taught me the skills that I need to succeed in my own life today. I love you, and you are the most amazing woman I know. Next, I want to thank my sisters. Stacie and Katie, you have been so supportive of my scientific endeavors, and it feels so good to know that I have you both cheering me on. Growing up, I always looked up to both of you and I would not be the woman I am today without you. I love you both and I am excited to see what the future brings for us. I also want to thank my brother-in-law, Paul. You have also been a huge supporter of my scientific interests and I enjoy being able to discuss them in just a little more detail with you.

Next, I want to thank my husband, Parker. There are no words for how much your love and support mean to me. You have been here every step of the way and I could not have done this without you. Thank you for always making me coffee and food when I need a boost. Thank you for believing in all my abilities and not taking my insecure BS. You always show me a different way of thinking and approaching my problems. Thank you for being my partner and encouraging all of my pursuits. You make me a better and well-rounded person and I love you for it all.

I also want to thank my grandparents, Kay and Harvey Stover. While they are no longer with us, they were my biggest supporters. I only wish they could have seen me, graduate. To Grandmother, you were a strong woman who pushed me to excel and succeed in whatever I set my mind to. Your influence on those around you will persevere through decades if not generations. I miss you so much but I work every day to make you proud. To Grandpa, your silent love and adoration never went unnoticed. You were an amazing man and thank you for being so proud of me. Your unfaltering support made this scary step of my life worth it. I know there are more than a handful of strangers out there who know way more about my life than they ever wanted to because you couldn't help but share how proud you are of your family.

I have so much support from all my family that is hard to enumerate it all in detail. To my Aunt Ginny, there are not enough words to express how much your support and love mean to me. You are a strong woman that I have the privilege of

calling family and thank you for always being there for me. To my Uncle David, you were always a solid presence that made the world a little less scary. I miss you but I will always remember the way you were determined to keep us all laughing. To my Aunt Pam and Uncle Karl, thank you for all of your love and for the support of my scientific interests. Thank you for taking the time to even watch an absolutely terrible quality video of my graduate research presentation. To Dave, it has been a joy getting to know you and now calling you family. Your laid-back attitude and kindness are a calming presence I miss when I am not with you all. Thank you for being an amazing officiant and for always taking care of my aunt. To my cousins, Courtney Lourdes and Nikki, thank you for all your support and kind words over the years.

Next, I obviously have to thank my best friend, Allyssa. You are truly my biggest hypeman. Without fail, there is no way to feel like I am not capable of anything when I am with you. You are such an amazing woman and I am so grateful to have you. Your laugh is infectious and my life is all the more interesting with you in it. Thank you for listening to my ranting and venting and making me feel a little more sane. Thank you for allowing us to stay so close even with so much distance. I also want to thank Julie and Mark, you two are like a second family to me. Thank you for basically treating me as your own and showing me so much love and support.

I also have to thank my best friend, Kaila. We have been friends for so many years so thank you for standing by me and showing me so much love and support. Your personality is so special and you bring so much joy to the people in your life. Thank you for always being yourself and encouraging me to do the same. I cannot wait till we are in the same county again and we can act like nerdy 12-year-old girls again. I also want to thank Denise, you were a huge part of my life growing up, and thank you for all you did to take care of me. I am so grateful that you treated me like your own and gave me so much unconditional love and support.

I think it is very important that I thank my undergrad PI, Steve. Thank you for being an unfaltering supporter of my scientific abilities and I am so grateful I had you to push me to reach for everything that I have achieved since then. I truly would not be graduating with my Ph.D. if it weren't for you.

Next, I want to thank the friends that I have here in Berkeley. To Brittany, I am so grateful that our friendship moved with us from Long Beach. Thank you for all your support and advice, and I can't wait to see what the future holds for our friendship. To Nicole, you have quickly become one of my best friends and biggest confidants. Thank you for all the 15-minute coffee breaks that turned into 2 hours and all the lunch dates of endless venting. You are an amazing woman I know I can always turn to when I need support. To the girls, Kayli, Marisol, Kerry, and Susanna, thank you for all the wine nights and fun outings. Starting graduate school

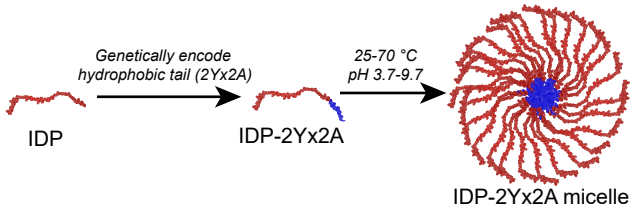
was a scary time and I am so glad that I got a chance to meet all of you and become fast friends.

Last but definitely not least, I want to thank all the lab members with whom I have overlapped. Starting with the boss who made this all possible: Matt, working with you has been a joy. You are a great mentor, boss, fellow coffee enthusiast, and wine snob. Your positivity and enthusiasm for what you do are contagious, and there was nothing that would brighten my day more than a 30-minute Matt Chat. The best ones were when we would get sidetracked about Dash, your 3D printer, or planning group events. You are one of the most supportive people I know, so thank you for just being you. Next up is Paul. I mean, there is so much I could say here. First of all, I would not be getting my Ph.D. without you. Thank you so much for our late-night venting sessions, boba runs, and doordash deliveries. Navigating the beginning of our graduate school experiences was just a little less daunting with you by my side. While starting off as a supportive colleague, you quickly became one of my best friends who will continue to be an essential part of my life even after we are both gone. I can't wait to see what your future holds, and I will be cheering you on from the sidelines. To Derek, thank you for being a fantastic friend and labmate. For being the person I could just vibe with and discuss x,y,z nerdy thing while loving on Cemi to give me my puppy fix. For being my hypeman and supporting me through all the difficult times over the past 4 years. Thank you for all the shared birthday memories and for cooking Puerto Rican food for me all the time. Sarah: you were an amazing mentor and someone I am so happy to be able to call my friend. Coming into the lab, having a brilliant female role model was crucial for me, and you definitely fit that bill. I am still in awe of all you are able to accomplish and thank you for all your support and for answering my endless questions. To Johnathan, thank you for making my rotation a wonderful experience. Working with you was a blast, and I am so glad I can call you my friend. Thank you for always taking an interest in my life, sharing photos of our cats, and allowing me to cry and go to boba when everything would get overwhelming. To Marco, I am so grateful to have spent as much time in the lab with you as I was able to. You were a true joy and always knew how to cheer up a room; I miss your random singing in the outpost and our coffee room chats, and I look forward to continuing to share pictures of our pets over Snapchat. To Dan, thank you for being a great colleague and friend. You supported me and helped me so much throughout my grad school experience. I am glad I spent time with you working on the climate survey and the diversity program. It was a blast trying to make the department a better place for our peers with you. To Amanda, thank you for being someone I am thrilled to call my friend. And for accepting Paul and me when we were being weird and sleep-deprived late at night because we were still in the lab. I hope we will continue to share book recommendations throughout

the years to come. To Alex, thank you for being an extraordinarily kind-hearted and down-to-earth individual. You were always willing to help me out and give me advice when I needed it. I am glad we got to work together for as long as we did. To Wendy, I am happy we got to join the lab together. The past 3-4 years of getting to know you have been really fun and I have enjoyed watching you grow into the amazing scientist you are today. To Vanessa, thank you for always lending an ear when needed and for giving such great advice. It was a blast getting to know you, and I am still in awe of your athletic capabilities. To Nick, I will never get over how you can eat so many pizza rolls. It took a while for us to get to know each other but moving to Stanley allowed that to happen, and I am very grateful. Thank you for having a great personality and being sometimes hilariously candid. To Paige, I was so impressed with your abilities during your rotation, and I was thrilled when you decided to join the lab. Your laugh is infectious, and I hope you continue to bring joy to those around you, especially in the lab when it is needed. To Deborah, thank you for being the happy and hilarious person we all needed in lab. Also, thank you for being one of the few coffee drinkers left! I am bummed we did not get to overlap more than we did. To Avery, your sneaky sense of humor has a way of lightening up a room, and I will miss seeing your amazing sense of style. To Arya, I am so glad that I got the chance to get to know you. You have been a great addition to the lab and thank you for kitty-sitting for us! To Kate and Hannah, I am so sad that we have only known each other for less than a year, but I can tell you both are going to be exceptional scientists because you already are. I am grateful that I got to know you both before moving on, and I hope this is only the beginning of our friendship!

Chapter 1

Insoluble Expression of Amphiphilic Intrinsically Disordered Proteins



Portions of this chapter are based on the following publication: Klass, S. H.*; Gleason J. M.*; Omole, A. O.; Onoa, B.; Francis, M. B. *Biomacromolecules* **2022**, *23*, 1462-1470.[1]

1.1 Abstract

Due to the widespread use of surfactants, there have been renewed efforts to develop bio-derived and bio-degradable alternatives. However, the large-scale production of designer bio-surfactants for medical or environmental purposes remains a challenge regarding economic viability compared to their synthetic counterparts. In this chapter, a scalable and versatile production method for bio-surfactants based on a hydrophilic intrinsically disordered protein (IDP) sequence with a genetically engineered hydrophobic domain (2Yx2A) is developed. This new expression system replaces the original solubilizing fusion protein, maltose binding protein (MBP), with the inclusion body directing fusion protein ketosteroid isomerase (KSI). By installing KSI at the N-terminus of IDP-2Yx2A, the entire fusion protein is sent to insoluble inclusion bodies during expression. Due to the insoluble nature of the fusion protein, it is easily purified by centrifugation after cell lysis. This system reduces the amount of solvent needed in the initial purification step and avoids stability issues encountered

with soluble proteins. A methionine residue is installed between the two protein domains to remove the KSI fusion protein from the IDP-2Yx2A sequence via cyanogen bromide (BrCN) cleavage. Overall, this represents a simplified and cost-effective method for producing the self-assembling IDP-2Yx2A sequence that we believe can also be applied to many different iterations and mutant sequences.

1.2 Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDRs) are defined by the lack of traditional secondary and tertiary structural motifs. Understanding these proteins has been vital in exploring their roles in cellular regulation, disease, and phase separation for forming membrane-less organelles.[2–9] Some of these behaviors have instigated use of IDPs as self-assembling proteins. In contrast to many self-assembling proteins, such as virus-like particles, IDPs do not rely on the direct interactions of amino acids with secondary, tertiary, and quaternary protein structures. Their assembly is driven largely by entropy and the hydrophobic effect.[10] Common classes of proteins gaining traction as assembling materials are elastin-like-polypeptides (ELPs) and the disordered regions of silk and reslin, which exhibit no defined secondary or tertiary structures at room temperature.[11] However, altering the environment (increasing the temperature, changing the pH, or conjugating a hydrophobic molecule) can induce these monomeric proteins to assemble into three-dimensional structures.[12] ELPs have well-understood properties and can be modified to adjust the temperature at which they self-assemble into spherical aggregates. Their repetitive sequence, Val-Pro-Gly-X-Gly, is derived from the hydrophobic domain of the natural biopolymer tropoelastin. The disordered nature is attributed to the abundance of proline and glycine residues, which are known to prevent the formation of stable hydrogen bonds and endow flexibility.[13–17].

IDPs mimic the disordered nature of the polymer chains of synthetic block-copolymers, thus making them promising building blocks for bio-derived and biodegradable alternatives. Unlike synthetic polymers, which generally comprise just one or two monomeric building blocks, IDPs can be produced from a toolbox consisting of the 20 naturally occurring amino acids, each with unique hydrophobicity, charge, and aromaticity properties. Additionally, protein-based polymers also benefit from precise monomer-level control over polymer chain length and position —a feat that is extremely difficult to achieve in synthetic polymerization reactions.[18] ELPs only constitute a small portion of all IDP sequences, which are abundant across all domains of life and display a diverse range of properties and functions.[19,20] One such IDP is a highly hydrophilic sequence derived from the neurofilament heavy arm side-

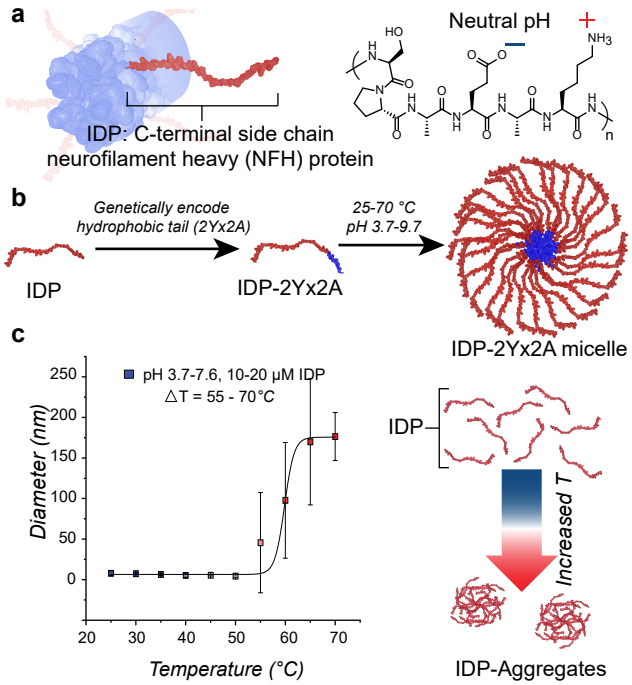


Figure 1.1: Generation of self-assembling IDP micelles. (a) The intrinsically disordered protein (IDP) domain derived from the human neurofilament protein protrudes from the neurofilament core. The IDP sequence is highly hydrophilic and is largely comprised of the repetitive motif SPXEXK, where X is G, A, or V. At neutral pH, this results in two oppositely charged residues per repeat: a negatively charged glutamic acid (E) and a positively charged lysine (K). (b) By genetically encoding a hydrophobic 16 amino acid segment (2Yx2A) at the C-terminus of IDP, the new amphiphilic protein IDP-2Yx2A was found to self-assemble spontaneously into ordered micelle structures in a range of buffer and temperature conditions.(c)When heated in aqueous solutions (pH 3.7-9.7, 0-200 mM PB), the monomeric IDP aggregates into larger structures.

chain protein found in human neurons (Figure 1.1a). This IDP has been shown to undergo expansion and contraction in response to charge alteration by pH, salt, and phosphorylation.[21–25] Furthermore, while the IDP protein exists as a monomer at room temperature, DLS analysis over a broad range of pH and buffer conditions indicates that the protein undergoes a thermo-responsive aggregation at temperatures above 50 °C (Figure 1.1c). This behavior is similar to commonly used hydrophilic polymers such as poly(ethylene oxide) (PEO) and ELPs whose thermally responsive self-assembling behavior have been characterized previously.[13–17,26]

Previous studies explored the IDP sequence as a hydrophilic domain for an amphiphilic biopolymer with a genetically encoded hydrophobic domain (2Yx2A) at the C-terminus of the protein. The production and characterization of this sequence

has been described previously.[27] Briefly, IDP-2Yx2A was solubly expressed in *E. coli* and after purification was found to undergo controlled self-assembly at room temperature to form nanoscale micellar structures in all tested buffer and temperature conditions (Figure 1.1b). The IDP-2Yx2A protein was expressed as a fusion with the solubilizing protein maltose binding protein (MBP).[28] The MBP fusion partner enabled the soluble expression of the MBP-thrombin cut site-IDP-2Yx2A sequence, which could undergo standard protein purification methods, including NiNTA chromatography, thrombin cleavage, and ion exchange to produce pure IDP-2Yx2A.[27] While these methods are suitable for initial laboratory-scale production, a simplified and less costly approach is necessary for industrial-scale production. To achieve this production scale, we replaced the MBP fusion partner with an inclusion body directing protein, ketosteroid isomerase (KSI) (Figure 1.2). The KSI fusion partner has shown promise in the insoluble expression of several shorter polypeptide sequences (<70 aa); however, this is the first instance where this protein has been applied to a longer self-assembling sequence.[29] By sending the IDP-2Yx2A sequence to inclusion bodies, issues such as proteolytic cleavage and stability that normally plague soluble protein expression are avoided. Sending the protein straight to inclusion bodies ensures that it is devoid of many contaminating proteins in the soluble fraction, leading to a relatively pure KSI-IDP-2Yx2A in the insoluble fraction by simple cell lysis and centrifugation steps, avoiding the need for extensive purification protocols. As an additional cost-saving measure, a cyanogen bromide (BrCN) cleavage site was installed in place of a thrombin cleavage site by encoding a methionine residue between the KSI and IDP-2Yx2A protein. As a final step, preparative HPLC purification is carried out on the crude material to yield the IDP-2Yx2A in high purity. Overall, this represents a simplified and cost-effective method for producing the self-assembling IDP-2Yx2A sequence that we believe can also be applied to many different iterations and mutant sequences.

1.3 Results and Discussion

To gain access to a wider variety of mutations, while also increasing yield, and lowering the cost of production, DNA encoding the previously explored IDP-2Yx2A sequence was inserted into a pET-31b(+) plasmid at the C-terminus of the inclusion body directing fusion protein KSI.[1,27] A single methionine residue was included between these regions to allow cleavage with BrCN (Figure 1.3a).[29] Protein expression was carried out in LB media at 18 °C for 16-24 h after induction by 0.5 mM IPTG. Due to the insoluble nature of the KSI fusion protein, protein purification was easily achieved as a series of centrifugation steps following cell lysis. This re-

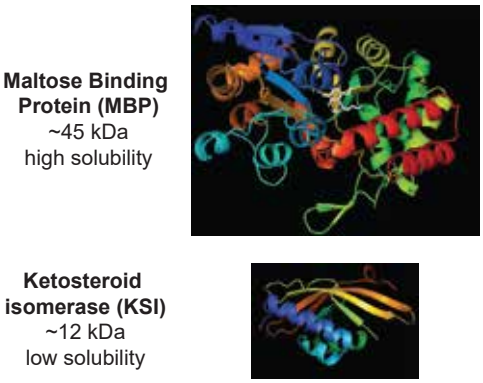


Figure 1.2: Structures for the Maltose Binding Protein (MBP) and Ketosteroid Isomerase (KSI) fusion partners. MBP is a highly soluble fusion partner of about 45 kDa used to aid expression and increase the solubility of protein in *E. coli*. KSI is a highly insoluble fusion partner of about 12 kDa that is used to direct an expressed protein to the inclusion bodies of *E. coli*. This protects bacteria from negative effects of the desired protein and allows for high yields of protein production.

sulted in an overabundance of the desired fusion protein KSI-IDP-2Yx2A in solution when analyzed by ESI-TOF-MS (Figure 1.3b). To remove the KSI fusion tag from IDP-2Yx2A, BrCN was used to cleave at the single Met residue between the two sequences. After cleavage, the desired IDP-2Yx2A protein was soluble in aqueous conditions while the KSI protein remained in the insoluble fraction and could be removed via centrifugation. The level of purity typically achieved after these subsequent rounds of precipitation and centrifugation is suitable for many applications, as it still shows assembly and surfactant behavior. Importantly, the purification process uses minimal solvent and is easily performed, making it amenable to scale up with protein yields between 100-300 mg/L. For this study, where precise measurements were sought, preparatory scale HPLC was performed as additional purification step (Figure 1.3c). When compared with the IDP-2Yx2A sequence produced by soluble expression, IDP-2Yx2A produced through this insoluble scheme showed similar size, morphology, and stability properties when analyzed by DLS, AFM, and pyrene fluorescence, respectively.[27]

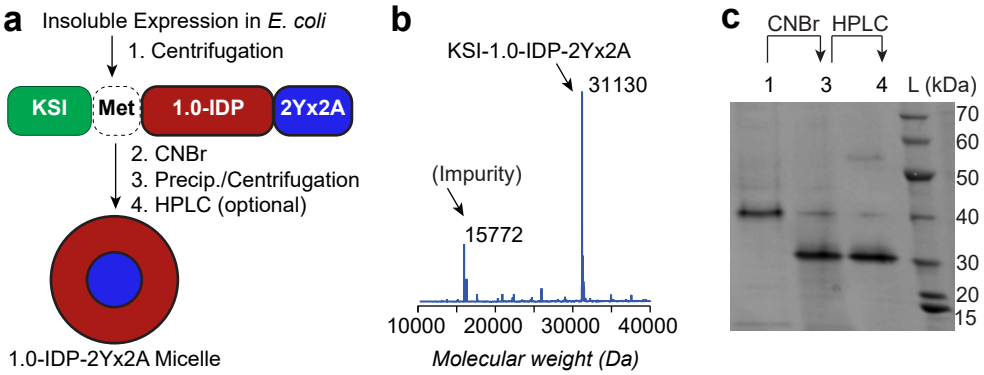


Figure 1.3: Insoluble expression of IDP-2Yx2A proteins. a) The IDP-2Yx2A is genetically encoded at the C-terminus of KSI following a single Met residue. Purification is achieved through a series of centrifugation and precipitation steps, followed by an optional HPLC purification step. b) LCMS analysis of the crude KSI-IDP-2Yx2A protein recovered from inclusion bodies after centrifugal purification is shown (pre-HPLC). Expected: 31132 Da, observed: 31130 Da. LCMS following KSI-cleavage and HPLC purification appear in Figure 3c. c) SDS PAGE gel analysis is shown for the purification steps outlined in (a). KSI-IDP-2Yx2A and IDP-2Yx2A have apparent molecular weights of 40 kDa and 30 kDa, respectively.

1.3.1 Protein Expression

KSI-IDP-2Yx2A was expressed using the following protocol. Plasmids containing the desired fusion proteins were transformed into Rosetta2 (pLysS) cells and plated onto carbenicillin(+)/chloramphenicol(+) LB Agar plates and grown overnight at 37 °C. Single colonies were selected and placed into 50 mL of LB with 50 µL of both 1000x stocks of carbenicillin and chloramphenicol antibiotics. The culture was grown overnight at 37 °C. The following morning, 30 mL of the 50 mL culture was used to innoculate 1000 mL of LB with 1000 µL 1000x stocks of carbenicillin and chloramphenicol in a 2.8 L baffled flask. Cultures were grown at 37 °C with shaking at 220 rpm until an OD_{600} of 0.8 was reached, at which point the temperature was lowered to 18 °C. After cooling for 20 min, IPTG was added to a final concentration of 0.25 mM. The protein was expressed overnight for 18-24 h. The culture was then centrifuged at 4000 xg for 20 min. The resulting protein pellets were transferred to 50 mL falcon tubes and frozen at -20 °C. A typical expression yields approximately 5-10 g cell pellets.

1.3.2 Protein Purification

KSI-IDP-2Yx2A was purified using the following protocol (Figure 1.4). The frozen cell pellet was resuspended in 30 mL of lysis buffer, 30 µL of 1% Trition-X, and

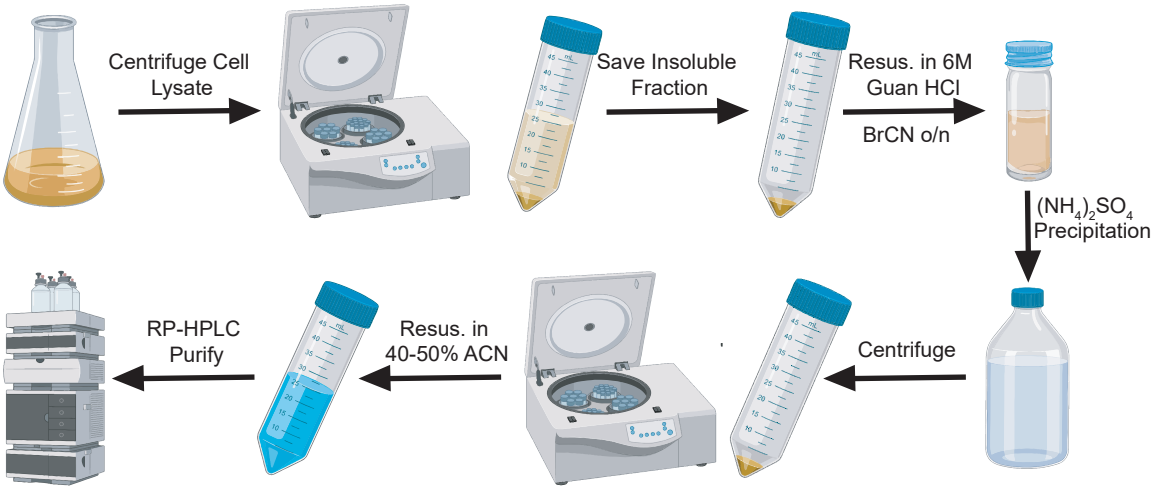


Figure 1.4: Schematic for insoluble expression and purification of IDP-2Yx2A. Rosetta2 (pLysS) cells containing the KSI-IDP-2Yx2A pet31b(+) plasmid were grown at 37 °C and protein production was induced and expressed at 18 °C. Following, the cells were lysed, centrifuged, and the insoluble fraction was save for resuspension in Guanidinium HCl and BrCN cleavage. All protein was precipitated in a saturated ammonia sulfate solution at -20 °C. Pelleted protein was collected and resuspended in 40-50% MeCN and purified using RP-HPLC.

300 µL of a 0.1 M PMSF solution (Sigma Aldrich). The cell pellet was vortexed until completely resuspended and then sonicated at 70% amplitude for 10 min (2 s on: 4 s off). The lysed cells were then centrifuged at 14,500 xg for 20 min. The supernatant containing the soluble proteins was discarded and the insoluble fraction was washed by resuspending it in 30 mL of lysis buffer and centrifuging again at 14,500 xg for 20 min. The washing step was repeated three times. After the third wash, 5 mL of 6 M guanidinium HCl (Guan HCl) was added to the insoluble fraction and the protein was resuspended by vigorous shaking for 1 h at room temperature and sonication at 70% amplitude for 10 min (2 s on: 4 s off). The protein solution was then centrifuged at 14,500 xg for 20 min at which point the KSI-IDP-2Yx2A fusion protein is solubilized, while the pelleted remains are largely composed of cell debris. The supernatant containing the fusion protein was transferred to a sterile 50 mL falcon tube. The approximate concentration was determined by measuring the UV absorbance at 280 nm on a NanoDrop 1000 (Thermo) and the calculated extinction coefficient (ExPASy Protein Parameters Tool https://web.expasy.org/protparam/) while the protein purity is assessed by SDS PAGE gel analysis.

The KSI protein was cleaved from the IDP-2Yx2A sequence by the slow addition of 3 M HCl, concentrated formic acid, 3 M BrCN to a final concentrations of 0.5 M HCl, 10% formic acid, and 100 molar equivalents of BrCN to protein. The solution

was stirred in the dark under a steady flow of nitrogen gas. After 24 h, the reaction was mostly complete when analyzed by ESI-TOF-MS. The protein was then precipitated from solution by the addition of cold saturated $(NH_4)_2SO_4$ solution. The precipitation solution was kept at -20 °C for 1 h. The precipitated protein was separated from the soluble components (BrCN, Guan HCl, and salts) by centrifugation at 14,500 xg for 20 min. The supernatant was discarded in a closed top waste receptacle specified for BrCN. The protein pellet was resuspended in H_2O or, for more hydrophobic constructs, up to 40% MeCN in H_2O by sonication at 70% amplitude for 10 min (2 s on: 4 s off) followed by a subsequent centrifugation at 14,500 xg for 20 min to remove any insoluble protein. The resulting protein solution was reasonably pure when analyzed by ESI-TOF-MS and SDS PAGE gel, however, an additional purification step using reverse phase HPLC (RP-HPLC) was performed to ensure maximum purity. RP-HPLC purification was performed in 1 mL injections on Waters XBridge Protein BEH C4 OBD Prep Column (300Å, 5 µm, 19 mm X 50 mm) with a 5-95% gradient of MeCN/H_2O +1% TFA over 30 min. Fractions containing the IDP-2Yx2A protein mutants eluted between 30-50% MeCN and were combined and lyophilized. Protein expression and purity in the soluble versus insoluble fractions was determined by NiNTA pull down as described previously and confirmed via SDS PAGE gel and ESI-TOF-MS analysis (Additional Figure 1.5).[27]

1.4 Conclusion

Although suitable for initial studies, the soluble expression system was not amenable to substantial mutations to the sequence space and thus limited the analysis of factors that impact the self-assembling properties of the micelles. By switching to an insoluble expression system, the yield, stability, and purification process improved. This system seems promising to have high tolerance to multisite mutations and deletions to the sequence space, resulting in significant changes to the protein net charge and molecular weight. The ability to alter and create a diverse sequence space is key to developing a systematic approach to understanding how alterations to the charge, relative hydrophobicity, and hydrophilic domain length affect the self-assembly behavior and physicochemical properties of these IDP-based bio-surfactants.

1.5 Materials and Methods

1.5.1 Materials

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Milli-Q (MQ) H_2O was purified to a resistivity of 18.2 M Ω • cm (at 25 °C) using a MQ Gradients ultrapure water purification system (Millipore). Itemized stock solutions, buffers, and reagents are detailed in the supporting information.

- 1000x ampicillin solution: 100 mg/mL of solid antibiotic were dissolved in Milli-Q H_2O. The solution was frozen at -20 °C until use.
- 1000x carbenicillin solution: 100 mg/mL of solid antibiotic were dissolved in Milli-Q H_2O. The solution was frozen at -20 °C until use.
- 1000x chloramphenicol solution: 25 mg/mL of slid antibiotic were dissolved in EtOH. The solution was frozen at -20 °C until use.
- LB agar plates: 12.5 g of LB broth (powder) from Thermo Fisher Scientific and 7.5 g agar was dissolved in 500 mL of Milli-Q H_2O and subsequently autoclaved. The solution was cooled to 50 °C and the desired antibiotic was added. The LB agar + antibiotic soultion was poured into sterile Petri dishes (100 mm x 15 mm) and solidified at room temperature. The plates were then stored at 4 °C until use.
- LB media: 20 g of Lauria broth (LB) powder (Sigma Aldrich) was dissolved into 1000 mL of Milli-Q H_2O and autoclaved.
- Lysis buffer: 20 mM Hepes, 300 mM NaCl, and 10 mM β-mercaptoethanol were added to Milli-Q H_2O.
- The pET-31b(+) vector was purchased from Millipore Sigma (catalog number: 69952-3).
- AlwN I (catalog #R0514S) and Xho I (catalog #R0146S) restriction endonucleases were purchased from New England BioLabs.
- Primers for overhang PCR were purchased from Integrated DNA Technologies (IDT).
- The Geneblock for IDP-2Yx2A was purchased from Integrated DNA Technologies(IDT) and was used as template DNA.

1.5.2 Instrumentation

1.5.2.1 Reverse Phase High Performance Liquid Chromatography (RP-HPLC)

RP-HPLC purification of IDP-2Yx2A protein mutants was performed on an Agilent 1260 Infinity series system equipped with a UV-Vis detector and a fraction collector, controlled by Open Lab CDS software (Agilent Technologies Inc. 2011 –2015). HPLC grade Milli-Q H_2O + 0.1% TFA and HPLC grade MeCN + 0.1% TFA were used as mobile phases A and B, respectively. Preparatory scale purification was performed on a Waters XBridge Protein BEH C4 OBD Prep Column (300 Å, 5 µm, 19 mm X 50 mm) with a flow rate of 5 mL/min over a 5-95% gradient over the course of 30 min. Fractions containing IDP-2Yx2A mutants were identified by their coinciding 280 nm absorbance and 350 nm fluorescence profiles.

1.5.2.2 Mass Spectrometry of Proteins (ESI-TOF-MS)

Time of flight electrospray ionization mass spectrometry (ESI-TOF-MS) of protein samples was performed using an Agilent 1260 series liquid chromatography (LC) system coupled with an Agilent 6224 TOF LC-MS (Santa Clara, CA). Protein samples were first passed through a 0.22 µM cellulose acetate filter and then placed in HPLC vials for analysis. Depending on sample concentration, 0.5 -10 µL was injected on to a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) analytical column with a flow rate of 0.30 mL/min and a gradient of 5-100% MeCN+0.1% (v/v) formic acid in Milli-Q H_2O+0.1% formic acid over the course of 8 min. The total ion chromatograms were extracted in the Agilent Mass Hunter Workstation software (Qualitative Analysis Version B.05.00, Build 5.0.519.13, Agilent Technologies Inc. 2011). Protein molecular weights were reconstructed using a maximum entropy deconvolution. The resulting spectra were exported as CSV files and plotted using the Chartograph web-tool(www.chartograph.com).

1.5.2.3 UV-VIS DNA and Protein Measurements

Estimates of DNA and protein concentration were calculated by measuring the sample absorbance at 260 nm and 280 nm, respectively, on a NanoDrop 1000 (Thermo). Using Beers law and the calculated extinction coefficients, the concentrations of both DNA and protein were approximated.

1.5.3 Methods of protein production

1.5.3.1 Methods for plasmid construction

PCR reactions were performed using the components as listed in Table 1.1. After mixing the reaction components in 0.5 mL PCR tubes, overhang PCR was performed on a BioRad S1000 Thermal Cycler with an annealing temperature of 55 °C. The PCR reaction was then removed from the thermocyler and 6 µL of 6X loading dye was added to the reaction tubes and mixed. The PCR reaction and a DNA standard(O'gene ruler 1 Kb) were then loaded into adjacent wells of a 1% agarose gel prestained with SYBR Safe (ThermoFisher). The gel was run at 120 V for 35 min in 1X TAE buffer. The resulting gel was imaged under UV fluorescence with a BioRad GelDoc EZ Imager. The DNA bands corresponding to the desired PCR product were excised from the gel and placed in a pre-weighed 1.5 mL Eppendorf tube. The amplified PCR product was removed from the gel using the quick gel extraction kit (Promega).

Table 1.1: Standard PCR Reaction Protocol

PCR Component	Volume (µL)
Template DNA[a] (50 ng/µL)	0.5
Forward Primer (10 µM)	1
Reverse Primer (10 µM)	1
5X phusion buffer	4
dNTPs (10 mM)	0.4
DMSO	0.6
Milli-Q H_2O	12
Phusion Enzyme[b]	0.5

[a] Geneblock (IDP-2Yx2A) [b] Phusion is added last

The PCR products and the pET-31b(+) entry vector were digested with 10 units of XhoI and AlwnI restriction enzymes for 1 h at 37 °C with a subsequent 20 min deactivation step at 80 °C. The cut entry vector and insert were then run on a 1% agarose gel prestained with SYBR Safe (ThermoFisher). The gel was run at 120 V for 35 min in 1X TAE buffer then imaged under UV fluorescence with a BioRad GelDoc EZ Imager. The DNA bands corresponding to the cut entry vector and insert were excised from the gel and placed in a pre-weighed 1.5 mL Eppendorf tube. The digested insert and vector were removed from the gel using the quick gel extraction

kit (Promega). The resulting complementary sticky ends of the entry vector and insert were then ligated together using the components listed in Table 1.2.

Table 1.2: Standard Ligation Reaction Protocol

Ligation Component	ratio insert : vector (0:1)	(3:1)	(5:1)
cut pET-31b(+)	50 ng	50 ng	50 ng
Insert (500 bp)	0 ng	15 ng	25 ng
10X T4 buffer	1 µL	1 µL	1 µL
Milli-Q H_2O	to 10 µL	to 10 µL	to 10 µL
T4 Ligase (400 units/µL) [b]	0.5 µL	0.5 µL	0.5 µL

[b] T4 ligase is added last

The ligation reaction was incubated at 16 °C for 16 h and then heat inactivated for 20 min at 80 °C. The entire ligation reaction was transformed into 50 µL of chemically competent XL1-Blue cells by first incubating the ligation reaction and cells on ice for 30 min and then incubating at 42 °C for 45 s. The cells were rescued by the addition of 1 mL of SOC media and incubated with rotary shaking at 37 °C. After 1 hour 200 µL of the transformed cells were plated on carbenicillin (+) LB agar plates and grown overnight at 37 °C. The number of colonies from on the plates where inset had been added (3:1) and (5:1) were compared to the growth of the control (0:1) to determine how many colonies to select for sequencing. An appropriate number of colonies was selected to account for background vector re-ligation and placed in individual 10 mL culture tubes containing approximately 5 mL of LB media and 5 µL of 1000x carbenicillin. Cultures were grown overnight at 37 °C. The following day, 1 mL of each culture was used to extract the plasmid DNA using a Miniprep Kit (Millipore), which were then sent for sequencing.

Primers for construction of KSI-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' GGTGGTCATATGCTGGCAGAGGCAAAG 3'
Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock IDP-2Yx2A: 5' AAAAATCAGATGCTGGCAGAGGCAAA
GAGTCCTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTC
GCGAAAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCG
GCAGAAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAA
AGCCCAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAG
CCAAGTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGC
CGAAGCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTC
CCCGGCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGT

AAAATCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGT
GACGGTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATC
ACCATACTGGAGTGCCTATGGCGCGTATGCGCAGTATGTGTATATTTAT
GCGTATTGGTATCTGTGACTCGAGTTTAAA 3'
AlwNI cut site shown in red, XhoI cut site shown in cyan

Corresponding amino acid sequence of fusion protein KSI-IDP-2Yx2A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYVYIYAYWYL*
KSI protein sequence shown in green

1.5.3.2 Optimizing protein expression

KSI-IDP-2Yx2A mutants expression was optimized using the following variables protocol. Plasmids containing the desired fusion proteins were transformed into Rosetta2 (pLysS) cells and plated onto carbenicillin(+)/chloramphenicol(+) LB Agar plates and grown overnight at 37 °C. Single colonies were selected and placed into 7 mL of LB or TB with 7 µL of both 1000x stocks of carbenicillin and chloramphenicol antibiotics. The culture was grown overnight at 37 °C. The following morning, 5 mL of the 7 mL culture was used to innoculate 100 mL of LB or 100 mL of TB with 100 µL 1000x stocks of carbenicillin and chloramphenicol in a 200m mL baffled flask. Cultures were grown at 37 °C with shaking at 220 rpm until an OD_{600} of 0.8 was reached, at which point the temperature was lowered to 18 °C. After cooling for 20 min, IPTG was added to a final concentration of 0.25 mM. The protein was expressed 6, 22, 30, and 48 h for each media type. The cell pellet collected for each time point was weighed and pellet weight was plotted against expression time (Additional Figure 1.6a). Each sample was lysed using the the method described previously and purified up to the BrCN step. Inclusion body pellets were weighed and the % of insoluble protein in the cell pellet was calculated (Additional Figure 1.6b). Approximately 100 µL of the protein solubilized in Guan HCl was then precitated into 1 mL of cold ethanol and resuspended into SDS for SDS-PAGE analysis (Additional Figure 1.7a). The molecular weight corresponding to the KSI-IDP-2Yx2A protein was confirmed by ESI-TOF-MS analysis (Additional Figure1.7b).

1.5.3.3 Protein Gel Analysis

Sodium dodecyl sulfate-polyacrylamide (SDS PAGE) gel analyses were run on a Novex apparatus using 4-12% precast linear gradient polyacrylamide gels (Bio-Rad, Hercules, CA). Protein samples were mixed with a 6X loading buffer and heated to 95 °C for 5 min. Samples were run for 30 min at 200 V in the SDS-MES buffer. Novex Sharp pre-stained protein standard was run on each gel for indication of molecular masses. Protein bands were stained with Coomassie Brilliant Blue R-250 (Bio-Rad) and gels were visualized on the Gel-Doc EZ Imager (Biorad).

1.6 Additional Figures

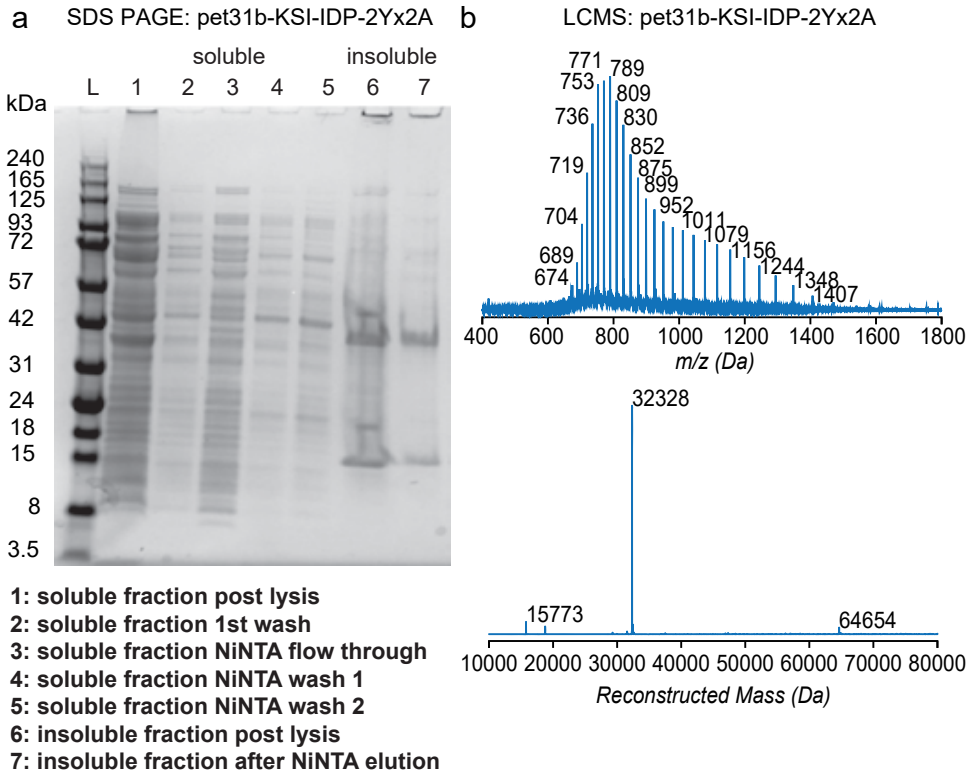


Figure 1.5: KSI-IDP-2Yx2A detection during protein purification. a) SDS PAGE analysis of KSI-IDP-2Yx2A-His_6. Lanes 1-5 are soluble protein fractions. Many soluble proteins are detected and small amounts of KSI-IDP-2Yx2a-His_6 are present. Lanes 6 and 7 are insoluble protein fractions where we see most of the KSI-IDP-2Yx2A-His_6 located. b) M/z and reconstructed mass (expected 32328 Da) spectra of the KSI-IDP-2Yx2A-His_6 fusion protein.

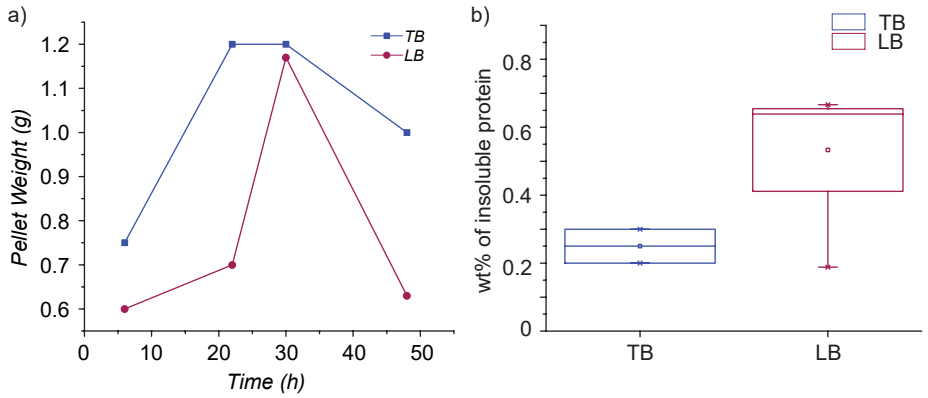


Figure 1.6: Yield of KSI-IDP-2Yx2A in LB vs. TB media. a) The weight of each collected cell pellet was measured and plotted against the corresponding expression time. LB: 6 h = 600 mg; 22 h = 700 mg; 30 h = 1170 mg; 48 h = 630 mg. TB = 6 h = 750 mg; 22 h = 1200 mg; 30 h = 1200 mg; 48 h = 1000 mg. b) The % by weight of protein in inclusion bodies in reference to the cell pellet was calculated for each expression and averaged. The average inclusion body content for TB expression was 25 ± 5 %. The average inclusion body content for LB expression was 53 ± 23 %. While the overall cell growth was higher for TB, LB was able to produce a higher percentage of insoluble protein material per cell pellet weight.

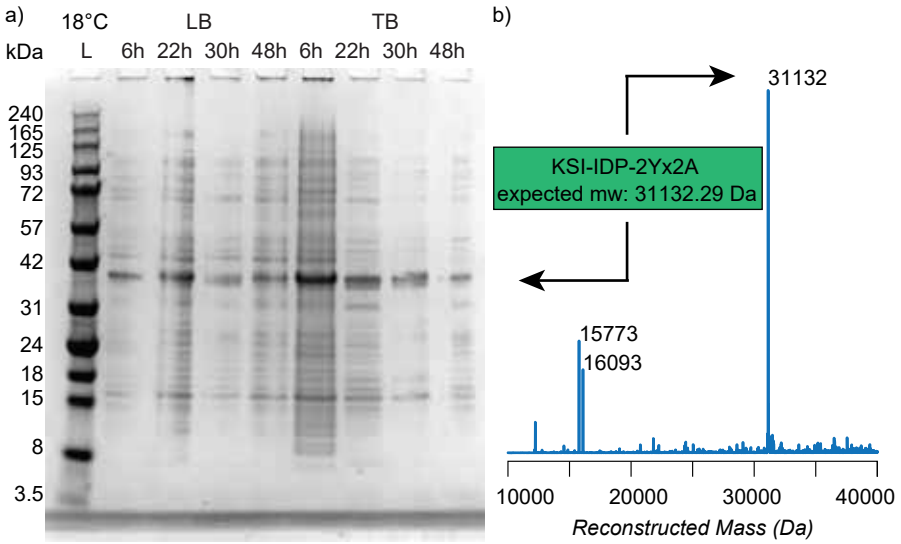


Figure 1.7: KSI-IDP-2Yx2A detection in LB vs. TB media. a) SDS PAGE analysis for determining presence of KSI-IDP-2Yx2A in the insoluble purification fraction at various time points of expression at 18 °C. Protein is detected at 6, 22, 30, and 48 h for both LB and TB media conditions. b) Mass spectra analysis confirms the gel band seen for KSI-IDP-2Yx2A is correct with an expected mw of 31132 Da. Suspected heat shock protein contaminants are seen at 15773 Da an 16093 Da.

1.7 Acknowledgements

This work was supported by the California research Alliance (CARA) by BASF, the UCSF Hana Jabsheh fund, National Institutes of Health Grant R01GM032543, and the Chemical Biology Graduate Program at UC Berkeley (NIH T32-GM066698). S.H.K. was supported by the University of California Cancer Research Coordinating Committee (UCCRC) predoctoral fellowship. J.M.G. was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program. Some figure elements were created with the online platform: BioRender.com.

1.8 References

(1) Klass, S. H.; Gleason, J. M.; Omole, A. O.; Onoa, B.; Bustamante, C. J.; Francis, M. B. *Biomacromolecules* **2022**, *23*, 1462–1470.

(2) Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L. M.; Cortese, M. S.; Lawson, J. D.; Brown, C. J.; Sikes, J. G.; Newton, C. D.; Dunker, A. K. *Bioinformatics* **2005**, *21*, 137–140.

(3) Varadi, M.; Vranken, W.; Guharoy, M.; Tompa, P. *Frontiers in molecular biosciences* **2015**, *2*, 45.

(4) Tompa, P. *Trends in Biochemical Sciences* **2012**, *37*, 509–516.

(5) Monastyrskyy, B.; Fidelis, K.; Moult, J.; Tramontano, A.; Kryshtafovych, A. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 107–118.

(6) Wright, P. E.; Dyson, H. J. *Nature Reviews Molecular Cell Biology* **2015**, *16*, 18–29.

(7) Uversky, V. N. *Current Opinion in Structural Biology* **2017**, *44*, 18–30.

(8) Uversky, V. N.; Kuznetsova, I. M.; Turoverov, K. K.; Zaslavsky, B. *FEBS Letters* **2015**, *589*, 15–22.

(9) Dzuricky, M.; Roberts, S.; Chilkoti, A. *Biochemistry* **2018**, *57*, 2405–2414.

(10) Gonzalez-Obeso, C.; González-Pérez, M.; Mano, J. F.; Alonso, M.; Rodríguez-Cabello, J. C. *Small* **2020**, DOI: `https://doi.org/10.1002/smll.202005191`.

(11) Garcia Quiroz, F.; Li, N. K.; Roberts, S.; Weber, P.; Dzuricky, M.; Weitzhandler, I.; Yingling, Y. G.; Chilkoti, A. *Science Advances* **2019**, *5*, eaax–5177.

(12) Girotti, A.; Reguera, J.; Arias, F. J.; Alonso, M.; Testera, A. M.; Rodríguez-Cabello, J. C. *Macromolecules* **2004**, *37*, 3396–3400.

(13) Keeley, F. W.; Bellingham, C. M.; Woodhouse, K. A. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2002**, *357*, 185–9.

(14) Wright, E. R.; Conticello, V. P. *Advanced Drug Delivery Reviews* **2002**, *54*, 1057–1073.

(15) Dreher, M. R.; Simnick, A. J.; Fischer, K.; Smith, R. J.; Patel, A.; Schmidt, M.; Chilkoti, A. *Journal of the American Chemical Society* **2008**, *130*, PMID: 18085778, 687–694.

(16) Ribeiro, A.; Arias, F. J.; Reguera, J.; Alonso, M.; Rodríguez-Cabello, J. C. *Biophysical Journal* **2009**, *97*, 312–320.

(17) Li, N. K.; Roberts, S.; Quiroz, F. G.; Chilkoti, A.; Yingling, Y. G. *Biomacromolecules* **2018**, *19*, 2496–2505.

(18) Ruzette, A. V.; Leibler, L. *Nature Materials* **2005**, *4*, 19–31.

(19) Holehouse, A. S.; Das, R. K.; Ahad, J. N.; Richardson, M. O.; Pappu, R. V. *Biophysical Journal* **2017**, *112*, 16–21.

(20) Peng, Z.; Yan, J.; Fan, X.; Mizianty, M. J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V. N.; Kurgan, L. *Cellular and Molecular Life Sciences* **2014**, *72*, 137–151.

(21) Adiga, S. P.; Brenner, D. W. *The Journal of Physical Chemistry C* **2010**, *114*, 5410–5416.

(22) Ackerley, S.; Thornhill, P.; Grierson, A. J.; Brownlees, J.; Anderton, B. H.; Leigh, P. N.; Shaw, C. E.; Miller, C. C. J. *The Journal of cell biology* **2003**, *161*, 489–95.

(23) Bhagawati, M.; Rubashkin, M. G.; Lee, J. P.; Ananthanarayanan, B.; Weaver, V. M.; Kumar, S. *Langmuir* **2016**, *32*, 5990–5996.

(24) Chang, R.; Kwak, Y.; Gebremichael, Y. *Journal of Molecular Biology* **2009**, *391*, 648–660.

(25) Lees, J. F.; Shneidman, P. S.; Skuntz, S. F.; Carden, M. J.; Lazzarini, R. A. *The EMBO journal* **1988**, *7*, 1947–55.

(26) Hocine, S.; Li, M.-H. *Soft Matter* **2013**, *9*, 5839–5861.

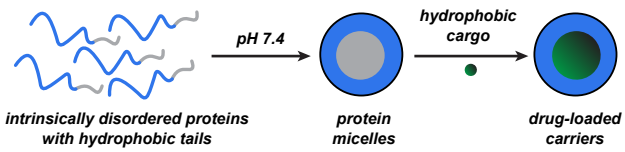
(27) Klass, S. H.; Smith, M. J.; Fiala, T. A.; Lee, J. P.; Omole, A. O.; Han, B.-G.; Downing, K. H.; Kumar, S.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 4291–4299.

(28) Raran-Kurussi, S.; Waugh, D. S. *Biotechnology and bioengineering* **2014**, *111*, 2407–11.

(29) Kuliopulos, A.; Walsh, C. T. *Journal of the American Chemical Society* **1994**, *116*, 4599–4607.

Chapter 2

In vivo Characterization of Intrinsically Disordered Proteins as Vehicles for Delivery to Glioblastoma Tumors



This chapter is based on the following publication: Gleason J. M.*; Klass, S. H.; Huang, P.; Ozawa, T.: Santos, R. S.; Fogarty, M. M.; Raleigh, D. R.; Berger, M. S.; Francis, M. B. *ACS Applied Bio Materials* **2022**.[1]

2.1 Abstract

Lipid and micelle-based nanocarriers have been explored for anticancer drug delivery to improve accumulation and uptake in tumor tissue. As an experimental opportunity in this area, our lab has developed a protein-based micelle nanocarrier consisting of a hydrophilic intrinsically disordered protein (IDP) domain bound to a hydrophobic tail, termed IDP-2Yx2A. This construct can be used to encapsulate hydrophobic chemotherapeutics that would otherwise be too insoluble in water to be administered. In this study, we evaluate the *in vivo* efficacy of IDP-2Yx2A by delivering a highly potent but water-insoluble cancer drug, SN38, into glioblastoma multiforme (GBM) tumors via convection-enhanced delivery (CED). The protein carriers alone are shown to elicit minimal toxicity effects in mice; furthermore, they can encapsu-

late and deliver concentrations of SN38 that would otherwise be lethal without the carriers. CED administration of these drug-loaded micelles into mice bearing U251-MG GBM xenografts resulted in slowed tumor growth and significant increases in median survival times compared to non-encapsulated SN38 and PBS controls.

2.2 Introduction

Many nanocarrier systems have been explored as drug delivery agents to increase drug accumulation in tumor tissue.[2–6] Of these, liposome and micelle structures have been used frequently due to their ability to house internal reservoirs of drug molecules.[7–11] Many of these platforms consist of self-assembling synthetic block copolymers, but these materials are often polydisperse and difficult to tune for optimal delivery. Furthermore, some of these formulations have been shown to have unpredictable and undesirable toxicity effects.[8,12–16] These challenges can potentially be bypassed by using protein-based nanoparticles, which can be produced reliably by inexpensive recombinant expression and designed to self-assemble into nanoscale structures composed of multiple copies of identical protein monomers.[17,18] An intrinsic advantage of protein-based nanoparticles is that they biodegrade in the body and have shown few toxicity issues.[19,20] Additionally, these materials can be specifically designed with control over every amino acid along the polypeptide chain using genetic methods.[21–29] Proteins can also be site-selectively modified with chemotherapeutic cargoes, targeting groups, and PEG polymers using mild bioconjugation reactions that do not interfere with nanoparticle self-assembly.[30–33] However, some drugs are not amenable to bioconjugation attachment strategies as they do not posses functional groups that are easily turned into attachment sites. Alternatively, chemotherapeutic cargoes can be loaded in a non-covalent fashion into the interior of protein-based nanoparticles through a number of approaches depending on the protein and drug properties. This approach, is especially attractive for the loading hydrophobic drugs into intrinsically disordered protein micelles that possess a hydrophobic core.[34]

By combining the hydrophilic IDP region with a hydrophobic protein sequence as a protein fusion, termed here as IDP-2Yx2A (Figure 2.1a), the proteins self-assemble into micelles that can solubilize and accumulate hydrophobic small-molecules within their cores (Figure 2.1b). Previously, the critical micelle concentration (CMC) of IDP-2Yx2A has been reported in the low µM range (2-10 µM).[34,35] These materials could serve as promising alternatives to synthetic liposome and micelle structures for the enhanced delivery of hydrophobic drugs into tumor tissue; however, there has been no previous characterization of their *in vivo* properties.

Glioblastoma multiforme (GBM) was selected as a specific target for these stud-

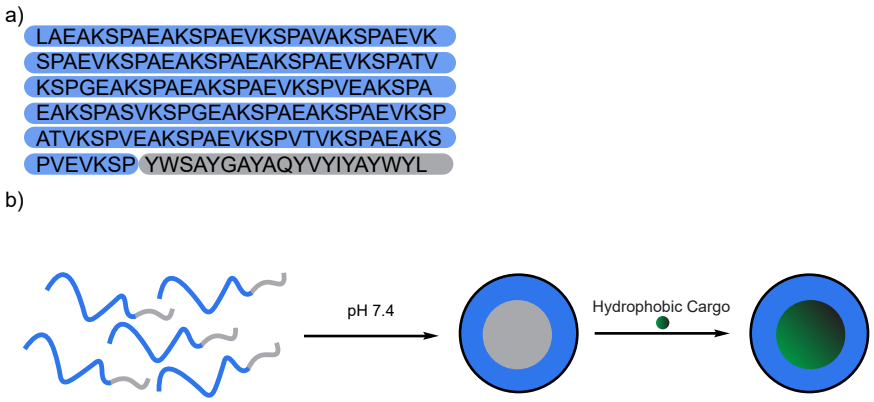


Figure 2.1: IDP micelles encapsulate hydrophobic cargo. a) The amino acid sequence of the IDP-2Yx2A protein is shown.[35] The blue sequence composes the hydrophilic domain and the gray segment indicates the hydrophobic amino acid tail. b) In aqueous solutions above the CMC, IDP-2Yx2A monomers self-assemble into spherical micelles approximately 30 nm in diameter. The hydrophobic cargo is loaded into the core of the micelles for delivery as a chemotherapeutic agent.

ies. GBM remains one of the most challenging cancers to treat despite the development of new therapeutics and a growing number of ways to administer them.[36–43] Convection-enhanced delivery (CED) has been investigated as a promising strategy to bypass the blood-brain barrier (BBB), and is the technique used in these studies.[44–47] Briefly, this method injects therapeutics directly at the tumor site via a constant pressure microfluidic pump. The continuous pressure creates fluid convection and aids the diffusion of therapeutics throughout the targeted tissue.[36,44,48] The delivery of micelle-based drug formulations via CED can greatly increase both the drug payload that reaches the tumor and its retention before clearance.[42,49–53] In this context, we believe low toxicity, biodegradability, and low CMC values offer a particular advantage for delivery with the IDP-2Yx2A system. Additionally, this system includes the potential for introducing precise site modifications. This allows for the introduction of specific GBM targeting groups as well as modification of the assembly sequence to control drug loading and release, making genetically engineered IDPs adaptable to a highly heterogeneous and fast-evolving cancer.

Here, we show that IDP-2Yx2A protein micelles are indeed compatible with the CED technique, showing little-to-no toxicity for the carriers themselves upon administration to mice. They allow quantities of SN38, a potent and hydrophobic anticancer drug, to be administered that would be lethal in the absence of the carriers.[54–58] Most importantly, statistically significant survival increases were observed when mice bearing GBM xenografts were treated with these agents. These results establish the IDP-2Yx2A carrier system as a promising new platform for directly

administering many drug molecules that would otherwise be too hydrophobic to use for cancer treatment.

2.3 Results and Discussion

2.3.1 IDP-2Yx2A *in vitro* characterization

The IDP-2Yx2A micelles were checked for cell toxicity before being used as drug delivery vehicles. Protein expression in *E. coli* invariably results in contamination with endotoxic lipopolysaccharides from the bacterial outer membrane, which must be removed from the protein sample prior to *in vivo* administration to prevent unwanted immunogenic effects.[59] Endotoxin levels were measured using the LAL detection kit as all protein was expressed and purified from *E. coli*. Endotoxin levels of less than 0.1 EU/mL, near the lower limit of detection, were observed for the RP-HPLC purified IDP-2Yx2A micelles (Figure 2.2a). Following this, we determined unformulated IDP-2Yx2A cell toxicity via an MTS cell viability assay. Non-cancerous embryonic kidney (HEK293) cells were used to examine innate toxicity. After a 72 h of incubation with IDP-2Yx2A, no significant decrease in cell viability was observed up to a concentration of 20 µM, and a minimal reduction was seen up to 40 µM (Figure 2.2b). These findings indicated that 20 µM could be an appropriate protein concentration for further cellular and *in vivo* studies.

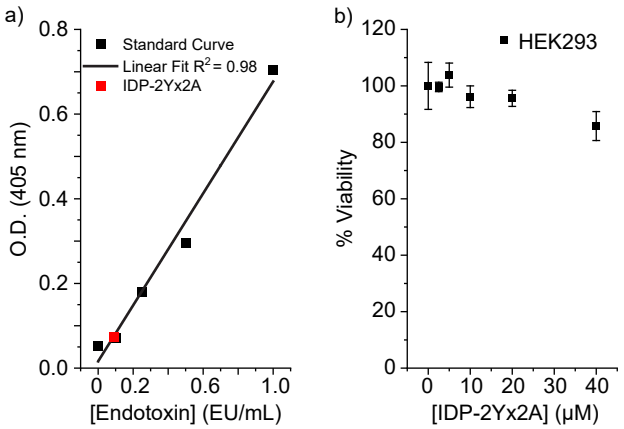


Figure 2.2: *In vitro* toxicity characterization of IDP-2Yx2A. a) Endotoxin quantification of IDP-2Yx2A post-HPLC purification was measured at less than 0.1 EU/mL, the lower detection limit for the Pierce LAL chromogenic endotoxin quantification kit. b) The cell toxicity of IDP-2Yx2A was determined using an MTS assay with HEK 293 cells. No significant changes in cell viability were seen up to 20 µM IDP-2Yx2A and a slight drop in cell viability was seen at 40 µM IDP-2Yx2A.

Release kinetics of IDP-2Yx2A were measured under controlled conditions (10% FBS, 37 °C) to monitor the release of hydrophobic cargo in the presence of relevant biological reservoirs. IDP-2Yx2A was covalently modified with Alexa Fluor(AF) 488 NHS ester on its lysine residues at approximately 12% labeling (Additional Figure 2.8c). Assembly was confirmed by dynamic light scattering (DLS, Additional Figure

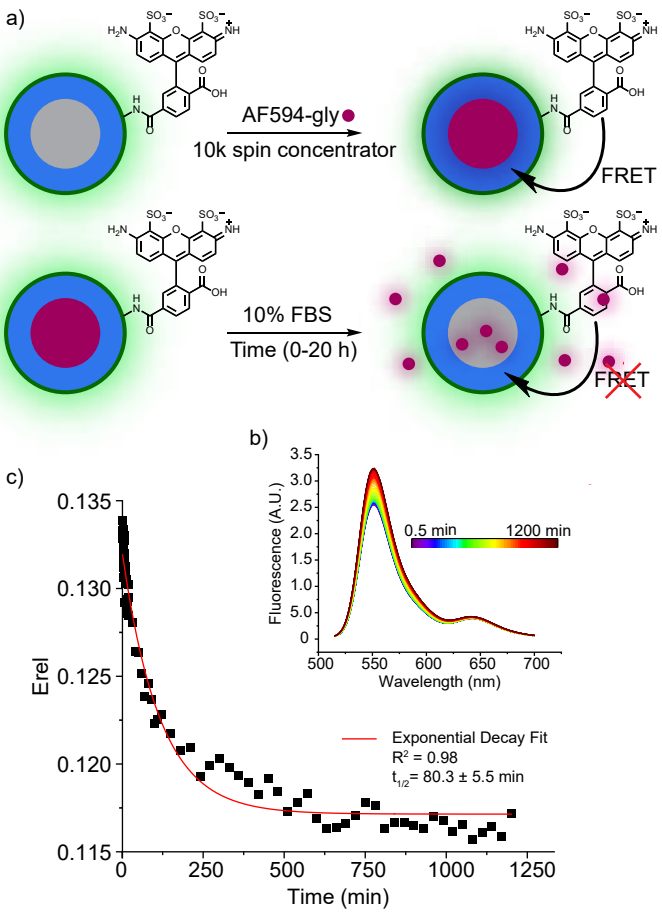


Figure 2.3: Hydrophobic cargo release kinetics of IDP-2Yx2A. a) AF594-gly was loaded into the IDP-2Yx2A-AF488 micelle core and excess dye was removed with a 10k spin concentrator. Fresh buffer with 10% FBS was introduced to instigate the release of the hydrophobic cargo dye. b) The fluorescence intensity (515-700 nm) was monitored for 20 h post FBS introduction. c) The FRET energy decrease was calculated and fit to an exponential decay function. The half-life of cargo release was determined to be approximately 80 min after FBS addition.

2.8b) and the micelles were loaded with a modified hydrophobic AF-594 (Additional Figure 2.9a). When the fluorescent green protein was loaded with a red fluorescent dye, fluorescence resonance energy transfer (FRET) occurred in the sample. However, as AF-594 was released from the micelle core, a decrease in FRET was observed. Dye-modified IDP-2Yx2A was introduced to 10% FBS to represent the conditions these micelles could experience in cell culture (Figure 2.3a). The micelle structures were stable in solution, but in the presence of a hydrophobic carrier, such as albumins in cell culture or cell membranes and lipophilic tissue *in vivo*, the hydrophobic cargo

would be expected to release from the core and transfer to other reservoirs. FRET was measured using a fluorometer (Figure 2.3b), and the FRET efficiency (E_{rel}) was calculated. The resulting data were then fit to an exponential decay function (Figure 2.3c). The half-life of dye release to an equilibrium state between IDP-2Yx2A and 10% FBS was estimated to be 80 min. The calculated half-life shows the release of cargo from the hydrophobic core to nearby hydrophobic reservoirs is not immediate and on a timescale suitable for biological applications under physiologically relevant conditions in the brain.[60]

2.3.2 Solubilizing SN38

The ability of IDP-2Yx2A to encapsulate and gradually release hydrophobic small-molecule cargo makes it an ideal carrier for drugs that observe high *in vitro* target binding but are precluded from clinical use due to their poor water solubility and resulting low bioavailability. One chemotherapeutic that has shown promise for treating many cancers is the topoisomerase inhibitor SN38, which has EC_{50} values in the low nM range for many cancer cell lines.[61–64] Unfortunately, SN38 exhibits extremely low solubility in water and therefore must be administered as a modified pro-drug, irinotecan (CPT-11). The pro-drug form is 100-1000 times less cytotoxic than SN38 until it is activated via hydrolysis by carboxylesterase enzymes in the liver (Figure 2.4a). Although this releases SN38, its activation requires systemic administration at much larger dosages, significantly increasing the risk of toxicity and off-target effects.[65–69] Lipid-based nanoparticle formulations have been developed to encapsulate CPT-11 to reduce systemic toxicity and have shown promise for treating neuroblastoma brain tumors. To reduce systemic toxicity, lipid-based nanoparticle formulations have been developed to encapsulate irinotecan[70] and have shown promise for the treatment of neuroblastoma brain tumors.[71] Additionally, a synthetic polymeric micelle formulation of SN38 (known as NK012) has been studied for its efficacy in Glioblastoma multiforme (GMB) brain tumors and was shown to increase survival and decrease the tumor volume for mice.[72] Additionally, SN38 has been shown to increase survival and decrease tumor volume in animal glioma models using a hydrolyzable covalent linkage to a synthetic block copolymer consisting of poly(ethylene glycol) and poly(glutamic acid) regions. In this case, the attachment of the drug molecule to the glutamate residues resulted in micelle formation.[51]. SN38 has a hydrophobic nature that makes delivery problematic, however it allows for encapsulation with the core of the IDP-2Yx2A micelle to create the formulation termed here as SN38-IDP-2Yx2A.

In this work, the formulation of SN38-IDP-2Yx2A was created by solubilizing SN38 into the core of the assembled IDP-2Yx2A micelles. Lyophilized IDP-2Yx2A

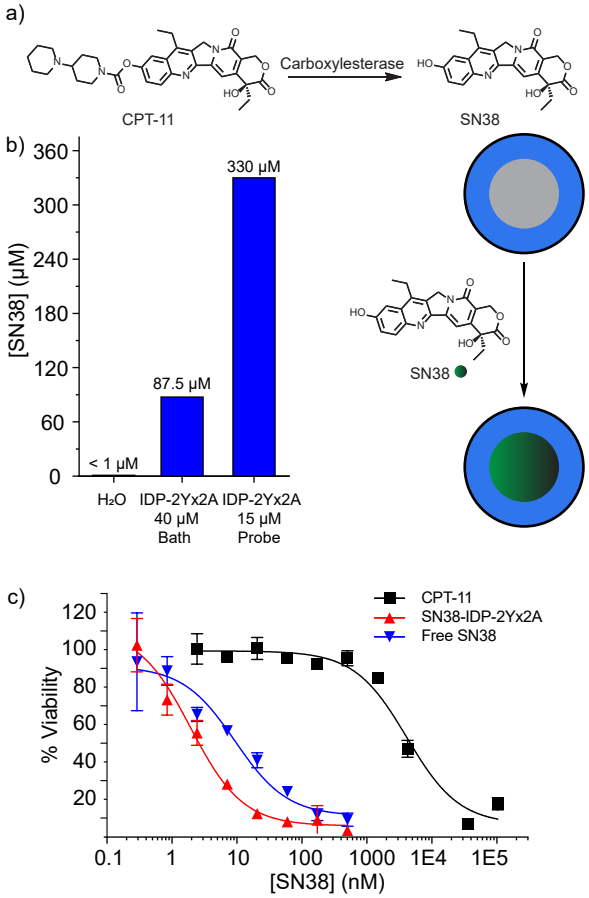


Figure 2.4: SN38 solubilization and formulated SN38-IDP-2Yx2A cell viability. a) CPT-11, the pro-drug for SN38, is metabolized by carboxylesterases in the liver to give the active metabolite. b) The solubility of SN38 increases in the presence of IDP-2Yx2A. In H_2O, the solubility of free SN38 is at the limit of detection by absorbance and is estimated to be less than one μM. Using a bath sonicator, up to 87.5 μM is solubilized by the micelles. Using a probe sonicator during the loading protocol increased the solubility to a maximum of 330 μM. c) The cell viability of U251-MG-FL cells after 72 h exposure to formulated SN38-IDP-2Yx2A was measured and fit to a logistic function. The EC_{50} value of SN38-IDP-2Yx2A was calculated to be 2.5 ± 0.5 nM. Free SN38 was calculated to have an EC_{50} value of 13.5 ± 4.3 nM while the CPT-11 was approximately 1000x less potent with an EC_{50} of 9720 ± 1650 nM.

was reconstituted in 50% MeCN:Milli-Q H_2O to a concentration of 10 μM and 500 μM SN38 from a DMSO stock was added dropwise to the solution. The solution was mixed using a probe sonicator (2 s on 4 s off, 65% amplitude) for 2 min. The MeCN was slowly evaporated under N_2, and the sample was lyophilized to remove the remaining H_2O and DMSO. The solution was reconstituted in DPBS

and filtered through a 0.22 µm cellulose acetate membrane to sterilize the solution and remove any aggregate particles. The concentrations of SN38 and IDP-2Yx2A were determined via absorbance using their calculated extinction coefficients (Figure 2.4b). The concentration of solubilized SN38 was compared to the concentration dissolved in DPBS with no IDP-2Yx2A. The maximum calculated concentration for SN38 with IDP-2Yx2A was 330 µM, but the concentration of SN38 in buffer alone was lower than could be calculated with confidence. This represents a greater than 300-fold increase in solubility when formulated with the IDP carriers. The drug loading and encapsulation efficiency values of SN38 were calculated to be 32.2% and 57.6%, respectively (Additional Table 2.1). In related works, DLS measurements confirmed that micelles were still assembling with a narrow polydispersity distribution at around 24 nm (Additional Figure 2.10a). The release and diffusion across a membrane of free SN38 versus SN38-IDP-2Yx2A are included as additinoal figures. The $t_{1/2}$ of SN38 diffusion was found to be 1.6 h and SN38-IDP-2Yx2A was approximately 2.8 times slower, with $t_{1/2}$ = 4.5 h (Additional Figure 2.10c). In previous reports, cryo-TEM and atomic force microscopy were used to confirm micelle assembly and apparent morphology for unloaded IDP-2Yx2A and in the presence of hydrophobic cargo, specifically SN38.[34,35]

2.3.3 SN38-IDP-2Yx2A *in vitro* studies

Cell viability experiments were conducted on human glioblastoma cells (U251-MG) to evaluate SN38-IDP-2Yx2A as a cytotoxic agent for glioblastoma. SN38-IDP-2Yx2A was compared to free SN38 and its pro-drug, CPT-11. Concentration gradients of SN38-IDP-2Yx2A, SN38, and CPT-11 were incubated with U251-MG-FL cells for 72 h. Cell viability was determined by an MTS assay, which quantifies the number of live cells in culture through the NAD(P)H-dependent conversion of the MTS tetrazolium compound by dehydrogenase enzymes in metabolically active cells. The half-maximal effective concentration, EC_{50}, was determined by fitting a logistic curve to the drug concentration vs. % viability data, which resulted in a calculated EC_{50} of 2.5 nM for SN38-IDP-2Yx2A. The EC_{50} value of SN38 with no carrier was 13.5 nM and CPT-11 had an EC_{50} of 9720 nM, about 1000x larger, as would be expected for the pro-drug form (Figure 2.4c). IDP-2Yx2A controls confirmed that unloaded protein micelles do not individually show cytotoxic effects to glioblastoma cancer cells and non-cancerous cell lines, as mentioned previously.

2.3.4 Evaluating SN38-IDP-2Yx2A *in vivo* convection-enhanced delivery

Pilot mouse toxicity studies were performed to determine if minimal cell cytotoxicity translated to biocompatibility *in vivo*. Three mice were injected intracranially with 10 µL of 20 µM IDP-2Yx2A DPBS and their weight was measured daily for one week to monitor for any adverse reactions to the injection (Figure 2.5a). No toxicity effects were seen in response to 20 µM IDP-2Yx2A, so it was determined safe to create drug formulations up to this concentration for efficacy studies. Additionally, toxicity studies with SN38 in 50% DMSO and vehicle controls of 50% DMSO with no drug (Figure 2.5a) were done to determine optimal loading concentrations for *in vivo* efficacy studies. Three healthy mice per arm were injected with 200 µM, 100 µM, and 75 µM free SN38 to determine the highest concentration safe for CED infusions. After 1-2 days, mice injected with 200 µM and 100 µM SN38 started to show neurological toxicity symptoms and were euthanized (Figure 2.5b). In contrast, mice injected with 75 µM SN38 showed no toxicity symptoms, and their body weight was monitored and determined to be in a normal range for one week.

Confident that drug and protein would cause no immediate toxicity to mice, two courses of a three-arm (n=6) efficacy study were performed on glioblastoma (U251-

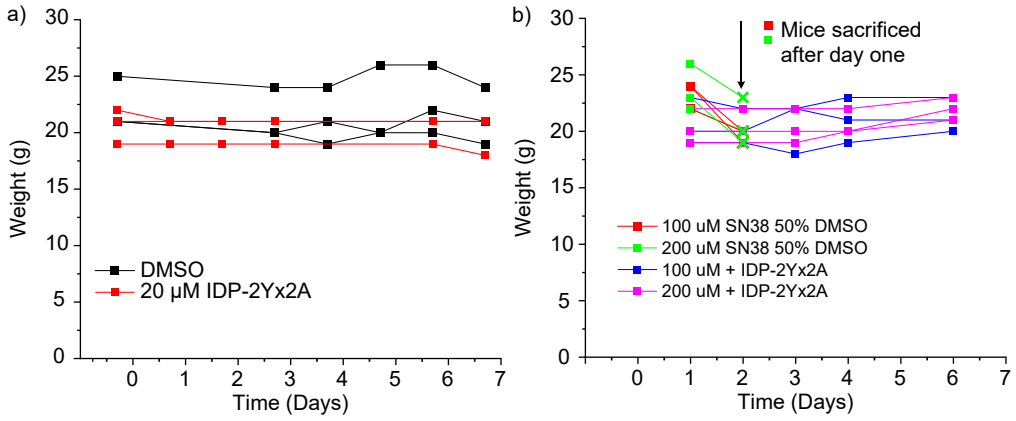


Figure 2.5: *In vivo* toxicity studies for intracranial injection of SN38-IDP-2Yx2A. a) Three healthy mice were injected with 10 µL of DMSO or 20 µM of HPLC-purified IDP-2Yx2A. They were monitored for one week to observe any adverse effects, such as weight loss or neurological symptoms. No significant decrease in body weight was observed. b) For each arm, three healthy female mice were injected with 10 µL of 100 µM SN38-50%DMSO, 200 µM SN38-50%DMSO, 100 µM SN38-IDP-2Yx2A, and 200 µM SN38-IDP-2Yx2A. Mice for both 100 µM SN38 and 200 µM SN38-50% DMSO displayed neurological toxicity symptoms within one day of injection and had to be euthanized. No effects were seen with formulated SN38-IDP-2Yx2A and body weight remained constant.

MG-FL) tumor-bearing mice. In the first round, SN38-IDP-2Yx2A was formulated with 75 µM drug and approximately 15 µM IDP-2Yx2A due to initial toxicity seen at 100 µM and 200 µM SN38. DPBS and 75 µM SN38 in 50% DMSO were included as controls for the study. CED infusions of 10 µL were conducted 15 days post-implantation, and subsequent infusions were performed every seven days for four weeks. Tumor size was monitored by bioluminescence intensity (BLI) (Figure 2.6a), and Kaplan-Meier analysis and log-rank tests were used to determine statistical significance for mouse survival (Figure 2.6b). Mice treated with SN38-IDP-2Yx2A showed statistically significant increased survival ($p = 0.032$) compared to mice treated with the SN38 in 50% DMSO control (Additional Table 2.2). The initial study showed promise for using formulated SN38-IDP-2Yx2A as a glioblastoma treatment to prolong mouse survival and inhibit tumor growth.

The toxicity study was repeated on healthy mice using 200 µM and 100 µM of SN38 formulated with IDP-2Yx2A instead of 50% DMSO to determine if loading in the micelles could protect against SN38 toxicity. Three mice per arm were injected intracranially with 10 µL of agent, and their weights were monitored for one week to determine if any adverse reactions or neurological symptoms developed. No toxicity effects were seen at SN38 concentrations well above tolerable levels for the free drug, suggesting that the IDP-2Yx2A protein does indeed protect mice against the toxicity of SN38 Figure 2.5b). These results suggested that mice can be treated with

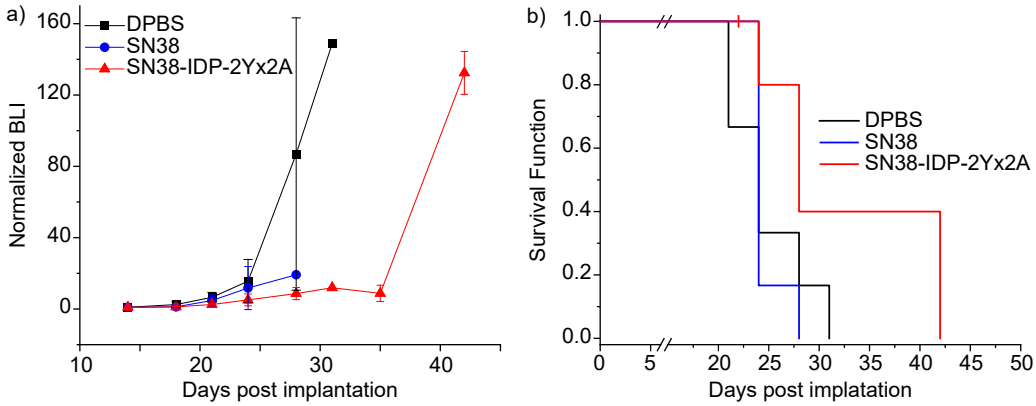


Figure 2.6: Preliminary efficacy study of U87 tumor bearing mice. a) On day 15 after tumor cell injection, 18 mice were distributed into three study arms (n=6) and injected via CED with 10 µL of DPBS, 75 µM SN38 in 50% DMSO, or 75 µM SN38-IDP-2Yx2A. Normalized BLI was monitored until tumor burden exceeded IACUC guidelines and animals were sacrificed. The average tumor size and standard deviation for the remaining mice were plotted. b) The Kaplan-Meier survival analysis for the three study arms showed a significant improvement ($p = 0.032$) in survival for formulated SN38-IDP-2Yx2A versus the SN38 DMSO control.

concentrations of 200 µM SN38-IDP-2Yx2A instead of 75 µM SN38-2Yx2A in 50% DMSO.

A full three-arm (n=9) efficacy study was repeated on glioblastoma (U251-MG-FL) tumor-bearing mice with higher SN38 dosing as determined by the toxicity data. The efficacy study was performed with formulated SN38-IDP-2Yx2A to determine if mouse survival increased and tumor growth slowed. SN38-IDP-2Yx2A was formulated with 150 µM drug and approximately 15 µM IDP-2Yx2A. DPBS and 75 µM SN38 in 50% DMSO were included as controls for the study. CED infusions of 10 µL were conducted six days post-implantation, and subsequent infusions were performed every seven days for four weeks. Tumor size was monitored by bioluminescence intensity (BLI) (Figure 2.7a, Table 2.3), and Kaplan-Meier analysis and log-rank tests were used to determine statistical significance for mouse survival (Figure 2.7b). Mice treated with SN38-IDP-2Yx2A showed statistically significant increased survival compared to mice treated with the DPBS control ($p = 0.014$) and SN38 DMSO control ($p = 0.021$, Additional Table 2.4).

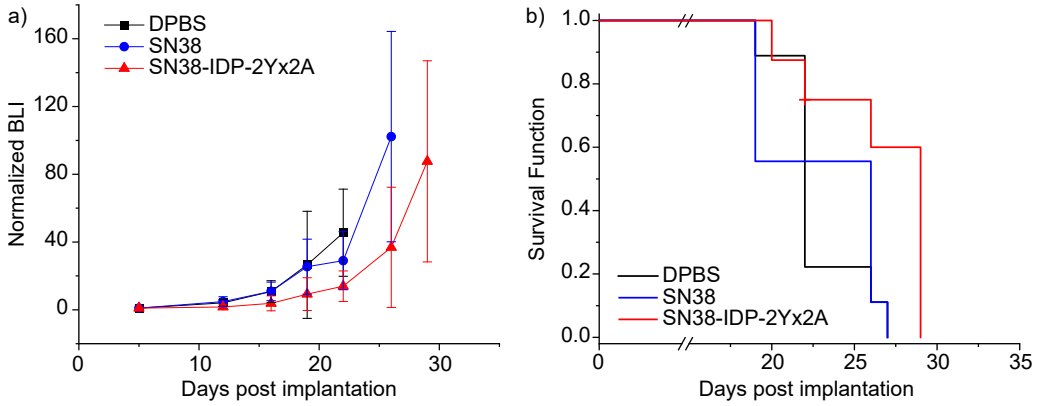


Figure 2.7: Efficacy study of formulated SN38-IDP-2Yx2A for CED treatment of GBM in mice. U251-MG-FL cells were implanted intracranially and grown until an average BLI indicated log phase growth. a) On day 6, 27 mice were distributed into three study arms (n=9) and injected via CED with 10 µL of DPBS, 75 µM SN38 in 50% DMSO, or 150 µM SN38-IDP-2Yx2A. Normalized BLI was monitored until tumor burden exceeded IACUC guidelines and animals were sacrificed. The average tumor size and standard deviation for the remaining mice were plotted. b) The Kaplan-Meier survival analysis for the three study arms showed a significant improvement ($p = 0.021$, $p = 0.014$) in survival for formulated SN38-IDP-2Yx2A versus both DPBS and SN38 DMSO controls, respectively.

2.4 Conclusion

This work evaluates the IDP-2Yx2A protein micelle construct as a delivery scaffold. The unloaded carrier system exhibited no cell toxicity up to 20 µM, and *in vivo* experiments indicated no adverse effects in mice with intracranial injections up to 20 µM. The highly insoluble but highly potent chemotherapeutic SN38 was loaded into the micelle interior in its active form. Using a protein-based platform for the delivery of chemotherapeutics via CED has many potential benefits, as these structures are biodegradable and have the opportunity for genetically optimizing properties and encoding targeting groups. When encapsulated, the cytotoxic effect of the drug was maintained for the U251-MG-FL glioblastoma cell line with EC_{50} values in the low nM. When SN38 was encapsulated, the IDP-2Yx2A delivery scaffold protected against neurological toxicity symptoms in healthy mice after intracranial injection. Mice with U251-MG-FL tumors were treated with DPBS, SN38 solubilized in a solution of 50% DMSO, and SN38 loaded IDP-2Yx2A micelles. Mice treated with the SN38-IDP-2Yx2A had slower tumor growth and significantly longer survival functions compared to the free drug. Given this initial success, future studies will expand the utility of IDP- 2Yx2A micelles as delivery scaffolds that can be easily tailored and engineered to improve the delivery of hydrophobic small molecule chemotherapeutics to tumor tissue.

2.5 Materials and Methods

2.5.1 Reagents and Instrumentation

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Milli-Q (MQ) H_2O was purified to a resistivity of 18.2 M Ω • cm (at 25 °C) using a MQ Gradients ultrapure water purification system (Millipore).

2.5.1.1 Mass Spectrometry

Proteins were analyzed by time of flight electrospray ionization mass spectrometry (ESI-TOF-MS) using an Agilent 1260 series liquid chromatography (LC) system that was connected in-line with an Agilent 6530 Q-TOF LC/MS system (Santa Clara, CA). Protein samples were passed through a 0.22 mM cellulose acetate filter prior to injection. Depending on sample concentration, 0.5 -10 mL was injected on to a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) analytical column with a flow rate of 0.40 mL/min and a gradient of 5-100% MeCN+0.1% (v/v)

formic acid (FA) in MQ H_2O+0.1% FA over the course of 8 min. The total ion chromatograms were extracted in the Agilent Mass Hunter Workstation software (Qualitative Analysis Version B.10.0, Build 10.0, Agilent Technologies Inc. 2020). Protein molecular weights were reconstructed using a maximum entropy deconvolution. The resulting spectra were exported as CSV files and plotted using the Chartograph web-tool (www.chartograph.com).

2.5.1.2 Reverse Phase High Performance Liquid Chromatography (RP-HPLC)

RP-HPLC purification of IDP-2Yx2A proteins mutants was performed on an Agilent 1260 Infinity series LC system controlled by Open Lab CDS software (Agilent Technologies Inc. 2011-2015) and equipped with a fraction collector and UV-Vis detector. Preparatory scale purification was performed on a Waters XBridge Protein BEH C4 OBD Prep Column (300 Å, 4 µm, 19 mm X 50 mm) with a flow rate of 5 mL/min and a gradient of 5-95% HPLC grade MeCN+0.1% (v/v) trifluoroacetic acid (TFA) in MQ H_2O+0.1% TFA over the course of 30 min. Fractions containing IDP-2Yx2A mutants were identified by their 280 nm and 350 nm fluorescence profiles and confirmed with ESI-TOF-MS.

2.5.1.3 Dynamic Light Scattering (DLS)

DLS was performed on a Malvern Instruments Zetasizer Nano ZS. Prior to analysis, lyophilized samples were rehydrated in the appropriate buffer and passed through a 0.22 µm cellulose acetate centrifugal filter. Protein concentration was determined and 80 µL of filtered sample was added to a DLS cuvette. All analyses were performed at 25 °C and measurement data were presented as averaged mean diameters determined by the %Number distribution showing the majority population in a sample. The standard deviations were calculated from an average of the %Number mean for each run. Heterogeneity was determined by the polydispersity index and %Intensity traces.

2.5.1.4 UV-VIS Spectroscopic Measurements

UV-Vis measurements were conducted on a Thermo Scientific nanoDrop 1000 scan benchtop spectrophotometer (Wilmington, DE). SN38 and protein concentrations were determined by measuring the absorbance at 380 nm and 280 nm, respectively, and were calculated using Beer's law and their respective extinction coefficients.

2.5.2 Protein Expression and Purification

Amino acid sequence of fusion protein KSI-IDP-2Yx2A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIRE
FYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFRFNGA
GKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEVKSPAEV
KSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPVEAKSPAEA
KSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKSPVTVKSPAEA
KSPVEVKSPYWSAYGAYAQYVVYIYAYWYL*
KSI protein sequence shown in magenta
IDP protein sequence shown in black
2Yx2A protein sequence shown in blue

Expression and purification of KSI-IDP-2Yx2A were adapted from a previously described protocol.[35] Rosetta2 (pLysS) cells transformed with a pET31(+) vector containing the KSI-IDP-2Yx2A sequence were grown overnight at 37 °C in 20 mL of Lysogeny broth (LB) with carbenicillin and chloramphenicol antibiotics. A 4 L flask with 2 L of LB/carbenicillin was inoculated with the overnight culture and grown at 37 °C, 220 rpm, until the OD600 reached 0.6-0.8. The temperature was lowered to 18 °C, and equilibrated for 30 min before protein expression was induced by the addition of IPTG (0.5 mM). Protein was expressed for 20-24 h before the culture was centrifuged at 8000 g for 20 min. The pelleted cells (5-10 g) were frozen at -20 °C overnight.

The cell pellet was resuspended in 20-35 mL of lysis buffer with PMSF, lysozyme, and RNase B. Cells were lysed by sonication at 75% amplitude (2 s on 4 s off) for 15 min. The lysis was centrifuged at 14,500 g for 30 min. The lysis supernatant was discarded, and the insoluble pellet was washed twice with lysis buffer followed by MQ H_2O by resuspending the insoluble fraction and centrifuging for 20 min at 14,500 g. Next, 5-10 mL of 6 M guanidinium HCl (Guan HCl) was added to resuspend all proteins from the insoluble lysis pellet. The Guan HCl suspension was vortexed, shaken at 80 rpm for 1 h, and then centrifuged for 20 min at 14,500 g to remove cell debris. SDS-PAGE and ESI-TOF-MS analysis determined the purity of the KSI-IDP-2Yx2A fusion protein. Protein yield was determined by measuring the A280 and calculated from the extinction coefficient (ExPASy Protein Parameters Tools https://web.expasy.org/protparam/).

A methionine residue followed the KSI fusion partner to allow for cyanogen bromide (BrCN) cleavage to obtain the final IDP-2Yx2A protein. The Guan HCl solution was adjusted to a final HCl concentration of 0.5 M with 10% FA. The cleavage reaction was carried out with 100 molar excess of BrCN to protein by adding a 3 M BrCN

stock in DCM. The solution was stirred under nitrogen in the dark for 24 h. ESI-TOF-MS confirmed the completed reaction. The Guan HCl solution was added to 30 mL of a prechilled saturated $(NH_4)_2SO_4$ solution and stored at -20 °C for 1-3 h to precipitate proteins. The soluble components (Guan HCl and BrCN) were retained in the supernatant after centrifugation (20 min, 14,500 g) and decanted from the precipitated protein pellet. The protein pellet was resuspended in 20% MeCN:H_2O by sonication (65% amplitude, 2 s on 4 s off, 10 min), and insoluble components were removed by centrifugation (20 min, 14,500 g). The supernatant was dialyzed (3kDa membrane) against MQ H_2O overnight to remove the remaining salts. The expression and purification methods have been optimized to obtain yields of 100-300 mg/L of IDP-2Yx2A at the end of this step.[35]

The IDP-2Yx2A protein solution was further purified using RP-HPLC. The protein was purified by 1 mL injections and eluted in fractions between 30-40% MeCN. SDS-PAGE and ESI-TOF-MS determined the purity of IDP-2Yx2A.

2.5.2.1 Protein Gel Analysis

Sodium dodecyl sulfate-polyacrylamide (SDS PAGE) gel analyses were run on a Novex apparatus using 4-12% precast linear gradient polyacrylamide gels (Bio-Rad, Hercules, CA). Protein samples were mixed with a 6X loading buffer and heated to 95 °C for 5 min. Samples were run for 30 min at 200 V in the SDS-MES buffer. Novex Sharp pre-stained protein standard was run on each gel for indication of molecular masses. Protein bands were stained with Coomassie Brilliant Blue R-250 (Bio-Rad) and gels were visualized on the Gel-Doc EZ Imager (Biorad).

2.5.3 IDP-2Yx2A Characterization

Each batch of RP-HPLC purified IDP-2Yx2A was fully characterized to confirm expected self-assembling behavior. Protein concentration was determined by measuring A280 and calculated using the extinction coefficient. (ExPASy Protein Parameters https://web.expasy.org/protparam/). For concentration measurements, the protein was first diluted 1:10 in 50% MeCN in H_2O. Dynamic light scattering measurements were used to confirm micelle size. Lyophilized IDP-2Yx2A was reconstituted in DPBS and passed through a 0.22 µm cellulose acetate centrifugal filter to remove any large particles from the solution prior to DLS measurements. The average diameter of micelles was determined by the %Number distribution based on the majority population in the sample.

2.5.4 Endotoxin Assay

The LAL endotoxin quantitation kit was used as indicated by the supplier. Serial dilutions were made using endotoxin-free water. The absorbance was measured at 405 nm using an Infinite 200 Pro (Tecan, Switzerland) plate reader, and the absorbance of the blank sample was subtracted from each well.

2.5.5 Dye Modification

All dye concentrations were determined using their appropriate nm absorbance and calculated using their extinction coefficients. A solution of 150 µM IDP-2Yx2A in 10 mM phosphate buffer pH 7.2 was mixed with AF-488 NHS ester from a DMSO stock (0.2 equivalents). The reaction was incubated for 2 h at room temperature and the excess AF-488 NHS ester was removed by using a 10 kDa MWCO spin concentrator. The final concentrations were 8 µM AF-488 dye and 80 µM IDP-2Yx2A. ESI-TOF-MS confirmed IDP-2Yx2A was covalently labeled with AF-488 NHS ester on its lysine residues at approximately 12% labeling (Additional Figure 2.8).

The AF-594 NHS ester was modified with a glycine tert-butyl ester to quench the lysine reactive NHS ester (AF-594 gly). A 1 mM solution of AF-594 in 50:50 H_2O:DMF was incubated overnight at room temperature with glycine tert-butyl ester (200 equivalents) and TEA (200 equivalents). The organic solvent was removed under reduced pressure and excess reagent was removed by using a Nap-25 desalting column. Reaction completion was confirmed by ESI-TOF-MS (Additional Figure 2.9b).

The AF-594 gly was loaded into the core of the covalently labeled AF-488-IDP-2Yx2A micelles. Excess AF-594 gly was added to a 30 µM solution of AF-488-IDP-2Yx2A protein and vortexed. Free AF-594 gly was removed by a 10kDa spin concentration and the final concentration was 24 µM, 2.4 µM, and 5.7 µM for the IDP-2Yx2A protein, AF-488 dye, and AF-504 gly dye, respectively (AF488-IDP-2Yx2A-AF594).

2.5.6 FRET Release Assay

Fetal bovine serum (FBS) was added to 24 µM AF488-IDP-2Yx2A-AF594 with a final concentration of 10%. The fluorescence intensity from 515-700 nm was monitored for 20 h starting at 30 s intervals from initial FBS introduction. The FRET efficiency (E_{rel}) was calculated using the time point values at 550 nm and 650 nm and fit to an exponential decay function to determine the half life of cargo release.

2.5.7 SN38 Release Study

The drug release assay was performed using a dialysis method. A 100 µL sample of 140 µM SN38 with 15 µM IDP was prepared in DPBS. In parallel, a 140 µM SN38 solution was prepared in DPBS + 5% v/v DMSO as a free drug control. Each sample was loaded into a Slide-a-Lyzer MINI Dialysis 10k MWCO cartridge, which was placed in a quartz cuvette containing 900 µL of DPBS + 10% FBS as a sink medium for drug release. The cuvette was kept at 37 °C, and the SN38 concentration in the cuvette was monitored regularly over 24 h by fluorescence at 370 nm excitation and 540 nm emission.

2.5.8 Cell Culture

Human embryonic kidney cells (HEK293) were obtained from the UC Berkeley Cell Culture Facility. U251-MG human glioblastoma cells bearing a luciferase reporter (U251-MG-FL) were acquired from UCSF. Cells were cultured in DMEM (ATCC, Manassas, VA, USA) + 10% fetal bovine serum (Omega Scientific, Tarzana, CA, USA) and 1% penicillin-streptomycin (Thermo) at 37 °C and 5% CO_2.

2.5.9 Cell Viability Assay

HEK293 and U251-MG-FL cells were trypsinized and diluted to a density of 50,000 cells/mL. An aliquot of 100 µL of cell stock was placed in each well of a 96 well plate (Corning, Corning, NY, USA) for a final density of 5,000 cells/well. The cells were incubated overnight at 37 °C, 5% CO_2. The media was aspirated from each well then replaced with 75 µL of fresh DMEM and 25 µL of the appropriate 4X sample stocks. The plate was incubated at 37 °C, 5% CO_2 for 72 h. Then, media was aspirated from each well, and 100 µL of DMEM containing 20% MTS was added. The plate was incubated at 37 °C, 5% CO_2 for 3 h. The 490 nm absorbance was measured using an Infinite 200 Pro plate reader (Tecan, Switzerland). Cell viability was calculated as an absorbance percent over the untreated cell control.

2.5.10 Convection-enhanced Delivery (CED Infusion)

For CED administration, our approach was similar to that previously described.[48] Briefly, infusion cannula were made with silica tubing fused to a 100 µL syringe (Plastic one, Roanoke, VA) with a 0.5 mm stepped-tip needle protruding from the silica guide base. The syringes were loaded with the appropriate agents and attached to the microinfusion pump (Bioanalytical Systems, Lafayette, IN, USA). The can-

nula was lowered 4 mm through a skull hole made by puncturing the skull with a coordination of 3 mm to the right from bregma and just on top of the coronal suture (the region where the tumor cells were injected). The samples were infused at a rate of 1 µL/min until a volume of 10 µL was delivered. The cannula was removed 5 min post-infusion.

2.5.11 *In vivo* Studies

All animal procedures were performed according to a protocol approved by the University of California San Francisco Institutional Animal Care and Use Committee (IACUC). Five-week-old athymic (homozygous, *nu/nu*) female mice weighing 18-22 g were purchased from Harlan-Envigo Labs (Livermore, CA, USA). For tumor inoculation, 3 x 10^5 human U251-MG-FL glioblastoma cells with the luciferase reporter gene were intracranially implanted as previously described.[73] Tumor size was monitored by measuring the bioluminescence from the luciferase reporter system using the IVIS Lumina imaging station (Caliper Life Sciences, Alameda, CA).

Mice were treated via CED infusion once bioluminescence indicated log phase growth. Mice were euthanized when tumor burden reached the levels determined by IACUC guidelines.

Mouse toxicity studies were performed to determine appropriate concentrations for CED samples. For each tested concentration, three healthy mice were injected intracranially with 10 µL of agent. Mice were monitored during one week for signs of cytotoxic or neurological side effects and weighed daily to check for weight loss that would indicate adverse reactions to the treatment.

2.6 Additional Figures

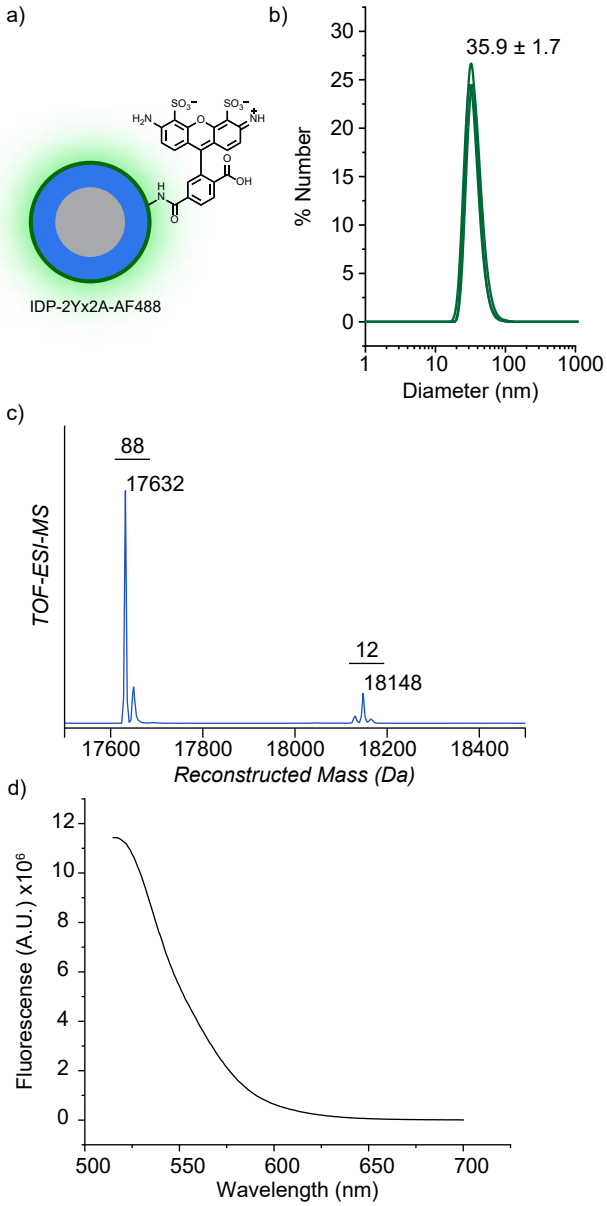


Figure 2.8: IDP-2Yx2A was covalently modified by AF-488 NHS ester. a) The lysine residues of IDP-2Yx2A were covalently labeled with AF488-NHS ester b) The assembly and size of IDP-2Yx2A-AF488 micelles were confirmed by DLS. c) LC-MS analaysis followed by integration of the IDP-2YXA and AF488-IDP-2Yx2A peaks by the Chartograph software showed 12% modification of the lysine residues on the protein. d) Fluorescence intensity spectrum of AF488-IDP-2Yx2A.

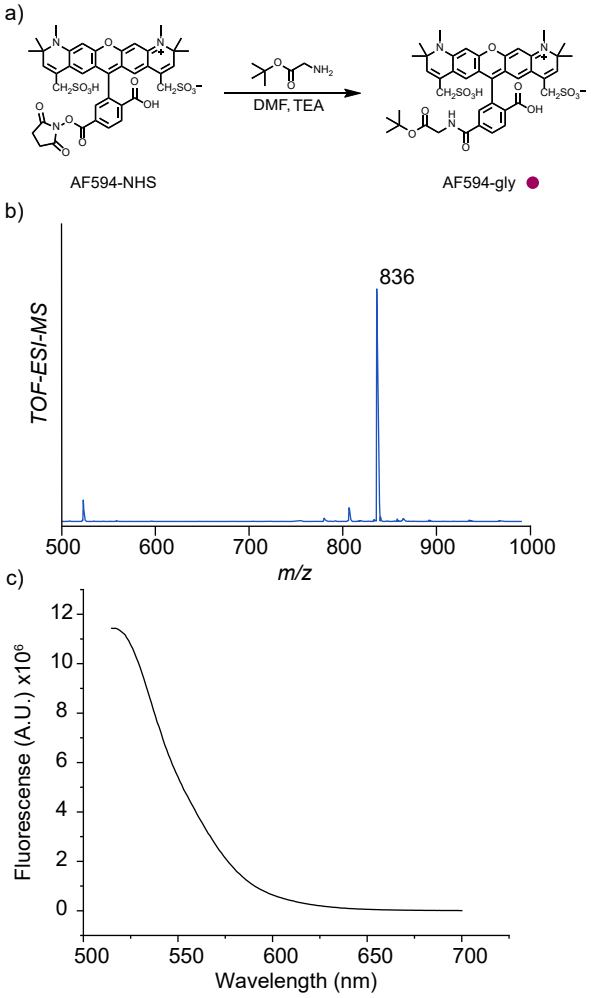


Figure 2.9: LC-MS analysis of AF-594 gly. a) AF594-NHS ester was modified to deactivate the NHS ester reactive group and create a hydrophobic dye non-covalent loading into the micelle core. b) The reaction of AF-594 NHS ester to AF-594 gly (825.9 g/mol) was confirmed by ESI-TOF-MS.

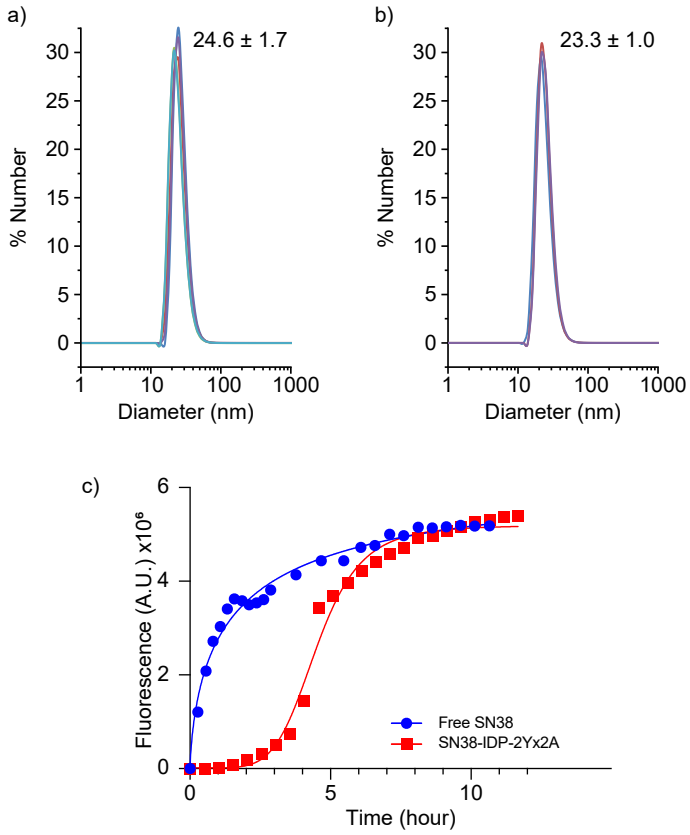


Figure 2.10: SN38-IDP-2Yx2A DLS and drug release study. a) DLS confirmed diameter of SN38-IDP-2Yx2A in PBS to be 24.6 ± 1.7 nm. b) DLS confirmed diameter of SN38-IDP-2Yx2A in PBS to be 23.3 ± 1.0 nm. c) Drug release kinetics of free SN38 ($t_{1/2}$ = 1.599 ± 0.447 h) vs. SN38-IDP-2Yx2A ($t_{1/2}$ = 4.517 ± 0.093 h).

Table 2.1: Drug loading and encapsulation efficiency of SN38 with IDP-2Yx2A micelles

[SN38] added (μM)	[SN38] loaded (μM)	[IDP] (μM)	Loading	Encapsulation
500	330	15	33%	66%
500	284	11	36%	57%
500	250	15	27%	50%
		Average	**32.2%**	**57.6%**
		stdev	**4.6**	**8.0**

Table 2.2: Median survival and log rank p-values for Kaplan-Meier Survival Analysis of preliminary efficacy study (n=6).

n=6	**DPBS Control**	**SN38 DMSO Control**	**SN38-IDP-2Yx2A**
Median Survival	24 Days	24 Days	42 Days
PBS Control	–	0.7948	0.07847
SN38 DMSO Control	0.9046	–	0.03198
SN38-IDP-2Yx2A	0.07847	0.03198	–

Table 2.3: Normalized bioluminescence intensity data of randomized mice for efficacy study (n=9). Blank entries indicates that all mice in the group were sacrificed.

Days after injection	**DPBS Control**	**SN38 DMSO Control**	**SN38-IDP-2Yx2A**
5	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0
12	4.11 ± 2.04	4.84 ± 2.88	1.70± 1.26
16	10.77 ± 6.411	10.88 ± 5.45	3.82 ± 4.41
19	26.54 ± 31.59	25.45 ± 16.24	9.25 ± 9.65
22	45.48 ± 25.72	29.01 ± 17.14	13.95 ± 9.02
26	–	102.22 ± 62.08	36.87 ± 35.48
29	–	–	87.65 ± 59.37

Table 2.4: Median survival and log rank p-values for Kaplan-Meier Survival Analysis of efficacy study (n=9).

n=6	**DPBS Control**	**SN38 DMSO Control**	**SN38-IDP-2Yx2A**
Median Survival	22 Days	26 Days	29 Days
DPBS Control	–	0.7948	0.01413
SN38 DMSO Control	0.7948	–	0.02093
SN38-IDP-2Yx2A	0.01413	0.02093	–

2.7 Acknowledgements

This work was supported by a generous donation from the Fishgold family through UCSF, and the Chemical Biology Graduate Program at UC Berkeley (NIH T32-GM066698). J.M.G was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program. S.H.K was supported by the University of California Cancer Research Coordinating Committee (UCCRC) predoctoral fellowship. P.H. was supported by the Immunotherapeutics and Vaccine Research Initiative

(IVRI) at UC Berkeley. Mouse Studies were conducted by Tomoko OZawa and David Raleigh at UCSF.

2.8 References

(1) Gleason, J. M.; Klass, S. H.; Huang, P.; Ozawa, T.; Santos, R. A.; Fogarty, M. M.; Raleigh, D. R.; Berger, M. S.; Francis, M. B. *ACS Applied Bio Materials* **2022**.

(2) Singh, S. *Journal of Nanoscience and Nanotechnology* **2010**, *10*, 7906–7918.

(3) Kiparissides, C.; Kammona, O. *The Canadian Journal of Chemical Engineering* **2013**, *91*, 638–651.

(4) Wang, A. Z.; Langer, R.; Farokhzad, O. C. *Annual Review of Medicine* **2012**, *63*, 185–198.

(5) Din, F. U.; Aman, W.; Ullah, I.; Qureshi, O. S.; Mustapha, O.; Shafique, S.; Zeb, A. *International Journal of Nanomedicine* **2017**, *12*, 7291–7309.

(6) Cha, G. D.; Kang, T.; Baik, S.; Kim, D.; Choi, S. H.; Hyeon, T.; Kim, D.-H. *Journal of Controlled Release* **2020**, *328*, 350–367.

(7) Croy, S. R.; Kwon, G. S. *Current Pharmaceutical Design* **2006**, *12*, 4669–4684.

(8) Sercombe, L.; Veerati, T.; Moheimani, F.; Wu, S. Y.; Sood, A. K.; Hua, S. *Frontiers in Pharmacology* **2015**, *6*, 286.

(9) De, R.; Mahata, M. K.; Kim, K.-T. *Advanced Science* **2022**, *9*, 2105373.

(10) Mei, H.; Cai, S.; Huang, D.; Gao, H.; Cao, J.; Bin, H. *Bioactive Materials* **2022**, *6*, 220–240.

(11) Zylberberg, C.; Matosevic, S. *Drug Delivery* **2016**, *23*, 3319–3329.

(12) Adams, M. L.; Lavasanifar, A.; Kwon, G. S. *Journal of Pharmaceutical Sciences* **2003**, *92*, 1343–1355.

(13) Savić, R.; Eisenberg, A.; Maysinger, D. *Journal of Drug Targeting* **2006**, *14*, 343–355.

(14) Raval, N.; Kalyane, D.; Maheshwari, R.; Tekade, R. K. In *Basic Fundamentals of Drug Delivery*; Elsevier: 2019, pp 173–201.

(15) Cabral, H.; Miyata, K.; Osada, K.; Kataoka, K. *Chemical Reviews* **2018**, *118*, 6844–6892.

(16) Lazzari, M.; Torneiro, M. *Polymers* **2020**, *12*, 869.

(17) Dedeo, M. T.; Finley, D. T.; Francis, M. B. *Progress in Molecular Biology and Translational Science* **2011**, *103*, 353–392.

(18) Saxena, R.; Nanjan, M. J. *Drug Delivery* **2015**, *22*, 156–167.

(19) Czapar, A. E.; Steinmetz, N. F. *Current Opinion in Chemical Biology* **2017**, *38*, 108–116.

(20) Wen, A. M.; Steinmetz, N. F. *Chemical Society Reviews* **2016**, *45*, 4074–4126.

(21) Dzuricky, M.; Roberts, S.; Chilkoti, A. *Biochemistry* **2018**, *57*, 2405–2414.

(22) MacEwan, S. R.; Weitzhandler, I.; Hoffmann, I.; Genzer, J.; Gradzielski, M.; Chilkoti, A. *Biomacromolecules* **2017**, *18*, 599–609.

(23) Simon, J. R.; Carroll, N. J.; Rubinstein, M.; Chilkoti, A.; López, G. P. *Nature Chemistry* **2017**, *9*, 509–515.

(24) Guler, M. O.; Claussen, R. C.; Stupp, S. I. *Journal of Materials Chemistry* **2005**, *15*, 4507.

(25) Zimenkov, Y.; Dublin, S. N.; Ni, R.; Tu, R. S.; Breedveld, V.; Apkarian, R. P.; Conticello, V. P. *Journal of the American Chemical Society* **2006**, *128*, 6770–6771.

(26) Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. *Science* **1998**, *281*, 389–392.

(27) Magnotti, E. L.; Hughes, S. A.; Dillard, R. S.; Wang, S.; Hough, L.; Karumbamkandathil, A.; Lian, T.; Wall, J. S.; Zuo, X.; Wright, E. R.; Conticello, V. P. *Journal of the American Chemical Society* **2016**, *138*, 16274–16282.

(28) Hassouneh, W.; Zhulina, E. B.; Chilkoti, A.; Rubinstein, M. *Macromolecules* **2015**, *48*, 4183–4195.

(29) Kim, W.; Xiao, J.; Chaikof, E. L. *Langmuir* **2011**, *27*, 14329–14334.

(30) Stephanopoulos, N.; Tong, G. J.; Hsiao, S. C.; Francis, M. B. *ACS Nano* **2010**, *4*, 6014–6020.

(31) Fiedler, J. D.; Brown, S. D.; Lau, J. L.; Finn, M. G. *Angewandte Chemie International Edition* **2010**, *49*, 9648–9651.

(32) Aanei, I. L.; ElSohly, A. M.; Farkas, M. E.; Netirojjanakul, C.; Regan, M.; Taylor Murphy, S.; O'Neil, J. P.; Seo, Y.; Francis, M. B. *Molecular Pharmaceutics* **2016**, *13*, 3764–3772.

(33) Finbloom, J. A.; Han, K.; Aanei, I. L.; Hartman, E. C.; Finley, D. T.; Dedeo, M. T.; Fishman, M.; Downing, K. H.; Francis, M. B. *Bioconjugate Chemistry* **2016**, *27*, 2480–2485.

(34) Klass, S. H.; Smith, M. J.; Fiala, T. A.; Lee, J. P.; Omole, A. O.; Han, B.-G.; Downing, K. H.; Kumar, S.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 4291–4299.

(35) Klass, S. H.; Gleason, J. M.; Omole, A. O.; Onoa, B.; Bustamante, C. J.; Francis, M. B. *Biomacromolecules* **2022**, *23*, 1462–1470.

(36) Allard, E.; Passirani, C.; Benoit, J.-P. *Biomaterials* **2009**, *30*, 2302–2318.

(37) Wohlfart, S.; Gelperina, S.; Kreuter, J. *Journal of Controlled Release* **2012**, *161*, 264–273.

(38) Jensen, S. A.; Day, E. S.; Ko, C. H.; Hurley, L. A.; Luciano, J. P.; Kouri, F. M.; Merkel, T. J.; Luthi, A. J.; Patel, P. C.; Cutler, J. I.; Rotz, M. W.; Meade, T. J.; Giljohann, D. A.; Mirkin, C. A.; Stegh, A. H. *Science Translational Medicine* **2013**, *5*, 209152–209152.

(39) Omuro, A. *JAMA* **2013**, *310*, 1842.

(40) Chen, W.; Zou, Y.; Zhong, Z.; Haag, R. *Small* **2017**, *13*, 1601997.

(41) Gregory, J. V.; Kadiyala, P.; Doherty, R.; Cadena, M.; Habeel, S.; Ruoslahti, E.; Lowenstein, P. R.; Castro, M. G.; Lahann, J. *Nature Communications* **2020**, *11*, 5687.

(42) Han, Y.; Park, J.-H. *Drug Delivery and Translational Research* **2020**, *10*, 1876–1887.

(43) Hsu, J.-F.; Chu, S.-M.; Liao, C.-C.; Wang, C.-J.; Wang, Y.-S.; Lai, M.-Y.; Wang, H.-C.; Huang, H.-R.; Tsai, M.-H. *Cancers* **2021**, *13*, 195.

(44) Bobo, R. H.; Laske, D. W.; Akbasak, A.; Morrison, P. F.; Dedrick, R. L.; Oldfield, E. H. *Proceedings of the National Academy of Sciences* **1994**, *91*, 2076–2080.

(45) Yamashita, Y.; Krauze, M. T.; Kawaguchi, T.; noble, C. O.; Drummond, D. C.; Park, J. W.; Bankiewicz, K. S. *Neuro-Oncology* **2007**, *9*, 20–28.

(46) Xi, G.; Robinson, E.; Mania-Farnell, B.; Vanin, E. F.; Shim, K.-W.; Takao, T.; Allender, E. V.; Mayanil, C. S.; Soares, M. B.; Ho, D.; Tomita, T. *Nanomedicine: Nanotechnology, Biology and Medicine* **2014**, *10*, 381–391.

(47) Mehta, A. M.; Sonabend, A. M.; Bruce, J. N. *Neurotherapeutics* **2017**, *14*, 358–371.

(48) Serwer, L.; Hashizume, R.; Ozawa, T.; James, C. D. *Journal of Visualized Experiments* **2010**, 1992.

(49) Inoue, T.; Yamashita, Y.; Nishihara, M.; Sugiyama, S.; Sonoda, Y.; Kumabe, T.; Yokoyama, M.; Tominaga, T. *Neuro-Oncology* **2009**, *11*, 151–157.

(50) Morshed, R. A.; Cheng, Y.; Auffinger, B.; Wegscheid, M. L.; Lesniak, M. S. *Frontiers in Pharmacology* **2013**, *4*, 157.

(51) Zhang, R.; Saito, R.; Mano, Y.; Sumiyoshi, A.; Kanamori, M.; Sonoda, Y.; Kawashima, R.; Tominaga, T. *Drug Delivery* **2016**, *23*, 2780–2786.

(52) Finbloom, J.; Aanei, I.; Bernard, J.; Klass, S.; Elledge, S.; Han, K.; Ozawa, T.; Nicolaides, T.; Berger, M.; Francis, M. *Nanomaterials* **2018**, *8*, 1007.

(53) Yeini, E.; Ofek, P.; Albeck, N.; Rodriguez Ajamil, D.; Neufeld, L.; Eldar-Boock, A.; Kleiner, R.; Vaskovich Daniellaand Dangoor, S. I.; Krivitsky, A.; Burgos Luna, C.; Shenbach-Koltin, G.; Goldenfeld, M.; Hadad, O.; Tiram, G.; Satchi-Fainaro, R. *Advanced Therapeutics* **2021**, *4*, 2000124.

(54) Araki, E.; Ishikawa, M.; Iigo, M.; Koide, T.; Itabashi, M.; Hoshi, A. *Japanese Journal of Cancer Research* **1993**, *84*, 697–702.

(55) Ma, M. K.; Zamboni, W. C.; Radomski, K. M.; Furman, W. L.; Santana, V. M.; Houghton, P. J.; Hanna, S. K.; Smith, A. K.; Stewart, C. F. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **2000**, *6*, 813–819.

(56) Sapra, P.; Zhao, H.; Mehlig, M.; Malaby, J.; Kraft, P.; Longley, C.; Greenberger, L. M.; Horak, I. D. *Clinical Cancer Research* **2008**, *14*, 1888–1896.

(57) Vejjasilpa, K.; Nasongkla, N.; Manaspon, C.; Larbcharoensub, N.; Boongird, A.; Hongeng, S.; Israsena, N. *Experimental Biology and Medicine* **2015**, *240*, 1640–1647.

(58) Tseng, Y.-Y.; Yang, T.-C.; Chen, S.-M.; Yang, S.-T.; Tang, Y.-L.; Liu, S.-J. *Pharmaceutics* **2020**, *12*, 479.

(59) Beutler, B.; Rietschel, E. T. *Nature Reviews Immunology* **2003**, *3*, 169–176.

(60) Singleton, W. G. B.; Bienemann, A. S.; Woolley, M.; Johnson, D.; Lewis, O.; Wyatt, M. J.; Damment, S. J. P.; Boulter, L. J.; Killick-Cole, C. L.; Asby, D. J.; Gill, S. S. *Journal of Neurosurgery: Pediatrics* **2018**, *22*, 288–296.

(61) Kawato, Y.; Aonuma, M.; Hirota, Y.; Kuga, H.; Sato, K. *Cancer Research* **1991**, *51*, 4187–4191.

(62) Wallin, A.; Svanvik, J.; Holmlund, B.; Ferreud, L.; Sun, X.-F. *Oncology Reports* **2008**, *19*, 1493–1498.

(63) Jandu, H.; Aluzaite, K.; Fogh, L.; Thrane, S. W.; Noer, J. B.; Proszek, J.; Do, K. N.; Hansen, S. N.; Damsgaard, B.; Stougaard, M.; Moreira, J.; Hamerlik, P.; Brünner, N.; Schrohl, A.-S.; Stenvang, J. *BMC Cancer* **2016**, *16*, 34.

(64) Jensen, N. F.; Agama, K.; Roy, A.; Smith, D. H.; Pfister, T. D.; Rømer, M. U.; Zhang, H.-L.; Doroshow, J. H.; Knudsen, B. R.; Stenvang, J.; Brünner, N.; Pommier, Y. *Journal of experimental & clinical cancer research: CR* **2016**, *35*, 56.

(65) Rivory, L. P.; Bowles, M. R.; Robert, J.; Pond, S. M. *Biochemical Pharmacology* **1996**, *52*, 1103–1111.

(66) Mathijssen, R. H.; van Alphen, R. J.; Verweij, J.; Loos, W. J.; Nooter, K.; Stoter, G.; Sparreboom, A. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **2001**, *7*, 2182–2194.

(67) Wang, W.; Ghandi, A.; Liebes, L.; Louie, S. G.; Hofman, F. M.; Schönthal, A. H.; Chen, T. C. *Journal of Neurosurgery* **2011**, *114*, 689–694.

(68) Santi, D. V.; Schneider, E. L.; Ashley, G. W. *Journal of Medicinal Chemistry* **2014**, *57*, 2303–2314.

(69) De Lisa, M.; Ballatore, Z.; Marcantognini, G.; Pierantoni, C.; Antognoli, S.; Pistelli, M.; Pagliacci, A.; Berardi, R. *Oncology and Therapy* **2020**, *8*, 147–160.

(70) Verreault, M.; Strutt, D.; Masin, D.; Anantha, M.; Waterhouse, D.; Yapp, D.; Bally, M. *Journal of Controlled Release* **2012**, *158*, 34–43.

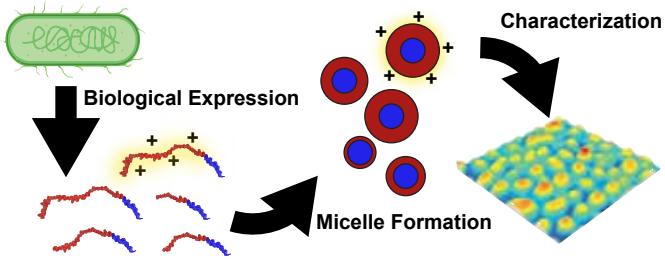
(71) Iyer, R.; Croucher, J. L.; Chorny, M.; Mangino, J. L.; Alferiev, I. S.; Levy, R. J.; Kolla, V.; Brodeur, G. M. *Cancer Letters* **2015**, *360*, 205–212.

(72) Kuroda, J.-i.; Kuratsu, J.-i.; Yasunaga, M.; Koga, Y.; Saito, Y.; Matsumura, Y. *International Journal of Cancer* **2009**, *124*, 2505–2511.

(73) Ozawa, T.; James, C. D. *Journal of Visualized Experiments: JoVE* **2010**, DOI: 10.3791/1986.

Chapter 3

Genetic Engineering of an Intrinsically Disordered Protein's Hydrophilic Domain



Portions of this chapter are based on the following publication: Klass, S. H.*; Gleason J. M.*; Omole, A. O.; Onoa, B.; Francis, M. B. *Biomacromolecules* **2022**, *23*, 1462-1470.[1]

3.1 Abstract

Surfactants, block-copolymers, and other types of micellar systems are used in a wide variety of biomedical and industrial processes. However, most commonly used surfactants are synthetically derived and pose environmental and toxicological concerns throughout their product life cycle. Because of this, bio-derived and bio-degradable surfactants are promising alternatives. For bio-surfactants to be implemented industrially, they need to be produced on a large scale and also have tailorable properties that match those afforded by the polymerization of synthetic surfactants. In this paper, a scalable and versatile production method for bio-surfactants based on a hydrophilic intrinsically disordered protein (IDP) sequence with a genetically engineered hydrophobic domain is used to study variables that impact their physico-

chemical and self-assembling properties. These amphiphilic sequences were found to self-assemble into micelles over a broad range of temperatures, pH values, and ionic strengths. To investigate the role of the IDP hydrophilic domain on self-assembly, variants with increased overall charges and systematically decreased IDP domain lengths were produced and examined for their sizes, morphologies, and critical micelle concentrations (CMCs). The results of these studies indicate that decreasing the length of the IDP domain and, consequently, the molecular weight and hydrophilic fraction, leads to smaller micelles. Additionally, significantly increasing the amount of charged residues in the hydrophilic IDP domain results in micelles of similar sizes, but with higher CMC values. This represents an initial step in developing a quantitative model for the future engineering of bio-surfactants based on this IDP sequence.

3.2 Introduction

Most commercially used surfactants are derived from hydrocarbon sources and encompass both conventional surfactants consisting of simple ionic head groups attached to hydrocarbon tails as well as macromolecular block-copolymers consisting of hydrophilic and hydrophobic domains. The properties of commonly used synthetic surfactants have been heavily researched, and they have well-understood physicochemical properties that can be tailored to address specific application needs.[2–5] By altering the hydrophobicities, chain lengths, charges, and geometries, the bulk self-assembly properties in a known context can be predictably modified.[6–11] Unfortunately, many synthetically derived surfactants show human or environmental toxicity due to their innate activities or those of their degradation products.[12–14] Due to the widespread use of surfactants, there have been renewed efforts to develop bio-derived and bio-degradable alternatives.[15,16] However, large-scale production of designer bio-surfactants for medical or environmental purposes remains a challenge regarding economic viability compared to their synthetic counterparts.[17,18] Furthermore, the design rules that govern the ability of bio-molecules to perform as surfactants are poorly understood, which impedes the development of novel constructs with direct applications.

Previous studies explored the IDP sequence as a hydrophilic domain for an amphiphilic biopolymer with a genetically encoded hydrophobic domain (2Yx2A) at the C-terminus of the protein. The production and characterization of this sequence has been described previously.[19] As this IDP sequence has not been extensively studied in an assembly context, the first variables we sought to examine were the nature of the charge and length of the IDP sequence itself. Herein, we discuss the development of IDP-2Yx2A mutational variants with systematically decreasing IDP lengths while

maintaining a neutral charge and the hydrophobic 2Yx2A assembly domain. Additionally, a "supercharged"[20] variant that maintained IDP length, but with double the overall positive charge, was also constructed. Finally, a variant with valine to proline mutations was constructed to look at the overall increase in hydrophilicity with a neutral charge. These variants were then used to investigate the role of the IDP length, charge, and hydrophilicity on physicochemical properties such as size, morphology, and critical micelle concentration (CMC). From these studies, several trends have emerged.

3.3 Results and Discussion

3.3.1 Exploring IDP length and charge

Using this insoluble expression and purification protocol described previously,[1] five additional variants of IDP-2Yx2A (further denoted as 1.0-IDP-2Yx2A) were produced. A positively charged version, (+)1.0-IDP-2Yx2A, was developed by genetically replacing all the glutamic acid residues (E) in the original sequence with lysine residues (K). This resulted in 45 total positively charged residues and an overall charge of +45 at neutral pH. Additionally, 3 variants with decreasing IDP lengths were produced, 0.75-IDP-2Yx2A, 0.5-IDP-2Yx2A and 0.25-IDP-2Yx2A, where the IDP domains are described as fractions of the full sequence in the naming scheme. Finally, a negative control variant (1.0-IDP), which contains the full-length IDP domain, but no assembly region was also expressed. This variant, which has no hydrophobic domain at its C-terminus, contains only one aromatic residue in the entire sequence for monitoring by UV-vis (Figure 3.1a). The Protein DisOrder prediction System (PrDOS) was used to create intrinsically disordered profiles for these variants and confirmed disorder probability predictions of >0.5 (Additional Figure 3.7).[21] While the positively charged variant maintains a similar molecular weight and hydrophilic fraction (molecular weight hydrophilic domain/molecular weight total protein), the variants with decreasing IDP lengths have a corresponding decrease in the hydrophilic fraction and the overall molecular weight (Figure 3.1b).

All variants were produced, isolated, and analyzed by mass spectrometry (Figure 3.1c). The spectra revealed no other identifiable species; therefore, purity is reported at greater than 95%. When analyzed by DLS in 100 mM PB pH 5.7, the 1.0-IDP negative control variant showed a small distribution centered about 7.9 nm for its mean diameter, and no apparent self-assembly behavior at room temperature (25 °C) by DLS in any analyzed conditions (pH: 1-12, phosphate buffer [buffer]: 0-200 mM) (Additional Figure 3.8). When appended with the 2Yx2A domain, 1.0-IDP-2Yx2A

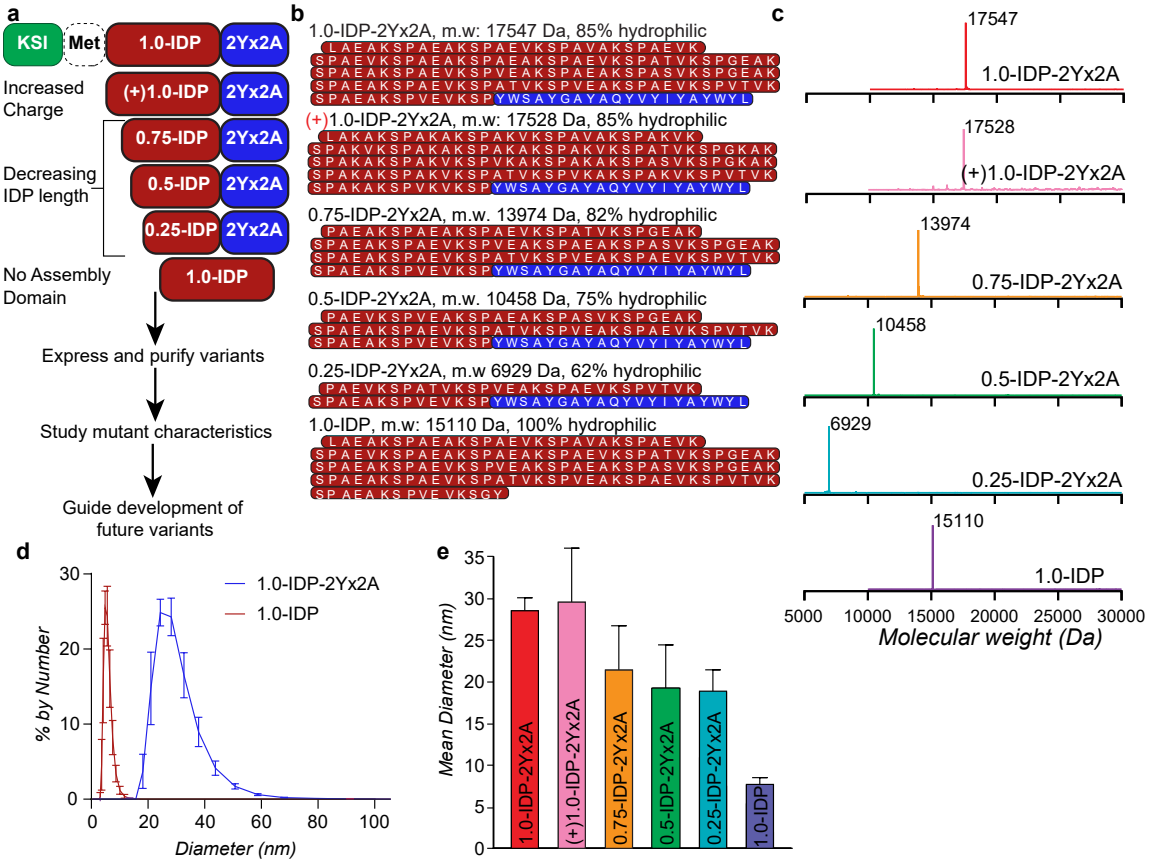


Figure 3.1: Generation of IDP variants with increased charges and decreasing lengths. (a) Using the insoluble expression and purification system, six protein constructs, including a non-assembling 1.0-IDP control, were expressed and purified. (b) The sequences varied in charge, length, and hydrophilic content. The parent construct, 1.0-IDP-2Yx2A, corresponds to the full length and charge-balanced 1.0-IDP domain appended with 2Yx2A. The (+)1.0-IDP-2Yx2A sequence has replaced every negatively charged E residue for a positively charged K. The 0.75-IDP-2Yx2A5, 0.5-IDP-2Yx2A, and 0.25-IDP-2Yx2A variants have decreasing fractions of the charge-balanced IDP domain and a corresponding decrease in mass fraction of the hydrophilic domain. (c) Mass spectra of the purified variants: 1.0-IDP-2Yx2A expected: 17547 Da, observed: 17547 Da; (+)1.0-IDP-2Yx2A expected: 17528 Da, observed: 17528 Da; 0.75-IDP-2Yx2A expected: 13975 Da, observed: 13974 Da; 0.5-IDP-2Yx2A expected: 10459 Da, observed: 10458 Da; 0.25-IDP-2Yx2A expected: 6929 Da, observed: 6929 Da; 1.0-IDP expected: 15110 Da, observed: 15110 Da. (d) Representative DLS spectra of 1.0-IDP-2Yx2A and the 1.0-IDP control in 100 mM PB, pH 5.7. (e) Averaged mean diameter values determined by DLS for each variant in 100 mM PB, pH 5.7: 1.0-IDP-2Yx2A: 28.9 ± 1.3 nm, (+)1.0-IDP-2Yx2A: 29.9 ± 6.2 nm, 0.75-IDP-2Yx2A: 21.7 ± 5.0 nm, 0.5-IDP-2Yx2A: 19.6 ± 4.9 nm, 0.25-IDP-2Yx2A: 19.1 ± 2.4 nm, 1.0-IDP: 7.9 ± 0.6 nm.

showed a large increase in its apparent diameter and size distribution, indicating self-assembly (Figure 3.1d). DLS analysis was performed for all variant proteins at multiple concentrations. Changing the overall charge of the variant showed only a slight, not significant, increase in overall size and an increase in polydispersity. Conversely, decreasing the IDP domain resulted in a noticeable, although not significant, decrease in average mean diameter (Figure 3.1e). Other DLS properties were analyzed, including %Intensity spectra (Additional Figure 3.10) and the averaged diffusion coefficients, Z-averages, and polydispersity indices (Additional Figure 3.9).

The decrease in size corresponding to a decrease in IDP length was also investigated by atomic force microscopy (AFM). The four IDP-2Yx2A variants with decreasing lengths and 1.0-IDP as a non-assembling control were deposited on atomically flat mica in low µM concentrations (1-5 µM) and subsequently imaged by AFM in air using soft tapping mode (TM-AFM). All the variants containing the 2Yx2A assembly domain showed the formation of small spherical particles, while 1.0-IDP alone showed no aggregation behavior. Using a watershed segmentation method (as defined in the ISO 25178-2 standard in Mountains Software) the maximum heights (Z-max) and diameters of the particles were obtained. Hill motifs were detected after pre-processing the images using a 3x3 filter size. Further height pruning was performed by merging motifs of <5% of the maximum height. The Z-max distributions showed that with each decreasing IDP length, there was a significant decrease ($p = 0.0045$ - <0.0001) in relative particle height between each consecutive variant. This was especially significant in comparison to the 1.0-IDP control, with representative images shown in (Figure 3.2a). Although the Z-spatial resolution in AFM is high (0.1 nm), the measured height might not reflect the actual micelle diameter as the material is chemo-adsorbed and dried on the surface. This has resulted in a flattened structure as in other micelle systems observed by AFM.[22,23] In addition, the XY-spatial resolution depends highly on the radius of curvature of the AFM tips, which could change upon repeated scanning. A comparison of the particle diameter distributions showed better consensus of relative diameters obtained by DLS size data. Applying the same analysis parameters to the 1.0-IDP control could not be accomplished as no distinct particle features were observed (Figure 3.2b).This analysis supports the formation of micelle structures for all of the IDP-2Yx2A variants containing an assembly domain.

The most interesting and perhaps useful property of the IDP-2Yx2A variants is their ability to encapsulate hydrophobic molecules into the micelle interior at low protein concentrations. To observe the effect of IDP length and charge on these properties, the critical micelle concentrations (CMCs) of the IDP-2Yx2A length variants were determined by using a solvatochromic pyrene assay. When encapsulated in the hydrophobic micelle interior, pyrene exhibits a unique I3 vibronic band emission at

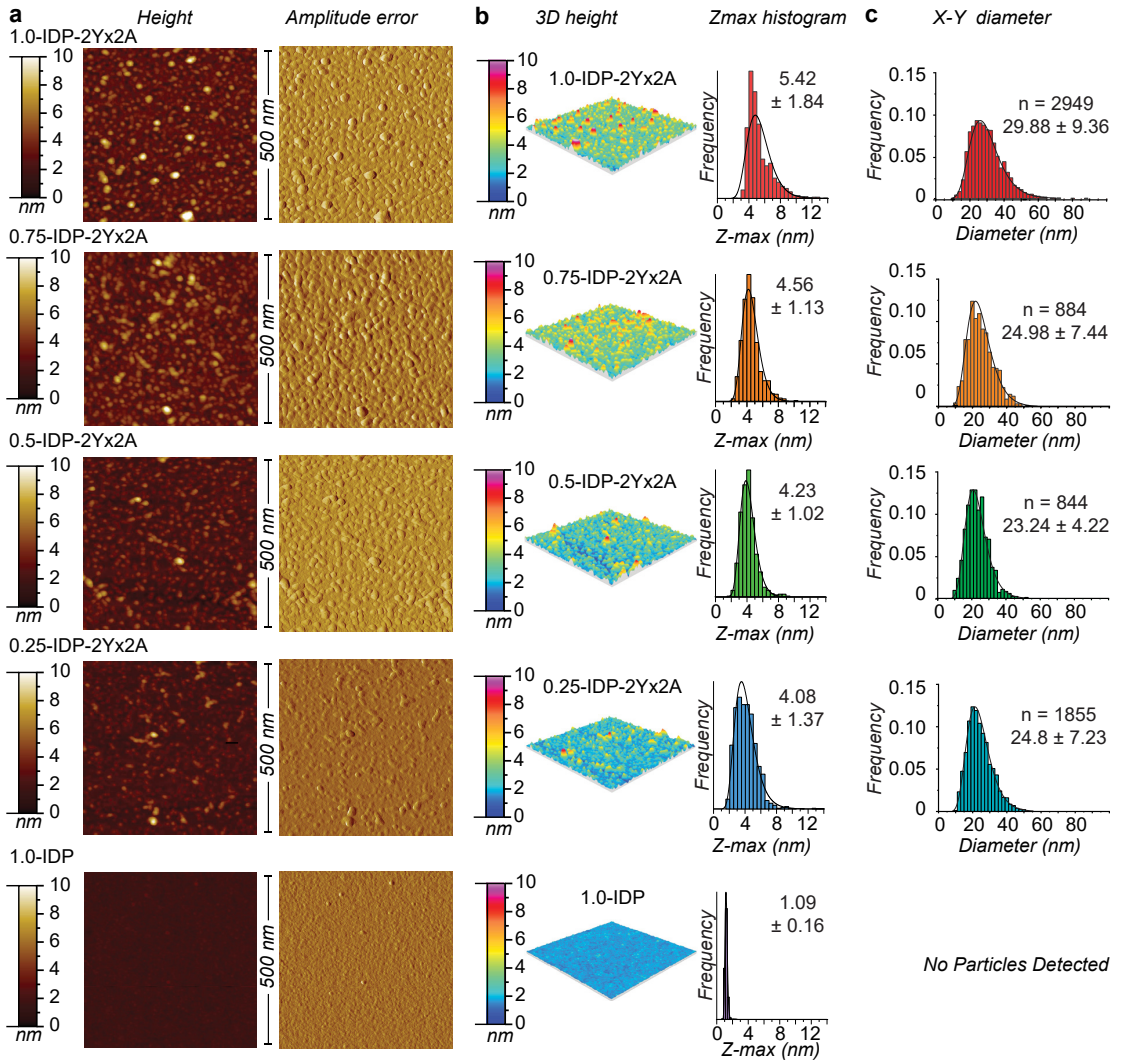


Figure 3.2: AFM analysis of IDP variants with decreasing lengths. (a) 2D height and amplitude error profiles showed tight packing of micelles with spherical morphologies for all variants with the 2Yx2A assembly domain. 2D amplitude error projections highlight the layering of the taller particles on top of the closely packed smaller particles. This effect is especially evident for the 1.0-IDP-2Yx2A variant. (b) Projection of height profiles in 3D along with particle analysis (watershed detection method) revealed decreasing Z-max values as IDP length decreased, $p < 0.0001$ for all variants, except for the transition between 0.5-IDP-2Yx2A to 0.25-IDP-2Yx2A where $p = 0.0045$. All variants with the 2Yx2A domain also had significantly larger z-max profiles than the IDP control ($p < 0.0001$). (c) The mean diameter in the X-Y plane for each IDP-2Yx2A variant was determined using a watershed detection method. 1.0-IDP-2Yx2A -> 0.25-IDP-2Yx2A: 29.88 ± 9.36 ($N = 2939$), 24.98 ± 7.44 ($N = 884$), 23.24 ± 4.22 ($N = 844$), 24.8 ± 7.23 ($N = 1855$). 1.0-IDP could not be evaluated for its mean diameter as no spherical particles were detected.

384 nm due to the solvatochromic nature of the molecule. Additionally, pyrene excimer formation (ex) at approximately 450 nm is observed when a large proportion of the pyrene in solution is encapsulated in close proximity in the interior of the micelle.[24–26] While the 1.0-IDP negative control shows no characteristic excimer formation at any protein concentration analyzed (0 - 150 µM), when resuspended in the same pyrene containing buffer all of the variants with the 2Yx2A domain exhibited prominent excimer peaks at protein concentrations high enough to support micelle formation (Figure 3.3a). All variants with the 2Yx2A domain also showed the characteristic I3 band emergence at high concentrations and its disappearance at lower concentrations as the micelles disassembled and the hydrophobic core was disrupted. The 1.0-IDP control sample showed no I3 band emergence at concentrations up to 100 µM (Additional Figure 3.13). At high concentrations, IDP variants showed I1/I3 ratios of 0.9 - 1.0 as compared to phosphate buffer alone (I1/I3 = 1.33). This is lower than common surfactants such as SDS (I1/I3 = 1.1) and Triton x 100 (I1/I3 = 1.0)[25].

CMC values for all the variants with the 2Yx2A domain were derived by plotting the ratio of the intensity of the I1 (372 nm) to the I3 (384 nm) band or the I1 to the excimer (444 nm) over a range of concentrations. This resulted in sigmoidal data plots. By fitting a logistic function to both the I1/I3 and I1/ex data sets, the CMC values could be interpreted as either the inflection points (EC_{50}) or the interceptions of the rapidly changing portion of the curve with the nearly-horizontal lower concentration regime. The latter values closely matched the logistic function derived EC_{20} value and are what we report.[27] In all cases the CMC values derived form the I1/ex were lower compared to the values derived from the I1/I3 band ratios. For all variants with the assembly 2Yx2A domain, the CMC values calculated using either method were in the low µM range. (Additional Tables 3.4, 3.5). CMC values were supported by representative contact angle measurements showing wetting behavior expected of micelles at low µM concentrations for the 1.0-IDP-2Yx2A and 0.5-IDP-2Yx2A mutants (Additional Figure 3.14). Notably, when comparing the pyrene fluorescence data for 1.0-IDP-2Yx2A to those of the positively charged variant (+)1.0-IDP-2Yx2A, a significant increase in the CMC (EC_{20}, EC_{50}) value was observed when calculated by both the I1/I3 ($p = 0.0029, 0.0103$) or I1/ex ($p = 0.00015, 0.0016$, Figure 3.3b). While still in the µM range, the increase in CMC for (+)1.0-IDP-2Yx2A is likely due to charge repulsion on the IDP side chain.

To determine if the harboring of hydrophobic molecules in the micelle interior disrupted micelle formation, DLS was performed on IDP-2Yx2A samples loaded with pyrene molecules. In comparison to the unloaded protein micelles, no substantial changes to the size were observed at high protein concentrations (Additional Figure 3.12). However, when the protein is diluted and the pyrene concentration is kept constant, micelle size increases for all variants. Consistent with (+)1.0-IDP-2Yx2A

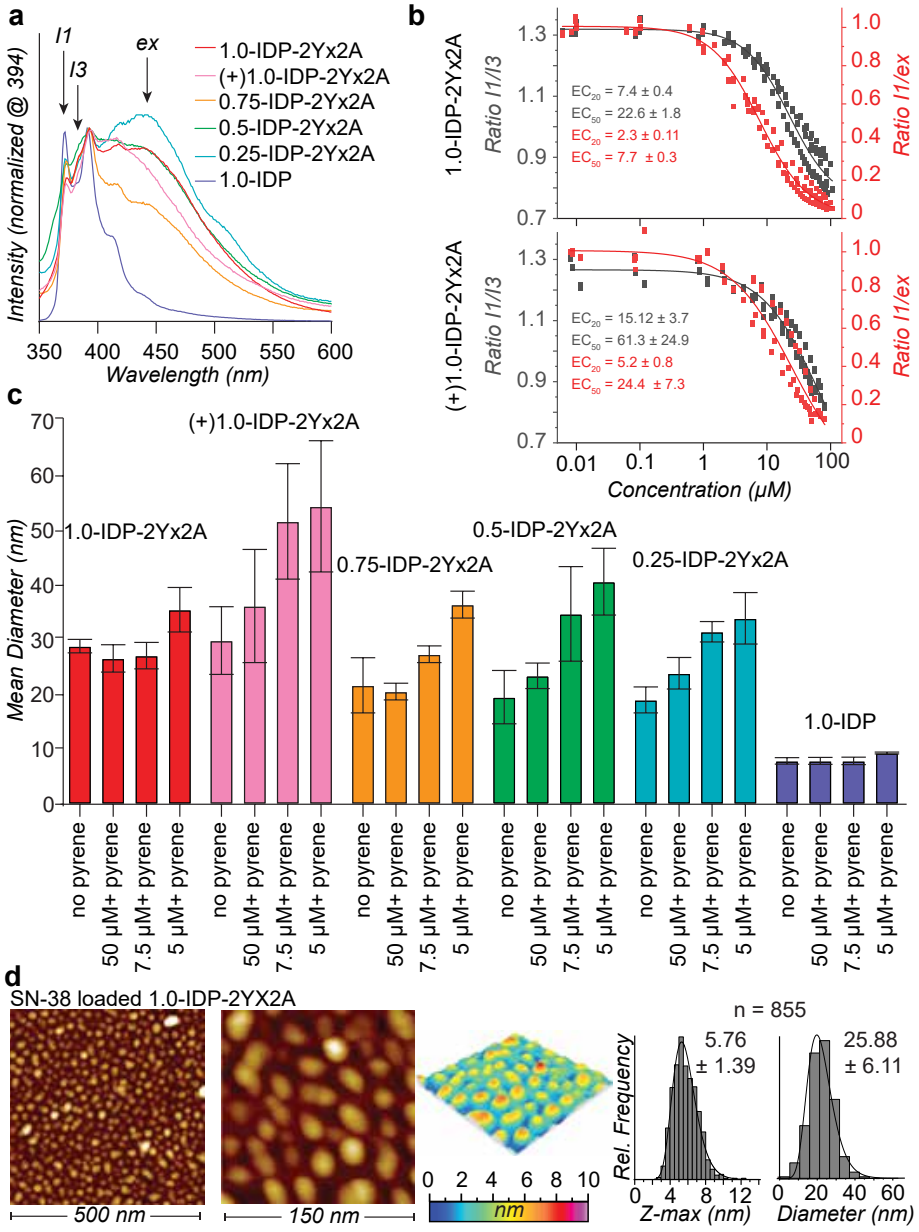


Figure 3.3: Encapsulation of hydrophobic molecules in the micelle core. (a) Representative fluorescent spectra of IDP-2Yx2A variants (30-50 μM) and 1.0-IDP control (100 μM) in the presence of 2 μM pyrene buffer are shown normalized for their emission at 394 nm. All variants except the 1.0-IDP control showed excimer fluorescence at 450 nm, which decreased with decreasing concentration. (b) I1/I3 and I1/ex ratios derived from the pyrene fluorescence spectra for 1.0-IDP-2Yx2A and (+)1.0-IDP-2Yx2A variants at concentrations of 0.01-100 μM. CMC values were determined by fitting a logistic function to three independently run data sets of both the I1/I3 ratio: 1.0-IDP-2Yx2A: (EC_{20}, EC_{50}) = I1/I3: ($7.4 \pm 0.4, 22.6 \pm 1.8$ μM), (+)1.0-IDP-2Yx2A: ($15.1 \pm 3.7, 61.3 \pm 24.9$ μM), and the I1/ex ratio: 1.0-IDP-2Yx2A: (EC_{20}, EC_{50}) = I1/I3: ($2.3 \pm 0.1, 7.7 \pm 0.3$ μM), (+)1.0-IDP-2Yx2A: ($5.2 \pm 0.8, 24.4 \pm 7.3$ μM). (c) DLS analysis of IDP-2Yx2A variants loaded with hydrophobic pyrene shows that size was unaffected by the presence of 2 μM pyrene at high protein concentrations. As protein concentration was decreased, an increase in size is observed for all variants. (d) AFM analysis of 1.0-IDP-2Yx2A loaded with SN38 exhibits spherical micelles with Zmax and diameter values of 5.76 ± 1.36 nm and 25.66 ± 6.11 nm, respectively.

having a larger CMC than the other variants, it also swells to form larger particles in comparison to the other variants at the same concentration (Figure 3.3c). This phenomenon is likely due to a swelling effect around the hydrophobic molecule core as the concentration of protein in the solution decreases relative to the hydrophobic molecule.

One major application of bio-surfactants is their use in the delivery and solubilization of hydrophobic compounds. To test if active drug molecules could be associated with the protein micelles, 1.0-IDP-2Yx2A (5 µM) was loaded with the hydrophobic chemotherapeutic agent SN38 (40 µM), a topoisomerase I inhibitor primarily used in the treatment of colorectal cancer (see Chapter 2).[28] These particles were examined by AFM to determine if the size or morphology was changed by the presence of SN38. Indeed, spherical micelles with a similar Zmax (loaded: 5.76 ± 1.36 nm, unloaded: 5.42 ± 1.84 nm) and diameter (loaded: 25.88 ± 6.11 nm, unloaded: 29.88 ± 9.36 nm) were observed, indicating that neither size nor morphology is appreciably altered by the presence of SN38 (Figure 3.3d). Additionally, the solubility of the hydrophobic compound Tinopal OB®, an optical brightener desired for its whitening properties, was measured for 1.0-IDP-2Yx2A and 0.25-IDP-2Yx2A to confirm comparable solubilization abilities among the length variants (Additional Figure 3.15).[29]

3.3.2 Exploring hydrophilic-lipophilic balance

For traditional surfactants, the hydrophilic-lipophilic balance (HLB) is a metric used to predict how these materials will perform in emulsion systems. The oil-to-water solubility level depends upon the portions of hydrophilic and hydrophobic groups present. These values indicate whether a surfactant will behave more appropriately as a water-in-oil emulsifier, oil-in-water emulsifier, wetting agent, detergent, or solubilizer (Table 3.1).[30,31] Surfactants with low HLB values are considered lipophilic, and those with high values are hydrophilic. In the 1950s, Griffin developed a method to calculate HLB values ranging from 1-20 based on the molecular weight of a molecule's hydrophilic and hydrophobic groups.[32,33] The HLB provides an essential metric for choosing and developing the appropriate surfactant for the desired application. No such metric has been developed for protein-based bio-surfactants; however, there is interest in how similar properties could be measured and manipulated.

Amino acids have also been classified based on their hydrophobic and hydrophilic nature. Multiple hydrophobicity scales for the 20 naturally occurring amino acids have been developed; however, Kyte and Dolittle developed a more commonly used scale.[34] In reference to the simplest amino acid, glycine, this scale denotes positive values for hydrophobic amino acids and negative values for hydrophilic amino acids. Using these values, we were interested to see if we could create our own HLB metric

Table 3.1: Ionic Surfactant HLB values. These values are proportional to the hydrophilic head group polarity.

HLB	Application
4-6	Water-in-Oil Emsulifier
7-15	Wetting Agent
8-18	Oil-in-Water Emulsifier
10-15	Detergent
10-18	Solubilizer

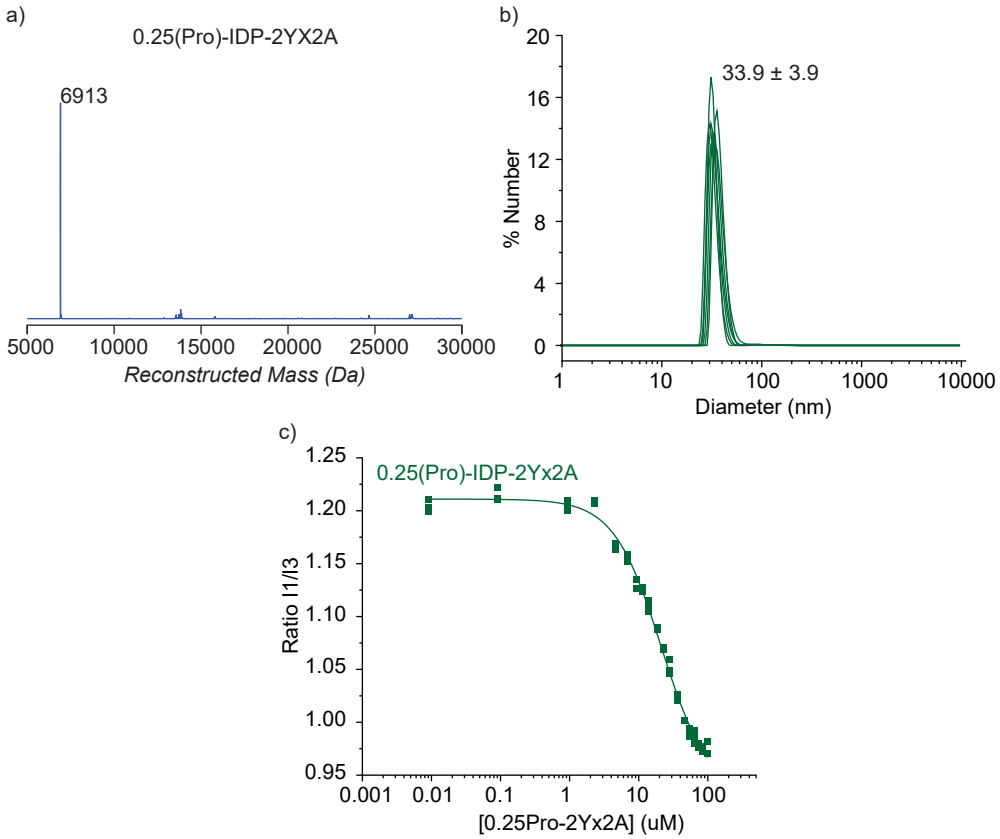


Figure 3.4: Characterization of 0.25(Pro)-IDP-2Yx2A. (a) Mass spectra of the purified variant: 0.25(Pro)-IDP-2Yx2A expected: 6913 Da, observed: 6913 Da. (b) Mean diameter values determined by DLS for 0.25(Pro)-IDP-2Yx2A in 100 mM PB, pH 5.7: 33.9 ± 3.9 nm. (c) I1/I3 ratio derived from the pyrene fluorescence spectra for 0.25(Pro)-IDP-2Yx2A variant at concentrations of 0.01-100 µM. CMC values were determined by fitting a logistic function to three independently run data sets of both the I1/I3 ratio. CMC = 7.0 ± 0.3 µM.

for proteins (HLB-P) to influence our design choices for bio-surfactant applications. Initially, we were interested in increasing the hydrophilic to lipophilic balance in the IDP sequence, which would correspond to a higher HLB value for traditional surfactants. We calculated our HLB-P values by summing the amino acid hydrophobicity values of our IDP domain and divided them by the total number of amino acids ($\mathrm{Index_{hydrophilic}/N_{aa}}$). We calculated the Index $_{\mathrm{hydrophilic}}/\mathrm{N_{aa}}$ of the 0.25IDP to be 3.7. We wanted to increase the overall hydrophilicity of this sequence, so we exchanged all valine amino acids in the sequence with prolines to preserve the disordered nature of the protein. The new Index $_{\mathrm{hydrophilic}}/\mathrm{N_{aa}}$ sequence was calculated as -19.

Using the insoluble expression system described previously,[1] a variant of 0.25-IDP-2Yx2A with valine (V) to proline (P) mutations (0.25(Pro)-IDP-2Yx2A) was expressed, purified, and characterized by MS (Figure 3.4a). To determine if this new variant had any assembling characteristics, DLS and pyrene assays were conducted as described previously to determine the diameter and CMC values. DLS measurements showed that the 0.25(Pro)-IDP-2Yx2A has a similar but slightly larger diameter (33.9 nm, Figure 3.4b) than its valine counterpart. Correspondingly the CMC value determined by the I1/I3 ratio of pyrene fluorescence was approximated to be 7 µM (Figure 3.4c). These results are promising and show that our amphiphilic IDPs still have assembling and surfactant-like behavior when the hydrophilicity of the soluble domain is significantly altered. Further investigations into the morphology, surface tension, and solubilizing capabilities are underway to understand this sequence's full scope and application. By rationally designing new IDP variants, we hope to begin to explore the effects the amino acid sequence of the hydrophilic domain can control or predict surfactant properties, much like the HLB for traditional surfactants.

3.4 Conclusion and Outlook

In summary, it was found that mutations to the hydrophilic IDP domain of IDP-2Yx2A protein micelles can impact the physicochemical properties of self-assembly. Using an insoluble expression system, IDP-2Yx2A variants with high positive charge densities and variable hydrophilic domain lengths were easily expressed and purified. In contrast to the 1.0-IDP protein alone, when the 2Yx2A hydrophobic domain was appended, self-assembly into micelle structures was observed for all variants. As the IDP domain length is decreased, spherical micelle morphology was maintained and the particle diameters remained similar in size.The ability to harbor hydrophobic molecules in the micelle interior without disrupting morphology was further confirmed for all variants with the 2Yx2A assembly domain. Importantly, it was found

that by increasing the charge of the hydrophilic IDP domain by replacing every gultamic acid residue with a lysine residue, the CMC value of the positively charged variant (+)1.0-IDP-2Yx2A was significantly higher than that of the neutral IDP sequence, indicating that this variant formed a less stable assembly likely due to charge repulsion. Preliminary results show that increasing the hydrophilicity of the IDP domain does not impede assembly behavior; however, further studies are being conducted to understand this parameter fully.

This study represents a systematic approach to investigating the impact of domain charge, length, and hydrophilicity on the self-assembly of protein micelles based on this IDP sequence. Overtime, data set such as these are envisioned to build a complete picture of the self-assembly behavior of these IDP-based bio-surfactants. However, further studies using this diverse insoluble expression system will aim to expand the sequence space and examine new parameters that may influence the physicochemical properties of these bio-surfactants.

3.5 Materials and Methods

3.5.1 Materials

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Milli-Q (MQ) H_2O was purified to a resistivity of 18.2 M Ω • cm (at 25 °C) using a MQ Gradients ultrapure water purification system (Millipore).

- 1000x ampicillin solution: 100 mg/mL of solid antibiotic were dissolved in Milli-Q H_2O. The solution was frozen at -20 °C until use.
- 1000x carbenicillin solution: 100 mg/mL of solid antibiotic were dissolved in Milli-Q H_2O. The solution was frozen at -20 °C until use.
- 1000x chloramphenicol solution: 25 mg/mL of slid antibiotic were dissolved in EtOH. The solution was frozen at -20 °C until use.
- LB agar plates: 12.5 g of LB broth (powder) from Thermo Fisher Scientific and 7.5 g agar was dissolved in 500 mL of Milli-Q H_2O and subsequently autoclaved. The solution was cooled to 50 °C and the desired antibiotic was added. The LB agar + antibiotic soultion was poured into sterile Petri dishes (100 mm x 15 mm) and solidified at room temperature. The plates were then stored at 4 °C until use.

- LB media: 20 g of Lauria broth (LB) powder (Sigma Aldrich) was dissolved into 1000 mL of Milli-Q H_2O and autoclaved.
- Lysis buffer: 20 mM Hepes, 300 mM NaCl, and 10 mM β-mercaptoethanol were added to Milli-Q H_2O.
- The pET-31b(+) vector was purchased from Millipore Sigma (catalog number: 69952-3).
- AlwN I (catalog #R0514S) and Xho I (catalog #R0146S) restriction endonucleases were purchased from New England BioLabs.
- Primers for overhang PCR were purchased from Integrated DNA Technologies (IDT).
- The Geneblock for IDP-2Yx2A was purchased from Integrated DNA Technologies(IDT) and was used as template DNA.

3.5.2 Instrumentation

3.5.2.1 Atomic Force Microscopy (AFM)

AFM samples were prepared in 30 mM phosphate buffer pH 3.7 at concentrations above 10 µM. Prior to analysis, the samples were filtered through a 0.22 µm cellulose acetate centrifugal filter. A 100 µL portion of the sample was saved and analyzed by DLS for a direct comparison of the size based on the two methods. The remaining sample was used for AFM analysis. Immediately prior to deposition on the freshly stripped mica grids, the sample was diluted with H_2O to a concentration of 1-5 µM. Using a sterile pipette, 2 µL of sample was placed in the center of the grid for 2 min. The tip of a Kimwipe was used to remove excess liquid and the sample was further dried under a steady stream of N_2 gas. The sample was loaded onto the AFM stage and imaged under soft tapping mode with the following tip: PPP-NCHAuD-10, PointProbePlus, tapping mode, 330 KHz w/Au coating detector side, 125 µm (NanoAndMore USA). Multiple images per sample were acquired. Images were acquired in the following order: bare mica, IDP, 1.0-IDP-2Yx2A, 0.25-IDP-2Yx2A, 0.5-IDP-2Yx2A, 0.75-IDP-2Yx2A.

3.5.2.2 Dynamic Light Scattering (DLS)

Dynamic light scattering was performed on a Malvern Instruments Zetasizer Nano ZS. Prior to analysis, lyophilized samples were re-hydrated in the appropriate buffer

and passed through a 0.22 µm cellulose acetate centrifugal filter. Protein concentration was determined by diluting the sample 1:10 in 50% MeCN in H_2O and measuring the absorbance at 280 nm on a NanoDrop 1000 (Thermo). The filtered samples were then diluted with buffer to the desired concentration and 80-100 µL was then added to the DLS cuvette. For room temperature analysis, (25 °C) was used. Data plots and standard deviations were calculated from an average of the %Number mean for each run. Measurement data was presented as averaged mean diameters determined by the %Number distribution. The %Number was used in this analysis because it shows the size of the majority population in a sample. Heterogeneity was evident in the micelle samples as evidenced by the polydispersity and %Intensity traces. It is likely that larger structures or inter-micelle interactions occur and form a small number of larger particles that contribute to larger diameters when analyzed by %Intensity, which heavily weights large particles in a sample as opposed to the most abundant.

3.5.2.3 Reverse Phase High Performance Liquid Chromatography (RP-HPLC)

RP-HPLC purification of IDP-2Yx2A protein mutants was performed on an Agilent 1260 Infinity series system equipped with a UV-Vis detector and a fraction collector, controlled by Open Lab CDS software (Agilent Technologies Inc. 2011 –2015). HPLC grade Milli-Q H_2O + 0.1% TFA and HPLC grade MeCN + 0.1% TFA were used as mobile phases A and B, respectively. Preparatory scale purification was performed on a Waters XBridge Protein BEH C4 OBD Prep Column (300 Å, 5 µm, 19 mm X 50 mm) with a flow rate of 5 mL/min over a 5-95% gradient over the course of 30 min. Fractions containing IDP-2Yx2A mutants were identified by their coinciding 280 nm absorbance and 350 nm fluorescence profiles.

3.5.2.4 Mass Spectrometry of Proteins (TOF-ESI-MS)

Time of flight electrospray ionization mass spectrometry (TOF-ESI-MS) of protein samples was performed using an Agilent 1260 series liquid chromatography (LC) system coupled with an Agilent 6224 TOF LC-MS (Santa Clara, CA). Protein samples were first passed through a 0.22 µM cellulose acetate filter and then placed in HPLC vials for analysis. Depending on sample concentration, 0.5 -10 µL was injected on to a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) analytical column with a flow rate of 0.30 mL/min and a gradient of 5-100% MeCN+0.1% (v/v) formic acid in Milli-Q H_2O+0.1% formic acid over the course of 8 min. The total ion chromatograms were extracted in the Agilent Mass Hunter Workstation software

(Qualitative Analysis Version B.05.00, Build 5.0.519.13, Agilent Technologies Inc. 2011). Protein molecular weights were reconstructed using a maximum entropy deconvolution. The resulting spectra were exported as CSV files and plotted using the Chartograph web-tool(www.chartograph.com).

3.5.2.5 UV-VIS DNA and Protein Measurements

Estimates of DNA and protein concentration were calculated by measuring the sample absorbance at 260 nm and 280 nm, respectively, on a NanoDrop 1000 (Thermo). Using Beers law and the calculated extinction coefficients, the concentrations of both DNA and protein were approximated.

3.5.2.6 Fluorescent Measurements

Fluorescent meaurements were made on a Fluoromax-4 Spectrofluorometer (HORIBA Scientific). A calibration curve of Tinopal OB®in 50% MeCN was created by measuring the emission at 435 nm after excitation at 350 nm with a 5 nm slit. A Tinopal OB®stock solution in DMSO was added in excess to 1.0-IDP-2Yx2A and 0.25-IDP-2Yx2A in 50% MeCN. The samples were sonicated at 65% amplitude for 2 min (15 s on: 15 s off). Samples were then lyophilized to remove all remaining organic solvents and resusended in 10 mM phostphate buffer at pH 5.7. To ensure full solubilization, samples were again sonicated at 65% amplitude for 2 min (15 s on: 15 s off). Samples were then diluted 1:100 in 50% MeCN and their fluorescence was measured and the concentration of Tinopal OB® was calculated from the calibration curve.

3.5.2.7 Goniometer Contact Angle Measurements

Contact Angle Measurements were performed on a Rame-Hart model 210 goniometer with DROPimage Pro. Contact angles were determined using the open-source Fiji software.[35] For each triplicate sample, a 4 µL drop was placed on clean glass coverslip and a snapshot was recorded by the DROPimage Pro software.

3.5.3 Methods for protein production

3.5.3.1 Construction of KSI-(0.25/0.5/0.75/1.0)-IDP-2Yx2A pET-31b(+) plasmids

Overhang PCR was used to construct four plasmids that encoded for mutants where the IDP sequence had been shortened to 75% (0.75-IDP-2Yx2A) 50% (0.5-IDP-2Yx2A) and 25% (0.25-IDP-2Yx2A) of the original sequence (1.0-IDP-2Yx2A). The

Geneblock for the full length sequence 1.0-IDP-2Yx2A was used as template DNA for all mutants. For overhang PCR, all mutants used the same reverse primer while the annealing location of the forward primers was moved further up sequence for each subsequent mutant. For the construction of the plasmid KSI-1.0-IDP, QuikChange PCR was performed on the KSI-1.0-IDP-2Yx2A pET-31b(+) plasmid to insert a stop codon at the end of the 1.0-IDP sequence prior to the 2Yx2A hydrophobic domain.

Primers for construction of KSI-1.0-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' GGTGGTCATATGCTGGCAGAGGCAAAG 3'
Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock 1.0-IDP-2Yx2A: 5' AAAAATCAGATGCTGGCAGAGGC
AAAGAGTCCTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGC
CGTCGCGAAAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAG
CCCGGCAGAAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGT
TAAAAGCCCAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCA
GAAGCCAAGTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGC
CGGCCGAAGCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTA
AGTCCCCGGCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCA
CCGTAAAATCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCC
GGTGACGGTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAA
ATCACCATACTGGAGTGCCTATGGCGCGTATGCGCAGTATGTGTATATT
TATGCGTATTGGTATCTGTGACTCGAGTTTAAA 3'
AlwNI cut site shown in red, XhoI cut site shown in cyan

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-2Yx2A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYVYIYAYWYL*
KSI protein sequence shown in magenta

Primers for construction of KSI-(+)1.0-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' TTCCAAAAAAAATCAGATGCTGGCA 3'
Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock (+)1.0-IDP-2Yx2A: 5' AAAAATCAGATGCTGGCAAAGG
CAAAAAGTCCGGCCAAAGCTAAATCTCCTGCAAAAGTCAAGTCTCCAGC
CGTCGCAAAGAGCCCCGCAAAAGTAAAAAGTCCAGCTAAGGTAAAGAG

CCCTGCAAAAGCAAAAAGCCCAGCGAAGGCTAAGAGTCCCGCTAAAGTT
AAAAGCCCGGCCACAGTAAAGTCCCCTGGAAAGGCAAAATCTCCGGCAA
AAGCCAAATCACCTGCCAAAGTGAAGTCACCTGTAAAAGCCAAATCGCC
GGCTAAGGCGAAAAGCCCAGCGAGCGTTAAGTCCCCCGGTAAGGCAAA
ATCGCCAGCGAAAGCAAAGTCTCCCGCAAAAGTTAAGAGTCCTGCAACC
GTGAAATCTCCAGTTAAGGCCAAGTCGCCAGCGAAAGTTAAATCTCCGG
TGACAGTCAAATCCCCTGCGAAAGCGAAATCCCCCGTGAAGGTGAAGTC
ACCGTACTGGAGCGCATATGGAGCGTATGCGCAGTATGTGTATATATA
TGCGTATTGGTACCTATGACTCGAGTTTAAA 3'
AlwNI cut site shown in red, XhoI cut site shown in cyan

Corresponding amino acid sequence of fusion protein KSI-(+)1.0-IDP-2Yx2A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAKAKSPAKAKSPAKVKSPAVAKSPAK
VKSPAKVKSPAKAKSPAKAKSPAKVKSPATVKSPGKAKSPAKAKSPAKVKS
PVKAKSPAKAKSPASVKSPGKAKSPAKAKSPAKVKSPATVKSPVKAKSPAK
VKSPVTVKSPAKAKSPVKVKSPYWSAYGAYAQYVYIYAYWYL
KSI protein sequence shown in magenta

Primers and template for KSI-0.75-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' AAAAATCAGATGCTGAAAATGCCGGCAGAAGCAAAAA 3'
Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock 1.0-IDP-2Yx2A

Corresponding amino acid sequence of fusion protein KSI-0.75-IDP-2Yx2A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLKMPAEAKSPAEAKSPAEVKSPATVKS
PGEAKSPAEAKSPAEVKSPVEAKSPAEAKSPASVKSPGEAKSPAEAKSPAEV
KSPATVKSPVEAKSPAEVKSPVTVKSPAEAKSPVEVKSPYWSAYGAYAQYV
YIYAYWYL*
KSI protein sequence shown in magenta

Primers and template for KSI-0.50-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' AAAAATCAGATGCTGAAAATGCCTGCCGAAGTGAAGT 3'
Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock 1.0-IDP-2Yx2A

Corresponding amino acid sequence of fusion protein KSI-0.50-IDP-2Yx2A :
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLKMPAEVKSPVEAKSPAEAKSPASVKS
PGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKSPVTVKSPAEAKSPVEV
KSPYWSAYGAYAQYVYIYAYWYL*
KSI protein sequence shown in magenta

Primers and template for KSI-0.25-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' AAAAATCAGATGCTGAAAATGCCAGCGGAAGTAAAGAGT 3'
Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock 1.0-IDP-2Yx2A

Corresponding amino acid sequence of fusion protein KSI-0.25-IDP-2Yx2A :
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLKMPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYVYIYAYWYL*
KSI protein sequence shown in magenta

Primers for construction of KSI-1.0-IDP pET-31b(+) plasmid
KSI-1.0-IDP mutant was prepared by performing a site directed mutagenesis to encode a stop codon after the IDP sequence. Aglients QuikChange mutagenesis kit and protocols were used. Primer 1: 5' CGCGCCATAGGCACTCTAGTATCCTGATTTCACCTCCACGG 3'
Primer 2: 5' CGCGCCATAGGCACTCTAGTATCCTGATTTCACCTCCACGG 3'
Template: KSI-1.0-IDP-2Yx2A pET-31b(+) plasmid
Corresponding amino acid sequence of fusion protein KSI-1.0-IDP :
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSGY*
KSI protein sequence shown in magenta

Template for KSI-0.25(Pro)-IDP-2Yx2A pET-31b(+) plasmid
Forward: 5' AAAAATCAGATGCTGAAAATGCCAGCGGAAGTAAAGAGT 3'

Reverse: 5' TTTAAACTCGAGTCACAGATACC 3'
Template: Geneblock 0.25(Pro)IDP-2Yx2A: 5' CCAGCGGAACCAAAGAGTCCG GCCACCCCAAAATCTCCGCCAGAGGCCAAGAGTCCGGCGGAACCAAAAT CTCCGCCAACGCCAAAGTCCCCAGCGGAAGCGAAGTCCCCCCCAGAGCC AAAATCACCATACTGGAGTGCCTATGGCGCGTATGCGCAGTATGTGTAT ATTTATGCGTATTGGTATCTGTGACTCGAGTTTAAA 3'

Corresponding amino acid sequence of fusion protein KSI-0.25-IDP-2Yx2A :
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR FNGAGKVVSIRALFGEKNIHACQMLKMPAEPKSPATPKSPPEAKSPAEPKSP PTPKSPAEAKSPPEPKSPYWSAYGAYAQYVYIYAYWYL*
KSI protein sequence shown in magenta

3.5.3.2 PCR reaction conditions

PCR reactions were performed using the components as listed in Table 3.2. After mixing the reaction components in 0.5 mL PCR tubes, overhang PCR was performed on a BioRad S1000 Thermal Cycler with an annealing temperature of 55 °C. The PCR reaction was then removed from the thermocyler and 6 µL of 6X loading dye was added to the reaction tubes and mixed. The PCR reaction and a DNA standard(O'gene ruler 1 Kb) were then loaded into adjacent wells of a 1% agarose gel prestained with SYBR Safe (ThermoFisher). The gel was run at 120 V for 35 min in 1X TAE buffer. The resulting gel was imaged under UV fluorescence with a BioRad GelDoc EZ Imager. The DNA bands corresponding to the desired PCR product were excised from the gel and placed in a pre-weighed 1.5 mL Eppendorf tube. The amplified PCR product was removed from the gel using the quick gel extraction kit (Promega).

3.5.3.3 PCR product ligation reaction with pET-31b(+) plasmid

The PCR products and the pET-31b(+) entry vector were digested with 10 units of XhoI and AlwnI restriction enzymes for 1 h at 37 °C with a subsequent 20 min deactivation step at 80 °C. The cut entry vector and insert were then run on a 1% agarose gel prestained with SYBR Safe (ThermoFisher). The gel was run at 120 V for 35 min in 1X TAE buffer then imaged under UV fluorescence with a BioRad GelDoc EZ Imager. The DNA bands corresponding to the cut entry vector and insert were excised from the gel and placed in a pre-weighed 1.5 mL Eppendorf tube. The

Table 3.2: Standard PCR Reaction Protocol

PCR Component	Volume (µL)
Template DNA[a] (50 ng/µL)	0.5
Forward Primer (10 µM)	1
Reverse Primer (10 µM)	1
5X phusion buffer	4
dNTPs (10 mM)	0.4
DMSO	0.6
Milli-Q H_2O	12
Phusion Enzyme[b]	0.5

[a] Geneblock (1.0-IDP-2Yx2A) [b] Phusion is added last

Table 3.3: Standard Ligation Reaction Protocol

Ligation Component	ratio insert : vector (0:1)	(3:1)	(5:1)
cut pET-31b(+)	50 ng	50 ng	50 ng
Insert (500 bp)	0 ng	15 ng	25 ng
10X T4 buffer	1 µL	1 µL	1 µL
Milli-Q H_2O	to 10 µL	to 10 µL	to 10 µL
T4 Ligase (400 units/µL) [b]	0.5 µL	0.5 µL	0.5 µL

[b] T4 ligase is added last

digested insert and vector were removed from the gel using the quick gel extraction kit (Promega). The resulting complementary sticky ends of the entry vector and insert were then ligated together using the components listed in Table 3.3.

3.5.3.4 transformation of ligation products into competent cells

The ligation reaction was incubated at 16 °C for 16 h and then heat inactivated for 20 min at 80 °C. The entire ligation reaction was transformed into 50 µL of chemically competent XL1-Blue cells by first incubating the ligation reaction and cells on ice for 30 min and then incubating at 42 °C for 45 s. The cells were rescued by the addition of 1 mL of SOC media and incubated with rotary shaking at 37 °C. After 1 h 200 µL of the transformed cells were plated on carbenicillin (+) LB agar plates and grown overnight at 37 °C. The number of colonies from on the plates where inset had been added (3:1) and (5:1) were compared to the growth of the control (0:1) to determine how many colonies to select for sequencing. An appropriate number of colonies was selected to account for background vector re-ligation and placed in individual 10 mL culture tubes containing approximately 5 mL of LB media and 5

µL of 1000x carbenicillin. Cultures were grown overnight at 37 °C. The following day, 1 mL of each culture was used to extract the plasmid DNA using a Miniprep Kit (Millipore), which were then sent for sequencing.

3.5.3.5 Protein Expression

KSI-IDP-2Yx2A mutants were expressed using the following protocol. Plasmids containing the desired fusion proteins were transformed into Rosetta2 (pLysS) cells and plated onto carbenicillin(+)/chloramphenicol(+) LB Agar plates and grown overnight at 37 °C. Single colonies were selected and placed into 50 mL of LB with 50 µL of both 1000x stocks of carbenicillin and chloramphenicol antibiotics. The culture was grown overnight at 37 °C. The following morning, 30 mL of the 50 mL culture was used to innoculate 1000 mL of LB with 1000 µL 1000x stocks of carbenicillin and chloramphenicol in a 2.8 L baffled flask. Cultures were grown at 37 °C with shaking at 220 rpm until an OD_{600} of 0.6 was reached, at which point the temperature was lowered to 18 °C. After cooling for 20 min, IPTG was added to a final concentration of 0.25 mM. The protein was expressed overnight for 18-24 h. The culture was then centrifuged at 4000 xg for 20 min. The resulting protein pellets were transferred to 50 mL falcon tubes and frozen at -20 °C. A typical expression yields approximately 5-10 g cell pellets.

3.5.3.6 Protein Purification

The frozen cell pellet was resuspended in 30 mL of lysis buffer, 30 µL of 1% Trition-X, and 300 µL of a 0.1 M PMSF solution (Sigma Aldrich). The cell pellet was vortexed until completely resuspended and then sonicated at 70% amplitude for 10 min (2 s on: 4 s off). The lysed cells were then centrifuged at 14,500 xg for 20 min. The supernatant containing the soluble proteins was discarded and the insoluble fraction was washed by resuspending it in 30 mL of lysis buffer and centrifuging again at 14,500 xg for 20 min. The washing step was repeated three times. After the third wash, 5 mL of 6 M guanidinium HCl (Guan HCl) was added to the insoluble fraction and the protein was resuspended by vigorous shaking for 1 h at room temperature and sonication at 70% amplitude for 10 min (2 s on: 4 s off). The protein solution was then centrifuged at 14,500 xg for 20 min at which point the KSI-IDP-2Yx2A fusion protein is solubilized, while the pelleted remains are largely composed of cell debris. The supernatant containing the fusion protein was transferred to a sterile 50 mL falcon tube. The approximate concentration was determined by measuring the UV absorbance at 280 nm on a NanoDrop 1000 (Thermo) and the calculated extinction

coefficient (ExPASy Protein Parameters Tool https://web.expasy.org/protparam/) while the protein purity is assessed by SDS PAGE gel analysis.

The KSI protein was cleaved from the IDP-2Yx2A sequence by the slow addition of 3 M HCl, concentrated formic acid, and cyanogen bromide(CNBr) to a final concentrations of 0.5 M HCl, 10% formic acid, and 100 molar equivalents of CNBr to protein. The solution was stirred in the dark under a steady flow of nitrogen gas. After 24 h, the reaction is mostly complete when analyzed by TOF-ESI-MS. The protein was then precipitated out of solution by the addition of cold saturated $(NH_4)_2SO_4$ solution. The precipitation solution was kept at -20 °C for 1 h. The precipitated protein was separated from the soluble components (CNBr, Guan HCl, and salts) by centrifugation at 14,500 xg for 20 min. The supernatant was discarded in a closed top waste receptacle specified for CNBr. The protein pellet was resuspended in H_2O or, for more hydrophobic constructs, up to 40% MeCN in H_2O by sonication at 70% amplitude for 10 min (2 s on: 4 s off) followed by a subsequent centrifugation at 14,500 xg for 20 min to remove any insoluble protein. The resulting protein solution was reasonably pure when analyzed by TOF-ESI-MS and SDS PAGE gel, however, an additional purification step using reverse phase HPLC (RP-HPLC) was performed to ensure maximum purity. RP-HPLC purification was performed in 1 mL injections on a Waters XBridge Protein BEH C4 OBD Prep Column (300Å, 5 µm, 19 mm X 50 mm) with a 5-95% gradient of MeCN/H_2O +1% TFA over 30 min. Fractions containing the IDP-2Yx2A protein mutants eluted between 30-50% MeCN and were combined and lyophilized. Protein purity was assessed by SDS PAGE gel and TOF-ESI-MS analysis.

3.6 Additional Figures

3.6.1 DLS analysis of IDP domain

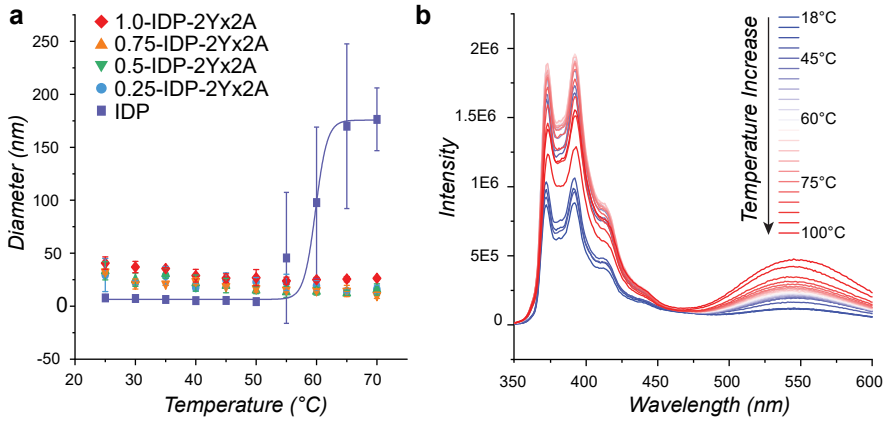


Figure 3.5: Effect of increased temperature on IDP hydrophilic domain. (a) The 1.0-IDP hydrophilic control mutant aggregated into larger structures when heated above 50 °C as observed by DLS. This behavior was not observed for the mutants with the 2Yx2A hydrophobic domain up to 70 °C, the range of detection for this DLS instrument (b) Pyrene excimer florescence was observed when the 1.0-IDP mutant is heated in the presence of a 2 µM pyrene buffer.

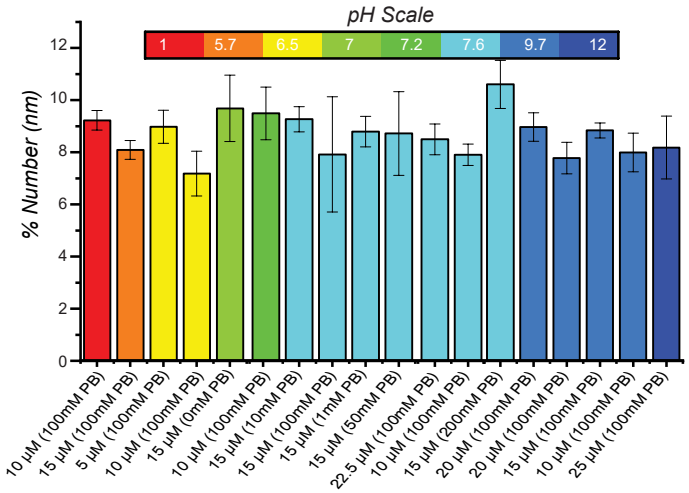


Figure 3.6: DLS of the averaged mean diameters by %Number of the 1.0-IDP hydrophilic control. No evidence of aggregation at 25 °C over a broad range of pH and buffer conditions was observed for the 1.0-IDP hydrophilic control.

3.6.2 Disorder Profiles

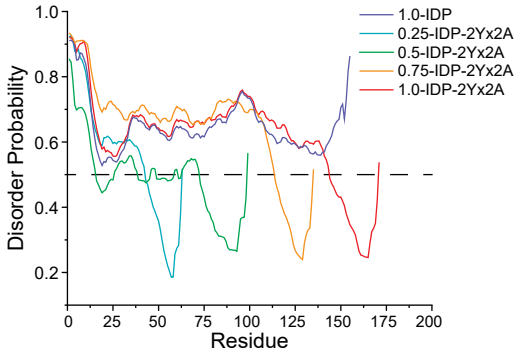


Figure 3.7: Intrinsically disordered profiles of IDP variants. The Protein DisOrder prediction System (PrDOS) was used to create profiles for these variants and confirmed disorder probability predictions of >0.5.

3.6.3 DLS analysis of IDP-Mutants

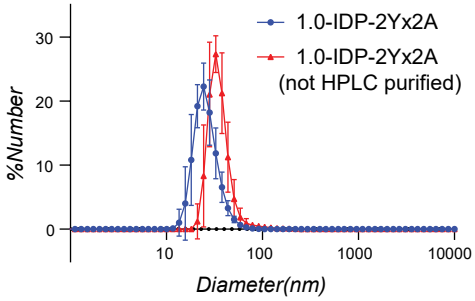


Figure 3.8: DLS traces by %Number of the 1.0-IDP-2Yx2A mutant before and after HPLC purification. The unpurified sample shows a slightly larger diameter by DLS which is likely due to minor protein or small molecule impurities.

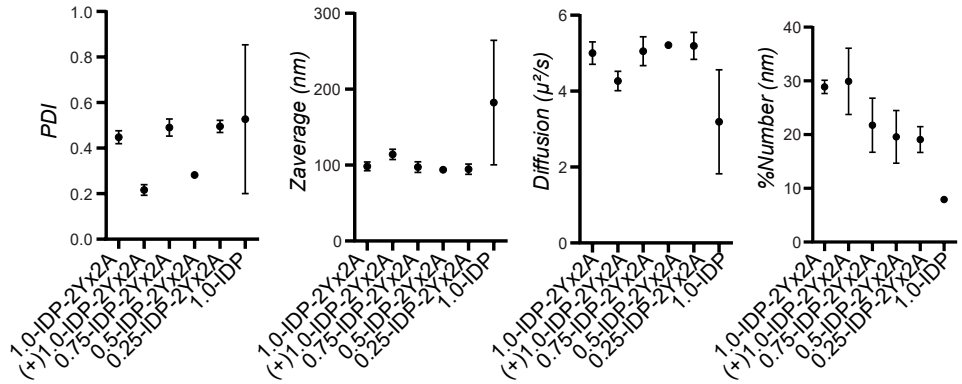


Figure 3.9: PDI, Z-average, and diffusion for all variants. The PDI for each sample indicates varying degree of polydispersity among the sample population. the 1.0-IDP control in particular is difficult to analyze by DLS due to its unstructured and dynamic nature. The Z-average size is corresponds to the intensity weight mean hydrodynamic size and is skewed larger than the diameters reported by the % Number DLS traces likely due to a small amount of dust or aggregated particles that become over represented in the %Intensity traces. The Z-average reported for biological samples are generally 2 times the actual particle size due to a lower than expected diffusion coefficient add citation. The trend of size decreasing while IDP length decreases is observed in the Zaverage and corresponding diffusion for all IDP-2Yx2A mutants. Due to the unstructured nature of the 1.0-IDP, the Z-average and diffusion coefficient appears to be inconsistent for the size of the mutant, likely due to its disordered nature.

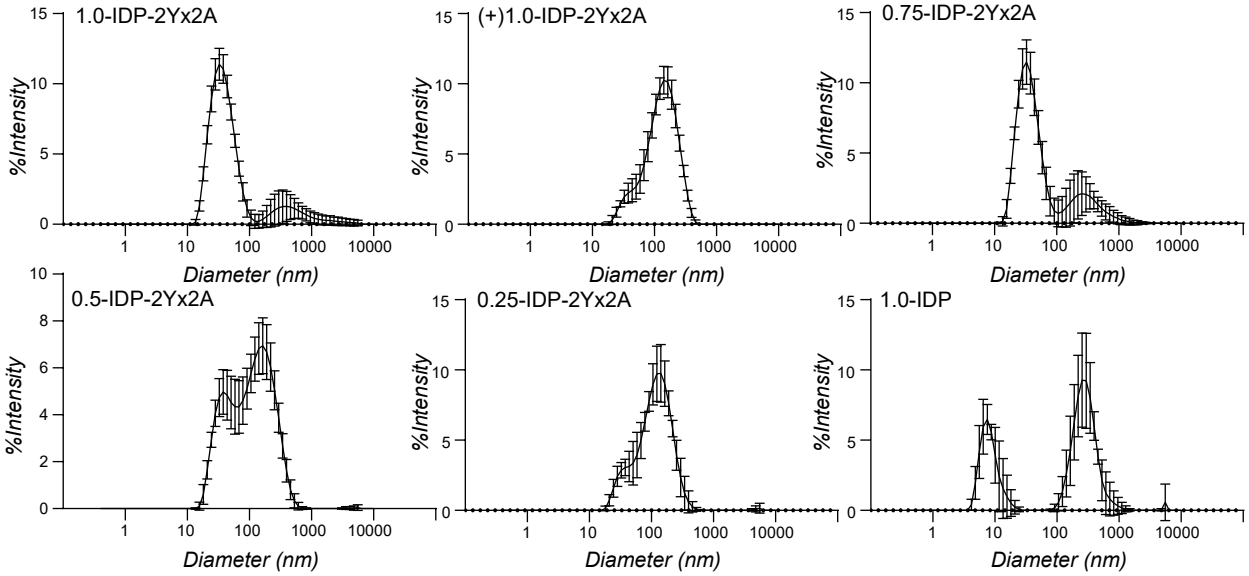


Figure 3.10: DLS traces by %Intensity of the IDP-2Yx2A mutants. Larger amounts of aggregation are observed or mutants with shorter IDP domains. The small diameter peaks in the intensity traces becomes the solely represented peak in the corresponding %Number DLS traces and the AFM traces, which show no evidence of $>$ 100 nm sized particles in any sample.

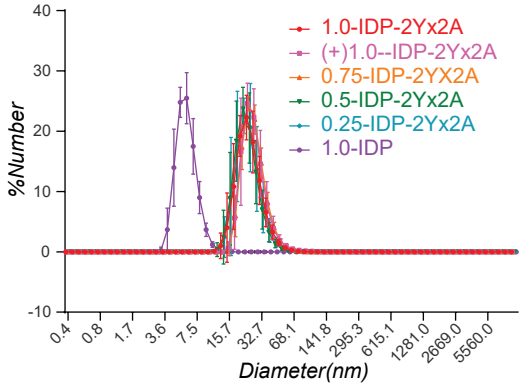


Figure 3.11: DLS traces by %Number of the IDP-2Yx2A mutants. Representative DLS traces by %Number show a clear increase in size for mutants with the 2Yx2A hydrophobic domain.

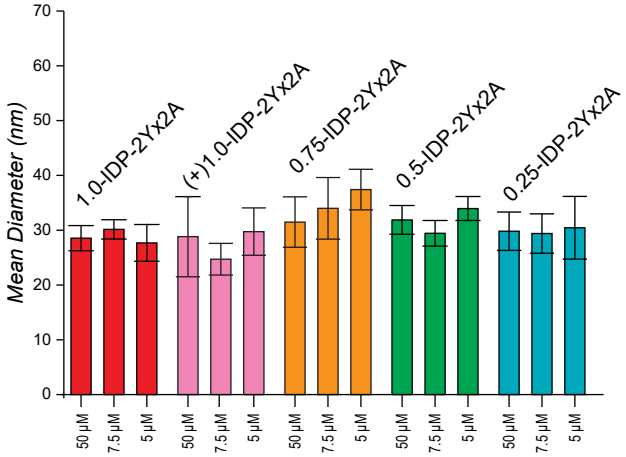


Figure 3.12: DLS dilution of IDP variants. IDP solutions in 100 mM PB, pH 5.7 were diluted from 50 µM to 5 µM. No increase in diameter was observed as the protein was diluted.

3.6.3.1 Critical Micelle Concentration

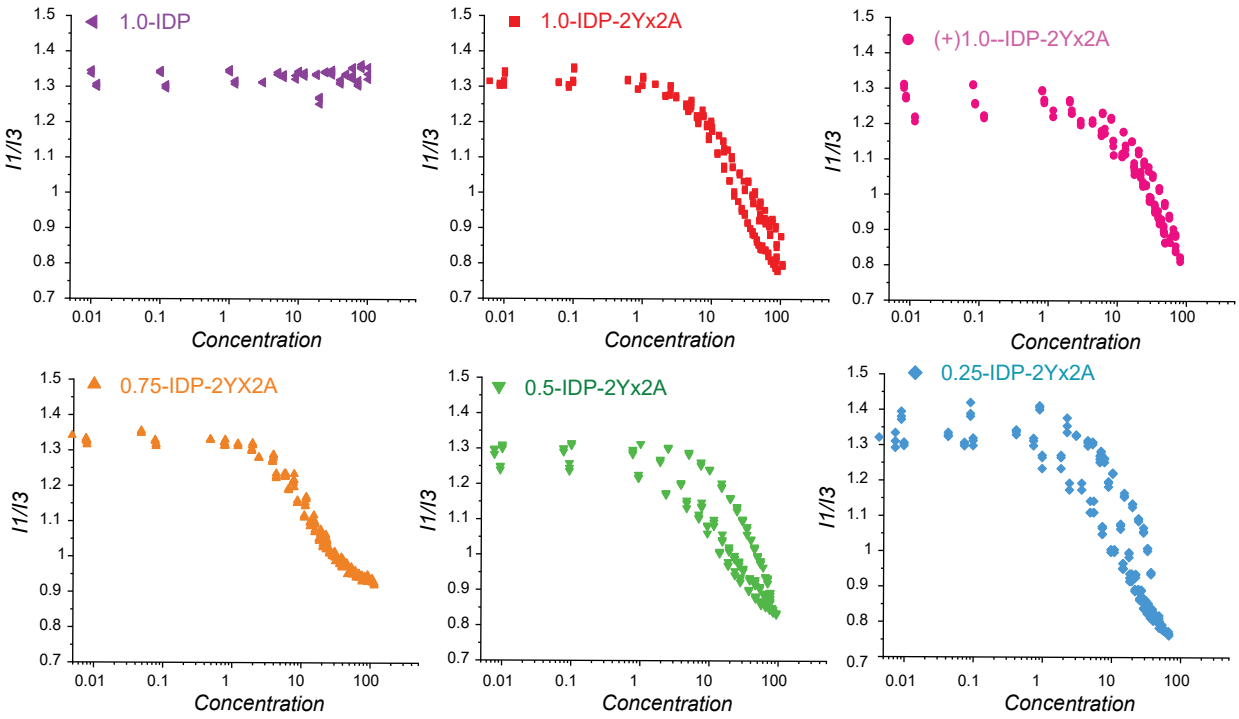


Figure 3.13: Pyrene Fluorescence Assay for the determination of the CMC. For each mutant the I1/I3 ratios are plotted against concentration. Shown are 3 independent replicates for each mutant fitted to a logistic function.

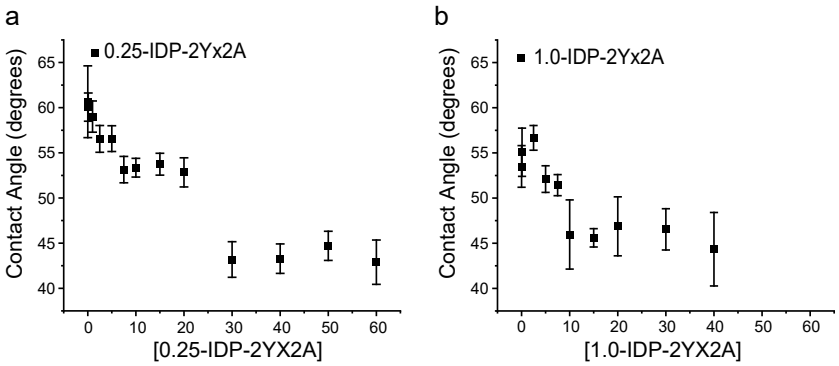


Figure 3.14: Representative contact angle measurements on glass microscope slides. (a) The contact angles of 4 µL of 0.25-IDP-2Yx2A at decreasing concentrations are shown. A sharp increase in surface wetting is seen at 20 µL and below. (b) The contact angles of 4 µL of 1.0-IDP-2Yx2A at decreasing concentrations are shown. An increase for surface wetting around 10 µM is shown.

3.6.4 Tinopal OB Solubilization

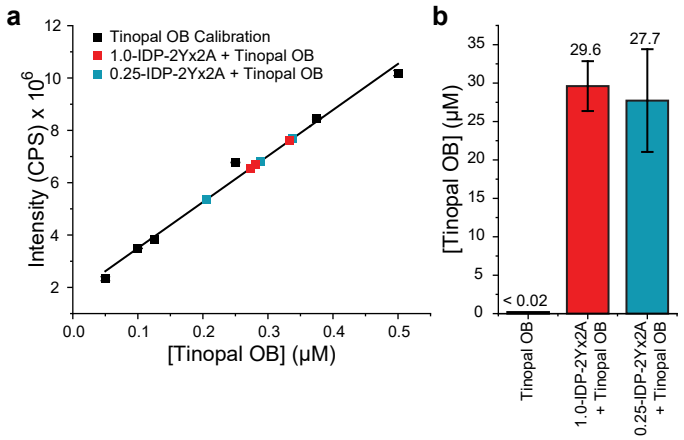


Figure 3.15: Tinopal OB®solubilization. (a) A calibration curve of fluorescent intensity at 435nm of Tinopal OB®in 50% MeCN was created. Intensities measured for a 1:100 dilution of 1.0-IDP-2Yx2A (red) and 0.25-IDP-2Yx2A (blue) in 50% MeCN are included on the fitted curve to determine initial concentration.(b) The increases in solubilized Tinopal OB®in buffer are reported for 1.0-IDP-2Yx2A (29.5µM ± 3.2) and 0.25-IDP-2Yx2A (27.7µM ± 6.7).

Table 3.4: CMCs derived from logistic fit of I1/I3 data.

I1/I3	1.0-IDP-2Yx2A	(+)1.0-IDP-2Yx2A	0.75-IDP-2Yx2A	0.5-IDP-2Yx2A	0.25-IDP-2Yx2A
EC_{20}	7.4 ± 0.4	15.1 ± 3.7	4.7 ± 0.1	7.5 ± 1.8	6.0 ± 3.6
EC_{50}	22.6 ± 1.8	61.3 ± 24.9	12.5 ± 0.3	36.8 ± 17.5	19.2 ± 3.6

Table 3.5: CMCs derived from logistic fit of I1/ex data.

I1/ex	1.0-IDP-2Yx2A	(+)1.0-IDP-2Yx2A	0.75-IDP-2Yx2A	0.5-IDP-2Yx2A	0.25-IDP-2Yx2A
EC_{20}	2.3 ± 0.1	5.2 ± 0.8	4.2 ± 0.2	2.3 ± 0.3	0.3 ± 0.1
EC_{50}	7.7 ± 0.3	24.4 ± 7.3	19.2 ± 1.5	10.0 ± 1.2	2.2 ± 0.4

3.7 Acknowledgements

This work was supported by the California research Alliance (CARA) by BASF, the UCSF Hana Jabsheh fund, National Institutes of Health Grant R01GM032543, and the Chemical Biology Graduate Program at UC Berkeley (NIH T32-GM066698). S.H.K. was supported by the University of California Cancer Research Coordinating Committee (UCCRC) predoctoral fellowship. J.M.G. was supported by the National Science Foundation (NSF) Graduate Research Fellowship Program. Some figure elements were created with the online platform: BioRender.com.

3.8 References

(1) Klass, S. H.; Gleason, J. M.; Omole, A. O.; Onoa, B.; Bustamante, C. J.; Francis, M. B. *Biomacromolecules* **2022**, *23*, 1462–1470.

(2) Ruzette, A. V.; Leibler, L. *Nature Materials* **2005**, *4*, 19–31.

(3) Deschênes, L.; Bousmina, M.; Ritcey, A. M. *Langmuir* **2008**, *24*, 3699–3708.

(4) Binder, K.; Müller, M. *Current Opinion in Colloid and Interface Science* **2000**, *5*, 314–322.

(5) Mayes, A. M.; de la Cruz, M. O. *Macromolecules* **1988**, *21*, 2543–2547.

(6) Jäger, C. M.; Hirsch, A.; Schade, B.; Böttcher, C.; Clark, T. *Chemistry - A European Journal* **2009**, *15*, 8586–8592.

(7) Stephenson, B. C.; Goldsipe, A.; Beers, K. J.; Blankschtein, D. *Journal of Physical Chemistry B* **2007**, *111*, 1025–1044.

(8) Vishnyakov, A.; Lee, M. T.; Neimark, A. V. *Journal of Physical Chemistry Letters* **2013**, *4*, 797–802.

(9) Nagarajan, R.; Ganesh, K. *The Journal of Chemical Physics* **1989**, *90*, 5843–5856.

(10) Ianiro, A.; Patterson, J.; González García, Á.; van Rijt, M. M. J.; Hendrix, M. M. R. M.; Sommerdijk, N. A. J. M.; Voets, I. K.; Esteves, A. C. C.; Tuinier, R. *Journal of Polymer Science Part B: Polymer Physics* **2018**, *56*, 330–339.

(11) Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **1976**, *72*, 1525–1568.

(12) Lechuga, M.; Fernández-Serrano, M.; Jurado, E.; Núñez-Olea, J.; Ríos, F. *Ecotoxicology and Environmental Safety* **2016**, *125*, 1–8.

(13) Cowan-Ellsberry, C.; Belanger, S.; Dorn, P.; Dyer, S.; McAvoy, D.; Sanderson, H.; Versteeg, D.; Ferrer, D.; Stanton, K. *Critical Reviews in Environmental Science and Technology* **2014**, *44*, 1893–1993.

(14) Fischer, D.; Li, Y.; Ahlemeyer, B.; Krieglstein, J.; Kissel, T. *Biomaterials* **2003**, *24*, 1121–1131.

(15) Gozlan, C.; Deruer, E.; Duclos, M. C.; Molinier, V.; Aubry, J. M.; Redl, A.; Duguet, N.; Lemaire, M. *Green Chemistry* **2016**, *18*, 1994–2004.

(16) Zhang, Q. Q.; Cai, B. X.; Gang, H. Z.; Yang, S. Z.; Mu, B. Z. *RSC Advances* **2014**, *4*, 38393–38396.

(17) Foley, P.; Kermanshahi Pour, A.; Beach, E. S.; Zimmerman, J. B. *Chemical Society Reviews* **2012**, *41*, 1499–1518.

(18) Douglas G. Hayes Daniel K. Solaiman, R. D. A., *Biobased Surfactants: Synthesis, Properties, and Applications*.

(19) Klass, S. H.; Smith, M. J.; Fiala, T. A.; Lee, J. P.; Omole, A. O.; Han, B.-G.; Downing, K. H.; Kumar, S.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 4291–4299.

(20) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. *Journal of the American Chemical Society* **2007**, *129*, PMID: 17665911, 10110–10112.

(21) Ishida, T.; Kinoshita, K. *Nucleic Acids Research* **2007**, *35*, W460–W464.

(22) Ouanezar, M.; Guyomarc'h, F.; Bouchoux, A. *Langmuir* **2012**, *28*, 4915–4919.

(23) Di Maria, F.; Blasi, L.; Quarta, A.; Bergamini, G.; Barbarella, G.; Giorgini, L.; Benaglia, M. *Journal of Materials Chemistry B* **2015**, *3*, 8963–8972.

(24) Guler, M. O.; Claussen, R. C.; Stupp, S. I. *Journal of Materials Chemistry* **2005**, *15*, 4507.

(25) Piñeiro, L.; Novo, M.; Al-Soufi, W. *Advances in Colloid and Interface Science* **2015**, *215*, 1–12.

(26) Basu Ray, G.; Chakraborty, I.; Moulik, S. P. *Journal of Colloid and Interface Science* **2006**, *294*, 248–254.

(27) J. Aguiar P. Carpena J.A. Molina-Bolivar, C. C. R. *Journal of Colloid and Interface Science* **2003**, *258*, 116–122.

(28) Nguyen, F.; Alferiev, I.; Guan, P.; Guerrero, D. T.; Kolla, V.; Moorthy, G. S.; Chorny, M.; Brodeur, G. M. *Clinical Cancer Research* **2018**, *24*, 2585–2593.

(29) Optical Brightener OB, Uvitex Tinopal OB, 7128-64-5 Mfg, en-US, (accessed Jan 28, 2022).

(30) Egito, E. S. T.; Machado, L. A.; Farias, I. E.; Farias, I. E.; Silva, K. G. H.; Oliveira, A. G. *Biomedical Journal of Scientific & Technical Research* **2018**, *10*, 001–002.

(31) Krishna, G.; Wood, G. C.; Sheth, B. B. *PDA journal of pharmaceutical science and technology* **1998**, *52*, 331–336.

(32) Griffin, W. C. *Journal of the Society of Cosmetic Chemists* **1954**, *5*, 249–256.

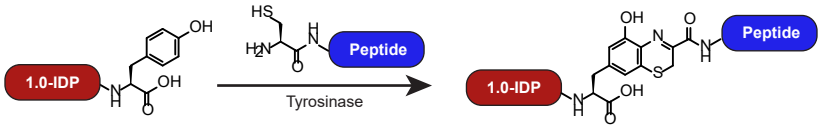
(33) Zheng, Y.; Zheng, M.; Ma, Z.; Xin, B.; Guo, R.; Xu, X. In *Polar Lipids*; Elsevier: 2015, pp 215–243.

(34) Kyte, J.; Doolittle, R. F. *Journal of Molecular Biology* **1982**, *157*, 105–132.

(35) Schindelin, J. et al. *Nature Methods* **2012**, *9*, 676–682.

Chapter 4

Sequence Exploration of the Amphiphilic Intrinsically Disordered Protein's Hydrophobic Domain



This chapter is based on current unpublished work in the Francis Group, which is ongoing under the direction of Kate Dolph and Arya Lall.

4.1 Abstract

Many biomedical and industrial processes use synthetic surfactants, block copolymers, and other amphiphilic polymer systems. However, many of these are produced using non-renewable and non-recyclable materials that pose environmental and toxicological concerns. Because of this, bio-derived and bio-degradable surfactants are promising alternatives. However, most naturally occurring bio-surfactants lack the tailorability afforded by chemically derived polymers. To be implemented industrially, promising bio-surfactants must have properties that mimic currently used surfactants based on charge, molecular weight, and hydrophilic-lipophilic balance (HLB). Previously, we reported a scalable and versatile bio-surfactant production method based on a hydrophilic intrinsically disordered protein (1.0-IDP) sequence with a genetically engineered hydrophobic domain (2Yx2A). These amphiphilic sequences were found to self-assemble into micelles over a broad range of temperatures, pH values, and ionic strengths. While we investigated the role of the 1.0-IDP hydrophilic domain by altering its length and charge, the hydrophobic domain was left

unchanged. To examine the role of the hydrophobic domain on self-assembly properties such as morphology, size, and critical micelle concentration (CMC) values, we developed a bioconjugation method to attach N-terminal cysteine peptides of varying hydrophobicities to the C-terminal tyrosine residue of a 1.0-IDP mutant. The reaction between the C-terminal tyrosine and N-terminal cysteine created a unique linkage akin to the naturally observed biosynthesis of pheomelanin. This represents a facile method for creating a large number of hydrophobic library members to be screened as these IDP-based bio-surfactants.

4.2 Introduction

Diblock copolymers and amphiphilic bio-surfactants are composed of a hydrophilic domain and a hydrophobic domain which our lab has seeked to imitate with an intrinsically disordered protein (1.0-IDP) derived from the human neurofilament heavy arm side chain protein. This sequence, referred to as 1.0-IDP herein, is based on the repetitive sequence SPXEXK (X = G, A, or V). This protein has two oppositely charged residues per repeat at a neutral pH: a negatively charged glutamic acid (E) and a positively charged lysine (K). The genetically engineered hydrophobic domain at the C-terminus of this highly hydrophilic sequene is a 16 amino acid sequence denoted as 2Yx2A. This modification created the amphiphilic protein 1.0-IDP-2Yx2A that assembles into protein micelles at room temperature in a wide range of buffer conditions.[1] The previously described insoluble expression system was highly amenable to substantial mutations to the sequence space and mutants with increased charge and increased length of the hydrophilic 1.0-IDP domain were produced.[2] These mutants provided insight on the role of the hydrophilic domain charge and length on the physicochemical properties of their self-assembly. However, no mutations to the hydrophobic 2Yx2A domain have been attempted. As the hydrophobic domain length, aromaticity, and hydrophobicity could greatly impact properties such as critical micelle concentration (CMC), size, and morphology we sought to screen a library of hydrophobic domains and determine their impact on the self-assembly of these amphiphilic IDP proteins.

To develop a hydrophobic domain library, two approaches were contemplated. One approach was to design a handful of genetically encoded sequences to study the specific variables of length, hydrophibicity, and aromaticity and create these constructs using the insoluble expression system. This approach, while feasible, was quickly deemed to be costly and time consuming as more mutants were added to the library and thus restricted the sequence space that could be screened and characterized. Alternatively, a second approach to develop a bioconjugation strategy

to attach a hydrophobic domain to the C-terminus of the 1.0-IDP sequence has shown promise and allows for the expression and purification of only one protein mutant and and thus favors screening a larger hydrophobic domain library.

In an attempt to find a fast bioconjugation reaction that can provide high conjugation yields of N and C-terminal peptide fragments, we took inspiration from the pheomelanin biosynthetic pathway. During pheomelanin biosynthesis, tyrosine residues are oxidized by the enzyme tyrosinase to form dopaquinone, which then reacts with free cysteine residues forming cysteinyldopas that undergo a reductive cyclization to 1,4-benzothiazines (Figure 4.1).

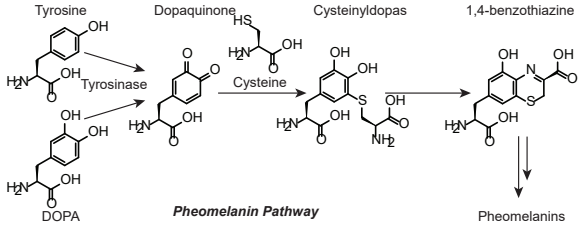


Figure 4.1: Bio-synthetic pathway for the production of pheomelanin from tyrosine. The first biosynthetic step is catalyzed by the tyrosinase enzyme which oxidized tyrosinase to dopaquinone. This dopaquinone is poised for nucleophilic attack by free cysteine residues resulting in cysteinyldopas which cyclize to form 1,4-benzothiazines which react further to produce pheomelanin.

It has been previously shown that flexible linkers with C-terminal tyrosines on proteins could be oxidized with the tyrosinase enzyme and then conjugated to aniline functionalized cargo.[3] Additionally, it has also been shown that free thiol compounds can react rapidly with quinone species produced by the enzymatic oxidation of tyrosine.[4] We hypothesized that a C-terminal tyrosine on the hydrophilic 1.0-IDP domain could be oxidized in a similar manner to the dopaquinone species. Then it could undergo nucleophilc attack by a range of N-terminal cysteine peptides creating a pheomelanin type linkage. As similar oxidative reactions have been proven to be rapid, selective, and high yielding, this could be a viable method for the conjugation of N and C-terminal fragments.[3–5] Successful conjugations of the 1.0-IDP-hydrophobic chimeras would then be tested on a small scale for favorable bio-surfactant properties such as size, CMC, and solublizing ability. Trends in the properties will be analyzed with respect to the hydrophobic domain molecular weight (MW), length, aromaticity, and hydrophobicity allowing for the prediction of future hydrophobic domain sequences with desired properties. Additionally, bioconjugates with favorable properties will then be implemented into an expression scheme and produced on large scale for industrial testing. Additionally, using this method, one could imagine that the incorporation of any molecule containing a free and solvent

accessible thiol, such hydrophobic drugs and hydrocarbons, could be attached to the C-terminus of the IDP domain. This represents a facile and rapid approach to creating a large number of hydrophobic library members for these IDP based bio-surfactants.

4.3 Results and Discussion

Initially, we set out to genetically engineer and express a library of 7 additional variants of the 2Yx2A C-terminal tail. This sequence has been explored and characterized previously.[2] We sought to vary the length, hydropobicity, or aromaticity of the original 16 amino acid sequence systematically and determine how these variables affect the properties of these materials (Table 4.1). All plasmids were successfully cloned for each of the 7 constructs. Due to cost and time restrictions, not all variants were able to be expressed and characterized. However, some insight was drawn for DLS and pyrene measurements taken for the handful of expressed IDPs (Additional Figure 4.6).

Table 4.1: Sequences for the hydrophobic domain library for genetically engineered IDPs.

Name	**Variable**	**Sequence**
2Yx2A	Original	YGAYAQYVYIYAYWYL
2Vx2A	(-) Aromatic	VGAVAQAVAIVAVLAL
2Yx2F	(+) Aromatic	YGFYFQYVYFYFYWYL
2Yx2T	(-) Hydrophobic	YGTYTQYTYGYTYWYG
2Yx2L	(+) Hydrophobic	YILYLQFVFIYIYWFL
1Yx1A	(-) Length	YAQVYAYWYL
4Yx4A	(+) Length	YGAYAQYVYIYAYWYLYGAYAQL YVYIYAYWY
6Yx6A	(++) Length	YGAYAQYVYIYAYWYLYGAYAQ YVYIYAYWYLYGAYAQYVYIYAYWYL

The original IDP sequence, 1.0-IDP-2Yx2A, was measured as a control. It had a diameter of 27.7 ± 5.3 nm indicating consistency with the previously expected spherical micelle assembly (Figure 4.2a). The 1.0-IDP-2Yx2F variant had an increased number of aromatic residues present in its sequence. The 2Yx2F variant showed a lower average diameter of 7.49 ± 0.58 nm suggesting that this variant was not able to assemble into spherical structures like the original sequence (Figure 4.2b). The 1.0-IDP-2Vx2A variant showed a similar average diameter of 23.1 ± 1.8 nm indicating that this variant, which lacks aromatic amino acids, could assemble

into spherical micelles similar to 1.0-IDP-2Yx2A (Figure 4.2c). The variant, 1.0-IDP-2Yx2T, also showed a similar average diameter of 29.5 ± 1.4 nm suggesting this less hydrophobic variant could also exhibit micelle assembly behavior (Figure 4.2d). The truncated 1.0-IDP-1Yx1A variant showed highly polydisperse DLS traces with an average diameter of 163 ± 38 nm. It is unclear if this diameter measurement was accurate due to the large polydispersity value; however, this led us to believe that the 1Yx1A variant did not behave similarly to 1.0-IDP-2Yx2A (Figure 4.2e). The 1.0-IDP-2Yx2L variant has an overall increased hydrophibicity with an equivalent amount of aromatic residues to the original assembly sequence. The average diameter for this variant was determined to be 10.1 ± 0.6 nm, suggesting that this variant also does not exhibit micelle assembly behavior (Figure 4.2f). The expression of all the mutants, including 1.0-IDP-4Yx4A, and 6Yx6A, was time-consuming, and did not result in desired protein yields for easy purification and characterization. Additional experiments can be done on these 7 sequences, and it is promising that insightful information can be gained. However, further rounds of expression are needed for future experiments. From this, we deemed that genetic engineering and expression of all interesting mutants may not be the most efficient process and turned towards an alternative bioconjugation strategy.

In preparation for oxidative coupling with hydrophobic moieties, the hydrophilic 1.0-IDP domain was genetically engineered to have a C-terminal tyrosinase residue. This protein was produced with the insoluble expression system outlined previously.[2,6] After cleavage by BrCN, the resulting 1.0-IDP-Y sequence is easily purified by centrifugation as all the other proteins in the solution, which are still insoluble, precipitate in aqueous conditions while 1.0-IDP-Y is highly hydrophilic and remains in the soluble fraction. This results in the production of a hydrophilic 1.0-IDP-Y domain without the use of any columns, with minimal solvent, and with yields between 100-300 mg (Figure 4.3b).

For bioconjugation of IDP-Y to a hydrophobic peptide, a solution of 20 µM 1.0-IDP-Y protein in phosphate buffer (PB) pH 6.5 was oxidized to the quinone species in 30 min by the addition of 0.2 µmol tyrosinase enzyme. However, due to the abundance of reactive and flexible lysine residues on the IDP backbone, undesired intra and inter-molecular reactions were observed, resulting in aggregation. To combat this, the tyrosinase reaction was alternatively carried out on ice for 3 h. To the tyrosinase reaction mixture, 100 eq of an N-terminal cysteine peptide $CG(WAG)_3$ (1121 Da) was added. The solution was mixed thoroughly and analyzed by ESI-TOF-MS. The reaction between an *ortho*-quinone and free thiol is known to happen rapidly, while the kinetics of the cyclization observed for this reaction are not known.[4] If the reaction is carried out on ice, the un-cyclized thiol addition is observed. In contrast, if the tyrosinase reaction is carried out a 4 °C a mixture of cyclized and un-cyclized

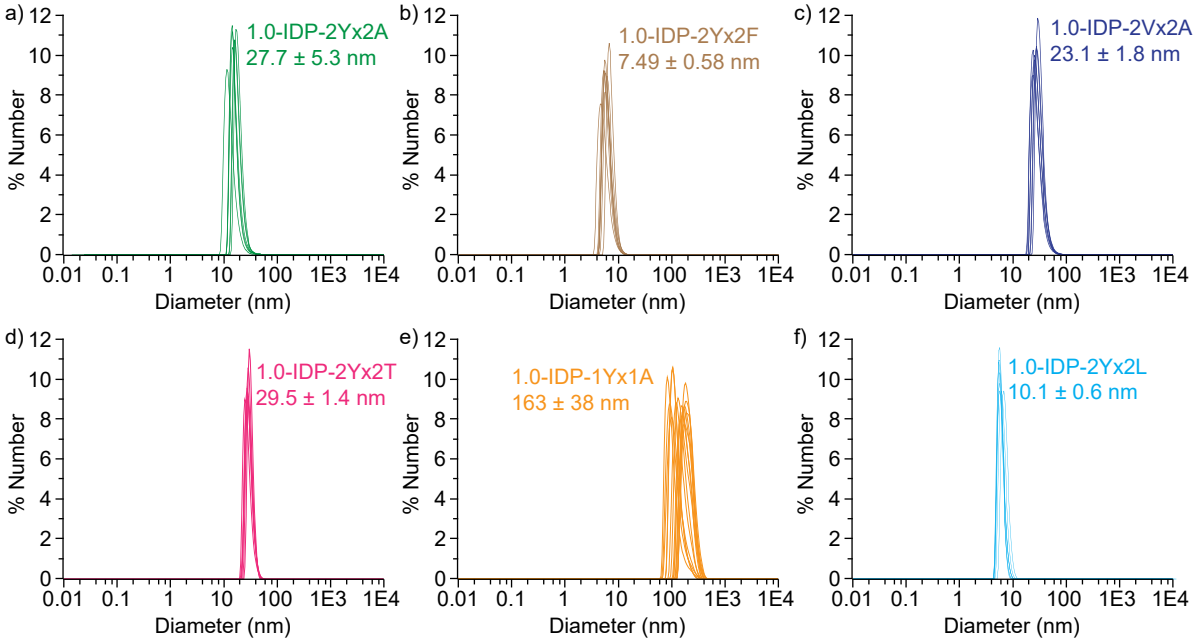


Figure 4.2: DLS analysis and diameter average based on %Number for the hydrophobic domain library. a) 1.0-IDP-2Yx2A = 27.7 ± 5.3 nm b) 1.0-IDP-2Yx2F = 7.49 ± 0.58 nm c) 1.0-IDP-2Vx2A = 23.1 ± 1.8 nm d) 1.0-IDP-2Yx2T = 29.5 ± 1.4 nm e) 1.0-IDP-1Yx1A = 163 ± 38 nm f) 1.0-IDP-2Yx2L 10.1 ± 0.6 nm

products are present (Additional Figure 4.7). To reduce the unwanted double addition, 100 equivalents of N-terminal cysteine peptide were added, and the reaction was monitored by ESI-TOF-MS for completion. Once complete, the reaction is then stopped by dialysis into fresh buffer. Using these reaction conditions, 100% modification of the 1.0-IDP-Y peptide with no double addition was observed by ESI-TOF-MS (Additional Figure 4.8). To determine if other physiologically relevant thiol molecules could add in to this pheomelanin type linkage, a stability study was carried out on the 1.0-IDP-CIIIG conjugate in the presence of 20 mM glutathione. Over the course of 14 h, no glutathione addition nor conversion back to the un-conjugated 1.0-IDP protein was observed. (Additional Figure 4.9).

Investigations into the activity and tolerance of tyrosinase enzymes showed previously that the enzyme found in *Bacillus megaterium* (megaTYR) has much broader substrate tolerance.[7] Due to this finding, we were insterested in seeing if the megaTYR enzyme would be useful for developing hybrid sufactants with IDP-Y and thiol containing hydrophobic small moleules. Many of these molecules have low solubility in water and require the addition of an organic co-solvent to be accessible in an aqueous buffer. Therefore, we were interested in determining the solvent compati-

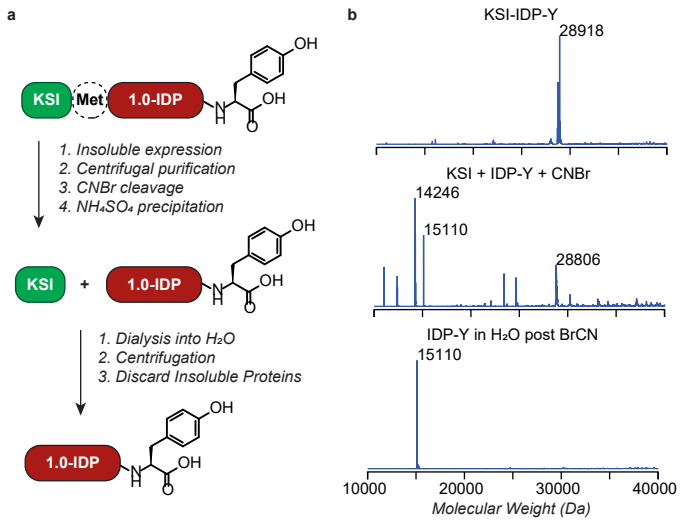


Figure 4.3: Production of IDP-Y with a C-terminal tyrosine. (a) Schematic for the expression and purification of the IDP sequence with a C-terminal tyrosine residue. This method uses no columns and relies on changes in solubility to produce a pure final protein product. (b) ESI-TOF-MS analysis of the purification steps used to produce pure IDP-Y protein (expected: 15110 Da observed: 15110 Da)

bility of megaTYR with three commonly used organic solvents, MeCN, DMF, and DMSO. These reactions were carried out in 20 mM PB pH 6.5 with the addition of 10-50% organic solvent. The oxidation of IDP-Y to the *ortho*-quinone was measured by ESI-TOF-MS after 0.2 nmol of the megaTYR enzyme was added on ice for approximately 40 min (Additional Figure 4.10). The % conversion of IDP-Y to the *ortho*-quinone was graphed to compare the tolerance of the enzyme in each organic solvent (Figure 4.4a). For all solvents, except for MeCN, the ability of the enzyme to oxidize IDP-Y significantly decreased as the solvent added approached 50% (Figure 4.4b.) A test run of 1-dodecanethiol conjugated to IDP-Y was carried out at 100 equivalents in 50% MeCN with the conditions stated previously. A control of glutathione addition to IDP-Y in 0% MeCN was used to confirm enzyme activity (Figure 4.5a). Dodecanethiol was successfully conjugated to IDP-Y; however, both the single (ex: 15326 Da) and double (15526 Da) addition products were seen at appreciable amounts (Figure 4.5b). This result was promising for creating a diverse library of hybrid surfactants using almost any desired molecule with an accessible thiol. Future experiments to look at the properties of this molecule are underway and it will be interesting to determine if these molecules will exhibit micelle or surfactant-like behavior.

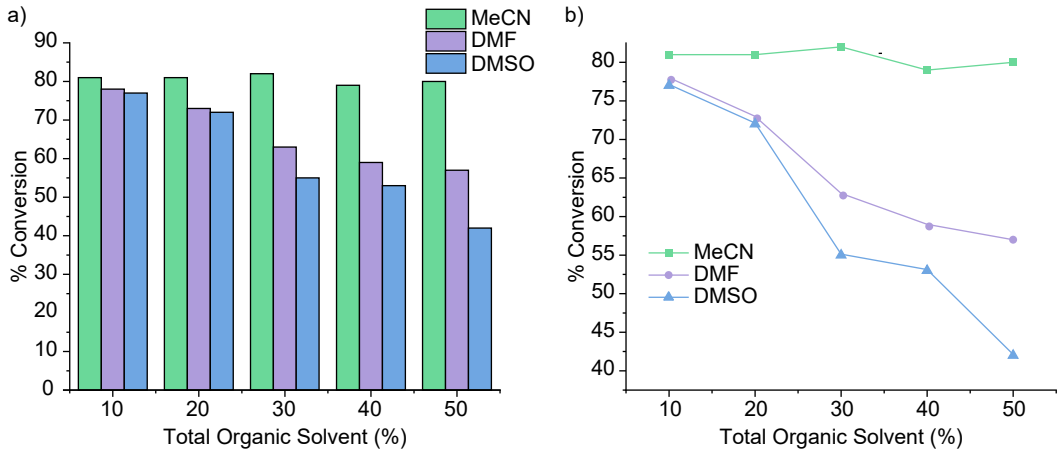


Figure 4.4: Organic solvent screen for megaTYR compatibility. a) The bar graph shows % conversion values of IDP-Y (15110 Da) to oxidized *ortho*-quinone IDP-Y (15124 Da) grouped by total % of organic solvent in the reaction. The organic solvents for MeCN, DMF, and DMSO are shown in green, purple, and blue, respectively. b) The plot shows % conversion trends for the organic solvents MeCN, DMF, and DMSO in green, purple, and blue, respectively.

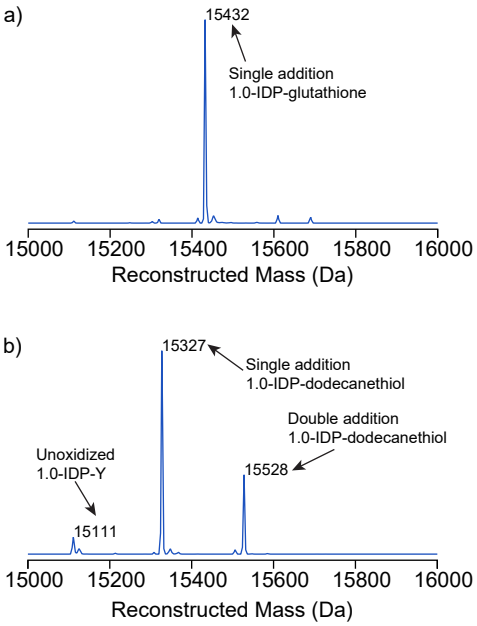


Figure 4.5: Mass Spectra of glutathione and dodecanethiol addition to IDP-Y. a) The spectrum shows quantitative conversion of IDP-Y (ex: 15110 Da) to the single addition product of glutathione (ex:15431 Da). b) The spectrum shows 93% conversion of IDP-Y (ex: 15110 Da) to both the single (70%) and double (30%) additions of dodecanethiol (ex: 15326 Da, 15526 Da). This reaction was carried out in 50% MeCN.

4.4 Conclusion

To study the impact of the hydrophobic domain on the assembly of IDP-based biosurfactants, we aimed to develop a facile bioconjugation method for the creation of a library of amphiphilic peptides based on the 1.0-IDP sequence with varying hydrophobic domains. We explored attempts at genetically engineering and expressing this library; however, efforts determined this strategy was not efficient enough for continued exploration. An alternative bioconjugation method was developed to attach N-terminal cysteine peptides to the C-terminal tyrosine of the 1.0-IDP protein sequence through an oxidative coupling strategy mediated by the tyrosinase enzyme. The bioconjugation reaction between these two fragments results in a pheomelanin type linkage with complete modification of the N-terminal protein fragment. This reaction is carried out in mild conditions (PB pH 6.5) and takes less than 3 h to complete. Additionally, the megaTYR enzyme was able to successfully conjugate molecules initially thought unusable due to their hydrophobic nature. This allowed for the production of a hybrid surfactant with a biologically derived soluble domain and hydrocarbon-based hydrophobic tail. Overall, this represents a rapid and versatile method for the production of amphiphilic IDPs with varying hydrophobic domains, which can be used in preliminary screens to identify sequences with favorable properties.

4.5 Materials and Methods

4.5.1 Materials

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Milli-Q (MQ) H_2O was purified to a resistivity of 18.2 M Ω • cm (at 25 °C) using a MQ Gradients ultrapure water purification system (Millipore).

- 1000x ampicillin solution: 100 mg/mL of solid antibiotic were dissolved in Milli-Q H_2O. The solution was frozen at -20 °C until use.
- 1000x carbenicillin solution: 100 mg/mL of solid antibiotic were dissolved in Milli-Q H_2O. The solution was frozen at -20 °C until use.
- 1000x chloramphenicol solution: 25 mg/mL of slid antibiotic were dissolved in EtOH. The solution was frozen at -20 °C until use.

- LB agar plates: 12.5 g of LB broth (powder) from Thermo Fisher Scientific and 7.5 g agar was dissolved in 500 mL of Milli-Q H_2O and subsequently autoclaved. The solution was cooled to 50 °C and the desired antibiotic was added. The LB agar + antibiotic soultion was poured into sterile Petri dishes (100 mm x 15 mm) and solidified at room temperature. The plates were then stored at 4 °C until use.
- LB media: 20 g of Lauria broth (LB) powder (Sigma Aldrich) was dissolved into 1000 mL of Milli-Q H_2O and autoclaved.
- Lysis buffer: 20 mM Hepes, 300 mM NaCl, and 10 mM β-mercaptoethanol were added to Milli-Q H_2O.
- The pET-31b(+) vector was purchased from Millipore Sigma (catalog number: 69952-3).
- AlwN I (catalog #R0514S) and Xho I (catalog #R0146S) restriction endonucleases were purchased from New England BioLabs (NEB).
- Gibson Assembly Mastermix was purchased from New England BioLabs (NEB).
- All Geneblocks were purchased from Integrated DNA Technologies(IDT) and used as template DNA.

4.5.2 Instrumentation

4.5.2.1 Reverse Phase High Performance Liquid Chromatography (RP-HPLC)

RP-HPLC purification of IDP-2Yx2A protein mutants was performed on an Agilent 1260 Infinity series system equipped with a UV-Vis detector and a fraction collector, controlled by Open Lab CDS software (Agilent Technologies Inc. 2011 –2015). HPLC grade Milli-Q H_2O + 0.1% TFA and HPLC grade MeCN + 0.1% TFA were used as mobile phases A and B, respectively. Preparatory scale purification was performed on a Waters XBridge Protein BEH C4 OBD Prep Column (300 Å, 5 µm, 19 mm X 50 mm) with a flow rate of 5 mL/min over a 5-95% gradient over the course of 30 min. Fractions containing IDP-2Yx2A mutants were identified by their coinciding 280 nm absorbance and 350 nm fluorescence profiles.

4.5.2.2 Mass Spectrometry of Proteins (ESI-TOF-MS)

Time of flight electrospray ionization mass spectrometry (ESI-TOF-MS) of protein samples was performed using an Agilent 1260 series liquid chromatography (LC) system coupled with an Agilent 6224 TOF LC-MS (Santa Clara, CA). Protein samples were first passed through a 0.22 µM cellulose acetate filter and then placed in HPLC vials for analysis. Depending on sample concentration, 0.5 -10 µL was injected on to a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) analytical column with a flow rate of 0.30 mL/min and a gradient of 5-100% MeCN+0.1% (v/v) formic acid in Milli-Q H_2O+0.1% formic acid over the course of 8 min. The total ion chromatograms were extracted in the Agilent Mass Hunter Workstation software (Qualitative Analysis Version B.05.00, Build 5.0.519.13, Agilent Technologies Inc. 2011). Protein molecular weights were reconstructed using a maximum entropy deconvolution. The resulting spectra were exported as CSV files and plotted using the Chartograph web-tool(www.chartograph.com).

4.5.2.3 UV-VIS DNA and Protein Measurements

Estimates of DNA and protein concentration were calculated by measuring the sample absorbance at 260 nm and 280 nm, respectively, on a NanoDrop 1000 (Thermo). Using Beers law and the calculated extinction coefficients, the concentrations of both DNA and protein were approximated.

4.5.2.4 Dynamic Light Scattering (DLS)

Dynamic light scattering was performed on a Malvern Instruments Zetasizer Nano ZS. Prior to analysis, lyophilized samples were re-hydrated in the appropriate buffer and passed through a 0.22 µm cellulose acetate centrifugal filter. Protein concentration was determined by diluting the sample 1:10 in 50% MeCN in H_2O and measuring the absorbance at 280 nm on a NanoDrop 1000 (Thermo). The filtered samples were then diluted with buffer to the desired concentration and 80-100 µL was then added to the DLS cuvette. For room temperature analysis, (25 °C) was used. Data plots and standard deviations were calculated from an average of the %Number mean for each run. Measurement data was presented as averaged mean diameters determined by the %Number distribution. The %Number was used in this analysis because it shows the size of the majority population in a sample. Heterogeneity was evident in the micelle samples as evidenced by the polydispersity and %Intensity traces. It is likely that larger structures or inter-micelle interactions occur and form a small number of larger particles that contribute to larger diameters when analyzed by

Table 4.2: Standard Gibson Assembly Protocol

Gibson component	insert:vector (5:1)
cut pET-31b(+)	50 ng
Geneblock insert (500 bp)	22 ng
Gibson Assembly Mastermix	10 µL
Milli-Q H_2O	to 10 µL

%Intensity, which heavily weights large particles in a sample as opposed to the most abundant.

4.5.3 Methods for protein production

4.5.3.1 IDP plasmid construction

The following constructs were created using a Gibson Assembly protocol with the pET 31b(+) entry vector and individual geneblocks designed with 20-30 overlapping basepairs with the template DNA.[8] The pET31b(+) entry vector was digested with 10 units of KhoI and AlwnI restriction enzymes for 1 h at 37 °C. Following a 20 min deactivation step at 80 °C was performed. The cut entry vector were run on a 1% agaraose gel prestained with SYBR Safe (ThermoFisher). The gel was run at 120 V for 35 min in 1X TAE buffer then imaged under UV fluorescnece with a Biorad GelDoc Ez Imager. The DNA band corresponding to the cut entry vector was excised from the gel and plaved in 1.5 mL Eppendorf tube. The digested vector was removed from the gel using the quick gel extraction kit (Promega).

The Gibson Assembly was compiled using the components listed in Table 4.2 and the reaction was incubated at 55 °C for 30 min and then kept at 4 °until further processing. Half of the Gibson Assembly (10 µL) was transoformed into 100 µL of chemically competent XL1-Blue cells. The reaction was incubated with the cells on ice for 30 min and then heat shocked at 42 °C for 45 s. The cells were recovered by adding 900 mL of SOC media and incubation at 37 °C, 220 rpm for 1 h. Following, 150 µL of the transformed cells were plated on carbenicillin (+) LB agar plates and grown overnight at 37 °C. For each plate 2-4 colonies were selected placed in culture tubes of 5 mL of LB media and 5 µL of 1000x carbenicillin. Cultures were grown overnight at 37 °C. Following this, the plasmid DNA was extracted using a Miniprep Kit (Millipore) and sent for sequences.

Geneblock for the construction of the KSI-1.0-IDP-2Yx2T pET-31b(+) plasmid

5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG
AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC
CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCTATGGCACTTATACACAGTATACATATGGTTATACTT
ATTGGTATGGATGACTCGAGCACCACCACCACCACC 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-2Yx2T:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYVYIYAYWYL*
KSI protein sequence shown in green
2Yx2T protein sequence shown in magenta

Geneblock for the construction of the KSI-1.0-IDP-1Yx1A pET-31b(+) plasmid
5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG
AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC
CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCTATGGCGCGTATGCGCAGTATTGGCTGTGACTCGAG
CACCACCACCACCACCACTGA 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-1Yx1A:

MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYWL*
KSI protein sequence shown in green
1Yx1A protein sequence shown in magenta

Geneblock for the construction of the KSI-1.0-IDP-2Vx2A pET-31b(+) plasmid
5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG
AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC
CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCGTAGGCGCGGTTGCGCAGGCTGTAGCTATTGTTGCG
GTTCTTGCTCTTTGACTCGAGCACCACCACCACCACCA 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-2Vx2A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAVGAVAQAVAIVAVLAL*
KSI protein sequence shown in green
2Vx2A protein sequence shown in magenta

Geneblock for the construction of the KSI-1.0-IDP-2Yx2F pET-31b(+) plasmid
5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG

AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC
CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCTATGGCTTTTATTTTCAGTATGTGTACTTCTACTTTT
ATTGGTATCTGTGACTCGAGCATCACCACCACCACC 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-2Yx2F:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGFYFQYVYFYFYWYL*
KSI protein sequence shown in green
2Yx2F protein sequence shown in magenta

Geneblock for the construction of the KSI-1.0-IDP-2Yx2L pET-31b(+) plasmid
5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG
AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC
CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCTATATTCTCTATCTACAGTTTGTGTTCATTTATATCT
ACTGGTTCCTGTGACTCGAGCACCACCACCACCACC 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-2Yx2L:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV

KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYILYLQFVFIYIYWFL*
KSI protein sequence shown in green
2Yx2L protein sequence shown in magenta

Geneblock for the construction of the KSI-1.0-IDP-4Yx4A pET-31b(+) plasmid
5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG
AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC
CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCTATGGCGCGTATGCGCAGTATGTGTATATTTATGCGT
ATTGGTATCTGTACGGTGCTTACGCACAATACGTTTATATATACGCTTA
CTGGTACCTTTGACTCGAGCACCACCACCACCACCACTGAGATCCG 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-4Yx4A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYVYIYAYWYLYGAYAQYVYIY
AYWYL*
KSI protein sequence shown in green
4Yx4A protein sequence shown in magenta

Geneblock for the construction of the KSI-1.0-IDP-6Yx6A pET-31b(+) plasmid
5' CGAGAAGAATATTCACGCATGCCAGATGCTGGCAGAGGCAAAGAGTC
CTGCGGAAGCTAAGTCGCCAGCCGAAGTCAAGTCGCCCGCCGTCGCGA
AAAGCCCCGCAGAGGTAAAAAGCCCGGCTGAAGTTAAGAGCCCGGCAG
AAGCAAAAAGTCCAGCAGAAGCCAAAAGTCCTGCTGAGGTTAAAAGCC

CAGCAACCGTCAAATCCCCTGGAGAGGCAAAATCTCCGGCAGAAGCCAA
GTCCCCTGCCGAAGTGAAGTCACCTGTCGAAGCCAAGTCGCCGGCCGAA
GCGAAGAGCCCAGCGAGCGTGAAAAGTCCTGGTGAGGCTAAGTCCCCG
GCGGAAGCGAAATCTCCAGCGGAAGTAAAGAGTCCGGCCACCGTAAAA
TCTCCGGTTGAGGCCAAGAGTCCGGCGGAAGTTAAATCTCCGGTGACG
GTCAAGTCCCCAGCGGAAGCGAAGTCCCCCGTGGAGGTGAAATCACCA
TACTGGAGTGCCTATGGCGCGTATGCgcAGTATGTGTATATTTATGCGT
ATTGGTATCTGTACGGTGCTTACGCACAATACGTTTATATATACGCTTA
CTGGTACCTTTATGGAGCTTACGCCCAGTACGTGTATATCTATGCTTAC
TGGTATCTATGACTCGAGCACCACCACCACCACCACTGAGA 3'

Corresponding amino acid sequence of fusion protein KSI-1.0-IDP-6Yx6A:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSPYWSAYGAYAQYVYIYAYWYLYGAYAQYVYIY
AYWYLYGAYAQYVYIYAYWYL*
KSI protein sequence shown in green
6Yx6A protein sequence shown in magenta

The KSI-1.0-IDP-Y plasmid was created by performing a site directed mutagenesis to encode a stop codon after the 1.0-IDP sequence resulting in a C-terminal tyrosine residue. The KSI-1.0-IDP-2Yx2A plasmid was used as a genetic template.[2] Aglient's QuikChange mutagenesis kit and protocols were used.

Primers for construction of KSI-1.0-IDP-Y pET-31b(+) plasmid:
Primer 1: 5' CGCGCCATAGGCACTCTAGTATCCTGATTTCACCTCCACGG 3'
Primer 2: 5' CGCGCCATAGGCACTCTAGTATCCTGATTTCACCTCCACGG 3'
Template: KSI-1.0-IDP-2Yx2A pET-31b(+) plasmid

Corresponding amino acid sequence of KSI-IDP-Y:
MHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAI
REFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFR
FNGAGKVVSIRALFGEKNIHACQMLAEAKSPAEAKSPAEVKSPAVAKSPAEV
KSPAEVKSPAEAKSPAEAKSPAEVKSPATVKSPGEAKSPAEAKSPAEVKSPV
EAKSPAEAKSPASVKSPGEAKSPAEAKSPAEVKSPATVKSPVEAKSPAEVKS
PVTVKSPAEAKSPVEVKSGY*

KSI protein sequence shown in cyan

4.5.3.2 Protein expression of IDP variats

All IDP variants were expressed using the following protocol. Plasmids containing the desired fusion proteins were transformed into Rosetta2 (pLysS) cells and plated onto carbenicillin(+)/chloramphenicol(+) LB Agar plates and grown overnight at 37 °C. Single colonies were selected and placed into 50 mL of LB with 50 µL of both 1000x stocks of carbenicillin and chloramphenicol antibiotics. The cultures were grown overnight at 37 °C. The following morning, 30 mL of the 50 mL cultures was used to innoculate 1000 mL of LB with 1000 µL 1000x stocks of carbenicillin and chloramphenicol in a 2.8 L baffled flask. Cultures were grown at 37 °C with shaking at 220 rpm until an OD_{600} of 0.6 was reached, at which point the temperature was lowered to 18 °C. After cooling for 20 min, IPTG was added to a final concentration of 0.25 mM. The proteins were expressed overnight for 18-24 h. The cultures were then centrifuged at 4000 xg for 20 min. The resulting protein pellets were transferred to 50 mL Falcon tubes and frozen at -20 °C. A typical expression yields approximately 5-10 g cell pellets.

4.5.3.3 Protein purification of IDP variants

The frozen cell pellets were resuspended in 30 mL of lysis buffer, 30 µL of 1% Trition-X, and 300 µL of a 0.1 M PMSF solution (Sigma Aldrich). The cell pellets were vortexed until completely resuspended and then sonicated at 70% amplitude for 10 min (2 s on: 4 s off). The lysed cells were then centrifuged at 14,500 xg for 20 min. The supernatant containing the soluble proteins was discarded and the insoluble fraction was washed by resuspending it in 30 mL of lysis buffer and centrifuging again at 14,500 xg for 20 min. The washing step was repeated three times. After the third wash, 5 mL of 6 M guanidinium HCl (Guan HCl) was added to the insoluble fraction and the proteins were resuspended by vigorous shaking for 1 h at room temperature and sonication at 70% amplitude for 10 min (2 s on: 4 s off). The protein solutions were then centrifuged at 14,500 xg for 20 min, at which point the fusion proteins were solubilized and the pelleted remains were largely composed of cell debris. The supernatants containing the fusion proteins were transferred to a sterile 50 mL Falcon tube. The approximate concentration was determined by measuring the UV absorbance at 280 nm on a NanoDrop 1000 (Thermo) and the calculated extinction coefficient (ExPASy Protein Parameters Tool https://web.expasy.org/protparam/) while the protein purity was assessed by SDS PAGE gel analysis.

The KSI protein was cleaved from the IDP sequence by the slow addition of 3 M HCl, concentrated formic acid, and cyanogen bromide(BrCN) to a final concentrations of 0.5 M HCl, 10% formic acid, and 100 molar equivalents of BrCN to protein. The solution was stirred in the dark under a steady flow of nitrogen gas. After 24 h, the reaction was mostly complete when analyzed by ESI-TOF-MS. The proteins were then precipitated from solution by the addition of cold saturated $(NH_4)_2SO_4$ solution. The precipitation solutions were kept at -20 °C for 1 h. The precipitated proteins were separated from the soluble components (BrCN, Guan HCl, and salts) by centrifugation at 14,500 xg for 20 min. The supernatant was discarded in a closed top waste receptacle specified for BrCN. The protein pellets were resuspended in H_2O by sonication at 70% amplitude for 10 min (2 s on: 4 s off) followed by a subsequent centrifugation at 14,500 xg for 20 min to remove any insoluble protein. The resulting solution of IDP was pure when analyzed by ESI-TOF-MS and SDS PAGE gel.

4.6 Additional Figures

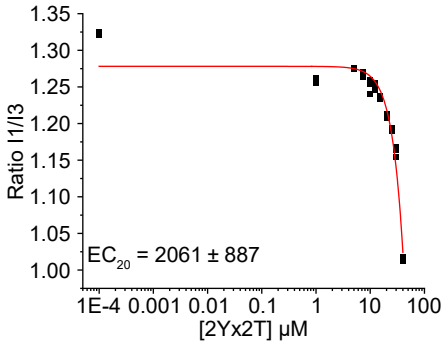


Figure 4.6: Pyrene Fluorescence Assay for the determination of 1.0-IDP-2Yx2T CMC. The I1/I3 ratios are plotted against concentration. A logistic function was fit to the curve and an EC_{20} value of 2061 ± 887 µm. Additional data points at higher concentrations are needed to generate a reliable logistic curve fitting value.

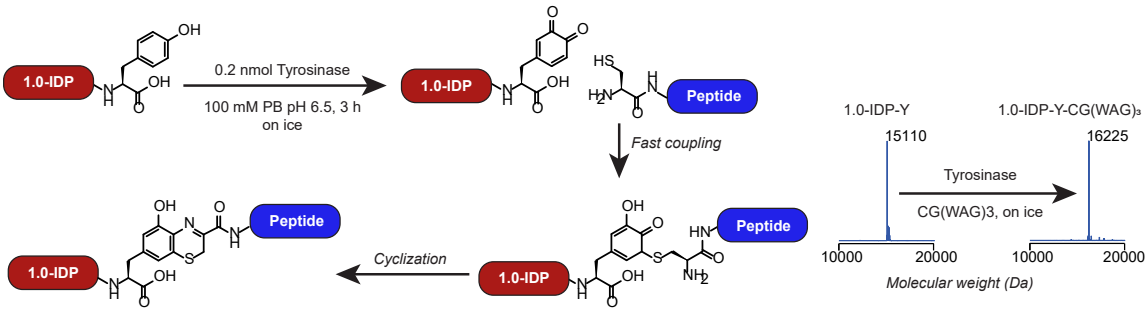


Figure 4.8: C-terminal tyrosine residue with N-terminal cysteine peptides. The C-terminal tyrosine of 1.0-IDP-Y is oxidized to the *ortho*-quinone IDP-Y and 100 eqs of a N-terminal cysteine peptide are added to the reaction and mixed on ice. Once the IDP-Y is oxidized to the quinone, it rapidly reacts with the thiol group of the cysteine residue. This is followed by reoxidation and subsequently a condensation reaction with the N-terminal amine of the peptide and the oxygen of the semi quinone, resulting in a cyclic product.

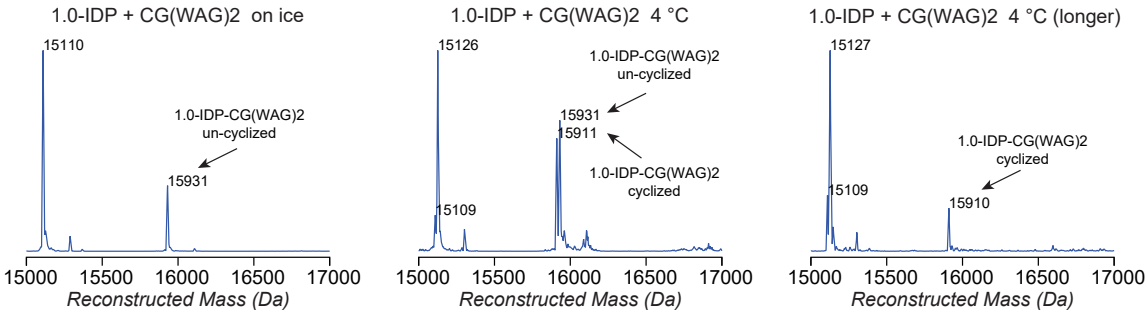


Figure 4.7: Analysis of fast and slow steps of phenomelanin type linkage formation. Conducting the reaction between 1.0-IDP and the peptide $CG(WAG)_2$ (MW = 806.7 Da) on ice resulted in only the nucelophilic addition of the thiol to the aromatic ring (expected: 15931 , observed: 15931 Da). At 4 °C both the condensation reaction product, forming the pheomelanin type cyclic linkage, and the thiol only addition were observed (15911 Da and 15931 Da). If the reaction was conducted for extended periods of time (24 h) or at elevated temperature, complete conversion to the cyclized product was observed (expected: 15911 Da, observed: 15910 Da).

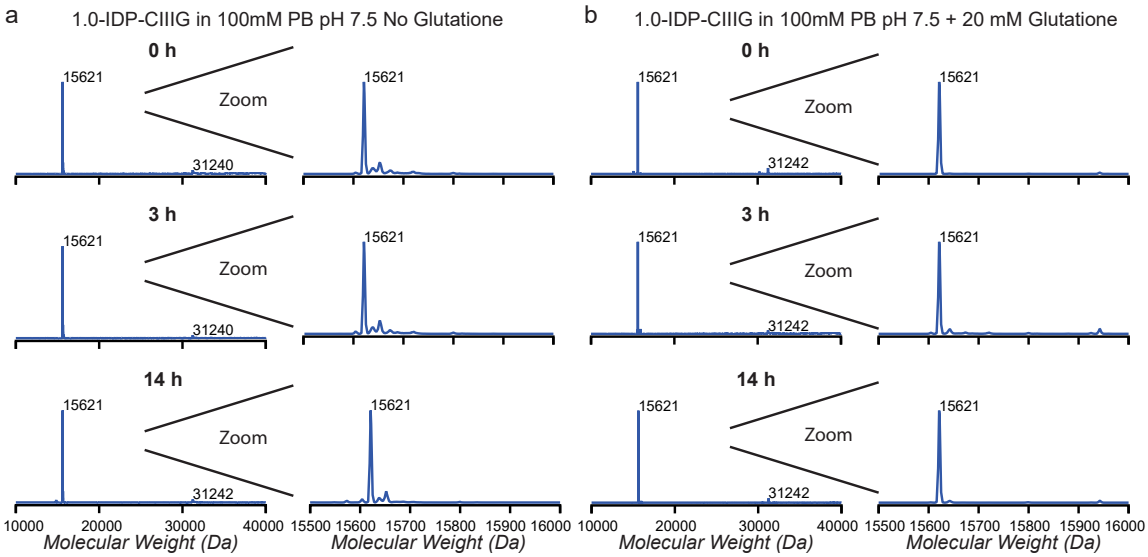


Figure 4.9: Stability of the 1.0-IDP-CIIIG conjugate in the presence of glutathione. (a) Aliquots of the 1.0-IDP-CIIIG (MW = 15621 Da) conjugate were incubated in 100 mM PB pH 7.5 for 14 h and then analyzed by ESI-TOF-MS. Over the course of 14 h the only observable peaks are those corresponding to 1.0-IDP-CIIIG at 15621 Da and a small amount of dimer at 31242 Da, which appears to be either an artifact of deconvolution or a non-covalent association between monomers. (b) The same 1.0-IDP-CIIIG sample was also incubated in 100 mM PB pH 7.5, plus 20 mM glutathione (MW = 307.3 Da) for 14 h. The sample was analyzed by ESI-TOF-MS and, as in the case without glutathione, only peaks corresponding to the conjugate and its dimer were observed with no discernible difference between the spectra corresponding to the sample without glutathione.

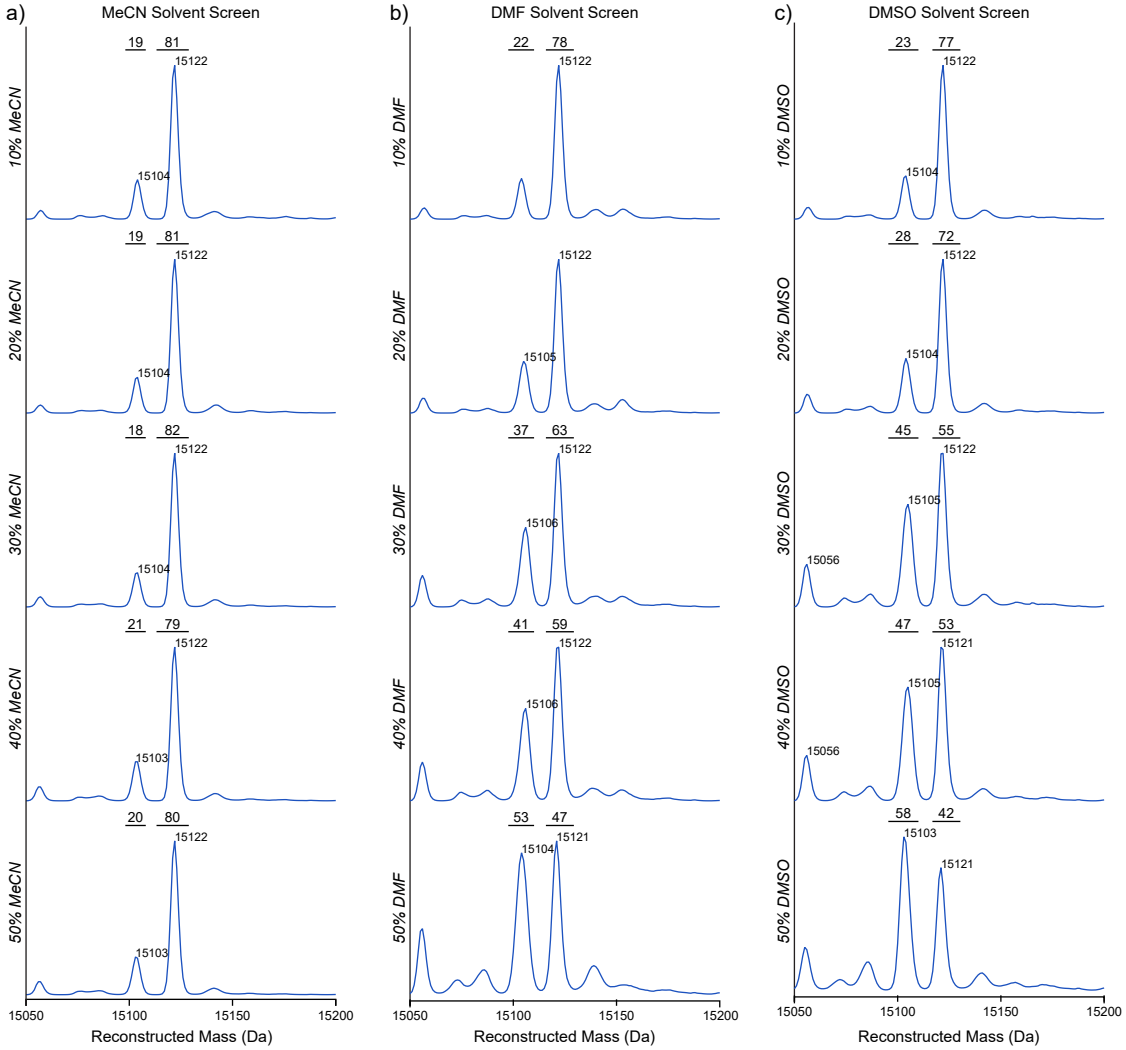


Figure 4.10: Mass spectra for the organic solvent screen of IDP-Y oxidation with megaTYR. Mass sectra showing integrated conversion values of IDP-Y (ex: 15110 Da) to the oxidized *ortho*-quinone IDP-Y (ex: 15124 Da). If their is no external nucleophile, the *ortho*-quinone IDP-Y will undergo an intramolecular reaction and produce a species with a molecular weight of 1522 Da. a) MeCN screen from 10-50% of total organic solvent. The oxidation levels stay constant around 80% for all samples. b) DMF screen from 10-50% of total organic solvent. The oxidation conversion starts at 78% but continues to drop with a final value of 47% oxidation at 50% DMF. c) DMSO screen from 10-50% of total organic solvent. The oxidation conversion is 77% at 10% DMSO; however, the oxidation drops as the DMSO is increased, with 50% DMSO only showing 42% oxidation.

4.7 Acknowledgements

This work was supported by the BASF-CARA program. Additionally, portions of this work were conducted and written by Sarah Klass and Arya Lall.

4.8 References

(1) Klass, S. H.; Smith, M. J.; Fiala, T. A.; Lee, J. P.; Omole, A. O.; Han, B.-G.; Downing, K. H.; Kumar, S.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 4291–4299.

(2) Klass, S. H.; Gleason, J. M.; Omole, A. O.; Onoa, B.; Bustamante, C. J.; Francis, M. B. *Biomacromolecules* **2022**, *23*, 1462–1470.

(3) Marmelstein, A. M.; Lobba, M. J.; Mogilevsky, C. S.; Maza, J. C.; Brauer, D. D.; Francis, M. B. *Journal of the American Chemical Society* **2020**, *142*, 5078–5086.

(4) Lobba, M. J.; Fellmann, C.; Marmelstein, A. M.; Maza, J. C.; Kissman, E. N.; Robinson, S. A.; Staahl, B. T.; Urnes, C.; Lew, R. J.; Mogilevsky, C. S.; Doudna, J. A.; Francis, M. B. *ACS Central Science* **2020**, *6*, 1564–1571.

(5) Maza, J. C.; Bader, D. L. V.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Krska, S. W.; Parish, C. A.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 3885–3892.

(6) Kuliopulos, A.; Walsh, C. T. *Journal of the American Chemical Society* **1994**, *116*, 4599–4607.

(7) Mogilevsky, C. S.; Lobba, M. J.; Brauer, D. D.; Marmelstein, A. M.; Maza, J. C.; Gleason, J. M.; Doudna, J. A.; Francis, M. B. *Journal of the American Chemical Society* **2021**, *143*, PMID: 34382787, 13538–13547.

(8) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. *Nature Methods* **2009**, *6*, 343–345.