

NORTHWESTERN UNIVERSITY

Designing Dynamic and Modular Biomolecules and Assays to Interrogate and Control Protein  
Fate

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biomedical Engineering

By

Daniel J. Sykora

EVANSTON, ILLINOIS

December 2022

© Copyright by Daniel J. Sykora 2022  
All Rights Reserved

## Abstract

Cancer has long been the second-leading cause of death in the United States and represents the leading cause of death in midlife (age 40-60). While the prognosis for many cancers has vastly improved over the last thirty years, many cancers remain elusive due to the late-onset of symptoms, the specific organ systems they affect, the primary sites of metastasis, and, of course, the type of tumor (e.g. solid v. blood) and the subsequent oft-immunosuppressive tumor microenvironment. Biological therapeutics (i.e. biologics) have revolutionized the way we treat cancer due to their inherent ability to successfully target overexpressed antigens – often, proteins expressed on the surface of cancer cells – while minimally affecting healthy cells. The most common biologic is the immunoglobulin G (IgG) monoclonal antibody (mAb), a Y-shaped protein secreted by plasma B cells of the adaptive immune system. However, there is an inherent inability to easily optimize the structure of an IgG for maximal efficacy, and this lack of programmability can contribute to issues biologics often face such as low tumor penetration, nonspecific immunogenic responses, rapid clearance, and high dosage requirements. To modulate the structure-function of biologics to improve cancer treatment, mitigating the dosage of non-discriminatory traditional chemotherapy in the process, our lab has developed a protein assembly platform technology known as ‘megamolecules’ (**Chapter 1**) which uses rapid, specific, and irreversible enzyme-inhibitor reaction chemistries to covalently bring fusion proteins together. The megamolecule approach provides atomic-level precision over the synthesis of protein scaffolds, and these scaffolds can modulate inherent properties of biologics such as binding specificities, affinities, orientations, and stoichiometries with relative ease.

In **Chapter 2**, this next-generation, modular assembly strategy was utilized to develop a library of therapeutics towards breast cancer research, building off our lab’s initial demonstration

of synthesizing, characterizing, and utilizing megamolecules to create mimics of the mAb trastuzumab. While trastuzumab – often in combination with the mAb pertuzumab – has shown moderate success in the clinic for HER2+ breast cancer patients, immune tolerance typically results, leading to a transiently efficacious drug. Thus, there is sufficient room to improve upon this well-researched mAb. I used megamolecules to investigate how HER2-targeting scaffolds can be modulated to interrogate biologic properties such as binding affinity, avidity, net internalization rate of the megamolecule-receptor complex, and downstream inhibition of cell proliferation. Increasing the binding valency of our megamolecule scaffolds from 2 to 3 only modestly improved binding affinity and had no effect on increasing megamolecule-HER2 endocytic rate nor the inhibition of cell proliferation. Creating bispecific (biparatopic) scaffolds that targeted two different epitopes on HER2 was the only way to significantly increase net internalization rate by cross-linking domains I and IV on the HER2 extracellular domain. Interestingly, scaffolds that *only* presented the trastuzumab Fab domains were the only candidates that showed significant inhibition of proliferation. Here, even adding an extra nanobody towards domain I within scaffolds that had two trastuzumab Fabs completely abrogated the inhibition of cell proliferation seen with scaffolds that had two trastuzumab Fabs alone.

Next, **Chapter 3** explores the utility of the megamolecule platform as a proof-of-concept reversible protein switch. Here, we utilized synthetic chemistry to build terpyridine-terminated small molecules that irreversibly reacted with one of our megamolecule enzymes, cutinase. Once incorporated into a megamolecule scaffold, two terpyridine groups could reversibly coordinate upon addition of bivalent transition metals (e.g.  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Zn}^{2+}$ ). Strategically positioning each terpyridine group at opposing ends of a linear megamolecule scaffold allowed for quaternary-scale domain cyclization, which could be quantitatively discerned through Förster Resonance Energy

Transfer (FRET). Ultimately, I demonstrated that terpyridine coordination – and therefore, FRET signal – was dependent on addition of specific divalent transition metals, which could be reversibly sequestered by addition of excess ethylenediaminetetraacetic acid (EDTA). The specific FRET response was unique to the length of each sensor as well as the individual metal ion; the data strongly correlated with long-standing literature of terpyridine-metal and EDTA-metal coordination kinetics. Longer scaffolds had faster coordination kinetics (i.e.  $k_{on}$ ) towards the bidentate complex, which, again, were unique to each individual metal. Coarse-grain modeling and small-angle X-ray scattering (SAXS) showed good agreement with experimental results, suggesting that the megamolecule platform's flexibility for synthesis of various protein scaffolds could be utilized within a protein sensor framework.

While the treatments for cancer are varied and complex, appropriate diagnosis and favorable prognoses rely on early and accurate detection. In **Chapter 4**, I utilized surface chemistry techniques to pattern single cells into specific shapes that, when stained for their actin cytoskeleton, could discriminate between cancer and non-cancer cells with a feature-extraction machine learning algorithm. High-resolution (60X) confocal microscopy imaging against the actin cytoskeleton without any patterning was sufficient to discriminate between two cell populations in the instances where phenotypes were quite distinct, which ran against our initial hypothesis of always requiring shape normalization *a priori*. In fact, patterning cells into shapes for algorithmic discrimination was only effective when cell lines had similar, overlapping phenotypes. This work demonstrates a compelling proof-of-concept incorporation of high-resolution confocal microscopy into quantitative machine learning workflows.

In **Chapter 5**, I present a co-author project from earlier in my PhD, which provided a necessary breadth to my five years at Northwestern. This published work interrogated phosphatase

activity and specificity from various cancer and non-cancer cell lysate utilizing our platform technology known as SAMDI. Here, high-throughput, modular peptide arrays were treated with cell lysate, and we were able to demonstrate that phosphatase activity and specificity were conserved across cell lines, cancer states, and species. Furthermore, phosphatases in the lysate were universally more active towards phosphorylated threonine than serine on our peptide arrays, which may contribute to the reported differences in phosphorylation seen across the phosphoproteome. This work is important because most research in the field focuses on activity and specificity of kinases.

In **Chapter 6**, I shortly reflect on my PhD, the major conclusions of my work, and discuss potential research projects for future students.

## Acknowledgements

Completing a PhD truly takes a village, and I would be remiss to not acknowledge every person who made all this possible. Unlike some lifelong students, I did not map out my life on a five or ten-year plan in my early 20s with a neatly-defined area of interest that I wanted to pursue a doctorate in. Rather, I found that I enjoyed the pace of research as a stark contrast to the hectic life that a class-based curriculum provided undergraduates. I also found that I enjoyed the process of thinking “like a scientist” – i.e. objective, analytical, creative, and with an eye for both the forest and for the trees. Thankfully, my scientific interests did naturally mature over time, and I hope some of that learned passion can come through in this work. As a result, I was never sure if a PhD was in the cards or if it was something I would (or should) even pursue. So, the purpose of this section is to properly thank everyone who helped my precarious journey towards a doctorate be less arduous, more rewarding, and more fun.

First and foremost, I need to thank my advisor **Dr. Milan Mrksich**, who has been nothing but supportive during my five years at Northwestern. Milan, you gave me the confidence and reassurance that I was making progress on projects, even if it was unclear to me, and you provided superior freedom and funding to try new experimental protocols, tools, and facilities to expand and understand the capabilities of our research. Due to the support you’ve shown us all, I’ve never been more confident about the prospect of megamolecules, and I’m excited to see how ModuMab develops! Additionally, thank you for letting me partake in the INVO Practicum during my 4<sup>th</sup> year; participation in this internship ultimately played a critical role in elucidating my career interests post-grad. Working with the INVO team – **Dr. Sonia Kim, Dr. Lisa Dhar, John Haugen, JD, Dr. Iwona Maciagiewicz, Dr. Wenting Luo, Dr. George Chellapa, and Dr. Alexandra de Paz** – was one of my highlights of graduate school, and I learned so much about

the intersection of law, venture capital, and scientific development. It would have been far more difficult to break into that space without your support.

I want to thank my other dissertation committee members, **Dr. Kathy Green**, **Dr. Evan Scott**, and **Dr. Neha Kamat**, for lending their scientific expertise and support across many disparate projects! Finally, I also want to thank **Dr. Wendy Murray** for being my first connection and resource at Northwestern. It was our first conversation in your office regarding the Biomedical Engineering program, department, and my background that motivated me to formally apply. I'm forever fortunate that Cairo and Clarence crossed paths.

I also want to thank members from previous institutions who have been instrumental in my journey. While an undergraduate at Duke University, the bulk of my research interests were inspired by a class in cellular mechanotransduction I took with my academic advisor, **Dr. Brent Hoffman**. Here, I was taught to view much of cell biology as physical phenomena, rather than just a bunch of abstract chemical reactions, and this meshed with my predilection towards visual learning. Brent, you were the first professor I felt like I could just *be myself* around and talk to. Your high expectations you had for your students was matched by the effort you put into your course, and I wish I had more professors at Duke like you.

In parallel to these budding academic interests, I had the opportunity to conduct cell mechanobiology research in the laboratory of **Dr. Farshid Guilak** with postdoctoral mentor **Dr. Christopher Gilchrist**. Here, I helped Chris understand how varying the spatial architecture of micropatterned 2D substrates (here, stacked 1D lines) influenced mesenchymal stem cell (MSCs) migration speed, persistence, and ultimately, the degree of aligned fibrillar collagen I deposition. This applied tissue engineering work awarded me a Pratt School of Engineering Fellowship where I was able to conduct this research full-time over my junior summer. This happened to coincide

with my fondest memories of Duke (shoutout to the Monmouth Ave. apartment). Getting my hands wet with conducting research full-time and understanding the consistent effort and dedication it takes to even get one piece of useful data helped shape my perspective on being a successful scientist. Further, this experience used methodologies developed by **Dr. Andrew Doyle** at the National Institutes of Health; one serendipitous phone call with him my senior spring set up my next two years as a postbaccalaureate researcher in his research group.

I moved to Washington, DC to pursue this two-year postbac in **Dr. Kenneth Yamada's** research group in the National Institute of Dental and Craniofacial Research (NIDCR) with Andrew as my primary mentor. These two years were easily the most pivotal towards my PhD journey, and I don't think I can say enough about my colleagues here. I was the youngest person in the group by a decade, but through the effort and consistency I learned to put into my research, I felt like they began to treat me as an equal colleague despite my youth and inexperience. Andrew was always willing to help or spitball about science, despite being pulled into fifteen different directions every day. He encouraged my creativity and allowed me to explore different ways of acquiring, analyzing, and quantifying data. Most importantly, he helped me forge my skills as a microscopist, which has been a skillset that has obviously proven invaluable to the bulk of my dissertation research. Ultimately, our work on understanding the mechanobiology behind 3D mesenchymal cell migration was published in *Developmental Cell* during Summer 2020, and it still remains, to this day, the work I am most proud of due to the sheer amount of growth it personally represents for me. Other members of the Yamada group – **Kaz Matsumoto, Dr. Josh Collins, Dr. Brian DuChes, Dr. Shaohe Wang**, and, of course, Ken himself, were always willing to have conversations with me about anything – science or not – and helped encourage me along the way. Josh, I wish we could catch up on the opposing – and perhaps, canonical – trajectories

the Cubs and Cardinals seem to be heading, and I always appreciated the friendly banter and your genuine happiness for me when the Cubs (finally!) won the 2016 World Series. Think the Cards are a true contender this year the way Arenado, Goldschmidt, and Edman are playing.

I also want to thank every other member of the Mrksich Group who has shaped and guided me throughout my PhD. First, thank you to **Dr. Justin Modica** for having the determination and resolve to single-handedly develop the megamolecule platform into what it is today. Literally, none of us would be here and doing what we do if it wasn't for you, and we're all very grateful you've decided to stick around to develop this technology, despite wanting to drop everything and move to California more than once. You're a great scientific mentor and friend, and I look forward to catching up with you and Grace over the ensuing years.

**Sraeyes Sridhar**, thank you for being my best lab friend and a source of unrivaled consistency over the last four years. You are truly the best "idea guy" I've ever met, and I know you'll help invent something game changing one day. I'm so interested to see how your career arc develops after you defend next year, and I'm confident that after pondering literally every possibility of interest within finance, government, industry research, and academia, you'll settle on the right path. Thanks for hosting such an incredible engagement weekend in Oklahoma City, and I wish you and Preeya the best coordinating your formal wedding next year! Here's to hoping we can try Oriole (or a restaurant of equivalent magnitude) together before we move to New York City.

**Dr. Blaise Kimmel**, thanks for being the epitome of scientific inspiration the four years we shared together. I don't think I've ever met anyone as passionate about science nor as capable to execute their own research program at such a young age as you. The fact that you landed your dream job as an assistant professor at *The Ohio State University* at just 26 years old would blow

me away...except that I know you and I know how hard you've worked for this and what you're capable of, and so I'm really nothing but proud of all you've accomplished. I still can't believe there was a three month stretch when you'd get into lab at 4AM so that you and John could navigate the single-car situation. I look forward to (hopefully!) visiting you in Nashville and (definitely!) one day in Columbus!

To everyone else in lab – notably **Dr. Lindsey Pearson (Szymczak)**, **Dr. Patrick O'Kane**, **Dr. Alexei Ten**, and **Juliet Roll**, thanks for being great colleagues and friends, especially in the pre-pandemic years. You all were very kind to me when I joined the group and helped me find my place in graduate school, which is something I really appreciate. Juliet, I can't wait to see a production of yours in New York, and I'm proud of you for taking that leap of faith after fast-tracking your Masters! And Lindsey – I can't wait to see how much you and Steve love your future daughter! You'll be an awesome Mom. I also wanted to thank my collaborator, **Zeynab Mousavikhamene**, for her work on the machine learning project. Looking back on it all, I feel like we got lucky with finishing and publishing that manuscript given your change of laboratory halfway through, and I'm grateful for all the hard work that you put into it. Thank you also to **Yael Mayer**, **Brianna Bullock**, and **Dr. Eric Berns** for handling the administrative responsibilities for the group, which I hope never went unappreciated. You three help lab run as efficiently as it does, and as a graduate student, I am very thankful that we don't have to worry about handling order processing, lab scheduling, or writing grants. Thanks especially to Brianna for helping me create the new-and-improved [Mrksich Group](#) website during the initial pandemic lockdown – it was fun to build some cursory web development skills alongside you! Finally, I wanted to thank **Dr. Dina Arvanitis**, **Dr. Jessica Hornick**, and **Dr. Elena Antonova** for their help with training and

troubleshooting various microscopy aspects of my dissertation at their respective core facilities here on campus (Center for Advanced Microscopy; Biological Imaging Facility).

Those who know me also know that I grew up in the Chicago suburbs, and so I've been beyond fortunate to have such a robust and supportive network of family and friends in the area while I completed my PhD. I still can't believe I formed and maintained childhood friendships that I know will be lifelong, and I hope I don't ever take it for granted because I know how uncommon (and frankly weird) it is. You all made being an only child a lot less lonely. This will be long-winded, which is on-brand for me, so consider skipping to the end if you'd like. Here goes nothing:

**David Reiss**, you're my oldest friend (almost 25 years!), and I'm so happy we've been able to remain close all these years. Mike and Sally's generosity every year hosting us at Green Lake is honestly what I look forward to most every single summer. You have an amazing sense of humor, and your upbeat optimism is a natural foil to my brooding cynicism. I love you and Taylor (and Ruby!), and I can't wait to see what the future holds for you guys. Let's try not to hurt our wallets too much in the bleachers this summer...though, I've been mostly selling my tickets with how horrible the Cubs are...

**Mike Davis**, your undying curiosity to further understand the world and the people who inhabit it really inspires me to never be intellectually complacent with anything I ever do. You really do have a mind for a PhD. From all our Thrice covers, Coheed concerts, and trips to the cabin, my fondest memories of adulthood almost always include you and music. I'll never forget when you visited me freshman year at Duke when I was having a tough time, and it really meant the world to me. I think, to this day, that was the most considerate thing anyone has ever done for me, and I'll love you forever for it. I can't wait to get back to making music with you, and I'm super excited to see you in Denver!

**Peter Bernardi**, it feels like we've been close friends for as long as we've known each other, which probably speaks to the sheer density of great times we've managed to have. Bulge Wedding #1 and #2, smoking the field in water badminton, bar crawls in Bloomington, ski trip to Copper, and of course, every Draft Weekend™. While the first time you visited DC really cemented our friendship at Dan's Cafe, the defining aspect has sadly been Hearthstone. Man, screw that game. But it allowed us to keep in touch so effortlessly and casually, even if just for fifteen minutes a day. You make it so easy on everyone to keep in touch and to love you, and that effort and drive you put into your relationships has clearly allowed your entrepreneurial career to be as wildly successful as it already has. Can't wait to visit you and Laura very soon (in your new house!) for Bulge Wedding #3!

**John Spacapan**, my friendship with you easily has the most intensity and passion, and those feelings are what make life worth living. Your boundless energy to experience new things, meet new people, and discuss matters of our fragile world provides such needed intellectual stimulation from mundane day-to-days. Though we have often naturally fought over trivialities, it always came from a place of love – even though you were (and are) always wrong. You and your family have been so generous over the years, and you know I consider Ed and Lise to be my second parents. We had incredible time in DC together trying to navigate the post-college world, and there have been many times during grad school I wish I could go back to them, if only for a day. Simpler times. And you even remarked that we'd feel that way about those times when we were going through them; that's what I really appreciate about you. I can't wait to see where your career takes you, I just hope it can support your love of antiques.

**Henry Taylor**, thanks for always being the most likable Taylor grandkid. I'll always appreciate having your place to crash at in Glenview during my first year of grad school, it was

always a wonderful time turning our brains off. Your resilience and determination have been nothing short of inspiring. I vividly remember talking to you on the phone outside of Tech and almost crying after you got into W&M Law because I knew how hard you worked for that moment. It's naturally fitting that you've had as much success as you have had, and you know how proud I am of you. Thanks also for sharing a sacred love for Wrigley Field, and I hope to make it up to the Northwoods at some point with you.

**Jack Lapinski**, thank you for being the most relentlessly positive person I've ever known. You always put a smile on my face, even when you're kicking everyone's butts at literally every game possible. I also know I speak for literally all our friends, but you are easily the most enjoyable person to golf with, and your casual advice on quick fixes always make my rounds more enjoyable and prevent me from being the worst iteration of myself. I'm so proud of who you've become, and your steadfast dedication to getting your degree will assuredly pay off. I'll be front and center the day you become the clinical psychologist you've always meant to be. Plus, we need more men in that field. I'm so happy you've found Yesi, and I look forward to some more games of Yacko after I finish writing this thing.

**Dr. John Denlinger, DPT** thank you for being often the only person who was down to hangout during the early days of the pandemic, even it was just us coming over to hang and play Jackbox games for the 53<sup>rd</sup> time. Your persistence and dedication to your career – and to pay off your student loans – is inspiring (and grounding for those who don't have any). I'm so happy you've been able to finally switch jobs (and have that sweet 5:30AM wake-up), and I can't wait until you have a clinic of your own one day. While I look forward to your upcoming 7-6 fantasy football season that'll culminate in a league championship, I'm looking even more forward to Bulge Wedding #4 to watch you and Maggie say your "I Do. She's been such a wonderful addition

to our lives, and her parents' generosity to host up in Saugatuck over the summers is appreciated by everyone!

**Brad Snider**, you are currently in Europe on a much-needed soul-discovery galivant, and I'm so, so happy that you are. Your upbeat optimism is always infectious, and you always seem to have an ability to calm my neuroticism whenever I'm in a funk. I'll always remember when you showed me Anberlin during sophomore year cross country, and that was *my band* for at least a year. Also...cheers to making stupid Fresca videos in my basement. Not sure why I primarily associate you with those because we are at fault for creating them. Can't wait to catch up and swap stories when you get back!

**Phil Zarcone**, my birthday twin! I'll be forever grateful for how often your folks would let us run wild in your basement/backyard, though perhaps a lot of it was to give Peyton exercise. It's amazing to see what a virtuoso you've become on the guitar, and I'm just so happy I could be the first person who got you to play One by Metallica. My trip to LA with Mike remains one of my fondest memories of adulthood, and I dream of that jam room quite often. Your pursuit of designing and building your own video game takes such everlasting focus and determination, which is something I so deeply admire about you. And on top of it all? It's a sick game. I can't wait to see what the final product looks like. I'm also so proud of the thorough self-reflection you underwent while navigating COVID out in LA by yourself; clearly, you rose to the challenge, and emerged a happier and more fulfilled person.

**Josh Hutchings**, your never-ending desire to get off your ass and go conquer something every single day is something I wish I could relate to. Only your optimism and zest for life could make living in Rapid City as a 23-year-old a great time. It has been awesome to see how much you've enjoyed being an uncle, and it was a blast from the past to jam in your basement like old

times this past winter. Sorry to be moving almost right after you move back to Chicago, but I'm sure we'll have some great times, nonetheless. And come to NYC! Of all my close friends, I know you visiting will be nothing short of a blast. We do have to arrange another trip to the Ontario cabin one of these next few summers...please plant that seed in Barry's mind.

**Sarah Guminski**, you are one of the wisest and most compassionate people I've ever met. You're always quick to give needed perspective to situations, and your maturity has always been a hallmark of your personality. I've always appreciated you and your parents' generosity with giving us tickets to Bears games, despite my love-hate (okay, fine, mostly hate) relationship for that franchise. I'm so proud to see how you've carved your own path in your career, and it makes me happy to see how much you love what you do. Can't wait to get the band back together to celebrate Andrew and Cayla's love at the end of July!

**Brian Johnson**, thanks for always being yourself. Your re-discovered pursuit of creative passions has clearly fulfilled you more than anything else could have, and I deeply respect your decision to do so. I admire your ability to stand firm on choices around your health and to never allow yourself to be tempted to compromise. We had some incredible times with the band (s/o Andrew and Danny), but I always appreciated how you and your parents would let me crash in your basement those summers in college when I had no Hinsdale home and was dealing with a lot of inner anxiety about being a root-less adult. I can't wait to see how your movie turns out!

**Andrew Regnery**, I had to save the best for last. You have been my best friend for 18 years, and I could probably write an entire separate dissertation on how much you have meant and will continue to mean to me. I really don't even know what to say or where to start, but you have been the textbook definition of what unwavering love and support look like. Like, look at your bachelor party weekend. The sheer number of kind, ambitious, and generous guys who you

effortlessly surround yourself with is insane. I really don't know what kind of people we all would be without you, and I think I can speak for everyone when I say that. We have so many inspiring friends in our life (see: above), but no one inspires me to be a better person every single day than you do. I love you so much, I love Cayla so much, and I can't wait to stand beside you both on your wedding day this summer. PS – thanks **Jake Regnery** for always being the funnier Regnery sibling, and you threw the most unbelievable bachelor party. Proud of you for landing that new valve engineering job! Diana is an awesome woman, don't screw it up. Can't wait to see what's in store for you two as well!

Before I wrap this up, there are more people I'd like to acknowledge, but I have an entire dissertation to write. In no particular order: thank you to my ex-roommates over the years: **Geoffrey King, Jasper Dittmar, and Matt Vasher**. I know I can be kind of a pain to live with at times, but we made it work. Best of luck to you, Matt, as you wrap up your dissertation, too! Special thanks to **Rebekka Nickman** for being such a great friend, a pseudo-roommate, and for playing double coverage with John. **Dr. Maggie Boyd** and **Michael Kinsey**, thanks for being such wonderful positive friends the last five years; that trip to Montana is easily one of my fondest memories of graduate school. I can even look back on the goat incident and laugh now. And congratulations to Maggie for recently defending! Hopefully, we'll all be able to join you on the other side of that bridge in time. Also looking forward to your wedding, whenever and wherever that is! Cannot wait to visit you guys in San Diego; I'm just worried I won't want to leave once I do.

Further thanks to the other members of my fantasy football league, YWDI;NB - **PJ Gallo, Eric Jaugey, Connor Tatooles, Tate Schneiderman, and Corey Devine**. You guys rock, and Draft Weekend™ is one of the highlights of my year, every year, even though our collective

interest in the NFL is clearly waning. Thanks to **Tom Barrett** and **Nate Brulport** for always making me cringe, roll my eyes, and laugh – in that order. Thanks to **Analisa** and **Sophia Bernardi** for always sending the dankest memes.

Thanks to more childhood friends: **Danny Delaney, Jack Ciolino, Billy Fayette, Jenn Oetter, Allie Stratis, Morgan Sherlock, Ashley Reimann, Zafar Bandukda, Jack Miller, Evan Roglich, Monika Jones, Maggie Langhamer, Connor Burch, Cole Adams, Kelsey Ramsey, Gina Conti, Dr. Ana Pervan, Dr. Olivia Purcell, and Michelle Scarpino.** Thanks to college friends **Eddy Hernandez, David Spruill, Richy Yun, John Bralich, Alex Gutierrez, Ben Spilsbury, Derek Schocken, Dr. Anna Knight, Molly Linabarger, Dr. Paige Mihalsky, Dr. Caitie Cristante, Dutch Waanders, Christine Schindler, Kendall Covington, Dr. Drew Lynch, Mikayla Wickman, and Chelsea Pieroni.** Thanks to college roommates **Jack Mercola, Daniel Ding, and Xavier de Gunten** – I was the worst version of myself during college, thanks for putting up with me. I swear you'd like me more now. Thanks to DC friends **Baxter Carr, Andy Vogt, Brock Jacobi, Conner Geery, Dan Jacobson;** and other friends I met along the way **Erik Nerison, Jeremy Kulbarsh, Alex Head, Andrew McHugh, Tyler Decker.** You've all had an impact on me in some form and helped shape who I am today; and for this, I am grateful.

Thanks also to various family members who have supported me over the years – **Chris and Joe Albrecht, Cindy Parisoli, Aimee and Greg Soyka,** and my cousins **David, Emily, Liz, Stephanie, and Lindsay.** To my parents, **Scott and Jean Sykora** (and Cairo!): You have been role models of unwavering consistency, love, and support; I can only hope to repay you over the years as I finally start my big boy job and learn how to relax a little bit. I love you. And as always, thank you so much for all the support with vacations, with rent, with everything – I hope you see yourself within this dissertation, as literally none of this would have been remotely possible without you

two. I also want to acknowledge my grandmother, **Judy Sykora**, who I know is very proud of this accomplishment. Your collective focus on STEM education clearly resonated, and I just hope I can continue to make you proud throughout the rest of my life. I wish you were still here to see my defense.

And, finally, **Kelli Pape**, my love, thank you for being you and loving me. I don't really know how you do it. You've unlocked something within me to be the best version of myself, and I never want to take it for granted. You are the most naturally intelligent, kind, hard-working, and yet still humble person I've ever met. We survived pandemic lockdowns, binged educational YouTube channels, took some incredible trips (eleven states and counting!), and got the goobiest cat Galaxy. I'm so beyond lucky to have you in my life, and I can't wait for our journey together to New York. I'm not very good at being kind to myself...or at celebrating myself...but you always ensure I do it anyway. I also want to thank your family for taking me in as one of their own over several family vacations, and I can't wait to create more memories alongside them. Especially once the kids have more agency, heh. I want to write more about how much I love you, but I have this funny aqueous salty substance emanating from my eyeballs, and I still have about 200 pages to write.

## List of Abbreviations

1 – 2Rs15d	Cutinase-2Rs15d monomer, N-terminal fusion
2 – 2Rs15d	Cutinase-2Rs15d dimer, N-terminal fusion
3 – 2Rs15d	Cutinase-2Rs15d trimer, N-terminal fusion
1 – 5F7	Cutinase-5F7 monomer, C-terminal fusion
2 – 5F7	Cutinase-5F7 dimer, C-terminal fusion
3 – 5F7	Cutinase-5F7 trimer, C-terminal fusion
1C	Cutinase-trastuzumab Fab monomer, heavy-chain C-terminal fusion
2C	Cutinase-trastuzumab Fab dimer, heavy-chain C-terminal fusion
3C	Cutinase-trastuzumab Fab trimer, heavy-chain C-terminal fusion
1N	Cutinase-trastuzumab monomer, heavy-chain N-terminal fusion
2N	Cutinase-trastuzumab dimer, heavy-chain N-terminal fusion
3N	Cutinase-trastuzumab trimer, heavy-chain N-terminal fusion
CDRs	Complementarity-determining regions
Da	Dalton
DPBS	Dulbecco's phosphate-buffered saline
<i>E. coli</i>	Escherichia coli
EC <sub>50</sub>	Half-maximal activating concentration
EDTA	Ethylenediaminetetraacetic acid
EG	Ethylene glycol
EGFR	Epidermal growth factor receptor
eq	Equivalents
ESI-MS	Electrospray-Ionization Mass Spectrometry

Fab	Antigen-binding fragment
FBS	Fetal bovine serum
Fc	Fragment crystallizable region
FPLC	Fast-protein liquid chromatography
FRET	Forster resonance energy transfer
GFP	Green fluorescent protein
H <sub>2</sub> O	Water
HCl	Hydrochloric acid
HER2	Human epidermal growth factor 2
HisTag	Histidine affinity tag
IC <sub>50</sub>	Half-maximal inhibitory concentration
IgG	Immunoglobulin protein G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
$K_D$	Equilibrium binding constant
$k_e$	Effective internalization rate constant
$k_{1/2}$	Half-life or half-time determined kinetic rate constant
$k_{obs}$	Observed kinetic rate constant
$k_{off}$	Kinetic off rate constant for dissociation
$k_{on}$	Kinetic on rate constant for association
kDa	Kilodalton
mg	Milligram
mL	Milliliter
mM	Millimolar

MPI	Mean pixel intensity
mRNA	Messenger ribonucleic acid
MW	Molecular weight
nL	Nanoliter
nm	Nanometer
nmol	Nanomole
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDMS	Polydimethylsiloxane
PEG	Poly(ethylene glycol)
pH	Potential of hydrogen
rpm	Rotations per minute
SAM	Self-assembled monolayers
SAMDI	SAM for matrix-assisted laser desorption/ionization
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide electrophoresis
SEC	Size exclusion chromatography
$\mu\text{L}/\text{uL}$	Microliter (probably inconsistent with u or $\mu$ )
$\mu\text{m}/\text{um}$	Micrometer
$\mu\text{M}/\text{uM}$	Micromolar

*To my parents, Jean and Scott Sykora. None of this would have been possible without you.*

## Table of Contents

<b>Copyright Page</b> .....	2
<b>Abstract</b> .....	3
<b>Acknowledgements</b> .....	7
<b>Abbreviations</b> .....	20
<b>Dedication Page</b> .....	23
<b>Table of Contents</b> .....	24
<b>List of Figures</b> .....	26
<b>List of Tables</b> .....	30
<b>Chapter 1 The Megamolecule Platform for Assembly of Protein Scaffolds</b>	
1.1 Synthetic Biology and Structure-Function Relationships.....	31
1.2 Limitations of Assembly of Higher-Ordered Structures.....	34
1.3 Megamolecules .....	34
1.4 Megamolecules in Literature .....	36
1.5 Biologics for Cancer Therapeutics .....	41
1.6 Biologics for Protein Sensors / Diagnostics .....	43
1.7 Overview of Dissertation .....	44
<b>Chapter 2 Megamolecules to Probe Binding Affinities, Internalization Rates, and Inhibition of Proliferation for HER2-Targeting Cancer Therapy</b>	
2.1 Introduction.....	47
2.2 Results.....	51
2.3 Conclusions.....	67
2.4 Experimental.....	72
<b>Chapter 3 Megamolecules as Sensors for Reversible Control of Protein Quaternary Structure Using Dynamic Covalent Linking Groups</b>	

3.1 Introduction.....	84
3.2 Results.....	87
3.3 Conclusions.....	106
3.4 Experimental.....	106
<b>Chapter 4 Single Cell Morphological and Cytoskeletal Features Can Discriminate Cancer from Non-Cancer Cell Lines</b>	
4.1 Introduction.....	111
4.2 Results.....	116
4.3 Conclusions.....	127
4.4 Experimental.....	130
<b>Chapter 5 Profiling Phosphatase Arrays and Quantifying Focal Adhesion Dynamics</b>	
5.1 Introduction.....	139
5.2 Protein Phosphorylation in the Proteome .....	141
5.3 Results.....	143
5.4 Conclusions of Peptide Work .....	148
5.5 Quantifying Focal Adhesion Dynamics.....	150
5.6 Experimental.....	154
<b>Chapter 6 Conclusions, Insights, and Future Directions</b>	
6.1 Megamolecule Thoughts.....	162
5.6 Final Research Thoughts.....	164
<b>References .....</b>	<b>165</b>

## List of Figures

<b>Figure 1.1</b> Major aspects of <i>de novo</i> protein design .....	33
<b>Figure 1.2</b> Our two primary enzyme-inhibitor pairs: cutinase and SnapTag.....	35
<b>Figure 1.3</b> Growing megamolecule chains show rapid kinetics .....	37
<b>Figure 1.4</b> Cyclic megamolecules form via homobifunctional benzylguanine linker; this was also verified by SDS-PAGE .....	38
<b>Figure 1.5</b> Synthesis of megamolecule protein scaffolds from solid-phase support.....	39
<b>Figure 1.6</b> Library of megamolecule-based trastuzumab mimics; inhibition of cell proliferation and individual scaffold binding affinities were quantified for each .....	40
<b>Figure 1.7</b> General structure of a mAb; binding can induce cell death directly or indirectly through immune cell recruitment and complement cascade.....	42
<b>Figure 1.8</b> Current and projected sales of small molecules and biologics .....	43
<b>Figure 2.1</b> HER2 activation through homo and heterodimerization .....	48
<b>Figure 2.2</b> HER2 and its various domains; different drugs target different epitopes .....	50
<b>Figure 2.3</b> Synthesis of megamolecule antibody mimics as well as TZB cleavage.....	53
<b>Figure 2.4</b> Analytical SEC demonstrates protein globularity in solution .....	54
<b>Figure 2.5</b> ESI-MS shows good agreement with theoretical masses below 150kDa.....	55
<b>Figure 2.6</b> Megamolecules selectively bind cells expressing HER2; binding affinities can also be quantified via flow cytometry.....	56
<b>Figure 2.7</b> Megamolecule scaffolds internalize and can be quantified by pH-sensitive rhodamine derivative dye; plots at 5nM confirm similar uptake rates across all scaffolds	58
<b>Figure 2.8</b> Megamolecule internalization negligible in HER2 $-/-$ cells after two days .....	59
<b>Figure 2.9</b> Antibody internalization vastly inhibited by clathrin inhibitor after two days.	59
<b>Figure 2.10</b> Low affinity nanobody C8 and its respective megamolecule structures do not internalize efficiently; only the trimer shows any internalization above background.....	60
<b>Figure 2.11</b> 2C and 2-Nb do not dissociate from cell surface after one day in culture, demonstrating that there is not a significant difference in off-rate that is affecting their	

different internalization rates .....	61
<b>Figure 2.12</b> Previous co-authored work demonstrate only bivalent scaffolds show slow dissociation compared to monovalent fusions .....	61
<b>Figure 2.13</b> Preparation of biparatopic scaffolds; internalization curves show 2-3 fold faster uptake of the megamolecule-receptor complex .....	63
<b>Figure 2.14</b> 2Rs15d monoparatopic scaffolds internalize slowly like 5F7 scaffolds .....	64
<b>Figure 2.15</b> Cell viability curves after six days in culture; only Fab-based monoparatopic megamolecule structures demonstrate inhibition of cell proliferation .....	66
<b>Figure 2.16</b> Normalized internalization rate for all bivalent scaffolds shows the 2-3 fold rate increase in the net internalization over trastuzumab.....	67
<b>Figure 2.17</b> Flow cytometry of trastuzumab-based biparatopic megamolecules.....	68
<b>Figure 2.18</b> Flow cytometry of 5F7-based biparatopic megamolecules .....	68
<b>Figure 2.19</b> MDA-MB-453 cells do not show sustained internalization, which is supported by literature suggesting weaker receptor recycling .....	69
<b>Figure 3.1</b> Schematic and synthesis of terpyridine-functionalized megamolecule scaffold; increase in FRET efficiency can be seen with increased metal concentration; treatment with different divalent transition metals show metal-dependent cyclization response; backed up by SAXS data.....	90
<b>Figure 3.2</b> TCS and NCS plain and functionalized building blocks' theoretical masses correlate with ESI-MS .....	91
<b>Figure 3.3</b> TCS, NCS, and three scaffold lengths can be seen by SDS-PAGE .....	91
<b>Figure 3.4</b> Scaffolds only undergo bidentate cyclization and increased FRET efficiency when both cutinase domains are functionalized with terpy ligands and both NCS, TCS joined .....	93
<b>Figure 3.5</b> Synthesis and FRET efficiencies of intermediate and long scaffolds demonstrate both scaffold and metal-specific relationships.....	95
<b>Figure 3.6</b> Kinetic schemes of transition metals with free terpy as well as with our scaffolds; rate of formation slower for short scaffold than longer scaffolds, second-order rate constants scale with length, and FRET efficiencies show bis-monodentate species at high concentration .....	99

<b>Figure 3.7</b> Dissociation of cobalt-cyclized scaffolds show no length dependence; utilizing cycles of cobalt and EDTA, demonstrated scaffold reversibility .....	104
<b>Figure 3.7</b> Nickel-coordinated scaffolds demonstrate negligible dissociation .....	105
<b>Figure 4.1</b> Cell cytoskeleton comprised of actin, microtubules, and intermediate filaments that can be labeled with dyes or fluorescent antibodies .....	113
<b>Figure 4.2</b> During EMT, cytoskeleton undergoes dynamic shift from epithelial to more migratory, mesenchymal phenotype .....	114
<b>Figure 4.3</b> Automated image processing, feature-extraction framework quantifies features of cellular cytoskeletal and morphological structures from single cell images .....	118
<b>Figure 4.4</b> Each image has cytoskeletal and morphological features quantified including actin intensity and cell roughness, respectively .....	119
<b>Figure 4.5</b> Single HFF and HT-1080 cells are segmented, features extracted; differences quantified by Fisher Discriminant Analysis for discrimination by ML.....	120
<b>Figure 4.6</b> Accuracies for all whole cell pairwise comparisons; Fisher Analysis repeated for all 28 comparisons.....	122
<b>Figure 4.7</b> Example of microcontact printing-based image analysis; pairwise accuracy increases when hard-to-discriminate cells are micropatterned into squares .....	124
<b>Figure 4.8</b> The algorithm demonstrates generalizability towards cancer cells that were not utilized during model training; Fisher Analysis repeated for whole cell class .....	126
<b>Figure 5.1</b> General SAMDI-MS workflow to profile phosphatase activity.....	140
<b>Figure 5.2</b> Phosphorylation balanced by kinases and phosphatases .....	141
<b>Figure 5.3</b> Global phosphatase activity from NIH/3T3 cell lysate .....	144
<b>Figure 5.4</b> Ionization efficiencies consistent across peptide libraries.....	145
<b>Figure 5.5</b> ATP required for Src kinase activity; thus no new phosphorylation seen when quantitating phosphatase activity on phosphor-peptides .....	146
<b>Figure 5.6</b> Activity and specificity of phosphatases consistent across disparate cell lines .....	147
<b>Figure 5.7</b> Phosphatase activity does not significantly vary across each peptide library; largest deviation is around 20% .....	149

- Figure 5.8** A hepta-saccharide protecting group released when exposed to 405 nm light, revealing cyclic RGD for cells to bind ..... 151
- Figure 5.9** Deprotected surfaces present both linear and cyclic RGD ..... 151
- Figure 5.10** Protected cRGD peptide gets deprotected, quantified by mass spec..... 152
- Figure 5.11** HT-1080 transfected with GFP-paxillin allows for spatiotemporal quantification of focal adhesion dynamics..... 152
- Figure 5.12** Cells increased protrusion into regions of cyclic RGD, quantified by counting number of focal adhesions into deprotected region ..... 154

## List of Tables

<b>Table 2.1</b> Observed binding $EC_{50}$ values via flow cytometry of monoparatomic megamolecules on BT-474 cancer cells .....	57
<b>Table 2.2</b> Observed internalization rates of monoparatomic megamolecules on BT-474 cancer cells .....	62
<b>Table 2.3</b> Observed internalization rates and binding $EC_{50}$ values via flow cytometry of bipolaratopic megamolecules on BT-474 cancer cells .....	65
<b>Table 2.4</b> Calculated and experimental molecular weights via ESI-MS with extinction coefficients .....	79
<b>Table 3.1</b> Kinetic and Thermodynamic Parameters for Formation and Dissociation of Transition Metal Complexes of Terpyridine and EDTA at 25° .....	97
<b>Table 4.1</b> Description and representative images of cell lines used in <b>Chapter 4</b> .....	117

## Chapter 1

### The Megamolecule Platform for Assembly of Proteins

#### 1.1 Synthetic Biology and Structure-Function Relationships

At the turn of the 21<sup>st</sup> century, the field of synthetic biology was born with the invention of two gene regulatory networks – the toggle switch<sup>1</sup> and the repressilator<sup>2</sup> – which used a promoter-repressor bistable<sup>1</sup> or tristable<sup>2</sup> gene circuit system. For the first time, engineers demonstrated direct control over the central dogma, whose simplified version states that DNA gets transcribed into mRNA, which then gets translated into a string of amino acids, which then folds into a functional protein. Over two decades of biological engineering later, complex genetic circuits can now be built into cell systems that elegantly and spatiotemporally compute specific outputs from various analog inputs<sup>3</sup>, therefore controlling for protein expression. Controlling the computational outputs of a living cell will continue to revolutionize biotechnology by improving existing processes and/or enabling novel biological applications.

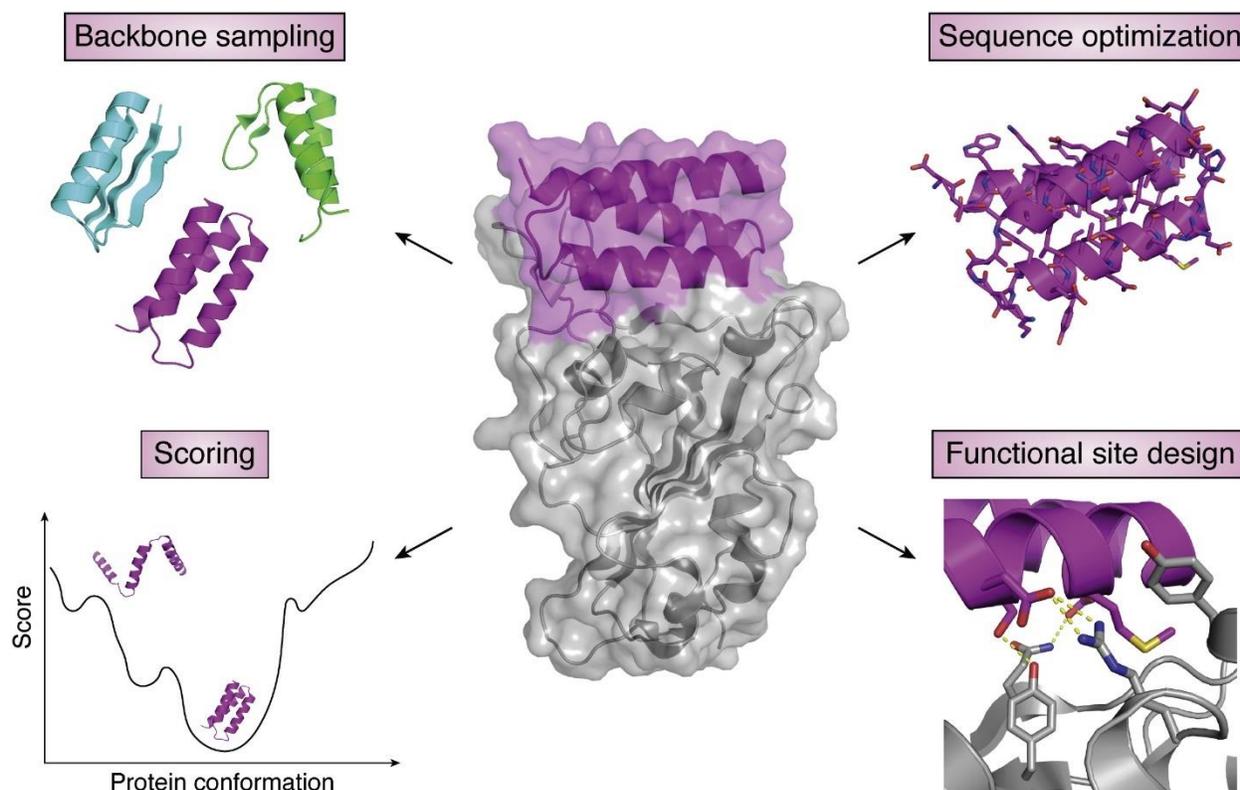
Advancements in both DNA synthesis<sup>4-6</sup> and DNA sequencing<sup>7</sup> have led to decreased cost with increasing demand, expediting rapid progress in the development of modern molecular biology techniques now employed by research groups across the world<sup>8</sup>. Synthetic biology extends past biomedicine, too – it now impacts agriculture<sup>9</sup>, civil engineering<sup>10</sup>, and architecture<sup>11,12</sup>. Nevertheless, precise utility over these biological building blocks – DNA, RNA, oligonucleotides, peptides, and proteins – has proven difficult, and part of this is due to their inherently complex structure-function relationships. Within the central dogma, biomolecule physical structure-function is most famously seen with the unfolding of DNA from histones for transcription<sup>13</sup>, the folding of mRNA for proper translation<sup>14</sup>, as well as the coordination of rRNA and ribosomal

proteins for effective mRNA translation into properly folded active proteins<sup>15</sup>. Without these precise biophysical phenomena, the downstream biochemistry would simply not be as efficient (or would take place entirely).

While structure-function relationships within synthetic biology have been explored through the self-assembly of rationally-designed DNA<sup>16</sup> and RNA<sup>17</sup>, proteins hold more potential due to the sheer functional diversity of 20 naturally occurring amino acids (plus modified or non-canonical amino acids like pyrrolysine, selenocysteine, and hydroxyproline). As David Baker often quips in his talks, a standard protein of 200 amino acids has  $20^{200}$  different sequences it could adopt. Proteins and their amino acid sequences can be created *de novo* – or from scratch – with novel or improved properties, often contributing to the understanding of native protein folding and behavior<sup>18-20</sup> itself. *De novo* protein engineering has facilitated development of protein switches that can change conformation via external stimuli<sup>21</sup>, protein circuits that can bypass gene regulation<sup>22,23</sup>, higher-ordered protein cages for drug delivery<sup>24,25</sup>, higher-ordered protein-based biomaterials<sup>26</sup>, and synthetic biologics for therapeutic applications<sup>27</sup>.

Traditionally, the study and development of novel proteins was a combination of painstaking labor and serendipity. Nowadays, *de novo* protein engineering can be expedited with high-throughput *in silico* methods, which integrate the mathematical energetics and folding probabilities of each amino acid, peptide strand, or secondary structure (e.g., alpha helices, beta sheets) (**Figure 1.1**). This is most famously demonstrated by Rosetta, developed by David Baker's group at University of Washington<sup>28</sup>, and AlphaFold2, developed by DeepMind Technologies under Alphabet, Inc.<sup>29</sup>. As powerful as these platforms are, each protein's DNA still must be synthesized, amplified, cloned, transformed, and expressed to validate *in silico* predictions to

verify novel behavior or function, as *de novo* proteins may not express, fold, or behave as anticipated by purely computational methods.



**Figure 1.1** Major aspects of the *de novo* protein design, which incorporate statistical likelihoods of various primary and secondary structure amino acid sequences/folds, leading to idealized sequence scoring, optimization, and, as an example here, binding pocket structure<sup>31</sup>.

Specific properties can subsequently be improved or imparted on the protein of interest by directed evolution efforts, which can be used either in parallel or in place of the *in silico* rational design. Directed evolution can impart or improve properties of *de novo* proteins including protein stability<sup>30</sup>, binding affinity<sup>31</sup>, enzyme activity<sup>32</sup>, and enzyme specificity<sup>33</sup>. Regardless, the computational boon over the past decade has helped facilitate a deeper understanding of how a protein's amino acid sequence impacts its downstream, three-dimensional folding conformation(s). This, in turn, can inform computational biologists and protein engineers how to best design and evolve proteins with novel and/or improved functionality.

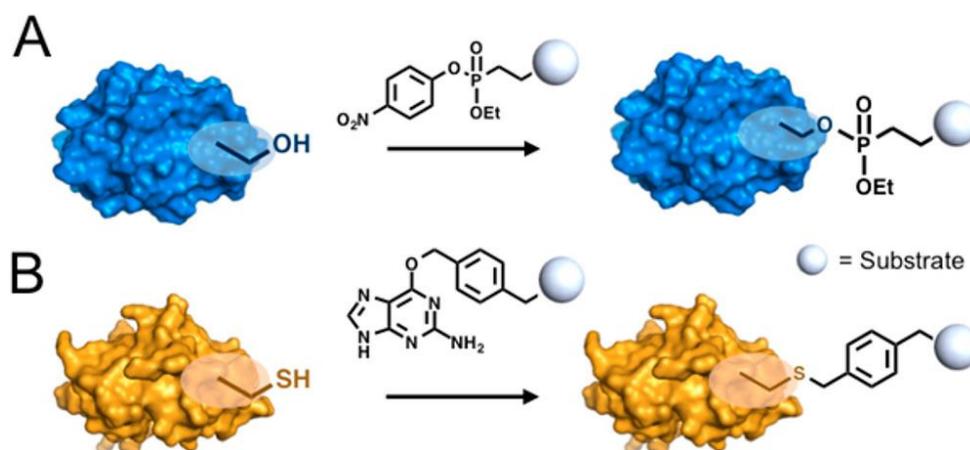
## 1.2 Limitations of Assembly of Higher-Ordered Structures

Large proteins and large protein assemblies possess their own set of challenges towards successful expression; namely, there is a metabolic burden on the cellular hosts after transfection/transformation that correlates with both protein size and plasmid copy number when proteins possess long folding times. As a result, issues of low yield and nonspecific aggregation or assembly can result<sup>34</sup>. Many have attempted to assemble larger protein structures from smaller, most easily expressed constituent parts in solution. Yet, these assemblies often rely on noncovalent protein-protein and protein-ligand interactions that can either fall apart during purification or result in nonspecific, heterogeneous products<sup>35</sup>. To combat this, covalent conjugation methods have arisen, which often either incorporate non-natural amino acids presenting reactive handles like azides for slow click chemistry-labeling<sup>36,37</sup> or utilize primary amines and thiols on the solvent-exposed surface of a protein through nonspecific conjugation<sup>38</sup>. It is therefore imperative to integrate these two approaches and develop a platform of rapid, atomically precise covalent protein assemblies for expression of large proteins, and more importantly, extended protein structures and assemblies.

## 1.3 Megamolecules

To address these existing limitations, our group has developed a set of orthogonal enzyme-inhibitor reaction pairs that selectively, rapidly, and covalently dock an electrophilic ligand of interest in the active site of a nucleophilic enzyme to form a stable enzyme-inhibitor intermediate. When these enzymes are presented within a fusion protein context, the fusion domains can then be brought together by bifunctional linkers that present the ligands of interest on either end. Typically, these linkers have poly(ethylene) glycol (PEG) backbones of 7-13 repeats. This

strategy, developed primarily by Justin Modica, is known as the **megamolecule** assembly approach, and it rapidly yields atomically-precise, covalent protein scaffolds.



**Figure 1.2** Our two primary enzyme-inhibitor pairs<sup>39</sup>. **A)** Cutinase, a serine esterase, reacts specifically with *p*-nitrophenyl phosphonate-terminated inhibitors while **B)** SnapTag, an O<sup>6</sup>-alkylguanine alkyltransferase, reacts with a benzylguanine-terminated inhibitor.

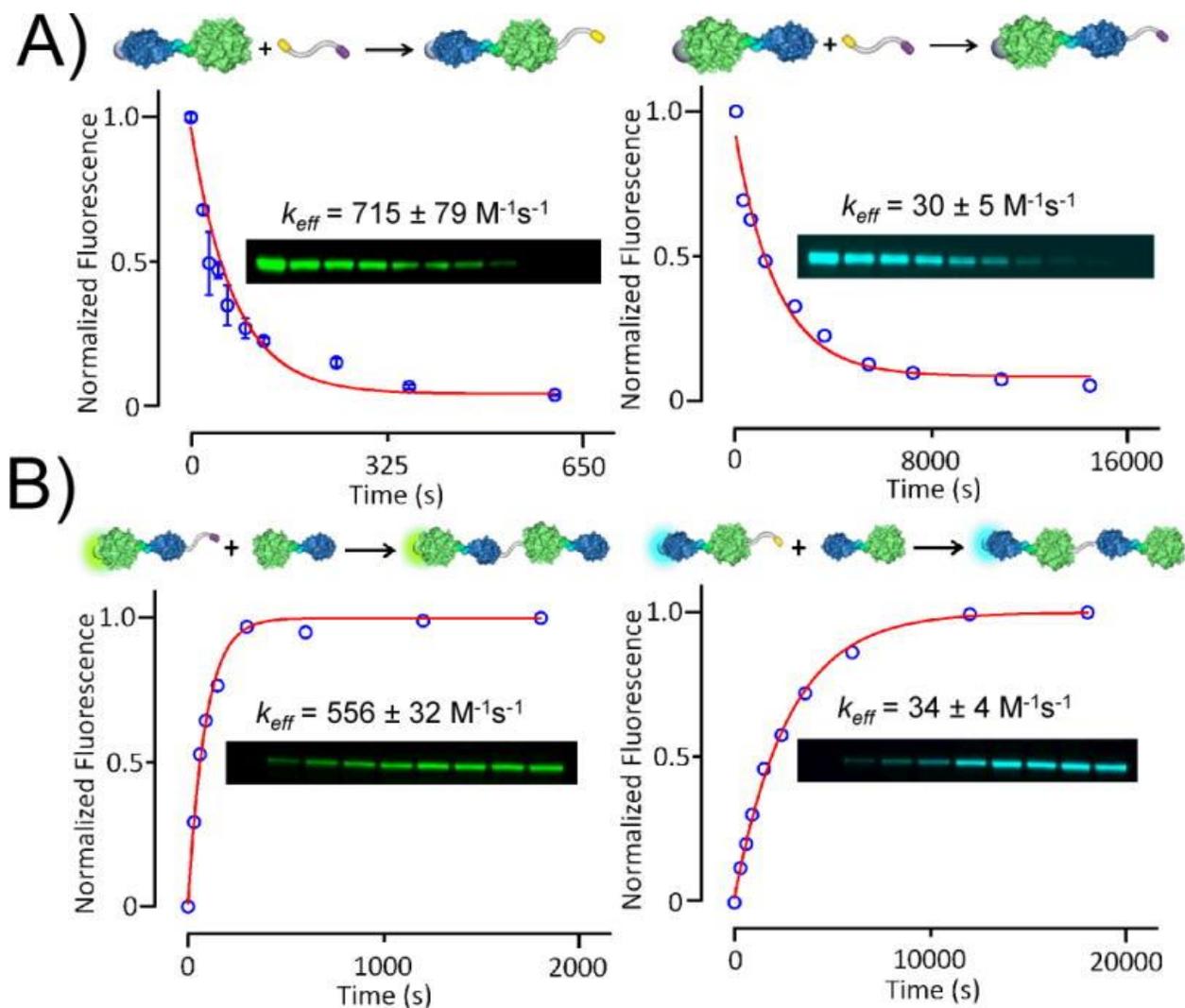
Our group primarily utilizes two orthogonal chemistries to bring fusion proteins together. First, cutinase is a serine esterase derived from phytopathogenic fungi<sup>40</sup> that site-specifically reacts with a *p*-nitrophenyl phosphonates (pNPP) at an active site serine (**Figure 1.2A**). The second is the use of SnapTag, an engineered, humanized version of the mammalian enzyme AGT that underwent directed evolution for increased efficiency and selectivity<sup>41</sup>. SnapTag site-specifically reacts with benzylguanine derivatives with an active site cysteine (**Figure 1.2B**). There is a third orthogonal chemistry just recently added to our toolbox – CRABTag<sup>42</sup>, derived by Blaise Kimmel from cellular retinoic acid binding protein II (CRABPII) – but it has not yet undergone widespread adoption and will not appear in this dissertation. Each of these enzymes are monomeric in solution – they do not aggregate or dimerize. They are also quite small and are well tolerated and well expressed by traditional, rapid *E. coli* expression systems, which is what we use in our group. Their

small size allows for synthesis of fusions on the N and/or C-terminus of the enzyme, creating functional dimers and trimers in the process. A fourth chemistry our group has reported and utilized – HaloTag, an engineered dehalogenase which reacts specifically with chloroalkanes – was found to be less stable than the other enzyme-inhibitor pairs in that it began to hydrolyze the inhibitor-enzyme complex after about a week, while the other pairs are stable at 4C for over a month. It, too, will not be presented in any original research here.

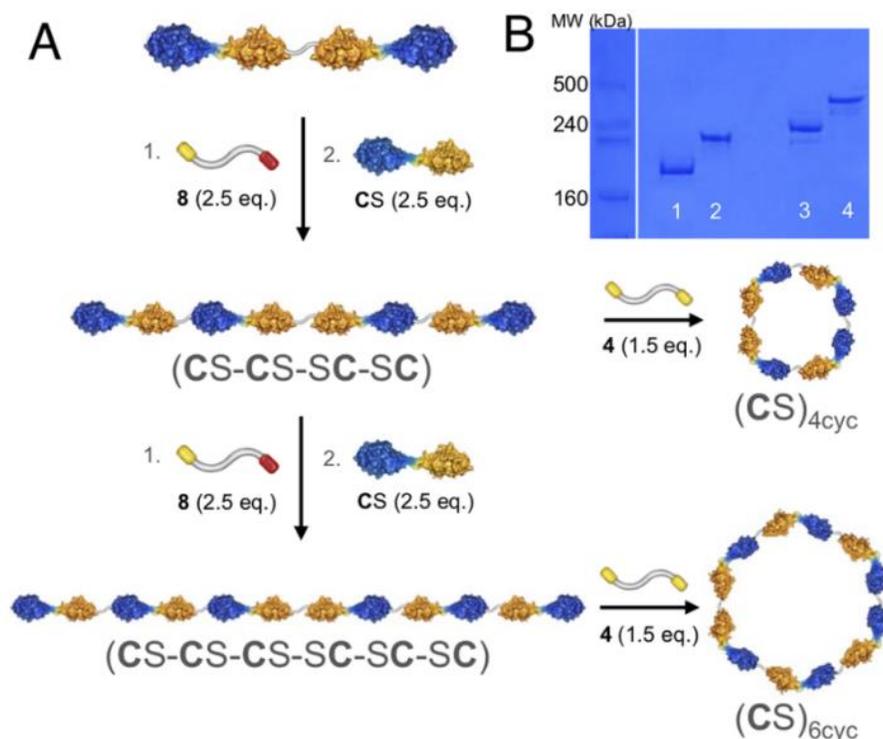
#### 1.4 Megamolecules in Literature

The first two demonstrations of the megamolecule assembly platform were proof-of-concept studies published by Justin Modica in 2012 and 2018, respectively. The 2012 publication<sup>43</sup> detailed the conjugation of a 300 kDa linear chain of cutinase-HaloTag fusion proteins. Here, heterobifunctional linkers were reacted with the HaloTag-exposed end while the cutinase end was blocked. After HaloTag reacted with the hexachloride end of the heterobifunctional PEG linker, the pNPP end could then react with the open cutinase end of a new cutinase-HaloTag fusion, extending the structure. Excitingly, the kinetics associated with each step could be estimated with an inhibitor-fluorophore conjugate, and either gain-of-signal or loss-of-signal could be modeled by a pseudo first-order curve. HaloTag reacted about 20x faster than cutinase, which makes its active site hydrolysis issue a real disappointment. Ultimately, the kinetics of each subsequent step did not dramatically decrease despite the increase in entropy of the growing strand (**Figure 1.3**), where HaloTag had an effective initial rate constant of  $715 \pm 79 \text{ M}^{-1}\text{s}^{-1}$  with its ligand and cutinase had an effective initial rate constant of  $30 \pm 5 \text{ M}^{-1}\text{s}^{-1}$  with its ligand. The second major publication<sup>39</sup> was the first to incorporate SnapTag, whose reaction rates were much faster than HaloTag or cutinase (about  $10^4 \text{ M}^{-1}\text{s}^{-1}$ ). This publication utilized SnapTag – in lieu of HaloTag – to create

cyclic megamolecules of four or six monomeric units with a homobifunctional benzylguanine linker to join the two terminal ends of the growing chain (**Figure 1.4**).



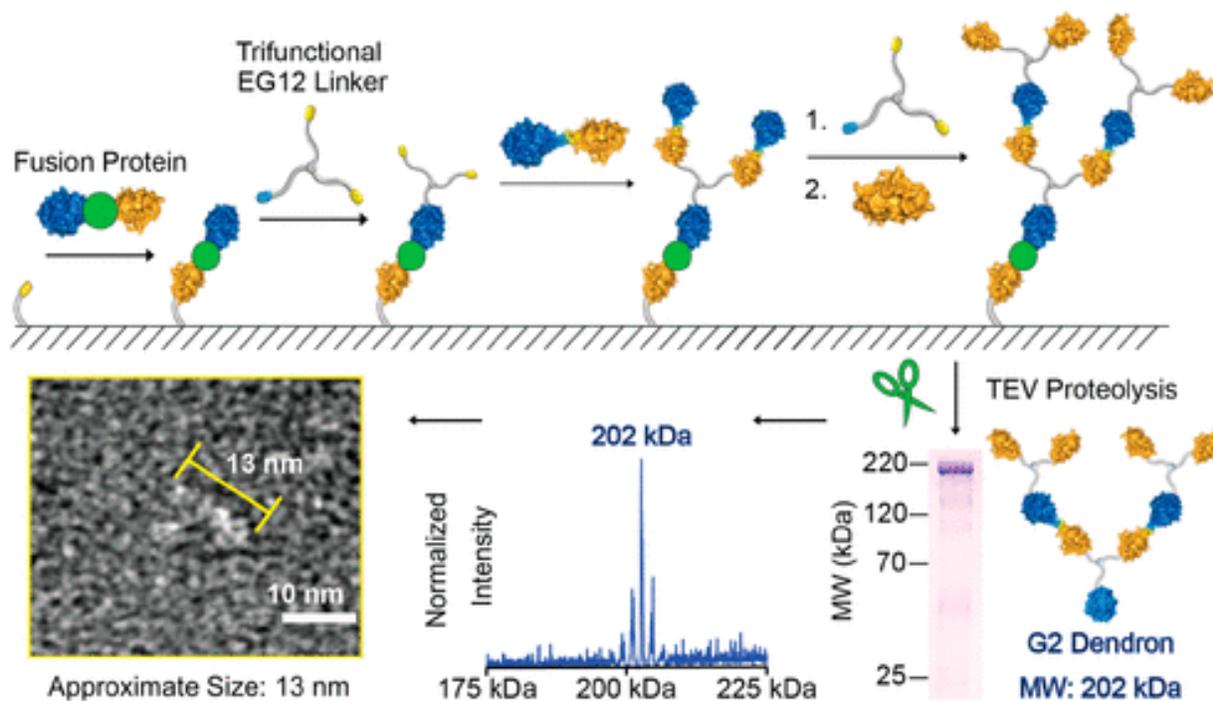
**Figure 1.3** Growing chains of megamolecules from either cutinase or HaloTag show rapid kinetics from both **A)** one fusion protein and a **B)** two fusion protein chain, demonstrating that the kinetics of the growing chain do not dramatically decrease<sup>42</sup>.



**Figure 1.4** A) Incorporation of SnapTag into cutinase-SnapTag allows for rapid and stable formation of a cyclic megamolecule with a benzylguanine homobifunctionalized linker. B) This was verified with SDS-PAGE purification after purification by SEC<sup>39</sup>.

Two more significant manuscripts were published back-to-back in 2020, both in the *Journal of the American Chemical Society (JACS)*. The first<sup>44</sup> by Blaise Kimmel demonstrated the utility of building full protein chains from megamolecule assembly chemistry in the same way protein engineers utilize solid-phase to build growing peptide chains<sup>45</sup>. Here, benzylguanine-functionalized resin beads were reacted with a SnapTag-TEV-cutinase fusion protein. TEV is a peptide sequence that gets cleaved by Tobacco Etch Virus (TEV) protease, which we have also used to cleave HisTags after protein purification when necessary (like in **Chapter 3**). From the exposed cutinase, pNPP ligands could irreversibly inhibit the active site, and further fusions could be built off from the surface (**Figure 1.5**). Once the chain was fully built, a TEV protease cleaved the product from the bead. The greatest benefit of solid-phase synthesis of megamolecules is

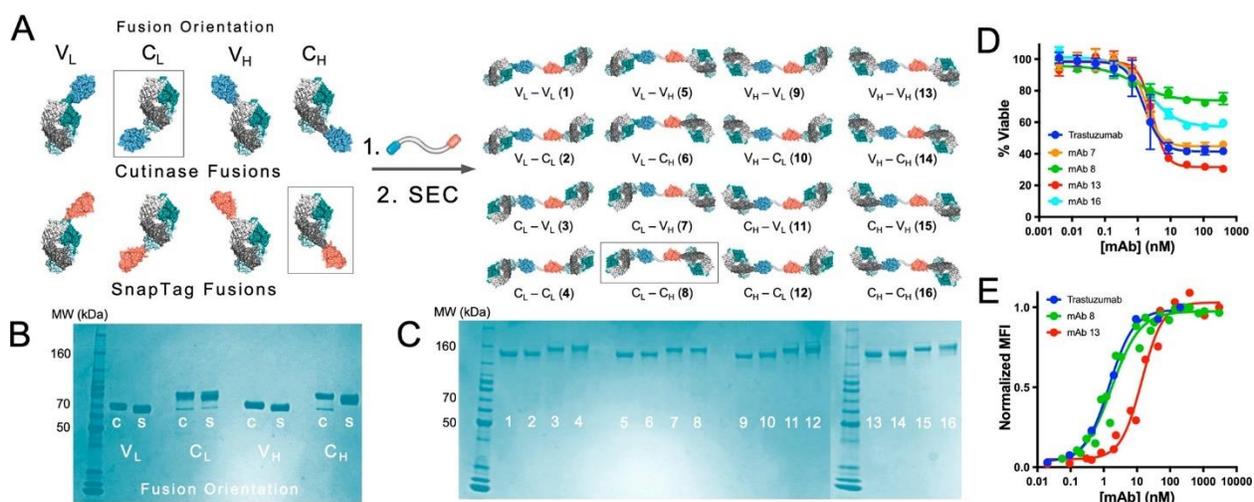
mitigating the need for purification steps after every reaction, which greatly reduces the time and material load it takes to build extended structures in solution.



**Figure 1.5.** Synthesis of megamolecule protein scaffolds build from a solid-phase support<sup>43</sup>. Here, G2 dendrimer scaffolds could be synthesized without intermediate steps of purification, saving time and material cost. The size of the scaffold matches ESI-MS and TEM, seen below.

The second and arguably most important demonstration by, again, Justin Modica<sup>46</sup> was the first therapeutic application of megamolecules. Here, a human epidermal growth factor receptor 2 (HER2)-targeting megamolecule was created by conjugating a cutinase-Fab fusion to a SnapTag-Fab fusion with a heterobifunctional linker presenting both a pNPP and benzylguanine group (**Figure 1.6**). These Fab domains had the same sequence as the Fab domain of the monoclonal antibody trastuzumab (Herceptin®), which has long been a mainstay in the clinic for HER2+ breast cancer. Importantly, each protein was fused through either the N- or C-terminus of the heavy chain or the light chain of the trastuzumab Fab. In this way, each enzyme had four possible fusion sites, so a total of sixteen scaffolds could be made. These perfectly-defined biologics ultimately retained

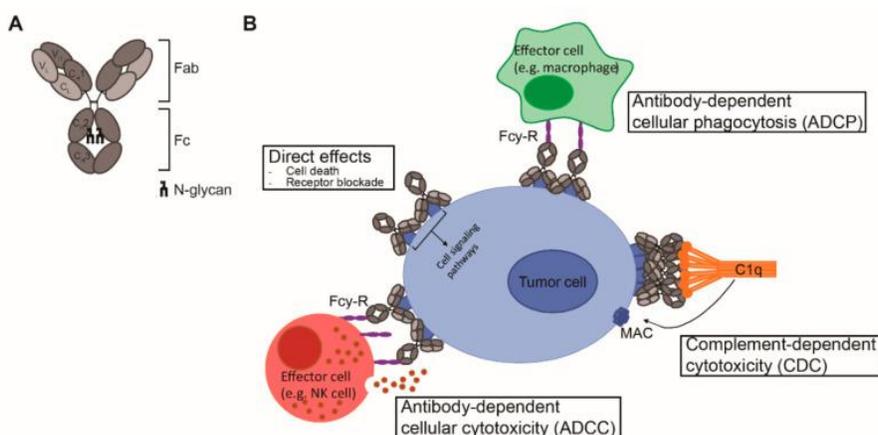
much of the ability of trastuzumab to bind HER2+ cells, to inhibit proliferation of HER2+ cells *in vitro*, and mitigate tumor growth *in vivo*. However, the strength of each biologic depended on the orientation of each fusion, which is a characteristic of modular biologics that megamolecules can sufficiently probe. Variable heavy chain ( $V_H$ ) or variable light chain ( $V_L$ ) fusions close to the Fab CDRs mitigated the binding affinity of the overall scaffold, yet the scaffold that led to the largest downstream inhibition of cell proliferation utilized fusions through the constant heavy chain  $V_H$  domain of the Fab (**Figure 1.6D**). This study demonstrated how the megamolecule assembly platform can be used to investigate how different fusion orientations can affect the efficacy of antibody mimics for targeted cancer therapy in solid tumors, and overall efficacy may not follow intuition, as we expected the best binding scaffolds to also lead to the largest inhibition of cell proliferation. Our “toolbox” of synthetic PEG linkers that irreversibly inhibit enzymes with specific warheads can thusly be used like LEGO® building blocks to rapidly build perfectly-defined, functional protein therapeutics with high-yield.



**Figure 1.6** A library of trastuzumab mimics synthesized from **A**) trastuzumab Fab-cutinase and trastuzumab Fab-SnapTag fusions of each possible orientation, creating sixteen scaffolds in the process. **B+C**) Despite having the same molecular weight, each dimer pair migrated on SDS-PAGE slightly differently. Both the **D**) *in vitro* inhibition of cell proliferation and **E**) apparent cell-binding capacity of each fusion was quantified for each candidate.<sup>46</sup>

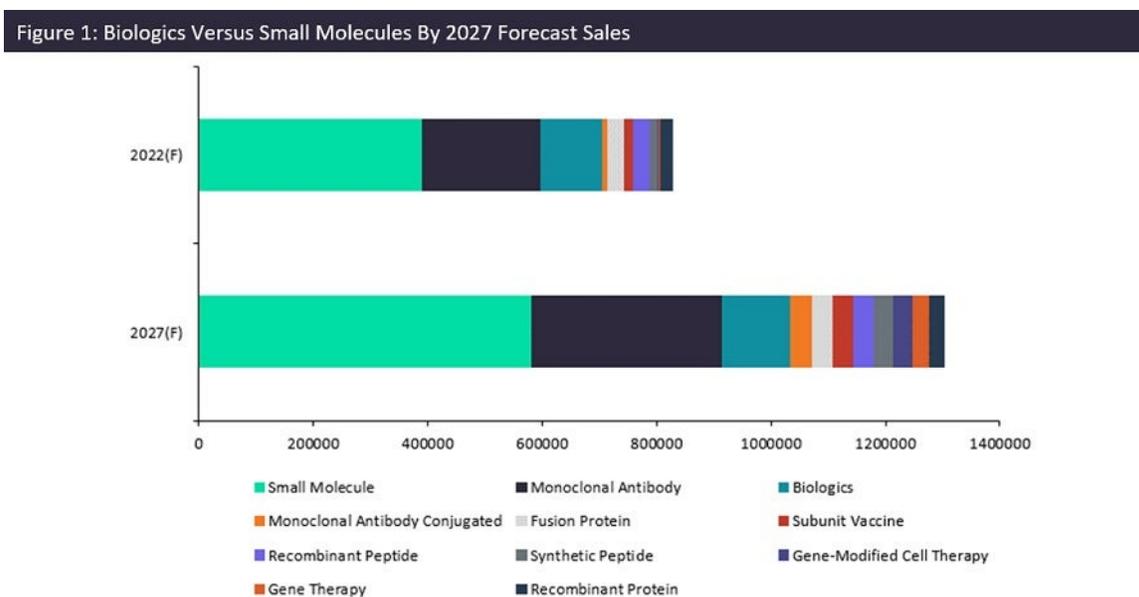
## 1.5 Biologics for Cancer Therapeutics

Cancer is the leading cause of death in midlife (40-65)<sup>47</sup>, and traditional chemotherapy relies on chemical agents that preferentially target rapidly dividing cells, like tumors. While chemotherapy is certainly powerful at killing cancer cells, it also non-discriminatorily kills dividing cells in healthy tissue, causing deleterious side effects in cancer patients. Biological therapeutics (biologics), like monoclonal antibodies (mAbs), utilize their discriminatory binding capacity for specific overexpressed antigens to target cancer while sparing healthy cells. Through the synergy of mAbs and chemotherapy, combination therapy with both could lower the dosage of chemotherapy necessary to eradicate the tumor of interest. When antibodies bind their target, ideally, they neutralize the specific overexpressed pathway of interest that is sustaining unchecked cell proliferation. As I have just shown, our group has mimicked the therapeutic potential of a traditional immunoglobulin G (IgG) mAb by bringing together two fusion proteins that present trastuzumab Fab domains. However, in addition to two Fab domains, IgGs also possess a fragment crystallizable (Fc) region, which can bind the Fc receptor present on effector cells, such as macrophages and natural killer (NK) cells, inducing antibody-dependent cellular phagocytosis and cytotoxicity, respectively. Furthermore, the Fc domain can also bind C1q and undergo classical complement dependent cytotoxicity<sup>48</sup>. It is by these two pathways that antibodies largely induce cell death *in vivo* (**Figure 1.7**).



**Figure 1.7** **A)** The general structure of a monoclonal antibody (mAb), which possess two Fab domains, composed of the variable and constant light chain region ( $V_L$  and  $C_L$ ) and the variable and first half constant heavy chain ( $V_H$  and  $C_{H1}$ ). The Fc region is composed of the second half of the heavy chain constant region ( $C_{H2}$  and  $C_{H3}$ ). **B)** mAbs can directly neutralize receptors and molecules via their Fab domains, which can induce cell death directly, as well as recruit immune cells and the complement cascade to indirectly induce cell death.<sup>48</sup>

While our group could interrogate the effects of modulating the Fc domain within our antibody scaffolds – especially since the Fc domain typically confers extended half-life *in vivo*<sup>48</sup>, for the sake of this dissertation, I mainly focus on modulating the Fab binding regions *in vitro*. The direct responses of antibody binding – receptor neutralization and the subsequent pathway inhibition – depend on the biochemistry of the Fab domains, and there is still much disputed about even well-documented mAbs like trastuzumab<sup>49</sup>. These protein-protein interactions have variable affinity and specificity for the three-dimensional epitope the CDRs bind, which can be affected by the avidity (# of domains) within the antibody (two for an IgG) as well as the number of receptors present on a cell surface. Further interrogating these structure-function relationships of antibody mimics with the utility of the megamolecule assembly platform is desirable, especially since biologics are poised to overtake small molecules as the leading pharmaceuticals by 2027<sup>50</sup> (**Figure 1.8**). Within this cohort, there exist mAbs, antibody-drug conjugates (ADCs) – which utilize mAb backbones to deliver cytotoxic agents – and novel therapies like chimeric antigen receptor-T cell (CAR-T) therapy.



**Figure 1.8** Current and projected sales of small molecules and various biologics<sup>50</sup>

## 1.6 Biologics for Protein Sensors / Diagnostics

In addition to synthesizing biologics for therapeutic aims, we can also utilize our protein assembly strategy for developing novel sensors/switches. Dynamic and reversible control of a protein's quaternary structure is a hallmark of many endogenous biological processes, like signal transduction<sup>51,52</sup> and generation of mechanical force<sup>53,54</sup>, both of which cascade from the macromolecule's microenvironment. However, artificially synthesizing various protein assembly sensors through synthetic biology has remained a challenge, again both because of limitations of assembling higher-order structures (e.g. low yield, aggregation), or because non-covalent and/or non-specific interactions result in heterogeneous products that can lack a well-defined on/off switch state. With the megamolecules assembly platform, I synthesized various atomically-defined linear scaffolds with covalently-bound reactive functionalities to cutinase that could dynamically and reversibly cyclize upon addition or removal of select divalent transition metals. Characterization of these scaffolds was mostly accomplished through Förster resonance energy

transfer (FRET). Even though the work highlighted in this dissertation yielded a well-characterized, proof-of-concept for a megamolecule sensor, it certainly sets the stage for potential synthesis of a diagnostic that can detect more clinically relevant analytes of interest due to the flexibility, modularity, and precision of the platform.

## 1.7 Overview of Dissertation

The majority of this background has thus far focused on elucidating the megamolecule assembly strategy to probe the structure-function relationships of rational biologic design, which will be the focus of the first part of my dissertation. This coincides with my most recent, and in my opinion, my impactful first-author work I have done in graduate school.

Initially, I expanded upon the work established by Justin Modica in his 2020 *JACS* paper – as well as a 2021 paper I co-authored in *Bioconjugate Chemistry* – which established megamolecules as a viable alternative to rational synthesis of antibody mimics. Here, I created a library of modular biologics to decouple the relationships between scaffold affinity, avidity, antibody-receptor internalization, and downstream inhibition of cancer cell proliferation. In the process, I created a standardized methodology via confocal microscopy for simply and continuously quantifying antibody internalization through labeling of primary lysines on the surface of proteins. This is a stark contrast to many other internalization studies, which have used either fusion proteins<sup>55</sup> or multi-step surface quenching<sup>56</sup> to reliably elucidate internalized antibodies at a few discrete time points. While probing the effects of valency, affinity, and specificity of HER2-targeting megamolecule-based antibodies, I successfully decoupled how these properties affect binding, internalization, and inhibition of cell proliferation. While this manuscript

has been written and is in preparation for submission, there are a few more experiments I plan on doing before I graduate that may make the final paper that are not present in this dissertation.

Next, I will present data of a project that establishes megamolecules as a viable platform to synthesize quaternary-scale protein switches, which I briefly referred to in **Chapter 1** as a focus of some *de novo* protein engineering efforts. Here, with co-first author Justin Modica, we assembled linear megamolecule fusions terminated with a fluorescent protein FRET pair, mNeonGreen and mTurquoise2. Upon addition of a cutinase-reactive terpyridine small molecule at each end, I demonstrated how the FRET pair cyclizes upon addition of a bivalent transition metal ion (e.g.,  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ), which coordinates with two terpyridine groups, and leads to an increase in mNeonGreen emission from local mTurquoise2 emission. I also demonstrated how cyclization is reversible, as the FRET efficiency drops back down to steady state with addition of excess metal-sequestering EDTA in a metal-dependent context. At the time of writing this dissertation, this manuscript is also in preparation to be submitted – we’re thinking at swinging for the fences at *Nature Chemistry*.

In the second part of my dissertation, I will switch from megamolecules work and present a first-authored project from the first half of my PhD. Here, I utilized self-assembled monolayer (SAM) surface chemistry to selectively pattern cancer and non-cancer cells into discrete shapes. Here, the cells were stained with a rhodamine dye to label their actin cytoskeletons, concurrently revealing their overall morphologies on a 2D surface. Collaborating with machine learning experts (the Bagheri Group, University of Washington), I took high-resolution (60X) confocal images of single cells at their basal layer – either unpatterned or patterned – and created a reproducible algorithm that could successfully discriminate between single-cell images based on their actin

cytoskeleton and cell morphology. This project was published in *Scientific Reports* in 2021 and was well-received as a talk to the Department of Cell & Developmental Biology in March 2022.

Finally, in the last part of my dissertation, I will present work from the first year or two of my PhD. This includes co-authored work published in 2019 in *Chem. Eur. J.* Here, I collaborated with another graduate student, Lindsey Pearson (Szymczak), to probe the differences in bulk phosphatase activity and specificity across various cancer and non-cancer cell lines. Here, phosphatase activity in cell lysate de-phosphorylated phosphorylated serine, threonine, and tyrosine peptide libraries presented on self-assembled monolayer (SAM) surfaces. We found that the activity of phosphatases towards phosphor-threonine peptides was much greater than phosphoserine peptides across five mammalian cell lines, despite similar specificities. These conclusions may help elucidate why phosphoserine is so much more abundant in the phosphoproteome than phosphor-threonine, as the majority of phosphorylation research focused on the activity of kinases. This was a concise study that provided necessary breadth to my PhD and gave me an initial project to complete with the laboratory skills I had possessed *a priori*. I will then conclude with a small, unpublished project that similarly utilizes SAMs to present specific ligands for interaction by cells with work I did over my first summer as a full-time researcher before discussing potential follow-up projects for a new student or postdoc in **Chapter 6**.

I want to emphasize how much I have enjoyed having a wide breadth of experience within my PhD. I did not foresee how joining Milan's group at such a scientific inflection point from SAMDI to megamolecules would allow me to engage and creatively apply my skillset acquired from my time at the NIH in cell biology, confocal microscopy, and image analysis across various disciplines in surface chemistry, cell biology, organic chemistry, synthetic biology, and protein engineering. If you've made it this far, thanks for reading.

## Chapter 2

### Megamolecules to Interrogate and Decouple Binding Affinities, Internalization Rates, and Inhibition of Proliferation for HER2-Targeting Cancer Therapy

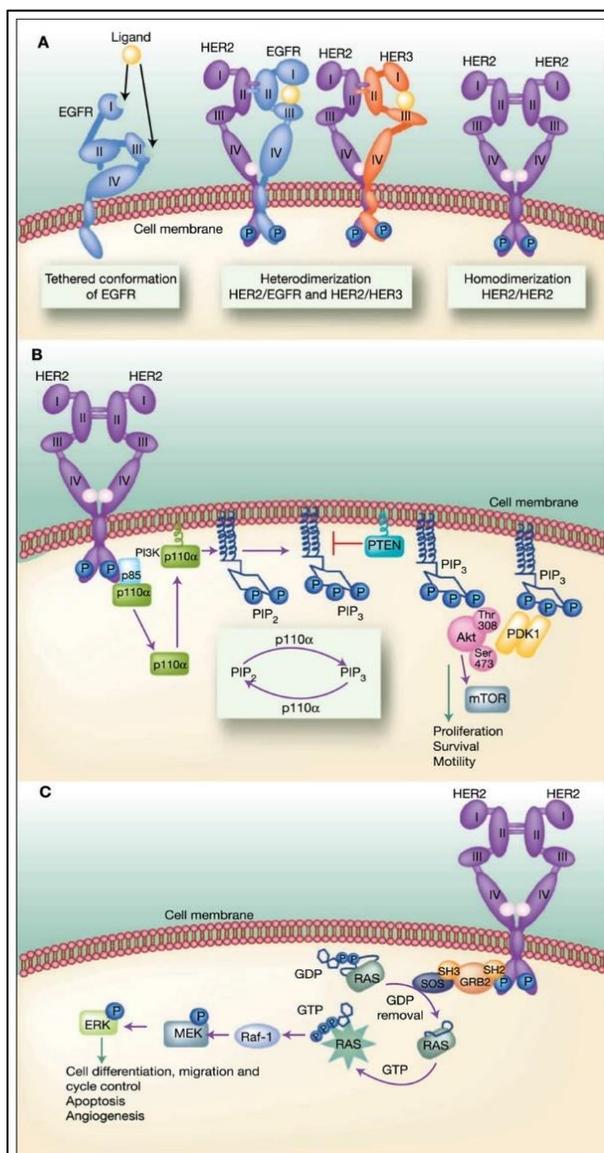
Research and figures presented in this chapter are adapted from work published (or soon to be published) in these two papers:

**Sykora, D. J.\***, Sridhar, S., Modica, J. A., Gu, Z. & Mrksich, M. Internalization kinetics of modular antibody scaffolds. *In Preparation* (2022).

\*Metcalf, K. J., \*Kimmel, B. R., **Sykora, D. J.**, Modica, J. A., Parker, K., Berens, E., Dai, R., Dravid, V., Werb, Z. & Mrksich, M. Synthetic tuning of domain stoichiometry in nanobody-enzyme megamolecules. *Bioconjugate Chemistry* **32**, 143-152 (2021).

#### 2.1 Introduction

Cancer is the leading cause of death in midlife<sup>57</sup>, and development of monoclonal antibodies (mAbs) against overexpressed receptors on tumors has enabled effective targeting and downstream treatment of cancer. The most common cancer characteristic is sustained cell proliferation, which is often induced by receptor tyrosine kinase (RTK) dysregulation. Arguably the most common RTK is *ERBB2*, encoding for human epidermal growth factor receptor 2 (HER2), which is known to play a critical role in many aggressive types of breast (15-30%) and gastric/gastroesophageal cancers (10-30%)<sup>58</sup>. HER2's intracellular tyrosine kinase signaling is induced by heterodimerization to other *ERBB* family members (EGFR, HER3, HER4) in either a ligand-dependent or ligand-independent manner as well as homodimerization to other HER2 receptors, of which there is no natural ligand (**Figure 2.1**)<sup>59</sup>.



**Figure 2.1** HER2 activation. HER2, with domains shown, has no natural ligand, and needs to dimerize to activate its oncogenic pathways. Here, it can dimerize with EGFR, HER2, HER3, and HER4 (not shown) with EGFR, HER3, and HER4 being either ligand-induced or ligand-independent dimerization<sup>72</sup>.

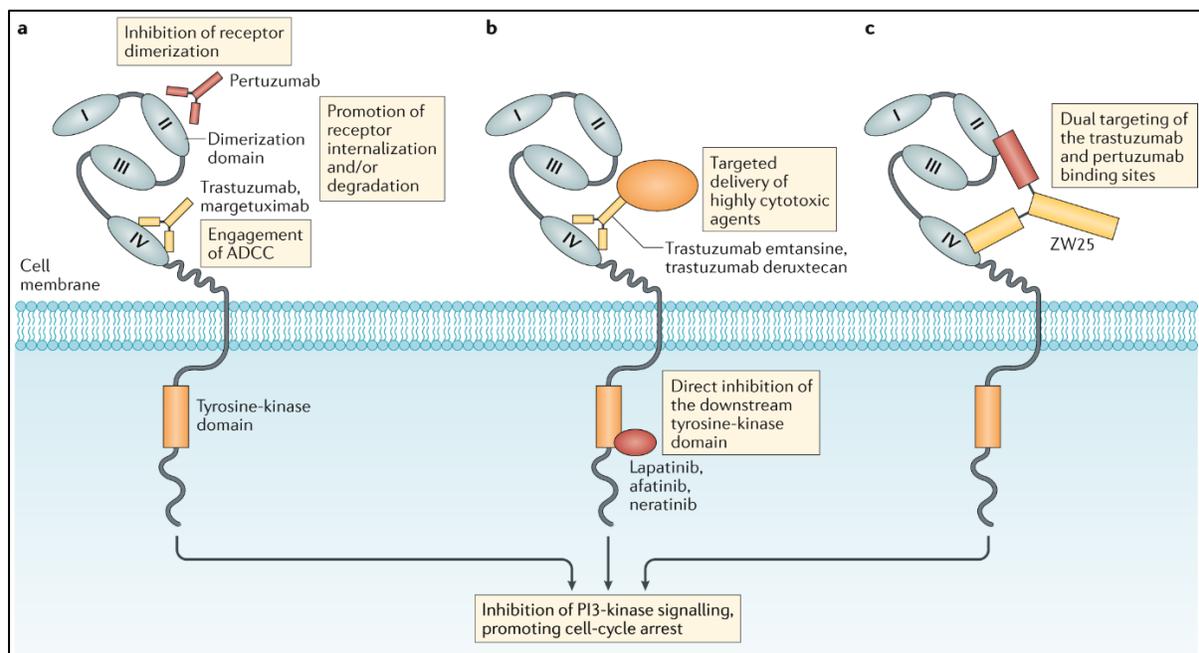
The five approved mAbs against HER2 are trastuzumab (TZB), margetuximab (MXB), pertuzumab (PZB), as well as antibody-drug conjugates (ADCs) of TZB with the cytotoxic warhead emtansine (Kadcyla®, T-DM1) and deruxtecan (Enhertu®, T-DXd). TZB (and MXB, which possess the same Fab domains as TZB) blocks dimerization of HER2 by neutralizing

domain IV (D4) proximal to the cell membrane<sup>60</sup>. PZB, on the other hand, blocks dimerization of HER2 by neutralizing domain II (D2) on the dimerization arm<sup>61</sup>. Due to the complementary mechanisms of action, co-administration of TZB and PZB has shown synergy in the clinic<sup>62</sup>. However, receptor neutralization, receptor-antibody internalization and degradation, and the associated dimerization inhibition has only proven to be transiently efficacious and can often lead to antibody resistance. This has led to the development of potent ADCs, like T-DM1 and T-DXd, which require robust cellular internalization by the target cell to release an apoptosis-inducing drug (**Figure 2.2**). However, cancer cells may still develop therapeutic resistance to ADCs<sup>63</sup>, and some have suggested that Kadcyła® specifically fails to reveal any improvements in patient outcomes<sup>64</sup>. Due to the litany of HER2-dependent cancers<sup>65</sup> most prevalent in women under 40<sup>66</sup> and the inconsistent treatment outcomes of existing biologics, HER2 remains a well-studied clinical target for improved antibody engineering efforts.

As I discussed in **Chapter 1**, Justin Modica demonstrated that our modular megamolecule assembly platform can recapitulate the therapeutic activity of trastuzumab by fusing the trastuzumab Fab to the enzymes cutinase or SnapTag and reacting them rapidly, selectively, and covalently with a heterobifunctional, ethylene glycol spacer (**Figure 1.6**)<sup>67</sup>. As that study focused on bivalent scaffolds, it remained to be seen how broadly modulating scaffold shape, valency, and specificity impacted scaffold binding, internalization, and inhibition of cell proliferation.

As you will read in **Chapter 2**, I used the megamolecule platform to prepare bivalent, trivalent, and bispecific (biparatopic) scaffolds that target HER2 and compared their efficacy against the parent trastuzumab mAb, its Fc-cleaved F(ab')<sub>2</sub> structure, and each monomeric fusion protein. Initially, I anticipated that increasing the valency of an antibody scaffold from two to three would show an additive effect towards not only binding affinity via increased avidity but would

also improve internalization of the receptor and the downstream inhibition of cell proliferation. Yet, while trivalent scaffolds did modestly improve binding, they did not demonstrate increased net receptor internalization rate nor growth inhibition *in vitro* towards HER2-overexpressing BT474 cells. Rather, biparatopic scaffolds that targeted both D1 and D4 showed a ~3-fold increase in internalization rate. Furthermore, growth inhibition of BT474 cells was only achieved with a monoparatopic scaffold presenting the trastuzumab Fab; in fact, despite the increased internalization rate achieved with the biparatopic scaffolds presenting the trastuzumab Fab, the addition of the D1-targeting nanobody 2Rs15d abrogated the inhibition of cell proliferation in all cases. These results demonstrate the power of how a modular biologic assembly platform can non-intuitively decouple properties that affect overall antibody efficacy.



**Figure 2.2** Schematic of HER2 and the regions in which various drugs target the receptor **A)** naked mAbs, pertuzumab, trastuzumab, and margetuximab target D2 and D4, **B)** ADCs target D4 by using trastuzumab as a backbone for the cytotoxic drug and small molecule inhibitors target the intracellular tyrosine kinase domain, and **C)** the experimental biparatopic mAb, ZW25 engages both D2 and D4, potentially mimicking the therapeutic synergy seen in clinic when co-administrating pertuzumab and trastuzumab<sup>65</sup>.

## 2.2 Results

I first prepared megamolecule antibody scaffolds as demonstrated previously<sup>68,69</sup> to investigate how increasing valency impacted efficacy. First, *E. coli* optimized fusion proteins presenting either the trastuzumab Fab fused to cutinase via the C<sub>H</sub> domain (**Figure 2.3A**) or via the V<sub>H</sub> domain (**Figure 2.3B**) were reacted with homobifunctional or homotrifunctional linkers overnight at room temperature and purified via size exclusion chromatography (SEC). Similarly, a previously used<sup>69</sup> C-terminal fusion of cutinase to the nanobody 5F7 (also specific for D4)<sup>70</sup> was reacted with the same homobifunctional and homotrifunctional linkers (**Figure 2.3C**) and are termed 1-5F7, 2-5F7, and 3-5F7. All monomeric fusion proteins and megamolecule scaffolds were compared against the mAb trastuzumab as well as its pepsin-cleaved F(ab')<sub>2</sub> (**Figure 2.3D**) since our structures also lacked an Fc domain. Faster protein scaffold elution correlated with an increase in scaffold size, and analytical FPLC on select scaffolds demonstrated globularity (**Figure 2.4**). Furthermore, select electron spray ionization mass spectrometry (ESI-MS) data revealed excellent agreement with predicted protein fusion and megamolecule scaffold sizes (**Figure 2.5**).

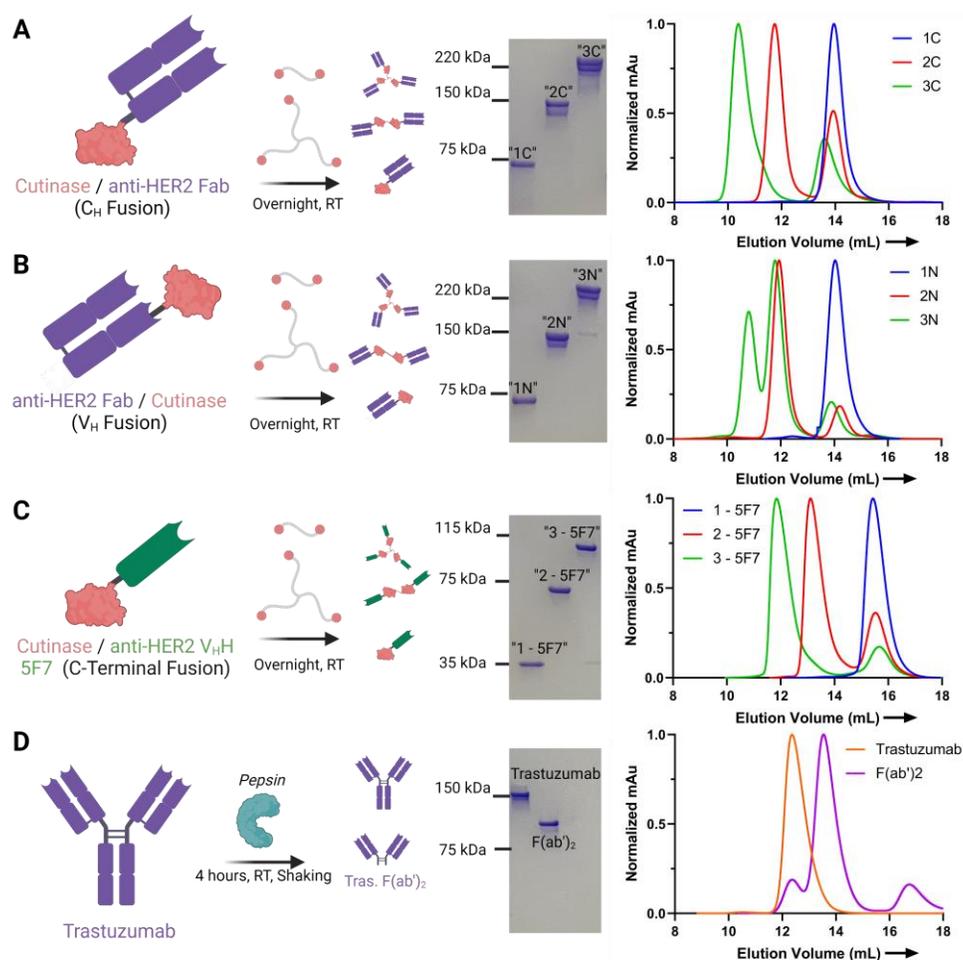
Initially, to investigate how increasing scaffold valency impacted binding affinity, I quantified scaffold specificity to HER2 using confocal microscopy for HER2<sup>high</sup> BT474 human breast ductal carcinoma cells, HER2<sup>medium</sup> MDA-MB-453 human breast mammary carcinoma cells, and HER2<sup>low</sup> MDA-MB-231 human breast mammary adenocarcinoma cells<sup>71</sup>. The scaffolds highly labeled the BT474 cells, modestly labeled the MDA-MB-453 cells, and did not label the MDA-MB-231 cells, which was expected based on previous work in our group<sup>68,69</sup> (select scaffolds in **Figure 2.6A**). Further analysis of cell binding by flow cytometry against BT474 cells demonstrated a diminishing return on binding affinity from two to three domains (**Figure 2.6B**). Flow cytometry-based apparent binding affinities for monoparatopic scaffolds can be seen in

**Table 2.1.** For megamolecules formed from C-terminal trastuzumab fusions, I observed that overall scaffold affinity correlated strongly with that of the parent trastuzumab mAb as well as its  $F(ab')_2$ , demonstrating that our fusions do not significantly abrogate binding of the native Fab. However, megamolecules formed from N-terminal fusions of trastuzumab had somewhat lower binding affinities, such that the trivalent N-terminal fusion (i.e., 3N) had a similar effective  $EC_{50}$  ( $\sim 7.2$ nM) as the monovalent C-terminal fusion (i.e. 1C). As the N-terminus of trastuzumab is the  $V_H$  domain and the C-terminus is the  $C_H$  domain, these results show that fusing cutinase close to the binding CDRs of the Fab adversely impacts binding, which affirms what Justin had demonstrated previously with only bivalent scaffolds.

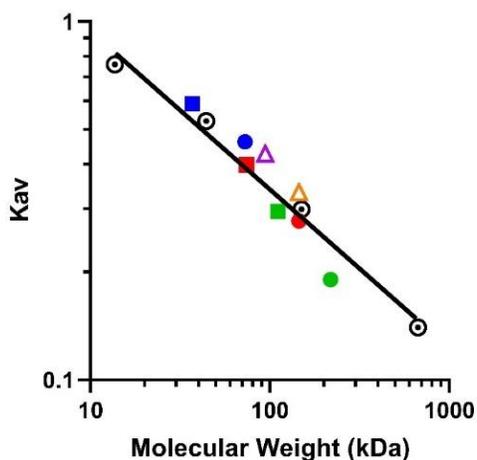
The apparent cell-binding  $EC_{50}$  values for trastuzumab and its  $F(ab')_2$  (3.8nM and 2.2nM, respectively) correlate with previously reported values ( $\sim 1.6$ nM and  $\sim 1.2$ nM, respectively), and the apparent  $EC_{50}$  for the  $C_H$  monomeric fusion (i.e. 1C, 6.8nM) correlates with what has been reported for the monovalent Fab ( $\sim 5.7$ nM)<sup>73</sup>. The nanobody 5F7 fusions showed the largest avidity-based decrease in apparent binding  $EC_{50}$ , where the bivalent 5F7 showed a  $\sim 5$ -fold increase in affinity. Again, the trivalent scaffold showed only a modest affinity increase.

Next, I asked how the rate of antibody-scaffold internalization correlated with valency and binding affinity. Again, robust antibody-scaffold internalization and degradation is thought to be the mechanism to which antibody binding mitigates proliferation and can induce cytotoxicity, not to mention the mechanism by which ADCs release cytotoxic drug cargo. To quantitatively monitor internalization, megamolecules were reacted with a pH sensitive dye, pHAb<sup>TM</sup> (Promega), which utilizes an NHS-ester handle to react with primary lysines on the surface of proteins. After excess dye was filtered out, labeled megamolecule scaffolds were purified again via SEC (**Figure 2.7A**). As scaffolds undergo receptor-mediated endocytosis, they generally get trafficked from the early

endosome (pH ~6.5) either back to the surface for receptor recycling or to the late endosome (pH ~5.5). Once late endosomes fully mature into the lysosomes (pH ~4.5), the receptor-antibody complex is degraded<sup>74</sup>. It is often by this mechanism that HER2 is degraded and downregulated by targeting antibodies. This rhodamine-based dye shows negligible fluorescence at neutral pH (7.0-7.4) and significantly increases as the early endosome matures into the late endosome<sup>75</sup> (**Figure 2.7B**).



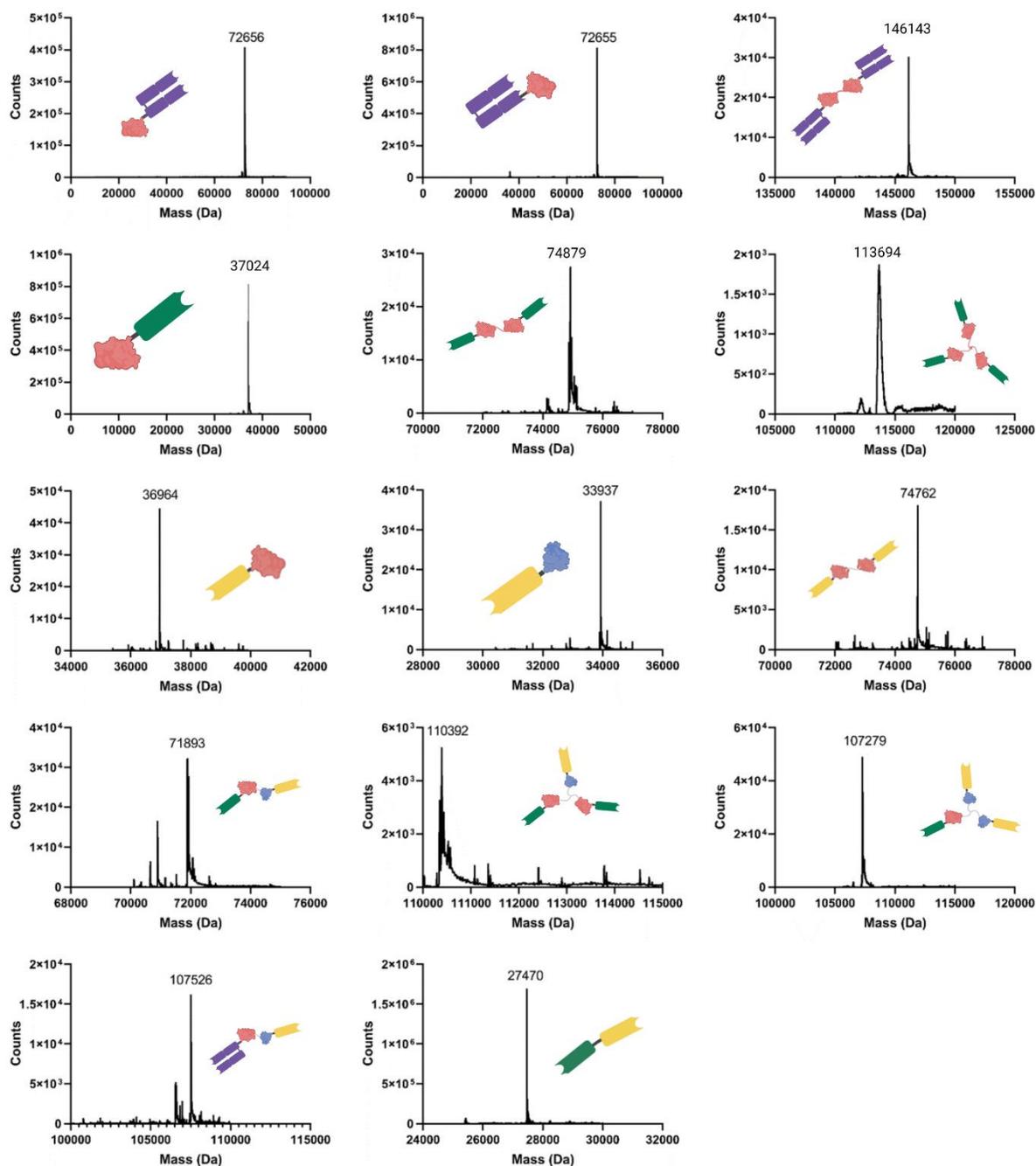
**Figure 2.3 Synthesis of Megamolecule Antibodies.** **A)** Cutinase fusions through the heavy chain C-terminus of the trastuzumab Fab, **B)** Cutinase fusions through the heavy chain N-terminus of the trastuzumab Fab, closer to CDRs, **C)** N-terminal cutinase fusions with HER2-binding nanobody 5F7, **D)** Preparation of trastuzumab's F(ab')<sub>2</sub> via pepsin cleavage. Shown are non-reducing SDS-PAGE gels and size-exclusion chromatography (SEC) purification traces to demonstrate the linear increase in size from monomer (e.g. 1C) to dimer (e.g. 2C) to trimer (e.g. 3C).



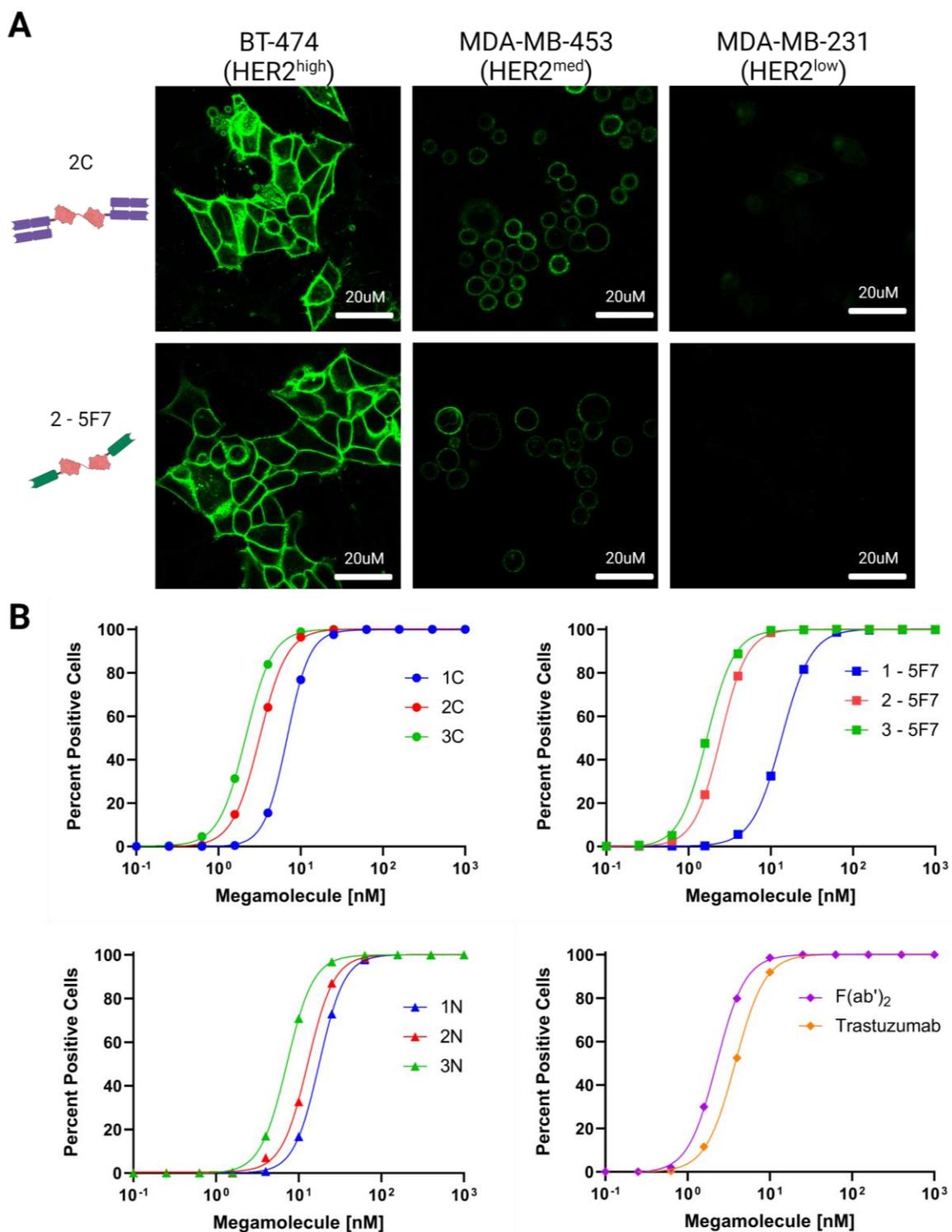
**Figure 2.4 Analytical SEC Demonstrates Globularity.** Here, 1C (Blue Circle), 2C (Red Circle), 3C (Green Circle), 1 – 5F7 (Blue Square), 2 – 5F7 (Red Square), 3 – 5F7 (Green Square), Trastuzumab (Orange Triangle), and Trastuzumab F(ab')<sub>2</sub> (Purple Triangle) demonstrate no intermolecular dimerization. Shorter column retention times (thus, larger proteins) are indicated by smaller values of  $K_{av}$ .

Using an automated fluorescent microscope, internalization was quantitatively monitored for each of my structures at 5nM over a 48-hour period (**Figure 2.7C & 2.7D**). At 5nM, I found that the signal intensity was bright enough to quantify internalization of the monovalent fusions, where lower concentrations were too dim. Conversely, concentrations above 20nM led to high background noise, which ultimately reduced signal after processing. The intensity profiles at 5nM could generally be modeled by a first-order association curve, although the absolute mean pixel intensity (MPI) of each structure directly correlated with the degree of labeling of dye as well as the amount of antibody internalized. No internalization was seen in cell lines that lacked HER2 overexpression (**Figure 2.8**), and internalization was massively reduced when scaffolds were co-cultured with a small molecule inhibitor of clathrin (100 $\mu$ M Pitstop2®, **Figure 2.9**). Further, a HER2-targeting nanobody with low affinity (C8<sup>76</sup>), only showed detectable internalization within a trimer scaffold at 50nM amidst high background (**Figure 2.10**). Taken together, this

demonstrates that all antibody scaffolds undergo receptor-mediated endocytosis. Observed internalization rates can be seen in **Table 2.2**.



**Figure 2.5** Select fusion proteins and megamolecules show strong agreement between theoretical and predicted masses via electron spray ionization mass spectrometry (ESI-MS). Key: red globular protein = cutinase; blue globular protein = SnapTag; green V<sub>H</sub> domain = 5F7; yellow V<sub>H</sub> domain = 2Rs15d; purple Fab = trastuzumab Fab. Masses above 150 kDa were generally noisy and were omitted from this figure.



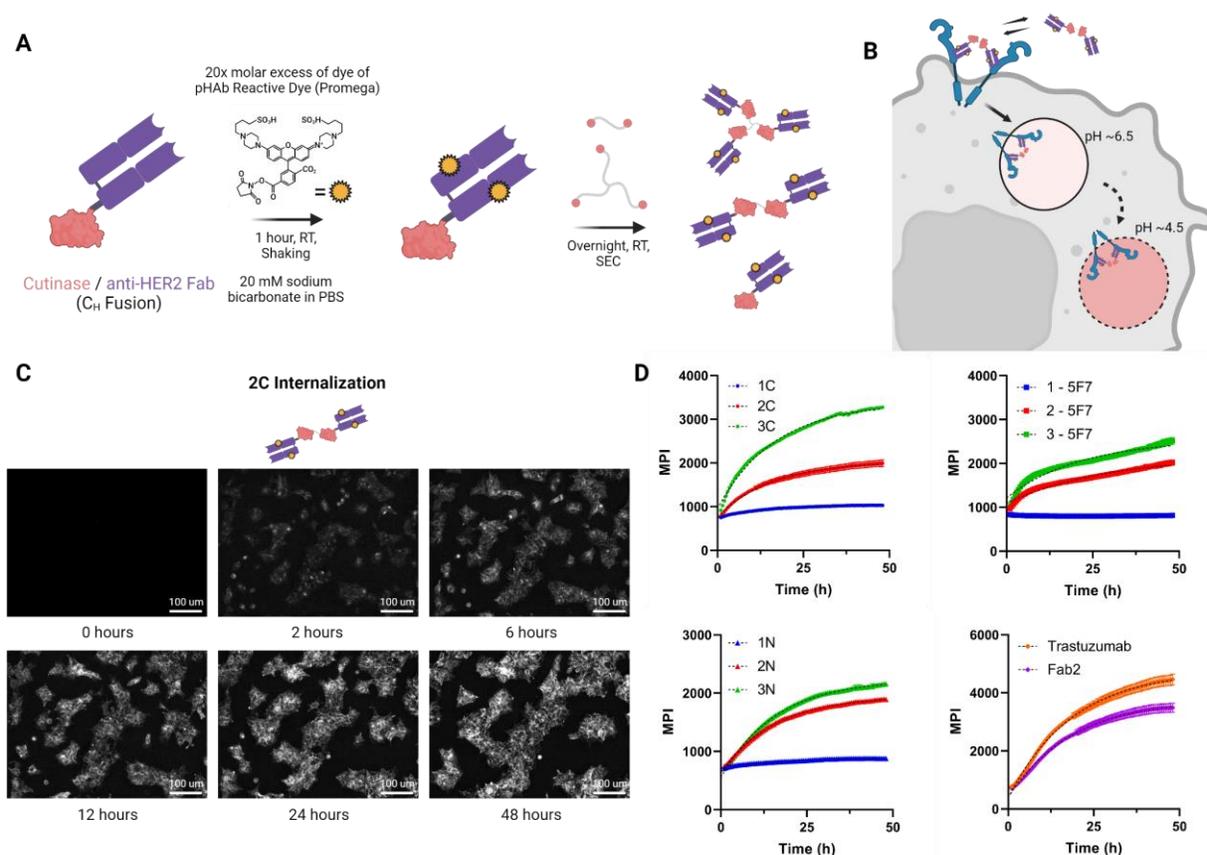
**Figure 2.6 Megamolecules Selectively Bind HER2.** A) Confocal images showing examples of megamolecules binding HER2-positive cells with increasing intensity based on HER2 expression levels, B) HER2 binding curves on BT474 cells (percent positive) via flow cytometry. Error bars are in SEM, though most are too small to be seen.

**Table 2.1** Observed Binding EC<sub>50</sub> Values of Monoparatopic HER2 Targeting Scaffolds via Flow Cytometry on BT474s

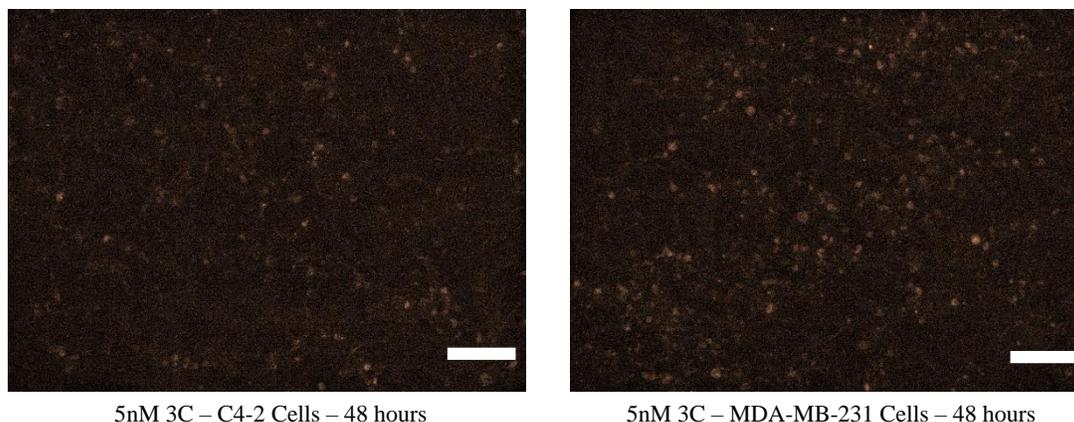
Scaffold	% Positive EC <sub>50</sub> (nM)
Trastuzumab	3.8 ± 0.1
Trastuzumab F(ab') <sub>2</sub>	2.2 ± 0.1
1C	6.8 ± 0.1
2C	3.2 ± 0.1
3C	2.1 ± 0.1
1N	17.7 ± 0.2
2N	12.9 ± 0.6
3N	7.2 ± 0.1
1 – 5F7	13.6 ± 0.3
2 – 5F7	2.5 ± 0.1
3 – 5F7	1.7 ± 0.1

In general, all of the Fab-based megamolecule scaffolds internalized at about the *same observed rate*,  $k_c$  (~0.06-0.07 h<sup>-1</sup>), with the exception of monomeric V<sub>H</sub> fusion (i.e. 1N), whose slower internalization rate may correlate to its high effective binding EC<sub>50</sub> (almost 20nM), which was much higher than the 5nM concentration used in this experiment. As expected, trastuzumab and its F(ab')<sub>2</sub> internalized at approximately the same rate (~0.053-0.054 h<sup>-1</sup>), matching an observed rate published by K. Dane Wittrup's group (~0.054 h<sup>-1</sup>)<sup>77</sup>. Interestingly, the nanobody 5F7 megamolecules did not internalize as quickly and may not necessarily follow a first-order rate constant (denoted by an asterisk[\*] in **Table 2.2**) despite having low nanomolar binding affinity (**Figure 2.6B**; **Tables 2.1 & 2.2**). Furthermore, the monomeric fusion protein (1-5F7) signal was just above the background, showing very little internalization after 48 hours. I compared dissociation rates ( $k_{off}$ ) of 2C to 2 – 5F7 and did not see any difference between the Fab and nanobody; in fact, both remained strongly bound even after 24 hours (**Figure 2.11**), which I had only previously found with bivalent scaffolds in our *Bioconjugate Chemistry* publication (**Figure 2.12**)<sup>69</sup>, suggesting little difference in the observed off rate ( $k_{off}$ ) of the megamolecule.

Interestingly, others have reported examples of non-Fab binding proteins (e.g. DARPin) against HER2 demonstrating tight binding and robust inhibition of proliferation without significant internalization<sup>78,79</sup>.



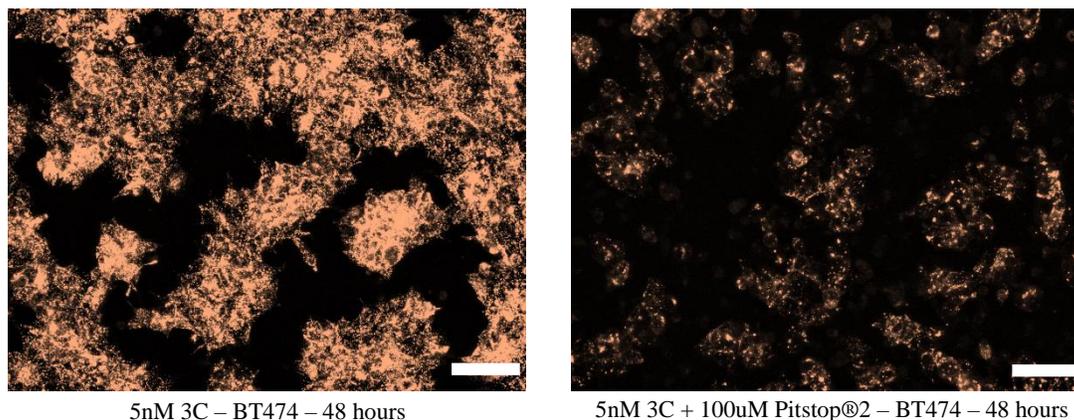
**Figure 2.7 Internalization of Megamolecule Scaffolds.** **A)** Labeling scheme with pHAb, a pH-sensitive dye from Promega; each monomeric fusion protein is labeled with 1-2 dyes **B)** A scheme of antibody-receptor internalization and how fluorescent signal increases with lysosome trafficking, **C)** Timelapse of internalization of 5nM 2C on BT474 cells over 48 hours, **D)** Net internalization curves for each scaffold (5nM) as well as trastuzumab and the trastuzumab F(ab')<sub>2</sub>. Error in MPI is SEM.



5nM 3C – C4-2 Cells – 48 hours

5nM 3C – MDA-MB-231 Cells – 48 hours

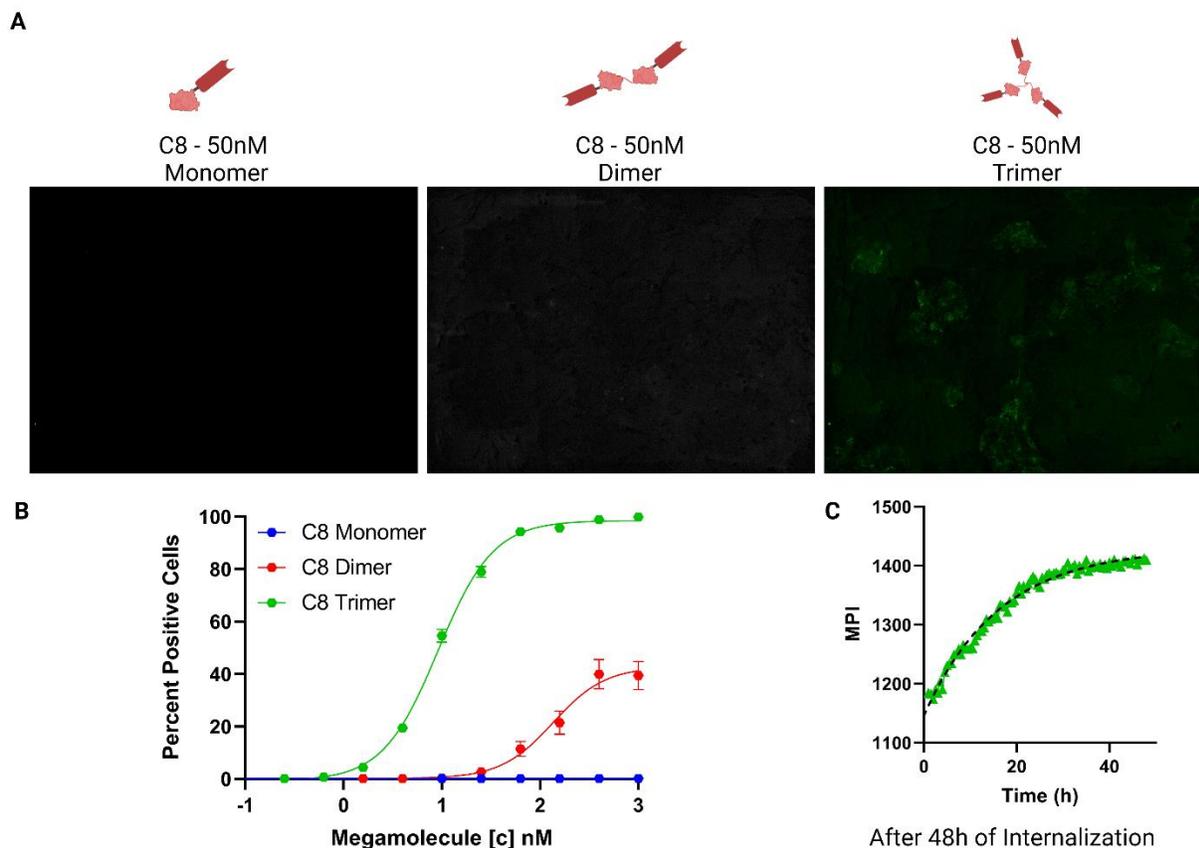
**Figure 2.8 Antibody Internalization Negligible in HER2<sup>-/-</sup> Cells.** After 48h of addition of 5nM construct 3C in culture, there is negligible internalization of scaffold into cells that do not overexpress HER2. Scale bar = 100um.



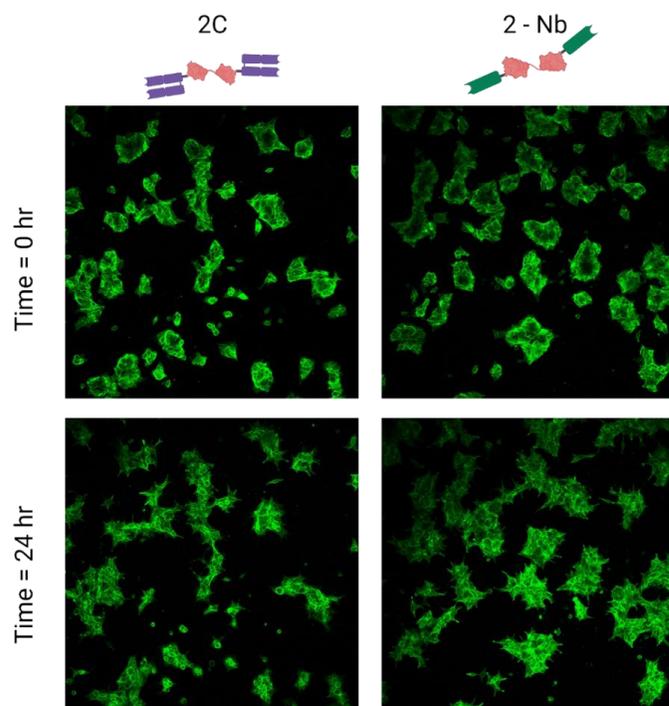
5nM 3C – BT474 – 48 hours

5nM 3C + 100uM Pitstop@2 – BT474 – 48 hours

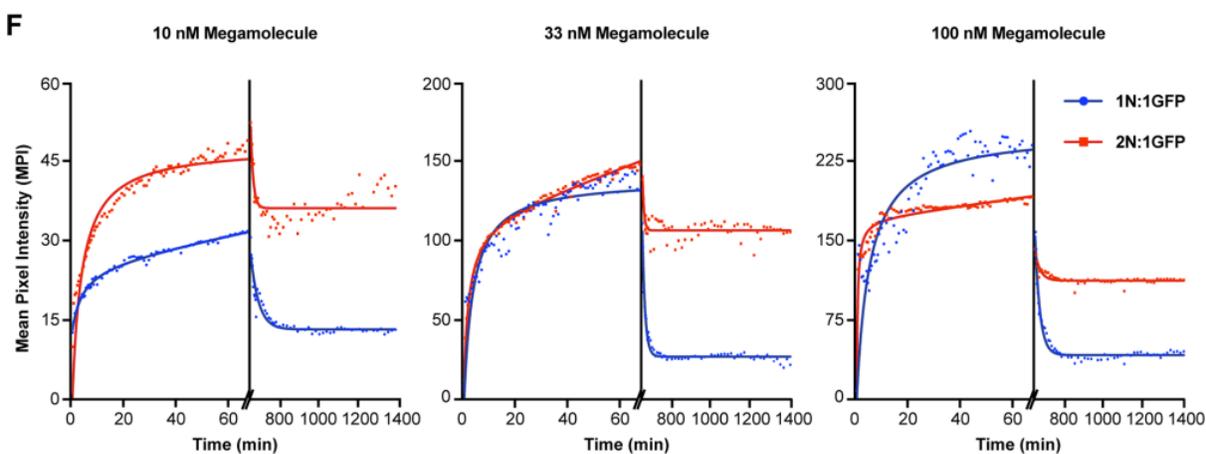
**Figure 2.9 Antibody Internalization Inhibited by Clathrin Inhibitor.** After 48h of addition of 5nM construct 3C in culture, there is clear inhibition of internalization with addition of 100uM Pitstop@2, a small molecule inhibitor of clathrin. Scale bar = 100um.



**Figure 2.10 Low affinity nanobody C8 does not internalize efficiently. A)** Monomeric, dimeric, and trimeric cutinase-C8 megamolecule constructs show negligible internalization into BT474 cells and only at high concentrations around 50nM does the trimeric scaffold achieve enough internalization signal to quantify. **B)** Binding curves from flow cytometry for the three scaffolds giving effective  $EC_{50}$  values around 136nM for the C8 dimer and 10nM for the trimer. The C8 monomer does not noticeably bind even at 1uM, which may be an artifact of post-staining wash steps before running flow cytometry **C)** Internalization curve for the C8 trimer at 50nM yields a rate around  $0.062 \pm 0.022 \text{ h}^{-1}$ .



**Figure 2.11 2C and 2-Nb Do Not Dissociate from Cells After 24 h.** After letting both 2C and 2 - Nb bind BT474 cells for two hours, cells were washed 3x with PBS and then re-cultured with fresh media, letting bound megamolecule dissociate for 24 h. Neither show significant unbinding, suggesting strong avidity effects for both, and the lack of rapid internalization seen in the 5F7 constructs do not depend on the off-rate of megamolecule ( $k_{off}$ ).



**Figure 2.12 Bivalent megamolecule scaffolds show significantly slower dissociation than monovalent scaffolds.** Adapted from a co-authored manuscript<sup>63</sup>, the estimated dissociation constant ( $K_D$ ) from kinetic binding data for a bivalent 5F7 scaffold (i.e. 2N:1GFP) was 4-fold lower than monovalent (i.e. 1N:1GFP), about 13nM to 3nM, which compares very favorably to the data acquired via flow cytometry.

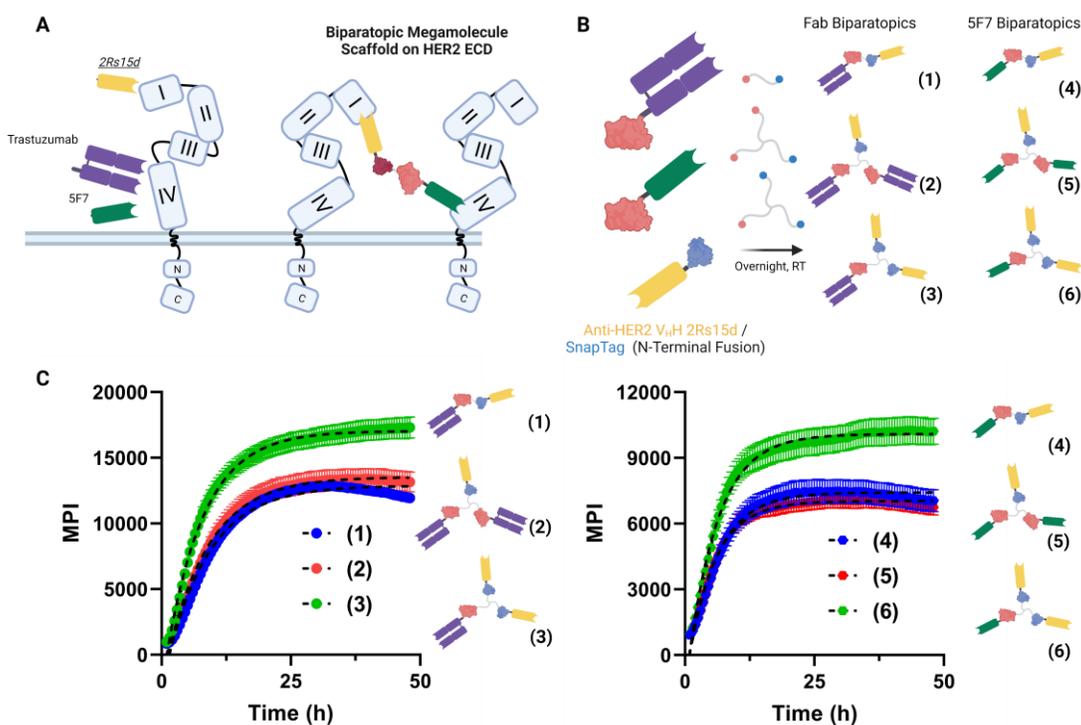
**Table 2.2** Observed Internalization Rates of HER2-Targeting Monoparatopic Scaffolds on BT474s

Scaffold	Internalization Rate, $k_e$ ( $\text{h}^{-1}$ )	Half-Time, $t_{1/2}$ (h)
Trastuzumab	$0.053 \pm 0.004$	~13
Trastuzumab F(ab') <sub>2</sub>	$0.054 \pm 0.004$	~12.9
1C	$0.072 \pm 0.008$	~9.6
2C	$0.064 \pm 0.007$	~10.9
3C	$0.061 \pm 0.003$	~11.5
1N	$0.050 \pm 0.005$	~13.9
2N	$0.064 \pm 0.003$	~10.8
3N	$0.062 \pm 0.003$	~11.1
1 – 5F7	N/A	
2 – 5F7*	$0.034 \pm 0.008$	~20.2
3 – 5F7*	$0.046 \pm 0.007$	~15

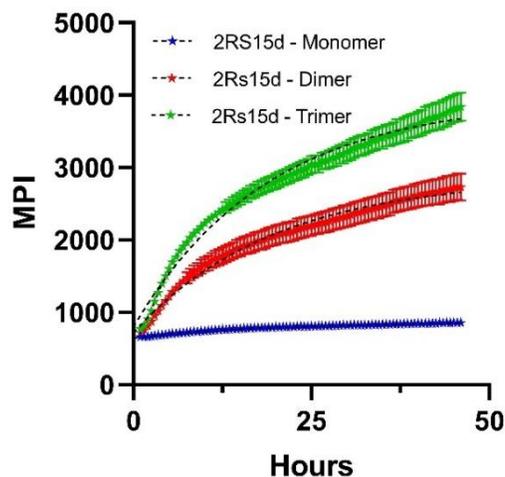
Motivated by recent publications describing the development of how biparatopic antibody scaffolds that target two different epitopes against HER2 lead to increased effective internalization<sup>80-82</sup>. I created scaffolds that presented the nanobody 2Rs15d, a nanobody selective for D1 on the HER2 ECD<sup>83-85</sup>. In this way, I aimed to validate the mechanism proposed, which stipulated that the substantial crosslinking of HER2 necessary to induce cytotoxicity is best induced by specifically crosslinking D1 and D4 (**Figure 2.13A**)<sup>78,82</sup>. Following the previous approach, a panel of biparatopic scaffolds presenting 2Rs15d with either 5F7 or 1C were synthesized, labeled, and purified (**Figure 2.13B**). Here, 2Rs15d was fused to the N-terminus of SnapTag, an enzyme in our megamolecule toolbox that reacts irreversibly with benzylguanine or benzylchloropyrimidine groups<sup>24</sup>. In this way, I could generate scaffolds in a one-pot reaction with a heterobifunctional or heteotrifunctional linker as long as the constituent fusions were added at stoichiometric equivalencies.

Importantly, I found significant increases in effective internalization rate across all scaffolds, which can be seen in **Figure 2.13C** and **Table 3**. The C-terminal trastuzumab Fab scaffolds demonstrated a 2-to-3 fold increase in effective internalization rate compared to

trastuzumab, and the 5F7 nanobody fusions demonstrated a 3-to-4 fold increase in effective internalization rate, which was even more striking considering the slow uptake of monoparatomic 5F7 megamolecules alone. This rate increase is consistent with a chimeric biparatomic IgG against the receptor CEA; the monoparatomic IgG had a half-time of around 13 hours (like trastuzumab) while the biparatomic IgG had a half-time of around 5 hours, reflecting a ~3-fold increase in effective internalization rate<sup>86</sup>. Again, I see no effective internalization rate increase from dimeric to trimeric scaffolds, further suggesting that two binding domains is sufficient when targeting extracellular receptors for internalization. Furthermore, control scaffolds composed of a 2Rs15d-ectinase dimers demonstrated a similarly slower uptake like 2-5F7 (**Table 2.3, Figure 2.14**), demonstrating that 2Rs15d alone does not possess enhanced net internalizing properties.



**Figure 2.13 Preparation of Biparatomic Scaffolds.** **A)** A schematic of intermolecular cross-linking of HER2 receptors by biparatomic megamolecule scaffolds, where both trastuzumab and the nanobody 5F7 target domain IV while 2Rs15d targets domain II, **B)** Utilizing megamolecule chemistries, dimeric and trimeric biparatomic scaffolds were synthesized with either trastuzumab Fab or 5F7 targeting D4, **C)** Internalization curves for each biparatomic scaffold after treating BT474s at 5nM for 48 hours. Error in mean pixel intensity (MPI) is SEM.



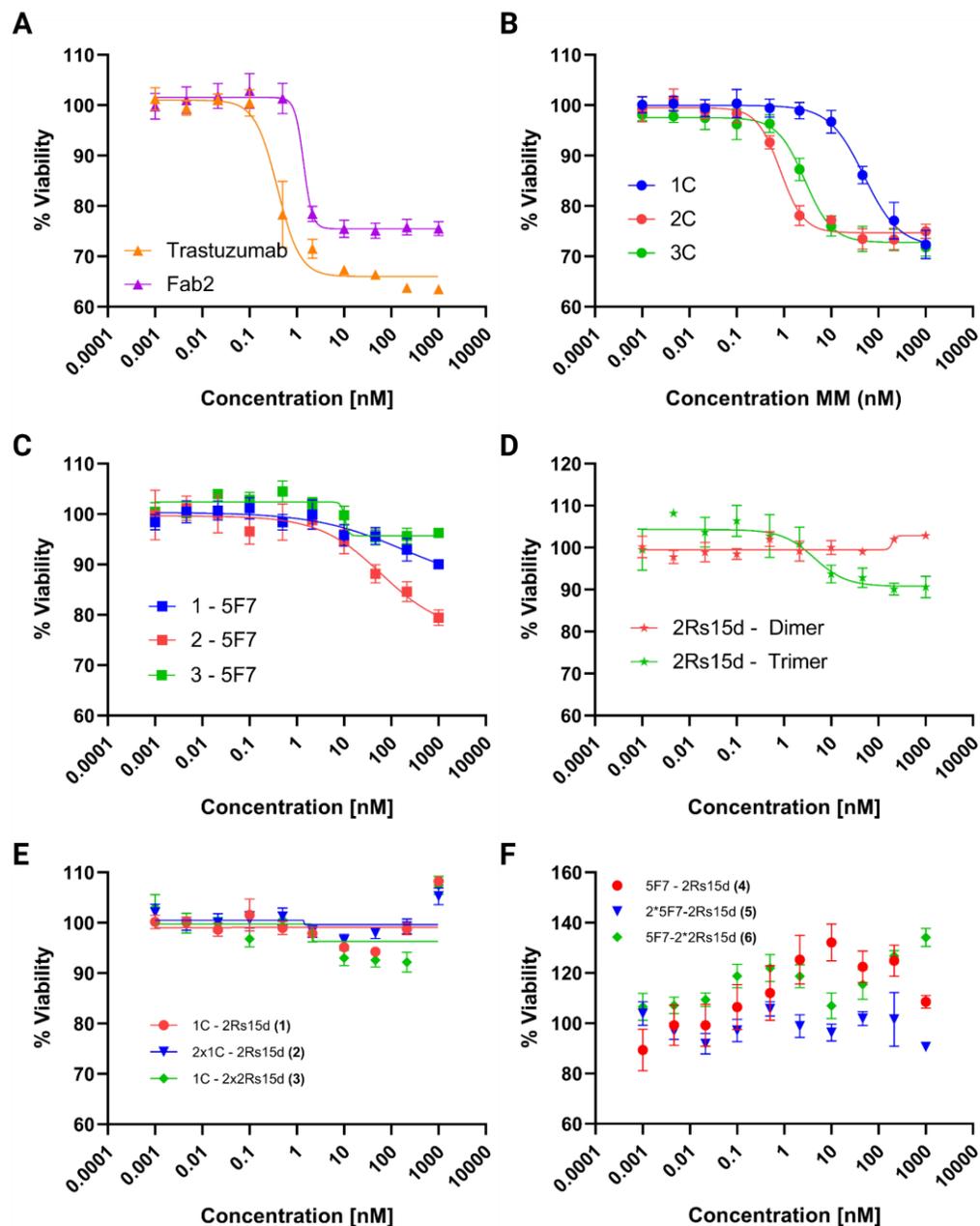
**Figure 2.14 2Rs15d Scaffolds Internalize Similarly to 5F7 Scaffolds.** After 48h of addition of 5nM construct, internalization of the 2Rs15d scaffolds can be similarly quantified as other scaffolds. Internalization rates are ( $\text{h}^{-1}$ ): **2Rs15d Monomer** – **0.054 +/- 0.007**; **2Rs15d Dimer** – **0.051 +/- 0.011**; **2Rs15d Trimer** – **0.053 +/- 0.007**. This is significantly slower than the biparatopic scaffolds, which demonstrates the synergy in targeting multiple epitopes.

Finally, I wanted to interrogate the effects that increased net internalization via epitope cross-linking had on viability of BT474 cells. Here, cells were cultured in opaque, clear-bottom 96-well plates, treated with concentrations ranging from 1pM to 1  $\mu\text{M}$ . After 6 days in culture, overall viability was measured using the alamarBlue™ reagent. For the monoparatopic Fab-based scaffolds, high concentrations inhibited cell proliferation with  $\text{EC}_{50}$  values similar to those our group has previously reported<sup>68</sup> (1-3nM), with 2C and 3C approximating trastuzumab and the trastuzumab  $\text{F}(\text{ab}')_2$  (**Figure 2.15A & 2.15B**). However, the monoparatopic nanobody scaffolds did not noticeably inhibit cell growth, despite having roughly equivalent-or-better binding affinities (**Figure 2.15C & 2.15D**). Similarly, despite the increased internalization rate of our biparatopic scaffolds, none of the biparatopic scaffolds displayed inhibition of cell growth (**Figure 2.15E & 2.15F**); in fact, the nanobody-based biparatopic fusions may demonstrate an agonistic effect towards cell proliferation (**Figure 2.15F**). This was not entirely surprising, as many nanobodies are matured and selected for purely their binding capacity, while trastuzumab was

evolved to both bind and inhibit cell proliferation. However, the biparatopic scaffolds that incorporated the trastuzumab Fab domain (i.e. 1C) also did not inhibit proliferation (**Figure 2.15E**), which is in contrast to the bivalent or trivalent monoparatopic trastuzumab structures. This result was surprising and unexpected; as a result, these experiments successfully decoupled cell binding, internalization rate, and inhibition of cell proliferation. Explanation of these differential effects will require further investigation of downstream gene expression through western blots (e.g. pAkt) – perhaps through a crystallography core to elucidate how these binding domains change the conformation of HER2 upon binding.

**Table 2.3.** Observed Internalization Rates and Binding EC<sub>50</sub>s of HER2-Targeting Biparatopic Scaffolds on BT474s

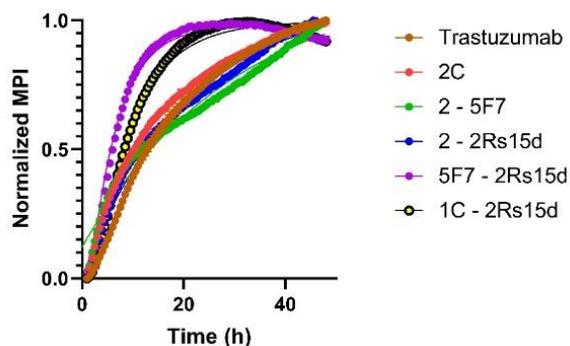
Scaffold	Int. Rate, $k_e$ (h <sup>-1</sup> )	Half-Time, $t_{1/2}$ (h)	% Positive (nM)	EC <sub>50</sub>
1C – 2Rs15d <b>(1)</b>	0.12 ± 0.01	~5.7	3.8 ± 0.1	
2*1C – 2Rs15d <b>(2)</b>	0.12 ± 0.01	~5.7	3.2 ± 0.1	
1C – 2*2Rs15d <b>(3)</b>	0.14 ± 0.01	~5.1	1.7 ± 0.1	
5F7 – 2Rs15d <b>(4)</b>	0.18 ± 0.02	~3.9	1.4 ± 0.1	
2*5F7 – 2Rs15d <b>(5)</b>	0.19 ± 0.01	~3.6	0.7 ± 0.1	
5F7 – 2*2Rs15d <b>(6)</b>	0.17 ± 0.01	~4.0	1.2 ± 0.1	
2 - 2Rs15d	0.051 ± 0.01	~13.6	4.6 ± 0.1	



**Figure 2.15 Cell Viability.** Viability curves were generated after treating BT474 cells with **A)** Trastuzumab and its F(ab')<sub>2</sub>, **B)** C-terminal trastuzumab Fab megamolecule fusions, **C)** Megamolecules presenting the nanobody 5F7, **D)** Megamolecules presenting the nanobody 2Rs15d, **E)** Biparatopic megamolecules presenting one or two 5F7 nanobodies and one or two 2Rs15d nanobodies, **F)** Biparatopic nanobodies presenting one or two C-terminal trastuzumab Fabs and one or two 2Rs15d nanobodies. Results demonstrate that *only* monoparatopic trastuzumab scaffolds show inhibition of cell proliferation. Error bars in SEM.

## 2.3 Conclusions

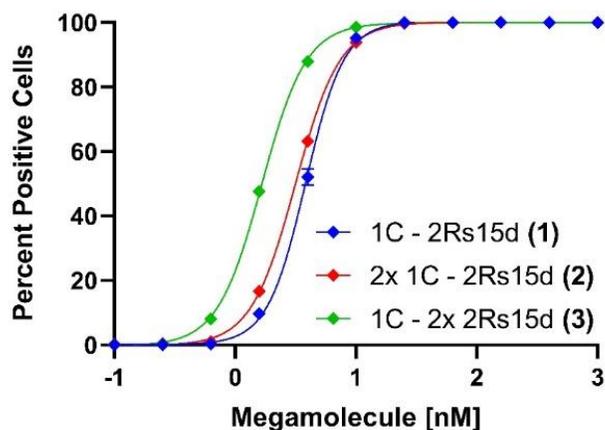
In **Chapter 2**, I demonstrated how megamolecules can be utilized to decouple binding affinity, valency, and specificity from effective internalization rate and inhibition of cell proliferation. Scaffolds bound with excellent affinity with both Fab and nanobody domains, yet the added avidity from two to three domains only modestly improved binding, which was even true with the biparatopic scaffolds. The monoparatopic fusions internalized ~20% faster than the parent trastuzumab and F(ab')<sub>2</sub>, which may be due to the increased flexibility in our scaffolds compared to an IgG. The cleavage of the Fc domain from trastuzumab did not significantly affect the binding, internalization, or cell growth inhibition of the mAb.



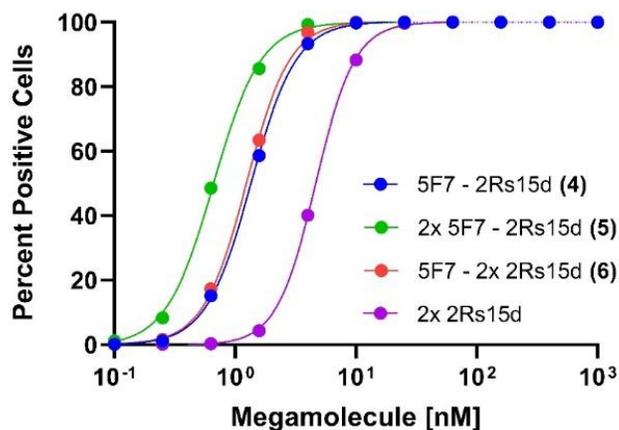
**Figure 2.16. Normalization of Internalization Rate of All Bivalent Scaffolds.** Here, the biparatopic scaffolds demonstrate a 3-4-fold rate increase in net internalization over trastuzumab.

Furthermore, the monoparatopic Fab scaffolds did not show an increase in effective internalization rate with added valency. I repeated this experimental workflow with 2Rs15d scaffolds, which also did not induce rapid internalization with increasing valency (**Figure 2.14**), demonstrating that increased net internalization was due to the synergy between targeting the two epitopes. The 2Rs15d scaffolds internalized similarly to the 5F7 scaffolds in that they did not plateau after 48h even though it bound cells with an affinity comparable to trastuzumab (~4.6nM) (**Table 2.3**). Only the subsequently prepared biparatopic megamolecule scaffolds dramatically increased net internalization rate over the parent trastuzumab mAb (**Table 3**, normalized bivalent

internalization in **Figure 2.16**). This mechanism persists even though the observed cell binding affinities via flow cytometry are all comparable (**Table 2.3, Figure 2.17 & 2.18**). This behavior is similar to that reported by previous groups<sup>80</sup> and helps validate the D1/D4 cross-linking mechanism of increased internalization demonstrated by Andreas Plückthun's group<sup>22,26</sup>.

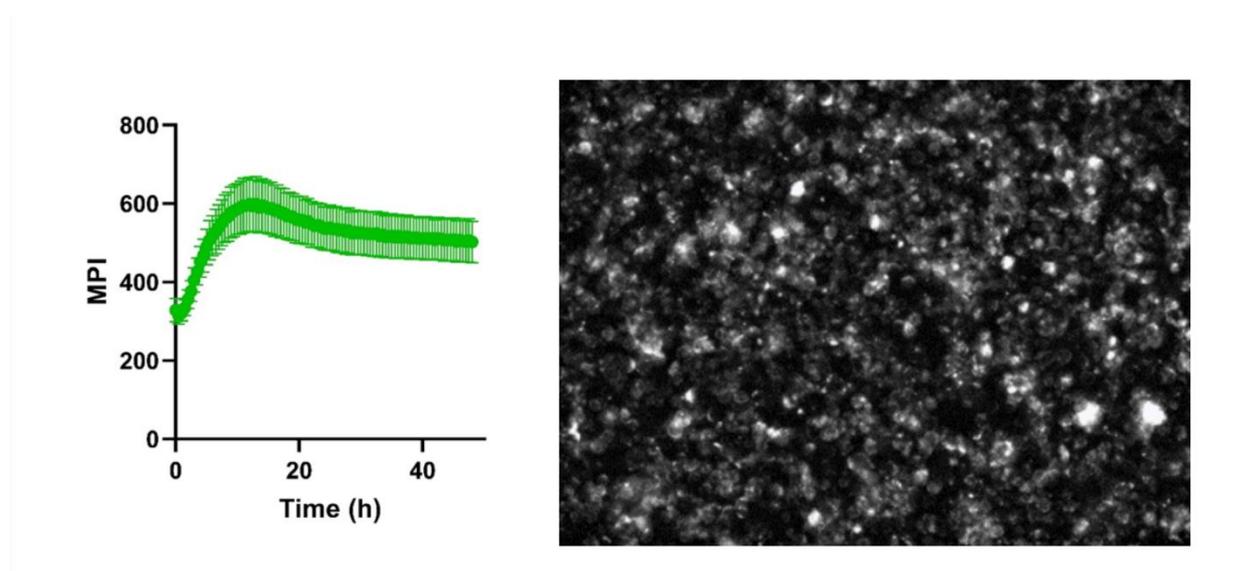


**Figure 2.17** Flow Cytometry Data of 1C-Based Biparatopic Megamolecules. The 1C biparatopics have effective binding  $EC_{50}$  values of 3.8 nM (1), 3.2 nM (2), and 1.7 nM (3).



**Figure 2.18** Flow Cytometry Data of 5F7-Based Biparatopic Megamolecules. The biparatopic megamolecules have tighter binding affinities than the monoparatopic 2Rs15d-cutinase dimer; effective  $EC_{50}$  affinity values are around 1.4 nM (4), 0.7 nM (5), 1.2 nM (6), and 4.6 nM (2Rs15d dimer).

At the later time points, some of the biparatopic scaffolds demonstrated a slight decrease in net internalization signal, which may be indicative of a loss of receptor expression on the cell surface. This is striking, as previous reports have suggested that high-expressing HER2 cell lines demonstrate robust receptor recycling after trastuzumab treatment whereas medium-to-low expressing HER2 cell lines show substantial decreases in HER2 surface expression after anti-HER2 antibody treatment<sup>87</sup>. Indeed, MDA-MB-453 cells may have demonstrated this decreased surface expression on HER2, as I do see a decrease in net internalization signal after about 12 hours (**Figure 2.19**). However, further research is necessary to confirm this hypothesis.



**Figure 2.19 MDA-MB-453 Cells Do Not Show Sustained Internalization.** MDA-MB-453 cells, which only moderately overexpress HER2 (~200K per cell), do not show the same level of sustained internalization that BT474 cells (~2\*10<sup>6</sup> per cell) do (here, **internalized 3C shown**); after about 12 hours, here is a drop in internalization, suggesting that HER2 does not get recycled as efficiently in this cell line.

Finally, cell proliferation was only inhibited in the monoparatopic, Fab-based scaffolds. The effective EC<sub>50</sub> values for 2C, 3C, trastuzumab, and the F(ab')<sub>2</sub> were all similar – from 1-5nM – while 1C's effective EC<sub>50</sub> was closer to 50nM, demonstrating how scaffold avidity can improve efficacy. Curiously, the biparatopic scaffolds which incorporated either one (i.e. 1C) or two (i.e.

2C) Fab domains did *not* display any inhibition of cell proliferation after 6 days in culture. This suggests that incorporation of the D1-targeting 2Rs15d nanobody concurrently abrogates the anti-proliferative ability of linked trastuzumab Fabs while increasing the net internalization rate of the antibody-receptor complex. Further research is required to elucidate the mechanisms of action (MOAs) of each individual scaffold; this is important, as even the MOAs of trastuzumab binding continue to be disputed. While it is generally accepted that trastuzumab has no effect on ligand-induced HER2 heterodimerization with EGFR, HER3, and HER4<sup>88-90</sup> ligand-independent HER2 heterodimerization remains disputed<sup>88,89,91</sup>. Furthermore, there is evidence that trastuzumab activates HER2 homodimerization<sup>88,92</sup>. Another group's molecular dynamics simulations suggested that trastuzumab binding concurrently has agonistic *and* antagonistic properties regarding inhibition of HER2 dimerization depending on whether HER2 is the activator or receiver of phosphorylation on the dimer pair<sup>93</sup>.

The MOA of HER2-trastuzumab internalization is also controversial; some suggested a downregulation of HER2 on the surface<sup>94-98</sup> while others demonstrated little-to-no effect<sup>99-101</sup> with robust receptor recycling for overexpressing cells<sup>99</sup>, which correlates with our data within the first 48 hours. A recent review<sup>102</sup> comprehensively elucidates these specific controversies surrounding the trastuzumab-HER2 MOAs that contribute to its clinical efficacy. Taken together, the specific MOAs of trastuzumab need further investigation to resolve many of these inconsistencies.

Adding a third binding domain against another non-overlapping epitope on HER2 utilizing a third orthogonal linking chemistry, such as our recently-published CrabTag<sup>42</sup>, could be used to investigate whether a triparatopic antibody scaffold (e.g., adding a DII-targeting pertuzumab nanobody) has any improved effective internalization over a biparatopic (or, if a third epitope can even bind concurrently with the other two). This is the focus of my final work that will not make

this dissertation but may make the final draft of this paper. Here, I found another nanobody towards DII, the same domain that pertuzumab targets. This nanobody, 47D5, was found through a patent published by Ablynx<sup>103</sup>, a subsidiary of Sanofi. In addition to similar branched scaffolds demonstrated in this work, I am also synthesizing N and C-terminal triple fusions of cutinase and every permutation of two of 5F7, 2Rs15d, and 47D5. Initial insights reveal sufficient cross-linking for increased net internalization rate with D1/D4 and D2/D4 nanobody fusions; however, D1/D2 fusions cannot sufficiently cross-link and undergo the slower internalization of monoparatopic nanobodies. In addition, branched triple fusions may demonstrate increased internalization rate with increased valency, which is not something seen previously or present in this work.

Further, addition of the engineered Fab against DI (e.g., hA21G<sup>82</sup>) in a biparatopic scaffold could be used to investigate whether use of this Fab can concurrently increase effective internalization rate while retaining the antiproliferative properties of trastuzumab. Initial clinical trials of the biparatopic zanidatamab (ZW25) seem to suggest an improved receptor internalization over trastuzumab in biliary tract cancers<sup>104</sup>, though more work needs to be done to confirm this. Further, testing biparatopic scaffolds in an ADC format *in vivo* could help elucidate whether the demonstrated *in vitro* gains in efficacy could lead to increased cytotoxicity and tumor reduction. The small size of nanobodies may aid tumor penetration, which can be difficult to achieve when scaffolds have potent binding affinities<sup>105</sup>. Potentially, co-administration of a biparatopic ADC at a lower concentration with a naked antibody at a high concentration may portend greatest tumor penetration<sup>106</sup>. In sum, the modular assembly platform of megamolecules allowed me to interrogate the effects of size, affinity, and specificity on antibody scaffold efficacy *in vitro*, which could potentially inform the rational design of biologics in the clinic.

## 2.4 Experimental

**Materials.** BT474, MDA-MB-453, MDA-MB-231, and C4-2 cell lines were purchased from the American Tissue Culture Collection (ATCC). All primers and genes (gBlocks) were purchased from Integrated DNA Technologies (IDT). DH5 $\alpha$ , SHuffle Express competent *E. Coli* were purchased from New England Biolabs along with Golden Gate master mix (BsaI). Inducing agents anhydrotetracycline (aTc; for Fab fusions) and isopropyl B-D-1-thiogalactopyranoside (IPTG; for nanobody fusions) were purchased from Sigma Aldrich. Phosphate-buffered saline (PBS), Dulbecoo's Modified Eagle Medium (DMEM), Fetal Bovine Serum (FBS), Penicillin/Streptomycin (PenStrep), FluoroBrite™ DMEM, Trypsin-EDTA (0.25%), Accutase® Cell Detachment Solution, alamarBlue™, Alexa Fluor™ 488 Antibody Labeling Kit, and 96-Well Black/Clear Bottom, TC Surface plates were purchased from Thermo Fisher Scientific. Bio-Rad Spin 6 (Tris) columns were purchased from Bio-Rad. u-Slide 8 Well Chamber slides (No. 1.5 polymer glass bottom, TC prepared) were purchased from ibidi.

### Protein Sequences:

#### Trastuzumab Sequence

(V<sub>L</sub> / (kappa) C<sub>L</sub>)

DIQMTQSPSSLSASVGDRTTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVP  
SRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPPS  
DEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSSTYLSSTL  
TLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

(V<sub>H</sub> / C<sub>H</sub> / Fc Domain)

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYT  
RYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGT  
LVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSQVHTFP  
AVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKEPKSCDKTHTCPPCPA  
PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEVHNAKTK  
PREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQV  
YTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLY  
SKLTVDKSRWQQGNVFNCSVMHEALHNHYTQKSLSLSPGK

Forms Dimer

Formula: C<sub>6460</sub>H<sub>9998</sub>N<sub>1724</sub>O<sub>2011</sub>S<sub>44</sub>

Theoretical Mass (account for 16 disulfide bonds): 145403 Da

$\epsilon_{280} = 215380 \text{ M}^{-1} \text{ cm}^{-1}$

### F(ab')<sub>2</sub> Trastuzumab Sequence

(V<sub>L</sub> / (kappa) C<sub>L</sub>)

MDIQMTQSPSSLSASVGDRVITICRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGV  
PSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPP  
SDEQLKSGTASVCLLNNFYPRKAVQWKVDNALQSGNSQESVTEQDSKDYSLSSST  
LTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

(V<sub>H</sub> / C<sub>H</sub>)

EVQLVESGGGLVQPGGSLRLSCAASGFRNIKDTYIHWVRQAPGKGLEWVARIYPTNGYT  
RYADSVKGRFTISADTSKNTAYLQMNSLR AEDTAVYYCSRWGGDGFYAMDYWGQGT  
LVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSQVHTFP  
AVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSC

Forms Dimer

Formula: C<sub>4182</sub>H<sub>6474</sub>N<sub>1120</sub>O<sub>1325</sub>S<sub>28</sub>

Theoretical Mass (account for 10 disulfide bonds): 94285 Da

$\epsilon_{280} = 143810 \text{ M}^{-1} \text{ cm}^{-1}$

### Trastuzumab Fab – Cutinase Sequence (C<sub>H</sub> Fusion “1C”)

(V<sub>L</sub> / (kappa) C<sub>L</sub>)

MDIQMTQSPSSLSASVGDRVITICRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGV  
PSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPP  
SDEQLKSGTASVCLLNNFYPRKAVQWKVDNALQSGNSQESVTEQDSKDYSLSSST  
LTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

(V<sub>H</sub> / C<sub>H</sub> / XTEN / Cutinase / Thrombin)

MEVQLVESGGGLVQPGGSLRLSCAASGFRNIKDTYIHWVRQAPGKGLEWVARIYPTNGY  
TRYADSVKGRFTISADTSKNTAYLQMNSLR AEDTAVYYCSRWGGDGFYAMDYWGQGT  
TLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSQVHTFP  
PAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCSGSETPGTSES  
ATPESGLPTS NPAQELEARQLGRTRRDDLINGNSASCADVIFIYARGSTETGNLGLTLP  
SI  
ASNLESAFGKDG VVIQGVGGAYRATLGDNALPRGTSSAAIREMLGLFQQANTKCPDAT  
LIAGGYSQGAALAAASIEDLDSAIRDKIAGTVLFGYTKNLQNRGRIPNYPADRTKVFCNT  
GDLVCTGSLIVAAPHLAYGPDARGPAPEFLIEKVRVARGSALVPRGSHHHHHH

Formula: C<sub>3189</sub>H<sub>4984</sub>N<sub>888</sub>O<sub>1015</sub>S<sub>20</sub>

Theoretical Mass (account for 7 disulfide bonds): 72638 Da

Experimental Mass (ESI-MS): 72655 Da [M + H<sub>2</sub>O]

$\epsilon_{280} = 86595 \text{ M}^{-1} \text{ cm}^{-1}$

**Trastuzumab Fab – Cutinase Sequence (V<sub>H</sub> Fusion “1N”)**(V<sub>L</sub> / (kappa) C<sub>L</sub>)

MDIQMTQSPSSLSASVGDRVTTICRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGV  
 PSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPP  
 SDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDYSLSSST  
 LTLSKADYEEKHKVYACEVTHQGLSSPVTKSFNRGEC

(Cutinase / XTEN / V<sub>H</sub> / C<sub>H</sub> / Thrombin)

MGLPTSNPAQELEARQLGRTRRDDLINGNSASCADVIFIYARGSTETGNLGLTGPSIASNL  
 ESAFGKDG VWIQGVGGAYRATLGDNALPRGTSSAAIREMLGLFQQANTKCPDATLIAG  
 GYSQGAALAAASIEDLDSAIRDKIAGTVLFGYTKNLQNRGRIPNYPADRTKVCNTGDL  
 VCTGSLIVAAPHLAYGPDARGPAPEFLIEKVRAVRGSA **SGSETPGTSESATPESEVQLVES**  
**GGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVK**  
**GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTQVTVSSAS**  
**TKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGL**  
**YLSVVTVPSSSLGTQTYICNVNHKPSNTKVKDKKVEPKSCLVPRGSHHHHHH**

Formula: C<sub>3189</sub>H<sub>4984</sub>N<sub>888</sub>O<sub>1015</sub>S<sub>20</sub>

Theoretical Mass (account for 7 disulfide bonds): 72638 Da

Experimental Mass (ESI-MS): 72656 Da [M + H<sub>2</sub>O] – don't see methionine cleavage in this fusionε<sub>280</sub>: 86595 M<sup>-1</sup> cm<sup>-1</sup>**V<sub>H</sub>H 5F7 – Cutinase Sequence (C-Term Fusion)**(Cutinase / XTEN / V<sub>H</sub>H 5F7)

MG LPTSNPAQELEARQLGRTRRDDLINGNSASCADVIFIYARGSTETGNLGLTGPSIASNL  
 ESAFGKDG VWIQGVGGAYRATLGDNALPRGTSSAAIREMLGLFQQANTKCPDATLIAG  
 GYSQGAALAAASIEDLDSAIRDKIAGTVLFGYTKNLQNRGRIPNYPADRTKVCNTGDL  
 VCTGSLIVAAPHLAYGPDARGPAPEFLIEKVRAVRGSA **SGSETPGTSESAEVQLVESGGG**  
**LVQAGGSLRLSCAASGITFSINTMGWYRQAPGKQRELVALISSIGDTYYADSVKGRFTIS**  
**RDNAKNTVYLQMNSLKPEDTAVYYCKRFRTAAQGTDYWGQGTQVTVSSHHHHHHH**

Formula: C<sub>1615</sub>H<sub>2555</sub>N<sub>473</sub>O<sub>515</sub>S<sub>10</sub>

Theoretical Mass (account for 3 disulfide bonds and methionine cleavage): 37024 Da

Experimental Mass (ESI-MS): 37024 Da [M] and 37066 [M + ACN]

ε<sub>280</sub>: 36245 M<sup>-1</sup> cm<sup>-1</sup>**V<sub>H</sub>H 2Rs15d – Cutinase Sequence (N-Term Fusion)**(V<sub>H</sub>H 2Rs15d / XTEN / Cutinase)

**MQVQLQESGGGSVQAGGSLKLTCAASGYIFNSCGMGWYRQSPGRERELVSRISGDGDT**  
**WHKESVKGRFTISQDNVKKTLYLQMNSLKPEDTAVYFCAVCYNLETYWGQGTQVTVS**  
**SSGSETPGTSESA**GLPTSNPAQELEARQLGRTRRDDLINGNSASCADVIFIYARGSTETGN  
 LGTLGPSIASNLESAFGKDG VWIQGVGGAYRATLGDNALPRGTSSAAIREMLGLFQQAN

TKCPDATLIAGGYSQGAALAAASIEDLDSAIRDKIAGTVLFGYTKNLQNRGRIPNYPADR  
TKVFCNTGDLVCTGSLIVAAPHLAYGPDARGPAPEFLIEKVRAVRGSAHHHHHH

Formula: C<sub>1604</sub>H<sub>2525</sub>N<sub>469</sub>O<sub>513</sub>S<sub>12</sub>

Theoretical Mass (account for 4 disulfide bonds): 36968 Da

Experimental Mass (ESI-MS): 36964 Da

$\epsilon_{280}$ : 40380 M<sup>-1</sup> cm<sup>-1</sup>

### V<sub>H</sub>H 2Rs15d – SnapTag Sequence (N-Term Fusion)

(V<sub>H</sub>H 2Rs15d / XTEN / SnapTag)

MQVQLQESGGGSVQAGGSLKLTCAASGYIFNSCGMGWYRQSPGRERELVSRISGDGDT  
WHKESVKGRFTISQDNVKKTYLQMNSLKPEDTAVYFCAVCYNLETYWGQGTQVTVS  
SSGSETPGTSESADKDCMKRTTLDSPGLKLELSGCEQGLHEIIFLGKGTSAADAVEVPA  
PAAVLGGPEPLMQATAWLNA YFHQPEAIEFPVPALHHPVFQQESFTRQVLWKLKVV  
KFGEVISYSHLAALAGNPAATAAVKTALSGNPVPIIPCHR VVQGDLDVGGYEGGLAVK  
EWLLAHEGHRLGKPLGHHHHHH

Formula: C<sub>1508</sub>H<sub>2337</sub>N<sub>417</sub>O<sub>453</sub>S<sub>12</sub>

Theoretical Mass (account for 3 disulfide bonds): 33938 Da

Experimental Mass (ESI-MS): 33936.85 Da [M + Na]

$\epsilon_{280}$ : 46785 M<sup>-1</sup> cm<sup>-1</sup>

### V<sub>H</sub>H C8 – Cutinase Sequence (C-Term Fusion)

(Cutinase / XTEN / V<sub>H</sub>H C8)

MGLPTSNP AQELEARQLGRTRDDLINGNSASCADVIFIYARGSTETGNLGTLP SIASNL  
ESAFGKDG VWIQGVGGAYRATLGDNALPRGTSSAAIREMLGLFQQANTKCPDATLIAG  
GYSQGAALAAASIEDLDSAIRDKIAGTVLFGYTKNLQNRGRIPNYPADR TKVFCNTGDL  
VCTGSLIVAAPHLAYGPDARGPAPEFLIEKVRAVRGSA SSGSETPGTSESAVQLQASGGGF  
VQPGGSLRLSCAASGDSYNESSMGWFRQAPGKEREFVSAISARGNHPLY YADSVKGRF  
TISRDN SKNTVYLQMNSLRAEDTATYYCASM PMPKWK KYWGQGTQVTVSHHHHHH

Formula: C<sub>1618</sub>H<sub>2536</sub>N<sub>476</sub>O<sub>511</sub>S<sub>12</sub>

Theoretical Mass (account for 3 disulfide bonds and methionine cleavage): 37083 Da

Experimental Mass (ESI-MS): 37082 Da [M], 37123 [M + ACN]

$\epsilon_{280}$ : 41745 M<sup>-1</sup> cm<sup>-1</sup>

### V<sub>H</sub>H 2Rs15d – 5F7 Fusion Dimer Sequence

(V<sub>H</sub>H 2Rs15d / XTEN / 5F7)

MQVQLQESGGGSVQAGGSLKLTCAASGYIFNSCGMGWYRQSPGRERELVSRISGDGDT  
WHKESVKGRFTISQDNVKKTYLQMNSLKPEDTAVYFCAVCYNLETYWGQGTQVTVS  
SSGSETPGTSESAEVQLVESGGGLVQAGGSLRLSCAASGITFSINTMGWYRQAPGKQRE  
LVALISSIGD TTYADSVKGRFTISRDN AKNTVYLQMNSLKPEDTAVYYCKRFRTAAQGT  
DYWGQGTQVTVSSHHHHHH

Formula: C<sub>1194</sub>H<sub>1849</sub>N<sub>343</sub>O<sub>382</sub>S<sub>11</sub>

Theoretical Mass (account for 3 disulfide bonds): 27470 Da

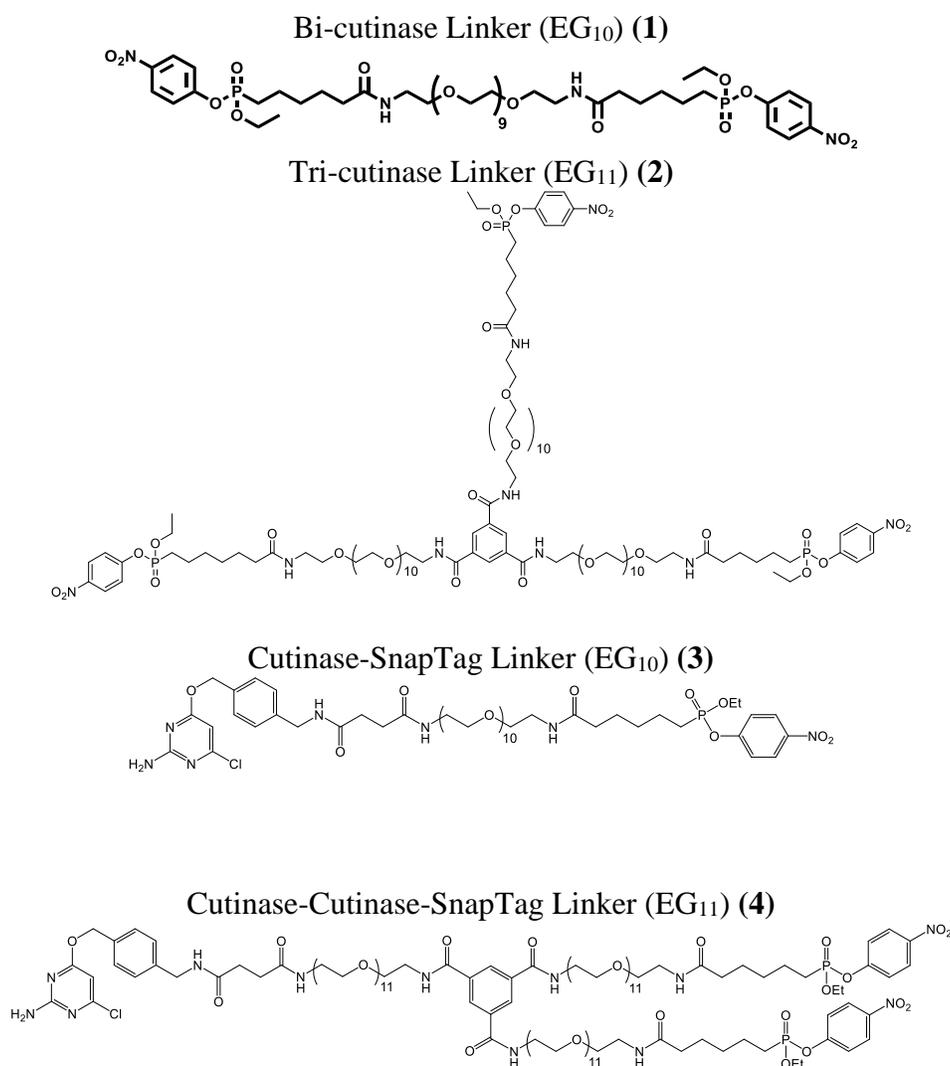
Experimental Mass (ESI-MS): 27470 Da [M]

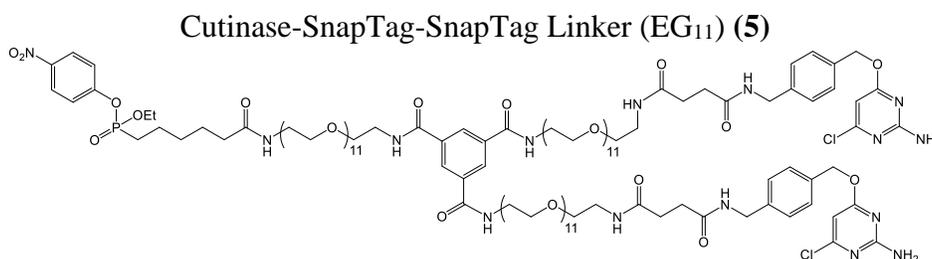
$\epsilon_{280}$ : 47245 M<sup>-1</sup> cm<sup>-1</sup>

**Protein Expression and Purification.** In short, all constructs were cloned and expressed in *E. Coli*. Fab-cutinase fusions were prepared as previously reported<sup>68</sup>. Nanobody fusions to megamolecule building block enzymes were cloned into pET-28b(+), a kanamycin-resistant and IPTG inducible expression vector, via Golden Gate compatible restriction enzyme sites (BsaI), containing a C-Terminus 6xHisTag for purification. Nanobodies were fused to Megamolecule building blocks using a 12-mer XTEN linker (SGSETPGTSESA). For all cloning and plasmid propagation, NEB DH5a cells were used. Fusions containing nanobodies 5F7 and C8 were cloned at the C-Terminus of the fusion protein while the megamolecule linking enzyme was placed at the N-terminus. 2Rs15d-fusions contained an N-Terminus nanobody and a C-Terminus enzyme. Sequence confirmed plasmids containing the nanobody-megamolecule fusions were transformed into SHuffle® T7 Express Competent *E. coli*. Cultures were grown in 2x YT culture medium containing 50ug/ml of kanamycin. Cultures were inoculated with a stab of cells and cultured overnight at 30C shaking at 250RPM in 5mL growth medium. After 12 hours, cultures were transferred into 2L baffled flasks containing 500mL growth medium with 1:1000 antibiotic and cultured in an Innova 44R (New Brunswick Scientific) incubator. Once cultures reached an OD600 of 0.8, they were induced with 2.5mM (final concentration) of IPTG and cultured 16 hours at 20C. The following day, cultures were pelleted at 5000RPM and were resuspended with lysis buffer (2XPBS containing DNase) and sonicated on ice for 2min. To remove cell debris from lysates, samples were centrifuged at 10K RPM for 10min, and supernatants were transferred to 50ml Kontes Flex columns (Kimbal Kontes Glassware) and were incubated with HisPur cobalt resin to

capture 6xHis Tagged proteins at 4C on a shaker for 2hours. Resin was washed three times with 1xPBS to remove any weakly bound proteins and desired products were eluted with 150mM imidazole. Samples were further purified by size exclusion chromatography using an AKTA FPLC system on a HiLoad 16/600 Superdex column with PBS + 0.05% NaN<sub>3</sub>.

**Linker Synthesis.** Linkers and similar compounds were previously synthesized and stocked in DMSO at 10mM. I am grateful for Justin and Blaise who made and stocked these in anticipation for future projects. Linkers:





**Scaffold Synthesis.** Generally, all proteins were stored in PBS unless otherwise noted. Typically, megamolecule scaffold reactions were carried out at a 10-50 nmol scale. For example, to create 2C, a solution of cutinase-Fab fusions (C<sub>H</sub>) (2mL, 20uM) was reacted with the appropriate EG<sub>10</sub> bis-(4-nitrophenyl)-ethyl phosphonate linker (1.8uL, 10mM in DMSO) in slight molar excess to ensure two fusions reacted per di-cutinase linker. The reaction was let to go overnight at room temperature, and then was purified the next day by an Akta FPLC on a Hi-Load 16/60 Superdex 200 column with an elution buffer of PBS + 0.05% NaN<sub>3</sub>.

**Antibody Mass Calculation.** The mass of the total fusion proteins is added to the mass of the linker of interest minus the mass of the leaving groups of the reaction. For each linker, the net mass is as follows:

Di-cutinase Linker (EG<sub>10</sub>) (1): +835 Da<sup>108</sup>  
 Tri-cutinase Linker (EG<sub>11</sub>) (2): +2328 Da<sup>109</sup>  
 Cutinase-SnapTag Linker (EG<sub>10</sub>) (3): + 936 Da<sup>107</sup>  
 Cutinase-Cutinase-SnapTag Linker (EG<sub>11</sub>) (4): +2400 Da<sup>69</sup>  
 Cutinase-SnapTag-SnapTag Linker (EG<sub>11</sub>) (5): +2386 Da<sup>107</sup>

These are reported from the original publication of use. Furthermore, one proton is lost from each nucleophilic residue in the active site of cutinase or SnapTag. For example, to make trimeric biparatopic scaffold (3) with 2x 5F7 nanobody-cutinase (37024.32 Da) and 1x 2Rs15d nanobody-SnapTag (33938.39 Da), take the mass from each, add the linker without the leaving groups (2400), and subtract out 3 protons. Mass = 37024.32\*2 + 33938.39 + 2400 – 3 = 110384 Da, which is within 8 Da of the experimental mass (**Table 2.4**).

**Table 2.4** Calculated and Experimental Molecular Weights (via ESI-MS) of Fusion Scaffolds with Extinction Coefficients

Scaffold	Calculated MW (Da)	Exper. MWs (Da)	Ext. Coeff. (M <sup>-1</sup> cm <sup>-1</sup> )
Tras. Fab-Cutinase (C <sub>H</sub> ) ( <b>1C</b> )	72638	72655 [M+H <sub>2</sub> O]	86595
2x Tras. Fab-Cutinase (C <sub>H</sub> ) ( <b>2C</b> )	146110**	146143	173190
3x Tras. Fab-Cutinase (C <sub>H</sub> ) ( <b>3C</b> )	220240**	n/a	259785
Cutinase-Tras. Fab (V <sub>H</sub> ) ( <b>1N</b> )	72638	72656 [M+H <sub>2</sub> O]	86595
2x Cutinase-Tras. Fab (V <sub>H</sub> ) ( <b>2N</b> )	146110**	n/a	173190
3x Cutinase-Tras. Fab (V <sub>H</sub> ) ( <b>3N</b> )	220240**	n/a	259785
Cutinase-5F7 ( <b>1-5F7</b> )	37024	37024, 37066 [M+ACN]	36245
2x Cutinase-5F7 ( <b>2-5F7</b> )	74881	74879, 74921 [M+ACN], 74963 [M+2*ACN]	72490
3x Cutinase-5F7 ( <b>3-5F7</b> )	113396	113694 [M]	108735
2Rs15d-Cutinase	36968	36964 [M]	40380
2x 2Rs15d-Cutinase ( <b>2-2Rs15d</b> )	74770	74762 [M]	80760
2Rs15d-SnapTag	33938	33937 [M]	46785
Tras. Fab-Cutinase (C <sub>H</sub> ) – 2Rs15d-SnapTag ( <b>1</b> )	107510	107526 [M + H <sub>2</sub> O]	133380
2x 1C – 2Rs15d-SnapTag ( <b>2</b> )	181611**	n/a	219975
1C – 2x 2Rs15d-SnapTag ( <b>3</b> )	142897**	n/a	180165
Cutinase-5F7 – 2Rs15d-SnapTag ( <b>4</b> )	71896	71893 [M], 71935 [M+ACN]	83030
2x Cutinase-5F7 – 2Rs15d-SnapTag ( <b>5</b> )	110384	110392 [M]	119275
Cutinase-5F7 – 2x 2Rs15d-SnapTag ( <b>6</b> )	107283	107279 [M]	129815
Trastuzumab	145407*	148223	215380
Tras. F(ab) <sub>2</sub>	94285*	97290	143810
2Rs15d – 5F7 Dimer Fusion	27470	27470 [M]	47245

\*Calculated MWs do not include glycoforms observed in mammalian monoclonal antibody production

\*\* ESI-MS data is generally noisy and inconclusive around 150 kDa and above

**Analytical Size-Exclusion Chromatography.** Similar quantitative assessment was done previously<sup>69</sup>. In short, peak elution volumes,  $V_e$ , were determined by SEC over at least duplicate injections. A protein standard mixture of four proteins ranging from 15 to 600 kDa (Sigma) was analyzed by the same method. The void volume,  $V_0$ , was determined by injection of 2 MDa dextran blue (Sigma), and the column volume,  $V_t$ , was determined by injection of 2% acetone in water. The partition coefficient,  $K_{av}$ , is calculated using Equation S1. Data for the protein standard mixture were fit using a power law model.

Calculated Void Volume = 8.2 mL

Column Volume = 20.7mL

$$K_{av} = (V_e - V_0) / (V_t - V_0) \quad (S1)$$

**Electrospray Ionization Mass Spectrometry.** ESI spectra were obtained on an Agilent 6230 LC/TOF mass spectrometer equipped with an Agilent 1260 Infinity Binary Pump. Samples were prepared at 1 $\mu$ M in nuclease-free water. A volume of 7 $\mu$ L sample was injected into a 1cm C18 guard column (Waters) and eluted using a gradient from 0% to 100% acetonitrile over 7.5 minutes and 0.1% formic acid in water at a flow rate of 0.4 mL/min. Data were analyzed using the Agilent MassHunter BioConfirm software. The protein molecular weight was obtained using a maximum entropy deconvolution calculation.

**Cell Culture.** All cell lines were cultured at 5% CO<sub>2</sub> and 37C. BT474, MDA-MB-453, C4-2, and MDA-MB-231 cell lines were cultured at 10% FBS, 1% P/S in DMEM. Cells were sub-cultured every 3-4 days.

**Confocal Microscopy.** For quick imaging on our Nikon Ti Eclipse, samples were plated in the ibidi u-Slide 8 well chambers and let to adhere overnight. Typically, samples were in FluoroBrite™ DMEM for imaging so that the phenol red typically present in cell culture media would not add significant autofluorescence to the background. Images were either taken on a 10X air objective or a 60X oil objective for stronger resolution.

**Flow Cytometry.** Samples were initially at 1 mg/mL in PBS, 100mM sodium bicarbonate. Each sample was reacted with one vial Alexa Fluor™ 488 antibody labeling dye with 100 $\mu$ L of sample. Reactions proceeded for one hour at room temperature on a tabletop rocker. Reaction mixtures were then filtered out with the resin filter with the kit. Concentrations and degree of labeling were

quantified by a NanoDrop™ benchtop spectrophotometer. The following equations, from the kit manual, were used (factor of 10 to adjust for our 1 mm path length):

$$[\text{Scaffold Concentration in } \mu\text{M}] = (A_{280} - 0.11 * A_{494}) / \text{Ext. Coeff} * 10000000 \quad (\text{S2})$$

$$\text{Degree of Labeling} = A_{494} / (71,000 * \text{Scaffold Concentration}) * 10 \quad (\text{S3})$$

All adherent cells were lifted from polystyrene culture flasks using Accutase™ dissociation reagent to retain intact surface HER2. Alexa Fluor™ 488-labeled scaffolds were serially diluted in 10% FBS/PBS from 1000nM to 100pM for monoparatomic scaffolds and 1000nM to 39pM for bipolaratomic scaffolds. 200,000 cells were stained at each concentration for 2 hours at room temperature and subsequently quenched and washed three times in 10% FBS/PBS. Samples were acquired using a BD LSRFortessa SORP Cell Analyzer™. Data was processed with FACSDiva™ Software and fit to 4-parameter regression curve in GraphPad Prism™ to determine EC50 values. Samples were prepared and ran in triplicate for monospecific scaffolds and in duplicate for bipolaratomic scaffolds.

**pHAb Labeling.** To label our scaffolds for internalization, we prepared each individual monomer fusion at 1 mg/mL in 20mM sodium bicarbonate in PBS. Then, the appropriate amount of dye, based on the user guide, was added. Each dye vial was resuspended in 25uL 1:1 DMSO:H<sub>2</sub>O just before use but was otherwise kept at -80C. One vial (25uL) sufficiently labeled 2mg of protein. After the dye was added to the mildly basic solution of protein, it was allowed to react for 1 hour at room temperature on a tabletop rocker. After an hour, the crude mixture was filtered through Bio-Rad Spin 6 columns to remove the bulk of unreacted dye. From here, megamolecule scaffolds were allowed to form overnight after addition of requisite linker. Then, each megamolecule (as well as the labeled monomeric fusion) were purified on the Akta FPLC on a Hi-Load 16/60

Superdex 200 column with an elution buffer of PBS + 0.05% NaN<sub>3</sub>. Concentration and degree of labeling were quantified by a NanoDrop™ benchtop spectrophotometer; the following equations from the kit manual (extra factor of 10 to adjust for our 1 mm path length):

$$[\text{Scaffold Concentration in } \mu\text{M}] = (A_{280} - 0.11 * A_{550}) / \text{Ext. Coeff} * 10000000 \quad (\text{S4})$$

$$\text{Degree of Labeling} = A_{550} / (75,000 * \text{Scaffold Concentration}) * 10 \quad (\text{S5})$$

**Internalization.** The day before internalization, 100,000 BT474s were plated in each well on an 8-well ibidi u-Slide and allowed to equilibrate overnight in FluoroBrite DMEM supplemented with 10% FBS, 1% P/S. Chamber slides were placed on a Lionheart FX Automated Microscope. Using the Gen5 software, a Z-stack throughout the entire cell was acquired (at 10x resolution) every 30m for 48 hours. Before acquisition started, 5 nM of pHAb labeled scaffold was added to the cells. After acquisition, a *focused projection* timelapse was created, which emphasized the planes most in focus for the whole cell height. Then, a general background subtraction was done on each cell stack. Images were stored as TIFFs and analyzed with a MATLAB script that only averaged pixel intensities above the background fluorescence (usually, around 350-450 intensity units). Internalization performed in duplicate. The first two time points were deleted in calculating rate, as most of the image was just noise. Rates were graphed and calculated in GraphPad Prism using a 1<sup>st</sup> order association curve. Experiments were performed in duplicate with at least six technical replicates.

## MATLAB Script to Quantify Internalization Signal.

```

%% Load in Image
for i = 1:9
    TiffFilename = strcat('A3ROI1_-3_2_1_Tsf[ZProj[RFP 531,593]]_00', num2str(i), '.tif');
    I = imread(TiffFilename);
    mask = (I > 400);
    I(~mask) = 0;
    MPI(i) = sum(sum(I)) ./ sum(sum(mask));
end

for i = 10:97
    TiffFilename = strcat('A3ROI1_-3_2_1_Tsf[ZProj[RFP 531,593]]_0', num2str(i), '.tif');
    I = imread(TiffFilename);
    mask = (I > 400);
    I(~mask) = 0;
    MPI(i) = sum(sum(I)) ./ sum(sum(mask));
end

MPI = MPI;

```

**Cytotoxicity.** For cytotoxicity, 10,000 BT474s were plated in each well in a 96-well opaque, clear bottom plate. The outer ring of wells demonstrated edge effects that induced artifacts in our data, so those were omitted and filled with PBS to maintain humidity across the plate. A  $\log_{10}$  range of concentrations from 1 $\mu$ M to 1pM were added to the first 5 rows, while the 6<sup>th</sup> served as a no treatment control. Total volume was 100 $\mu$ L. Each scaffold was buffer exchanged with pure PBS thoroughly before use to remove the 0.05% NaN<sub>3</sub> buffer. Cells were placed in a plastic box with wet paper towels (to ensure minimal evaporation) for six days. Then, 10 $\mu$ L of alamarBlue™ was added to each well (10% v/v) and fluorescence (ex: 560nm, em: 590nm) was read on a benchtop plate reader. Decreases in fluorescence correlated with fewer cells in the well, which suggested inhibition of cell proliferation. Percent viability was determined as a ratio of fluorescence from the no treatment control lane. Cytotoxicity performed in duplicate with at least ten technical replicates.

## Chapter 3

### Megamolecules as Sensors for Reversible Control of Protein Quaternary Structure Using Dynamic Covalent Linking Groups

Research and figures presented in this chapter are adapted from work from the following manuscript that is in preparation to be submitted:

Modica, J. A.\*, Sykora, D. J.\*, He, P., Weigand, S. Kimmel, B. R., Voth, G. A. & Mrksich, M. Reversible control of protein quaternary structure using dynamic covalent linking groups. *In Preparation* (2022).

#### 3.1 Introduction

Dynamic regulation of protein quaternary structure is a hallmark of numerous biological processes such as signal transduction<sup>110,111</sup> and the generation of mechanical force<sup>112,113</sup>. These systems rely on the association and dissociation of protein subunits through the presence or absence of a chemical modification like phosphorylation<sup>114</sup> or an external stimulus such as a cofactor<sup>115</sup>, allosteric ligand<sup>116</sup>, or light<sup>117</sup>. As a result, these reversible systems are a paradigm for the design of sensors, materials, and molecular motors that can respond dynamically to their environment. An ongoing challenge in engineering artificial dynamically controlled protein assemblies is the incorporation of functional units that exhibit well-defined state switching behavior that can also be successfully coupled to the conformational or bound/unbound state of the component proteins. Several examples of such systems have been developed using natural stimulus-responsive protein domains such as calmodulin (CaM)<sup>118</sup> for calcium responsive materials and sensors<sup>119-125</sup> light-oxygen-voltage sensing (LOV) domains for light-driven reversible interactions<sup>123, 126-129</sup>, small molecule-sensing protein domains<sup>130-132</sup> and *de novo*

designed proteins<sup>133-135</sup>. However, their use is often limited to cases where their own inherent structure is irreversibly specific to the dynamics of the system.

In contrast, the use of chemical functionalities or ligands as the basis for the state switching behavior of a dynamic protein assembly not only more closely mimics the majority of reversible interactions found in nature, but it also enables the potential of accessing the vast chemical space of small molecules in their design. While true synthetic allostery of modified protein functionality remains a challenge<sup>136</sup>, chemists and protein engineers have relied on both the mutagenesis of protein surface residues and/or the incorporation of non-canonical amino acids (ncAAs) as the means to introduce chemical groups that respond to various stimuli. Mutagenesis offers the most direct way of introducing this capability; the use of judiciously placed histidine residues has enabled the preparation of pH- and metal-responsive assemblies of naturally oligomeric proteins<sup>137-139</sup>, while the use of cysteines enables reversible behavior based on the redox state of the system<sup>140</sup>. While these approaches are powerful, their synthetic space is limited to the canonical set of amino acids. The incorporation of stimulus-responsive ncAAs into proteins has been used in various examples to reversibly control antibody binding<sup>141</sup> and enzyme activity<sup>142,143</sup>. With improved methods of designing and evolving tRNA synthetase systems<sup>145</sup> that tolerate a wider set of substrates<sup>146-149</sup> and exhibit enhanced ncAA incorporation efficiency<sup>150,151</sup>, these approaches offer enormous potential to engineer stimulus responsive behavior into protein assemblies. Finally, the direct modification of protein surface residues using chemical and chemoenzymatic methods – like labeling nucleophilic lysine and cysteine<sup>152</sup> – have been used to install numerous stimulus-responsive functionalities into protein assemblies, though labeling can often result in heterogeneous products<sup>153</sup>.

these approaches likely offer the possibility of incorporating the widest diversity of substrates into protein structures, but they also require reaction optimization to achieve acceptable yields and selectivity<sup>152</sup>.

Here is where the megamolecule approach can be utilized to synthesize a well-defined state-switching protein sensor. As mentioned and demonstrated in both **Chapter 1** and **Chapter 2**, cutinase<sup>154</sup>, CrabTag<sup>42</sup>, and SnapTag<sup>155,156</sup>, can be used to synthesize non-natural linear<sup>43</sup>, cyclic<sup>39</sup>, and branched<sup>44,109</sup> protein complexes, antibody-mimics<sup>46</sup>, and antibody-enzyme conjugates<sup>69</sup>. We envisioned that megamolecule assembly could synthesize a linear protein scaffold that, when functionalized at distal sites with a pair of ligands (linking groups) that bind intramolecularly upon the addition of an external stimulus, could undergo a quaternary structural change into a cyclic conformation. Upon select removal of this stimulus, the ligands would dissociate and return the scaffold into a linear conformation. Needless to say, we achieved this system, and **Chapter 3** describes the synthesis and characterization of such protein sensors.

We synthesized linear scaffolds of two fusion proteins – mNeonGreen-cutinase-SnapTag (NCS) and mTurquoise2-cutinase-SnapTag (TCS), containing terpyridine (terpy) groups located at the internal cutinase domains. The formation of covalent bidentate complexes of 2,2':6',2''-terpyridines with transition metals is very favorable in aqueous solvents at low concentrations and the kinetics of their formation and dissociation have been experimentally determined<sup>157,161-163</sup>. These two terpy-functionalized triple fusion proteins were covalently linked at their SnapTag domains using one of the previously-demonstrated PEG linkers. The synthesis of this linear scaffold allowed us to use the extent of Forster resonance energy transfer (FRET) between the terminal fluorescent proteins as a reporter of the distance between the two ends of the scaffold. Upon treatment of these scaffolds with varying concentrations of divalent transition metals (e.g.

cobalt, nickel), the two internal terpy groups coordinate the metal, resulting in intramolecular bidentate complexes, which brings the two ends of the protein scaffold closer together in a cyclic manifold. This results in an increase of the observed FRET efficiency (**Figure 3.1A**). Using ethylenediamine tetraacetic acid (EDTA) as a competitive chelator of each transition metal, the terpyridine groups, in theory, dissociate and cause a relaxation of the structure into its original linear conformation. Yet, dissociation was nevertheless metal-dependent. We examined the effect of the transition metal and length of scaffold on the efficiency and observed kinetics of this process, simulated the dynamics and FRET efficiency of the system using coarse-grained modeling, and characterized the structure of the scaffolds using small- and wide-angle X-ray scattering (SAXS / WAXS). We also demonstrated the reversibility of the conformational change over several cycles by treating the scaffold with pulses of cobalt and EDTA in successive reactions.

### 3.2 Results

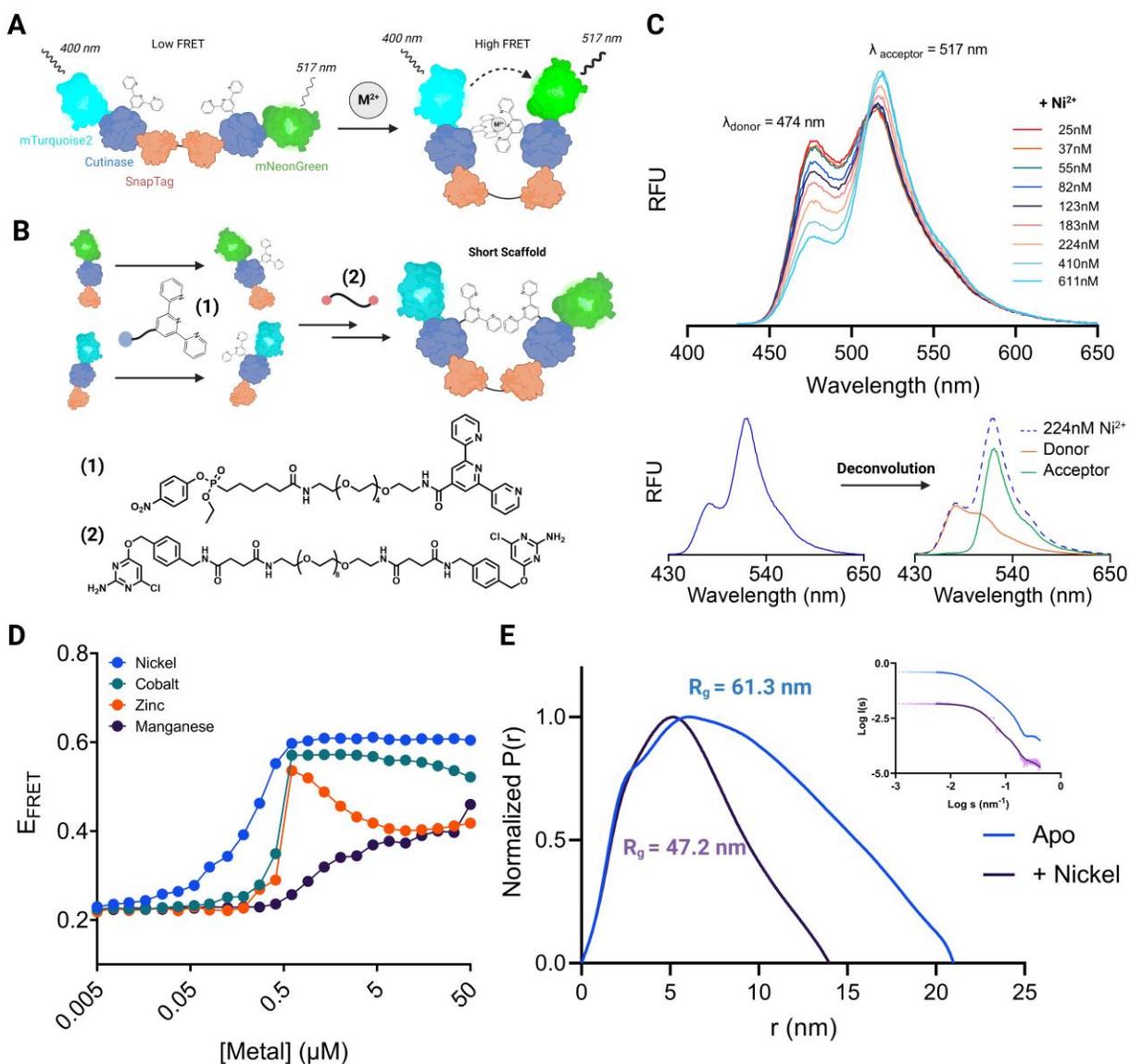
First, these two fusion proteins – TCS and NCS – were similarly expressed in SHuffle T7 Express *E. coli* and both had tobacco etch virus (TEV) cleavable C-terminal 6x-histidine (His) tags. MTurquoise2 and mNeonGreen were chosen as a FRET pair because they are monomeric, do not form an appreciable number of intramolecular dimers because they originate from different species, and because of the reported high efficiency of energy transfer between the two<sup>158</sup>. The fusions were purified the two by cobalt isolated metal affinity chromatography (Co-IMAC) and the C-terminal His-tags were removed by treatment with Ac-TEV protease. The fusions were then further purified by passing the His-tag free proteins over fresh Co-IMAC resin to remove any uncleaved species, and then by size-exclusion chromatography (SEC) in phosphate buffered saline containing 1 mM EDTA (PBS-EDTA). Using solid phase synthesis, a cutinase-reactive ethyl(p-

nitrophenyl)-phosphonate-(hexa)ethyleneglycol-2,2':6',2''-terpyridine (pNPP-EG6-terpy, compound **1**) half linker was made (**Figure 3.1B**). To synthesize the terpy conjugates of each fusion, each unlabeled fusion at 20  $\mu\text{M}$  was incubated with 2 equivalents (eq.) of pNPP-EG6-terpy for 3 hours and purified using SEC in PBS-EDTA. To prepare heterodimers of the two fusion proteins, we first treated TC'S at 5  $\mu\text{M}$  with 5 eq. of an EG10-bis-(benzyl-O<sup>6</sup>-chloropyrimidine, EG10-bis-(benzyl-CP), compound **2**) linker for 15 min to create a monofunctionalized fusion (**Figure 3.1B**), TC'S-EG10-benzyl-CP, containing a pendant SnapTag reactive group, removed the excess linker by SEC and then treated the resulting product with 1 eq. of NC'S for an hour at 20  $\mu\text{M}$ . The bis-terpy functionalized heterodimer T-C<sup>terpy</sup>-S-S-C<sup>terpy</sup>-N (TC'S-SC'N) was then purified by SEC in PBS-EDTA. This linear protein assembly was known as the *short scaffold*, and general synthesis procedure described above can be seen in **Figure 3.1B**. Using the same process, we also prepared a scaffold containing no terpy groups (TCS-SCN) and two containing only one terpyridine functionality (TC'S-SCN and TCS-SC'N) as negative controls to ensure intramolecular bidentate coordination only occurs when both terpy groups are present within the scaffold. Further, ESI-MS (**Figure 3.2**) and SDS-PAGE (**Figure 3.3**) confirmed the masses of our scaffold and its constituent parts.

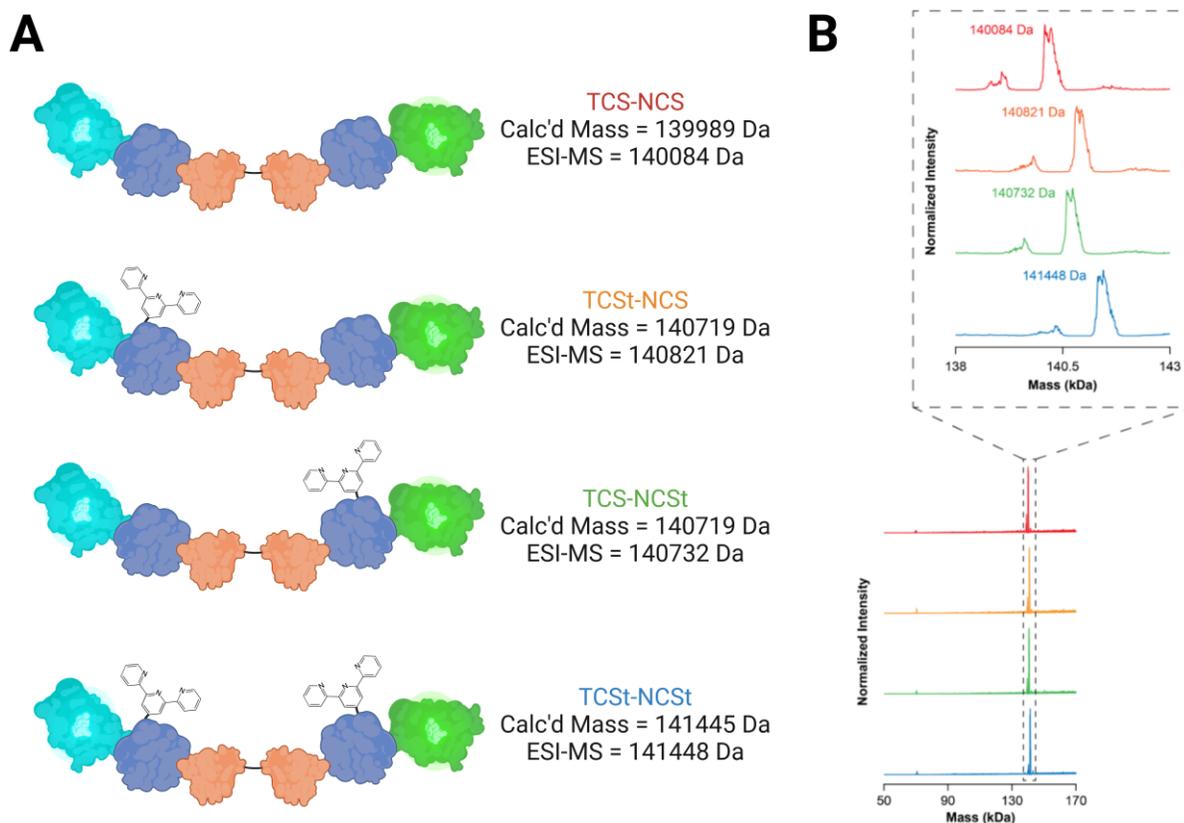
Prior to experimentation, each scaffold was buffer exchanged into HEPES containing a small amount of EDTA (10  $\mu\text{M}$ ). We used this as the reaction medium because HEPES is compatible with high concentrations of transition metals<sup>159</sup> and because in initial experiments, omitting a low concentration of EDTA resulted in products that showed a high degree of energy transfer, presumably due to the presence of trace iron contamination in the buffer from the FPLC.

To determine the FRET response of the system to a metal stimulus, I treated the fully assembled short scaffold, TC'S-SC'N, at 100 nM with concentrations of Ni<sup>2+</sup>, Co<sup>2+</sup>, Zn<sup>2+</sup>, and Mn<sup>2+</sup>

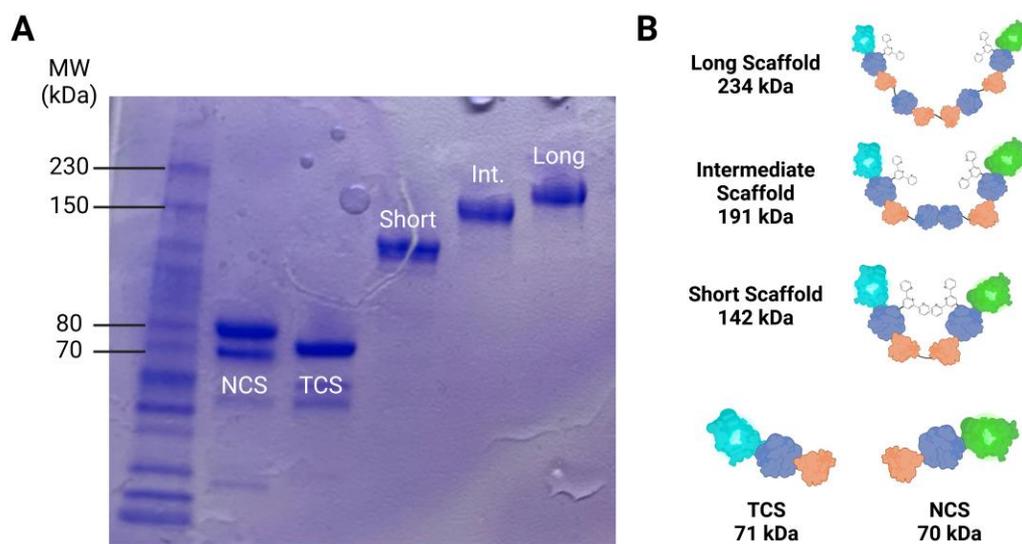
from 0.005 to 50  $\mu\text{M}$  for 24 hr (**Figure 3.1D**). I excluded  $\text{Fe}^{2+}$  and  $\text{Cu}^{2+}$  from our study due to the instability of  $\text{Fe}^{2+}$  in aqueous solutions as well as  $\text{Cu}^{2+}$  because Justin had previously observed significant fluorescence quenching in experiments using copper. I then took fluorescence emission spectra of each reaction by exciting the donor fluorophore at 400 nm (mTurquoise2  $\lambda_{\text{ex/em}} = 434 / 474$  nm) and then scanning the region between 430 and 650 nm (mNeonGreen  $\lambda_{\text{ex/em}} = 506 / 517$  nm) using a Synergy plate reader. I chose these reaction conditions because in previous work, our lab demonstrated that cyclization reactions performed on a scaffold of similar length at concentrations less than 500 nM suppressed intermolecular reaction pathways to an undetectable level. Further, the signal intensity of the resulting spectra fell within an appropriate dynamic range<sup>107</sup> of the plate reader. I found that the resulting spectra did not change appreciably after 24 hours, so all further reactions were allowed to proceed for this period. An example of the raw data for the reactions with nickel is shown in **Figure 3.1C**. I then took the raw spectra, deconvoluted the donor and acceptor peaks using a script written in MATLAB by Blaise, integrated areas under each curve and used these integrated intensities to calculate observed FRET efficiencies for each reaction. In these data, I observed a clear increase in the FRET efficiency with increasing metal concentration, although efficiency did tend to decrease at higher concentrations of zinc and cobalt. As a result, I anticipated cyclization curves whose response would track with the known formation constants of the various terpy-metal complexes ( $\beta [\text{M}(\text{terpy})_2]^{2+} = \text{Ni}^{2+} > \text{Co}^{2+} > \text{Zn}^{2+} > \text{Mn}^{2+}$ )<sup>160</sup>. However, after experimentation, I found that the concentration required to reach peak FRET efficiency (i.e. bidentate cyclization) were similar ( $\sim 500\text{nM}$ ) in all cases except for  $\text{Mn}^{2+}$ , which had a much weaker response. I will discuss the differences in the shape of the curve/response further in **Chapter 3**.



**Figure 3.1.** **A)** Schematic of terpyridine-functionalized megamolecule scaffold. Addition of divalent transition metal creates bidentate coordination by terpyridine (terpy) groups, demonstrating an increase in FRET efficiency; **B)** Synthesis of the short scaffold. mNeonGreen-cutinase-SnapTag (NCS) and mTurquoise2-cutinase-SnapTag (TCS) triple fusions are reacted with pNPP-terpy small molecules before the triple fusions are conjugated by a homobifunctional benzylguanine linker; **C)** Increase in FRET efficiency can be seen by increase in mNeonGreen emission with corresponding decrease in mTurquoise2 emission. Here, 100nM short scaffold was treated with  $\text{Ni}^{2+}$  from 25nM to 611nM and plotted accordingly. FRET efficiency is calculated through deconvolution of the emission curves of the two fluorophores; **D)** Treatment of the short scaffold with four different transition metals from 5nM to 50 $\mu\text{M}$  demonstrates metal-dependent cyclization response; **E)** Small-angle X-ray scattering (SAXS) analysis of short scaffold with and without addition of  $\text{Ni}^{2+}$  shows that the radius of gyration ( $R_g$ ) decreases upon addition of metal, demonstrating that cyclization and no large intermolecular species occur.

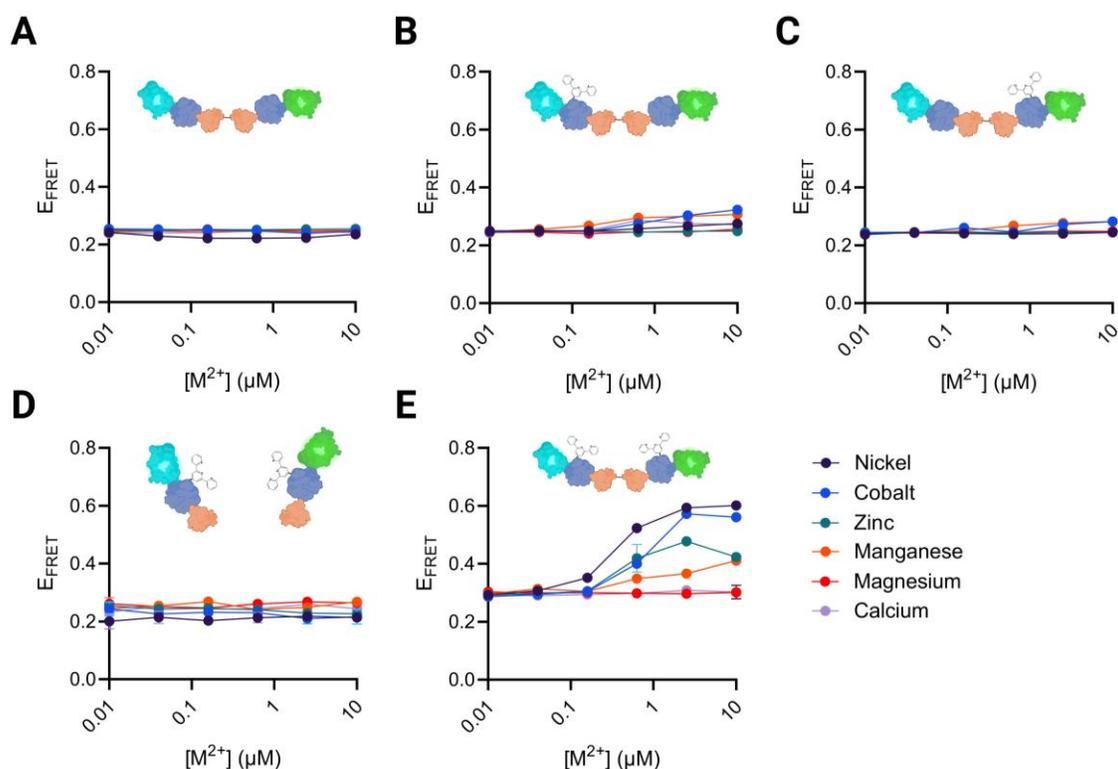


**Figure 3.2.** A) Predicted masses of the short scaffold with and without one or both terpy groups correlate with experimental data obtained by ESI-MS



**Figure 3.3** Both building blocks – TCS and NCS – as well as each scaffold length (intermediate and long generated later in **Chapter 3**) separated by predicted mass via SDS-PAGE.

To verify that the metal-dependent increase in the FRET signal was in fact due to the intramolecular bidentate binding of the terpy groups and not some other feature of the system, I then performed similar reactions on the aforementioned negative control scaffolds, TCS-SCN, TC<sup>t</sup>S-SCN, and TCS-SC<sup>t</sup>N; and on an equimolar mixture of TC<sup>t</sup>S and NC<sup>t</sup>S using concentrations of Ni<sup>2+</sup>, Co<sup>2+</sup>, Zn<sup>2+</sup>, and Mn<sup>2+</sup> from 0.01 to 10 μM (and also the divalent cations Mg<sup>2+</sup> and Ca<sup>2+</sup> over the same concentration range as a control for any non-specific cross linking of protein surface residues). In these experiments, I did not observe any FRET response in the terpy-free scaffold nor with calcium and magnesium and a very weak increase in FRET in reactions with mono-functionalized variants across the various concentrations of the three transition metals (**Figure 3.4**). We did, however, observe a small but noticeable increase in the signal in an equimolar mixture of the two terpy-functionalized fusions treated with nickel but of much lower magnitude than that observed for the linked heterodimer. From these data, we conclude that the metal-induced linking reactions are specific to the terpy functionalities and that the magnitude of the metal-dependent increase in FRET is due to the presence of both terpy groups in the scaffold.

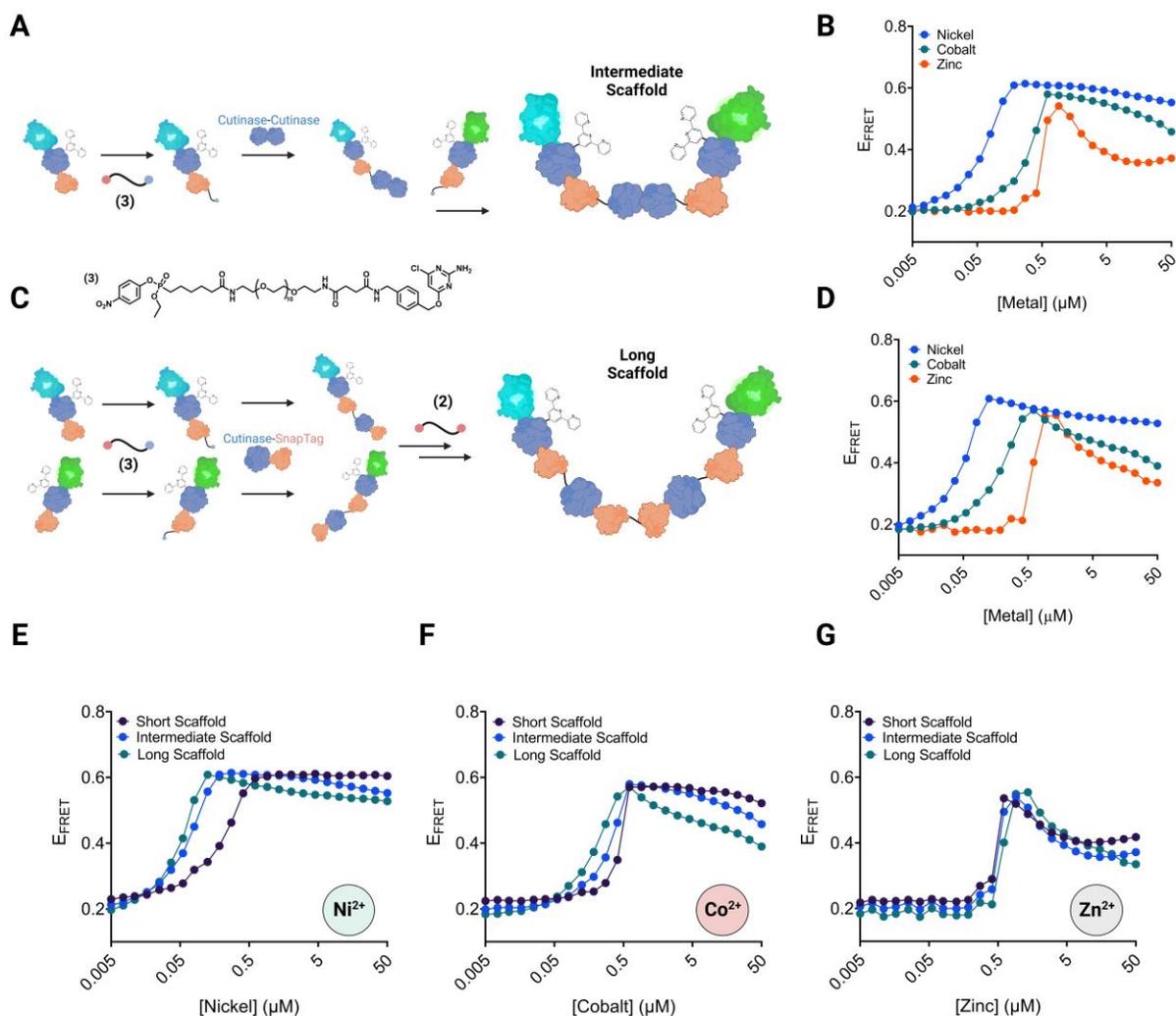


**Figure 3.4** Control short scaffolds demonstrate that increased FRET efficiency does not occur with zero (A) one (B & C) terpyridine ligand, nor does it occur without SnapTag-SnapTag conjugation (D); only with all constituent parts does the scaffold cyclize upon addition of divalent transition metals (E).

While the FRET data indicated that the two fluorophores were being brought closer together in space upon treatment with metal, it did not rule out the possibility that some intermolecular product could give rise to the observed signal. If the products of the reaction are indeed cyclic, they should be more compact than the linear starting materials. To determine and compare the size and shape of the starting material and products, SAXS / WAXS scattering profiles of the apo (+EDTA) and nickel-treated scaffold were obtained. To prepare the nickel-bound scaffold, the short scaffold was diluted to 500 nM in HEPES-EDTA and treated with 10 eq.  $\text{Ni}^{2+}$  for 24 hr. After this period, it was concentrated to  $\sim 4$  mg / mL and five 2-fold serial dilutions were made for test samples. Nickel was chosen because the FRET response curves indicated this metal showed the least sensitivity to excess metal concentrations and also because it has the highest

reported formation constant for the bidentate complex among all the metals we tested (**Figure 3.1D**). For the apo scaffold, ~7 mg / mL sample of the short scaffold in HEPES containing 1 mM EDTA was to ensure any trace metal would be sequestered by the large excess of the chelating agent. Then, five 2-fold serial dilutions of this scaffold were made as test samples. Experimentation was performed by a collaborator, Dr. Steven Weigand, at Argonne National Laboratory. The buffer-subtracted scattering profiles of the apo and nickel-bound complexes at ~ 1 mg / mL are shown in **Figure 3.1E** (inset). The software package DATGNOM was then used to generate pair distance distribution curves, known as  $P(r)$ , for each data set over the indicated range(s). The scattering profiles at these concentrations were chosen for each because they yielded the highest quality fits using DATGNOM.  $P(r)$ , in this case, reports on the frequency of distances between pairs of atoms in each protein scaffold and therefore reveals information about the three-dimensional shape of each molecule.

In **Figure 3.1E**, the  $P(r)$  curves for the apo and nickel bound scaffolds both show the highest frequency of distances around 5 nm – presumably near the average length of the interdomain spacing between the protein subunits yet show much different features at longer distances with the apo scaffold having a long trailing distribution of frequencies up to ~20 nm and the nickel bound species showing a sharper decline in frequency terminating near ~15 nm. Values of the radius of gyration,  $R_g$ , obtained using DATGNOM were 6.13 nm for the apo scaffold and 4.72 nm for the nickel bound species. Taken together, these data show that the apo scaffold is more extended and flexible than the nickel-bound scaffold, suggesting that the reaction products are indeed a more compact cyclic form of the protein and do not form oligomeric products via intermolecular pathways.



**Figure 3.5** A) Synthesis of intermediate length scaffold that has an extra cutinase-cutinase fusion protein between the NCS and TCS domains seen in the short scaffold; B) FRET efficiency of the intermediate scaffold upon addition of  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Zn}^{2+}$  from 5nM to 50 $\mu\text{M}$ ; C) Synthesis of long scaffold that has two cutinase-SnapTag fusion proteins between the NCS and TCS domains seen in the short scaffold; D) FRET efficiency for the long scaffold upon addition of  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Zn}^{2+}$  from 5nM to 50 $\mu\text{M}$ ; E-G) FRET efficiency plotted by metal species demonstrates both scaffold and metal-specific relationships.

Because the cyclization reaction occurs intramolecularly, the efficiency of the reaction should be affected by the effective molarity of the two terpy groups in the protein scaffold and possible cyclic strain in the products. To examine these effects, two longer megamolecule scaffolds were synthesized -  $\text{TC}^{\text{I}}\text{S-CC-SC}^{\text{I}}\text{N}$ , an *intermediate scaffold* that separates the two terminal triple fusions by a cutinase-cutinase (CC) fusion protein and  $\text{TC}^{\text{I}}\text{S-CS-SC-SC}^{\text{I}}\text{N}$ , a *long scaffold* that

further increases the distance between the two ends by inserting two cutinase-SnapTag (CS) domains between the terminal triple fusions. These syntheses are outlined in **Figure 3.5A & C**. For brevity, each of the products described were purified by SEC in PBS-EDTA following reaction. The intermediate scaffold was prepared by first reacting TC'S at 20  $\mu$ M with 2 eq. of compound **3**, a cutinase-SnapTag linker described in previous work and shown in **Figure 3.5A**, for 15 min. The resulting product, TC'S-EG12-pNPP, was then concentrated to 10  $\mu$ M, and treated with 10 eq. cutinase-cutinase fusion protein for 5 hr. A large excess of CC was used in the reaction to favor the formation of the monofunctionalized product TC'S-CC. The mNeonGreen triple fusion, NC'S, was then treated with 2 eq. of linker **3** to give NC'S-EG12-pNPP, and then reacted with TC'S-CC overnight to yield the final product TC'S-CC-SC'N. To prepare the longest scaffold, TC'S-CS-SC-SC'N, TC'S and NC'S were treated in separate reactions at 20  $\mu$ M with linker **3** for 15 min. The products of both reactions, TC'S-EG12-pNPP and NC'S-EG12-pNPP, were then concentrated to 20  $\mu$ M and treated with 2 eq. of a cutinase-SnapTag fusion protein (CS) for 5 hr to yield TC'S-CS and NC'S-CS. Then, TC'S-CS was reacted at 5  $\mu$ M with 5 eq. of the Snap-Snap linker **2** for 15 min to yield TC'S-CS-EG10-benzyl-CP, and resulting product was treated with 1 eq. of NC'S-CS for an hour at 20  $\mu$ M to give the final product, TC'S-CS-SC-SC'N. SDS-PAGE, again in **Figure 3.3**, confirms their masses.

To determine the response of the scaffold of intermediate (TC'S-CC-SC'N) and longest (TC'S-CS-SC-SC'N) length to metal, I again treated each with concentrations of Ni<sup>2+</sup>, Co<sup>2+</sup>, and Zn<sup>2+</sup> from 0.005 to 50  $\mu$ M in separate reactions for 24 hr, took fluorescence spectra, and determined the FRET efficiency for each reaction as I did for the short scaffold. These data in **Figure 3.5B and D** show that as the length of the scaffold increases, the difference in the amount of each metal required to reach the cyclic state becomes more pronounced, especially in the case

of zinc and cobalt. In other words, lengthening the scaffold increases the ability to effectively discriminate between the three metals and highlights the benefit of being able to vary structure to modulate response to a given analyte. When I compare the responses of the three scaffolds versus each metal individually in **Figure 3.3E-G**, I discovered that the length has a large effect on the efficiency of reaction. There is a large difference with nickel going from the shortest to the intermediate length, a smaller overall effect among the three lengths with cobalt, and almost no effect across the three scaffolds in the case of zinc. This behavior points to differences in the reported kinetics of the reactions of each metal with free terpyridine groups (**Table 3.1**<sup>157,161-163</sup>) and also suggests some influence of the scaffold itself on the observed efficiency, specifically in the case of nickel in the shortest scaffold.

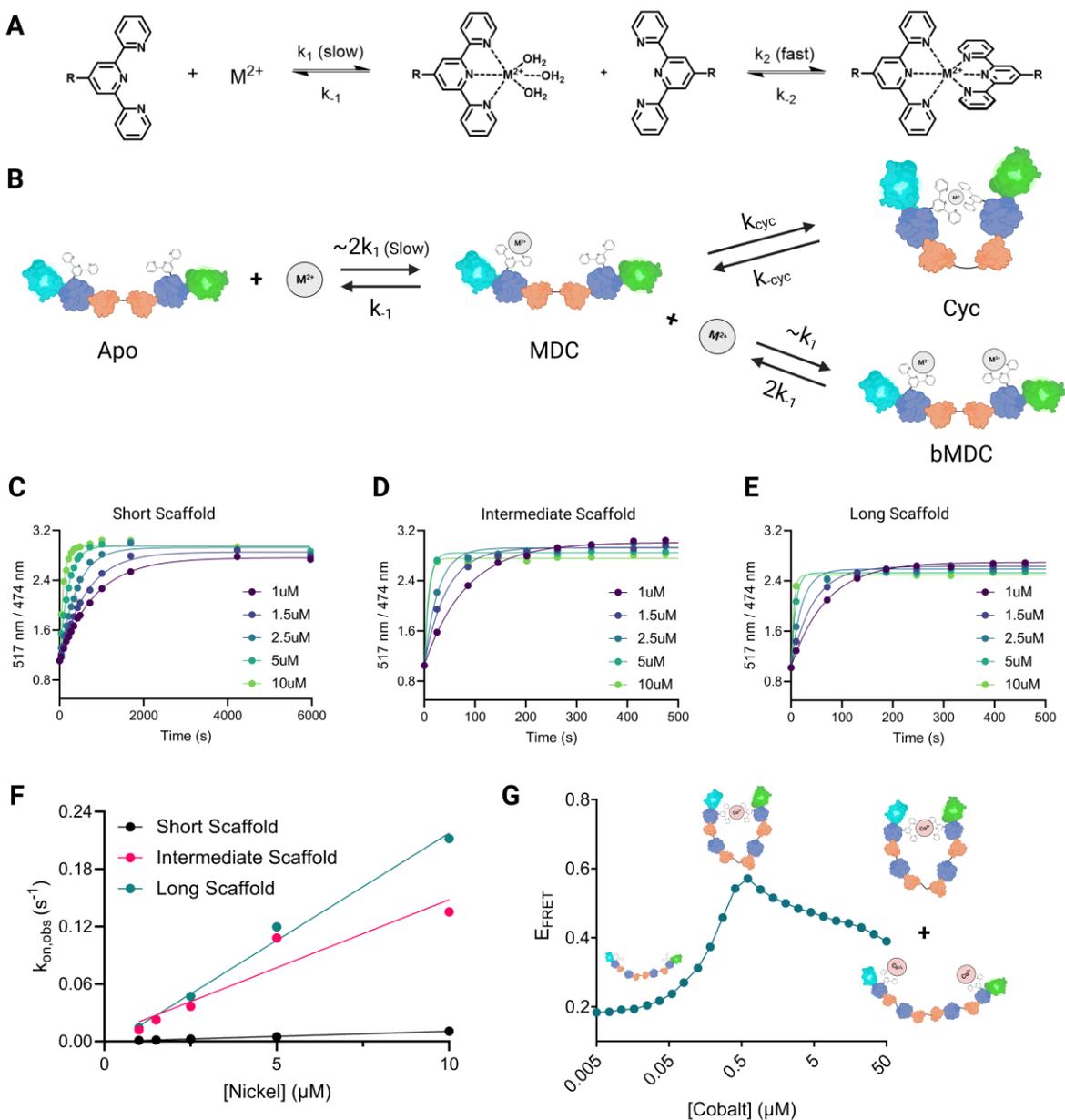
**Table 3.1.** Kinetic and Thermodynamic Parameters for Formation and Dissociation of Transition Metal Complexes of Terpyridine and EDTA at 25° C<sup>157, 161-163</sup>

Complex	$k_1$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{-1}$ (s <sup>-1</sup> )	$k_2$ (M <sup>-1</sup> s <sup>-1</sup> )	$k_{-2}$ (s <sup>-1</sup> )	log $K_f$	log $\beta$
Ni(Terpy) <sup>2+</sup>	1.4 x 10 <sup>3</sup>	1.6 x 10 <sup>-8</sup>	-	-	10.7	-
Co(Terpy) <sup>2+</sup>	2.5 x 10 <sup>4</sup>	1.0 x 10 <sup>-4</sup>	-	-	8.4	-
Zn(Terpy) <sup>2+</sup>	1.3 x 10 <sup>6</sup>	1.3	-	-	6.0	-
Ni(Terpy) <sub>2</sub> <sup>2+</sup>	-	-	2.0 x 10 <sup>5</sup>	1.6 x 10 <sup>-6</sup>	-	21.8
Co(Terpy) <sub>2</sub> <sup>2+</sup> <sup>a</sup>	-	-	5.0 x 10 <sup>6</sup>	6.3 x 10 <sup>-4</sup>	-	18.3
Ni(EDTA) <sup>2-</sup> <sup>b</sup>	1.1 x 10 <sup>4</sup>	(4.2 x 10 <sup>-15</sup> ) <sup>d</sup>	-	-	18.4	-
Co(EDTA) <sup>2-</sup> <sup>c</sup>	1.0 x 10 <sup>7</sup>	(3.6 x 10 <sup>-10</sup> ) <sup>d</sup>	-	-	16.5	-
Zn(EDTA) <sup>2-</sup>	-	-	-	-	16.5	-

<sup>a</sup> Reaction Performed at 5° C. <sup>b</sup> Reaction performed at 0° C in acetate buffer. <sup>c</sup> Reaction performed in acetate buffer. <sup>d</sup> Values were calculated using the relationship  $k_{off} = k_{on} / K_f$ . Reactions performed in water unless stated otherwise.

Justin and I proposed a kinetic scheme for the reactions of the short scaffold with metal in **Figure 3.6**. Free in solution, the reactions of transition metals with terpyridine proceed first through a slow association step to form the monodentate complex,  $M(\text{terpy})^{2+} \bullet \text{H}_2\text{O}$ , with rate constant  $k_1$  followed by rapid association of this intermediate with another equivalent of terpy to form the bidentate complex  $M(\text{terpy})_2^{2+}$  with rate constant  $k_2$ . Each process has a corresponding

dissociation rate constant:  $k_{-1}$  for dissociation of the monodentate complex and  $k_{-2}$  for dissociation of the bidentate complex (**Figure 3.6A**). In our system, we assume that formation of the protein-bound monodentate complex proceeds similarly with rate constant  $\sim 2k_1$ . This intermediate, however, can then go on to form the cyclic product with rate constant  $k_{\text{cyc}}$ , a bis-monodentate metal complex with rate constant  $\sim k_1$  or intermolecular products. From the control scaffolds and the SAXS data, we have determined that under these reaction conditions, intermolecular pathways are mostly suppressed, therefore most species present in the reactions are the apo scaffold, the monodentate complex, the cyclic product, or the bis-monodentate complex at higher concentrations with select scaffolds (**Figure 3.6B**). Using this scheme and the known rates of formation and dissociation of the metal-terpy complexes, Justin and I could then begin to understand the response curves of the various scaffolds.



**Figure 3.6** **A)** Kinetic scheme of the reactions of transition metals with free terpyridines; **B)** Kinetic scheme for short scaffold cyclization upon addition of divalent transition metal; **C-E)** Cyclization kinetics of each scaffold length reveal a 10-fold slower rate of formation for the short scaffold compared to the intermediate and long scaffolds; **F)** Second order rate constants for cyclization reveal length dependence; **G)** FRET efficiency of the longest scaffold upon addition of cobalt demonstrates concentration-dependence towards cyclized and bis-monodentate complex.

For  $\text{Ni}^{2+}$  at protein concentrations of 100 nM and similar metal concentrations, the rate of formation of the monodentate complex would be on the order of  $10^{-4} \text{ s}^{-1}$  or  $\sim 10^4$  faster than the dissociation rate (**Table 3.1**). For the free ligands (i.e. **3.6A**),  $k_2$  is  $\sim 200\times k_1$  and its corresponding dissociation constant is negligible in comparison. If these relationships hold in the context of the protein scaffold, essentially any monodentate complex would lead to rapid formation of a persistent cyclic bidentate product. Furthermore, the increased effective molarity of the two terpy groups or electrostatic interactions with specific residues may accelerate the cyclization rate. For  $\text{Ni}^{2+}$ , this explains the smooth titration of the signal from low to high FRET at substoichiometric metal concentrations. This does not, however, explain why for the intermediate and long scaffold, the signal peaks at approximately 1:1 stoichiometry while the shortest scaffold requires a higher metal concentration to reach the same state. In contrast,  $\text{Co}^{2+}$  elicits a sharper response. At 100 nM, the rate of formation of the monodentate complex is only  $\sim 25\times$  faster than the rate of dissociation and  $k_2$  is again  $\sim 200\times k_1$ . In similarity to nickel, this indicates that any monodentate cobalt complex formed would rapidly lead to a cyclic product, but that a higher concentration of metal would be required to form the monodentate intermediate. This may explain why among all scaffold lengths, the signal for cobalt peaks at  $\sim 5$  eq. metal. For  $\text{Zn}^{2+}$ , formation and dissociation of the monodentate complex are orders of magnitude faster than for nickel and cobalt and consequently may explain its very sharp response to increasing metal concentration up to  $\sim 5$  eq. At higher metal concentrations, I began to observe a decrease in the FRET efficiency and this effect is more pronounced as the length of the scaffold and lability of the metal increases ( $\text{Zn}^{2+} \gg \text{Co}^{2+} > \text{Ni}^{2+}$ )<sup>157</sup>. Presumably, this decrease arises from the formation of linear bis-monodentate metal complexes (bMDC) of the megamolecules, as suggested in **Figure 3.6B**. A larger fraction of these products are formed in the longer scaffolds potentially due to an increase in  $k_1$  with respect

to  $k_{cyc}$  at higher metal concentrations (and lower effective terpy molarity with the groups further apart).

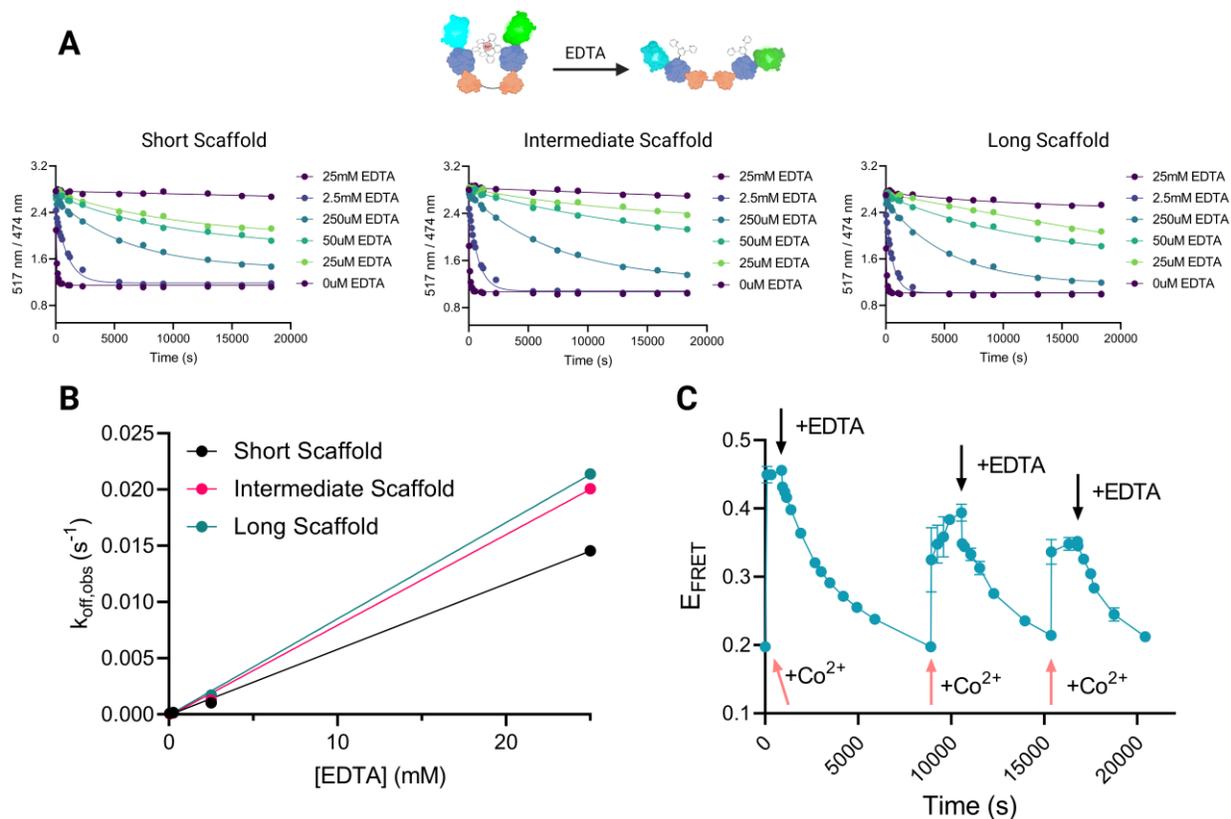
To investigate the effect of scaffold length on the rate of  $Ni^{2+}$ -MDC formation, I took each protein at 100 nM, treated each in separate reactions with concentrations of  $Ni^{2+}$  from 1 – 10  $\mu$ M and measured the ratio of the donor and acceptor fluorescence intensities over time. These data, shown in **Figures 3.6C-E**, were then fit to first order exponentials to obtain pseudo-first order rate constants and plotted vs nickel concentration to obtain second-order rate constants for the process (**Figure 3.6F**). The determined values of  $k_1$  to be  $5.4 \times 10^2$ ,  $7.1 \times 10^3$ , and  $1.1 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ , for the short, intermediate, and long scaffold respectively, revealing that as the scaffold length increases, so does the rate of  $Ni^{2+}$ -MDC formation. A plausible explanation for this behavior is that in the shortest scaffold, the terpy groups are located closer to the core of the molecule and therefore less solvent accessible than in the longer variants. Interestingly, the experimentally determined  $k_1$  for nickel the longest scaffold is  $\sim 10$ x greater than the reported value for the free ligands. The possible explanations for this behavior are numerous but, again, may arise from a rapid pre-concentration of the metal via a protein surface residue such as histidine. The buffers used in our reactions contained a final EDTA concentration of 500 nM. Using the rate constants for  $Ni^{2+}$ -MDC formation and the reported formation rates for  $Ni(EDTA)^{2-}$  listed in **Table 3.1**, we can explain the response curves for nickel at sub-stoichiometric metal concentrations. At 100 nM protein and 500 nM EDTA, formation of the EDTA complex is 50, 5, and  $\sim 3$  fold faster than formation of  $Ni^{2+}$ -MDC for the shortest, intermediate, and long scaffolds, respectively and therefore accounts for the observed shifts in reaction efficiency. In other words, in reactions with the shortest scaffold and nickel, EDTA must be nearly completely titrated before reaching maximum yield of the cyclic product. Reactions with cobalt and zinc were too rapid to obtain similar rate data for formation of

$M^{2+}$ -MDC, yet for cobalt, dissociation of  $\text{Co}(\text{terpy})^{2+}$  is 10-fold faster than that of nickel, while its reaction with EDTA is nearly a thousand-fold faster. The increased lability of cobalt in its reactions with terpy and its accelerated rate with EDTA likely explains the increased sharpness of its response curves, the weaker dependence of length on reaction efficiency, as well as why the maximum yield occurs for each species at 500 nM. Zinc, being the most dynamic of all three metals, consequently, exhibits a very sharp response and almost no dependence on the length of the scaffold on reaction efficiency.

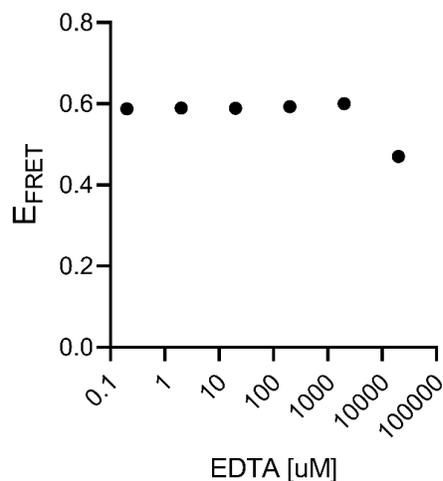
Except in the reactions of nickel with the short scaffold, the FRET efficiency of the products decreases as the concentration of metal exceeds the value at which the maximum yield occurs. We attribute the decreased FRET to the formation of the linear bis-monodentate metal complexes ( $M^{2+}$ -bMDC) of the scaffolds. As the length of the molecule increases, the effective molarity of the two terpy groups decreases and therefore also the rate of cyclization. As we showed previously, the rate of  $M^{2+}$ -MDC formation increases with length and this rate should also increase going from nickel to zinc. Therefore, as the rate of cyclization decreases and the rate of metal-terpy reaction increases, there should be a greater propensity to form  $M^{2+}$ -bMDC than the cyclic product. This is borne out in the behavior of the response curves in **Figures 3.5E-G**. Nickel shows a very gradual decrease in the signal that becomes steeper with scaffold length, and cobalt shows an even more pronounced decline in the efficiency with length. Zinc appears to show complex behavior past 1:1 stoichiometry but this has been reported elsewhere<sup>164</sup>. I summarized our findings about the distribution of products along the longest scaffold response curve with cobalt in **Figure 3.6G**, as this scaffold / metal combination formed the highest amount of bMDCs.

To examine the effect of length on the dissociation rate – and to interrogate the possibility of cyclic strain of each scaffold – I took each scaffold at 100 nM, reacted each with cobalt at 5  $\mu\text{M}$

for 24 hr, then treated each with concentrations of EDTA from 0.05 to 50 mM and measured the ratio of the donor and acceptor fluorescence over time (**Figure 3.7A**). These data were then fit to a first order exponential decay and plotted vs EDTA concentration to obtain second order rate constants for decomposition (**Figure 3.7B**). I chose cobalt to measure the rates of dissociation because reactions with zinc were too rapid to measure and because reactions with nickel were too slow, even at concentrations of EDTA > 100 mM. In fact, even after 24 h, the nickel-scaffold complex does not dissociate (**Figure 3.8**). Here, we assume that dissociation of the bidentate terpy complex is required prior to reaction with EDTA and that this process is rate-limiting. From this experiment, we determined values of  $k_{\text{off}}$  to be  $0.3 \text{ M}^{-1}\text{s}^{-1}$  for the shortest scaffold and  $0.4 \text{ M}^{-1}\text{s}^{-1}$  for both the intermediate and longest scaffold. The similar rates of dissociation suggest that there is not significant ring strain involved, as ring strain should – in theory – lead to faster dissociation kinetics if the linear form was more energetically favored.



**Figure 3.7** **A)** Here, 100 nM short scaffold was treated with 2.5  $\mu\text{M}$  cobalt and let sit overnight. Then, various concentrations of EDTA were added to this complex. Similar to the rate of formation experiments, FRET efficiencies could be quantified, demonstrating that scaffold length has negligible effect on scaffold dissociation (**B**); **C)** Scaffold reversibility could be demonstrated by cycling the short scaffold with cobalt and EDTA.



**Figure 3.8** 100 nM short scaffold was treated with 500 nM nickel and let sit overnight. Then, various concentrations of EDTA were added to this complex and also let to sit overnight. From here (above), FRET efficiencies could be quantified, demonstrating that the bidentate complex does not noticeably dissociate.

Finally, I wanted to demonstrate that our scaffold formation – dissociation with divalent transition metals/EDTA could be a reversible process. To do this, I cyclized 100 nM of the short scaffold with 2.5  $\mu\text{M}$  cobalt and attempted to quantify FRET efficiency by estimating the FRET ratio within the first minute of total cyclization. Then, I added 1 mM EDTA and observed dissociation similar to what I did in **Figure 3.7A**. However, after full dissociation, I added 1  $\mu\text{M}$  of cobalt back into the well to fully titrate the EDTA added with enough left over from the original cyclization to re-coordinate the scaffold into a bidentate complex. This was achieved, and the rapid kinetics of cobalt association / slow kinetics of dissociation can be seen in **Figure 3.7C**. Overall decrease in FRET efficiency at subsequent steps may be due to the increased volume within the well itself, mildly diluting the 100 nM scaffold in the process.

### 3.3 Conclusions

Here, I demonstrated synthesis and high-quality characterization of a quaternary-scale protein switch that undergoes bidentate coordination upon addition of a divalent transition metal stimulus. Fundamentally, length of the megamolecule scaffold affected the rate of association, demonstrating how emphasizing a modular design of a well-defined state switch can be useful. Furthermore, we excitingly discovered that the kinetic – and “equilibrium” – data generated perfectly matched 50-to-60 year old inorganic chemistry papers from Japan with coordination of both free EDTA and terpyridine in solution with each transition metal. Our proposed mechanism of formation, in our opinion, is robust and the rates are backed by this data. While this work may represent more of a proof-of-concept protein switch, than to say something like a full-blown biologic diagnostic, it is not a far stretch of the imagination to surmise what a diagnostic within this framework could look like, as the terpy groups could be exchanged with any small molecule of interest for design purposes as long as it did not affect the pNPP’s reaction with the cutinase active site.

### 3.4 Experimental

**Protein Expression and Purification.** All constructs were cloned and expressed in *E. Coli*. Fusions to megamolecule building block enzymes were cloned into pET-28b(+), a kanamycin-resistant and IPTG inducible expression vector, via Golden Gate compatible restriction enzyme sites (BsaI), containing a TEV-cleavable C-Terminus 6xHisTag for purification. Fusions to megamolecule building blocks using a 12-mer XTEN linker (SGSETPGTSESA). For all cloning and plasmid propagation, NEB DH5a cells were used. Sequence confirmed plasmids containing the megamolecule fusions were transformed into SHuffle® T7 Express Competent *E. coli*. Cultures were grown in 2x YT culture medium containing 50ug/ml of kanamycin. Cultures were

inoculated with a stab of cells and cultured overnight at 30C shaking at 250RPM in 5mL growth medium. After 12 hours, cultures were transferred into 2L baffled flasks containing 500mL growth medium with 1:1000 antibiotic and cultured in an Innova 44R (New Brunswick Scientific) incubator. Once cultures reached an OD600 of 0.8, they were induced with 2.5mM (final concentration) of IPTG and cultured 16 hours at 20C. The following day, cultures were pelleted at 5000RPM and were resuspended with lysis buffer (2XPBS containing DNase) and sonicated on ice for 2min. To remove cell debris from lysates, samples were centrifuged at 10K RPM for 10min, and supernatants were transferred to 50ml Kontes Flex columns (Kimbal Kontes Glassware) and were incubated with HisPur cobalt resin to capture 6xHis Tagged proteins at 4C on a shaker for 2hours. Resin was washed three times with 1xPBS to remove any weakly bound proteins and desired products were eluted with 150mM imidazole. Samples were treated with TEV protease and then further purified by size exclusion chromatography using an AKTA FPLC system on a HiLoad 16/600 Superdex column with PBS + 0.05% NaN<sub>3</sub>.

**Electrospray Ionization Mass Spectrometry.** ESI spectra were obtained on an Agilent 6230 LC/TOF mass spectrometer equipped with an Agilent 1260 Infinity Binary Pump. Samples were prepared at 1uM in nuclease-free water. A volume of 7uL sample was injected into a 1cm C18 guard column (Waters) and eluted using a gradient from 0% to 100% acetonitrile over 7.5 minutes and 0.1% formic acid in water at a flow rate of 0.4 mL/min. Data were analyzed using the Agilent MassHunter BioConfirm software. The protein molecular weight was obtained using a maximum entropy deconvolution calculation.

**Transition Metal Stocks:** Transition metals (Sigma-Aldrich) were first diluted to 100mM in deionized H<sub>2</sub>O. From here, 1:10 dilutions were made in 1X HEPES down to 100μM. From 100μM, ~33% dilutions were made (in 1X HEPES) such that a 24-log step titration of each metal was achieved down to 10nM. This gave the following concentrations - 100μM, 67μM, 45μM, 30μM, 20μM, 13.5μM, 9μM, 6μM, 4μM, 2.7μM, 1.8μM, 1.2μM, 820nM, 550nM, 367nM, 246nM, 165nM, 111nM, 74nM, 50nM, 33nM, 22nM, 15nM, 10nM.

**Scaffold Equilibrium Binding Experiments:** To deduce the experimental EC<sub>50</sub> of cyclization is for each metal/scaffold combination, scaffolds were first diluted to 200nM in 1X HEPES buffer with 1μM ethylenediaminetetraacetic acid (EDTA). This 5x excess of EDTA sequestered trace metal binding to terpyridine from scaffold purification by FPLC. Then, 75μL of this dilution was added to each well in a 96-well black, opaque polystyrene plate (Costar®). Then, 75μL of each bivalent transition metal of each concentration generated from 100μM to 10nM – manganese (Mn<sup>2+</sup>), nickel (Ni<sup>2+</sup>), cobalt (Co<sup>2+</sup>), and zinc (Zn<sup>2+</sup>) was added to each well. Each plate was then sealed with Parafilm and scaffolds were allowed to coordinate with each metal overnight. This gave final well concentrations of 100nM scaffold, 500nM EDTA, and then 50μM to 5nM of each metal. Then, plates were scanned on a Synergy H1 (BioTek) plate reader via spectral scan at 400nm from 430nm to 650nm with a 1nm step size. Fluorescence data was then transferred over to Excel where FRET efficiency could be calculated via MATLAB and subsequently plotted in GraphPad Prism.

**Scaffold Kinetic On-Rate Binding Experiments:** Due to its measurable on-rate kinetics, nickel was chosen to model coordination of both terpyridines for each scaffold (cobalt was too quick).

Here, similar stocks and 96-well plates were used as above – 200nM scaffold, 1 $\mu$ M EDTA in 1X HEPES – and 75 $\mu$ L was added to each well. Then, 75 $\mu$ L of nickel was added to each well, giving final nickel concentrations from 1 $\mu$ M to 10 $\mu$ M. Due to both the fast coordination kinetics of nickel binding terpyridine and the slow acquisition of the full spectral scan for calculation of FRET efficiency, only the two endpoint emission wavelengths of mTurquoise2 (474nm) and mNeonGreen (517nm) from 400nm excitation were acquired. The ratio of mNeonGreen/mTurquoise2 gave a reliable approximation of FRET efficiency, and this could be modeled as a first-order association curve. We found on-rates scaled linearly with concentration of nickel and that the intermediate and longer scaffold cyclized 10-20-fold faster than the short scaffold, which may elucidate differences in the effective EC<sub>50</sub> of scaffold binding.

**Scaffold Kinetic Off-Rate Binding Experiments:** Due to its measurable off-rate kinetics, cobalt was chosen to model EDTA sequestering of pre-coordinated scaffolds. Here, 2x, pre-coordinated stocks were used in similar 96-well plates – 200nM scaffold, 1 $\mu$ M EDTA, 2 $\mu$ M Co<sup>2+</sup> in 1X HEPES – and 75 $\mu$ L was added to each well. Then, 75 $\mu$ L of EDTA was added to each well, giving final EDTA concentrations from 25 $\mu$ M to 25mM. Similar estimations of FRET efficiency were calculated from the two endpoint emission wavelengths of mTurquoise2 (474nm) and mNeonGreen (517nm) via 400nm excitation. The ratio of mNeonGreen/mTurquoise2 could be modeled as a first-order dissociation curve. We found off-rates scaled linearly with concentration of EDTA; here, however, the dissociation and sequestering of coordinated metal by EDTA did **not** vary by scaffold length.

**Scaffold Reversibility Experiments:** After kinetic experimentation, it was found that cobalt was the best metal to use to demonstrate reversibility of scaffold coordination. Manganese did not form stable dimers, zinc required far too precise of metal concentration necessary for bidentate coordination, and nickel did not dissociate at reasonable levels of EDTA (<100mM). Here, 75 $\mu$ L of 200nM short scaffold, 1 $\mu$ M EDTA in 1X HEPES was added to each well. An initial FRET efficiency was taken (here, 3nm step size). Then, 75 $\mu$ L of 5 $\mu$ M Co<sup>2+</sup> was added to each well, giving final concentrations of 100nM scaffold, 500nM EDTA, 2.5 $\mu$ M Co<sup>2+</sup>. Based on the dissociation kinetics observed above, mM-level EDTA concentration was required to completely sequester cobalt from the bidentate complex. Next, 3 $\mu$ L of 50mM EDTA stock was added to each well. This does not significantly change the overall volume of the well, giving final concentrations of ~98nM short scaffold, 1mM EDTA, and ~2.45 $\mu$ M Co<sup>2+</sup>. FRET efficiencies were taken. Then, molar-equivalent cobalt (here, 15 $\mu$ L of 10mM Co<sup>2+</sup>) was added to the well to quench the EDTA, freeing up cobalt to re-coordinate the scaffold. FRET efficiencies were taken. This cycle was then repeated once more. The final volume after the last EDTA addition was ~189 $\mu$ L, and this 25% dilution correlated with a 25% decrease in fluorescence.

## Chapter 4

### Single Cell Cytoskeletal and Morphological Features Can Discriminate Cancer from Non-Cancer Cell Lines

Research and figures presented in this chapter are adapted from work published in the following manuscript:

Mousavikhamene, Z.\*, **Sykora, D. J.\***, Mrksich, M. & Bagheri, N. Morphological features of single cells enable accurate automated classification of cancer from non-cancer cell lines. *Sci. Rep.* **11**, 24375 (2021).

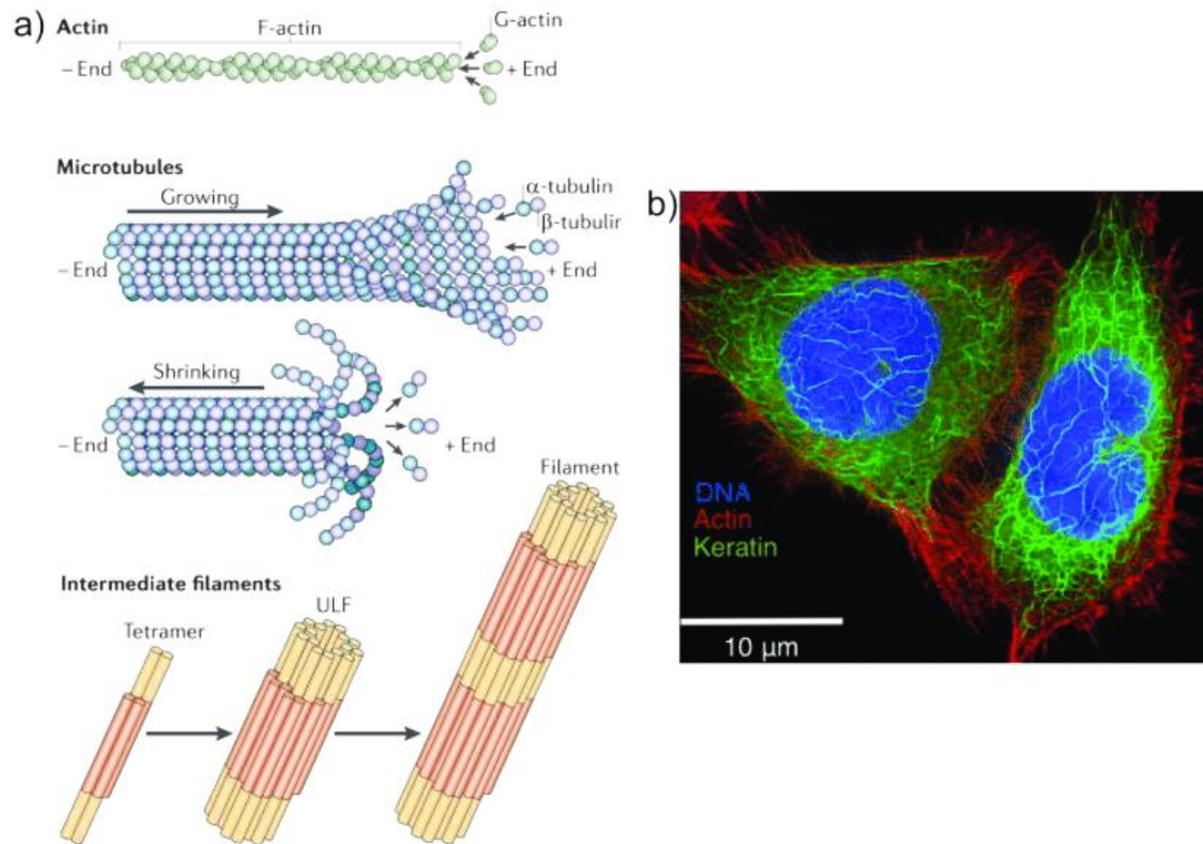
#### 4.1 Introduction

While **Chapter 2** discussed treating cancer from a therapeutic perspective with our modular megamolecules assembly platform, accurate diagnosis of cancer remains a vital cog in the healthcare workflow. Diagnosis rates of many cancers have continued to rise with the obesity epidemic<sup>165,166</sup> despite the overall decrease (~26%) in American cancer mortality, demonstrating that healthier lifestyles (e.g. reduced tobacco use) and improved detection capabilities can lead to better clinical outcomes<sup>166</sup>. In the healthcare workflow, image classification is a fundamental step in successful cancer detection, and automated technologies have been used to complement expert pathologists<sup>167</sup>. However, the accurate description of cancer remains a significant challenge. Not only is it difficult to obtain enough material to run robust image and genetic analysis, tumor microenvironments (TMEs) also possess heterogeneous phenotypes – especially primary to secondary tumor sites<sup>168</sup> – that potentially obfuscate computer-based analysis. Furthermore, variability in specimen preparation can lead to varying diagnoses, even amongst experts<sup>169,170</sup>.

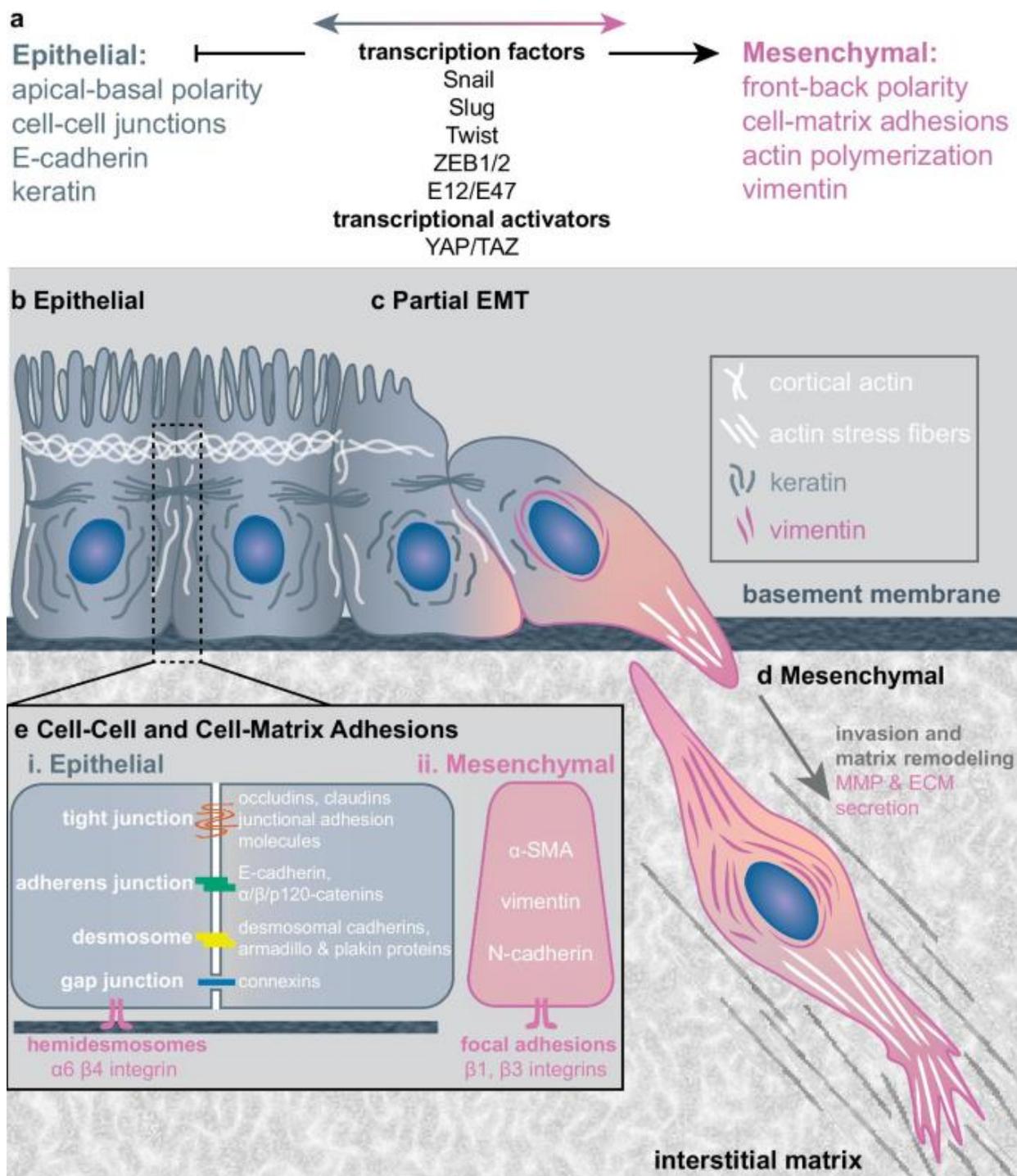
Cancer classification is often empowered by imaging procedures such as hematoxylin & eosin<sup>171</sup>, high-resolution microendoscopy<sup>171</sup>, immunohistochemistry<sup>172</sup>, or radiomics<sup>173</sup> that are run in parallel to genetic analysis (e.g. flow cytometry, RNA-Seq, qPCR, western blots). While rich in tissue-level information, these imaging techniques can often overlook heterogeneity that is necessary to describe complex diseases like cancer. At the cellular level, image classification usually relies on brightfield imaging that enables analysis of large fields of view<sup>174</sup>, limiting the single-cell level information necessary to fully describe biological heterogeneity. Continual improvement of automated cell classification frameworks for the diagnosis and characterization of malignancy in cancer remains an unresolved high priority.

The structure and integrity of a cell's cytoskeletal network has long been known to play an important role in cancer progression (**Figure 4.1**). The cytoskeleton is composed of actin microfilaments of globular actin subunits that bind non-muscle myosin II to create actomyosin stress fibers, intermediate filaments (e.g. vimentin, keratin), and microtubules composed of  $\alpha$  and  $\beta$ -tubulin. During cancer, this network generally shifts from an ordered and rigid state to an irregular and compliant one to support increased proliferation and motility<sup>175</sup>. More specifically, successful epithelial-to-mesenchymal transitions (EMT) and downstream metastasis are dependent upon reorganization of actin microfilaments via Rho GTPases (Rac1, RhoA, Cdc42), which ultimately guide phenotype<sup>176,177</sup>. During EMT, the actin bundles of epithelial cells shift from thin, cortical bundles to thicker, parallel, contractile bundles<sup>178</sup>, allowing for actomyosin contraction and subsequent migration (**Figure 4.2**). Thus, some metastatic cancer cells can adopt a similar actomyosin phenotypic profile of healthy migratory mesenchymal cells (e.g. fibroblasts). Yet, metastasis is still more complicated than this overarching model, as some circulating tumor cells (CTCs) do not need to engage in EMT to metastasize. Additionally, CTCs can display a spectrum

of both epithelial and mesenchymal biomarkers<sup>179</sup> and the subsequent intermediate phenotypes<sup>180</sup>. Finally, the cytoskeleton serves as a regulator of gene expression, which can ultimately lead to cell proliferation and activation of various oncogenes<sup>181</sup>.



**Figure 4.1** The cell's cytoskeleton is comprised of three constituent biomacromolecule building blocks **A)** actin filaments (F-actin), which assemble from globular monomers (G-actin); microtubules, which are assembled of alpha and beta tubulin monomers, and various intermediate filaments, such as vimentin; **B)** these structures can be stained with dyes or labels (e.g. fluorescent antibodies) for imaging by confocal microscopy<sup>213</sup>.



**Figure 4.2** During epithelial-mesenchymal transition (EMT), the cell's cytoskeleton undergoes a dynamic shift from an epithelial phenotype to a migratory, mesenchymal phenotype more fitting for metastasis<sup>214</sup>.

In a similar spirit to elucidating structure-function of biomacromolecules elucidated in **Chapter 1**, macro-scale cell shape also plays a role in its subsequent function<sup>182-183</sup>; one of the best-studied examples is how cells physically bisect their longest axis during symmetric division. An earlier paper from our group has previously shown that cell shape can orchestrate the mechanochemical signals that direct mesenchymal stem cells to specifically differentiate<sup>183</sup>. Here, we utilized the soft lithographic technique of microcontact printing ( $\mu$ CP) on self-assembled monolayers (SAMs) that Milan helped develop while in the Whitesides lab to pattern cells in fixed shapes<sup>184-187</sup>. In essence, the changes in both the actin cytoskeletal structure as well as the overall cell shape can serve as vital markers in cancer diagnostics and progression, potentially improving characterization and diagnosis from automated classification frameworks analyzing inherently heterogeneous cancer cell populations.

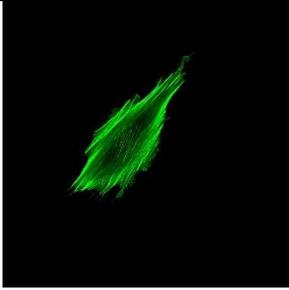
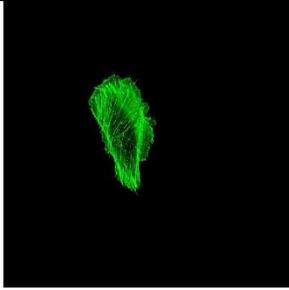
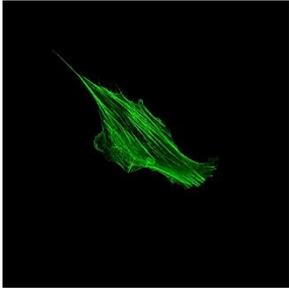
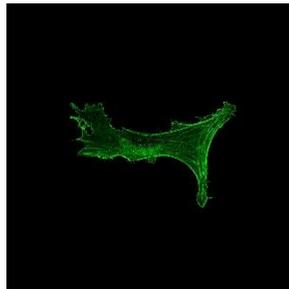
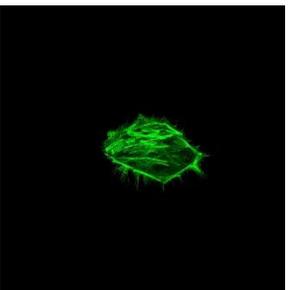
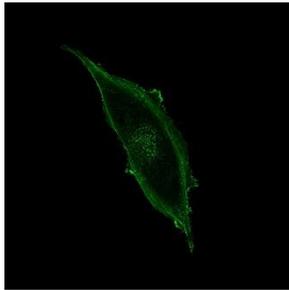
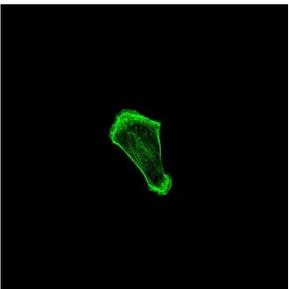
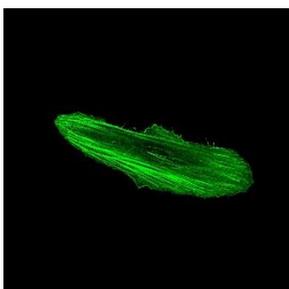
In the rest of **Chapter 4**, I will describe how actin cytoskeletal and morphological data of eight cell lines by high-resolution confocal microscopy provided ample discriminatory data to support an accurate and generalizable classification model. I first quantified single-cell actin cytoskeletal confocal images by defining features based on the spatial configuration and morphology of both cancer and non-cancer cells. With my collaborator Zeynab, we applied supervised feature extraction techniques to identify the cytoskeletal and morphological features with the most discriminating power between cell types. Next, support vector machines (SVM) with various kernels successfully classified cell types in a pairwise approach. Specifically, we applied SVM on all binary cell line combinations to explore classification outcomes between cancer/non-cancer cases against both cancer/cancer and non-cancer/non-cancer cases. Furthermore, certain pairwise comparisons demonstrated improved classification when spatially restricting and normalizing cell shape via  $\mu$ CP on SAMs. Finally, we tested the model's ability to distinguish a

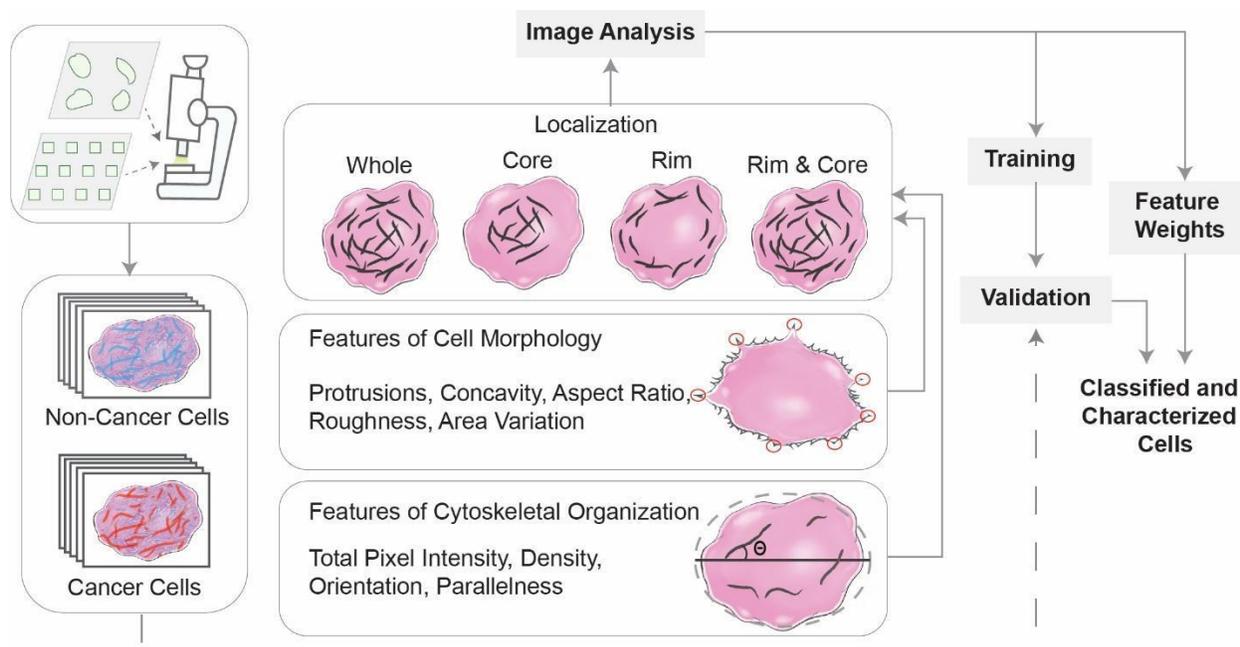
new, dissimilar cell line from a different tissue. This approach confirmed the generalizability of this proof-of-concept model to make predictions for completely novel cell lines.

## 4.2 Results

Hundreds of confocal images from eight commonly used cancerous and non-cancerous cell lines were stained for their actin cytoskeletons and collected (**Table 4.1**). We spatially segmented cells into various groups (whole, rim, core, and rim & core) in order to analyze the spatial data in different ways – we called these localization classes. (**Figure 4.3**). First, the actin cytoskeleton from each image localization class was quantified; this included actin fiber intensity, density, orientation, and parallelness (determined by normalized variance of overall fiber angle within the cell). The “whole” image class underwent additional morphological analysis, quantifying shape-based features of the whole cell: protrusions, concavity, aspect ratio, roughness, and area variance (**Figure 4.4**). In cancer, malignant, migratory cells are guided by protrusions and parallel contractile actomyosin bundles to help promote efficient cell migration<sup>188</sup>, which often differ from their healthy epithelial precursors. Furthermore, as the cytoskeleton of cancer cells is generally less ordered compared to non-cancer cells<sup>189</sup>, cancer cells often have greater surface area variation known as pleomorphism, which is a hallmark of cancer detection<sup>189-190</sup>. In this way, we could comprehensively describe both the actin distribution and organization at the basal surface as well as the overall morphology of the cell. We took two approaches to validate our model: our *pairwise* approach compared two cell lines in a binary fashion, while our *combinatorial* approach combined multiple cell lines into “cancer” and “non-cancer” categories, later validating novel cell lines to demonstrate generalizability.

**Table 4.1.** Description of Cell Lines Used in **Chapter 4**

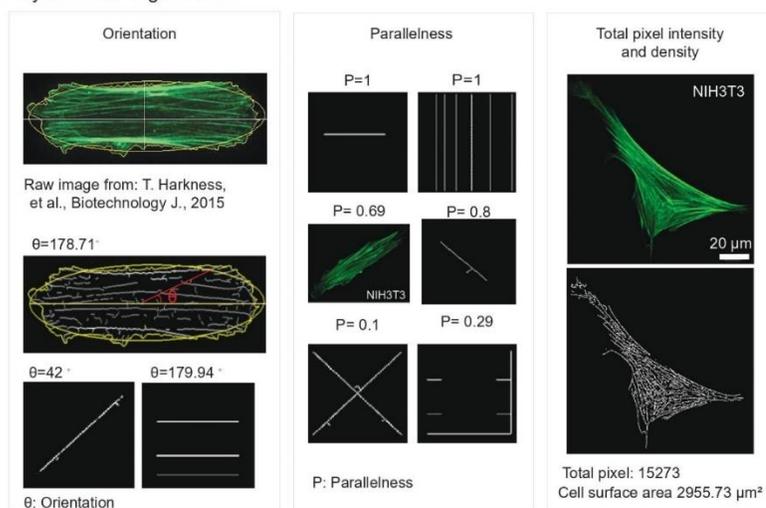
Cell Line Brief Description	Representative Image	Cell Line Brief Description	Representative Image
<p>HFF- 1, Human</p> <p>Non-cancerous fibroblast from foreskin. Typically elongated with aligned stress fibers, may or may not have strong basal fibers in the middle of the cell. Jagged cell shape</p>		<p>HT-1080, Human</p> <p>Cancerous fibrosarcoma cells. Have well-defined stress fibers. Typically less aligned than non-cancerous counterparts.</p>	
<p>NIH/3T3, Murine</p> <p>Non-cancerous fibroblasts from Swiss albino mice. Typically elongated with aligned stress fibers. Seem to have slightly more pronounced basal stress fibers in the middle of the cell.</p>		<p>B16-F1, Murine</p> <p>Melanoma cells, medium metastatic potential. Mix of spindle shaped and epithelial shaped cells. Large area variation. May have stress fibers within the interior, most often do not.</p>	
<p>MCF10A, Human</p> <p>Non-cancerous breast epithelial cells. Less elongated than fibroblasts but still display thick stress fibers. Often have hairy filopodia protruding from the exterior. Generally have a more rigid cell shape.</p>		<p>B16-F10, Murine</p> <p>Melanoma cells, high metastatic potential. Mix of spindle shaped and epithelial shaped cells. Large area variation. Very few stress fibers in the interior of the cell.</p>	
<p>MDA-MB-231, Human</p> <p>Triple negative basal breast cancer cells. Typically, smaller than other cells with sparse aligned stress fibers interior to the cell; typically have ruffled lamellipodia.</p>		<p>HeLa, Human</p> <p>Cervical adenocarcinoma cells. Epithelial cells when confluent/as islands but can appear elongated and mesenchymal when single cell. Have aligned stress fibers with rounded cell shape. Typically fewer defined fibers in the middle of the cell.</p>	



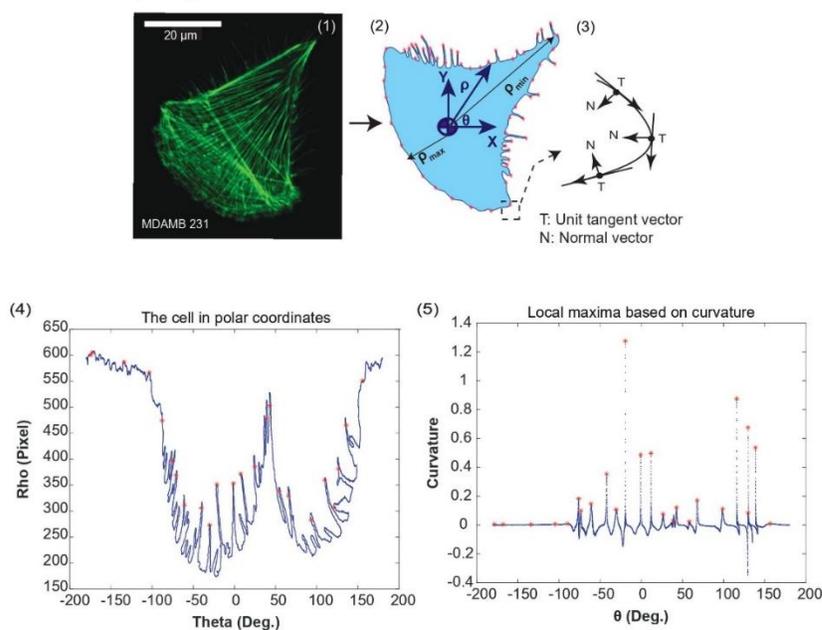
**Figure 4.3** An automated image processing framework quantifies features of cellular cytoskeletal and morphological structure from single cell images. These features were used to train parameters of a classification model and its performance was evaluated using validation data. The algorithm was able to accurately discriminate cancer cells from non-cancer cells and identify individual features that had the greatest influence on classification outcome.

In order to determine whether the model could successfully classify single-cell images by their cytoskeleton and morphology alone, the feature extraction model was trained by quantifying features of interest in a *pairwise* manner amongst eight commonly used human and murine cell lines. Non-cancerous cells included human foreskin fibroblasts (HFF-1), murine fibroblasts (NIH/3T3), and human breast epithelial cells (MCF10A). Cancerous cell lines included human fibrosarcoma (HT-1080), human breast cancer (MDA-MB-231), murine melanoma of moderate (B16-F1) and high (B16-F10) metastatic potential, and human cervical adenocarcinoma (HeLa) cells. Pairwise training identified the optimal hyperplane in an N-dimensional space (N is the number of features) that best discriminates between two cell types. A 10-fold cross validation procedure was used for model training and testing.

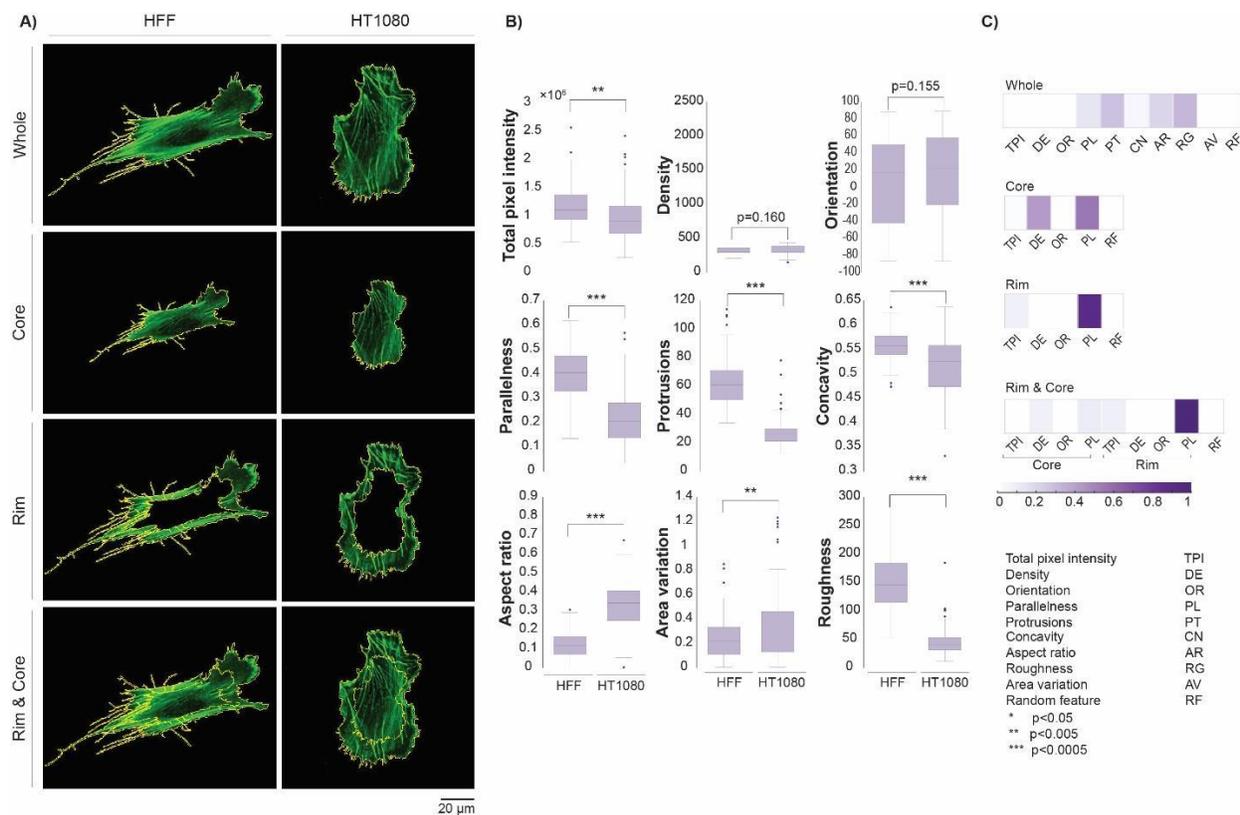
## A) Features of Cytoskeletal Organization:



## B) Features of Cell Morphology:



**Figure 4.4 Quantified features for each single cell image.** A) Features of single cell cytoskeletal organization including orientation, parallelness, and total actin intensity & density are calculated (46). B) Features of single cell morphological features are calculated from the (B1) initial image. B2 is the cell with the transferred boundary to the centroid. An example of minimum and maximum radii ( $\rho_{\min}$  and  $\rho_{\max}$ ) used for the “area variation” feature is presented. B3 illustrates an element of the border curve: unit tangent and normal vectors used to calculate protrusions and concavities. B4 represents the cell in polar coordinates. B5 illustrates local maxima of the corresponding protrusions in B2 (red asterisks). These steps are taken for each individual single cell image.

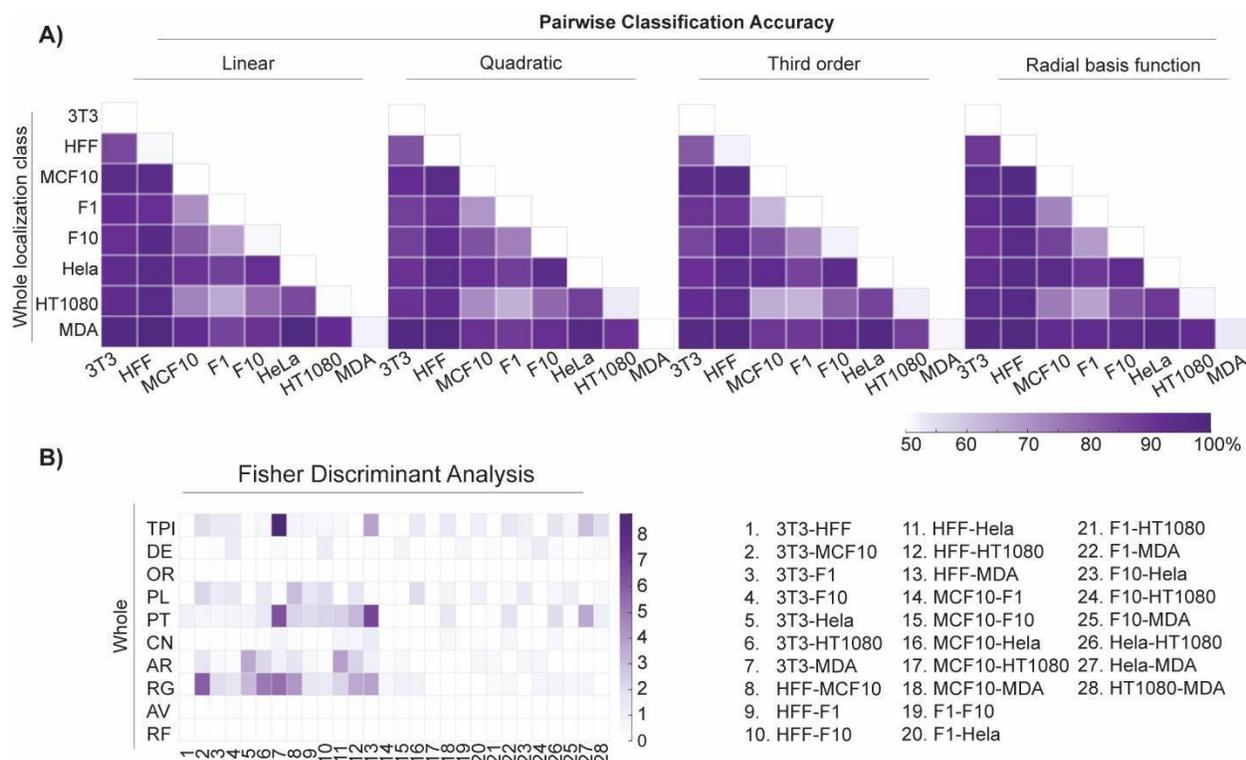


**Figure 4.5** **A)** Single HFF-1 and HT-1080 cells are segmented based on defined localization classes. **B)** Features were extracted from the entire data set of HFF-1s and HT-1080s images (whole cell, ~100 cells per line). Differences between these classes can be observed by eye and by statistically significant differences in features summarizing morphological and organizational properties. **C)** Fisher discriminant analysis identified the most discriminatory features between these two data sets for all four localization classes.

Within the set of all possible cell-cell comparisons, a physiologically relevant comparison is the HFF-1 against the HT-1080. While the HFF-1 and HT-1080 specifically would not be found in the same tissue, the HFF-1 represents a model human fibroblast, and the HT-1080 represents a model human fibrosarcoma cell - that is, a cancerous cell line derived from similar mesenchymal tissue. Both cell lines are commonly used in cell migration studies<sup>191</sup>. Each raw confocal image was spatially segmented into each of the four primary localization classes, skeletonized for quantification via ImageJ/FIJI, and analyzed to extract features most critical for discrimination (**Figure 4.5**).

Looking at the whole cell, the non-cancerous HFF-1 displayed a more elongated aspect ratio, increased parallelism, more protrusions, less area variation, greater pixel intensity, and greater roughness than the HT-1080 (**Figure 4.5B**). Yet, statistical significance of quantified features may not necessarily play an equivalent role in classification. An unbiased Fisher Discriminant Analysis identified the combination and contribution of features on classification. Results highlighted the relative impact (weight) of each feature for robust cell discrimination (**Figure 4.5C**). Not surprisingly, when analyzing the whole cell, roughness, protrusion count, aspect ratio, and parallelism of actin stress fibers were most discriminatory. When analyzing other localization classes, different features were highlighted: density and parallelness were most discriminatory for the core class, while parallelness alone was sufficient to discriminate both the rim and rim & core localization classes. A Support Vector Machine (SVM) was used to perform the classification. SVMs separate classes by solving for the hyperplane that maximizes their distance using various kernels. Four common kernels were employed: linear, quadratic, third-order polynomial, and radial basis function (RBF). Looking at the whole cell comparison between HFF-1s and HT-1080s, the classification algorithm yielded about 97% accuracy across the four kernels, a promising indication for our methodology.

This approach was extended to evaluate the remaining pairwise comparisons provided by the data. Accuracy of pairwise classification is outlined in **Figure 4.6** (whole localization). Here, 22 of the 28 pairwise model training and testing demonstrated accurate classification rates of greater than 90%. An unbiased Fisher Discriminant Analysis revealed that total actin intensity and stress fiber parallelness along with the morphological protrusion count, aspect ratio, and roughness often played the most significant role in discrimination. However, the latter two features were less relevant when discriminating cancer/cancer pairings.



**Figure 4.6 Accuracies of all whole cell pairwise comparisons.** **A)** Heat maps summarize classification accuracies for each SVM kernel; 22 of the 28 pairwise combinations resulted in 90% accuracy and above. **B)** Fisher Discriminant Analysis reveals the features most important for pairwise discrimination across all 28 comparisons.

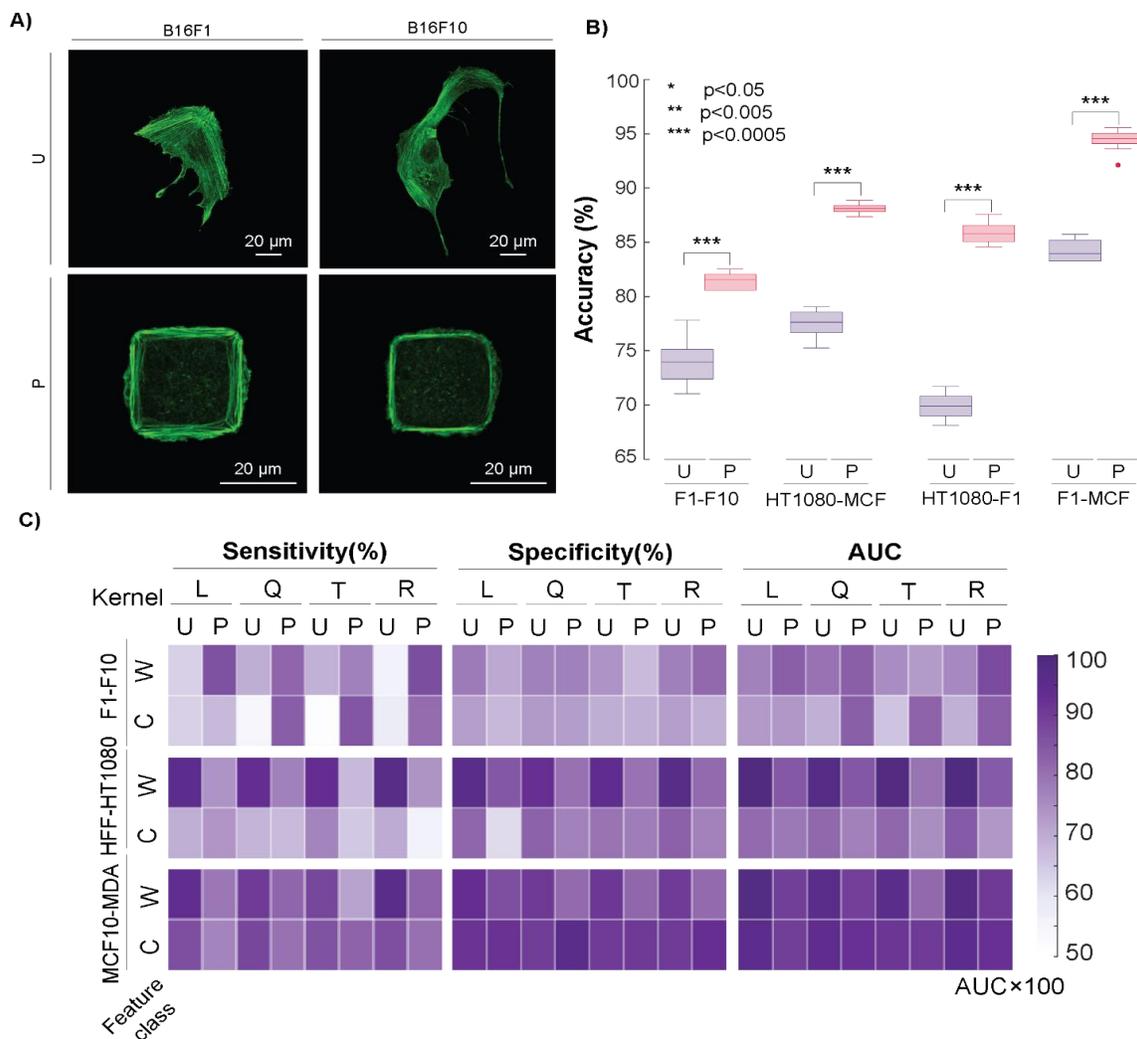
In addition to the HFF-1/HT-1080 comparison, both the MDA-MB-231/MCF-10A and B16-F1/B16-F10 comparisons are somewhat physiologically relevant. The model discriminated between MDA-MB-231 and MCF-10A cells with ~94% accuracy but discriminated between B16-F1 and B16-F10 cells with only ~74% accuracy. Based on the images acquired via confocal microscopy, it was not surprising that the latter pairwise comparison was more challenging to discriminate; the cells represent the same cancer type from the same species (murine melanoma) and differ only in metastatic potential/aggressiveness<sup>192</sup>. This result marked a true test for the model, as the significant population overlap between these slightly differing phenotypes was difficult to discern. In addition to B16-F1/B16-F10, there were four other pairwise comparisons that were not as accurate (<85%) as the rest. These comparisons often had similar phenotypes even

if the cells came from different species/tissues. For these few weaker pairwise comparisons, a secondary approach to collect and discriminate cell image data was clearly necessary.

To improve the predictive accuracy of not only the B16-F1 v. B16-F10 pairwise comparison, but also the other weaker comparisons, we sought to minimize the population overlap between these similar phenotypes by normalizing cell shape. We hypothesized that the phenotypic plasticity seen in cancer<sup>193</sup> may lead to a wider distribution of phenotypes, potentially confounding discrimination attempts when cells are left unpatterned. We posited that normalization of shape could effectively mitigate the spread of phenotypes that made the previous unpatterned cell analysis unsuccessful. Phenotype normalization was done by the microcontact printing ( $\mu$ CP) approach on self-assembled monolayers (SAMs) previously described<sup>185-187,194</sup>. Here, 900  $\mu\text{m}^2$  square islands of octadecanethiol (ODT) were stamped on gold-evaporated glass slides and backfilled with tri(ethylene glycol)-terminated alkanethiols to prevent nonspecific protein adsorption. Fibronectin was specifically adsorbed to these ODT islands, which gave the individual cells a platform for attachment and spreading. We found that this area provided cells adequate room to fully spread (i.e. cover the entire square island) in a short amount of time ( $< 8$  hours) without allowing the highly proliferative cancer cells to divide. Coupled with nuclear staining, this approach ensured that we analyzed one spread cell per spot (**Figure 4.7**).

After collecting enough data for each cell type, we re-ran the model with micropatterned (“patterned”) cells. What we found was illuminating: patterning improved the predictive accuracy from 74% to 84% for the B16-F1 v. B16-F10 comparison (**Figure 4.7A**). In fact, patterning improved the accuracy of classification for three of the least accurate ( $< 85\%$ ) pairwise comparisons (**Figure 4.7B**), demonstrating that shape normalization with methods like  $\mu$ CP can be effective in improving image classification in certain situations where native, spread phenotypes

are difficult to distinguish. Ultimately, due to the known phenotypic plasticity of cancer, normalizing cell shape may help play a role in effectively analyzing biological properties or monitoring drug response of similar cancer cell types. Overall, 27 of the 28 pairwise comparisons through either unpatterned or patterning methods were able to achieve discriminatory accuracies over 82% with most at or above 90%.



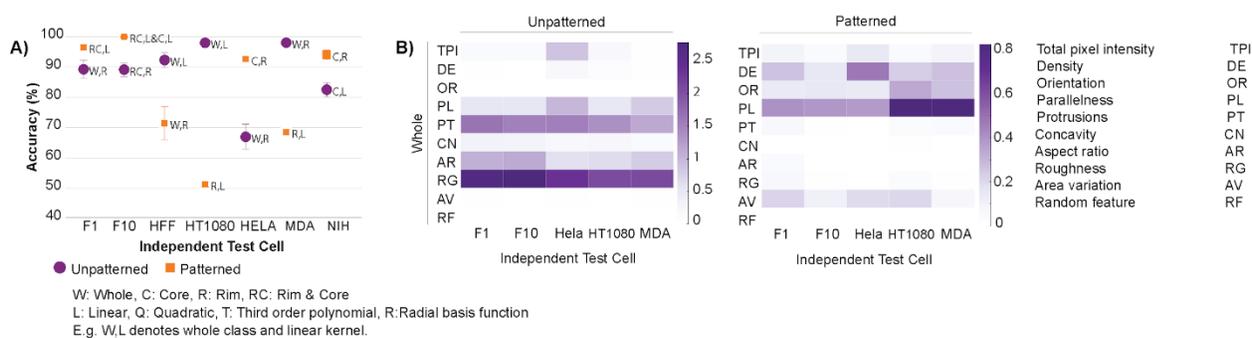
**Figure 4.7 Example of microcontact printing-based image analysis. A)** Comparison of B16-F1 and B16-F10 murine melanoma cells when unpatterned (U) or patterned (P) in 900  $\mu\text{m}^2$  islands. **B)** Improvement of pairwise predictive accuracy when hard-to-discriminate cell lines are patterned. **C)** Sensitivity, Specificity, and AUC of physiologically relevant comparisons B16-F1/B16-F10, HFF-1/HT-1080, and MCF10A/MDA-MB-231 for both patterned (P) and unpatterned (U) cells across both whole (W) and core (C) classes and all four SVM kernels.

Further quantitative validation can be seen in **Figure 4.7C**, which displays the sensitivity, specificity, and Area Under the Receiver Operating Characteristics (AUC) of the three main physiologically relevant pairwise comparisons. Sensitivity describes the rate at which the model could successfully identify the cancerous (HT-1080 and MDA-MB-231) and more metastatic (B16-F10) cell out of the cancerous/highly metastatic data set. Specificity describes the rate at which the model could successfully identify the non-cancerous (HFF-1 and MCF10A) and less metastatic (B16-F1) cell out of the non-cancerous/less metastatic data set. The AUC curve is a performance metric that quantifies how well our model can discriminate either cancer from non-cancer (HT-1080 & MDA-MB-231 v. HFF-1 & MCF10A, respectively) or less metastatic (B16-F1) from more-metastatic (B16-F10) cells.

We found that for the B16-F1 v. B16-F10 comparison (“whole” localization), sensitivity and AUC are improved as a result of patterning cells in comparison with the same kernel of unpatterned data; however, this was not the case for HFF-1 v. HT-1080 or MCF10A v. MDA-MB-231 comparisons, which were already highly robust when left unpatterned (**Figure 4.7C**). This outcome provided a fundamental conclusion that we found compelling: normalizing cell shape only helped enhance pairwise discrimination in instances where the native, spread phenotypes were highly similar. Initially, we had expected patterning to aid discrimination across most pairwise combinations, but due to the quality of our single cell images, unpatterned cells that were already well discriminated did not require shape normalization.

Finally, we evaluated the generalizability of this model on both patterned and unpatterned cells. In this approach, we pooled all five cancer cell lines as a bulk “cancer” class and the two mesenchymal non-cancer cell lines as a “non-cancer” class, as epithelial cancer cells generally adopt a more migratory, mesenchymal phenotype than their epithelial non-cancer precursors (in

the case of the HT-1080, it is already derived from mesenchymal cells). Thus, the MCF10A cell line was withheld from model training. As the number of cancer cells was greater than the number of non-cancer cells, the training set was balanced by random sampling from cancer labeled data points to make it equal to the number of non-cancer labeled data points. To evaluate the algorithm's sensitivity to sampling, we repeated the process (sampling, training, testing) ten times, withholding the same Independent Test Cell, and evaluated (**Figure 4.8**), including leftover images from cell lines utilized in model training (i.e. Consistent Test Cell) The algorithm was ultimately validated on the independent cell line, demonstrating - in most cases - accurate predictive outcomes. In most cases, some combination of image type (patterned or unpatterned), SVM kernel choice (linear, quadratic, third-order, RBF), and localization class (whole, core, rim, rim & core) yielded a predictive accuracy of 90% or greater. Fisher analysis revealed that while discriminating features differed by cell type, all cases made use of stress fiber parallelness as the most impactful feature. This outcome reinforces the observation that cancer and non-cancer cells have consistent differences in cytoskeletal structure and overall morphology that can be generally predicted by a robust classification framework.



**Figure 4.8** The model demonstrates generalizability towards cancer cells that were not used in model training. **A)** Maximum predictive accuracy for both unpatterned and patterned cell lines by localization class and kernel. **B)** Fisher Discriminant Analysis reveals the most discriminatory features for both unpatterned and patterned cells for the whole localization class.

### 4.3 Conclusions

I sought to determine whether differences in cytoskeletal structure and overall cell morphology and organization could provide robust discrimination between cancer and non-cancer cell phenotypes. Feature extraction-based machine learning was utilized, augmented by SVM, to classify images of single cancer and non-cancer cells. My single-cell imaging technique provided high resolution, in-depth analysis of cell morphology and cytoskeletal organization in both patterned and unpatterned shapes. We trained a machine learning model with only 100 images per cell type, demonstrating accurate prediction of dissimilar cell lines across many tissues; every independent test cell achieved an accuracy >90%. Additionally, I was able to identify key morphological and actin-based cytoskeletal features that were instrumental in discriminating cancer/non-cancer cells across most cell lines. Differences in the density and alignment of the actin cytoskeletal network coupled with differences in overall morphology (roughness, protrusion count, and aspect ratio) of the cell often played the largest role in successful image classification. These differences correlate to many of the known, overarching morphological changes that occur during cancer progression and subsequent EMT<sup>175-178,193</sup>. Ultimately, heterogeneity within and between cell types reinforces the utility of feature extraction models. As advanced imaging techniques that incorporate high, single-cell resolution with increased throughput are made more accessible to both scientists and clinicians, the strengths of unbiased machine learning algorithms in detection, diagnosis, and prognosis of cancer are poised to improve patient outcomes. Additional machine learning highlighting the actin cytoskeleton has previously been done using convolutional neural networks (CNNs)<sup>195</sup>. While CNNs are a powerful tool often used for image classification of large datasets, interpretability is a known challenge, limiting insight on identifying elements of the cytoskeleton that have the greatest impact on discrimination and their biological interpretation.

Both our work and this work of Oei *et al.* highlight the great promise that machine learning portends in cancer classification, and we anticipate that a larger and a more comprehensive single-cell, high resolution data set acquired by a dedicated imaging team would improve model training even further.

Initially, comparing individual cell types allowed us to identify features that could be used as quantitative parameters in both translational and basic cancer research. While the intersection of cell biology and physics is well-studied (e.g. mechanotransduction), most cancer treatments mainly focus on alterations in biochemical signaling. Actin filaments themselves represent a poor drug target due to their ubiquitous role in cardiac, renal, and skeletal tissue homeostasis<sup>196</sup>, but the overall organization of the actin cytoskeleton could be used as a phenotypic guide for drug studies and development<sup>197-199</sup>. Few chemotherapeutic mainstays target microtubules (e.g. taxanes like Paclitaxel), and novel cytoskeletal agents may end up targeting actin-associated proteins that show distinct differential expression in tumors (e.g. gelsolin<sup>200-202</sup>). We believe there are many relevant scenarios where such an approach could help investigators and clinicians answer critical questions.

First, this classification framework could interrogate how the actin cytoskeleton and overall morphology of the cell is altered or disrupted. Cancer cells have a distribution of gene expression and downstream phenotypes - it may be informative to discern how cancer cells respond to both existing or new therapeutic agents relative to physiologically relevant healthy cells in the same TME. As a corollary, it may be informative to observe and document consistent phenotypic patterns based on the molecular target within various cancer and non-cancer cells. We speculate that there may be an oncological therapeutic that could remodel the cytoskeleton of a malignant cell (i.e. disordered) to that of a healthy or more benign counterpart. In effect, these sorts of computationally inexpensive analyses should be integrated into traditional biomarker panels that

help predict likelihood of therapeutic response. Ultimately, false negatives in oncological screens lower the chance of an early diagnosis with an effective treatment plan and promising prognosis. This powerful generalizability suggests that a similar framework could be useful for developing detailed protocols or kits that can train relevant cell types within a specific TME to ultimately test against a patient biopsy not used in training. While the cell lines used in our study are commonly used immortalized cells and do not represent the types of cells found in primary tumor biopsies, we anticipate that similar discriminatory cytoskeletal and morphological properties would arise between and within different cancers depending on the location and grade. Immense progress has been recently made in increasing the quantitative throughput of pathology at both the tissue-level and single-cell level<sup>203</sup>, which is concurrently reflected within products and services offered by key industry players (e.g. TissueGnostics, Spatial Transcriptomics, ACD Bio). This suggests that incorporating cell culture and high-resolution microscopy in the workflow may no longer be as time or cost prohibitive as it once was, especially if the high-resolution analysis can be acquired in the native, three-dimensional biopsy.

This classification model has the capacity to be extended (both in series and in parallel) to the other cytoskeletal elements that have been observed to be altered in cancer. Not only does this include microtubules, which are targeted by some cancer agents like taxanes, but also includes intermediate filaments, actin-associated proteins, and even the nucleus. While the nucleus would require a different workflow that does not highlight filamentous structures, it is known to face morphological alterations like size, aspect ratio, shape, and number per cell<sup>204,205</sup>. In addition to new phenotypic elements potentially captured by high-resolution confocal microscopy, this type of model could be extended to primary healthy and cancerous cells found within the TME across various biopsy methods. Since fine needle syringe extraction typically yields cell counts on the

order of millions<sup>206</sup>, building up a robust model on the order of hundreds of cells would be quite achievable for a team of clinical technicians. Finally, we utilized a microscale patterning technique to normalize cell shape. With the recent advancement of various nano-scale patterning techniques, direct control of subcellular cytoskeletal components on a two-dimensional surface is now more feasible than ever before<sup>207</sup>. These nano-sized adhesive cues guide subsequent cell response, allowing researchers to potentially uncover designs with even better discriminatory power. Ultimately, reliably controlling cell shape in three dimensions in a high-throughput manner is a holy grail in the field of *in vitro* cell microenvironments, and the continued development of such may greatly facilitate further analysis of native tumor biopsies.

#### **4.4 Experimental**

**Cell Culture.** All cell lines were acquired from American Type Culture Collection (ATCC) and cultured at 37C at 5% CO<sub>2</sub>. Non-cancerous murine NIH/3T3 fibroblast cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% bovine calf serum (BCS) and 1% Penicillin/Streptomycin (Thermo Fisher Scientific). Non-cancerous human MCA10A breast epithelial cells were cultured in mammary epithelial cell growth medium (MEBM) supplanted with 4 µL/mL bovine pituitary extract, 1 µL/mL human epidermal growth factor, 1 µL/mL insulin, 1 µL/mL hydrocortisone (all from Lonza), and 100 ng/mL cholera toxin (Sigma-Aldrich). Human foreskin fibroblasts (HFF-1s), B16-F1, B16F-10 murine melanoma cells, HeLa human adenocarcinoma cells, HT-1080 human fibrosarcoma cells, and MDA-MB-231 human breast adenocarcinoma cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) and 1% Penicillin Streptomycin (Thermo Fisher Scientific). All cell lines

were routinely subcultured every 1-3 days to avoid over-confluence and potential cellular quiescence.

**Preparation of Monolayers and Microcontact Printing.** No. 1.5 glass coverslips (Fisher Scientific) were sonicated in ethanol, then water, then ethanol for 30 minutes per cycle before being dried with nitrogen gas to clean the surface. Titanium (50Å) and then gold (200Å) were evaporated onto these coverslips using an electron beam evaporator (Thermionics) at 0.2 and 0.5 nm/s, respectively, at  $10^{-6}$  Torr. The patterned surface of polydimethylsiloxane (PDMS) stamps (previously fabricated in-house<sup>30</sup>) were spotted with 10mM octadecanethiol (ODT) (Sigma Aldrich) in ethanol and allowed to air dry for at least 10 minutes. Further drying via nitrogen gas was performed to ensure the entire stamp surface was completely dry. Stamps were very carefully placed face-down onto the cut gold-evaporated coverslip with a flat, 15 g weight for 45-50 seconds before being briefly washed with ethanol, water, and then ethanol. This yielded selective thiol-gold semi-covalent bonding of hydrophobic ODT on the gold slide in the shape of the pattern used. Then, slides were placed in 10mM triethylene glycol mono-11-mercaptoundecyl ether (Sigma-Aldrich) in ethanol overnight at 4C to backfill the unpatterned regions with hydrophilic alkylthiol. The following day, slides were washed in ethanol, dried with a nitrogen stream, and placed in a 6-well plate (Fisher Scientific) in 1mL of PBS. Here, fibronectin was directly pipetted into the PBS solution to adsorb onto the hydrophobic ODT pattern and allowed to sit in a humidified chamber for 1 hour at 37C. Wells were triple rinsed with excess PBS, but being careful to never go dry before the third rinse so excess fibronectin did not nonspecifically adsorb to the slide.

**Cell Seeding.** An 8-well chamber slide with #1.5 glass bottom (Ibidi) had a solution of 25  $\mu\text{g}/\text{mL}$  fibronectin (Sigma-Aldrich) in PBS (Sigma-Aldrich) adsorbed to the slide for 1 hour at 37C to promote cell attachment and spreading for cytoskeletal analysis. After a triple PBS rinse, 5,000 cells were seeded in each well in culture media and allowed to adhere and spread overnight. For patterned cells on the microcontact printed surfaces, 25,000 cells were seeded in each well and allowed to attach and spread for 6-8 hours.

**Immunofluorescence.** After either unpatterned cells spread overnight or patterned cells spread for 6-8 hours, all cells were fixed in 4% paraformaldehyde for 15 minutes and then permeabilized with 0.3% Triton-X 100 for 5 minutes at room temperature. Cells were blocked in 1% BSA (Santa Cruz Biotechnologies) with 22 mg/mL glycine for 30 minutes at room temperature. Then, cells were treated with a 1:40 dilution of AlexaFluor 488 phalloidin (Thermo Fisher Scientific) to label intracellular actin in 1% BSA for 30 minutes at room temperature with NucBlue nuclear staining (Thermo Fisher Scientific). Cells were triple rinsed with PBS after staining and imaged immediately to ensure high image fidelity. Single cells were taken on a 60X oil immersion objective on a Nikon Ti Eclipse confocal microscope (Nikon Instruments) with the corresponding NIS Elements software. For image acquisition, only the 488 nm (actin) channel was taken, but the 405 nm channel (nucleus) was viewed *a priori* to ensure that only one cell was in the field of view, which was not always obvious when imaging patterned cells. Laser power was typically between 2-5%. Only the basal layer of the cell was taken, as the underlying cell surface demonstrates the most robust change in cancer cell progression and has the strongest planar F-actin profile. It also allowed us to expedite image acquisition and algorithm design in a reasonable time frame. Images

were stored as JPEGs with 2048x2048 resolution (0.1  $\mu\text{m}/\text{pix}$ /0.03  $\mu\text{m}/\text{pix}$  for unpatterned/patterned, respectively).

**Data Pre-Processing.** *Boundary Detection and Masking:* Single-cell images' external boundaries were detected and extracted from the background based on the threshold of their pixel values. Pixel values that are larger than the threshold are filtered against the black background with very low pixel value (almost zero). Dilation and erosion, which are fundamental morphological procedures, are used to prevent detection of intracellular areas while the external boundary is detected. The detected external boundaries are plotted on the original image to check the validity of the boundary detection. Each image is masked to make sure the image background, which is not part of the cell and might have some background noise like cell debris, is excluded, so that the analysis is merely performed on the inner cellular region of the image.

**Feature Extraction.** Features are extracted from each single-cell image within each cell line.

*Cytoskeletal Features (Figure 4.4):*

Total Pixel Intensity and Density. Total pixel intensity is calculated by the summation of the actin pixel values of each single-cell image. Some cells have more concentrated actin with more robust stress fibers while others, particularly cancer cells, are more diffuse. Actin density is calculated by taking the total actin count and dividing it by the surface area of the spread cell. Pixel lengths are reported from the microscope (e.g. 0.1  $\mu\text{m}/\text{pix}$  for 60x objective image at 2048x2048 resolution, 0.03  $\mu\text{m}/\text{pix}$  for the patterned images).

Directionality and Randomness. For directionality, we fitted an ellipse to each cell. Directionality is the average angle of actin stress fibers with respect to the largest diameter of the fitted ellipse. Angles were appropriately weighted based on length of stress fiber, so larger stress fibers contributed more weight to directionality calculation than smaller, more diffuse fibers. Randomness is used to track the overall variation of actin fibers angles. It ranges from 0 to 1 for fully random to fully parallel, respectively. The FIJI LPX plugin was used to calculate the directionality and randomness<sup>208,209</sup>. We developed a framework to make this plugin fully automated.

*Morphological Features (Figure 4.4):*

Protrusions and Concavity. To calculate protrusion and concavity of each cell, the curvature of the cell boundary is found using MATLAB's 2D Line Curvature and Normals Package<sup>210</sup>.

$$\frac{d\vec{T}}{dS} = \kappa\vec{N} \quad (1)$$

where  $\vec{T}$  is unit tangent vector,  $dS$  is differential element of border curve of the cell,  $\kappa$  is curvature and  $\vec{N}$  is the normal vector. The number of local maxima for curvature of each single cell is stored as “protrusion.” The sign of the curvature indicates convexity/concavity at any point along the cell border. “Concavity” is defined as the number of times the curvature’s sign changes along the boundary, which is further normalized with the total number of points in the boundary.

Roughness and Aspect Ratio. The origin of the coordinate system is coincident with the centroid of the cell and transformed to polar coordinates, which assigns radial values to each point along the perimeter. Roughness, or the presence of non-smooth surfaces, is another indicator of projections in the cell boundary. It is defined as the standard deviation of the radii along the boundary. Aspect ratio is simply the ratio of the radii minimum to maximum along the cell boundary, which provides a normalized measure between 0 and 1. Here, a lower aspect ratio means the cell has a longer, more uniaxial phenotype.

Area Variation. The standard deviation of spread surface areas of all cells was calculated for each cell line. In general, cancer cells indeed had more variation in phenotype than non-cancer cells.

Localization. Synonymous with feature classes, the spatial heterogeneity in actin organization was defined accordingly: The “core” of the cell is the inner 65% of the cell area from the centroid to the boundary. The “rim” constitutes the remaining 35%. The whole cell is further defined as “rim” + “core” or “whole”, which takes the entire cell without summing constituent parts. The latter “whole” classification allows morphological features to be calculated and applied accordingly. These percentages were determined by trial-and-error in order to significantly change the outcomes of classification.

**Feature Weights.** To investigate the impact of each feature in the classification outcome, Fisher scoring - a supervised feature ranking method - was applied. Features are standardized with their mean and standard deviation before applying the feature ranking and classification algorithm. This was done with MATLAB's Feature Selection Library<sup>211</sup>.

Fisher Discriminant Analysis. In this algorithm, a projected subspace where the data is well-separated is found by minimizing the within-class variances and maximizing the between-class variances<sup>212</sup>. All data are transformed to the new subspace to explore weights of each feature. Feature weight is the ratio of between-class variances to within-class variances using the following equation:

$$\frac{(\mu_1 - \mu_2)}{STD_1^2 + STD_2^2} \quad (2)$$

Here,  $\mu_1$ ,  $STD_1$ ,  $\mu_2$  and  $STD_2$  are the means and standard deviations of projected values for two classes (e.g. cancer vs non-cancer). Higher difference of mean with tighter distributions result in higher discriminative power. A random feature with uniform distributions was generated and applied to the feature ranking algorithms as a negative control. The higher weight indicates the more discriminating power, which allows us to rank features for cytoskeletal and morphological characterization.

**Feature Classification.** We used the Support Vector Machine (SVM) classification algorithm that can identify the optimal hyperplane that maximizes the separation of classes. However, data that was not linearly separable was mapped into a new space that does make the data linearly separable. This allows the SVM to perform an efficient non-linear classification with the so-called "Kernel trick". We applied four types of common kernels in SVM classification: linear, quadratic, third-

order polynomial, and the radial basis function (RBF) kernel. From here, we had two main approaches for classification.

**Pairwise Approach.** Classification with ten-fold cross validation was performed for all binary combinations of labeled cell lines. For each cell pair, each kernel was used with each feature class (4x4). Sampling the data points was repeated ten times and average prediction accuracy values were reported. Prediction accuracy is defined as the number of cell classes predicted correctly to the total number of cells in the test set. For B16F1-B16F10, HFF-HT1080, MCF10A and MDAMB231 sensitivity, specificity and area under the curve (AUC) are also reported. Sensitivity is the number of cases with cancer (HT1080 and MDAMB231) or a more aggressive type of cancer (B16F10) that was predicted correctly out of the total number of known cells in the test set. Specificity is the total number of non-cancerous (HFF and MCF10A) and less aggressive (B16F1) classifications against the total number of non/less aggressive-cancer cells in the test set. AUC is the area under the curve of the Receiver Operating Characteristics (ROC) curve. The ROC curve investigates the performance of the model for all possible classification thresholds. The closer AUC is to 1, the better classification outcome we have.

**Combination Approach.** Instead of individually labeling cancer cells by their respective cell line, all but one cancer cell line were pooled into one bulk cell line labeled “cancer”. The remaining cancer cell line was left out of the training step to be used as the test set (i.e. “Independent Test Set”). Thus, the test set, unlike the pairwise approach, is made of a cell line entirely not used during training. To balance the larger cancer cell set vs the smaller non-cancer cell set, sample sizes were matched into one comprehensive training set where the number of cancer cells matched the pool

of non-cancer cells by randomly selecting a subset of combined cancer cells. The remaining cancer cells were also used in testing (i.e. “Consistent Test Set”). Again, classification was performed on all four kernels and four feature classes. Random sampling was repeated twenty times and the average true positive rates of both the independent test set and consistent test set was reported.

**Statistical Analysis.** Student’s two-tailed t-test was used for statistical analysis. Pooled and unpooled t-test were used when the variances of two populations are equal and unequal, respectively. An F-test and a Levene test were used to study the homogeneity of the variances when both populations are normally distributed or otherwise, respectively. One-sample Kolmogorov-Smirnov test was used to study whether the data in each sample comes from a standard normal distribution.

## Chapter 5

### Profiling Phosphatase Arrays and Quantifying Focal Adhesion Dynamics

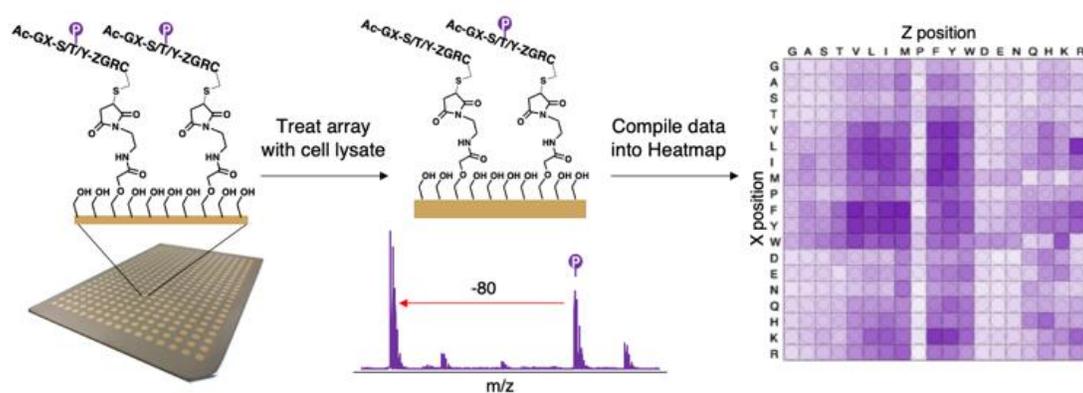
Research and figures presented in this chapter are adapted from work published in the following manuscript:

Szymczak, L. C.\*, **Sykora, D. J.** & Mrksich, M. Using peptide arrays to profile phosphatase activity in cell lysates. *Chem. Eur. J.* **26**, 165 (2020).

#### 5.1 Introduction

As I have mentioned previously, I joined Milan's group at a significant inflection point. The decade from 2005-2015 saw the graduate students and postdocs in the group work to create an assay platform that could quantitate biological processes in a high-throughput, label-free manner. This work resulted in a platform technology called self-assembled monolayers for matrix-assisted desorption/ionization mass spectrometry (SAMDI-MS). Here, alkylthiolates semi-covalently bond to a pure gold surface through a terminal cysteine. When these alkylthiolates present a hydrophilic group at the other end – a group such as poly(ethylene glycol) (PEG) – these chains self-assemble into a monolayer where the hydrophilic, protein adsorption-resistant PEG group extends out from the surface that contacts an aqueous environment. Within these alkylthiolates, reactive handles can be interspersed into the self-assembled monolayer (SAM), and these handles can serve as the basis for monitoring biological reactivity via mass spectrometry to quantitate the masses of compounds that form after reaction. After several dozen publications in the group around the development and utility of SAMDI, Milan founded the company SAMDI

Tech, Inc., which is based out of Illinois Institute of Technology's campus. An example of this workflow can be seen in **Figure 5.1**, where phosphorylated, cysteine-terminated peptides were reacted with exposed maleimides, and activity of a phosphatase enzyme was measured by the loss of phosphate mass at the surface from the ionized SAM.

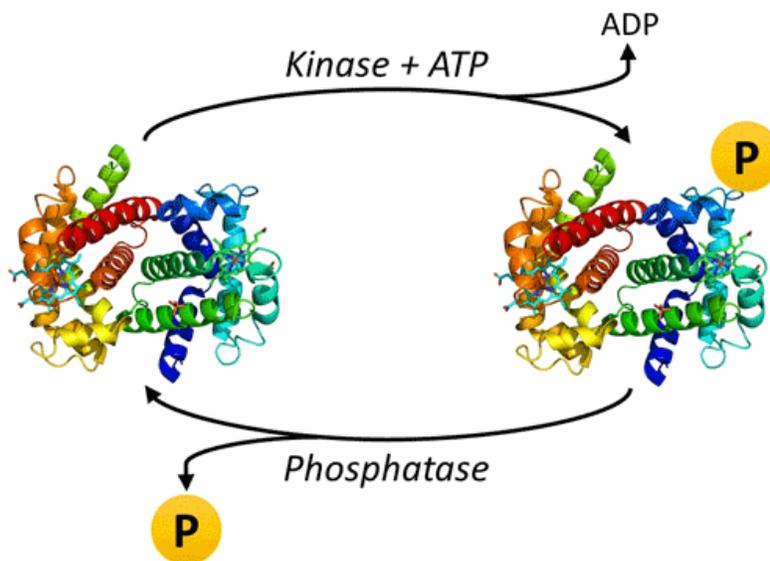


**Figure 5.1** General SAMDI-MS workflow to profile phosphatase activity on phospho-peptides self-assembled on a surface where the de-phosphorylation can be quantitated by a loss of 80 Da via MS.

Yet, as Justin started making headway with the development of megamolecules from 2012-onward, the focus of the lab began to shift away from SAMDI (as SAMDI Tech could now sufficiently handle development of commercially-viable ideas). This happened to coincide with many older graduate students finishing up remaining projects and defending their PhDs. One student, Lindsey Pearson (Szymczak), needed help with designing and executing a project involving a phospho-peptide library she had synthesized two years prior. Because of SAMDI's ability to probe biological reactions from cell lysate, we decided to utilize my several years' worth of experience maintainin many mammalian cell lines in parallel to probe phosphatase activity from several different cancer and non-cancer cell lines. This collaboration led to a co-author communication in *Chemistry – A European Journal* at the end of 2019.

## 5.2 Protein Phosphorylation in the Proteome

Protein phosphorylation, regulated by kinases (+ PO<sub>4</sub>) and phosphatases (- PO<sub>4</sub>) (**Figure 5.2**), is the most prominent reversible post-translational modification and is involved in almost all cellular processes including signaling, migration, proliferation, apoptosis, differentiation, and metabolism<sup>215-217</sup>. Furthermore, dysregulation of phosphorylation contributes to cancer, diabetes, and neurodegenerative and inflammatory disorders<sup>218</sup>. Proteomic studies have shown that phosphorylation occurs on over 30% of cellular proteins on serine (Ser, 86.4%), threonine (Thr, 11.8%), and tyrosine (Tyr, 1.8%) residues<sup>219,220</sup>. Serine and threonine phosphatases typically comprise one family while tyrosine phosphatases have been thought to concurrently evolve from another common precursor<sup>220</sup>.



**Figure 5.2** Kinases, in the presence of ATP, add a phosphate group to a protein while phosphatases remove a phosphate group from a protein.

The reasons and functional implications for this distribution of phosphorylated amino acids remain poorly understood. Tyrosine phosphorylation is known to be an important regulator of dynamic signaling events, and the lower levels of phosphotyrosine (pTyr) are consistent with the

comparable number of kinases (90) and phosphatases (107)<sup>221</sup>. The functional roles of Ser/Thr phosphorylation are less understood, and the reasons behind the abundance of phosphoserine (pSer) compared to phosphothreonine (pThr) are still poorly understood. The number of serine/threonine kinases (428) is far greater than the number of phosphatases (only 30)<sup>221</sup>. Furthermore, Ser/Thr phosphatases act as holoenzymes that form various complexes with large numbers of regulatory subunits to gain specificity<sup>222</sup>. The inherent challenges of studying Ser/Thr phosphatases and the lack of assays that can quantitate both phosphatase activity and substrate specificity in complex samples, such as cell lysates, help to explain why the majority of work is directed towards the roles of kinases<sup>223-228</sup>. In fact, most explanations for the regulation and distribution of phosphorylation sites have often emphasized or solely addressed changes in kinase activity with little consideration for the roles that phosphatases play. However, as the extent of phosphorylation depends on the balance of kinase and phosphatase activities, increases in kinase activity or corresponding decreases in phosphatase activity can both lead to greater phosphorylation.

In this first part of **Chapter 5**, phosphopeptide arrays were used to interrogate phosphatase activities in cell lysate with the aim of identifying general trends in the relative activities on pSer, pThr, and pTyr substrates. Peptide arrays have been important for the study of substrate specificity of enzymes, epitope mapping of antibodies, and protein binding interactions<sup>229-233</sup>. We prepared peptide arrays on SAMs suitable for SAMDI-MS, providing a label free assay to semi-quantitate phosphatase activity on phosphor-peptide substrates<sup>234-237</sup>. The most striking insight from this work is that Ser/Thr phosphatases are globally more active for the pThr peptide array than the peptides containing pSer. This work profiles Ser/Thr and Tyr phosphatase activity in cell lysates

using hundreds of distinct peptide substrates and has revealed activity trends that will be important for understanding the roles these phosphatases play in cellular processes.

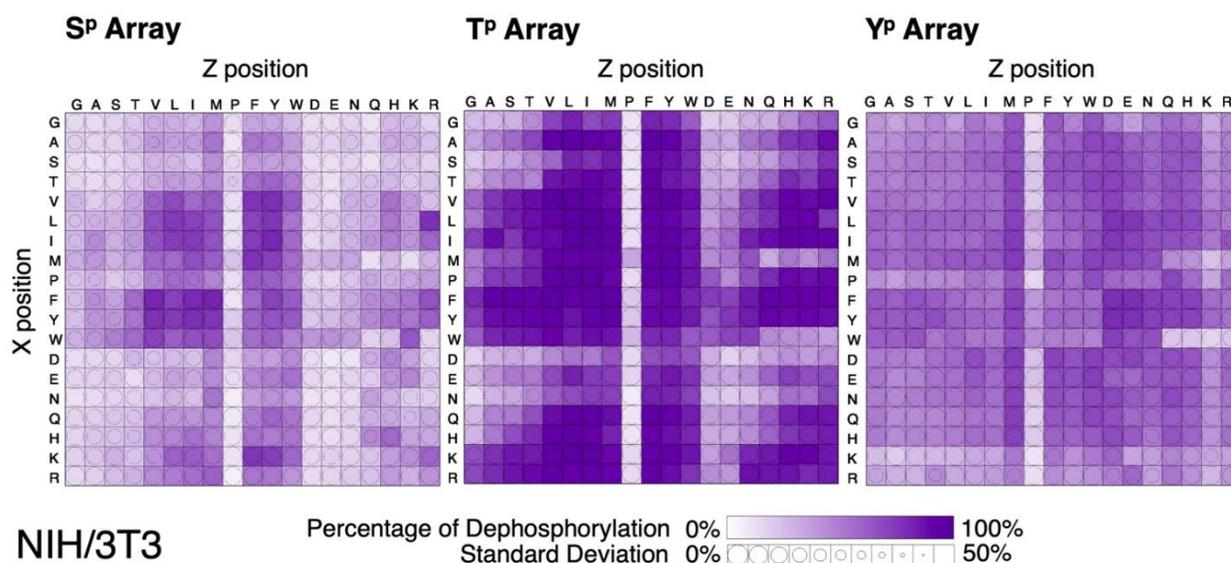
### 5.3 Results

We prepared phosphopeptide arrays on steel plates with 384 gold islands arranged in the standard microwell plate geometry, each modified with a SAM presenting a maleimide group against a background of protein-resistant EG3 groups, as described previously<sup>238,239</sup>. Lindsey used standard Fmoc-based solid-phase peptide synthesis to prepare pSer, pThr, and pTyr peptide libraries based on the following sequence: Ac-GXS<sup>P</sup>/T<sup>P</sup>/Y<sup>P</sup>ZGRC (where X and Z are variable positions and represent all natural amino acids except for cysteine). In this way, the peptides represent a broad distribution of sequences and can resolve very general specificities of phosphatases.

We next used each peptide array to profile phosphatases that were present in lysates derived from cultured NIH/3T3 cells. I applied the lysate to each spot on the array (1.5  $\mu$ L at 0.5 mg/mL total protein) and incubated for 15 minutes at 37 °C before the peptide array was rinsed with ethanol, water, ethanol and dried. During the incubation, endogenous phosphatase enzymes could dephosphorylate their corresponding active peptide substrates. I then applied MALDI matrix to the array plate and measured the extent of dephosphorylation of each peptide using SAMDI mass spectrometry (**Figure 5.3**).

The extent of dephosphorylation of each peptide was determined by dividing the area under the curve (AUC) of the dephosphorylated product peak by the sum of the AUC of the dephosphorylated product and phosphorylated substrate peaks, which is a standard methodology to quantify mass spectra data in our group. The extents of dephosphorylation of each peptide in

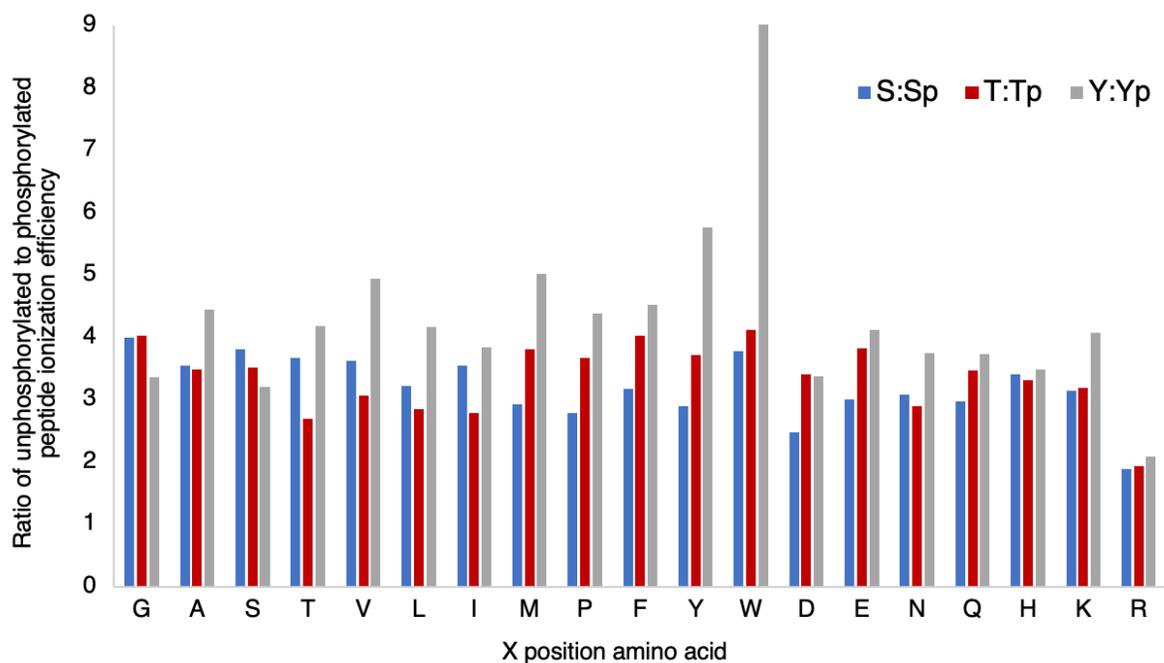
the array are displayed in heatmaps, where each square represents an individual peptide with variable X and Z residues – both proximal to the phosphorylate residue – on the horizontal and vertical axes, respectively. The heatmaps report the percent of dephosphorylation on a color scale, where dark purple represents 100% dephosphorylation, and the standard deviation from three replicates for each peptide is displayed by circle size within each peptide square—where a smaller circle corresponds to a larger standard deviation.



**Figure 5.3** Heatmaps showing global phosphatase activity in cell lysate from NIH/3T3 cells. The average extent of dephosphorylation of each peptide is plotted in the heatmap, where complete dephosphorylation is denoted by dark purple, and each square represents a peptide of sequence Ac-GX-S<sup>P</sup>/T<sup>P</sup>/Y<sup>P</sup>-ZGRC, where the X and Z residues are denoted on the vertical and horizontal axes, respectively. The standard deviation of the dephosphorylation of each peptide is displayed by circle size in each peptide square, with larger standard deviations resulting in smaller circles.

Inspection of the heatmaps reveals striking differences in activity and specificity between the pSer, pThr, and pTyr arrays. While Ser/Thr phosphatases act on both pSer and pThr substrates, as I mentioned previously, the levels of activity observed on the arrays are notably different. The substrate specificities observed on these two arrays are quite similar, but the activity on the pThr array is clearly higher than on the pSer array. The pSer and pThr arrays show preferential Ser/Thr phosphatase activity on substrates containing hydrophobic and/or aromatic amino acids in both the

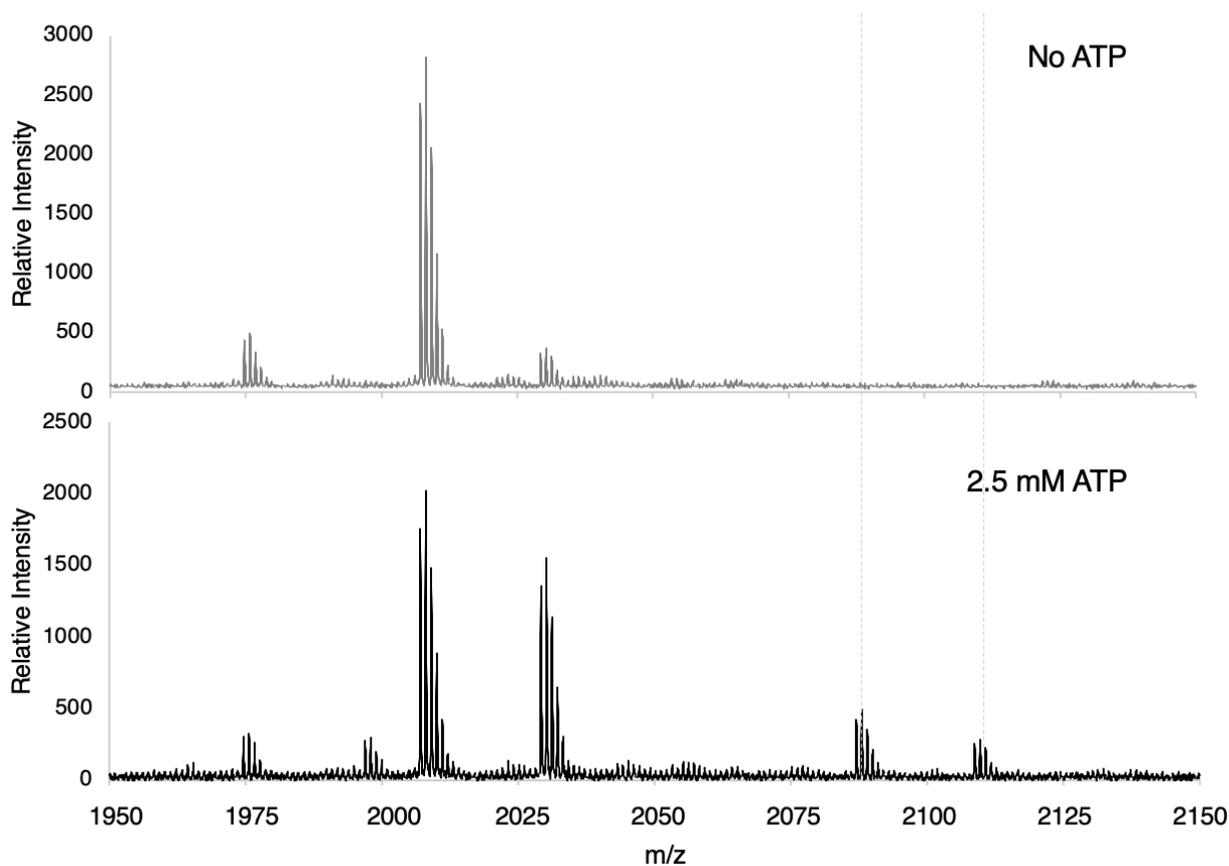
X and Z positions (V, L, I, M, P, F, Y, and W). Additionally, there were disfavored Ser/Thr phosphatase activity on peptides with combinations of G, A, S, T, D, E, N, and Q in both X and Z positions. The pTyr array revealed low Tyr phosphatase specificity where most peptides were equally dephosphorylated between 40-60%; however, there was slightly lower Tyr phosphatase activity on substrates with lysine or arginine in either variable position, which is consistent with previous findings<sup>240</sup>. Additionally, on all three arrays, there was little activity on peptides containing proline in the Z position, which likely introduces steric kinking and hindrance of the phosphorylated residue with the surface, resulting in low phosphatase activity.



**Figure 5.4** There are differences in ionization efficiencies of each peptide depending on the specific sequence; however, these differences seem to be largely consistent across each peptide library (i.e. phospho-serine, threonine, and tyrosine). The major exception is when tryptophan is in the X position within the phosphorylated tyrosine peptide; this phospho-peptide seems to ionize much worse than the unphosphorylated peptide.

It should be noted that there are differences in ionization efficiency between unphosphorylated and phosphorylated peptides; however, these differences are consistent between

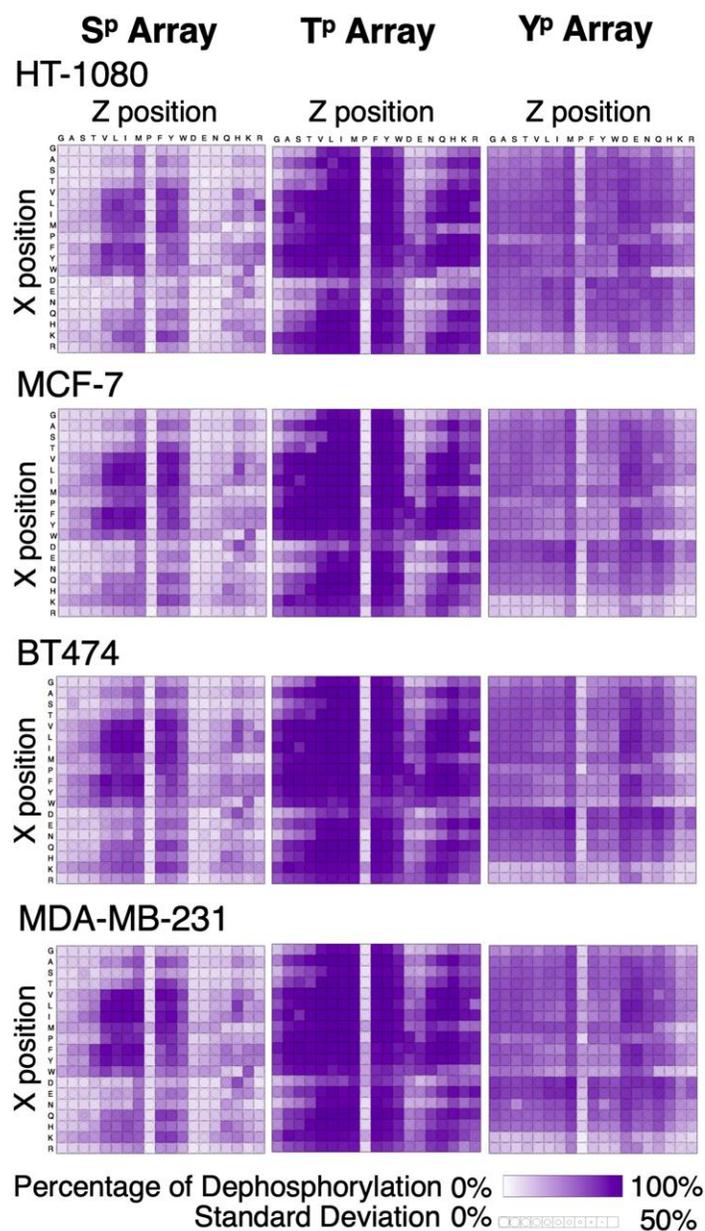
peptides of various sequences and therefore have a minimal effect on the trends in activity and sequence specificity that we observe (**Figure 5.4**) Additionally, we observe no kinase activity in our cell lysates without the addition of ATP, a required co-factor for kinase activity (**Figure 5.5**), demonstrating new phosphorylation is absent from the reaction of lysate at the surface.



**Figure 5.5** ATP is required for phosphorylation of peptides on SAMs; here, no new phosphorylation was seen with Src kinase or kinase from cell lysate towards its constitutive Src peptide sequence as well as one of our own peptides without ATP. This demonstrates that adding cell lysate to the SAM without large amounts (2.5mM) ATP is not introducing new phosphorylation against a backdrop of phosphatase-induced dephosphorylation.

Following analysis of NIH/3T3 cell lysates, we extended our study to measure phosphatase activity in cell lysates from four additional mammalian cell lines: HT-1080, MCF-7, BT474, and MDA-MB-231. I chose these cell lines because they compare specificity and activity from a

murine non-cancer cell line (NIH/3T3) to human cancer cells, including two different cancer types (mesenchymal – HT1080 vs. epithelial) and breast cancer cells of varied gene expression and metastatic potential. Each cell lysate was applied to the pSer, pThr, and pTyr peptide arrays, and the arrays were again analyzed by MALDI-MS and reported in heatmaps (**Figure 5.6**).



**Figure 5.6** The activity and specificity of phosphatases towards each peptide library are consistent across several disparate cell lines.

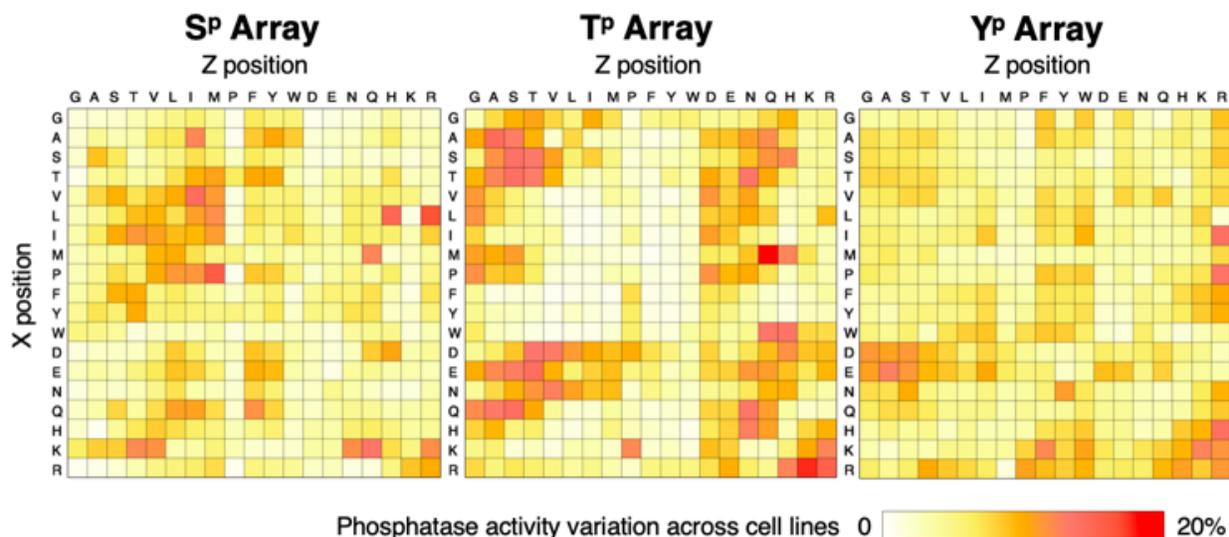
The global phosphatase profiles for each lysate again show preferential activity on pThr compared to pSer substrates. As mentioned earlier, several proteomic studies have found that in various cellular conditions, phosphorylation occurs with ~86% on Ser, ~12% on Thr, and ~2% on Tyr residues<sup>3</sup>. Proteomic studies for the detection of phosphorylated sites offer an unbiased view of the *in vivo* proteome; however, they do not provide mechanistic information such as the enzymes responsible for the addition and removal of the phosphorylation modification or the stability and regulatory roles of each phosphorylation site. Previous reports suggest the phosphoproteome distribution across Ser, Thr, and Tyr residues likely results from the higher number of Ser/Thr kinases and their preferences for Ser and Thr as phosphoacceptor residues<sup>241-243</sup>. While this may be true, these results suggest that phosphatases, and more specifically the Ser/Thr phosphatases, may play a more significant role in determining the phosphoproteome distribution than previously believed and can help explain the high levels of pSer observed.

#### **5.4 Conclusions of Peptide Work**

The high specificity and lower activity of phosphatases on the pSer array may indicate that phosphorylation on Ser is more regulated than on Thr and/or that Ser/Thr phosphatases generally have a stronger preference for pThr over pSer substrates. Either of these conclusions could explain why proteomic studies report the highest level of phosphorylation on Ser. Merlevede and coworkers found that members of the PP2A family of Ser/Thr phosphatases preferred pThr over pSer substrates and that phosphatase activity on pSer could be increased by changing the proximal amino acid sequences surrounding the pSer, while activity on pThr was less affected by changes in proximal amino acid sequences<sup>244</sup>. Additionally, Burgess and coworkers observed in a proteomic study that pThr sites with proline in the +1 position were dephosphorylated at a higher

rate than pSer-proline motifs<sup>245</sup>. These results are consistent with these findings that phosphatase activity is lower on pSer than on pThr substrates. This example is a significant reminder that kinases and phosphatases both play important roles in determining global phosphorylation.

I was surprised by the similar phosphatase profiles across five different cell lines. To visualize the variance in phosphatase activity across the five cell lysates, the standard deviation of the average percentages of dephosphorylation from all five cell lysates for each peptide was calculated. Unsurprisingly, variation in phosphatase activity on all peptides was less than 20%, with most peptides having a standard deviation between 1-10% (**Figure 5.7**). The standard deviation of the average percentages of dephosphorylation from each cell lysate was calculated for each peptide and are shown on a color scale range of white-yellow-orange-red, where white represents 0% and red represents 20% standard deviation.



**Figure 5.7** Phosphatase activity did not significantly vary across each individual peptide library; the largest deviation of activity was about 20%.

I naively anticipated that comparing cell lines across species (e.g. murine NIH/3T3 vs human HT-1080) or across different breast cancer lines (e.g. MCF-7 v. MDA-MB-231) would reveal a class/family of phosphatases uniquely targeting a specific peptide group that was

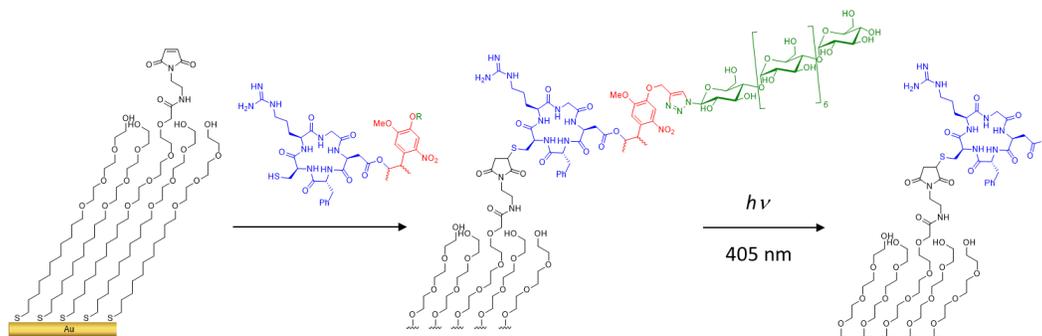
over/under expressed. However, many proteomic studies that have monitored global phosphorylation in various cellular processes found small variations across a myriad of conditions<sup>246</sup>. For instance, proteomic studies that examined global phosphorylation changes in various mouse tissues<sup>247</sup> and during mitotic exit<sup>245</sup>, epidermal growth factor stimulation<sup>218</sup>, and DNA damage response<sup>248</sup> all found that *global* phosphorylation was only altered between 10-15% and is consistent with our observation of little variation in phosphatase activity in the five different cell lysates. The similar, but specific phosphatase profiles that we observed likely indicate that phosphatase activity is similarly regulated in different cell types and could also suggest that many phosphorylation sites are non-functional. The mechanisms behind phosphatase specificity remain largely unknown; however, it is clear from our results, as well as from others, that phosphatases are highly regulated enzymes<sup>249</sup>.

Here, peptide arrays and SAMDI-MS were utilized in tandem to observe and differentiate phosphatase activity on Ser, Thr, and Tyr residues on more than 1,000 peptides for lysate that I derived from five cell lines. It was found that phosphatase activity is lower on pSer in comparison to pThr substrates, which may suggest that the phosphorylation distribution across Ser, Thr, and Tyr residues is largely impacted by phosphatase activity, rather than differential activity of kinases alone. It is clear that phosphatases have significant regulatory roles in the cell; however, further studies on the substrate specificities and dynamics between kinases and phosphatases are necessary to fully decipher the mechanisms behind phosphorylation and the regulatory roles of each site.

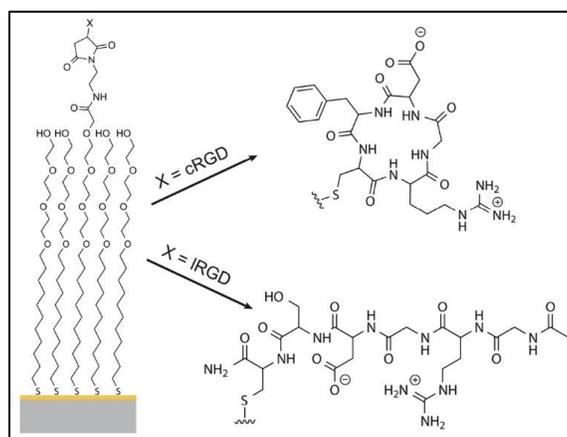
## 5.5 Quantifying Focal Adhesion Dynamics

One final project I wanted to briefly address is work I completed my first summer that was not novel enough for sole publication, but still marked, for me, the first time I explored a project

to an adequate level of completion. I worked on this project somewhat concurrently with the time I spent with Lindsey on the phosphatase project. Here, I similarly utilized SAMs to present the cell-binding peptide arginine-glycine-aspartic acid (RGD). RGD can present itself in a linear form (IRGD) and a cyclized form (cRGD); cell integrins have ~100-fold greater binding affinity for cRGD over IRGD<sup>250</sup>. A previous graduate student in the group, Pradeep Bugga, had synthesized a sugar-protected version of cRGD that is activated through photoactivation (**Figure 5.8**). He elegantly demonstrated<sup>251</sup> that this peptide could be selectively deprotected into shapes, allowing cells to spread into these specific shapes. This is not unlike my use of  $\mu$ CP for patterning cells in **Chapter 4**. However, I wanted to analyze how shifting the peptide from linear RGD to cyclic RGD in real-time affected cell adhesion dynamics.

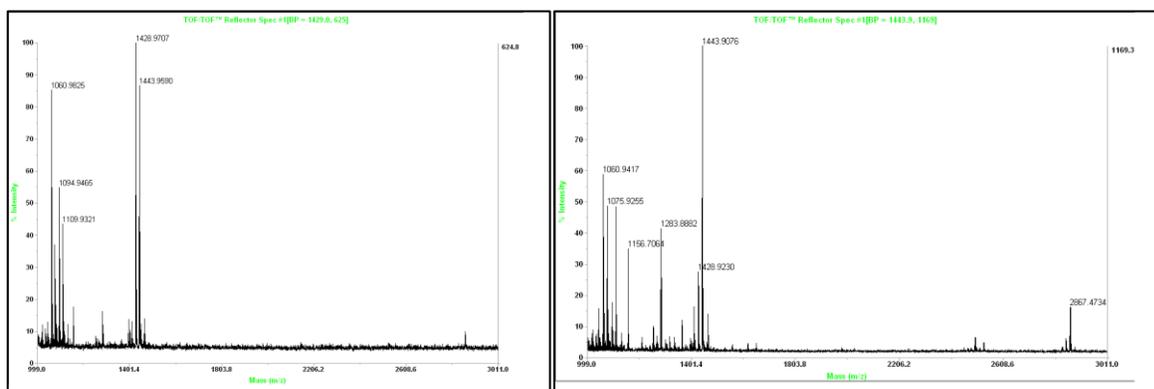


**Figure 5.8** A nitrophenyl-based hepta-saccharide protecting group is released when exposed to 405nm light, revealing cRGD to cells to bind.

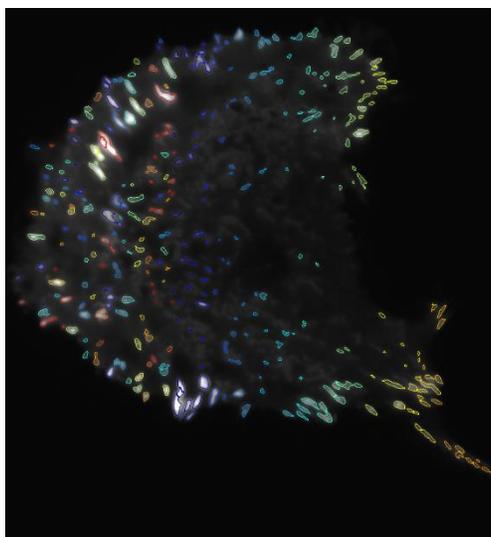


**Figure 5.9** Once photo-deprotected, cells can concurrently sense both linear and cyclic RGD.

By mixing IRGD and this photo-protected cRGD on the same SAM, I could culture cells and allow them to initially adhere to IRGD. Then, using 405nm light, I could de-protect the cRGD, and subsequently cells could concurrently sense both forms of RGD (**Figure 5.9**). Then, after transfecting cells *a priori* with a lentiviral GFP-paxillin gene, I could quantify cell adhesion size via total internal reflection fluorescence (TIRF) microscopy after this substrate switch. Mass spectrometry by MALDI revealed how 405nm deprotection uncages cRGD for binding (**Figure 5.10**).



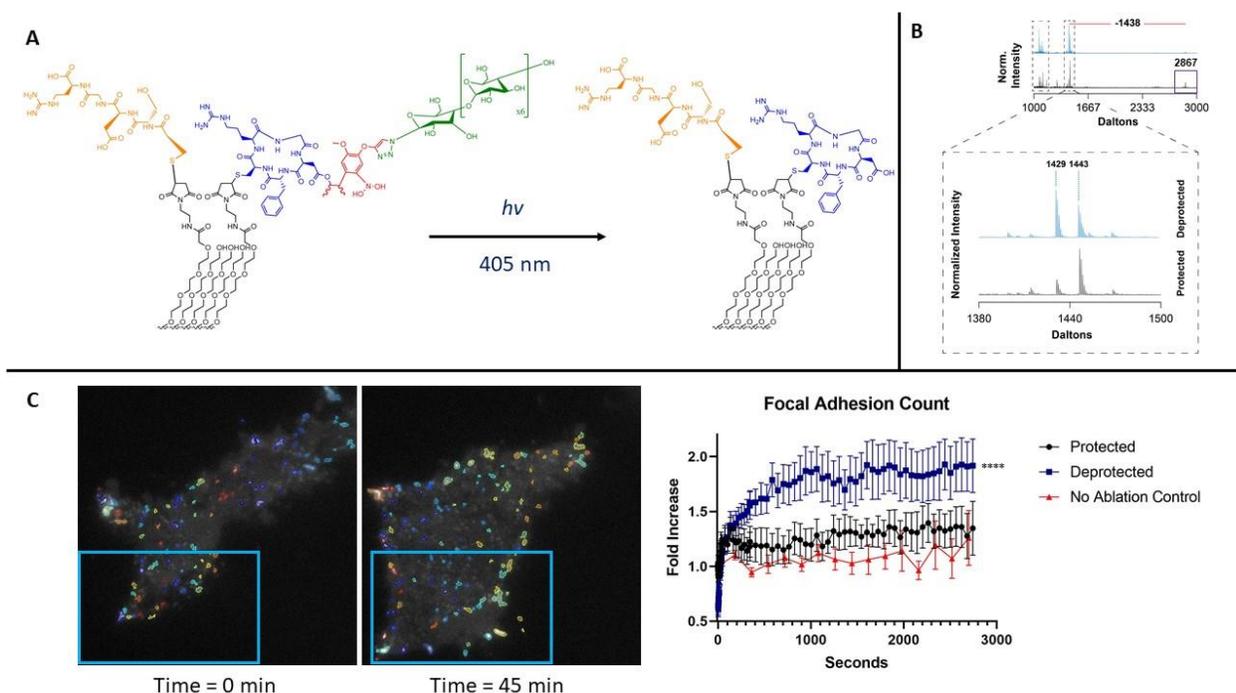
**Figure 5.10** The protected cRGD peptide (2867 Da, left, ionizes poorly) gets deprotected, revealing cRGD (1429 Da) in roughly equivalent amounts to linear RGD (1444 Da) (right).



**Figure 5.11** A GFP-paxillin transfected HT-1080 can have its focal adhesions spatiotemporally quantified.

The HT-1080 fibrosarcoma cell line was used to quantify cell adhesion dynamics, and they were transfected with GFP-Paxillin lentiviral particles with 8 ug/mL polybrene and let to incubate overnight for 16-20 hours. Then, cells were plated on the mixed linear RGD / photoprotected cRGD presented by SAMs on a gold monolayer over a No. 1.5 glass slide. I let them adhere overnight as well. Through TIRF, cells were quantified before and after photo-deprotection, and their focal adhesions were quantified with the *Focal Adhesion Analysis Server (FAAS)*<sup>252</sup>. An example image quantifying focal adhesions can be seen in **Figure 5.11**, and specific traits included size, aspect ratio, and lifetime (i.e. how long does it last).

Ultimately, I was able to demonstrate how cells immediately protruded into the region of deprotection, which did not happen if I used a high wavelength of light (e.g. 564nm) or did not deprotect at all (**Figure 5.12**). Through utility of our SAMs presenting ligands for cell binding alongside lentiviral particle transfection and an online quantitative server, I was able to demonstrate a novel way to quantify cell adhesion. Sadly, this project did not progress further, as the impact of this work was not high enough to justify the high cost to synthesize the sugar-protected cRGD, which comprised the majority effort of Pradeep's dissertation to make.



**Figure 5.12** **A**) Photoactivation reveals both linear and cyclic RGD for cells to bind, **B**) MALDI reveals the loss of mass of the sugar-protected RGD (which ionizes poorly), similar to Figure 5.10, **C**) Cells underwent an increase in adhesion (by count) and protruded into the region of deprotected cRGD as opposed to cells with no deprotection.

## 5.6 Experimental

**Peptide Library Synthesis.** Peptide libraries Ac-GXS<sup>P</sup>ZGRC, Ac-GXT<sup>P</sup>ZGRC, and Ac-GXY<sup>P</sup>ZGRC were synthesized using standard Fmoc-based solid-phase peptide synthesis<sup>[1]</sup> on Fmoc-Rink Amide MBHA resin (Anaspec). Synthesis was performed in 96-well filter plates (Arctic White) with the use of a 96-well plate vacuum manifold (Millipore). Fmoc- and side chain-protected amino acids and N-acetyl-glycine were purchased from either Anaspec or Sigma-Aldrich. Phosphorylated amino acids: Fmoc-Ser(PO(OBzl)OH)-OH, Fmoc-Thr(PO(OBzl)OH)-OH, and Fmoc-Tyr(PO(OBzl)OH)-OH were purchased from Sigma-Aldrich. Coupling reagents Pybop and N-methylmorpholine (NMM), as well as dimethylformamide (DMF), were purchased from Fisher or Sigma-Aldrich.

For each library, 15 mg of resin was put into 361 wells in four 96 well plates. Resin was swelled for 1 hour in DMF. The resin was Fmoc-deprotected with a solution of 20% piperidine in DMF for 20 minutes. Using vacuum filtration, the wells were drained, and the resin was rinsed with DMF. The first coupling reaction was performed for 20 minutes with Fmoc-Cys(Trt)-OH, Pybop, and NMM in 250  $\mu$ L DMF at a molar excess of 8:8:16, respectively, to the molar loading capacity of the resin. The wells were then drained, and the resin was rinsed with DMF. The coupling was repeated once more to ensure complete cysteine addition. The wells were again drained, and the resin was rinsed with DMF. Fmoc-deprotection, washing, and double coupling reactions with respective amino acids were repeated until the last amino acid (N-acetyl-glycine) was coupled. We note that  $\beta$ -elimination can occur on protected phospho- threonine and serine amino acids during synthesis. To prevent  $\beta$ -elimination after coupling, the protected phosphorylated serine and threonine, the remaining Fmoc-deprotections, and couplings were performed very carefully and strictly for only 20 minutes. After the last coupling reaction, the resin was rinsed with DMF and dichloromethane and dried under vacuum for 1 hour. Amino acid side chain deprotection and peptide cleavage from the resin was performed in tandem for 2 hours using a solution of 2.5% triethylsilane, 2.5% H<sub>2</sub>O, and 95% trifluoroacetic acid (TFA). The cleavage solutions for each peptide were then transferred into new 96-well plates and then evaporated under nitrogen. The peptides were resuspended in 0.1% TFA in H<sub>2</sub>O and lyophilized. We found that phosphorylated peptides do not store well in water. After lyophilization the peptides were resuspended again in 0.1% TFA in H<sub>2</sub>O to a final concentration of 500  $\mu$ M, and 3  $\mu$ L of each peptide were transferred into several 50  $\mu$ L 384 plates. All peptide library plates were again lyophilized and stored at -80 °C until needed for peptide array preparation.

**Additional Peptide Synthesis.** Fifty-eight additional peptides were synthesized—one of which was used for ATP-dependence kinase activity control experiments with Src: Ac-IYGEFKKKC. The remaining fifty-seven peptides were used to calculate relative MALDI ionization efficiency ratios between the phosphorylated and unphosphorylated peptide peaks and consisted of the sequence Ac-GX-S/T/Y-GGRC, where X is a variable position and represents all natural amino acids except for cysteine. All peptides were synthesized similarly to the peptide libraries using standard Fmoc-based solid-phase peptide synthesis on Fmoc-Rink Amide MBHA resin. The resin was swelled for 60 minutes in DMF. Fmoc-deprotection and amino acid coupling reactions were performed as described above for the peptide library synthesis. The peptides were cleaved, and the amino acid side chains were deprotected in a solution of 2.5% triethylsilane, 2.5% H<sub>2</sub>O, and 95% TFA for 2 hours. The cleaved peptides were filtered to remove the resin and then precipitated with ethyl ether. The precipitated peptides were resuspended in 0.1% TFA in H<sub>2</sub>O and lyophilized. They were then resuspended in 0.1% TFA to a final concentration of 500 μM and stored at -20 °C until needed.

**Preparation of SAMDI Peptide Arrays.** Preparation of SAMDI peptide arrays have been described previously.<sup>[2]</sup> Briefly, using a Thermionics E-beam evaporator, 384 titanium spots were evaporated onto steel plates. Gold was then evaporated over the titanium spots, resulting in array plates of 384 gold spots. The plates were soaked in a 1 mM total disulfide monolayer solution of 0.8 mM tri(ethylene glycol) disulfide and 0.2 mM tri(ethylene glycol)-maleimide disulfide in ethanol, allowing for self-assembly of an alkanethiolate monolayer onto the gold surfaces. The monolayer presents a functional maleimide group at a density of 10% against a background of tri(ethylene glycol), which prevents non-specific protein adsorption to the surface. The lyophilized

peptide libraries were resuspended in 50 mM Tris pH 7.5 to a concentration of 50  $\mu$ M and 3  $\mu$ L of each peptide solution was pipetted onto the gold spots of array plates. The peptides were incubated on the surface of array plates in a humidified chamber at room temperature for 1 hour for immobilization. Peptide immobilization occurs through 1,4-Michael addition of the cysteine-thiol side chain to the maleimide. The arrays were then washed with H<sub>2</sub>O and then ethanol, dried under nitrogen, and stored in vacuum sealed bags at 4°C until ready for use.

**Cell Culture.** Cell lines were acquired by ATCC. For preparation of lysates, cells were grown and passaged in similar conditions. HT-1080, BT-474, MCF-7, and MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco). NIH/3T3 cells were supplemented with 10% bovine calf serum (Gibco) instead. All cells were grown at 37 °C and 5% CO<sub>2</sub>. DMEM for all cells came supplemented with 4.5 g/L D-glucose, 4 mM L-glutamine, and 110 mg/L sodium pyruvate.

**Cell Lysis.** Cells were grown to ~90% confluency for lysis. Cells were resuspended in cold phosphate-buffered saline (Fisher Scientific) on ice following trypsinization. After pelleting this solution, cells were lysed in cold lysis buffer (25 mM HEPES pH 7.4, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 0.5% Triton-X 100, EDTA-free protease inhibitor (Roche, Mini cOmplete™, 1 tablet per 10 mL lysis buffer) on ice for 10 minutes. The DTT and protease inhibitor tablets were added fresh before every lysate collection. The lysate solution was put into a 7 mL dounce homogenizer (KONTES), where it was sheared with 25 strokes on ice. Lysate was spun at 1500 g for 3 minutes, and the supernatant was removed. The supernatant was spun again for 20 minutes at 14,000 g and 4 °C. The remaining supernatant collected was the final lysate used in our

experiments and was stored on ice before use. To estimate lysate protein concentrations for experimentation, a BCA assay (Pierce) was performed using a NanoDrop™ Spectrophotometer (Thermo Scientific). Typical protein concentrations were around 2-4 mg/mL from a near-confluent 225 cm<sup>2</sup> flask. Cell lysates were flash frozen in liquid nitrogen and stored at -80 °C until needed for experiments.

**Lysate Array Experiments.** Cell lysates were thawed on ice and then diluted to a total protein concentration of 0.5 mg/mL in lysis buffer. 5 mM MgCl<sub>2</sub>, 2.5 mM MnCl<sub>2</sub>, and 1 mM CaCl<sub>2</sub> were added to the lysate solution. A multidrop Combi Reagent Dispenser was used to pipette 1.5 μL of lysate onto each peptide spot on the phospho- serine, threonine, and tyrosine SAMDI peptide arrays and incubated in a humidified chamber for 15 minutes at 37 °C. The arrays were then rinsed with a solution of 0.5% w/v of Alconox detergent (Sigma-Aldrich), then water and dried under nitrogen. The plates were stored in vacuum sealed bags at 4 °C until analyzed by SAMDI-MS. Cell lysate experiments were repeated on all three phosphorylated peptide arrays three times for all five cell lines.

**Measuring ionization efficiency of phosphorylated vs unphosphorylated peptides (Figure 5.4):** The 57 unphosphorylated peptides consisting of the sequence Ac-GX-S/T/Y-GGRC, where X represents all amino acids except for cysteine, were neutralized by dilution in 50 mM Tris, pH 7.5 to a final concentration of 50 μM. The unphosphorylated peptides and their phosphorylated counter peptides from the arrays were each immobilized onto individual gold spots on a SAMDI array plate as described above. The plate was rinsed with water, and then ethanol, and dried under nitrogen. Each peptide was analyzed by SAMDI-MS as described below. The area under the curve

(AUC) of all adducts of the peptide substrate was standardized against the area under the curve of the adducts from the EG3 background peaks. The EG3 peak is a good standard because it is at 90% density on every SAM, and its relative intensity to the immobilized peptide is constant from spectra to spectra. The area under the curve of the peptide was divided by the area under the curve of EG3, and ratios between the phosphorylated and unphosphorylated peptides were generated. We found that most unphosphorylated peptides had an ionization efficiency about 2-4 times greater than that of the phosphorylated peptide.

**Demonstrating that ATP is required for kinase activity in cell lysates (Figure 5.5):** We synthesized one peptide that is a known biological substrate of the common tyrosine kinase, Src: Ac-IYGEFKKKC.<sup>[3]</sup> The peptide was neutralized by dilution in 50 mM Tris, pH 7.5 to a final concentration of 50  $\mu$ M, and immobilized onto 6 spots on a SAMDI array plate. The peptide was treated with lysate solutions with and without ATP. NIH/3T3 cells were lysed in the same lysis buffer used to measure phosphatase activity with the addition of phosphatase inhibitors: 80 mM sodium fluoride, 100 mM  $\beta$ -glycerol phosphate, 50 mM sodium pyrophosphate, and 5 mM sodium orthovanadate, to prevent all phosphatase activity. The lysate was diluted to a total protein concentration of 0.5 mg/mL. To make sure kinase activity could be observed, we added 100 nM of active Src kinase (Sigma-Aldrich) and 20 mM  $MgCl_2$  to the lysate. The immobilized peptide was treated with the lysate and Src solution with and without the addition of 2.5 mM ATP for 1.5 hours at 37 °C. The SAMDI plate was rinsed with a solution of 0.5% w/v of Alconox detergent, then water, and dried under nitrogen. The plate was analyzed by SAMDI-MS as described below, and we only observed kinase activity with the addition of ATP.

**SAMDI-MS.** A matrix solution of 10 mg/mL 2,4,6-Trihydroxyacetophenone (THAP), 5 mg/mL ammonium citrate dibasic in 0.1% TFA in 50% acetonitrile (ACN) and 50% water was prepared fresh. 1  $\mu$ L of matrix solution was applied to each spot on all treated peptide array plates using a multidrop Combi Reagent Dispenser, and the matrix crystallized at room temperature and pressure over 15 minutes. The arrays were analyzed in reflector positive mode on an AB Sciex 5800 MALDI TOF/TOF. Spectra were collected from 900 laser shots with a stage velocity of 2000  $\mu$ m/sec.

**Data Analysis.** The proportion of dephosphorylation (or phosphorylation for control experiments) on each peptide was calculated by dividing the area under the curve (AUC) of the dephosphorylated peptide product divided by the AUC of both the dephosphorylated product and phosphorylated substrate. All replicates were averaged in presented results. The peptide array data is also displayed in heat maps, which are a graphical representation of the average and standard deviation of the proportion of dephosphorylation of each peptide in the array. Each square represents an individual peptide in the array with the X and Z position amino acids on the vertical axis and horizontal axis, respectively. Dark purple represents complete dephosphorylation and the circle size inversely represents standard deviation.

**SAM Preparation for Cell Adhesion Study.** Gold (200 $\text{\AA}$ ) was evaporated on a No. 1.5 glass slide after 50 $\text{\AA}$  of titanium as described previously<sup>37</sup>. These glass slides were soaked in EG<sub>3</sub>-maleimide (1% or 8% to EG<sub>3</sub>) solutions) overnight. Each peptide, as generated previously<sup>37</sup> were stored in 1:1 ACN/H<sub>2</sub>O and are diluted in 10% Tris, pH 8. Then, they are incubated at room temperature for 30m. After triple washing with ethanol, water, ethanol, they could be used for cell culture.

**RGD Surface SAM MALDI.** Positive-mode MALDI-MS was used to confirm both peptide identity and successful deprotection after peptide immobilization. MALDI matrix consisted of 2:1 THAP:ammonium citrate diluted in 1:1 ACN/H<sub>2</sub>O with 0.1% TFA. Deprotection was carried out *a priori* by focused 405nm light on out Nikon Ti Eclipse confocal microscope into an allocated region on the glass slide to be compared to a region that was not deprotected.

**Cell Culture.** Similar to all previous studies, HT-1080 cells were cultured at 5% CO<sub>2</sub> in 10% FBS in DMEM supplanted with L-glutamine and 1% pen/strep. For imaging, cells were cultured in FluoroBrite DMEM to minimize background noise in the 488nm channel.

**Cell Transfection.** HT-1080s (and briefly, BHK-21s) were transfected via LentiBrite™ GFP-paxillin lentiviral particles (Millipore Sigma) supplanted with 8 uL/mL polybrene for increased transfection efficiency. An MOI of 40 (e.g. 40 viral particles per cell) was used based on the tier of the vial. All labware that touched viral particles was bleached before being discarded.

**TIRF.** TIRF was carried out at the Center for Advanced Microscopy (CAM) on the Chicago campus with help from Dr. Dina Arvanitis with their Nikon A1R confocal microscope with TIRF add-on. Regions of deprotection to be illuminated by 405nm light were established on the NIS Elements software and images were acquired at 488nm every 5m for up to two hours.

## Chapter 6

### Conclusions, Insights, and Future Directions

#### 6.1 Megamolecule Thoughts

The five years I spent in the Mrksich group have been nothing short of rewarding, and I am eager to see where my colleagues next take megamolecules. I believe my work in **Chapter 2** decoupling antibody affinity, avidity, net internalization rate, and inhibition of cell proliferation may be the strongest demonstration of how a modular protein engineering platform like megamolecules can provide mechanistic insight into fundamental cellular processes. This toolbox of reactive enzymes for specific chemical warheads allows us to create an array of targeting protein scaffolds of various geometries, specificities, and functionalities. With the insights gained from the work – in both **Chapter 2 and Chapter 3** – we can really begin to explore the complex geometric space for programmability in *both* therapeutics and diagnostics.

I believe it will be vital to get mammalian expression off the ground. As I mentioned in **Chapter 1**, none of our antibody mimics have Fc domains; not only do these domains extend the half-life of antibodies and induce cytotoxicity and phagocytosis via NK cells and macrophages, respectively, but these constant regions are also heavily N-glycosylated<sup>253</sup>. The patterning of various sugar groups affects the solubility of the antibody, decreases aggregation, as influences how effector various cells interact with it. These sugar groups are absent from *E. Coli* expressed proteins, while most biologics expressed in industry use CHO or HEK cell lines do impart these carbohydrates onto expressed proteins, such as albumin<sup>354</sup>. Testing specific immune-compatible carbohydrate patterns from a similar structure-function point of view will be vital for furthering this platform to something potentially usable by a pharmaceutical company. Mammalian expression is also favorable when expressing larger proteins – like albumin – or proteins that have

many disulfides – like antibodies, though the trade-off is a slower throughput and higher cost to that of *E. coli* bacterial expression systems. I also believe it will be critical for Bethel’s lipid conjugation project to show robust improvement of *in vivo* half-life of megamolecule-based therapeutics.

Furthermore, I am interested to see other use-cases for megamolecules outside of the context of cancer, which is something that Justin, Blaise, Sraeyes, and myself have largely focused our efforts on. I am very excited to see how megamolecules could be used as a diagnostic, like in the quaternary-scale protein switch work discussed in **Chapter 3** or Tori’s prostate cancer imaging work with PSMA-targeting ligands. The modularity of domains that can plug-in a reagent of interest may portend greater flexibility and use (at a lower barrier of entry) than a full therapeutic, and it is worth developing megamolecules into this space. Additionally, as I have found with the nanobody-based biparatopic antibodies that rapidly internalize but do not inhibit proliferation on their own – creating an ADC platform with the various scaffolds I have synthesized would also be of interest – and potentially a low-hanging fruit project for a new graduate student! Sraeyes and I had also talked about creating cytokine fusions for inflammatory disorders, and we spent some cycles investigating macrophage polarization in the summer of 2019. I ultimately got a lot of sufficient qPCR data for polarizing macrophages (data not shown here), but we never took that project further as other projects took precedence. Looking into additional therapeutic directions beyond cancer (say, in the inflammation space) may also yield novel insight and exciting avenues to pursue.

Ultimately, most of this work will be dependent on the success of Justin’s “Megamolecules 2.0” work he has been pioneering since the advent of the COVID-19 pandemic. These small, humanized nanobody domains will not only replace the larger, patent-protected, non-totally-

human enzymes like cutinase and SnapTag, but having > 4 orthogonal, new chemistries will allow for greater freedom in the biologic design space that is enticing not only from a pure research perspective, but also from a commercial one as well. I wish him, and everyone else in lab, the best of luck at maintaining this newfound momentum we generated over the last year.

## 6.2 Final Research Thoughts

While joining the lab at an inflection point allowed me to get my hands wet with a wide array of disparate research, it also, sadly, came with the territory of watching a previously-robust academic space fizzle out. SAMDI was a platform in full swing when I joined the group in 2017; fast forward five years, and our group is in full megamolecules-mode, which is exciting! Yet, I'm confident that SAMDI could be still be used as an auxiliary method to probe directed evolution efforts of 1) evolved proteins incorporated into megamolecule scaffolds for specific antigen targeting, like seen in Adam's work<sup>255</sup>, or 2) new enzymes (or covalent nanobodies via MM 2.0) to use in our toolbox. Having a high-throughput methodology to complement the (currently) low(er)-throughput work of megamolecules may streamline best-in-class candidates for the in-depth QC that pharmaceutical research requires.

Again, I am grateful for Milan's support over the last five years, and I am quite excited to see how our groundwork laid for ModuMab will impress and pique the interest of life science investors.

## References

1. Gardner, T., Cantor, C. & Collins, J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
2. Elowitz, M. B., Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
3. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).
4. Smith, H.O., Hutchinson, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly:  $\phi$ X174 bacteriophage from synthetic oligonucleotides. *PNAS* **100(26)**, 15440–15445 (2003).
5. Tian, J., Gong, H., Sheng, N. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
6. Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
7. May, M. Engineering a new business. *Nat. Biotechnol.* **27**, 1112–1120 (2009).
8. Young, E. & Alper, H. Synthetic biology: Tools to design, build, and optimize cellular processes. *J. Biomed. Biotechnol.* **2010**, (2010).
9. Ryu, M.-H. *et al.* Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* **5**, 314–330 (2020).
10. Sun, G. L., Reynolds, E. E. & Belcher, A. M. Using yeast to sustainably remediate and extract heavy metals from waste waters. *Nat. Sustain.* **3**, 303–311 (2020).
11. Heveran, C. M. *et al.* Biomineralization and successive regeneration of engineered living building materials. *Matter* **2**, 481–494 (2020).
12. Smith, R. S. H. *et al.* Hybrid living materials: digital design and fabrication of 3D multimaterial structures with programmable biohybrid surfaces. *Adv. Funct. Mater.* **30**, 1907401 (2020).
13. Petesch, S. J. & Lis, J. T. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* **134(1)**, 74–84 (2008).
14. Faure, G., Ogurtsov, A. Y., Shabalina, S. A. & Koonin, E. V. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res.* **44**, 10898–10911 (2016).

15. Endo, Y. Structure and function of ribosomal RNA. *Tanpakushitsu Kakusan Koso*. **40**, 1485–1493 (1995).
16. Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
17. Geary, C., Rothmund, P. W. K. & Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799-804 (2014).
18. Brunette T. J., Parmeggiani, F., Huang P. S., Bhabha, G., Ekiert, D. C., Tsutakawa, S. E., Hura, G. L., Tainer, J. A. & Baker, D. Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580-584 (2015).
19. Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B. L. & Bradley, P. Rational design of  $\alpha$ -helical tandem repeat proteins with closed architectures. *Nature* **528**, 585-588 (2015).
20. Rocklin G. J., Chidyausiku, T. M., Goreshnik, I., Ford, A., Houliston, S., Lemak, A., Carter, L., Ravichandran, R., Mulligan, V. K., Chevalier, A., Arrowsmith, C. H. & Baker, D. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168-175 (2017).
21. Langan, R. A., Boyken, S. E., Ng, A. H., Samson, J. A., Dods, G., Westbrook, A. M., Nguyen, T. H., Lajoie, M. J., Chen, Z., Berger, S. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205-210 (2019).
22. Gao, X. J., Chong, L. S., Kim, M. S. & Elowitz, M. B. Programmable protein circuits in living cells. *Science* **361**, 1251-1258 (2018).
23. Ng, A. H., Nguyen, T. H., Gomez-Schiavon, M., Dods, G., Langan, R. A., Boyken, S. E., Samson, J. A., Waldburger, L. M., Dueber, J. E., Baker, D *et al.* Modular and tunable biological feedback control using a de novo protein switch. *Nature*. **572**, 265-269 (2019).
24. Edwardson, T. G. W., Mori, T. & Hilvert, D. Rational engineering of a designed protein cage for siRNA deliver. *J. Am. Chem. Soc.* **140**, 10439-10442 (2018).
25. Morris, C. Glennie, S. J., Lam, H. S., Baum, H. E., Kandage, D., Williams, N. A., Morgan, D. J., Woolfson, D. N. & Davidson, A. D. A modular vaccine platform combining self-assembled protein cages and immunogenic peptides. *Adv. Funt. Mat.* **29**, 1-12 (2019).
26. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365-1368 (2015).
27. Quijano-Rubio, A., Ulge, U. Y., Walkey, C. D. & Silva, D. A. The advent of de novo proteins for cancer immunotherapy. *Curr. Opin. Chem. Biol.* **56**, 119-128 (2020).

28. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
29. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
30. Zhao, H. & Arnold, F. H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Engineering, Design and Selection* **12(1)**, 47-53 (1999).
31. Pan, X. & Kortemme, T. Recent advances in *de novo* protein design: Principles, methods, and applications. *JBC Reviews* **296**, 100558 (2021).
32. Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D, Ban, N. & Hilvert, D. Evolution of a designated retro-aldolase leads to complete active site remodeling. *Nat. Chem. Bio.* **9(8)**, 494-498 (2013).
33. MacBeath, G., Kast, P. & Hilvert, D. Redesigning enzyme topology by directed evolution. *Science* **279**, 1958-1961 (1998).
34. Glick, B. Metabolic load and heterologous gene expression. *Biotech. Adv.* **13(2)**, 247-261 (1995).
35. Ringler, P. & Schulz, G. E. Self-assembly of proteins into designated networks. *Science* **302**, 106-109 (2003).
36. McKay, C. S. & Finn, M. G. Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem. Biol.* **21**, 1075–1101 (2014).
37. Lee, T. C., Kang, M., Kim, C. H., Schultz, P. G., Chapman, E. & Deniz, A. A. Dual unnatural amino acid incorporation and click-chemistry labeling to enable single-molecule FRET studies of p97 folding. *ChemBioChem* **17(11)**, 981-984 (2016).
38. Spicer, C. & Davis, B. Selective chemical protein modification. *Nat. Comm.* **5**, 4740 (2014).
39. Modica, J. A., Lin, Y. & Mrksich, M. Synthesis of cyclic megamolecules. *J. Am. Chem. Soc.* **140(20)**, 6391-6399 (2018).
40. Chen, S., Su, L. Chen, J. & Wu, J. Cutinase: characteristics, preparation, and application. *Biotech. Adv.* **31(8)**, 1754-1767 (2013).
41. Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H. & Johnsson, K. Directed evolution of O<sup>6</sup>-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem. Biol.* **10(4)**, 313-317 (2003).

42. Kimmel, B. K. & Mrksich, M. Development of an enzyme-inhibitor reaction using cellular retinoic acid binding protein II for one-pot megamolecule synthesis. *Chem. Eur. J* **27(71)**, 17843-17848 (2021).
43. Modica, J. A., Skarpathiotis, S. & Mrksich, M. Modular assembly of protein building blocks to create precisely defined megamolecules. *ChemBioChem*. **13(16)**, 2331–2334 (2012).
44. Kimmel, B. R., Modica, J. A., Parker, K., Dravid, V. P. & Mrksich, M. Solid-Phase Synthesis of Megamolecules. *J. Am. Chem. Soc.* **142(10)**, 4534-4538 (2020).
45. Canne, L. E. *et al.* Chemical Protein Synthesis by Solid Phase Ligation of Unprotected Peptide Segments. *J. Am. Chem. Soc.* **121**, 8720–8727 (1999).
46. Modica, J.A., Iderzorig, T. & Mrksich, M. Design and Synthesis of Megamolecule Mimics of a Therapeutic Antibody. *J. Am. Chem. Soc.* **142**, 13657-13661 (2020).
47. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A. & Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209-249 (2021).
48. van der Horst, H. J., Nijhof, I. S., Mutis, T. & Chamuleau, M. E. D. Fc-Engineered Antibodies with Enhanced Fc-Effector Function for the Treatment of B-Cell Malignancies. *Cancers (Basel)*. **12(10)**, 3041 (2020).
49. Maadi, H., Soheilifar, M. H., Choi, W.-S., Moshtaghian, A. & Wang, Z. Trastuzumab mechanism of action; 20 years of research to unravel a dilemma. *Cancers*. **13**, 3540 (2021).
50. GlobalData, Drugs Database 2022.
51. Navarro, G. *et al.* Interactions between intracellular domains as key determinants of the quaternary structure and function of receptor heteromers. *J Biol Chem* **285**, 27346–27359 (2010).
52. Paakinaho, V., Johnson, T. A., Presman, D. M. & Hager, G. L. Glucocorticoid receptor quaternary structure drives chromatin occupancy and transcriptional outcome. *Genome Res* **29**, 1223–1234 (2019).
53. Han, B. *et al.* Conversion of mechanical force into biochemical signaling. *J Biol Chem* **279**, 54793–801 (2004).
54. Preller, M. & Manstein, D. J. Myosin structure, allostery, and mechano-chemistry. *Structure* **21**, 1911–22 (2013).
55. Stuber, J. C., Kast, F. & Pluckthun, A. High-throughput quantification of surface protein internalization and degradation. *ACS Chem. Bio* **14(6)**, 1154-1163 (2019).

56. Li, Y., Liu, P. C., Shen, Y., Snavely, M. D. & Hiraga, K. A cell-based internalization and degradation assay with an activatable fluorescence-quencher probe as a tool for functional antibody screening. *J. Biomol. Screen.* **20(7)**, 869-875 (2015).
57. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A. & Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209-249 (2021).
58. Iqbal, N.; Iqbal, N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Mol Biol Int.* **2014**, 852748 (2014).
59. Tai, W.; Mahato, R.; Cheng K. The role of HER2 in cancer therapy and targeted drug delivery. *J Control Release.* **146(3)**, 264-275 (2010).
60. Cho, H. S.; Mason, K.; Ramyar, K. X.; Stanley, A. M.; Gabelli, S. B.; Denney, D. W.; Leahy, D. J. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature.* **421**, 6924 (2003).
61. Franklin, M. C.; Carey, K. D.; Vajdos, F. F.; Leahy, D. J.; de Vos, A.M.; Sliwkowski, M. X. Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell.* **4**, 317-328 (2004).
62. Richard, S.; Selle, F.; Lotz, J. P.; Khalil, A.; Gligorov, J.; Soares, D. G. Pertuzumab and trastuzumab: the rationale way to synergy. *An Acad Bras Cienc.* **88(1)**, 565-577 (2016).
63. Garrett, J. T.; Arteaga, C. L. Resistance to HER2-directed antibodies and tyrosine kinase inhibitors: mechanisms and clinical implications. *Cancer Biol Ther.* **11(9)**, 793-800 (2011).
64. Hunter, F. W.; Barker, H. R.; Lipert, B.; Rothé, F.; Gebhart, G.; Piccart-Gebhart, M. J.; Sotiriou, C.; Jamieson, S. M. F. Mechanisms of resistance to trastuzumab emtansine (T-DM1) in HER2-positive breast cancer. *Br J Cancer.* **122**, 603-612 (2020).
65. Oh, D.Y., Bang, YJ. HER2-targeted therapies — a role beyond breast cancer. *Nat Rev Clin Oncol.* **17**, 33–48 (2020).
66. Yu, S.; Liu, Q.; Han, X.; Qin, S.; Zhao, W.; Li, A.; Wu, K. Development and clinical application of anti-HER2 monoclonal and bispecific antibodies for cancer treatment. *Exp Hematol Oncol.* **6**, 31 (2017)
67. Anders, C. K.; Johnson, R.; Litton, J.; Phillips, M.; Bleyer, A. Breast Cancer Before Age 40 Years. *Seminars in oncology.* **36(3)**, 237-249 (2009).
68. Modica, J.A., Iderzorig, T. & Mrksich, M. Design and Synthesis of Megamolecule Mimics of a Therapeutic Antibody. *J. Am. Chem. Soc.* **142**, 13657-13661 (2020).

69. Metcalf, K. J., \*Kimmel, B. R., **Sykora, D. J.**, Modica, J. A., Parker, K., Berens, E., Dai, R., Dravid, V., Werb, Z. & Mrksich, M. Synthetic tuning of domain stoichiometry in nanobody-enzyme megamolecules. *Bioconjugate Chemistry* **32**, 143-152 (2021).
70. Pruszynski, M.; Koumariou, E.; Vaidyanathan, G.; Revets, H.; Devoogdt, N.; Lahoutte, T.; Zalutsky, M. R. Targeting breast carcinoma with radioiodinated anti-HER2 nanobody. *Nucl. Med. Biol.* **40**, 52-59 (2013).
71. Stanley, A.; Ashrafi, G. H.; Seddon, A. M.; Modjtahedi, H. Synergistic effects of various HER inhibitors in combination with IGF-1R, C-MET, and Src targeting agents in breast cancer cell lines. *Sci. Rep.* **7(1)**, 3964 (2017).
72. Pohlmann, P., Mayer, I. A. & Mernaugh, R. Resistance to trastuzumab in breast cancer. *Clin. Cancer Res.* **15(24)**, 7479-7491 (2009).
73. Scheer, J. M.; Sandoval, W.; Elliott, J. M.; Shao, L.; Luis, E.; Lewin-Koh, S.-C.; Schaefer, G.; Vandlen, R. Reorienting the Fab domains of trastuzumab results in potent HER2 activators. *PLOS ONE.* **7(12)**, e51817 (2012).
74. Hu, Y. B.; Dammer, E. B.; Ren, R. J.; Wang, G. The endosomal-lysosomal system: from acidification and cargo sorting to neurodegeneration. *Transl. Neurodegener.* **4**, 18 (2015).
75. Nath, N.; Godat, B.; Zimprich, C.; Dwight, S. J.; Corona, C.; McDougall, M.; Urh, M. Homogeneous plate-based antibody internalization assay using pH sensor fluorescent dye. *J. Immun. Methods.* **431**, 11-21 (2016).
76. Soler, M. A.; Fortuna, S.; de Marco, A.; Laoi, A. Binding affinity prediction of nanobody-protein complexes by scoring of molecular dynamics trajectories. *Phys. Chem. Chem. Phys.* **20**, 3438-3444 (2018).
77. Maass, K. F.; Kulkarni, C.; Betts, A. M.; Wittrup, K. D. Determination of cellular processing rates for a trastuzumab-maytansinoid antibody-drug conjugate (ADC) highlights key parameters for ADC design. *AAPS J.* **18(3)**, 635-646 (2016).
78. Jost, C.; Schilling, J.; Tamaskovic, R.; Schwill, M.; Honegger, A.; Plückthun, A. Structural basis for eliciting a cytotoxic effect in HER2-overexpressing cancer cells via binding to the extracellular domain of HER2. *Structure.* **21**, 1979-1991 (2013).
79. Tamaskovic, R.; Schwill, M.; Nagy-Davidescu, G.; Jost, C.; Schaefer, D. C.; Verdurmen, W. P. R.; Schafer, J. V.; Honegger, A.; Plückthun, A. Intermolecular biparatopic trapping of ErbB2 prevents compensatory activation of PI3K/AKT via RAS-p110 crosstalk. *Nat. Commun.* **7**, 11672 (2016).
80. Oganessian, V.; Peng, L.; Bee, J. S.; Li, J.; Perry, S. R.; Comer, F.; Xu, L.; Cook, K.; Senthil, K.; Clarke, L.; Rosenthal, K.; Gao, C. Damschroder, M.; Wu, H.; Dall'Acqua, W. Structural

- insights into the mechanism of action of a biparatopic anti-HER2 antibody. *J. Biol. Chem.* **293**(22), 8439-8448 (2018).
81. Cheng, J.; Liang, M.; Carvalho, M. F.; Tigue, N.; Faggioni, R.; Roskos, L. K.; Vainshtein, I. Molecular mechanism of HER2 rapid internalization and redirected trafficking induced by anti-HER2 biparatopic antibody. *Antibodies*. **9**, 49 (2020).
  82. Kast, F.; Schwill, M.; Stüber, J. C.; Pfundstein, S.; Nagy-Davidescu, G.; Rodriguez, J. M. M.; Seehusen, F.; Richter, C. P.; Honegger, A.; Hartmann, K. P.; Weber, T. G.; Kroner, F.; Ernst, P.; Piehler, J. Plückthun, A. Engineering an anti-HER2 biparatopic antibody with a multimodal mechanism of action. *Nat. Commun.* **12**, 3790 (2021).
  83. Vaneycken, I.; Devoogdt, N.; Van Gassen, N.; Vincke, C.; Xavier, C.; Wernery, U.; Muyldermans, S.; Lahoutte, T.; Caveliers, V. Preclinical screening of anti-HER2 nanobodies for molecular imaging of breast cancer. *FASEB J.* **25**, 2433-2446 (2011).
  84. Xavier, C.; Vaneycken, I.; D'Huyvetter, M.; Heemskerck, J.; Keyaerts, M.; Vincke, C.; Devoogdt, N.; Muyldermans, S.; Lahoutte, T.; Caveliers, V. Synthesis, preclinical validation, dosimetry, and toxicity of <sup>68</sup>Ga-NOTA-anti-HER2 nanobodies for iPET imaging of HER2 receptor expression in cancer. *J. Nucl. Med.* **54**(5), 776-784 (2013).
  85. D'Huyvetter, M.; De Vos, J.; Xavier, C.; Pruszynski, M.; Sterckx, Y. G. J.; Massa, S.; Raes, G.; Caveliers, V.; Zalutsky, M. R.; Lahoutte, T.; Devoogdt, N. <sup>131</sup>I-labeled anti-HER2 camelid sdAb as a theranostic tool in cancer treatment. *Clin. Cancer Res.* **23**(21), 6616-6628 (2017).
  86. Ackerman, M. E.; Pawlowski, D.; Wittrup, K. D. Effect of antigen turnover rate and expression level on antibody penetration into tumor spheroids. *Mol. Cancer Ther.* **7**(7), 2233-2240 (2008).
  87. Ram, S.; Kim, D.; Ober, R. J.; Ward, E. S. The level of HER2 expression is a predictor of antibody-HER2 trafficking behavior in cancer cells. *MAbs*. **6**(5), 1211-1219 (2014).
  88. Diermeier, S.; Horváth, G.; Knuechel-Clarke, R.; Hofstaedter, F.; Szöllösi, J.; Brockhoff, G. Epidermal growth factor receptor coexpression modulates susceptibility to Herceptin in HER2/neu overexpressing breast cancer cells via specific erbB-receptor interaction and activation. *Exp. Cell Res.* **304**, 604-619 (2005)
  89. Yu, X.; Wang, L.; Shen, Y.; Wang, C.; Zhang, Y.; Meng, Y.; Yang, Y.; Liang, B.; Zhou, B.; Wang, H.; et al. Targeting EGFR/HER2 heterodimerization with a novel anti-HER2 domain II/III antibody. *Mol. Immunol.* **87**, 300-307 (2017)
  90. Junttila, T.T.; Akita, R.W.; Parsons, K.; Fields, C.; Phillips, G.D.L.; Friedman, L.S.; Sampath, D.; Sliwkowski, M.X. LigandIndependent HER2/HER3/PI3K Complex Is Disrupted by Trastuzumab and Is Effectively Inhibited by the PI3K Inhibitor GDC-0941. *Cancer Cell.* **15**, 429-440 (2009)

91. Gaborit, N.; Larbouret, C.; Vallaghe, J.; Peyrusson, F.; Bascoul-Mollevis, C.; Crapez, E.; Azria, D.; Chardès, T.; Poul, M.-A.; Mathis, G.; et al. Time-resolved Fluorescence Resonance Energy Transfer (TR-FRET) to Analyze the Disruption of EGFR/HER2 Dimers. *J. Biol. Chem.* **286**, 11337–11345 (2011).
92. Maadi, H.; Nami, B.; Tong, J.; Li, G.; Wang, Z. The effects of trastuzumab on HER2-mediated cell signaling in CHO cells expressing human HER2. *BMC Cancer* **18**, 1–14 (2018).
93. Zhao, J.; Mohan, N.; Nussinov, R.; Ma, B.; Wu, W.J. Trastuzumab Blocks the Receiver Function of HER2 Leading to the Population Shifts of HER2- Containing Homodimers and Heterodimers. *Antibodies* **10**, 7 (2021).
94. Dokmanovic, M.; Hirsch, D.S.; Shen, Y.; Wu, W.J. Rac1 contributes to trastuzumab resistance of breast cancer cells: Rac1 as a potential therapeutic target for the treatment of trastuzumab-resistant breast cancer. *Mol. Cancer Ther.* **8**, 1557–1569 (2009).
95. Gijzen, M.; King, P.; Perera, T.; Parker, P.; Harris, A.; Larijani, B.; Kong, A. HER2 Phosphorylation Is Maintained by a PKB Negative Feedback Loop in Response to Anti-HER2 Herceptin in Breast Cancer. *PLoS Biol.* **8**, e1000563 (2010).
96. Di Cara, G.; Marengo, G.; Albanese, N.N.; Marabeti, M.R.; Musso, R.; Cancemi, P.; Pucci-Minafra, I. Proteomic profiling of Trastuzumab (Herceptin(R))-sensitive and -resistant SKBR-3 breast cancer cells. *Anticancer. Res.* **33**, 489–503 (2013).
97. Milella, M.; Trisciuglio, D.; Bruno, T.; Ciuffreda, L.; Mottolese, M.; Cianciulli, A.; Cognetti, F.; Zangemeister-Wittke, U.; Del Bufalo, D.; Zupi, G. Trastuzumab Down-Regulates Bcl-2 Expression and Potentiates Apoptosis Induction by Bcl-2/Bcl-XL Bispecific Antisense Oligonucleotides in HER-2Gene–Amplified Breast Cancer Cells. *Clin. Cancer Res.* **10**, 7747–7756 (2004).
98. Scaltriti, M.; Verma, C.; Guzman, M.; Jimenez, J.; Parra, J.L.; Pedersen, K.; Smith, D.J.; Landolfi, S.; Cajal, S.R.Y.; Arribas, J.; et al. Lapatinib, a HER2 tyrosine kinase inhibitor, induces stabilization and accumulation of HER2 and potentiates trastuzumab-dependent cell cytotoxicity. *Oncogene.* **28**, 803–814 (2008).
99. Austin, C. D.; De Maziere, A. M.; Pisacane, P. I.; van Dijk, S. M.; Eigenbrot, C.; Sliwkowski, M. X.; Klumberman, J.; Scheller, R. H. *Mol. Bio. of the Cell.* **15(12)**, 5268-5282 (2004).
100. Bagnato, P.; Castagnino, A.; Cortese, K.; Bono, M.; Grasso, S.; Bellese, G.; Daniele, T.; Lundmark, R.; Defilippi, P.; Castagnola, P.; et al. Cooperative but distinct early co-signaling events originate from ERBB2 and ERBB1 receptors upon trastuzumab treatment in breast cancer cells. *Oncotarget.* **8**, 60109–60122 (2017).
101. Mohsin, S.K.; Weiss, H.L.; Gutierrez, M.C.; Chamness, G.C.; Schiff, R.; DiGiovanna, M.P.; Wang, C.-X.; Hilsenbeck, S.G.; Osborne, C.K.; Allred, D.C.; et al. Neoadjuvant

- Trastuzumab Induces Apoptosis in Primary Breast Cancers. *J. Clin. Oncol.* **23**, 2460–2468 (2005).
102. Maadi, H.; Soheilifar, M.H.; Choi, W.-S.; Moshtaghian, A.; Wang, Z. Trastuzumab Mechanism of Action; 20 Years of Research to Unravel a Dilemma. *Cancers.* **13**, 3540 (2021).
  103. Revets, H. & Hoogenboom, H. R. J. M., Ablynx NV. Method for obtaining polypeptide constructs comprising two or more single domain antibodies. EP2215125A1. October 27, 2007.
  104. Pant, S.; Ducreux, M.; Harding, J. J.; Javle, M. M.; Oh, D.-Y.; Wasan, H. S.; Fortenberry, A.; Josephson, N. C.; Mwatha, A.; Wang, K.; Fan, J. A phase IIb, open-label, single-arm study of zanidatamab (ZW25) monotherapy in subjects with advanced or metastatic HER2-amplified biliary tract cancers. *J. Clin. Oncology.* **39(3)**, TPS352 (2021).
  105. Rudnick, S. I.; Lou, J. Shaller, C. C.; Tang, Y.; Klein-Szanto, A. J. P.; Weiner, L. M.; Marks, J. D.; Adams, G. P. Influence of affinity and antigen internalization on the uptake and penetration of anti-HER2 antibodies in solid tumors. *Cancer Res.* **71(6)**, 2250-2259 (2011).
  106. Evans, R.; Thurber, G. M. Design of high avidity and low affinity antibodies for in situ control of antibody drug conjugate targeting. *Sci. Rep.* **12**, 7677 (2022)
  107. Taylor, E. L., Metcalf, K. J., Carlotti, B., Lai, C.-T., Modica, J. A., Schatz, G. C., Mrksich, M. & Goodson, T. Long-range energy transfer in protein megamolecules. *J. Am. Chem. Soc.* **140**, 15731-15743 (2018).
  108. Modica, J. A., Skarpathiotis, S. & Mrksich, M. Modular assembly of protein building blocks to create precisely defined megamolecules. *ChemBioChem.* **13(16)**, 2331–2334 (2012).
  109. Zhou, S. *et al.* Synthesis, characterization, and simulation of four-armed megamolecules. *Biomacromolecules* **22(6)**, 2363-2372 (2021).
  110. Navarro, G. *et al.* Interactions between intracellular domains as key determinants of the quaternary structure and function of receptor heteromers. *J Biol Chem* **285**, 27346–27359 (2010).
  111. Paakinaho, V., Johnson, T. A., Presman, D. M. & Hager, G. L. Glucocorticoid receptor quaternary structure drives chromatin occupancy and transcriptional outcome. *Genome Res* **29**, 1223–1234 (2019).
  112. Han, B. *et al.* Conversion of mechanical force into biochemical signaling. *J Biol Chem* **279**, 54793–801 (2004).
  113. Preller, M. & Manstein, D. J. Myosin structure, allostery, and mechano-chemistry. *Structure* **21**, 1911–22 (2013).

114. Nishi, H., Shaytan, A. & Panchenko, A. R. Physicochemical mechanisms of protein regulation by phosphorylation. *Front Genet* **5**, 270 (2014).
115. Segura-Peña, D., Lutz, S., Monnerjahn, C., Konrad, M. & Lavie, A. Binding of ATP to TK1-like enzymes is associated with a conformational change in the quaternary structure. *J Mol Biol* **369**, 129–41 (2007).
116. Abrusán, G. & Marsh, J. A. Ligand-Binding-Site Structure Shapes Allosteric Signal Transduction and the Evolution of Allostery in Protein Complexes. *Mol Biol Evol* **36**, 1711–1727 (2019).
117. Lee, R. *et al.* A critical element of the light-induced quaternary structural changes in YtvA-LOV. *Protein Sci* **24**, 1997–2007 (2015).
118. Bradshaw, J. M., Kubota, Y., Meyer, T. & Schulman, H. An ultrasensitive Ca<sup>2+</sup>/calmodulin-dependent protein kinase II-protein phosphatase 1 switch facilitates specificity in postsynaptic calcium signaling. *Proc Natl Acad Sci* **100**, 10512–7 (2003).
119. Hall, W. P. *et al.* A calcium-modulated plasmonic switch. *J Am Chem Soc* **130**, 5836–7 (2008).
120. Meister, G. E. & Joshi, N. S. An engineered calmodulin-based allosteric switch for peptide biosensing. *ChemBioChem* **14**, 1460–1467 (2013).
121. Murphy, W. L., Dillmore, W. S., Modica, J. & Mrksich, M. Dynamic hydrogels: translating a protein conformational change into macroscopic motion. *Angew Chem Int Ed Engl* **46**, 3066–9 (2007).
122. Liu, L. *et al.* Cyclic Stiffness Modulation of Cell-Laden Protein–Polymer Hydrogels in Response to User-Specified Stimuli Including Light. *Advanced Biosystems* **2**, 1800240 (2018).
123. Luo, J. & Sun, F. Calcium-responsive hydrogels enabled by inducible protein-protein interactions. *Polymer Chemistry* **11**, 4973–4977 (2020).
124. Kellmann, S.-J., Dübel, S. & Thie, H. A strategy to identify linker-based modules for the allosteric regulation of antibody-antigen binding affinities of different scFvs. *MAbs* **9**, 404–418 (2017).
125. Pudasaini, A., El-Arab, K. K. & Zoltowski, B. D. LOV-based optogenetic devices: Light-driven modules to impart photoregulated control of cellular signaling. *Frontiers in Molecular Biosciences* **2** (2015).
126. Hepp, S. *et al.* An Optogenetic Tool for Induced Protein Stabilization Based on the *Phaeodactylum tricornutum* Aureochrome 1a Light-Oxygen-Voltage Domain. *J Mol Biol* **432**, 1880–1900 (2020).

127. Lalwani, M. A. *et al.* Optogenetic control of the lac operon for bacterial chemical and protein production. *Nature Chemical Biology* **17**, 71–79 (2021).
128. He, L. *et al.* Circularly permuted LOV2 as a modular photoswitch for optogenetic engineering. *Nat Chem Biol* **17**, 915–923 (2021).
129. Zajc, C. U. *et al.* A conformation-specific ON-switch for controlling CAR T cells with an orally available drug. *Proc. Nat. Acad. Sci.* **117**(26) 14926–14935 (2020)
130. Wu, C.-Y., Roybal, K. T., Puchner, E. M., Onuffer, J. & Lim, W. A. Remote control of therapeutic T cells through a small molecule-gated chimeric receptor. *Science* **350**, aab4077 (2015).
131. Hill, Z. B., Martinko, A. J., Nguyen, D. P. & Wells, J. A. Human antibody-based chemically induced dimerizers for cell therapeutic applications. *Nature Chemical Biology* **14**, 112–117 (2018).
132. Shui, S. *et al.* A rational blueprint for the design of chemically-controlled protein switches. *Nature Communications* **12**, (2021).
133. Foight, G. W. *et al.* Multi-input chemical control of protein dimerization for programming graded cellular responses. *Nature Biotechnology* **37**, 1209–1216 (2019).
134. Glasgow, A. A. *et al.* Computational design of a modular protein sense-response system. *Science* **366**, 1024–1028 (2019).
135. Quijano-Rubio, A. *et al.* De novo design of modular and tunable protein biosensors. *Nature* **591**, 482–487 (2021).
136. Fisher, S. A., Baker, A. E. G. & Shoichet, M. S. Designing Peptide and Protein Modified Hydrogels: Selecting the Optimal Conjugation Strategy. *Journal of the American Chemical Society*, **139** 7416–7427 (2017).
137. Aupič, J., Lapenta, F. & Jerala, R. SwitCCh: Metal-Site Design for Controlling the Assembly of a Coiled-Coil Homodimer. *ChemBiochem* **19**, 2453–2457 (2018).
138. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).
139. Rittle, J., Field, M. J., Green, M. T. & Tezcan, F. A. An efficient, step-economical strategy for the design of functional metalloproteins. *Nat Chem* **11**, 434–441 (2019).
140. Churchfield, L. A., Medina-Morales, A., Brodin, J. D., Perez, A. & Tezcan, F. A. De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J Am Chem Soc* **138**, 13163–13166 (2016).

141. van Deventer, J. A., Yuet, K. P., Yoo, T. H. & Tirrell, D. A. Cell surface display yields evolvable, clickable antibody fragments. *Chembiochem* **15**, 1777–81 (2014).
142. Zubi, Y. S. *et al.* Metal-responsive regulation of enzyme catalysis using genetically encoded chemical switches. *Nature Communications* **13**, (2022).
143. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nature Communications* **6**, (2015).
144. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53-60 (2017).
145. Saito, H. & Suga, H. A ribozyme exclusively aminoacylates the 3'-hydroxyl group of the tRNA terminal adenosine. *J Am Chem Soc* **123**, 7178–9 (2001).
146. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol* **7**, 28–33 (2000).
147. Ramaswamy, K., Saito, H., Murakami, H., Shiba, K. & Suga, H. Designer ribozymes: programming the tRNA specificity into flexizyme. *J Am Chem Soc* **126**, 11454–5 (2004).
148. Murakami, H., Ohta, A., Ashigai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat Methods* **3**, 357–9 (2006).
149. Schmied, W. H., Elsässer, S. J., Uttamapinant, C. & Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J Am Chem Soc* **136**, 15577–15583 (2014).
150. Jewett, M. C., Fritz, B. R., Timmerman, L. E. & Church, G. M. In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol Syst Biol* **9**, 678 (2013).
151. Liu, Y., Davis, R. G., Thomas, P. M., Kelleher, N. L. & Jewett, M. C. In vitro-Constructed Ribosomes Enable Multi-site Incorporation of Noncanonical Amino Acids into Proteins. *Biochemistry* **60(3)**, 161-169 (2021).
152. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat Commun* **5**, 4740 (2014).
153. Shadish, J. A. & DeForest, C. A. Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter*, **2(1)**, 50–77 (2020).
154. van Gemeren, I. A., Beijersbergen, A., van den Hondel, C. A. & Verrips, C. T. Expression and secretion of defined cutinase variants by *Aspergillus awamori*. *Appl Environ Microbiol* **64**, 2794–9 (1998).

155. Juillerat, A. *et al.* Directed evolution of O<sup>6</sup>-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem Biol* **10**, 313–7 (2003).
156. Mollwitz, B. *et al.* Directed evolution of the suicide protein O<sup>6</sup>-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **51**, 986–94 (2012).
157. Holyer, R. H., Hubbard, C. D., Kettle, S. F. A. & Wilkins, R. G. The Kinetics of Replacement Reactions of Complexes of the Transition Metals with 2,2',2'' - Terpyridine. *Inorganic Chemistry* **5**, 622–625 (1966).
158. Mastop, M. *et al.* Characterization of a spectrally diverse set of fluorescent proteins as FRET acceptors for mTurquoise2. *Scientific Reports* **7**, (2017).
159. Ferreira, C. M. H., Pinto, I. S. S., Soares, E. v. & Soares, H. M. V. M. (Un)suitability of the use of pH buffers in biological, biochemical and environmental studies and their interaction with metal ions-a review. *RSC Advances* **5(39)**, 30989–31003 (2015).
160. Shunmugam, R., Gabriel, G. J., Aamer, K. A. & Tew, G. N. Metal-ligand-containing polymers: terpyridine as the supramolecular unit. *Macromol Rapid Commun* **31**, 784–93 (2010).
161. Nobuyuki, T. & Yasuhiko, S. The Kinetics of the Complex-forming Reaction between Nickel(II) and Ethylenediaminetetraacetate. *Bulletin of the Chemical Society of Japan*. **32(6)**, 578-584 (1959).
162. Kimura, M. The Effect of Acetate Ion on the Rate of the Reaction between Cobalt(II) and Manganese(II)-Ethylenediaminetetraacetate Complexes. *Bulletin of the Chemical Society of Japan*. **42**, 404-406 (1969).
163. Martell, A. E., Smith, R. M., & Motekaitis, R. J. NIST Critically Selected Stability Constants of Metal Complexes. *NIST Standard Reference Database 46, Gaithersburg, MD* (2001).
164. Bi, X. & Pang, Y. Optical Response of Terpyridine Ligands to Zinc Binding: A Close Look at the Substitution Effect by Spectroscopic Studies at Low Temperature. *Journal of Physical Chemistry B* **120**, 3311–3317 (2016).
165. GBD 2015 Mortality and Causes of Death Collaborators "Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980-2015: A Systematic Analysis for the Global Burden of Disease Study. *Lancet*. **388(10053)**, 1459-1544 (2015).
166. Siegel, R. L., Miller, K.D. & Jemal, A. Cancer Statistics, 2019. *CA Cancer J Clin*, **69(1)**, 7-34 (2019).

167. Aeffner, F. *et al.* Introduction to Digital Image Analysis in Whole-Slide Imaging: A White Paper from the Digital Pathology Association. *J. Pathol. Inform.* **10**, 9 (2019).
168. Binnewies, M. *et al.* Understanding the Tumor Immune Microenvironment (Time) for Effective Therapy. *Nat. Med.* **24(5)**, 541-550 (2018).
169. Komaki, K., Sano, N. & Tangoku, A. Problems in Histological Grading of Malignancy and Its Clinical Significance in Patients with Operable Breast Cancer. *Breast Cancer* **13(3)**, 249-253 (2006).
170. Janowczyk, A. & Madabhushi, A. Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases. *J Pathol Inform* **7**, 29 (2016).
171. Muldoon, T. J. *et al.* Evaluation of Quantitative Image Analysis Criteria for the High-Resolution Microendoscopic Detection of Neoplasia in Barrett's Esophagus. *Journal of Biomedical Optics.* **15(2)**, 026027 (2010).
172. Odze, R.D., Marcial, M. A. & Antonioli, D. Gastric fundic gland polyps: a morphological study including mucin histochemistry, stereometry, and MIB-1 immunohistochemistry. *Hum Pathol.* **27(8)**, 896-903 (1996).
173. Tang, T. T., Zawaski, J.A., Francis, K.N., Qutub, A.A. & Gaber, M.W. Image-Based Classification of Tumor Type and Growth Rate Using Machine Learning: A Preclinical Study. *Scientific Reports* **9** (2019).
174. Tripathy, R. K., Mahanta, S. & Paul, S. Artificial Intelligence-Based Classification of Breast Cancer Using Cellular Images. *Rsc Advances.* **4(18)**, 9349-9355 (2014).
175. Guck, J. *et al.* Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence. *Biophysical Journal.* **88(5)**, 3689-3698 (2005).
176. Shankar, J. & Nabi, I.R. Actin Cytoskeleton Regulation of Epithelial Mesenchymal Transition in Metastatic Cancer Cells. *PLoS ONE.* **10(3)**, e0119954 (2015).
177. Izdebska, M., Zielinska, W., Grzanka, D. & Gagat., M. The Role of Actin Dynamics and Actin-Binding Proteins Expression in Epithelial-to-Mesenchymal Transition and Its Association with Cancer Progression and Evaluation of Possible Therapeutic Targets. *Biomed Res Int.* **4579373** (2018).
178. Haynes, J., Srivastava, J., Madson, N., Wittmann, T. & Barber, D.L. Dynamic actin remodeling during epithelial-mesenchymal transition depends on increased moesin expression. *Mol Biol Cell.* **22(24)** , 4750-4764 (2011).
179. Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science.* **339**, 580–584 (2013).

180. Gao, D. *et al.* Metastatic tumor cells - genotypes and phenotypes. *Front Biol (Beijing)*. **13(4)**, 277-286 (2018).
181. Gaspar, P., Holder, M.V., Aerne, B.L., Janody, F. & Tapon, N. Zyxin antagonizes the FERM protein expanded to couple F-actin and Yorkie-dependent organ growth. *Curr Biol*. **25(6)**, 679-689 (2015).
182. Haupt, A. & Minc N. How cells sense their own shape - mechanisms to probe cell geometry and their implications in cellular organization and function. *Journal of Cell Science*. **131** (2018).
183. Kilian, K.A., Bugarija, B., Lahn, B.T. & Mrksich M. Geometric cues for directing the differentiation of mesenchymal stem cells. *Proc. Natl. Acad. Sci*. **107(11)**, 4872-4877 (2010).
184. Kumar, A., Biebuyck, H.A., Whitesides, G.M. Patterning Self-Assembled Monolayers: Applications in Materials Science. *Langmuir*. **10**, 1498-1511 (1994).
185. Chen, C.S., Mrksich, M., Huang, S., Whitesides, G.M. & Ingber, D.E. Geometric Control of Cell Life and Death. *Science*. **276**, 1345-1347 (1997).
186. Mrksich, M., Dike, L.E., Tien, J.Y., Ingber, D.E. & Whitesides, G.M. Using Microcontact Printing to Pattern the Attachment of Mammalian Cells to Self-Assembled Monolayers of Alkanethiolates on Transparent Films of Gold and Silver. *Exp. Cell Res*. **235**, 305-313 (1997).
187. Chen, C.S., Mrksich, M., Huang, S., Whitesides, G.M. & Ingber, D.E. Micropatterned Surfaces for Control of Cell Shape, Position, and Function. *Biotech. Prog*. **14**, 356-363 (1998).
188. Hu, S., *et al.* Reciprocal regulation of actomyosin organization and contractility in nonmuscle cells by tropomyosins and alpha-actinins. *Mol Biol Cell*. **30(16)**, 2025-2036 (2019).
189. Travis, W.D., Brambilla, B., Burke, A.P, Marx, A. & Nicholson, A.G. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart*. Lyon: International Agency for Research on Cancer. ISBN 978-92-832-2436-5. (2015).
190. El-Naggar, A.K., Chan, J.C.K., Grandis, J.R., Takata, T. & Slootweg, P.J. *WHO Classification of Head and Neck Tumours*. Lyon: International Agency for Research on Cancer. ISBN 978-92-832-2438-9. (2000).
191. Doyle, A.D., **Sykora, D. J.**, Pacheco, G.G., Kutys, M.L. & Yamada K.M. 3D mesenchymal cell migration is driven by anterior cellular contraction that generates an extracellular matrix prestrain. *Developmental Cell*. **56**, 1-16 (2021).
192. Nakamura, K. *et al.* Characterization of mouse melanoma cell lines by their mortal malignancy using an experimental metastatic model. *Life Sci*. **70(7)**, 791-798 (2002)

193. Mathis, R.A., Sokol, E.S. & Gupta, P.B. Cancer cells exhibit clonal diversity in phenotypic plasticity. *Open Biol.* **7**, 160283 (2017).
194. Burbulla, L.F., Beaumont, K.G., Mrksich, M. & Krainc, D. Micropatterning Facilitates the Long-Term Growth and Analysis of iPSC-Derived Individual Human Neurons and Neuronal Networks. *Adv. Healthc. Mater.* **5(15)**, 1894-1903 (2016).
195. Oei, R. W. *et al.* Convolutional Neural Network for Cell Classification Using Microscope Images of Intracellular Actin Networks. *PLoS ONE* **14(3)**, e0213626 (2019).
196. Schatten, H. *The Cytoskeleton in Health and Disease*. ISBN: 978-1-4939-2903-0 (2015).
197. Hardeman, E.C. & Gunning P.W. Life and death agendas of actin filaments. *Nature Materials.* **19**, 135-136 (2020).
198. Nersesian, S. *et al.* Effects of Modulating Actin Dynamics on HER2 Cancer Cell Motility and Metastasis. *Sci Rep.* **8(1)**, 17243 (2018).
199. Bonello, T.T., Stehn, J.R. & Gunning, P.W. New approaches to targeting the actin cytoskeleton for chemotherapy. *Future Med. Chem.* **1(7)**, 1311-1331 (2009).
200. Huang, B. *et al.* Gelsolin-mediated activation of PI3K/Akt pathway is crucial for hepatocyte growth factor-induced cell scattering in gastric carcinoma. *Oncotarget.* **7**, 25391 (2016).
201. Tochwawng, L. *et al.* Gelsolin-Cu/ZnSOD interaction alters intracellular reactive oxygen species levels to promote cancer cell invasion. *Oncotarget.* **7**, 52832 (2016).
202. Zhuo, J., *et al.* Gelsolin induces colorectal tumor cell invasion via modulation of the urokinase-type plasminogen activator cascade. *PLoS ONE.* **7**, e43594 (2012).
203. Jackson, H.W., *et al.* The single-cell pathology landscape of breast cancer. *Nature.* **578**, 615-620 (2020).
204. Jevtic, P., Edens, L.J., Vukovic, L.D. & Levy, D.L. Sizing and shaping the nucleus: mechanisms and significance. *Curr. Opin. Cell Biol.* **28**, 16-27 (2014).
205. Harkness T., *et al.* High-content imaging with micropatterned multiwell plates reveals influence of cell geometry and cytoskeleton on chromatin dynamics. *Biotechnol. J.* **10(10)**, 1555-67 (2015).
206. Rajer, M. & Kmet, M. Quantitative analysis of fine needle aspiration biopsy samples. *Radiology and Oncology.* **39(4)** (2005).

207. Cabezas, M.D., Meckes, B., Mirkin, C.A. & Mrksich, M. Subcellular Control over Focal Adhesion Anisotropy, Independent of Cell Morphology, Dictates Stem Cell Fate. *ACS Nano*. **13(10)**, 11144-11152 (2019).
208. Higaki, T. Quantitative evaluation of cytoskeletal organizations by microscopic image analysis." *Plant Morphology*. **29.1**, 15-21 (2017).
209. Ueda, H. Myosin-dependent endoplasmic reticulum motility and F-actin organization in plant cells. *Proc. Natl. Acad. Sci.* **107(15)**, 6894-6899 (2010).
210. Kroon, D.-J. 2D Line Curvature and Normals. MATLAB Central File Exchange. <http://www.mathworks.fr/matlabcentral/fileexchange/32696-2d-line-curvature-and-normals>. (2011).
211. Roffo, G. Feature Selection Library (MATLAB Toolbox). *arXiv*. 1607.01327 (2016).
212. Bishop, C.M. Neural Networks for Pattern Recognition. *Oxford University Press, Inc.* ISBN: 0198538642 (1995).
213. Clément, H. Investigating Cellular Nanoscale with X-rays: From Proteins to Networks. *PhD Dissertation*. (2016).
214. Leggett, S.E., Hruska, A.M., Guo, M. *et al.* The epithelial-mesenchymal transition and the cytoskeleton in bioengineered systems. *Cell Commun. Signal.* **19**, 32 (2021).
215. Manning, G., Whyte, D. B., Martinez, T., Hunter, S. & Sundarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).
216. Manning, G., Plowman, G. D., Hunter, T. & Sundarsanam, S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27(10)**, 514-520 (2002).
217. Ubersax, J. A. & Ferrell Jr., J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8(7)**, 530-541 (2007).
218. Tan, C. S. H., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M. O., Jorgensen, C., Bader, G. D., Aebersold, R., Pawson, T. & Linding, R. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci. Signaling* **2(81)**, ra39 (2009).
219. Olsen, J. V., Blagoev, B., Gnäd, F., Macek, B., Kumar, C., Mortensen, P. & Mann, M. Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127(3)**, 635-648 (2006).
220. Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F. & White, F. M. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20(3)**, 301-305 (2002).

221. Shi, Y. Serine/threonine phosphatases: mechanism through structure. **139(3)**, 468-484 (2009).
222. Virshup, D. M. & Shenolikar, S. From promiscuity to precision: protein phosphatases get a makeover. *Mol. Cell.* **33(5)**, 537-545 (2009).
223. Bose, A. K. & Janes, K. A. A high-throughput assay for phosphoprotein-specific phosphatase activity in cellular extracts. *MCP.* **12(3)**, 797-806 (2012).
224. Killilea, S. D., Cheng, Q. & Wang, Z.-X. in *Protein Phosphatase Protocols, Vol. 93* (Ed.: J. Ludlow), Humana Press. 22-33 (1998).
225. McAvoy, T. & Nairn, A. C. Serine/threonine protein phosphatase assays. *Curr. Prot. In Mol. Bio.* **92**, 18.18.11-18.18.11 (2010).
226. Espanel, X., Huguenin-Reggiani, M. & van Huijsduijnen, R. H. SPOT technique: A novel tool to study protein tyrosine phosphatases. *Protein Sci.* **11**, 2326-2334 (2002).
227. Knight, Z. A. & Shokat, K. M. Features of selective kinase inhibitors. *Chem. Biol.* **12(6)**, 621-637 (2005).
228. Beck, J. R., Lawrence, A., Tung, A. S., Harris, E. N. & Stains, C. I. Interrogating endogenous protein phosphatase activity with rationally designed chemosensors. *ACS Chem. Biol.* **11**, 284-290 (2016).
229. Disney, M. D. & Seeberger, P. H. The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens. *Chem. Biol.* **11(12)**, 1701-1707 (2004).
230. Stadler, V., Felgenhauer, T., Beyer, M., Fernandez, S., Leibe, K., Guttler, S., Groning, M., Konig, K., Torralba, G., Hausmann, M., Lindsenstruth, V., Nesterov, A., Block, I., Pipkorn, R., Poustka, A., Bischoff, F. R. & Breitling, F. Combinatorial synthesis of peptide arrays with a laser printer. *Angew. Chem. Int. Ed.* **47**, 7132-7135 (2008).
231. Weber L. K., Isse A., Rentschler S., Kneusel R. E., Palermo A., Hubbuch J., Nesterov-Mueller A., Breitling F., and Loeffler F. F. Antibody fingerprints in lyme disease deciphered with high density peptide arrays. *Eng. Life. Sci.* **17**, 1078–1087 (2017).
232. Szymczak, L. C., Kuo, H.-Y. & Mrksich, M. Peptide arrays: development and application. *Anal. Chem.* **90(1)**, 266-282 (2018).
233. Meng, X., Wei, J. Wang, Y., Zhang, H. & Wang, Z. The role of peptide microarrays in biomedical research. *Anal. Methods.* **38(10)**, 4614-4624 (2018).
234. Gurard-Levin, Z. A. & Mrksich M. Combining self-assembled monolayers and mass spectrometry for applications in biochips. *Annu. Rev. Anal. Chem.* **1**, 767-800 (2008).

235. Mrksich, M. Mass spectrometry of self-assembled monolayers: A new tool for molecular surface science. *ACS Nano*. **2(1)**, 7-18 (2008).
236. Berns, E. J., Cabezas, M. D. & Mrksich, M. Cellular assays with a molecular endpoint measured by SAMDI mass spectrometry. *Small*. **12(28)**, 3811-3818 (2016).
237. Kuo, H.-Y., DeLuca, T. A., Miller, W. M. & Mrksich, M. Profiling deacetylase activities in cell lysates with peptide arrays and SAMDI mass spectrometry. *Anal. Chem.* **85(22)**, 10635-10642 (2013).
238. Houseman, B. T., Gawalt, E. S. & Mrksich, M. Maleimide-functionalized self-assembled monolayers for the preparation of peptide and carbohydrate biochips. *Langmuir* **19(5)**, 1522-1531 (2003).
239. Szymczak, L. C., Huang, C. F., Berns, E. J. & Mrksich, M. Combining SAMDI mass spectrometry and peptide arrays to profile phosphatase activities. *Methods Enzymol.*, Vol. 607 (Ed.: K. N. Allen), Academic Press. 389-403 (2018).
240. Ren, L., Chen, X., Luechapanichkul, R., Selner, N. G., Meyer, T. M., Wavreille, A.-S., Chan, R., Iorio, C., Zhou, X., Neel, B. G. & Pei, D. Substrate specificity of protein tyrosine phosphatases 1B, RPTPa, SHP-1, and SHP-2. *Biochemistry*. **50**, 2339-2356 (2011).
241. Chen, C., Ha, B. H., Thevenin, A. F., Lou, H. J., Zhang, R., Yip, K. Y., Peterson, J. R., Gerstein, M., Kim, P. M., Filippakopoulos, P., Knapp, S., Boggon, T. J. & Turk, B. E. Identification of a major determinant for serine-threonine kinase phosphoacceptor specificity. *Mol. Cell*. **53(1)**, 140-147 (2014).
242. Hunter, T. Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.* **21(2)**, 140-146 (2009).
243. Sharma, K., D'Souza, R. C., Tyanova, S., Schaab, C., Wisniewski, J. R., Cox, J. & Mann, M. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8(5)**, 1583-1594 (2014).
244. Agostinis, P., Goris, J., Pinna, L. A., Marchiori, F., Perich, J. W., Meyer, H. E. & Merlevede, W. Synthetic peptides as model substrates for the study of the specificity of the polycation-stimulated protein phosphatase. *Eur. J. Biochem.* **189**, 235-241 (1990).
245. McCloy R. A., Parker B. L., Rogers S., Chaudhuri R., Gayevskiy V., Hoffman N. J., Ali N., Watkins D. N., Daly R.J., James D. E., Lorca, T., Castro, A. & Burgess, A. Global phosphoproteomic mapping of early mitotic exit in human cells identifies novel substrate dephosphorylation motifs. *Mol. Cell. Proteomics*. **14**, 2194-2212 (2015).
246. Lienhard, G. E. Non-functional phosphorylations? *Trends Biochem. Sci.* **33(8)**, 351-352 (2008).

247. Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R. Beausoleil, S. A., Villen, J. Haas, W., Sowa, M. E., Gygi, S. P. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*. **143(7)**, 1174-1189 (2010).
248. Bennetzen, M. V., Larsen, D. H., Bunkenborg, J. Barket, J., Lukas, J. & Andersen, J. S. Site-specific phosphorylation dynamics of the nuclear proteome during the DNA damage response. *Mol. Cell. Proteomics*. **9**, 1314 (2010).
249. Li, X., Wilmanns, M. Thornton, J. & Kohn, M. Elucidating human phosphatase-substrate networks. *Sci. Signal*. **6(275)**, rs10 (2013).
250. Patel, P. R., Kiser, R. C., Lu, Y. Y., Fong, E., Ho, W. C., Tirrell, D. A. & Grubbs, R. H. Synthesis and cell culture adhesive properties of linear and cyclic RGD functionalized polynorborene thin films. *Biomacromolecules*. **13(8)**, 2546-2553 (2012).
251. Bugga, P. & Mrksich, M. Sequential photoactivation of self-assembled monolayers to direct cell adhesion and migration. *Langmuir*. **35(17)**, 5937-5943 (2019).
252. Berginski, M. E. & Gomez, S, M. The focal adhesion analysis server: a web tool for analyzing focal adhesion dynamics. *F1000Research* **2**, 68 (2013).
253. Hayes, J.M., Cosgrave, E. F., Struwe, W. B., Wormald, M., Davey, G. P., Jefferis, R. & Rudd P. M. Glycosylation and Fc receptors. *Curr Top Microbiol Immunol*. **382**, 165-199 (2014).
254. Iberg, N. & Flückiger, R. Nonenzymatic glycosylation of albumin in vivo. Identification of multiple glycosylated sites. *J Biol Chem*. **261(29)**, 13542-13545 (1986).
255. Pluchinsky, A. J., Wackelin, D. J, Huang, X., Arnold, F. H. & Mrksich, M. High throughput screening with SAMDI mass spectrometry for directed evolution. *J. Am. Chem. Soc.* **142(47)**, 19804-19808 (2020).