Smarter Molecular Probes for Disease Detection and Treatment

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE
SCHOOL OF THE UNIVERSITY OF MINNESOTA

BY

Sadie Maria Johnson

IN PARTIAL FUFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Benjamin J. Hackel

December 2018

Acknowledgements

There will never be enough time or space to thank all the people that have helped me on my journey to this point, but I will do my best here. I first must thank my family, my parents Neil and Kathy Johnson and my brother Nathan. Dad, you have and will always be the most important teacher in my life, and there is never a day that I don't learn from you. Mom, you have always shown what true love and support means, even when I didn't exactly deserve it. You both have taught me so much about the importance of hard work, family, and living an honest life. Nathan, it has been a blessing to be your sister and watch you become such a smart, hardworking, funny, and kind person. You touch so many people's lives every day, and they are all the better for it. But most importantly, you all have given me a home, and for all this I am eternally grateful.

I have also had the great privilege of working under Professor Ben Hackel. Ben, you embody the lead by example principle, and it does not go unnoticed. You set the standard of how to be a good researcher, teacher, mentor, and team member. Your enthusiasm for the science is infectious, and there was rarely a meeting that I did not leave more energized for the work. Perhaps most important was the importance you placed on personal and professional development of your students. I could not have completed my Ph.D. without your mentorship and support. I will always strive to meet your standards both inside and out of the lab.

In addition to Dr. Hackel, I have greatly enjoyed working with my lab members, both past and present. Danny, Brett, Larry, and Max were an amazing crew to begin my

research with, and the comradery made the late nights and stressful sequence dashes worth every minute. To rest of the past and current members, I have loved working with you all and I cannot wait to see what the future holds for you all and the growing Hackel lab. Keep up the good work!

And lastly, I must thank my extensive second family in CEMS. Larry, Brittany, and Seth have supported (and sometimes, dragged) me through the extremely rocky portions that accompany any Ph.D. program. My squad of Ralm, Scott, Koustav, Dayne, and Frankie has been such a solid group of friends to enjoy the brief respites away from Amundson Hall with. These and countless other CEMS members have made my time here some of the best a person could ask for. Thank you all.

Abstract

Molecular targeting is a powerful diagnostic and therapeutic tool for clinicians to combat disease. A molecular ligand with specific binding for a disease biomarker can deliver a detectable or active moiety. A vast majority of molecular targeting agents rely on a single ligand:target interaction to be both specific to the target of interest, specific to the diseased tissue, and have the desired potency or functional output. The goal of this thesis is to study and develop novel targeting mechanisms that can utilize multi-domain molecules to more efficiently detect and treat diseases based on their molecular profiles.

We first addressed the issue of biomarker uniqueness, where biomarkers upregulated in cancer cells are often still present in healthy cells and cause myriad side effects or diagnostic false positives. We studied heterobivalent fusions containing two domains of moderate affinity to separate biomarkers as a biological AND gate probe, whose localization requires overexpression of both biomarkers for localization. We demonstrated that within a model system that the enhanced selectivity can reduce high off-tumor tissue signal. We showed that yeast surface display could be used to engineer these moderate affinity ligands, but also found that many ligands used lacked modularity when expressed in the heterobivalent construct. We then showed *in vitro* that affinity modulations can confer improved specificity to EGFR^{high}/CEA^{high} cells over cells expressing only one of the two biomarkers.

We also applied this heterobivalent ligand concept to two distinct epitopes on a single target, via a protein — small molecule fusion, to enhance selectivity for

homologous proteins. We showed that for two nonspecific inhibitors of carbonic anhydrase and the GPCR CXCR4, we could engineer a protein—small molecule fusion that would preferentially bind to a single isoform within the target protein class. We were able to evolve these selective fusions in a robust and efficient manner using a simple cysteine—maleimide conjugation strategy in the context of yeast surface display.

To improve on non-invasive disease imaging technology, we partnered with Dr. Ekaterina Morgounova-Ippolito and Prof. Shai Ashkenazi to develop an activatable imaging probe using a novel imaging modality (photoacoustic lifetime imaging) and an enzyme-labile probe to create a probe that emits only in the presence of an enzymatic biomarker. We found that the contrast agent methylene blue could be efficiently dimerized by a hairpin peptide structure consisting a poly-glutamate and poly-arginine zipper sequence linked by a protease-cleavable site and activated by the MMP-2 enzyme.

Together, this work advances several new concepts for improving molecular targeting. These molecular designs can improve selectivity and sensitivity of therapeutics and diagnostics, which is crucial for the next generation of cancer and disease treatment.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1: Molecular Medicine

Molecular targeting has been used to improve diagnostic and therapeutic drug delivery to diseased tissues. Targeted therapies have contributed to the decrease in cancer deaths and increase in 5-year survival rate[1–3]. In addition to the treatment of cancer, molecular targeted therapies are also successfully used in the treatment of a host of autoimmune, cardiovascular, and infectious diseases[4]. One of the earliest examples of using molecular targeting as a therapy was the use of Herceptin (trastuzumab) to treat patients with HER2+ breast cancers. Researchers had identified the HER2/*neu* gene amplification and protein overexpression as a strong indicator of poor prognosis[5]. To utilize this knowledge, Genentech developed trastuzumab to bind selectively to the extracellular domain of the HER2 protein. Once introduced to the clinic, it has been instrumental in the treatment of aggressive breast cancers, both as a single agent therapeutic or as an adjuvant to non-selective chemotherapies[6,7]. Trastuzumab has also been explored as a targeting moiety for delivery of contrast agents in the non-invasive detection and characterization of breast cancer[8–11].

In its development and implementation, trastuzumab has been a shining example of the potential that precision medicine brings to the clinic, and newly engineered treatments and diagnostic tools have built off its success. In addition to HER2, estrogen receptor (ER) has been important molecular targets for breast cancer therapies[12–14]. Targeting of vascular endothelial growth factor (VEGF) has been successful in treating a

variety of diseases. VEGF binding antibodies bevacizumab (Avastin, Genentech) and ranibizumab (Lucentis, Genentech) that inhibit the binding of VEGF to its target receptors have been used in treatment of several severe cancer indications, such as glioblastoma and renal cell carcinomas[15–17], and in treatment of ophthalmological indications such as diabetic macular edema and proliferative diabetic retinopathy[18–20]. Treatment of autoimmune disorders has also vastly improved with the introduction of targeted molecular therapies. The tumor necrosis factor alpha (TNF-α) inhibitors adalimumab (Humira, Abbott Laboratories) and etanercept (Enbrel, Immunex/Amgen) are approved to treat rheumatoid arthritis[21], ulcerative colitis, Crohn's disease[22–24], and plaque psoriasis[25]. In addition to these examples, metadata analysis of clinical trial success rates from 2010-2015 have shown that drugs with a known biomarker have a higher chance of success in all disease indications reviewed[26]. Molecular targeting diagnostics have also shown great promise, improving the ability to make better treatment decisions via patient stratification[27–29], and have demonstrated the ability to improve the limits of disease detection over traditional diagnostic methods[30–33].

1.2: Engineering Molecular Targeting Agents

The shared property across all molecular targeting agents is that they must have an engineered surface (paratope) that can strongly and selectively interact with some portion of the target molecule (epitope) (Figure 1.1). The paratope can consist of optimized small molecule moieties that can insert into small pockets and grooves on the target

2

molecule, or a subset of residues on a protein's surface that can interact strongly with epitopes of smoother topologies. For many molecular targets, the small spaces are often active or allosteric sites, and small molecules are more amenable to binding these clinically important epitopes.
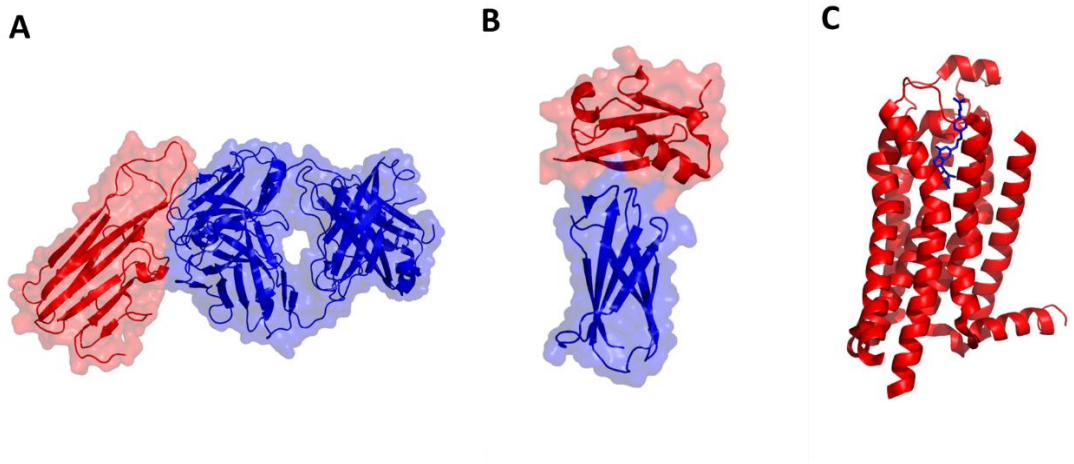
Figure 1.1: Engineered molecular targeting agents can exhibit a wide array of binding interfaces. A) Adalimumab (blue) Fab in complex with TNFα (red) (PDB:3WD5). B) Fn3 (blue) in complex Yes1-SH2 (red)PDB: 5mtj. C) Small molecule (blue) in complex with Adenosine A2A receptor (red) (PDB:4ug2).

Discovery of effective molecules can be pursued via rational design, though the complexity of the protein targets, the conformational diversity of binding interfaces, and the imperfect knowledge of binding energetics limit such approaches. Alternatively – or more powerfully, in concert – rationally-guided combinatorial approaches can evaluate libraries of many candidates to discover functional ligands. In this regard, proteins as targeting agents are superior to small molecules in the ability to efficiently generate and search a much larger chemical for novel, functional paratopes. Technologies that link protein phenotype to genotype empower high-throughput evaluation and isolation of

millions to billions of variants. Yeast display[34–36] , phage display[37–39], and similar

concepts have proven highly powerful in drug discovery and development.

1.3 Molecular Targeting Limitations: Epitopes

As revolutionary as molecularly targeted drugs have been, they are limited in their

efficacy at the start. The vast majority of these targeting agents and therapeutics rely on a

single drug-target interaction, which can lead to several issues. For many, the drug action

site is not unique to the biomarker being targeted. Due to the conserved nature of active

sites within many receptors, the most 'druggable' sites are often ones that are shared

within the family of receptors or enzymes that otherwise are excellent biomarker

candidates. For example, one large class of anticancer drugs, receptor tyrosine kinase

inhibitors, have been shown, post-FDA approval, to have activity across a broad

spectrum of receptor tyrosine kinases outside of the intended target[40], largely due to the

homology in the catalytic domains[41,42]. This results in a large array of side effects[43–46]

and can accelerate the development of drug resistance[47,48]. Drug development efforts

have used combinatorial chemistry[49–51] – for small molecules – or directed evolution[52–55]

– for proteins – or rational design – for both – to improve specificity and activity. Yet,

despite anecdotal advances[56–58] a multitude of epitopes remaining challenging.

1.4: Molecular Targeting Limitations: Biomarker Uniqueness

Even if a molecular targeting agent is highly specific, typical biomarkers naturally exists in low to moderate levels throughout the rest of the body; thus, targeting these can lead to undesirable side effects. Examples include non-mutated but upregulated receptors such as EGFR and HER2, where binding of targeted antibodies on off-target tissues can lead to serious side effects[44,59,60]. This on-target/off-tumor binding is also a significant problem in the emerging area of immunotherapy, where on-target, off-tumor binding can lead to signaling cascades resulting in treatment rejection and cytokine storm induction[61–64].

This difficulty in finding a biomarker unique to the disease is also a large issue in the burgeoning field of targeted imaging and early disease detection. For a variety of imaging agents delivered systemically, background signal can be high due to poor biodistribution and passive emission of the contrast agent. When systemically delivered, targeted contrast agents for PET and MRI imaging will have some signal in tissues without the biomarker of interest due to the contrast agent being 'always on'[11,65,66]. To reduce background for these agents, finely tuned biodistribution parameters and dosing regimens are required, which can be difficult to apply to a large patient population[8,11,28,29,67–71].

Potential solutions include identification of targets truly unique to diseased tissue using proteomics[72,73] and other discovery strategies[74,75]. Yet the presence of the full genome in all cells limits potential uniqueness to mutations – such as EGFRvIII [76,77] and unique post-translational modifications, which are rare. Other strategies include multi-layered targeting strategies in which a semi-selective targeting approach is coupled to a biological mode of action that is also semi-selective to the disease state.

1.5 Thesis Overview

The goal of my thesis is to address these challenges by considering how to improve specificity through multiple mechanisms of action, either via heterobivalent ligand development or dual-action imaging agent design. In Chapter 2, we propose a method of targeting tumor cells that utilizes a molecular fusion of two weak affinity ligands for two distinct biomarkers. We hypothesize that such a fusion will localize preferentially to cells that overexpress two disease biomarkers rather than just a single biomarker because of a multivalency advantage for the two-target cells; *i.e.* it will act as a biological AND gate. We mathematically model these fusions at the molecular and tissue scale and experimentally demonstrate enhanced selectivity. In Chapter 3, we use a modified perspective on this AND gate concept to develop a technology to improve the speed and efficacy of creating specific active molecules for challenging epitopes. We propose and validate a two-ligand fusion with a small molecule attached to a non-antibody protein scaffolds to provide specific binding to a single target via simultaneous engagement of two epitopes, including the active site. In Chapter 4, in collaboration with Dr. Morgounova and Prof. Shai Ashkenazi of the University of Minnesota Biomedical Engineering, we develop a novel imaging probe enhances selectivity by requiring target-selective activation. The probe emits signal only when activated by an enzymatic biomarker and was detected via the reporters' long photoacoustic lifetime. Collectively, these studies advance the capacity and understanding of molecular medicine.

Chapter 2: Biological AND gate

2.1: Introduction

2.1.1: Biomarker targeting lacks tissue selectivity

Molecular targeting for disease detection and treatment has been instrumental in improving outcomes for patients suffering from cancer and inflammatory disease. Small molecules and proteins targeting EGFR, VEGF, and HER2 have greatly improved the prognosis for patients with breast and colorectal cancers[33,78–83]. More recently, molecular targeting of immuno-oncology targets has seen clinical success, with PD-1-targeting antibodies pembrolizumab and nivolumab and CTLA-4-targeting antibody ipilimumab being approved for use in the U.S. and E.U. to treat several tumor types, including renal cell carcinoma, melanoma, non-small cell lung cancer, Hodgkin's lymphoma, and metastatic squamous cell carcinoma[84,85]. However, these drugs can come with a variety of side effects. EGFR-targeting therapies are often dose-limited due to high EGFR levels in the liver and epithelial tissues[43,44]. HER2-targeting antibodies such as trastuzumab have seen uptake in cardiac tissue lead to adverse cardiac events[45,59,78]. In colorectal cancers, CEA as a biomarker has been shown to be indicative of malignancy in colorectal cancers[86–88], but additional studies have shown that benign tumors also upregulate CEA[89,90].

These examples demonstrate the limitations of many biomarkers and therapeutic targets; while these molecular targeted drugs are specific to the biomarker of interest, the biomarkers themselves are not specific to the diseased tissue and exist at non-negligible

levels around the body. Due to these low but non-zero levels of biomarkers in healthy tissues, the molecular targeting agents do not always localize and act on the correct cells. In addition to causing harmful side effects, this on-target/off-tumor localization of molecular targeting agents can sequester the drugs and reduce the effective dose when systemically delivered. In non-invasive diagnostics, such as molecularly targeted contrast agents for PET or MRI scans, this can also obscure potential metastasis and tumor boundaries by localizing to healthy tissues that contain the disease biomarker of interest, and decrease the sensitivity of the diagnostic method[9,33,91–93].

2.1.2: Biological AND gate concept

To address this issue of biomarker uniqueness, we hypothesized that a more selective targeting agent can be made by requiring two upregulated biomarkers for strong binding and localization. This molecular design, called a biological AND gate, can be created by designing a heterobivalent ligand that has moderate affinity, but high avidity, to two biomarkers expressed on the diseased cells. The design of the individual ligands requires sufficiently weak affinity to be predominantly unbound in the presence of a single target yet predominantly bound — via multivalent avidity — in the presence of two targets (Figure 2.1). This is in contrast to many bispecific antibodies, where the goal is to engage two targets via high affinity monovalent interactions. While this might improve coverage of heterogeneous tumors (as an OR gate), it exacerbates the tissue specificity issues. This is also distinct from bispecifics designed to crosslink two targets to elicit a specific therapeutic function, such as immune cell recruitment.

Monovalent: strong affinity to one target   Bispecific OR gate : strong affinity to two targets

Heterobivalent AND gate: weak affinity to two targets

Monovalent Ligand

Weak Interactions

Target Biomarkers
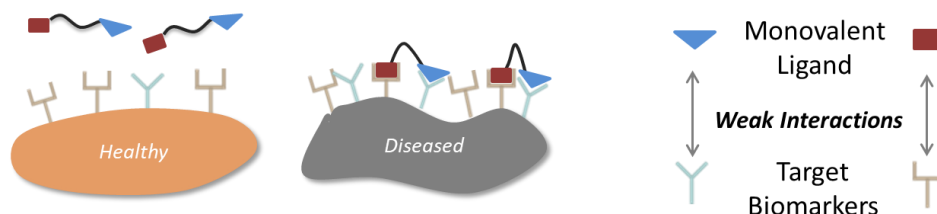
Figure 2.1: Concept of biological AND gate compared to other targeting mechanisms. Current molecular targeting agents consist of strong individual ligand:target interactions that cannot address the background expression level or overexpression in healthy tissue of the targeted biomarker. The biological AND gate would theoretically be able to bypass these cells and localize only to a cell with both biomarkers upregulates, greatly enhancing the selectivity over conventional targeting agents.

Examples of this AND gate concept can be seen in the literature, from fused natural ligands to bispecific antibodies. Some of the first modeling work in this area was the mathematical assessment of multivalent binding by Caplan and Rosca, who showed for both homobivalent and heterobivalent targeting agents, increasing the monovalent affinities does not necessarily increase specificity to the target cells[94,95]. A follow-up study showed that, for both flexible and rigid linkers connecting the two monovalent ligands, there is a theoretical optimal length to maximize the binding affinity of the second binding event[96]. Additional computational studies with functionalized nanoparticles and other highly multivalent targeting agents show that specificity is

improved when the monovalent affinity is weaker, and that this is related to the valency of the particle[97–99] .

There is experimental evidence to support these models as well. Heterobivalent ligands binding to melanocortin receptor 4 and cholecystokinin 2 receptor showed improved selectivity to dual positive cell lines over single-target expressing cells lines both *in vitro* and in *in vivo* mice models, and that this selectivity was more pronounced for the targeting ligand with a lower affinity[100,101]. A similar experiment with bispecific antibodies targeting EGFR/HER2 and CD4/CD70 showed that the reduction in affinity for any arm led to increased specificity towards dual target positive cell lines over single positive cell lines[102–105].

These studies demonstrate that the biological AND gate has the possibility of greatly reducing on-target/off-tumor localization, yet they also show the need for precision engineering and design to reach optimal specificity, while being robust to allow for heterogeneity within a patient. Many of these examples were not designed into the heterobivalent targeting molecules but were observed when various linkers and fusion methods were used to combine the monovalent binding agents. They show that the monovalent affinity of each binding arm matters, but also show that the surface expression level and linker length is crucial to increasing selectivity. Other factors that could be important are the orientation of the fused molecules and receptor biology/response to ligand engagement. With these elements in mind, we sought to develop a more comprehensive design model that could guide the design of

heterobivalent ligands into biological AND gate molecules. We then developed *in vitro* model systems to test these hypotheses.

2.2: Materials and Methods

2.2.1 Cell lines and cell culture methods

LS-174T cells were a kind gift from Dr. Tim Starr (UMN). A431 cells were a kind gift from Dr. Daniel Vallera (UMN). HEK-293T cells were a kind gift from Dr. Douglas Yee (UMN). Cells were cultured in DMEM with 10% v/v fetal bovine serum and 1% v/v penicillin and streptomycin (Gibco). Cells were grown at 37 °C and 5% CO_2. Cells were grown to 80-90% confluency and passaged at 20-fold dilution every 3-4 days. LS174T cells were plated at a 5-fold dilution and passaged every 3 days.

2.2.2 Generation of stable cell lines expressing CEA gene

Lentiviral transfer vector pLX304 containing the full-length CEA (CEACAM5) gene with C-terminal V5 epitope tag was purchased from DNASU. Packaging and envelope vectors were also a kind gift from Dr. Douglas Yee. HEK-293T cells were used to produce in-house lentiviral particles. Briefly, ~$5x10^6$ HEK-293T cells were transfected with 3 plasmids; vsv-g, dvpv, and a pLX304-CXCR4-V5 using Lipofectamine 3000 (Invitrogen) per manufacture protocol. Cells were allowed to

11

produce particles for 48 hours, and media harvested and stored at -20 C. Separate

batches of HEK-293T cells were then transduced with 1 mL lentiviral particle suspension

in media and 8 µg/mL polybrene. Cells were then maintained under selection pressure of

5 µg/mL blasticidin. Transduced HEK-293T cells were then subjected to FACS sort to

select for the top 5% of CEA-expressing clones.

2.2.3 Expression quantification

Cells to be tested were washed twice in cold 1% PBSA (10 g/L bovine serum

albumin, PBS). 50,000 cells were resuspended in 50 µL PBSA with 10 µg/mL mouse

anti-target antibody and incubated for 30 minutes at 4 °C. Cells were washed once in

PBSA and incubated in 20 µL of 10µg/mL goat anti-mouse IgG – FITC for 5 minutes at

4 °C. Cells were then washed in PBSA and analyzed via flow cytometry. In parallel,

Quantum™ Simply Cellular® anti-Mouse IgG beads (Bangs Labs, Fishers, IN) were

incubated with 10 µg/mL mouse IgG isotype control antibody (Abcam, Cambridge, MA).

Beads were washed and labeled in an identical fashion to cells being tested and evaluated

via flow cytometry.

2.2.4 Generation of EGFR- and CEA-binding libraries

The pCT plasmid was used as the yeast expression vector for the EGFR- and CEA-

binding libraries. Genes encoding for the E.6.2.6 (EGFR-binding)and C7.4.3 (CEA-

binding) Fn3 clones was mutated via error-prone PCR using nucleotide analogs[106]. PCR

products were then purified via gel electrophoresis and amplified. Final reactions were

then transformed into yeast with digested pCT plasmid to allow for homologous

recombination. Transformed yeast were cultured by shaking at 30 °C in SD-CAA media

(16.8 g/L sodium citrate dihydrate, 3.9 g/L citric acid, 20.0 g/L dextrose, 6.7 g/L yeast

nitrogen base, 5.0 g/L casamino acids), and expression of Fn3-PAS-Cys was induced for

>4 hours at 30 ºC with shaking by switching the yeast into SG-CAA media (10.2 g/L

sodium phosphate dibasic heptahydrate, 8.6 g/L sodium phosphate monobasic

monohydrate, 19.0 g/L galactose, 1.0 g/L dextrose, 6.7 g/L yeast nitrogen base, 5.0 g/L

casamino acids).

2.2.5 Target preparation

The extracellular domain of EGFR was purchased from Invitrogen. CEA was purchased

from Fitzgerald. Targets were biotinylated with 100 µg/mL biotin-NHS for 2 hours and

quenched with 1 M Tris HCl. Biotinylation was verified via flow cytometry. Briefly, b-

EGFR or b-CEA was incubated with either pCT-E.6.2.6-expressing yeast (for b-EGFR)

or pCT-C7.4.3-expressing yeast, respectively. Yeast were washed and incubated with

streptavidin-AlexaFluor647 (Invitrogen), then washed and analyzed on BD Accuri C6

(BD Biosciences).

2.2.6 Isolation of low affinity ligands

Induced EGFR-binding library and pCT-E6.2.6 were incubated with 10 µg/mL 9e10 (Santa Cruz Biotechnology) and 300 nM b-EGFR. CEA-binding library and pCT-C7.4.3 were incubated with 300 nM b-CEA. In parallel, pCT-E.6.2.6 and pCT-C7.4.3 expressing yeast were labeled 300 nM b-EGFR and 300 nM b-CEA, respectively. Yeast were washed and incubated with 10 µg/mL goat-anti-mouse-FITC (Sigma-Aldrich) and 2 µg/mL streptavidin-AlexaFluor647. Yeast were washed and sorted using FACSAria (BD Biosciences). This process was repeated, and the resultant clone DNA was isolated from yeast via Zymoprep plasmid extraction kit (Zymo Research, Orange County, CA). Plasmids from zymoprepped yeast populations were transformed into NEB5α bacterial cells and cultured in lysogeny broth with 100 µg/mL ampicillin. Amplified DNA was then extracted from clonal bacterial cultures via miniprep kit (Epoch Life Sciences, Houston, TX). Clonal plasmid isolates were then submitted for Sanger sequencing.

2.2.7 Cloning and DNA sequences

For production of individual Fn3 clones, the T7 bacterial expression vector pET-22b was used for production of clones in *E. coli*. pET-22b was digested using NheI and BamHI enzymes. Candidate clones were amplified via PCR and digested with NheI and BamHI enzymes. Digested pET and Fn genes were recombined using T4 DNA ligase enzyme. Briefly, 100 ng digested pET, 25 ng digested Fn gene, 1 µL T4 DNA ligase, 1.5 µL T4 DNA ligase buffer, and deionized H_2O were combined to a final volume of 15 µL and incubated on the benchtop for 20 minutes. Ligations were transformed into NEB5α

bacterial cells and plated on LB agar + 50 µg/mL kanamycin plates. Plasmids from

resulting colonies were extracted and sequenced as previously described.

For production of fusion proteins, an Fn-linker-Fn gene was expressed in pET-

22b plasmid (pET-FnFn, Figure 2.10). The translated sequence, AS-Fn-

GSGGGSGGGKGGGGT-Fn-ELRSHHHHH contained restriction enzyme cut sites

flanking both Fn genes. Restriction sites for NheI and BamHI enzymes flanked the N-

terminal Fn gene, and restriction sites for KpnI and SacI enzymes flanked the C-terminal

Fn gene. Fn genes were cloned into the pET-FnFn plasmid as previously described.

2.2.8 Production of ligands in E.coli

Clones transferred into pET vector were transformed into T7 Express cells from

NEB. *E. coli* and were grown in 100 mL cultures to an OD between 0.5 and 1.

Production was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 37 °C for

two hours. Cultures were then pelleted, and cells resuspended in a bacterial lysis buffer

(50 mM sodium phosphate (pH 8.0), 0.5 M sodium chloride, 5% glycerol, 5 mM CHAPS,

and 25 mM imidazole, cOmplete Mini EDTA-free Protease Inhibitor Tablet). Cells were

lysed via five or more freeze-thaw cycles, and the insoluble fraction was removed via

high speed centrifugation and filtered (0.22 µm filter paper).

Clones were purified and conjugated to small molecule via a modified metal

affinity chromatography protocol. Briefly, cell lysate was applied to a HisPur Cobalt

spin column (ThermoFisher Scientific), equilibrated with 15 mM imidazole + 50 mM sodium phosphate (pH 8.0), and centrifuged at 700g for 2 minutes. The column was then washed three times with 400 µL of 15 mM imidazole to remove protein lacking the His$_6$ purification tag. His$_6$-tagged protein was eluted from the column with 300 mM imidazole + PBS. Protein concentration and purity was analyzed via SDS-PAGE gel.

2.2.9 Cell-based affinity titrations

Detached cells were washed and labeled with varying concentrations of each Fn clone or Fn fusion for > 30 min at 4 °C. Cells were pelleted at 300g for 3 min and washed with 1 mL ice cold PBSACM, then labeled with 20 µL anti-His6 FITC conjugate (ab1206, Abcam, 13 µg/mL) for 5 min at 4 °C. Cells were again pelleted and washed with 1 mL ice cold PBSACM. Fluorescence was analyzed via flow cytometry, and median fluorescent values were obtained. Equilibrium dissociation constant (K_d) was determined using a least squared regression to fit to median fluorescence results assuming a 1:1 binding model. K_d values are presented as the average ± standard deviation.

2.3: Results

2.3.1: Quantification of K_d enhancement with ligand fusion

To quantify the benefit of the AND gate over monovalent targeting, we first described the dual binding mode of the heterobivalent ligand. The initial binding

probability is proportional to the monovalent ligand's association rate constant (k_{on}).

Once one of the two ligands are engaged, the local concentration of the second ligand is

increased to the space it can survey when tethered to the surface of the cell. For effective

linkers sufficiently long to reach a second target epitope this subsequently increases the

effective association rate of the unbound second ligand. Simultaneous engagement of

both targets also requires dissociation of two bound ligands, which increases the

residence time of at-least-singly-bound ligand thereby increasing effective affinity. .
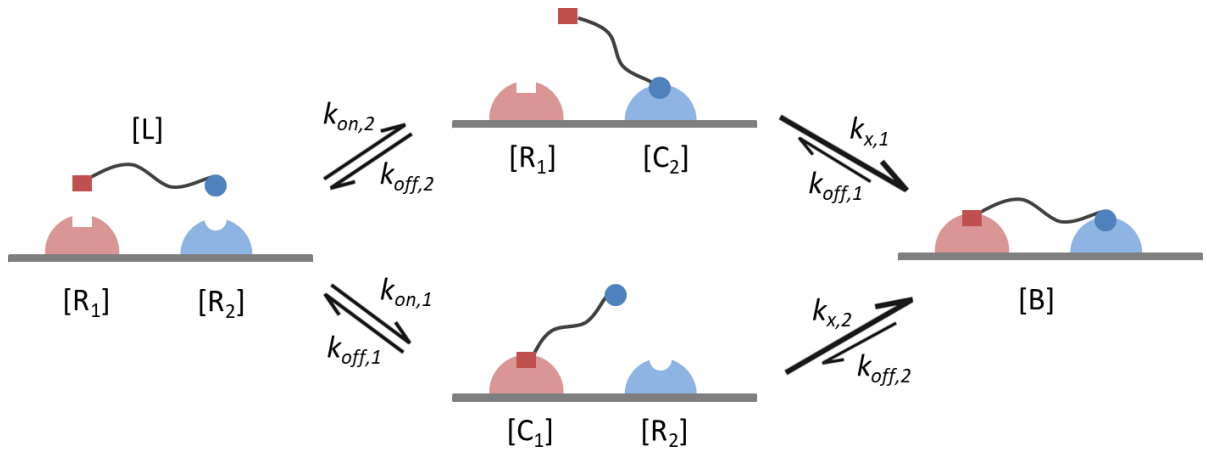
Figure 2.2: Binding mechanism for biological AND gate. Initial binding event is dependent on the association rate of the individual ligands within the fusion. Once the single ligand:receptor complex has been made, the association rate is now increased for the unbound ligand due to the increased local concentration of the tethered fusion. Upon crosslinking of receptors, two dissociation events must occur near simultaneously for the fusion to fully dissociate, greatly increasing the residence time of the fusion.

The k_{on} and k_{off} values for each initial binding event are intrinsic to the individual

ligands. Dissociation of the individual ligands also depends only on the strength of the

interaction with its target. However, the association rate of the second binding event

(cross-linking rate) will be different due to the change in accessible volume of the tethered

ligand. To estimate this variable, we assumed that the only change to the ligand is that it is

now in a constrained volume and that the rate of local atom rearrangement is unaffected by the initial binding event. The resulting relationship is that k_x/k_{on} is proportional to the ratio of accessible cell surface area to total surface area (Appendix A). To calculate the accessible surface area, a geometric model was developed to describe the system (Figure 2.3A). The receptor is assumed to be a hemisphere of radius r. The ligand has a fully flexible linker that does not affect ligand:receptor interactions and can survey the entire enclosed sphere of radius l with equal probability. The crosslinking rate enhancement, which is proportional to the concentration of the target receptor, can then be found as a function of the length of the linker (Figure 2.3B). For a target radius of 20 Å, this corresponds to a linker length of 22 amino acids for a maximum crosslinking rate enhancement of 171 times the monovalent k_{on} value. This result agrees with other published literature for fully flexible linkers[96,107].
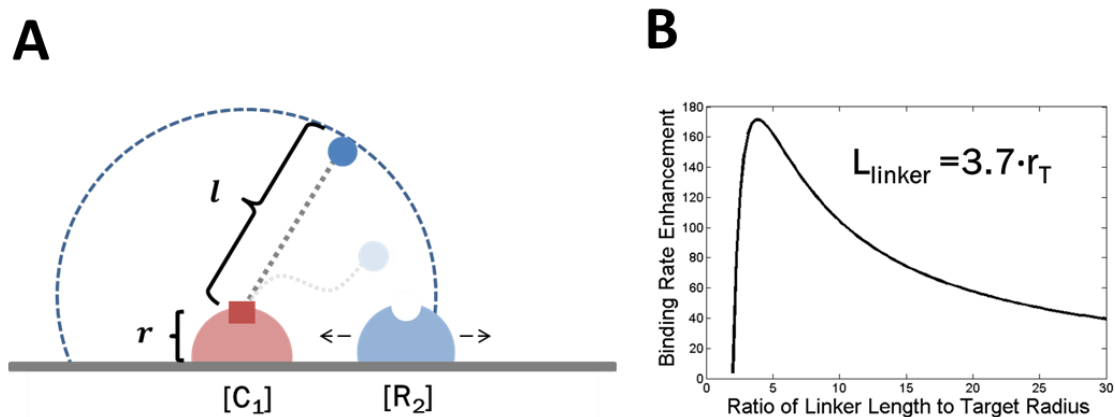
Figure 2.3: Increased local concentration of second ligand increases k_{on} of second binding event. A) Schematic of geometric model. Unbound ligand is allowed to survey available space defined by the length of the linker, l. and the receptor radius r. B) Effect of linker length on increase in k_{on} for second binding event. The maximum indicates the optimal linker length for the fusion molecule.

2.3.2 Specificity improvement estimation via in silico biodistribution model

In order to identify the effect of this second binding event in a transient system such as the human body, we incorporated this two-event binding mechanism into an *in silico* biodistribution model the system using a transport model adapted from previous literature[108,109] . (See Appendix for full model description). Comparison of tumor uptake of a monovalent and bivalent ligands of the same ligand:receptor affinity showed a large increase in selectivity of the bivalent ligand at 290 nM (Figure 2.4). The appearance of an affinity optima supports the AND gate hypothesis, demonstrating that for the given system, there is an individual binding strength that is strong enough to localize to cells with both receptors, but weak enough to minimize uptake to cells containing only one of the two receptors compared to a monovalent targeting ligand.

Comparison of selectivity over a range of receptor expression levels revealed that the optimal affinity was linearly related to the expression of the target receptor (Figure 2.4). This relationship does not hold at monovalent affinity values below 1 nM, which could be due to reaching a dissociation rate constant value that is much slower than the transport kinetics of the ligand through the tumor. Other considerations are internalization and recycle rates of the receptors. This demonstrates that while the AND gate concept is theoretically valid, there is no 'one size fits all' AND gate ligand design. However, depending on the target expression level, a specificity improvement of 200-fold is theoretically achievable.
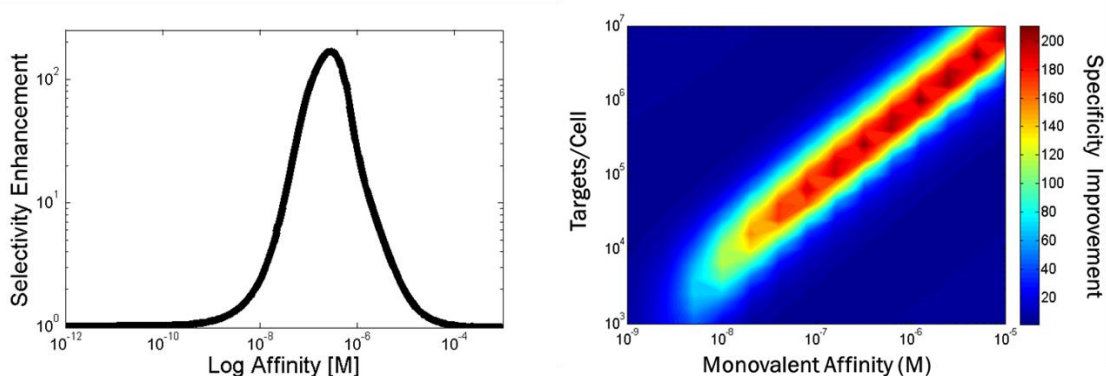
Figure 2.4: Heterobivalent targeting improves cell selectivity over 100-fold. (Left) Selectivity of an AND gate fusion molecule over an individual targeting ligand is over 100-fold higher for monovalent affinities of 270-500 nM. (Right) Selectivity is also a strong function of the target receptor surface expression. At 4 hours post-injection, a 200-fold selectivity of the targeting ligand is achievable.

2.3.3 Engineering of low-affinity ligands

With these results in hand, we moved forward to engineer ligands with a range of

affinities to test the accuracy of the model. To reduce variance from other system

elements, such as monovalent ligand size, charge, binding site, and binding orientation,

we used two high-affinity ($K_d \sim 1$ nM) ligands that were developed from the same

molecular scaffold: the tenth type 3 domain of fibronectin (Fn). This domain exists as a

genetic fusion to several homologous Fn domains, which indicated that it likely would be

a modular protein scaffold whose activity would be unaffected by fusion partners.

Moreover, a previous study demonstrated the ability of Fn-Fn fusions to provide bivalent

activity[110]. The Fn scaffold has been previously engineered to bind to a multitude of

targets, including VEGFR2[111,112], CD20[113], transferrin[55], EGFR[114], and CEA[115]. The

latter two targets, EGFR and CEA, are important biomarkers for many cancer indications,

20

specifically colorectal cancers[33,48,116]. For this reason, we chose to use Fn targeted

towards these molecules as our high-affinity parental ligands. The EGFR binding Fn

chosen was clone E6.2.6, which binds to the extracellular domain of EGFR with a K_d of

0.26 nM[110]. The CEA binding Fn chosen was clone C7.4.3, which binds membrane-

bound CEA at a K_d of 2 nM[117].

We evolved the ligands for reduced affinity by directed evolution. Each parental

clone was randomly mutated via error-prone PCR of the entire gene and displayed on the

surface of yeast to provide a phenotype-genotype linkage for high-throughput evaluation

and selection[34,106]. Previous work has shown that the yeast surface display system can

discriminate between similar ligand affinities[118]. We aimed to isolate clones with a range

of binding affinities, with a K_d of 300 nM as the optimal affinity (Figure 2.4). To do this,

we labeled our yeast populations at 300 nM biotinylated target (either EGFR or CEA),

and allowed the population to approach equilibrium. Target binding and Fn expression

per cell was quantified via flow cytometry and clones with reduced binding, normalized

to Fn expression, were collected in multiple gates (Figure 2.5). These populations were

then sorted a second time to remove any non-binding clones from the populations.

$$f_b = \frac{[Fn3]}{[Fn3] + K_d}$$
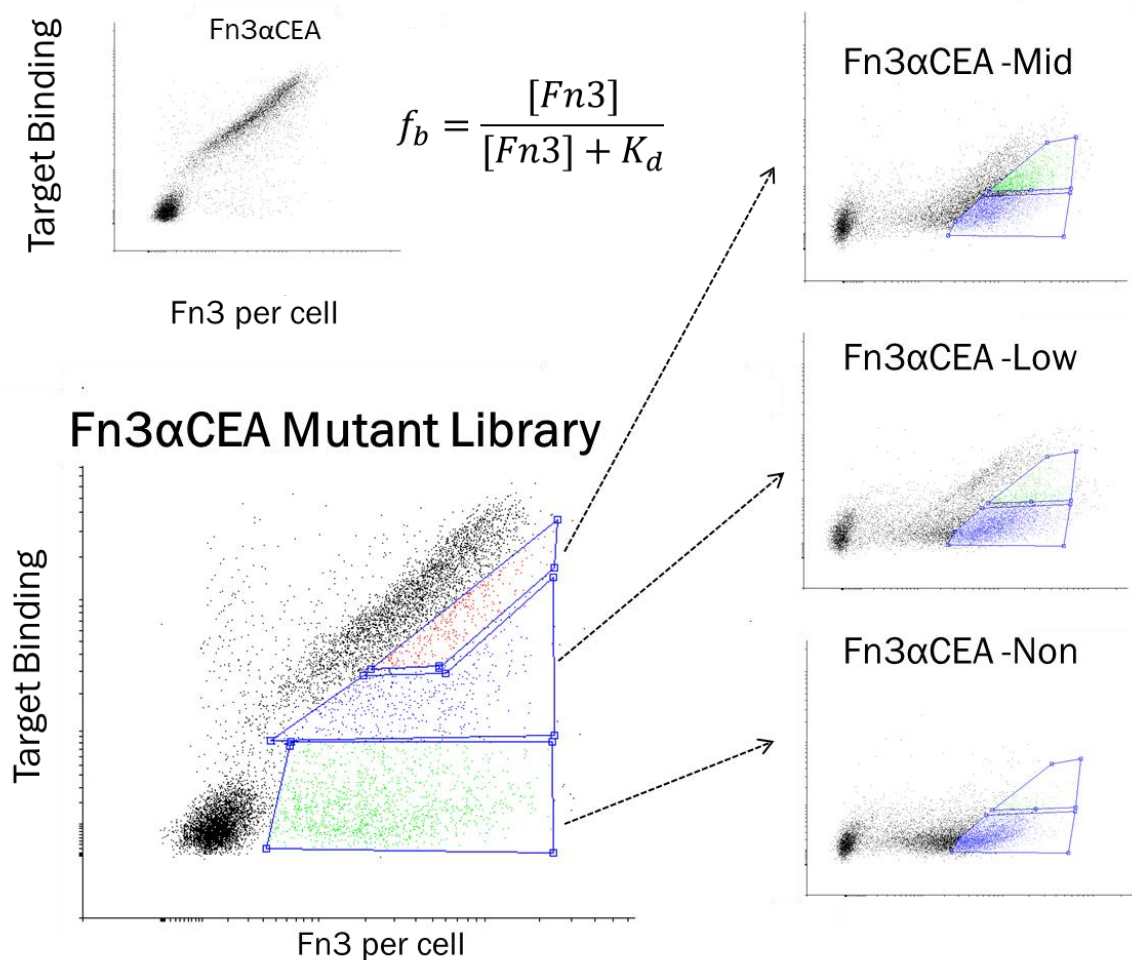
Figure 2.5: Isolation of low affinity clones. Population generated from a mutated CEA-binding or EGFR-binding Fn clone were sorted twice at 300 nM target. Multiple affinity ranges were collected, and clones were isolated and sequenced.

2.3.4 Affinity determination of isolated clones

Isolated ligands from mid-mid and low-mid populations were then produced and purified in *E. coli* both as single domains and as fusions to a non-binding Fn scaffold to mimic the heterobivalent structure. Binding affinity was measured via cell-based affinity

titrations (Figure 2.6). Interestingly, though the fine affinity discrimination has been

shown using yeast surface display, clones from the same populations demonstrated vastly

different affinities towards their target of interest (Figure 2.7 and 2.8). This result likely

emerged from the difference between target formats in the two modes of binding

analysis: soluble recombinant extracellular domain during flow cytometry selections of

yeast-displayed Fn vs. intact cellular-bound target in the context of affinity titrations.

This hypothesis motivates the use of detergent-solubilized cell lysate[119] or intact

cells[73,120] as targets during future evolution. Alternatively, it is plausible that we isolated

mutants that are less stable in the soluble form than in the context of the yeast display

construct.
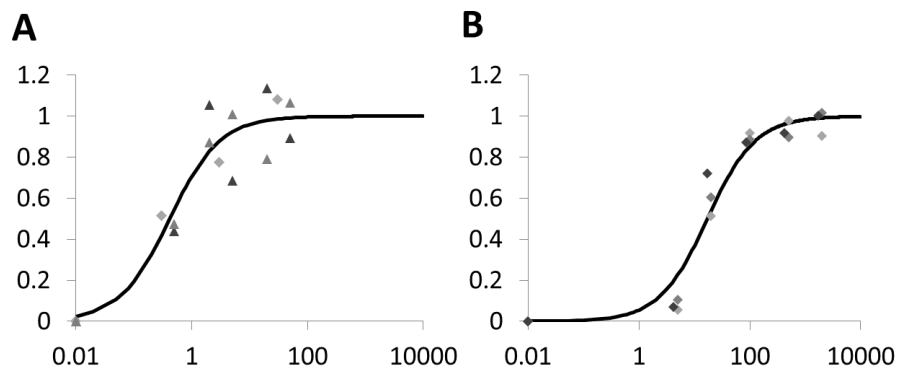
Figure 2.6: Example titrations. A) Clone E6.2.6. B) Clone eL3. Clones were produced with a 6-His repeat tag for purification and antibody detection. Each clone was titrated at multiple concentrations against either A431 (EGFR^{high}) or LS174T (CEA^{mid}) cells. Binding was detected using an anti-6-His fluorescent antibody and analyzed via flow cytometry. K_d values were determined by fitting individual experimental titrations and fitted K_d values were averaged across all replicates.

Mutant	Mutations	Affinity (nM)
E6.2.6	---	0.2
eL3	N78S	17 ± 4
eL6	G50S	15
eM3	E9G, S17G, N78D	60 ± 10
eM1	P80L	71
eM6	W51R	100
eL1	I54T, F81S	181 ± 134
eL5	T14A, V45A, N78S, I88V	3,100 ± 400
eM4	Y30C, I32V	3,000
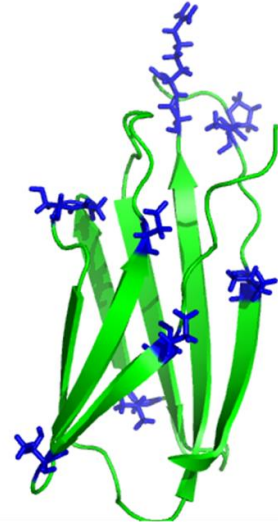
Figure 2.7: Affinity values for EGFR-binding Fn clones to A431 cells. Clones were individually produced and titrated on EGFR-expressing A431 cells. Clones with >3 replicates are reported with standard deviations. Structure shown is an Fn3 (1TTF) with clone mutations noted in blue.

Mutant	Mutations	Affinity (nM)
C7.4.3	--	13 ± 4
cLL3	T5A	2.5
cLL8	K64E, S80P	80
cML8	Y33H, S44T	140 ± 98
cLL9	K64E, T77A	238 ± 18
cML2	S5cP, H29Y	2,000
cML3	S25N, S27G, H29R, S81G	5,400
cLL10	Y26S, E39G	> 1000
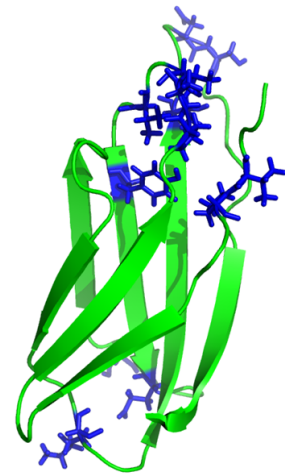
Figure 2.8: Affinity values of CEA-binding Fn clones to LS174T cells. Clones were individually produced and titrated on CEA-expressing LS174T cells. Clones with >3 replicates are reported with standard deviations. Structure shown is an Fn3 (1TTF) with clone mutations noted in blue.

2.3.5 Fusion of non-binding Fn3 modifies affinity

Though the yeast display / soluble target binding results were inconsistent with soluble Fn / cellular target K_d values, clones with a range of affinities were identified and moved into the heterobivalent AND gate expression construct (Figure 2.7). Due to the native structure of Fn as a genetic fusion, we chose to fuse the binding Fn domains in a similar fashion. A pET-22b production vector was modified to express a fusion protein where the N-terminal ligand was a EGFR-binding clone, and the C-terminal binding ligand was a CEA binding clone. The linker, GSGGGSGGGKGGGGT, was adapted from previous work using Fn fusions[110]. Select mutants were cloned into the expression vector and tested for their affinity to cell line containing one of two targets: A431 for EGFR and LS174T for CEA.
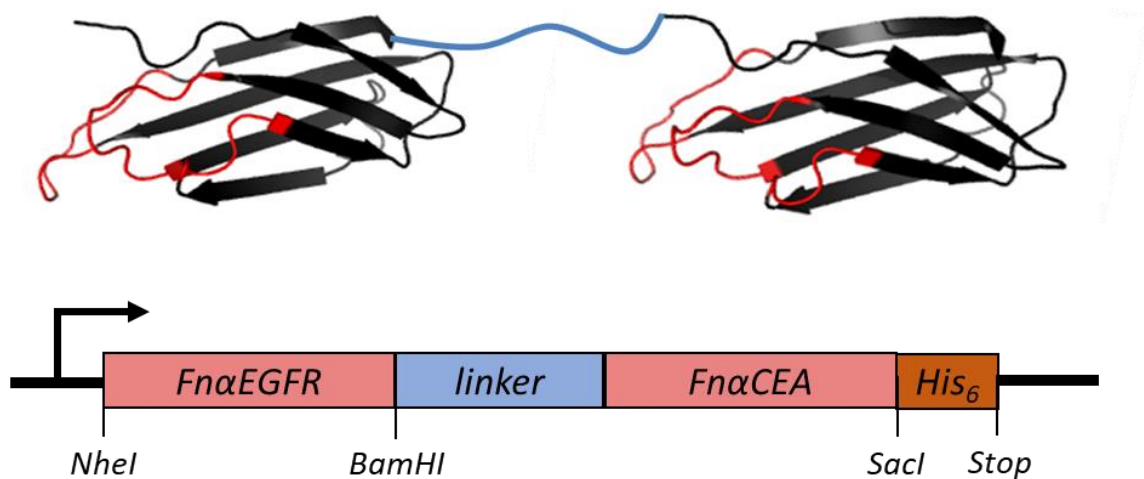
Figure 2.9: Fusion construct. Fn-linker-Fn clones were produced in E. coli with a C-terminal 6-His tag. Restriction digest sites flanking both the N-terminal and C-terminal Fn genes were used to facilitate efficient cloning.

The fusion of the non-binding Fn3 (RDG) had vastly different effects on the affinity of the clones (Table 2.1). For the EGFR-binding clones, the fusion of the linker-RDG was detrimental to three out of four clones tested, including the parental clone. In the case of clone eL3, the affinity was completely ablated at concentrations up to 1 µM. A similar story was seen with CEA binding mutants. As with clone eL3, all four RDG-linker-CEA clones showed a decrease in affinity to CEA, with all three mutants showing no binding at concentrations lower than 1 µM.

Table 2.1: Affinity of Fn clones in fusion construct

Fn3αEGFR mutants			Fn3αCEA mutants		
Fusion Clone	**Monovalent Affinity (nM)**	**Fusion Affinity (nM)**	**Fusion Clone**	**Monovalent Affinity (nM)**	**Fusion Affinity (nM)**
Fn3αEGFR-RDG	0.4 ± 0.03	21 ± 6	RDG-Fn3αCEA	15 ± 4	89 ± 30
eM3-RDG	26 ± 2	27 ± 5	RDG-cM2	2,000	10.000
eL3-RDG	17 ± 3	>10,000	RDG-cM8	140 ± 98	>10,000
eL5-RDG	$3,100 \pm 400$	$19,300 \pm 3000$	RDG-cL9	238 ± 18	>10,000

This effect of the Fn fusion on the monovalent affinity was also found when the binding Fn were moved to the alternate position within the construct, and when they were fused to other binding clones. (Table 2.2). Parental clone E6.2.6 had a marked drop in affinity with both RDG and C7.4.3 partner Fn at either the N- or C-terminal position, with its location in the C-terminal position being the most detrimental. Clone C7.4.3 affinity was not as influenced as E6.2.6 but did appear to consistently decrease in affinity when in the C-terminal position. As an additional comparison, an alternative EGFR-binding clone (D) was tested in the fusion construct, and interestingly showed a stronger

affinity when in the N-terminal position than in the monovalent context. These results

demonstrate the lack of modularity in the system, and that the effect of the fusion is clone

and orientation dependent.

Table 2.2: Effect of various fusion partners on clone affinity

Fn	$f_{'nb'}$	f_{nb}-Fn	Fn	Fn-f_{nb}
A'.101	RDG	18 ± 5 nM	0.4 ± 0.03 nM	21 ± 6 nM
	C7.101	9.5 ± 6 nM		102 ± 18 nM
C7.101	RDG	89 ± 30 nM	15 ± 4 nM	16 ± 4 nM
	A'.101	35 nM		23 nM
	NeLM3 (A'.1)	69 ± 10 nM		X
	eL3 (A'.2)	**800 ± 200**		X
D.101	RDG	1.6 nM	2.0 ± 0.9 nM	*0.38 ± 0.09 nM
	C7.101	1.2 ± 0.9		0.4 ± 0.2 nM

2.3.6 Selectivity is improved with reduced affinity

We then moved forward to test the selectivity of the new fusions. Three fusions, E6.2.6-

C7.4.3, eM3-C7.4.3, and eL3-C7.4.3 were produced. Each clone was titrated on one of

three cell lines: A431 (EGFR^{high}/CEA^{neg}), HEK-CEA+(EGFR^{low}/CEA^{high}, and LS174T

(EGFR^{mid}/CEA^{mid})(Figure 2.10A). The K_d values for each fusion on each cell line was

assessed (Figure 2.10B). Preliminary results show an increase in the selectivity of eL3-

C7.4.3 to LS174T cells over either A431 or HEK-CEA+ cells. However, the difference

is not statistically significant. Further potential AND gate candidates, particularly CEA-

binding mutants proved difficult to produce, and more rigorous extraction methods are

needed to isolate the ligands.
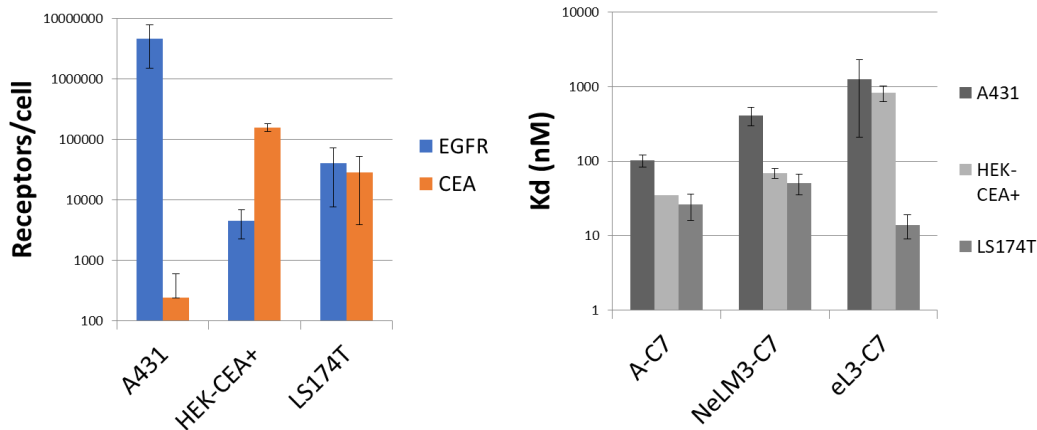
Figure 2.10: Low-affinity fusion is more selective to double-positive cells. (Left) Expression of 3 cell lines was determined using a fluorescent standard bead kit. Affinity values for each fusion were fit using a 1:1 binding model, and error bars represent standard deviation of 3 or more replicates. eL3-C7 shows the highest selectivity to LS174T cells over A-C7 and NeLM3-C7, both from the weaker affinity to A431 and HEK-CEA+ cells and a stronger affinity to LS174T cells than the alternative fusions.

2.4: Discussion

2.4.1: Biodistribution model validation of AND gate concept

The aim of this work was to address the issue of biomarker selectivity in molecular targeting of disease by requiring upregulation of more than one biomarker to be present for targeting molecule localization. The hypothesis was that a biological AND gate could be created by fusing two weak but non-zero affinity ligands that bound two different biomarkers, and that this fusion would preferentially bind to cells that upregulated both biomarkers over cells that upregulated only one or neither. *In silico* calculations showed that this was theoretically achievable, and that an optimal affinity for

28

the individual ligands existed. The model also demonstrated the importance of other biomarker properties. Receptor expression had a strong effect on the optimal affinity. At high receptor concentrations, a much weaker ligand is necessary. At high receptor concentrations, there is increased probability that the bivalent ligand can bind to the target, thus requiring a faster dissociation rate to remain selective. Additionally, the slower dissociation rate can increase the probability that the ligand will internalize with the receptor, which also ablates the chance of the second binding event to confer selectivity. This highlights the importance of biomarker characterization when designing an AND gate or any other targeting agent.

2.4.2: Fusion Modularity

The selected Fn scaffold for use as a modular system was based on previous experiments and its natural domain-oriented structure. However, throughout this study it became clear that the modularity of these fusions was not consistent with this hypothesis. This result is one that, while not entirely surprising, is often underappreciated when creating fusion molecules of any kind, from linked small molecules to dimeric scFv fusions. One should expect that modifying a molecule will confer some activity change, whether due to steric interactions of the modification with the target, or structural changes to obtain more stable configuration of the new molecule. Additionally, increasing the relative size of the fusion/modification will increase the effect on the initial molecule. This is particularly important when modifying small proteins, such as the Fn, natural ligands and low molecular weight molecules.

2.4.3 Affinity engineering of fusion ligands

One culprit could be that the mutagenesis strategy, rather than resulting in a library of stable but reduced affinity clones, led to a library of clones whose affinity was modulated by a reduction in stability. These two outcomes would manifest as the same phenotype under the sort conditions used. Additionally, the linkage of the second Fn via genetic transcription could also be a suboptimal method of linkage. One method of addressing this issue is to focus the mutagenesis strategy on a known paratope. This requires *a priori* knowledge of the paratope, which can be identified on the parent clone with many validated techniques[121–123]. In addition to focused mutagenesis, the construct expressed on the display organism (yeast, phage, etc.), should also be designed to more closely mimic the context in which the ligands will ultimately be utilized. It could be that, had the expression construct contained a non-binding Fn fused to the mutated Fn genes, the isolated clones would have activity more consistent with that of the corresponding soluble format. In addition, stability selections could be performed in addition to affinity selections to enhance robustness.

2.4.4: Additional AND gate design considerations

The mathematical model framework, directed evolution of reduced affinity, and experimental validation of enhanced selectivity set the stage for further development and study of deeper complexities of the system. For the AND gate ligand, the linker length

and structure likely impact overall performance. While our geometric calculations

suggested 22 amino acids as an optimal length, this is predicated on the assumptions of

stability, binding orientation, and permutation length of the linker. The extent to which

these assumptions are inaccurate are dependent on the ligands and targets being utilized.

Binding orientation of the ligand could greatly affect the space that the second unbound

ligand can survey, which in turn can change the k_x for either ligand. Linker flexibility is

also an important property of the AND gate. Inclusion of rigid portions of the linker

could improve the accessibility of the unbound ligand to the second receptor, thereby

increasing k_x. Future work to optimize and understand linker length and composition will

be valuable to advance technological utility and fundamental understanding.

2.5: Conclusion

In conclusion, biological AND gates have the potential to improve selectivity over

conventional mono-targeted therapies. We have shown *in silico* that there is an optimal

individual ligand affinity that can result in over 100-fold selectivity of a biological AND

gate molecule over a monovalent targeting ligand. We engineered clones to bind to two

clinically relevant targets at various affinities. We showed that the chosen Fn ligand

molecules are non-modular in fusion contexts, and that engineering molecules in the

context that they will be used is a key lesson in designing these heterobivalent systems.

Future work includes optimizing production of fusion molecules and testing selectivity

on cell lines expressing differential values of two biomarkers.

Chapter 3: Engineering protein-small molecule fusions to enhance target selectivity

3.1: Introduction

3.1.1 Therapeutic small molecules lack isoform specificity

Diagnostic biomarkers and targets of molecular therapeutics have been discovered in the emerging era of personalized and precision medicine. However, identifying selective, potent molecular targeting agents for these molecules has long been a difficult task for researchers. A variety of techniques, including combinatorial chemistry[50,124–126], structure-guided drug design[127–130], and repurposing of natural ligands[131–133], has been used to identify small molecule and protein binders. Yet, these molecules often exhibit off-target activity towards proteins within the same class/family[45,79,134,135], lack the necessary affinity for physiological efficacy[135,136], or bind without eliciting the desired activity (agonism or antagonism)[46,134].

One reason this method fails in many cases is due to the limitations of both small molecules and proteins as drug scaffolds. Small molecules often don't provide enough surface area to confer strong or unique interactions with a target[137,138], and screening techniques are often limited by size and ease of synthesis of new compounds[135,139–141]. Biologics, such as antibodies, can provide a much larger surface area for interaction and larger libraries for screening, yet their chemical and shape diversity is reduced relative to small molecules. Additionally, there are few high-throughput screening techniques for either small molecules or biologics that can select for a certain drug action rather than

passive binding[141–145]. Both antibodies and small molecules have various advantages in the types of epitopes they can target as well. Antibodies are superior at targeting large, flat topographies[146,147], while small molecules are able to fit in active sites and grooves within a receptor or enzyme than are inaccessible to an antibody[127,148–150]. Without prior structural knowledge, it is exceedingly challenging to know which scaffold would be most advantageous in developing a potent and selective drug towards a novel molecular target.

3.1.2 Engineering protein ligands to confer small molecule selectivity

Building off what was learned while developing a biological AND gate (Chapter 2), we believe that the concept of combining two distinct binding motifs into one ligand can be used to improve specificity and activity of a small molecule or biologic. We proposed that by evolving a protein scaffold binding 'partner' that is already linked with a starting ligand of interest, a highly functional and selective molecule could be engineered that could not be easily or efficiently achieved via *de novo* protein engineering or ligand-based drug design alone (Figure 3.1A). We hypothesized that we could use yeast surface display for high throughput discovery and evolution of these protein—small molecule fusions (PriSMs). For efficient synthesis and analysis, we proposed to use a C-terminal peptide linker with a terminal cysteine residue to attach a small molecule via a maleimide functional group to the protein. With this method of library conjugation, we hypothesized that various cell-sorting techniques could identify

ligands with improved affinity and selectivity while still retaining functional activity of the starting small molecule.
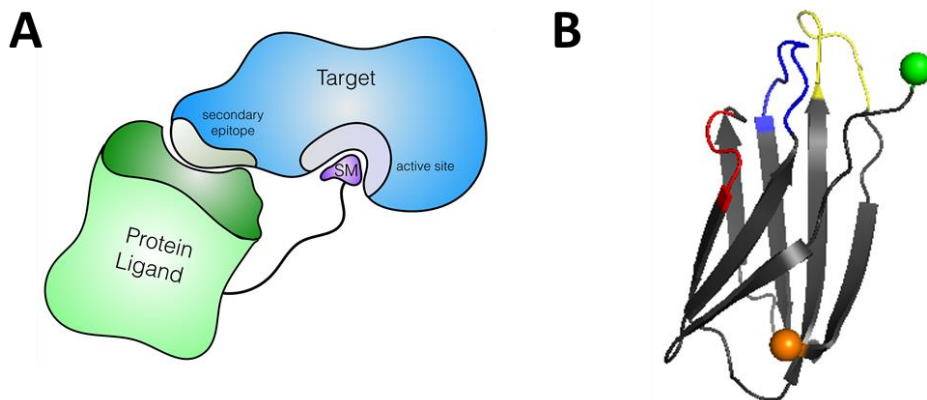
Figure 3.1: PriSM concept. A) Schematic of hypothesized binding mode. The small molecule of interest binds to the target protein's active site to convey agonist/antagonistic function but lacks specificity to the target isoform of interest due to active site homology between isoforms. The protein ligand will be engineered to provide selectivity by binding to a distal, unique epitope on the target protein. This additional binding event adds avidity and increases the activity of the fusion over the small molecule alone. B) Protein ligand scaffold. The 10th type III domain of fibronectin (Fn3) is a 94 amino acid protein that has previously been engineered to bind selectively to many targets[106,151–153]. A library of mutants was created by diversified the colored loops (blue: BC loop, red: DE loop, yellow: FG loop). A linker for small molecule attachment was added to the C-terminus (orange), and the N-terminus (green) linked to display construct.

To demonstrate that this can be used for a variety of druggable targets, we chose two target classes, one soluble enzyme and one G-protein coupled receptor (GPCR), with known small molecule inhibitors cross-reactive to other proteins within the target families as our model system. These classes of proteins have very different structures and mechanisms of function and are representative of a large array of important drug targets. Many enzymes and GPCRs are known to have conserved binding pockets and active sites across isoforms[154–156], and are thereby challenging to selectively target.

Efficient strategies to increase the selectivity and potency of drugs targeting these classes is applicable to a vast array of disease treatments as well as other biotechnology applications[157,158].

3.1.3 Carbonic anhydrase targeting

Carbonic anhydrases (CA) are a class of metalloenzymes that catalyze the reversible conversion of CO_2 to bicarbonate with a coordinated metal ion[159]. There are at least 14 known CA isoforms found in humans, and many are implicated in diseases from macular edema[160–162] to cancer[57,163,164]. There is a strong need for inhibitors that are selective towards a single isoform because there is high homology between CA isoforms. CAIX has been shown to be upregulated in hypoxic tumor environments, yet inhibition of other CAs can cause serious side effects[163,165]. Targeting of CAI and CAII to reduce intraocular pressure has also seen success, though with similar issues of side effects associated with the inhibition of other CA isoforms[162,166]. There are many known inhibitors towards the CA family. One class of inhibitors are sulphonamides, which bind to the coordinated metal ion within the enzymatic pocket. Members of this class of inhibitors have been used as antiglaucoma agents, diuretics, and antiepileptics[162,167]. Though they are known to be tight binding inhibitors, they are not selective towards any particular member of the CA family[160,167]. Indeed, the binding site of sulphonamides to CAs is so well conserved across the CA family that 14 of 18 residues within the active site are identical across all but one of 11 active CA isoforms[57]. One inhibitor of particular interest is acetazolamide (AAZ), which exhibits 38 nM affinity towards CAII

and 20 nM affinity towards CAIX[168] and has been previously crystallized in complex

with both CAII (Figure 3C) and CAIX (Figure 3D). In these and additional solved

CA:AAZ complex structures, the sulphonamide binds to the complexed Zn^{2+} ion deep

within the active site in almost all solved structures[169–171]. This along with the clinical

relevance of a potential isoform-specific CA inhibitor indicating that this is an exciting

model system for demonstrating the PriSM design capabilities.
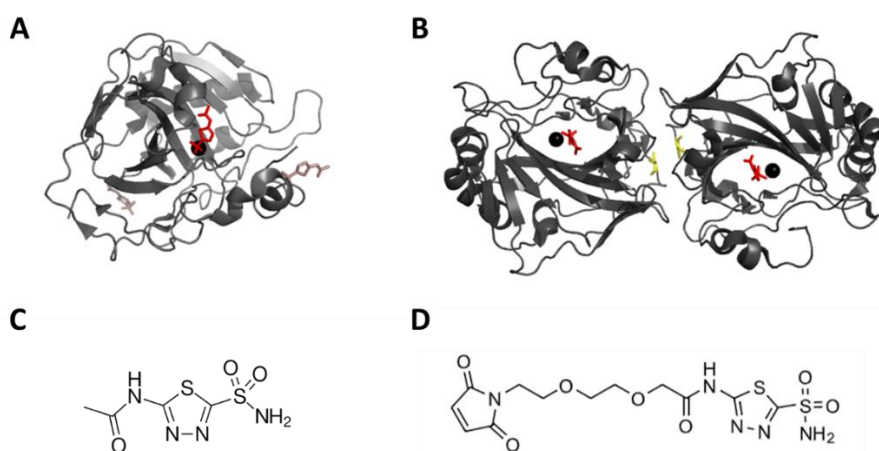
Figure 3.2: Structure of acetazolamide inhibitor with CAII and CAIX. A) CAII (PDB: 3HS4). Black: Zn^{2+} ion, Red: Acetazolamide bound to Zn^{2+} and active site residues. Pink: two additional acetazolamides in complex with regions outside the active site. B) CAIX catalytic domains (PDB: 3IAI). The CAIX catalytic domain forms a dimer that is stabilized via a disulfide bond at site 41 (yellow residues). Like CAII, acetazolamide binds to the Zn^{2+} ion and coordinating residues within the active site of CAIX.

3.1.4 GPCR targeting

GPCRs are the most abundant class of surface receptors in humans, with over 800 unique

members[157]. Many are drug targets for diseases ranging from cancer[172] to HIV[173,174]. As

ubiquitous and important as they are, they are generally difficult to engineer ligands and

36

drugs towards due to their hydrophobic transmembrane domains, lack of a separable extracellular domain, limited extracellular size, and membrane-dependent structure[147,175]. GPCRs are also challenging to produce in a soluble form that recapitulates the structure and behavior found in its native state, which has resulted in a lack of crystal structures to be used in rational drug design[176–179].

C-X-C chemokine receptor 4 (CXCR4) is a GPCR that is implicated in a vast array of pathologies, including angiogenesis[180,181] and chemotaxis in metastatic cells[182,183], HIV infection[173,184], and stem cell motility[185,186] (Figure 3.3A). With so many therapeutic applications, there is a large need for CXCR4-targeted binders and inhibitors. Adding to the difficulties of engineering proteins to a GPCR such as CXCR4, there is significant overlap of the extracellular domain between it and other CXC-motif receptors[175,187–190], making it difficult to engineer highly selective ligands or therapeutics (Figure 3.3B).
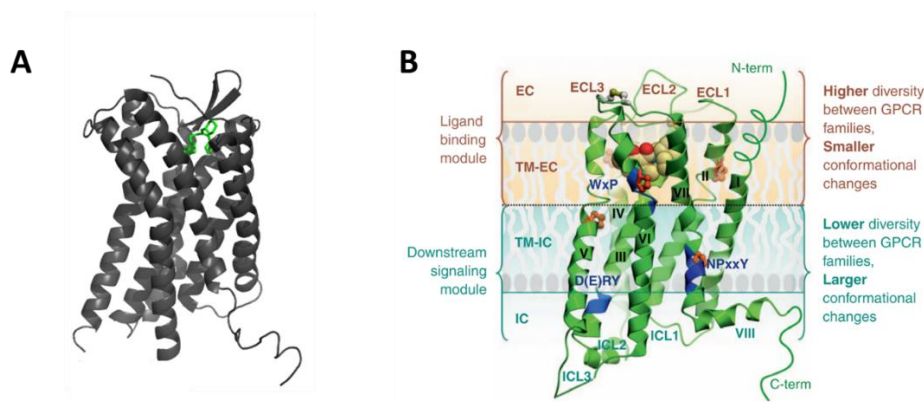
Figure 3.3: GPCR structure and binding pockets. A) Solved structure of CXCR4 in complex with the small molecule inhibitor IT1t (PDB: 3ODU). B) General GPCR structural features. Adapted from Katritch, et al[175]. The accessible surface area for GPCRs (including CXCR4) is a low percentage of the total structure but contains the highest diversity between isoforms. The ligand-binding pockets extend into the core of the protein and is a more conserved region.

3.1.5 Yeast surface display of PriSM libraries

In the current work, we generated a yeast surface display library of PriSMs to select for

functional PriSM molecules with enhanced selectivity relative to the small molecule

alone (Figure 3.4). Yeast display is a protein display format that links the genotype and

phenotype of the expressed variant[34,191]. The gene of interest is produced as a genetic

fusion with the Aga2p domain of a-agglutinin. This fusion is secreted and tethered to the

surface of each yeast cell, where it forms two disulfide bonds with the Aga1p domain.

There are approximately 10,000 copies of each variant on the surface of the yeast cell,

and the genetic information for the variant is found within the cell. For the PriSM

display format, a polypeptide linker comprising proline, alanine, and serine residues

(PAS) of either 25 or 50 amino acids was fused to the Fn3 variants (Figure 3.4B). At the

C-terminus of these linkers was a cysteine, and it is this thiol that was used to couple the

small molecule. Our initial small molecules were synthesized with a maleimide

functional group which couples specifically with free thiols, such as the ones on the

linker termini or, undesirably, other cell surface thiols such as on Aga1p (Figure 3.4C).

We have demonstrated that small molecules can be more functional when conjugated to

Fn-PAS-Cys than to random yeast surface thiols, and that various sorting methods

yielded highly specific and functional PriSM molecules.
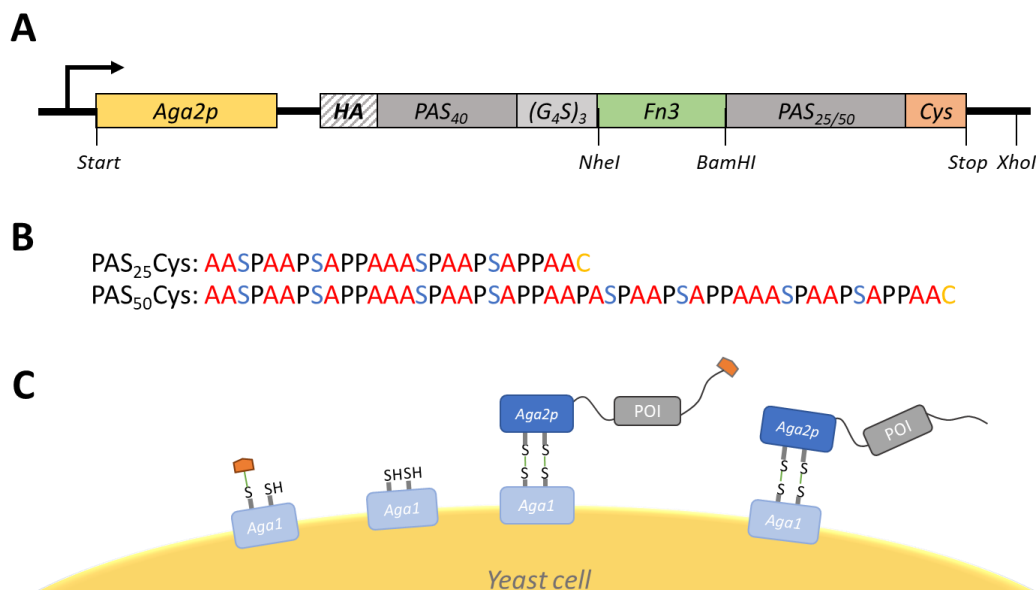
Figure 3.4: PriSM libraries design. A) Construct design for yeast display of PriSM proteins. The Aga2p protein is secreted with the display construct consisting of an expression tag (HA), PAS$_{40}$ and G$_4$S repeat spacer, the diversified Fn3 gene, and a PAS linker terminated with a cysteine. B) Sequence of PAS$_{25}$-Cys and PAS$_{50}$-Cys linkers. Linkers of proline, alanine, and serine repeats and terminated with a cysteine were added to the C-terminus of the expression construct. C) Hypothesized conjugation sites for small molecules on the surface of yeast. Small molecule (orange) can couple to any free thiol on the yeast, including any yeast surface proteins.

3.2 Materials and Methods

3.2.1 Small molecule synthesis

CA targeting molecules were synthesized by PepTech (Bedford, MA). Briefly, the inhibitor acetazolamide was synthesized with a 2-unit polyethylene oxide linker attached to and capped with the maleimide functional group for the final product of maleimide-PEO$_2$-AAZ (99.1% purity by HPLC). The CXCR4 targeting molecule Cyclo[D-Tyr-D-N-Me-Orn(maleimido-PEO7)-Arg-2-Nal-Gly] (96.1% purity by HPLC) was purchased

39

from Peptides International (Louisville, KY). Molecules were reconstituted in DMSO and stored at -20 C.

3.2.2 CA biotinylation

5 μM recombinant His-tagged human carbonic anhydrase 9 (CAIX, Sino Biological) was incubated with 500 μM EZ-link NHS-Biotin (Thermo) in PBS, pH 7.4 for 4 hours at room temperature. An average of 1.9 biotin molecules conjugated per CAIX molecule as evaluated by MALDI-TOF mass spectrometry. Carbonic anhydrase 2 (CAII, Sino Biological) was similarly biotinylated at 15 μM CAII with 135 μM NHS-biotin. Conjugation was confirmed by MALDI-TOF (1.1 biotins per CAII). Unreacted biotin was removed using two sequential Zeba 7K desalting spin columns (Thermo).

3.2.3 Cloning and library construction

Oligonucleotides for construction of the diversified Fn3 library and proline-alanine-serine (PAS) linker genes were purchased as gblocks from IDT DNA Technologies. The hydrophilic Fn3 gene was diversified as adapted from Woldring, et al [55] and is under publication review (Lewis, *et al*). PAS sequences of 25 and 50 amino acids were adapted from Schlapschy, et al[192]. The pCT-40 display vector used to display the Fn3-PAS-Cys library was adapted from Stern, et al[120]. Fn3 library oligonucleotides were assembled into a pool of full-length Fn3 genes via overlap extension PCR. Full-length Fn3 genes, PAS$_x$ linkers, and pCT-40 vector were transformed into the EBY100 strain of *S.*

cerevisiae and combined via homologous recombination. Transformation efficiency was quantified via plate counting of diluted library transformants on selective SD-CAA plates. Assembly fidelity was verified via Sanger sequencing. The PAS_{25} and PAS_{50} libraries were then pooled in an equimolar fashion for the remainder of the study.

3.2.4 Yeast preparation and conjugation to small molecule

Transformed yeast were cultured by shaking at 30 °C in SD-CAA media (16.8 g/L sodium citrate dihydrate, 3.9 g/L citric acid, 20.0 g/L dextrose, 6.7 g/L yeast nitrogen base, 5.0 g/L casamino acids), and expression of Fn3-PAS-Cys was induced for >4 hours at 30 ºC with shaking by switching the yeast into SG-CAA media (10.2 g/L sodium phosphate dibasic heptahydrate, 8.6 g/L sodium phosphate monobasic monohydrate, 19.0 g/L galactose, 1.0 g/L dextrose, 6.7 g/L yeast nitrogen base, 5.0 g/L casamino acids).

Induced yeast were washed in PBS (pH 7) and resuspended in PBS pH 7 with 100 µM small molecule. Yeast were incubated at room temperature for > 1 hour, and subsequently washed twice in PBSA. To verify conjugation, a small aliquot of yeast both with and without small molecule were washed twice in PBS and then incubated with 2 µM AlexaFluor590-PEO_5-Maleimide for 30 minutes, washed twice in PBSA and analyzed via flow cytometry.

3.2.5 Magnetic bead selection

Yeast, sufficient in number to provide 10-fold coverage of the population diversity, of each conjugated Fn3-PAS25Cys and Fn3-PAS50Cys library were pooled and incubated with 10 µL streptavidin-coated Dynabeads for 2 hours while rotating. Dynabeads and bound yeast were removed by placing samples against a magnet for 5 minutes and recovering unbound yeast. To the unbound yeast, another 10 µL of Dynabeads were added and the incubation and recovery process was repeated. The recovered unbound yeast were incubated with either 5 nM biotin-CAII or 5 nM biotin-CAIX for 1 hour, washed once in PBSA, and then incubated with 10 µL beads for 2 hours. Bead-bound yeast were recovered and incubated in 100 mL SD-CAA.

3.2.6 Fluorescence-activated cell sorting (FACS) affinity sort

Recovered yeast (>20-fold coverage) enriched for binding CAII or CAIX were separately induced and conjugated to small molecule per the protocol above. Yeast were then incubated with 50 nM of their respective targets (either CAII or CAIX) for 1 hour at room temperature, followed by mouse anti-HA-DyLight650 and streptavidin-AlexaFluor647. Cells that showed target binding above background were collected and cultured in 5 mL SD-CAA.

3.2.7 FACS specificity sort

Recovered yeast (>20-fold coverage) from the above affinity sort were separately induced and conjugated to small molecule per the protocol above. Each population was

incubated with both CAII-biotin and CAIX-DL650, followed by streptavidin-AlexaFluor647 and mouse anti-HA-DyLight550. Cells that showed either strong binding to both targets or binding only to their initially sorted targets were separately collected. These collected populations were cultured and induced, and their binding to CAII-biotin and CAIX-biotin with or without small molecule conjugation was then analyzed via flow cytometry via above protocols.

3.2.8 Construction of PAS_0 populations

The Zymoprep plasmid extraction kit (Zymo Research, Orange County, CA) was used to isolate pCT40-Fn3-PASx-Cys plasmids from the triply sorted yeast populations. These plasmid pools were amplified via PCR using primers flanking the Fn3 genes and transformed into yeast with the traditional pCT-40 plasmid via homologous recombination. Transformation and expression in yeast was verified using flow cytometry with HA and cMyc labeling.

3.2.9 Clone sequencing and production

Plasmids from zymoprepped yeast populations were transformed into NEB5α bacterial cells and cultured in lysogeny broth with 100 µg/mL ampicillin. Amplified DNA was then extracted from clonal bacterial cultures via miniprep kit (Epoch Life Sciences, Houston, TX). Clonal plasmid isolates were then submitted for Sanger sequencing.

3.2.10 Cloning into production vector

To leave the C-terminal cysteine and conjugated small molecule unblocked by a purification tag, the T7 bacterial expression vector pET-22b was modified to include the His$_6$ tag at the N-terminus of the expressed protein. This new plasmid, pETn, was then digested using NheI and XhoI enzymes. Lead clone genes (Fn3 + PAS tail) were amplified via PCR from miniprepped isolates. Each gene was combined with digested pETn vector via HiFi DNA Assembly Master Mix (New England Biolabs) at the following ratios: 1 µL gene PCR product, 1 µL digested pETn product, and 2 µL HiFi Master mix. The reaction was incubated for 1 hour at 50 °C and subsequently transformed into T7 Express competent *E. coli* (New England Biolabs) and streaked on lysogeny broth with agar plates containing 100 µg/mL kanamycin. Colonies were cultured and sequenced as previously described.

3.2.11 Protein-small molecule production and purification

Lead clones were produced in *E. coli* using the T7 Express cells from NEB. *E. coli* were grown in 100 mL cultures to an OD between 0.5 and 1, and production was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 37 °C for two hours. Cultures were then pelleted, and cells resuspended in a bacterial lysis buffer (50 mM sodium phosphate (pH 8.0), 0.5 M sodium chloride, 5% glycerol, 5 mM CHAPS, and 25 mM imidazole, cOmplete Mini EDTA-free Protease Inhibitor Tablet). Cells were lysed

44

via five or more freeze-thaw cycles, and the insoluble fraction was removed via high speed centrifugation and filtered (0.22 µm filter paper).

Clones were purified and conjugated to small molecule via a modified metal affinity chromatography protocol. Briefly, cell lysate was applied to a HisPur Cobalt spin column (ThermoFisher Scientific), equilibrated with 15 mM imidazole + 50 mM sodium phosphate (pH 8.0), and centrifuged at 700g for 2 minutes. The column was then washed three times with 400 µL of 15 mM imidazole to remove protein lacking the His$_6$ purification tag, followed by six washes with 50 mM sodium phosphate (PBS, pH ~7.5). The column was then incubated with 200 µL of 200 µM small molecule for 30 minutes at room temperature and washed four times with 15 mM imidazole + PBS. His$_6$-tagged protein was eluted from the column with 300 mM imidazole + PBS. Imidazole was removed from the elution via PD-10 desalting column. Protein conjugation was verified via MALDI, and concentration measured via A280 quantification.

3.2.12 CA activity assay

CA activity was measured via hydrolysis of 4-nitrophenyl acetate (4-NPA, Sigma). Briefly, conjugated clones were diluted to the desired concentrations in 10 mM HEPES buffer (pH 7.3) and combined with either CAII at 25 nM or CAIX at 75 nM and allowed to equilibrate at room temperature. The substrate 4-NPA was diluted in HEPES buffer and added to the mixture. Absorbance was measured at 348 nm and 400 nm every

minute on an H1-Synergy plate reader. Reaction velocity was calculated from the slope of absorbance at 400 nm per time for the first 30 minutes of the reaction.

3.2.13 Cell lines and cell culture methods

HEK-293T cells were a kind gift from Dr. Douglas Yee (UMN). Cells were cultured in DMEM with 10% v/v fetal bovine serum and 1% v/v penicillin and streptomycin (Gibco). Cells were grown at 37 °C and 5% CO_2. Cells were grown to 80-90% confluency and passaged at 20-fold dilution every 3-4 days.

3.2.14 Generation of stable cell lines expressing CXCR4 gene

Lentiviral transfer vector pLX304 containing the full-length CXCR4 gene with C-terminal V5 epitope tag was purchased from DNASU. Packaging and envelope vectors were also a kind gift from Dr. Douglas Yee. HEK-293T cells were used to produce in-house lentiviral particles. Briefly, ~5x10^6 HEK-293T cells were transfected with 3 plasmids; vsv-g, dvpv, and a pLX304-CXCR4-V5 using Lipofectamine 3000 (Invitrogen) per manufacture protocol. Cells were allowed to produce particles for 48 hours, and media harvested and stored at -20 C. Separate batches of HEK-293T cells were then transduced with 1 mL lentiviral particle suspension in media and 8 μg/mL polybrene. Cells were then maintained under selection pressure of 5 μg/mL blasticidin. Transduced HEK-293T cells were then subjected to FACS sort to select for the top 5% of CXCR4-expressing clones. These sorted populations were used for sorting experiments.

46

Analogous strategies were used to develop CEACAM5 and CXCR7-expressing HEK-293T cell lines.

3.2.15 Expression quantification

Cells to be tested were washed twice in cold 1% PBSA (10 g/L bovine serum albumin, PBS). 50,000 cells were resuspended in 50 µL PBSA with 10 µg/mL mouse anti-target antibody and incubated for 30 minutes at 4 °C. Cells were washed once in PBSA and incubated in 20 µL of 10µg/mL goat anti-mouse IgG – FITC for 5 minutes at 4 °C. Cells were then washed in PBSA and analyzed via flow cytometry. In parallel, Quantum™ Simply Cellular® anti-Mouse IgG beads (Bangs Labs, Fishers, IN) were incubated with 10 µg/mL mouse IgG isotype control antibody (Abcam, Cambridge, MA). Beads were washed and labeled in an identical fashion to cells being tested and ran with samples on flow cytometry.

3.2.16 CXCR4-SM naïve library enrichment via whole yeast-cell panning

HEK-293T cells were passaged as previously discussed and plated at 500,000 cells/well in 6-well plates previously treated with 0.1 % (w/v) poly-L-lysine solution (Sigma) per manufacturers protocol. The following day, plated cells were washed 3 times by gently adding 1 mL cold PBSACM (1x PBS, 1 g/L bovine serum albumin, 1 mM $CaCl_2$, 0.5 mM Mg_2SO_4) to the side of the well and gently rocking the plate before aspirating the PBSACM wash. 4x diversity of the naïve yeast Fn3-PAS$_x$ library were

47

conjugated with Maleimide-PEO$_7$-CXCR4 as previously discussed and resuspended to a concentration of 10^8 yeast/mL in PBSACM. 1 mL of the library was added to each well and allowed to incubate at 4 °C for >15 minutes. After that time, unbound yeast were aspirated, and wells were washed 5x with cold PBSACM. Wells were then scraped to remove HEK-293T-CXCR4+ cells and bound yeast and recovered cells were suspended and pooled together in SD-CAA growth medium. Yields were quantified by dilution plating on YPD agar plates. For the second and third rounds of sorting, 10^8 yeast were added to each well in triplicate against HEK-293T-CXCR4+ or HEK-293T-CEA+ cells and yields for each well were quantified separately by dilution plating. Error bars represent standard error in plate counts.

3.2.17 Clonal yeast panning

The triply-sorted CXCR4 yeast population was plated on SD-CAA agar plates and incubated ~3 days until yeast colonies were seen. Individual colonies were then picked and cultured in 1 mL SG-CAA overnight at 30 °C to induce protein expression. One day prior to panning, 96-well culture plates were treated with poly-L-lysine solution as previously mentioned. 50,000 cells were added to each well and allowed to adhere overnight. On the day of panning, clonal yeast cultures were washed twice in PBS (pH 7.5) and incubated in PBS with 100 µM mal-PEO$_7$-CXCR4 small molecule for 30 minutes at room temperature. Yeast were washed 3 times with PBSACM and resuspended in 250 µL PBSACM. To each well, 50 µL yeast plus 150 µL were added to HEK-293T cells and incubated at 4 °C for >15 minutes. 96-well plates were then gently

washed 3x with cold PBSACM. 100 µL of PBSACM was added to each well, and wells were imaged using EVOS FL Cell Imaging System (Thermo Fisher Scientific) at 40x magnification.

3.3: Results

3.3.1 AAZ-conjugated library characterization

We hypothesized that a fibronectin – linker – small molecule fusion would provide an effective molecular design for the PriSM concept (Figure 3.4). Two libraries were constructed containing a fibronectin domain with three solvent exposed loops with a sitewise gradient of diversity as a protein binding scaffold[153,193,194] and a C-terminal polypeptide chain consisting of either 25 or 50 amino acids and capped with a C-terminal cysteine (Figure 3.4A,B). This polypeptide chain consisted of proline, alanine, and serine amino acids arranged to mimic the properties of a polyethylene glycol chain of similar dimensions[192,195]. 4×10^8 total transformants were obtained. Sequencing indicates that 66% of the library matched the intended design.

The non-specific CA inhibitor AAZ was synthesized with a PEO_2 spacer and maleimide functional group to facilitate conjugation to cysteines on yeast cells under physiological/mild conditions (Figure 3.5). Yeast displaying the newly constructed library were then conjugated via thiol-maleimide chemistry to maleimide-PEO_2-AAZ.

300 nM Target   **50 nM Target**

CAII-
binding

CAIX-
binding

CA binding

Fn3/yeast
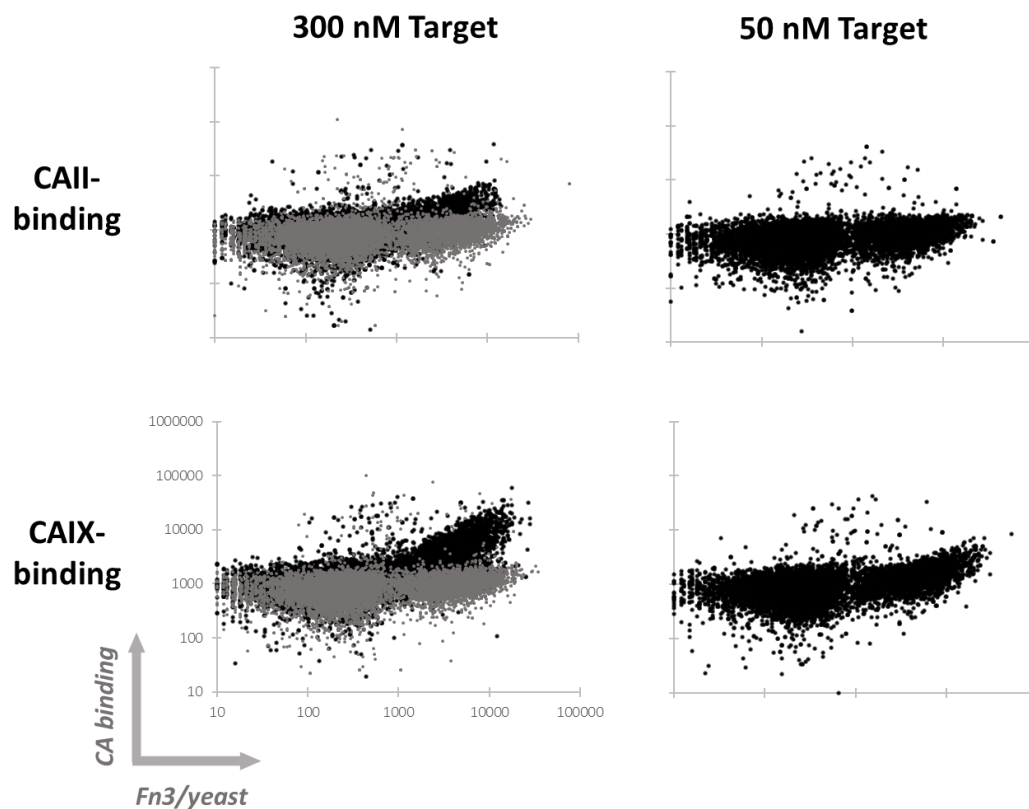
Figure 3.5: Unsorted Fn-PAS$_x$-Cys+AAZ variants bind CA. A) Structure of acetazolamide (AAZ). The sulphonamide binds to the Zn2+ ion in the active site of carbonic anhydrase. B) Structure of maleimide-(PEO)2-AAZ. The added maleimide functional group forms a stable thioether bond with free cysteines. The two-unit polyethylene oxide (PEO)$_2$ allows for spacing and flexibility for AAZ to bind the CA active site. C) Conjugated (black) and unconjugated (gray) Fn3-PAS$_x$-Cys libraries were incubated with 300 nM of either CAII or CAIX and anti-HA tag antibody, followed by secondary fluorophore conjugated antibodies. For the unconjugated library, no binding was seen to either target. For the AAZ-conjugated libraries, binding to CAII and CAIX was seen on HA-positive yeast.

The naïve library with and without AAZ conjugation – and with and without Fn-PAS-Cys display – was incubated with 300 nM CAII or CAIX. The naïve library showed binding to CAII and CAIX when conjugated to AAZ, but not without (Figure 3.5), and no binding was detected against CAII or CAIX in non-displaying yeast. This shows that while the AAZ can conjugate to any free cysteine on the yeast surface, there is either preferential conjugation or binding to Fn3-PAS-Cys-AAZ. Given the high affinity (>30 nM) of free AAZ, the lack of binding to non-displaying yeast is consistent with either drastically reduced affinity in the yeast-PEO-AAZ context or a lack of conjugation to non-Fn sites. To differentiate these hypotheses, we evaluated binding of Fn-PAS-Cys-AAZ to reduced concentrations of CA, which revealed binding only for rare clones (Figure 3.5), thereby supporting the hypothesis that small molecule affinity is reduced by the linker context. Yet, this may be ideal as it seemingly still targets the CA active site (based on the majority of clones binding at 300 nM, which should be driven by AAZ rather than the diverse array of Fn sequences) while providing opportunity for substantial avidity via the bivalent PriSM concept (Figure 3.1). Collectively, this shows that yeast display of protein—small molecule conjugates can be used to sort for binders. Also, while there is slight preferential binding of the majority of the library to CAIX over CAII, the selectivity is minimal, which agrees with previous AAZ experiments[129,164,168,169,196] and provides the opportunity for particular Fn sequences to provide selectivity.

3.3.2 Isolation and characterization of CAII- and CAIX-specific PriSM molecules

To identify PriSM molecules that were highly specific to either CAII or CAIX, the populations underwent several rounds of sorting (Figure 3.6A). The AAZ-conjugated library was sorted for binders to CAII or CAIX in parallel. For the initial sort, the conjugated library was labeled at 5 nM target, and binders were recovered with multivalent magnetic bead capture to remove most of the non-functional clones and enrich for ligands. The resultant populations were sorted via flow cytometry with 50 nM target. Each subsequent population showed enrichment for their respective targets that was AAZ-dependent (Figure 3.6B). However, because no isoform-selection pressure had been added, the twice-sorted populations contained clones with a wide range of specificity to the appropriate CA isoform. To push the selectivity towards either CAII or CAIX, a 3-color competitive FACS sort was done with 5 nM target and 5 nM off-target labeling on each population (Figure 3.6C). Yeast were collected that showed binding to either CAII or CAIX, while showing no binding to the alternative target.
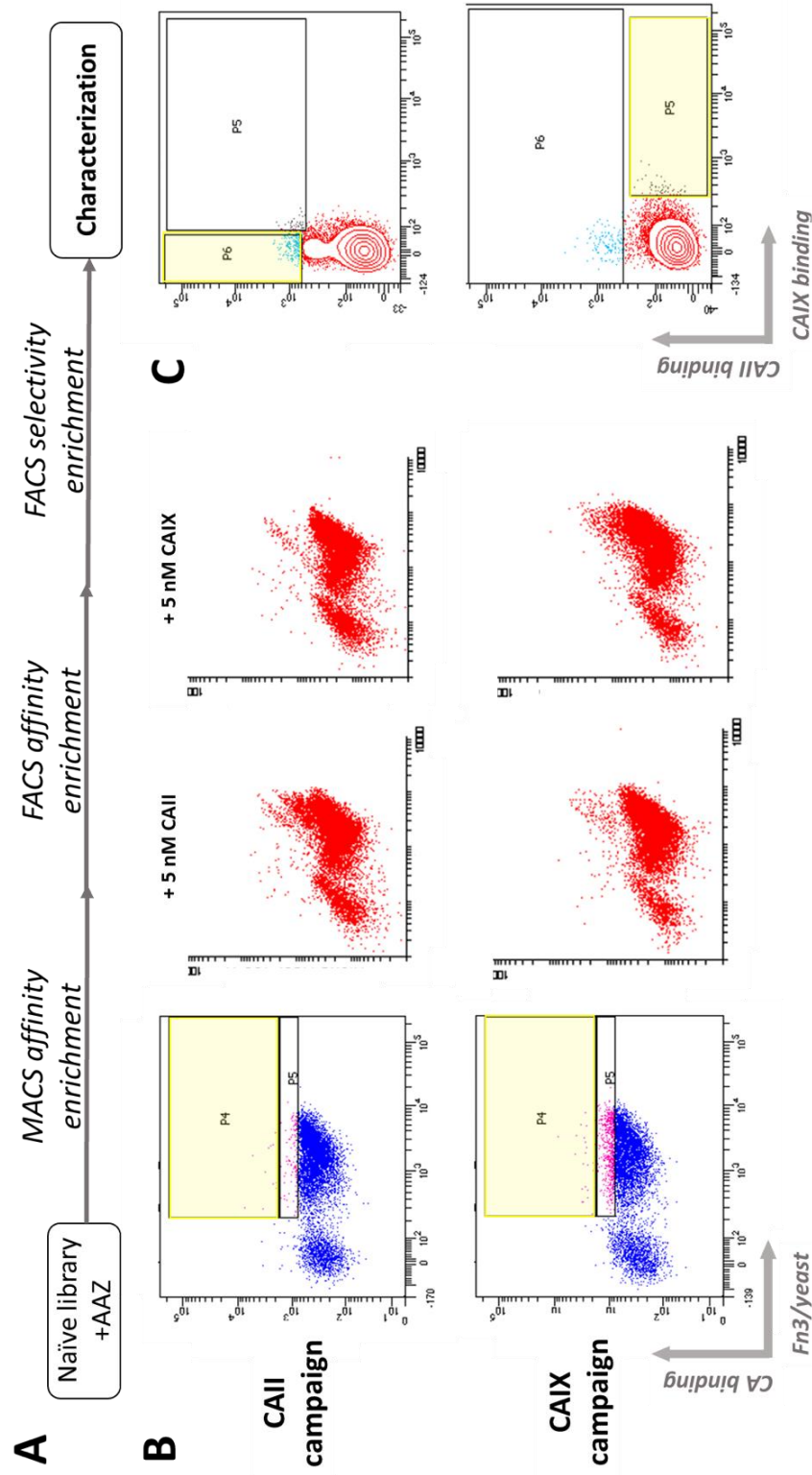
Figure 3.6: Sorting strategy for CA-isoform-selective PriSM molecules. A) Sorting flowchart. Naïve library was conjugated with Maleimide-(PEO)$_2$-AAZ and sorted in parallel for CAII and CAIX selective ligands three times. B) Second affinity sort and collected population selectivity. The CAII-binding and CAIX binding campaigns were first enriched for target-binding ligands via MACS and followed by a FACS enrichment at 50 nM target. Yellow gates were collected and analyzed for binding to CAII and CAIX. Each population contained isoform-specific ligands in addition to nonspecific ligands. C) Populations were then sorted via labeling with each target and collecting target-binding ligands that showed no binding to the off-target isoform. Yellow gates were collected.

53

After three rounds of sorting, populations were found to be of high affinity and high specificity to the target CA of interest (Figure 3.7). Both triply-sorted populations, either towards CAII or CAIX, demonstrated binding at concentrations as low at 0.5 nM. A subset of the CAII-sorted population appears to bind streptavidin, as is evident from the double-positive yeast when labeled with streptavidin-AlexaFluor488 only. Otherwise the CAII-sorted population shows very strong binding of all clones at 5 and 50 nM CAII. Similarly, a majority of the CAIX-sorted population shows binding at 0.5 nM CAIX, and no non-binding ligands seen at 5 nM or 50 nM. In addition to containing strong ligands, a majority of both populations are also highly specific to the isoform of interest. Aside from potential avidin-binding ligands, both CAII-sorted and CAIX-sorted populations show no binding above background. These results indicate that the initial library and sorting strategies can yield populations with high affinity and selectivity towards a specific enzyme isoform.

Analysis of non-conjugated Fn-PAS-Cys indicates that binding for both populations was AAZ-dependent for nearly all clones (Figure 3.8). This indicates the importance of AAZ in binding and suggests that while the sorting strategy is based on affinity alone, the inhibitory function has likely been preserved. Non-displaying cells lacked binding, which is consistent with Fn-dependent binding (Figure 3.8).
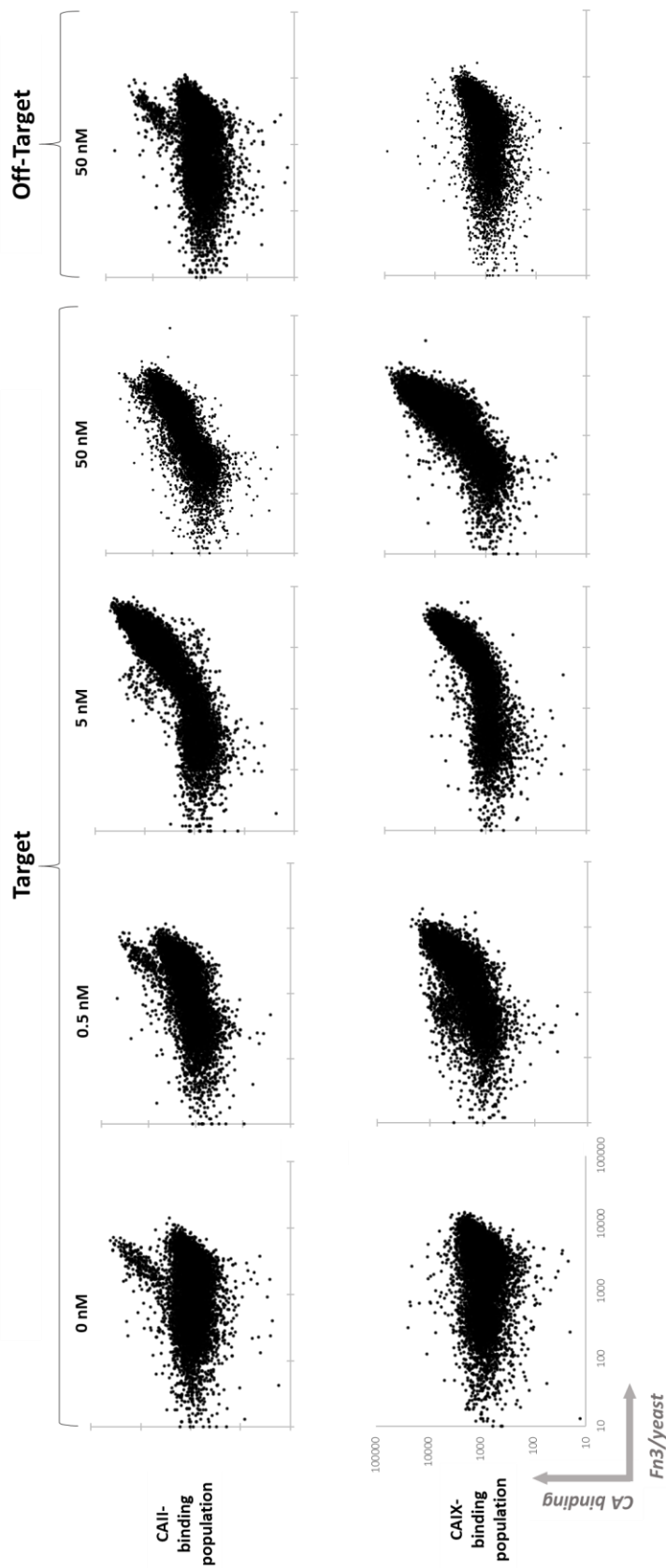
Figure 3.7: Triply-sorted populations show strong binding to the appropriate CA isoform. Maleimide-(PEO)$_2$-AAZ conjugated yeast populations were labeled for Fn3 expression and with increasing concentrations of target (CAII for the CAII sorted population, CAIX for the CAIX population).

55
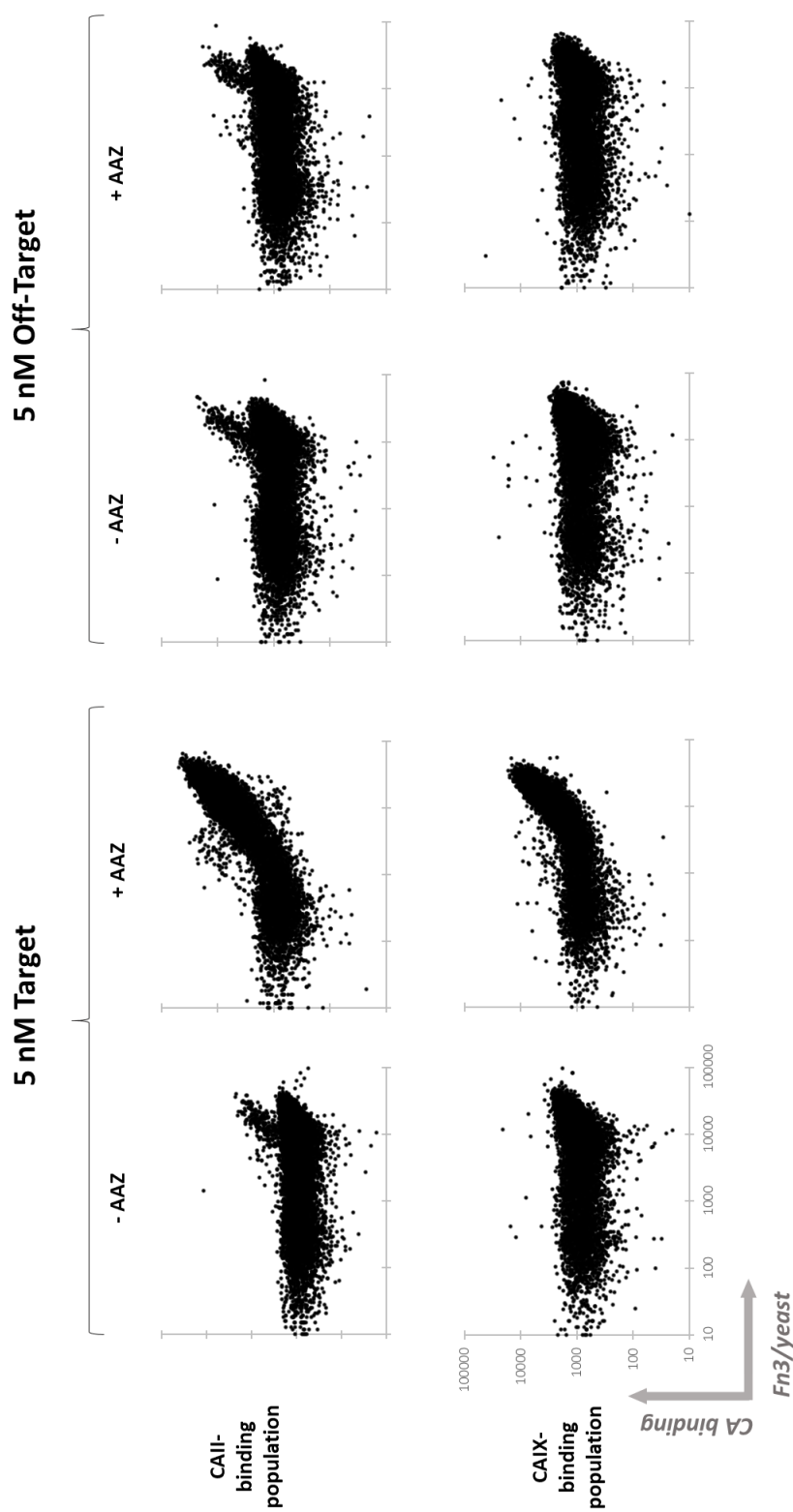
Figure 3.8: Triply-sorted populations show AAZ-dependent binding to the appropriate CA isoform. Unconjugated and Maleimide-(PEO)$_2$-AAZ conjugated yeast populations were labeled for Fn3 expression and with 5 nM of either CAII or CAIX. Non-avidin binding was only seen in the conjugated libraries labeled with the target isoform.

3.3.3 Clonal analysis of triply-sorted populations

Sequencing of the enriched populations revealed that the location of AAZ conjugation may not have only been on the intended C-terminal cysteine, but within the diversified paratope. All 11 unique clones from the CAII-sorted population contained a single additional cysteine in one of the diversified loops (Table 3.1). Eight of 11 clones had Cys at site 26 in the BC loop, two clones had Cys78, and one had Cys30. Conversely, five of six CAIX-binding clones were Cys-free within the Fn gene (Table 3.2). All but one clone exhibited strong isoform specificity (Figure 3.9, 3.10), consistent with the population analysis. Interestingly, the only non-specific clone tested from either population was the FnαCAIX-4, which was also the only CAIX-sorted clone to contain a loop cysteine (Table 3.2, Figure 3.10). Each population contained PAS-Cys linker lengths of 25 and 50 amino acids, suggesting both linkers are functional for creating a CA isoform-specific PriSM molecule. Based on these results, further characterization was needed to determine the mechanism by which these populations demonstrated AAZ-dependent specificity.

Table 3.1: CAII-binding population sequences

Name	BC	DE	FG	PAS length	Framework
FnαCAII-1	DYHCV-ALY	GY-TT	TDYNVH---SN	25	
FnαCAII-2	DYHCL-AYS	GY-TY	TDYGYD---SN	25	
FnαCAII-3	DYHCL-VYS	GY-TY	TDFGGY---SN	25	
FnαCAII-4	DYHCI-ALY	GT-TT	TTSNTQ---SN	25	
FnαCAII-5	DYPCI-ALY	GY-TN	TDVGHH---SN	25	
FnαCAII-6	DYHYL-ALY	GN-TY	TDCDYE---SN	25	L18R
FnαCAII-7	DYSCL-ALY	GN-TY	STENYY---SN	50	A57V
FnαCAII-8	DDYPHGACY	GTSYT	ADWSDDPY-SN	50	
FnαCAII-9	DYSCL-ALY	GN-TY	TTENYY---SN	50	
FnαCAII-10	DYHLL-ALY	GY-TS	TDCDSD---SN	50	
FnαCAII-11	DYHCL-VSS	GT-TS	TDFNGY---SN	25	
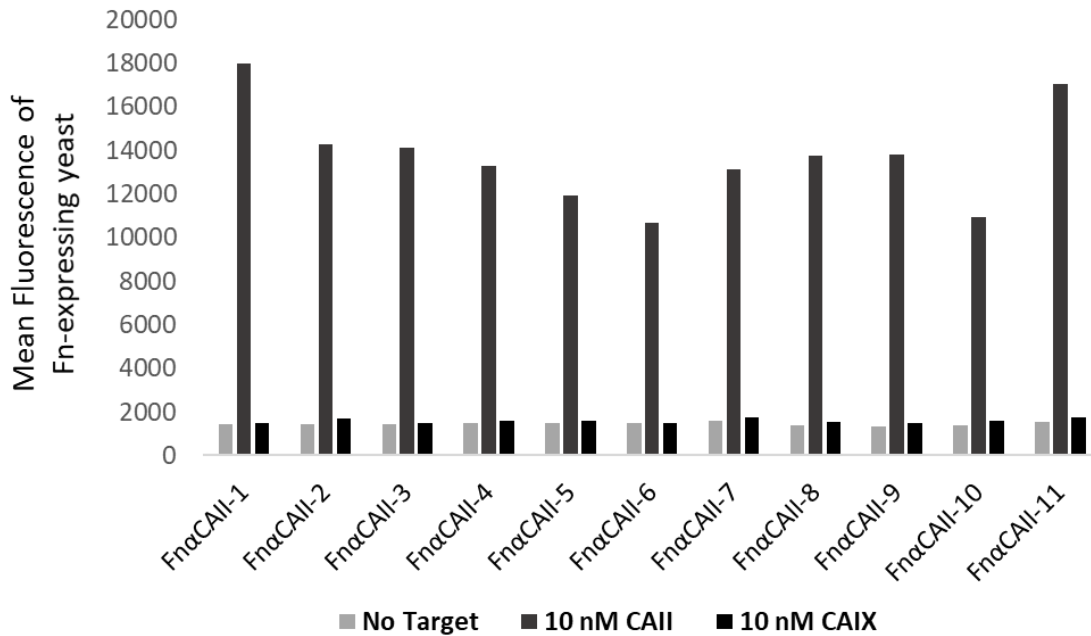
Figure 3.9: Clones from triply-sorted CAII populations show strong affinity and specificity for CAII. Each clone was incubated with 0 nM, 10 nM CAII, or 10 nM CAIX. Median fluorescence of yeast expressing the Fn-PAS$_x$-Cys construct were used to quantify target binding.

Table 3.2: CAIX-binding population sequences

Name	BC	DE	FG	Tail length	Framework
FnαCAIX-1	DYHHW--VRS	GYTNS	TGRNLRN--SN	50	
FnαCAIX-2	DYHRY--VIS	GYTNS	TSWSYFE--SN	25	
FnαCAIX-3	DYHRFYYAIS	GYYSS	TNYDSRG--SN	25	
FnαCAIX-4	DYSYN--ALY	GNTNN	AGFGWVC--SN	50	
FnαCAIX-5	DYHWY--VHS	GYYNT	TNRSYYPY-SN	25	
FnαCAIX-6	DAHVYHRASS	GYTTS	AGFGYSW--SN	25	
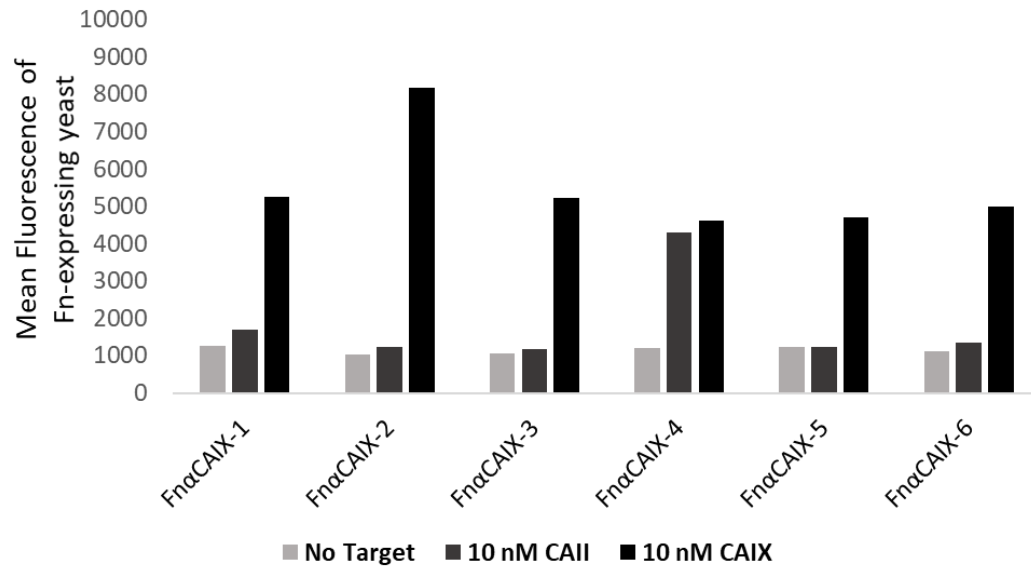
Figure 3.10: Clones from triply-sorted CAIX populations show strong affinity and specificity for CAIX. Each clone was incubated with 0 nM, 10 nM CAII, or 10 nM CAIX. Median fluorescence of yeast expressing the Fn-PAS$_x$-Cys construct were used to quantify target binding. FnαCAIX-4 was the only analyzed clone to bind to CAII at 10 nM.

3.3.4 Determination of PAS-Cys tail importance in PriSM populations

The triply-sorted population sequences suggest that the loop cysteines were integral to the binding of several clones within each population, which may emerge from AAZ conjugation or thiol-based binding. To identify whether the C-terminal cysteine, and its potential AAZ conjugation, played a role in CA binding, Fn sequences from the triply-sorted populations were moved into an expression vector lacking any C-terminal PAS_x-Cys linker to remove the site for conjugation. Binding analysis of the resulting population revealed both AAZ-dependent and independent binding clones (Figure 3.11). For both populations, removal of PAS_x-Cys resulted in a majority of clones losing all detectable binding to CA isoforms. This result is consistent with the clones benefiting from the C-terminal PAS_x-Cys-AAZ as designed, although some Fn-only binding would be expected to provide the specificity-driving avidity. Without AAZ conjugation, the PAS_x-Cys-less CAII-sorted population showed no binding at 50 nM CAII or CAIX, which is consistent with their inability to bind without AAZ even with PAS_x-Cys. In contrast, multiple PAS_x-Cys-less clones from the CAIX-sorted population bound CAIX without AAZ, which indicates that the Fn half of some PriSM clones are strong standalone binders (though likely lacking inhibitory activity). Notably, several clones in both target pools showed strong affinity when conjugated to AAZ, which is consistent with a more local, integrated protein-small molecule paratope. Such a paratope is achievable with the observed loop cysteines, which are prevalent in the CAII population but not the CAIX population. Additionally, almost all CAII- and CAIX-binding clones were specific although there was limited off-target binding for some AAZ-conjugated clones.

60

From this data, we can see that there are multiple mechanisms of conjugation that can result in high affinity, isoform-specific fusion molecules. To isolate dual-paratope binding PriSM molecules, the original triply-sorted populations were further sorted without conjugating AAZ (Figure 3.12A). In sorting without AAZ conjugation, the true PriSM molecules that can bind a site outside of the active site (where it is hypothesized all conjugated AAZ are binding) should be enriched. The mode of sorting was dictated by the average population affinity without AAZ. The CAIX-binding population, containing FACS-detectable ligands, was sorted monovalent FACS, and the much weaker affinity CAII-binding population that lacked FACS-detectable binding was sorted via highly-avid MACS.
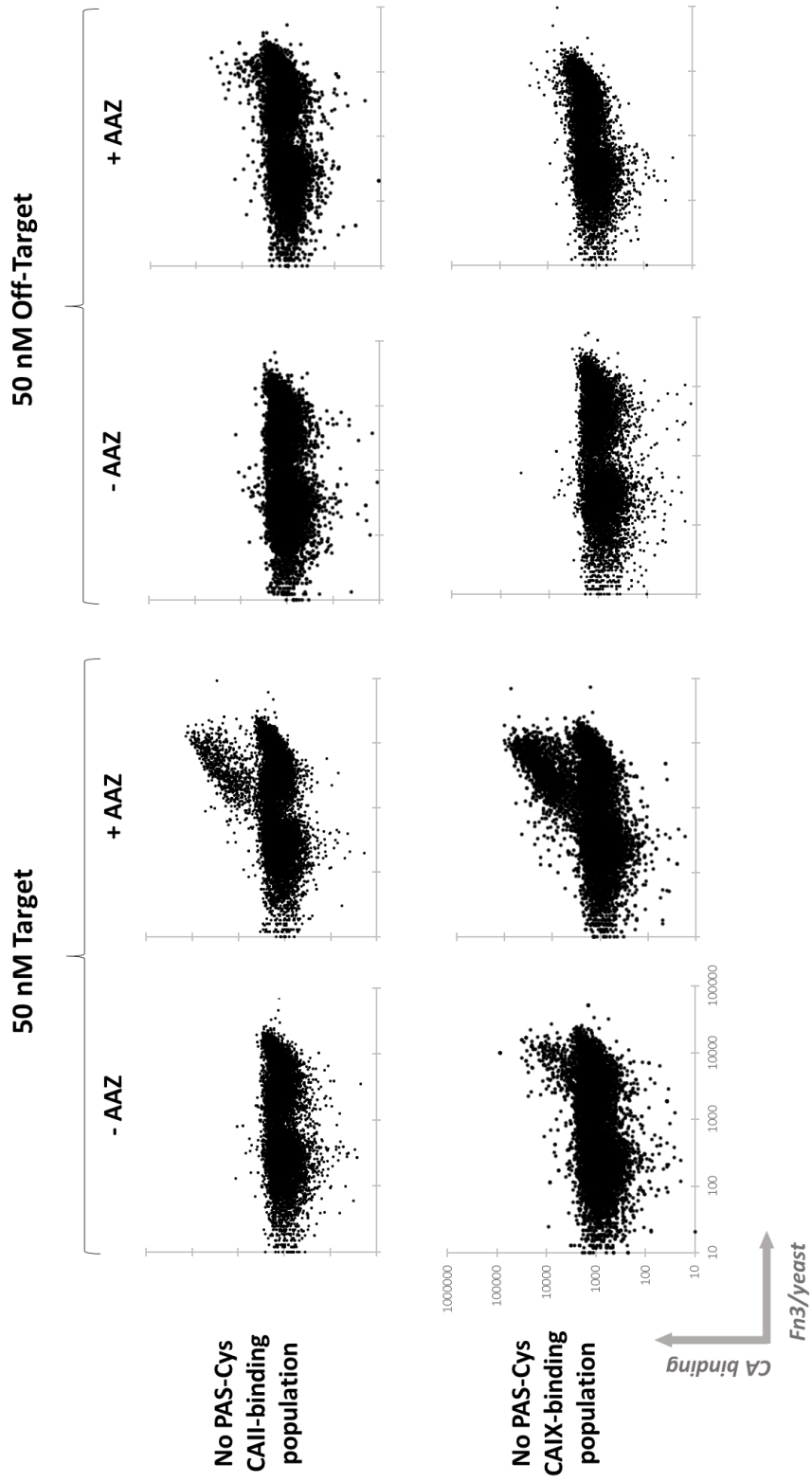
Figure 3.11: 'Tail-less' populations demonstrate diverse phenotypes. The CAII-binding population lacking the PAS$_x$-Cys tail show no binding to either CA isoform without AAZ. When conjugated to AAZ, a small fraction of Fn-expressing yeast bind to CAII, indicating that there are clones that have an AAZ-integrated paratope. The CAIX-binding population contains Fn clones that bind to CAIX without AAZ, as well as a large number of clones that require AAZ to bind, similar to the CAII-binding population. No off-target binding was seen with or without AAZ conjugation.

62
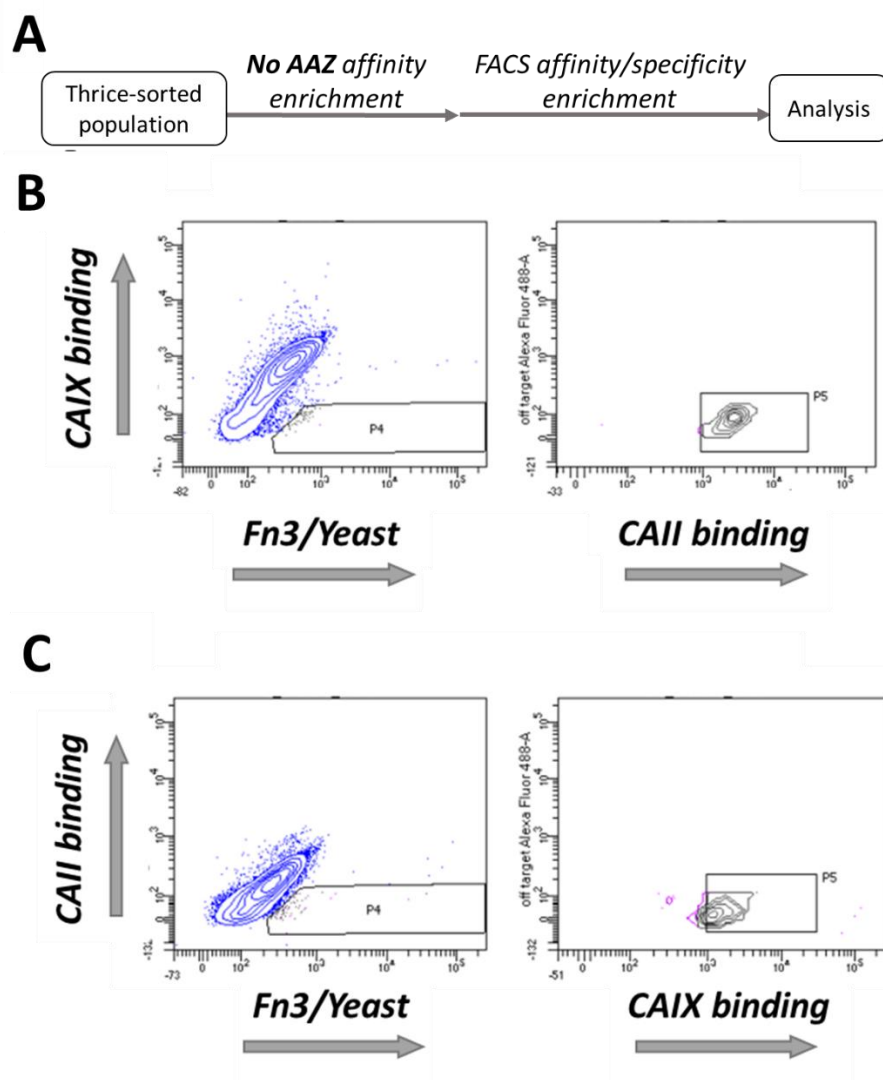
Figure 3.12: Sorting strategy for isolation of ultra-selective PriSM ligands. A) To isolate PriSMs with high-affinity protein ligand partners, thrice-sorted CAII and CAIX populations were sorted a fourth time without AAZ. Collected clones were then conjugated to AAZ and sorted with 0.2 nM target CA and 100 nM off-target CA to isolate the highest selectivity and affinity PriSM constructs.

To ensure that binding PriSMs were recovered, each population was then subjected to a high stringency three-color sort, labeling the target CA at 0.2 nM target CA and 100 nM non-target CA (Figure 3.12). The resulting population was then sequenced. For the CAIX binding population, two unique sequences of five total were identified (Table 3.3). Notably, clone FnαCAIX-5.2 is only two mutations away from clone FnαCAIX-5 (from the triply-sorted population), with identical BC and FG loop sequences. In the 5x-sorted CAII-binding population, three unique sequences (of five total) were found: FnαCAII-1, and two new sequences (Table 3.3). Notably, even with the MACS sort for binding in the absence of AAZ, the CAII sequences still contained cysteines within the loop structures. This suggests that the AAZ/Fn integrated PriSM ligands still retain CAII-specific affinity without AAZ conjugation.

Collectively, these results indicate that the yeast-displayed PriSM concept is a robust and efficient method of identifying isoform-specific binding ligands. The isolated clones also demonstrated additional mechanism of selectivity that can be further explored. Clones from the 5x-sorted and triply-sorted CAII and CAIX populations were selected for off-yeast characterization.

Table 3.3: 5x-sorted CAII and CAIX population sequences

Name	BC	DE	FG	Tail length	Framework
FnαCAII-5.1	DYHCV--ALY	GY-TT	TDYNVH---SN	25	–
FnαCAII-5.2	DYHCL--AYS	GY-TY	TAYGYD---SN	25	–
FnαCAII-5.3	DAHYDADAIS	GYSTY	AGDDCVT--SN	50	–
FnαCAIX-5.1	DYHWY--VHS	GY-YN	TNRSYYPY-SN	25	T57S
FnαCAIX-5.2	DYHIYR-AVY	GYYSN	TSTGHYIG-SN	25	–

3.3.5 Panning of CXCR4-SM conjugated yeast show enrichment.

In addition to developing PriSM molecules for CA isoform targeting, we also wished to determine whether this technique could be used to develop selective GPCR-targeted PriSM molecules, more specifically CXCR4-specific ligands. Of the available small molecule structures that block CXCR4's main ligand SDF-1 (CXCL-12) from binding and inducing signaling, we chose FC131-variant 9 (Figure 3.13A). This molecule was the only ligand that contained an extended residue or functional group that could allow for attachment of a PEO-maleimide linker without greatly disrupting the binding of the inhibitor. With this in mind, we had the SM4 molecule synthesized (Figure 3.13B)
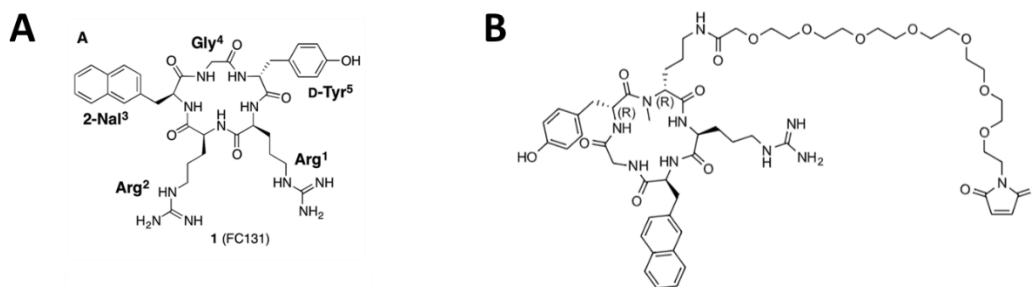
Figure 3.13 Small molecule inhibitor of CXCR4. A) CXCR inhibitor FC131-variant 9[197]. Small molecule engineered to bind CXCR4 and inhibit signaling via the SDF-1 ligand. B) SM4 molecule for conjugation. Like Maleimide-(PEO)2-AAZ the FC131-variant 9 Arg1 residue was modified to include a 7-unit PEO linker and maleimide functional group for conjugation.

Due to the complexity of the CXCR4 structure and difficulty recapitulating the binding site with soluble target or cell lysate, the SM4-conjugated Fn3 library was sorted via cell panning methods previously developed in our lab[120]. To sort for CXCR4 binding ligands via cell panning, adherent cell lines expressing the target and a related isoform at levels in excess of 100,000 molecules per cell are required. Since native expression of CXCR4 and the chosen related isoform CXCR7 are on the order of 10,000 per cell[198–200], we chose to modify HEK-293T cells with CXCR4 or CXCR7 genes under the control of the constitutive CMV promoter to empower overexpression. cDNA encoding either CXCR4 or CXCR7 was stably transfected into adherent HEK-293T cells, and expression validated and quantified via flow cytometry (Figure 3.14).
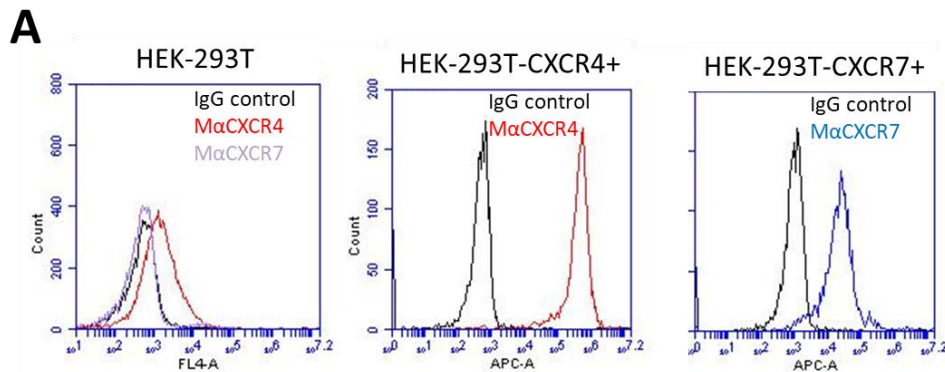
Figure 3.14: Modified HEK293T cells stably overexpress CXCR isoforms. Native HEK-293T cells (left panel) express minimal amounts of either CXCR4 or CXCR7. HEK-293T cells transduced with either CXCR4 (middle panel) or CXCR7 (right panel) constitutively express the GPCR of interest on the cell surface.

3.3.6 Isolation of CXCR4-specific PriSMs

SM_4 conjugation to the naïve library was verified via blocking of the conjugation of a fluorescein-PEO_5-maleimide (Figure 3.15A). The conjugated library was sorted thrice for binding to HEK293T-CXCR4+. Comparative analysis for binding to HEK293T-CXCR4+ vs. HEK293T parental cells exhibited differential yield (p = 0.11; Figure 3.15B). In comparison, blank EBY100 yeast conjugated with SM_4 showed negligible enrichment, consistent with enrichment driven by PriSM conjugates rather than conjugated but non-Fn-PAS_x-Cys-expressing yeast.
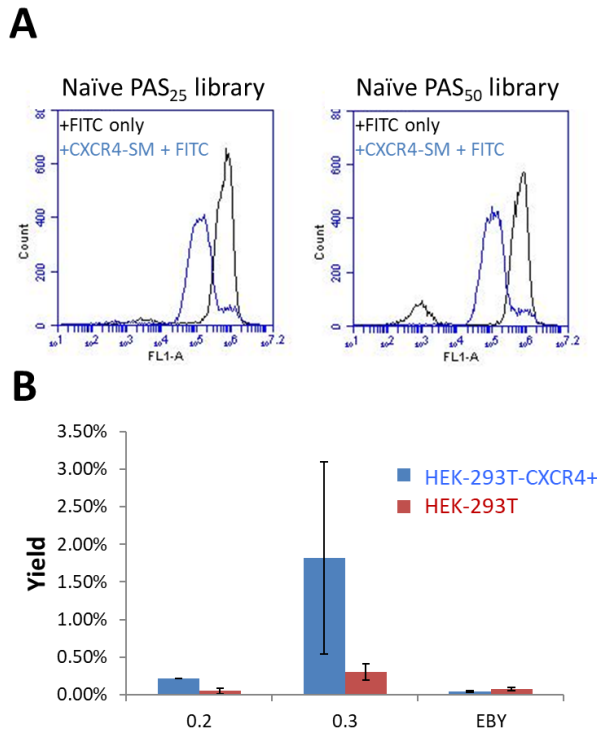
Figure 3.15: SM_4-conjugated PriSM library is enriched towards CXCR4-expressing cells. A) Validation of PAS_{25} and PAS_{50} conjugation to SM_4 molecule. Fluorescence shift indicated reduction in available cysteines for fluorescein-5-maleimide when conjugated with SM_4. B) Yield of SM_4-conjugated yeast populations panned against HEK-293T and HEK-293T-CXCR4+ cells. 0.2 indicates the second sequential sort, 0.3 the third, and EBY is nonexpressing yeast conjugated to SM_4.

3.3.7 Clonal characterization of triply-sorted CXCR4-specific PriSMs

Clones isolated from the triply-sorted CXCR4 binding population were then individually panned against HEK293T and HEK293T-CXCR4+ cells to identify potential selective clones. Individual clones and uninduced yeast were conjugated, panned in duplicate against either HEK-293T or HEK-293T-CXCR4+ cells, and imaged via bright-field microscopy (Figure 3.16). Binding was qualitatively characterized by the number of yeast present in the field of view (Figure 3.17). Several clones displayed preferential binding to HEK-293T-CXCR4+ cells over HEK-293T cells alone, which was consistent with cell panning yield results. Additionally, SM_4-conjugated yeast lacking the PriSM expression plasmid were tested in parallel, and no detectable binding was seen. Several of the clones tested were also sequenced (Table 3.4). Interestingly, two clones that appeared selective were truncated Fn3 genes, and likely acted as unstructured linkers for the SM_4 molecule. Like CAIX-specific clones, no free cysteines were found in any of the loop regions, however clone FnαCXCR4-0.3.12 did contain two sequential cysteines in

the BC loop. As with initial CA PriSM development, no selectivity pressure has been introduced to the population.
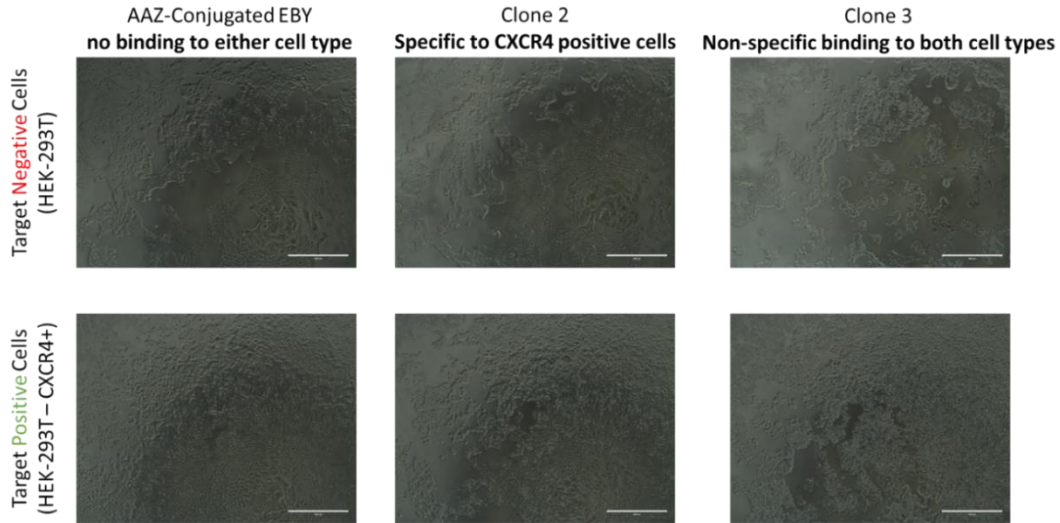
Figure 3.16: Example of clonal panning results. Shown above are representative images of a clonal panning with SM$_4$-conjugated yeast, a CXCR4-binding PriSM clone, and a non-specific PriSM clone. The amount of yeast within the frame was given a qualitative integer score from 0 (few to no yeast) to 5 (all cells decorated with yeast).
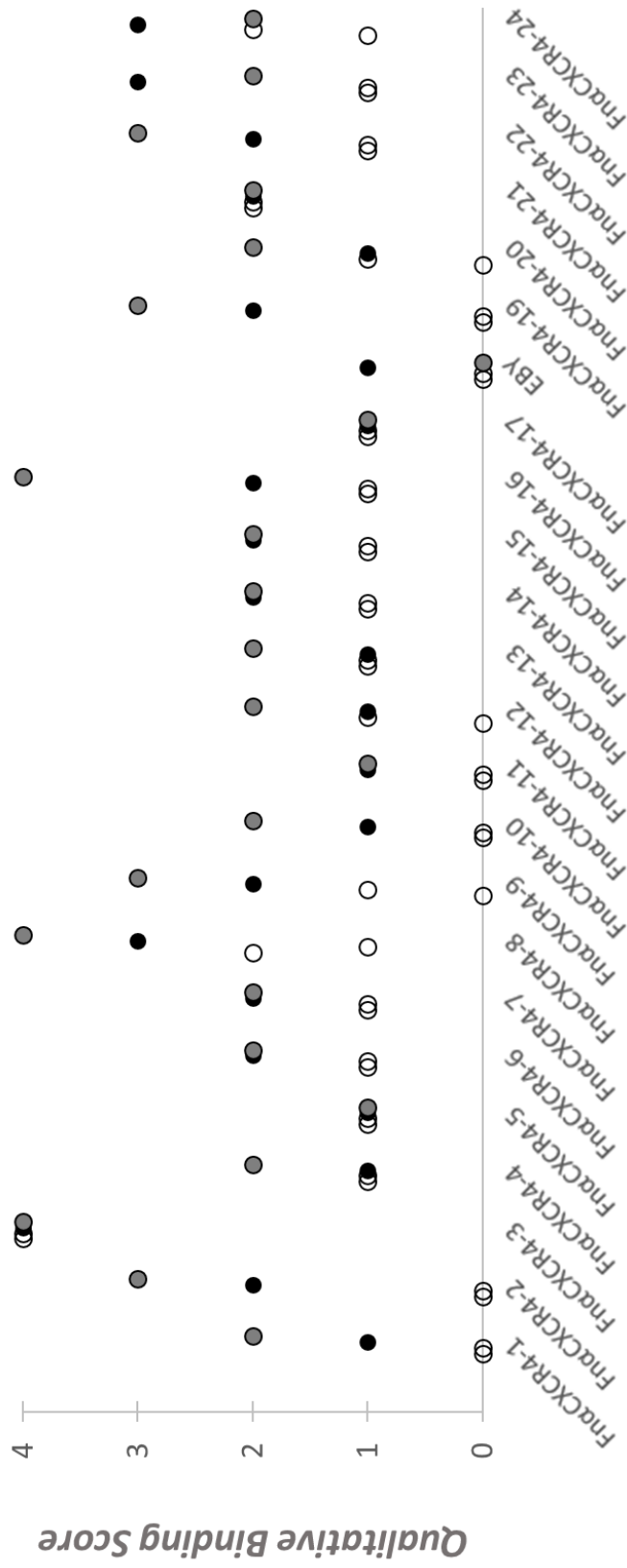
Figure 3.17: Clones from triply-sorted CXCR4 population bind preferentially to HEK-293T-CXCR4+ cells. From the sorted population, 23 clones and uninduced yeast were conjugated to SM_4 and panned individually in duplicate against HEK-293T (circles, unfilled) and HEK-293T-CXCR4+ (circles, solid fill) cells. An integer score between 0 and 5 was assigned to each well based on the number of yeast within a randomly selected portion of the well.

Table 3.4: Sequences of clones from CXCR4-binding population

Name	BC	DE	FG	Tail Length	Framework
FnαCXCR4-0.3.1	DDPDW--VHS	GYYYT	ATAGRV--SN	25	
FnαCXCR4-0.3.2	DDYST--VDS	GST-Y	TADSVDA-SN	50	
FnαCXCR4-0.3.4	DYPFA--AIS	GTN-T	TGTNSVV-SN	50	
FnαCXCR4-0.3.6	DAYAHVTSDY	GST-T	ASDSTHH-SN	25	
FnαCXCR4-0.3.8	DYHYH--VLY	GYYNS	ASVGYR--SN	50	
FnαCXCR4-0.3.9	DYHVYG-VYY	GTN-Y	AADDNE--SN	50	
FnαCXCR4-0.3.11	DDYVDV-AVS	GTTSY	TSESYA--SN	50	
FnαCXCR4-0.3.12	DDPVFCCALS	GNN-N	AGDSVLK-SN	50	
FnαCXCR4-0.3.14	DSYHS--AIS	GTS-N	TGVNAY--SN	25	
FnαCXCR4-0.3.15	DDYYQ--ALS	GSN-N	TAVDDAH-SN	50	
FnαCXCR4-0.3.16	-----			50	Truncation (45 AA)
FnαCXCR4-0.3.19	DDYSW--VRS	GYYYY	ADDGGR--SN	50	
FnαCXCR4-0.3.20	DAPAA--AIS	GYN-Y	TNGGHELYSN	50	
FnαCXCR4-0.3.23	-----			50	Truncation (40 AA)
FnαCXCR4-0.3.24	DDHADDDAIY	GYN-T	ADDGVYD-SN	50	

To identify if these clones are isoform-specific, the triply-sorted CXCR4 binding population was panned against HEK-293T-CXCR4+, HEK-293T-CXCR7+, and HEK-293T-CEA+ (a non-GPCR target) in triplicate, and resulting populations were submitted for deep sequencing. Clones will be selected from this population to be tested as soluble PriSM molecules and to assess their ability to block native ligand SDF-1 binding to CXCR4. We have demonstrated that the yeast-display PriSM method, in addition to discovering CA-isoform-specific molecules, can also identify selective molecules to a membrane-bound target such as CXCR4. Further characterization is needed to quantify the strength and inhibitory activity of isolated CXCR4 PriSM clones.

3.4: Discussion

3.4.1 Generation of isoform-specific ligands

Isoform specific carbonic anhydrase and GPCR inhibitors have long been a goal for many communities trying to treat various diseases. For these reasons, we first chose to develop protein-small molecule inhibitors that can selectively target a single CA isoform or a single CXC-receptor isoform. For targeting CAs, our starting small molecule was acetazolamide (AAZ). AAZ is a common CA inhibitor and also an FDA-approved treatment for macular edema[162]. Our goal was to develop unique protein-small molecules that can selectively inhibit the enzymatic activity of either CAII or CAIX. We were able to select for high-affinity and highly specific CAII and CAIX ligands, and that they also selectively inhibited the CA of choice at a much higher specificity than AAZ alone. For targeting CXC-receptor isoforms, we used a previously developed 5 amino acid cyclic peptide mimetic (SM_4)[201,202] that bound to both CXCR4 and CXCR7. We showed that conjugated libraries on yeast were able to be enriched for CXCR4 binding.

3.4.2 PriSM selection greatly increased speed to strong ligand isolation

In addition to validating the PriSM concept, we have shown here an efficient and robust method of identifying selective inhibitors using an Fn3 library and a non-selective inhibitor as a starting point for developing a bivalent intramolecular protein-small molecule fusion using yeast surface display and various sorting techniques. Notably,

initial naïve library and sorting analysis showed that this small molecule drastically increased the number of recovered yeast from the first sort over traditional naïve library only sorts. This increase in the affinity of the starting library towards the target allowed for stringent selectivity sorts to be conducted at a very early stage in the sorting process, and we believe this is what resulted in the ability to recover diverse sequences that matched the phenotypes of interest. It is possible that in alternative efforts to discover and evolve new ligands towards a target, an initial 'handle' of some other ligand could be used to improve the recovery of a diverse set of weak but selective ligands from a naïve library.

3.4.3 Alternative PriSM formats

In sequencing the triply-sorted CAII binding population, we found that all clones sequenced in the CAII binding populations contained PriSM molecules that contained at least one cysteine in one of the diversified loops of the Fn3 gene in addition to the C-terminal cysteine. In contrast, this genotype was only seen in one of the 6 unique sequences from the CAIX populations. This discrepancy could be due to the difference in oligomeric state of the targets. CAIX is a dimeric cell surface protein whose complex is mediated by a disulfide bond near the active site[203], whereas CAII is monomeric[170,196]. This difference could cause the 'double' AAZ conjugated PriSMs to win out over the single conjugates due to a potential increase in effective affinity if the CAII can dimerize when in forced proximity. This oligomerization could also be the reason behind the FACS results showing stronger binding to CAIX over CAII in the naïve conjugated

library. The dimerization of CAIX is within the catalytic domain of the enzyme and could introduce additional avidity to the system that results in a falsely strong binding signal. This high avidity on yeast could also explain the less than stellar results in the soluble protein inhibition experiments where no avidity can be utilized outside of the intramolecular bispecific binding of the PriSM molecule (data not shown). Because the cell sorting strategies were based on CA binding rather than inhibition, it is possible that the clones selected for are highly specific ligands, but not strong inhibitors.

An alternative hypothesis is that the AAZ-conjugated loops are binding an area away from the active site. Crystal structure data shows that AAZ binds to two other sites on CAII other than the catalytic pocket (Figure 3.2). While these complexes could be artifacts of the excess AAZ used in crystallization, it could explain how these dual-AAZ conjugated PriSMs can be highly selective.

3.4.4 Yeast-displayed PriSM libraries identify ligands to 'undruggable' targets

In targeting CXCR4 via the PriSM system, we did not see a similar strong affinity for CXCR4+ cells of the naïve library, though this is likely due to the weak binding of the SM_4 molecule. However, a similar trend in selection efficiency was seen. The yield divergence of SM_4 containing molecules at the 3rd sort was markedly sooner than many other cell-panning techniques[39,73,204,205]. Clonal panning results confirmed that many clones within the sorted population were specific to the CXCR4 receptor-containing cells. However, we have yet to identify whether the molecules are selective to CXCR4 over

CXCR7, which is a much more stringent requirement. Selectivity to CXCR4 over

CXCR7 is also a clinically relevant problem in cancer therapy. CXCR7 is a known

scavenger of the SDF-1 ligand[206,207], and this also holds for any inhibitor that blocks

SDF-1 binding, such as SM_4. If an SM_4 based PriSM could be validated, it would likely

improve the therapeutic window of the drug in addition to the potency.

3.5: Conclusion

In this study we demonstrated that the PriSM concept can create highly selective

ligands robustly and efficiently. CA isoform-selective ligands were developed that have

over 500-fold specificity to either CAII or CAIX over the alternative CA tested.

Characterization of the clones identified multiple mechanisms of AAZ:Fn binding that

could lead to the development of new inhibitors requiring only one paratope for

selectivity. We demonstrated that GPCR-specific ligands could be isolated using a

known small molecule and whole cell yeast panning. Further off-yeast characterization is

needed to verify these results; however, these findings demonstrate the utility of

engineering partner ligands to improve affinity, selectivity, and potency of small

molecules and other ligands.

Chapter 4: Activatable photoacoustic lifetime imaging probe for tumor detection

This work was done in collaboration with Dr. Ekaterina Morgounova-Ippolito and Professor Shai Ashkenazi of the Biomedical Engineering departing at the University of Minnesota. Their expertise in photoacoustic imaging was integral to this work and we are extremely grateful for their fruitful collaboration. This chapter was jointly written as a manuscript for submission for peer review, which is forthcoming.

4.1: Introduction

4.1.1 Activatable probes in cancer imaging

Photoacoustic molecular imaging is a rapidly growing imaging modality that combines the selectivity of light-absorbing contrast agents with the relatively high spatial resolution and penetration depth of ultrasound imaging[208,209]. Different types of photoacoustic imaging probes have been developed to visualize biological events[210,211]. Early probe designs for cancer research have relied on passive uptake through the enhanced permeability and retention (EPR) effect to enter the tumor site, but they lack specificity and show high off-target signal. Targeted photoacoustic probes can bind to a specific molecular marker such as cell surface receptors, transporters or enzymes, and accumulate at the target site, but since they are "always on" they cannot provide functional information about dynamic events. A third class of photoacoustic contrast agents are activatable smart probes that can detect specific events by switching from a silent state to a visible state upon cleavage by a target molecule, or conformational reorganization induced by environmental

factors. Activatable photoacoustic probes can therefore provide low background noise and high imaging contrast[212–222].

4.1.2 Matrix metalloproteinase activity as cancer biomarkers

A prominent example of molecular targets studied in both research and clinical settings are matrix metalloproteinases (MMPs), a family of enzymes that degrade proteins in tissue extracellular matrices (ECM)[223]. They are involved in normal physiological processes such as wound healing, but also in pathological states such as cancer progression and inflammation-related diseases[224]. Although MMP inhibitors have been extensively studied as drug targets for cancer therapy, treatments have met with limited success *in vivo* so far, possibly due to the complexity of their signaling activities[225]. Hence, developing molecular tools and techniques to visualize MMPs is still needed to better understand their role in disease initiation and progression.

4.1.3 Current research of activatable probes

Recently, several groups have developed activatable smart probes for photoacoustic imaging of enzymes[213,215,216,226], copper(II)[217], reactive oxygen species[218], pH[219,220], and temperature[221], or demonstrated the use of commercially available activatable fluorescent probes with photoacoustic detection[214,222]. Two common methods to detect the active probe signal include single and dual wavelength imaging. Single wavelength imaging maximizes photoacoustic emission by exciting at the maximum of absorption of the active probe signal[214,215,220,221]. Dual wavelength imaging relies on a change in the absorption spectrum

between the inactive and active forms of the probe to enhance contrast, using either spectral subtraction[213,226], or radiometric detection[216–219]. Both methods assume low and homogeneous absorption of background tissue absorbers and do not actively attempt to subtract the background noise. In practice, however, background absorption can vary significantly, and resolving the active probe signal may become challenging in the presence of highly absorbing and spatially heterogeneous tissue chromophores combined with low probe concentration. One strategy is to incorporate particles with high absorption capacity and high energy conversion efficiency into the probe design, such as semiconducting polymers[218,219] or gold nanoparticles[220]. Alternatively, activated probes may be designed to aggregate locally resulting in photoacoustic signal amplification[215,221]. Finally, multispectral optoacoustic tomography (MSOT) is capable of extracting the probe signal by decomposing, or unmixing, a photoacoustic image based on the unique spectral signature of its photoabsorbing components[222]. In its simplest application, MSOT assumes an a priori knowledge of the absorption spectrum of each endogenous and exogenous absorber and performs a linear fit of the photoacoustic signal acquired at multiple excitation wavelengths.

4.1.4 Photoacoustic lifetime imaging reduces endogenous background signal

Photoacoustic lifetime imaging (PALI) is an alternative imaging method that detects the excited-state lifetime of molecules using a two-step pump-probe approach. First, a laser pulse (pump) excites the molecule to a higher energy level. Then, a second laser pulse (probe) interrogates the transient excited state by generating a transient

photoacoustic signal. By scanning the time delay between the pump and probe pulses, we can capture the relaxation dynamics of this excited state. As endogenous tissue absorbers such as hemoglobin, melanin and albumin have small excited-state lifetimes in the 0.1 – 10 ns range[227–230], PALI can ignore their photoacoustic signal by suppressing any signal under a set lifetime threshold. Photosensitizers such as methylene blue (MB) and some porphyrin derivatives show potential as photoacoustic lifetime contrast agents due to their high optical absorption in the red region of the spectrum and their long-lived triplet excited state[231,232]. In addition, our collaborative team has previously shown that the dimerization of MB molecules led to the quenching of their photoacoustic lifetime in sodium sulfate and sodium dodecyl sulfate solutions[233].

4.1.5: Conceptual design of photoacoustic lifetime imaging probes

Here, we considered whether PALI could be employed to resolve the activation of a cleavable smart probe labeled with two MB chromophores arranged in a dimer configuration. The probe was based on a design by Jiang et al. who developed an activatable cell-penetrating peptide for fluorescence imaging comprising an electrostatic zipper[234]. After cleavage by the target enzyme, half of the probe enters and accumulates in the cells, resulting in signal amplification[235]. This design was later modified by Levi et. al who evaluated two chromophore-quencher pairs for dual-wavelength photoacoustic imaging[226]. More recently, Rood et al. developed a fluorescent activatable lifetime imaging probe using luminescence quenching[236]. Our lifetime-based activatable probe comprises a pair of MB chromophores conjugated to an electrostatic zipper and covalently linked by

the MMP-2 specific recognition sequence PLGLAG (Figure 4.1). MB dimerization is facilitated by two mechanisms: 1. The N-terminal proline of the PLGLAG sequence facilitates a turn in the peptide backbone[237]; 2. The electrostatic attraction between the two halves of the zipper bring MB molecules together. Cleavage of the amide backbone by MMP-2 enzyme separates the two halves, and the chromophores recover their long excited-state lifetime.
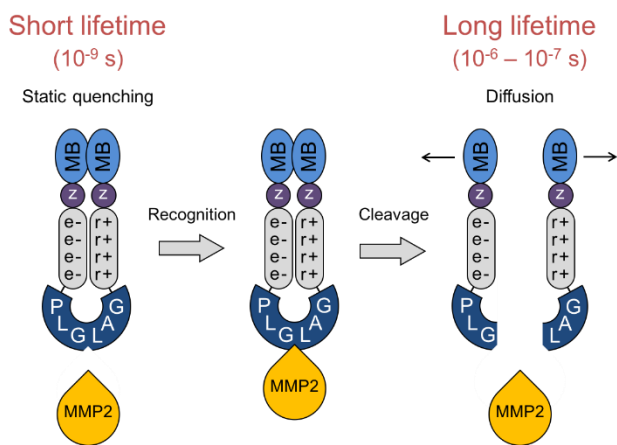
Figure 4.1. Photoacoustic probe design. The probe is composed of a MMP-2 recognition sequence PLGLAG, an electrostatic zipper comprising cationic poly-D-arginine (r+) and anionic poly-D-glutamate (e-) domains, a flexible linker (z) and a pair of MB chromophores. The intact probe is hypothesized to have a short lifetime of tens of nanoseconds due to static quenching. Upon cleavage of the recognition sequence, thereby elimination of the indirect covalent linkage, the electrostatic zipper is insufficient to maintain MB dimerization. MB monomers separate and exhibit long lifetime (several microseconds).

Here we describe the synthesis and evaluation of several conjugated agents. First, we assessed the dimerization and quenching efficiency of a pair of MB chromophores covalently linked by a single lysine (MB$_2$K). Then, we compared the basic functionality of 12 activatable peptide probes with general structure MB-Kz(e)$_n$[Ahx]PLGLAG(r)$_{mz}$K-MB, as well as pre-activated and non-activatable controls (Table 4.1). We studied the effect of

varying the length of the electrostatic zipper $((e)_n(r)_m)$ and introducing a flexible linker (z)

on dimerization/separation efficiency and activation rate in order to identify the optimal

probe sequence. Finally, we measured the lifetime contrast before and after incubation with

MMP-2 to determine the photoacoustic quenching efficiency of the activatable probes.

Probe	n (e-)	m (r+)	z	*Targeting sequence*
E4	4	-	-	PLGLAG
R5	-	5	-	PLGLAG
E0R0	0	0	-	PLGLAG
E0R1	0	1	-	PLGLAG
E1R2	1	2	-	PLGLAG
E2R3	2	3	-	PLGLAG
E3R4	3	4	-	PLGLAG
E4R4	4	4	-	PLGLAG
E4R5	4	5	-	PLGLAG
E5R6	5	6	-	PLGLAG
E4DR5	4	5	-	plglag
E4scrR5	4	5	-	Pgallg
GE4R5G	4	5	G	PLGLAG
XE3R4X	3	4	[Ahx]	PLGLAG
XE4R5X	4	5	[Ahx]	PLGLAG
XE5R6X	5	6	[Ahx]	PLGLAG

Table 4.1. Design of the activatable probe sequences. General form MB-Kz(e)$_n$[Ahx]PLGLAG(r)$_m$zK-MB, where n represents the number of D-amino glutamic acids, m represents the number of D-amino arginines, and z represents a flexible linker (G, 6-aminohexanoic acid [Ahx], or none). A simplified naming convention (zEnRmz) is used to describe the different probe designs. Pre-activated positive controls are identified as E4 and R5. Negative D-amino and scrambled controls are identified as E4DR5 and E4scrR5, respectively.

4.2: Materials and Methods

4.2.1 Materials

Methylene blue hydrate (purity $\geq$ 97%) and methylene blue N-hydroxysuccimide ester (MB-NHS) were acquired from Sigma-Aldrich and ATTO-TEC (Siegen, Germany), respectively. L-lysine was purchased from Sigma-Aldrich, and activatable peptides were purchased from UnitedPeptide (Herndon, VA). Recombinant human MMP-2 proenzyme was from Calbiochem. *p*-aminophenylmercuric acetate (APMA) (**Toxic!**) was from Sigma-Aldrich. TCNB buffer was prepared by mixing 50 mM Tris-HCl, 10 mM $CaCl_2$, 150 mM NaCl and 0.05% (w/v) Brij-35 (Technicon), and adjusting the pH to 7.5 by adding 1M HCl.

4.2.2 Synthesis of MB-lysine conjugates

MB-NHS ester and L-lysine were resuspended in dry DMSO to 20 mg/mL and 2.2 mg/mL, respectively. A 3-fold molar excess of L-lysine was added to the MB-NHS solution to ensure both MB-K and MB_2-K as products, along with 2% v/v trimethylamine. This mixture was left to react protected from light at room temperature for 2 hours. The samples were then purified using a C18 reverse-phase HPLC purification column with a solvent gradient of 22.5-67.5% acetonitrile in water with 0.1% trifluoroacetic acid. Individual peaks containing dually and singly labeled conjugates were collected and analyzed via electrospray ionization mass spectrometry (MALDI-TOF/TOF 5800, Applied Biosystems-Sciex Framingham, MA). Finally, the samples were lyophilized and stored at 4°C prior to testing.

4.2.3 Synthesis of activatable peptide probes and their controls

Peptides were resuspended in distilled water or buffer solutions of 1X PBS and 0.2 M sodium bicarbonate (20:1 v/v) with a final pH of 8.3. To conjugate each peptide terminus with MB, a 3-6 molar excess of MB-NHS resuspended in DMSO (20 mg/mL) was added to each peptide solution and allowed to incubate for > 24 hours protected from light at room temperature. The samples were then purified via reverse-phase HPLC using a C18 column with varying gradients of water/acetonitrile and 0.1% TFA. Peaks containing dually-labeled peptides were collected. Peak purity of >95% was verified via mass spectrometry, with lower purity samples undergoing further reverse-phase HPLC purification. Purified probes were lyophilized and stored at 4°C and resuspended before use. Prior to each experiment, the absorption spectra of the probes were verified to be invariant with concentration (data not shown), indicating that no significant cleavage took place prior to enzymatic incubation.

4.2.4 Absorbance spectra measurements

All MB-lysine conjugates and activatable peptide compounds were first solubilized in distilled water and then diluted to their working concentration in distilled water or 1X PBS before testing. Absorption spectra were measured using 384-well microplates with a SpectraMax Plus 384 spectrophotometer (Molecular Devices, CA) or a Synergy H1 spectrophotometer (BioTek, VT). The pure monomer and dimer spectra were obtained from a two-state linear regression model of MB dimerization described previously[233]. Total molar concentrations were calculated using Beer Lambert's law and the extinction

coefficient of MB at the isosbestic point (35,700 $M^{-1}.cm^{-1}$ at 625 nm) calculated with this model.

4.2.5 Enzyme kinetics

MMP enzymes become active after cleavage of their inhibitory N-terminal domain[238]. MMP-2 proenzyme was activated by incubating a 1:9:1 v/v ratio of MMP-2, TCNB buffer and 11 mM APMA freshly dissolved in 0.1 M NaOH at 37°C for 45 minutes. All experiments performed with APMA must follow the guidelines established for using organic mercuric compounds, including wearing appropriate protective equipment. The initial activation rate of the probes was measured by incubating 10 μM peptide probe with 568.6 nM (0.2 ng/ μL) activated MMP-2 in distilled water or in 1X PBS at 37°C for 3 hours. We monitored the reaction by recording the absorbance at 613 nm ($\lambda_{max, dimer}$), 668 nm ($\lambda_{max, monomer}$), and 630 nm (isosbestic point) over time. Absorbances were normalized by the total concentration calculated at the isosbestic point to account for a drift in concentration over time due to evaporation. The initial rate of activation was calculated by fitting a linear regression model between the 6th and 15th minute. Measurements were performed three times with 2-3 replicates each time.

4.2.6 Photoacoustic signal acquisition.

The photoacoustic experimental setup used to measure transient photoacoustic absorption was described in detail previously and reproduced in Figure 4.2a[233]. A pump-probe system comprising two Q-Switched Nd:YAG lasers (Phocus Mobile, Opotek, and Quantel, Brilliant), each equipped with a second harmonic generator, delivered 5 nanosecond pulses at 10 Hz repetition frequency. Each beam was directed into an optical parametric oscillator (OPO; MagicPRISM, Opotek) for wavelength conversion. The pump and probe wavelengths were tuned to the maximum of absorption of MB monomers (660 nm) and the maximum of absorption of the transient excited state (830 nm) to maximize signal generation. The beams were then combined into a bifurcated randomized fiber bundle to ensure a uniform illumination of a plastic tube containing the samples. The output of the light guide was inserted into a 3D-printed photoacoustic chamber and aligned with the tube and a focused ultrasound transducer (f = 1″, Panametrics, V311, 10 MHz) (Figure 4.2b). The optical energies after the light guide were 12 and 17 mJ for the pump and probe laser, respectively. A peristaltic pump (Dolomite Microfluidics, MA) was used to draw the solutions in alignment with the light guide and the transducer. Laser firing, ultrasound acquisition, and the delay between the two pulses were controlled by MATLAB via a field-programmable gate array (FPGA, KNJN, FX2 Saxo). Pump-probe delay times were randomized to correct for time-dependent photobleaching of the solutions. The photoacoustic signal was recorded by a digital oscilloscope (Lecroy, LC584) and processed with MATLAB. Noise reduction was achieved by averaging the signal over 100 measurements during acquisition.
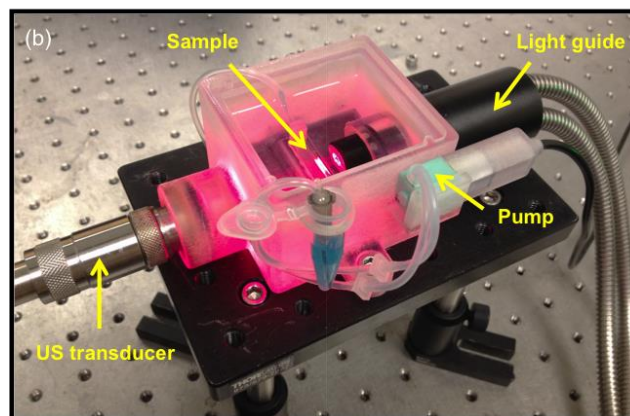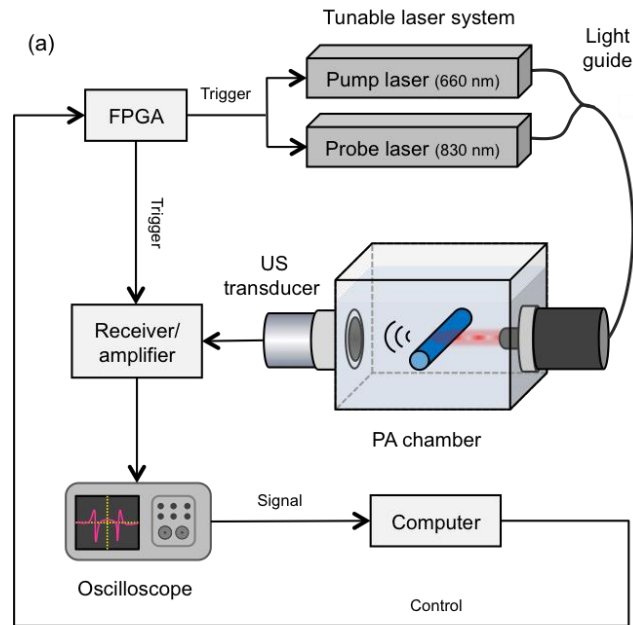
Figure 4.2: (a) Photoacoustic experimental setup. A 3D printed chamber ensures a repeatable and precise alignment between the light guide, the plastic tube filled with the sample, and the ultrasound transducer. (b) Photoacoustic chamber during laser firing.

4.2.7 Photoacoustic signal processing

For each pump-probe delay time, the transient photoacoustic signal was obtained

by subtracting the signal recorded with the pump laser only and the probe laser only from

the signal recorded with both lasers: $tPA = PA_{660+830} - PA_{660} - PA_{830}$. The signals were then processed with a Wiener filter followed by an averaging filter. The envelope was detected using the Hilbert transform method. Finally, the photoacoustic lifetime was evaluated by exponential fitting of the transient photoacoustic signal at the dye-plastic wall interfaces in function of pump-probe delay time.

4.2.8 Statistical analysis

An unpaired two-sampled t-test was used for statistical analysis of the probe activation data. The data were expressed as mean ± standard error (SEM) from 3 repeat activation experiments using 2-3 replicates each time. Each repeat experiment was conducted with freshly activated MMP-2. P-values < 0.001 were considered statistically significant.

4.3: Results and Discussion

4.3.1 Quenching of MB$_2$K conjugate is driven by MB dimerization

We first tested the hypothesis that intramolecular connection of two MB molecules by a short indirect covalent link results in stable dimerization and quenching. A single lysine (K) was chosen as the linker, and it was verified that the amino acid absorbs negligibly in the wavelength range of this experiment. Absorbance spectra were collected for free MB, singly conjugated control MBK, dually conjugated MB$_2$K and hydrolyzed MB-NHS ester (MB-COOH) at different dilutions. Increasing the concentration of free MB dye results in a hypsochromic shift of the maximum of absorption from 665 nm to 608 nm

as seen in sandwich-type (H-type) dimers (Figure 4.3a)[239]. By contrast, MB$_2$K displays a spectral shape that is invariant with concentration with an absorption peak at 608 nm (Figure 4.3b), suggesting that a single unimolecular dimer dominates the solution. To further demonstrate that this peak is caused by intramolecular MB pair interaction and not by intramolecular MB-lysine interaction, we measured the spectrum of the singly conjugated control MBK, which revealed that MBK undergoes aggregation/dissociation processes marked by a peak shift similar to that of free MB (Figure 4.3c), albeit with a lower dissociation constant. These combined results suggest that a pair of MB molecules held in close proximity by a covalent linker form a stable dimer in aqueous solution. Finally, the absorbance spectra of MB-COOH also shows a peak shift with increasing concentrations (Figure 4.3d), indicating that this contaminant product of the labeling reaction dimerizes like free MB and MBK in solution.
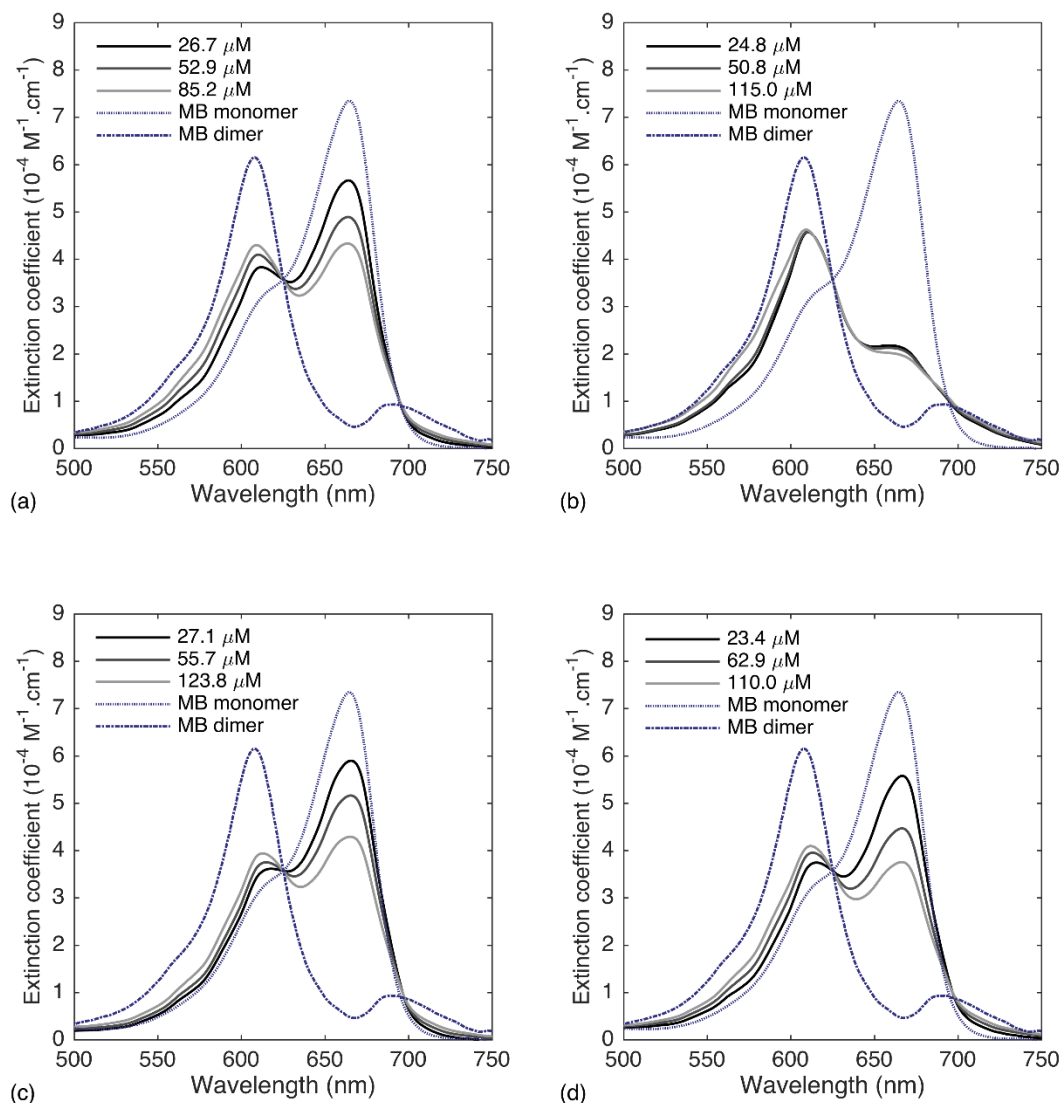
Figure 4.3: Extinction coefficient spectra of free MB dye (a), dually conjugated MB$_2$K (b), singly conjugated MBK (c), and MBCOOH (d) diluted at different concentrations in 1X PBS. Dashed lines represent the pure MB monomer and dimer spectra calculated from the dimerization model described in (*25*).

Interestingly, the absorption spectrum of MB$_2$K does not match the MB dimer spectrum calculated from the dimerization model used in our previous study, and exhibits

both a lower peak at 608 nm and a higher peak at 665 nm[233]. Characterization of MB_2K solution by mass spectrometry revealed that the solution contained 3% MB-COOH, which corresponds to a MB-COOH concentration ranging from 0.80 to 3.71 µM across the dilutions – a range at which monomers represents more than 97% of all species. This would translate into a spectral decline of up to 946 $M^{-1}.cm^{-1}$ at 608 nm and a spectral rise of up to 2,060 $M^{-1}.cm^{-1}$ at 665 nm, which cannot fully explain the extent of the observed deviation. An alternative explanation is that a portion (26% according to our model) of dually-conjugated MB_2K remain undimerized, resulting in a higher absorption in the monomer region. Finally, it is possible that MB_2K conjugates are completely dimerized but that their spectrum is different from that of pure MB dimer.

We assessed the quenching of MBK and MB_2K by measuring their photoacoustic lifetime signal. Figure 4.4a shows the transient photoacoustic signal of the two conjugates for increasing pump-probe delay times (0.25, 1 and 2 µs). The back and front dye/plastic interfaces are visible at x = 0.9 mm and x = 2.6 mm, respectively. The photoacoustic lifetime was obtained by fitting an exponential to the maximum of the PA signal after envelope detection in function of pump-probe delay time (Figure 4.4b,c). One can see that both conjugates present a photoacoustic lifetime, however the amplitude of the exponential fit decreases by 30-fold between MBK and MB_2K. The lifetimes were estimated to be 1.44 $\pm$ 0.23 µs ($R^2 = 0.982 \pm 0.022$) for MBK and 1.00 $\pm$ 0.60 µs ($R^2 = 0.806 \pm 0.126$) for MB_2K. The small signal observed for MB_2K may be due to the presence of a small amount of monomeric species, either in the form of MB-COOH monomers or undimerized MB_2K, however the lower R^2 values and higher standard deviation suggest that it could be the

result of curve fitting error brought by low signal-to-noise ratio (SNR). In conclusion, the decrease in lifetime amplitude for MB₂K combined with a similar photoacoustic lifetime for the two conjugates demonstrates that MB₂K is statically quenched.
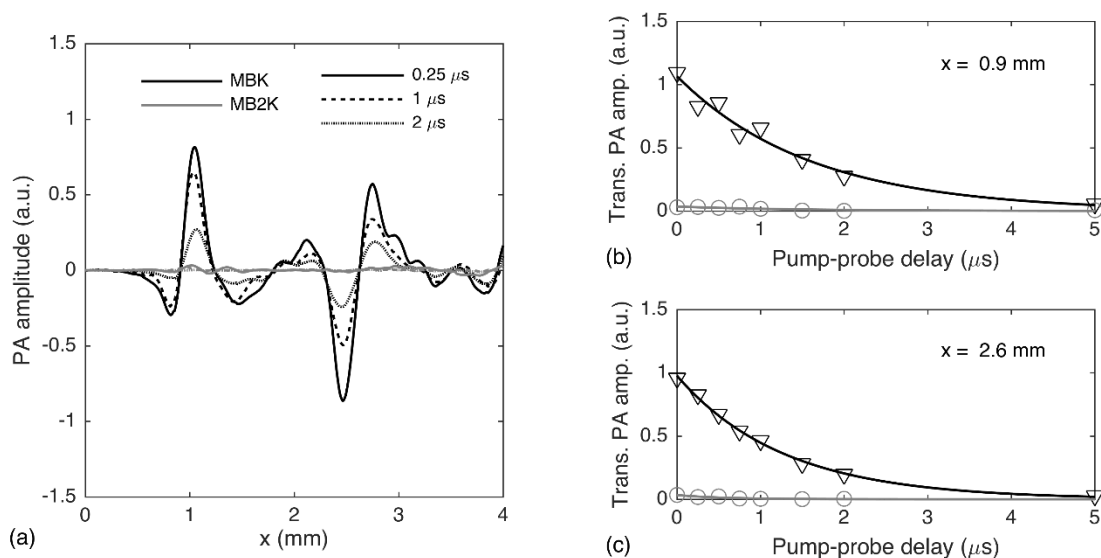
Figure 4.4: Dimerized MB exhibits long lifetime. (a) Transient photoacoustic signal of a plastic tube filled with MBK and MB₂K at 50 μM in 1X PBS at three different pump-probe delay times (0.25 μs, 1 μs and 2 μs). (b-c) Photoacoustic lifetime signal of the same solutions at the front (x = 0.9 mm) and back (x = 2.6 mm) dye-plastic interfaces. The lifetimes were calculated by applying an exponential fit to the maximum of the envelope signal.

We previously posed the hypothesis that MB dimers have a short, excited state lifetime of a few tens of nanoseconds. However, our results do not show that MB₂K conjugates have a short lifetime component in this time range. This observation was verified by measuring the transient absorption of MB₂K with an excitation wavelength at the isosbestic point (625 nm) using flash-photolysis. This suggests that MB₂K may not

undergo intersystem crossing to a triplet state but another type of de-excitation process such as direct relaxation from the singlet to the ground state.

4.3.2 Validation of E4R5 activation

The E4R5 probe (Figure 4.1, Table 4.1) was used to validate whether the general design was amenable for use with MB monomer dimerization and subsequent monomerization by MMP-2. E4R5 fully dimerized the conjugated MB molecules, as the extinction coefficient values were independent of probe concentration. E4R5 was activated by MMP-2 over the course of three hours (Figure 4.5a), and the spectra shifted to one of MB monomer composition (Figure 4.5b). Notably, the maximum of absorption of the uncleaved and cleaved form of the probes were red-shifted by about 4 nm compared to free MB dye and MB-lysine conjugates.
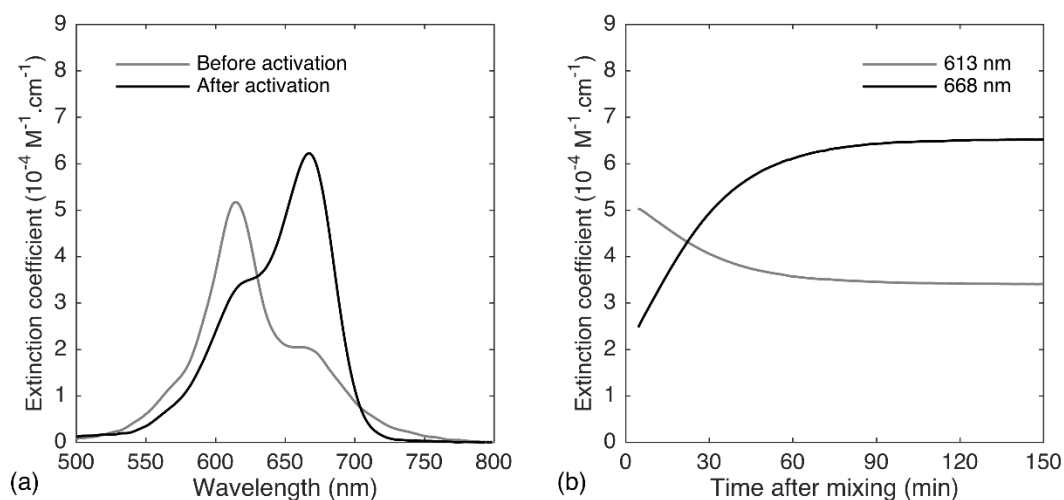
Figure 4.5: Spectral observation of peptide cleavage by MMP-2. (a) Absorption spectra before/after incubation. (b) Rate of monomerization.

4.3.3 Comparison of activatable peptide probes to optimize imaging contrast

In order to select the optimal probe design, we compared the dimerization state and quenching efficiency of the different probes before and after incubation with MMP-2. Before incubation, dually conjugated probes showed a dominant absorption peak in the dimer region but variable dimerization efficiencies (Figure 4.6). Apart from E0R0 and E1R2, probes without a flexible linker displayed a higher dimer-to-monomer ratio (D:M) than probes with a G or X linker ($P < 0.0005$). Furthermore, the D:M decreased significantly with linker size when comparing probes of the same zipper length (D:M$_{E4R5}$ > D:M$_{GE4R5G}$ > D:M$_{XE4R5X}$) ($P < 0.0001$). These results indicate a size-dependent disruption of MB dimerization by flexible linkers, which we posit may be mediated either by increased conformational freedom or by increased distance between the chromophores. Additionally, whereas the D:M did not differ significantly between E0R1, E2R3, E3R4, E4R5 and E5R6 probes, the D:M of XEnRmX probes was inversely related to their length ($r = 0.998$, $P < 0.05$). This suggests that flexible linkers may counteract the electrostatic attraction of the zipper, and that longer zippers may be more prone to unzipping resulting in lower dimer formation. On the other end, the zipper-free E0R0 probe displayed a significantly lower D:M than other EnRm probes apart from E1R2 ($P < 0.001$), indicating that a minimal zipper length is necessary for terminal MB chromophores to reach a configuration conducive to dimerization. Finally, the D:M of the E4R5 probe and its non-activatable D-amino and scrambled controls E4DR5 and E4scrR5 were not found to be significantly different ($P > 0.001$).

Incubation of E*n*R*m* probes with MMP-2 led to a 79.5 % average decrease in D:M with a final D:M comparable to that of split controls E4 + R5 (P > 0.001), indicating a complete separation of MB chromophores. In comparison, GE4R5G and XE*n*R*m*X probes showed a higher D:M than E*n*R*m* probes (except E5R6) after incubation (P < 0.001). We posit that the presence of a linker may partially hamper the separation of the probes or that the linker impacts MB's absorption spectrum. Interestingly, the D:M of non-activatable controls E4DR5 and E4scrR5 decreased by approximately 10 % each, indicating that cleavage took place to a lower extent over the incubation period, possibly due to enzymatic action or to reaction conditions.
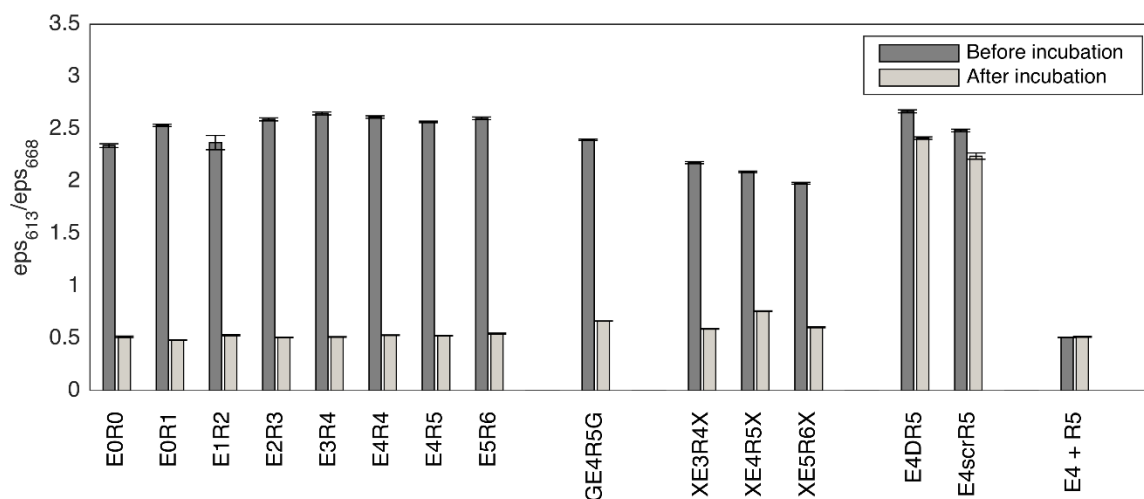
Figure 4.6. Peptide dimerization efficacy. Mean (± SEM) of the dimer-to-monomer absorption ratio (D:M) of the peptide probes (10 µM) before and after incubation with MMP-2 (0.2 ng/µL) for three hours.

4.3.4 Probe photoacoustic lifetime before and after MMP-2 cleavage

Two cleavable probes (E2R3 and E4R5) as well as the split and scrambled controls were assessed for their quenching efficiency (Figure 4.7). In accordance to the absorption results, E2R3 and E4R5 probes recovered their photoacoustic lifetime signal by more than 6 folds after cleavage to reach a signal intensity comparable to that of a mixed E4 + R5 solution. The photoacoustic lifetime signal amplitude before incubation was slightly higher for E4R5 than for E2R3, possibly due to a small amount of cleaved E4R5 initially present in solution as supported by a greater absorption in the monomer region, or to low SNR hampering the calculation of the lifetime and resulting in low R^2 (0.742 ± 0.128 µs, R^2 = 0.981 ± 0.022). In comparison, the scrambled control retained a low photoacoustic lifetime signal before and after enzymatic incubation. Interestingly, the lifetime of the split controls was found to be reduced by half after enzymatic incubation. An investigation into the origin of this difference excluded oxygenation variations and different buffer environments, leaving the incubation process – including the presence of APMA and MMP-2 and heating to 37°C – as a possible explanation.
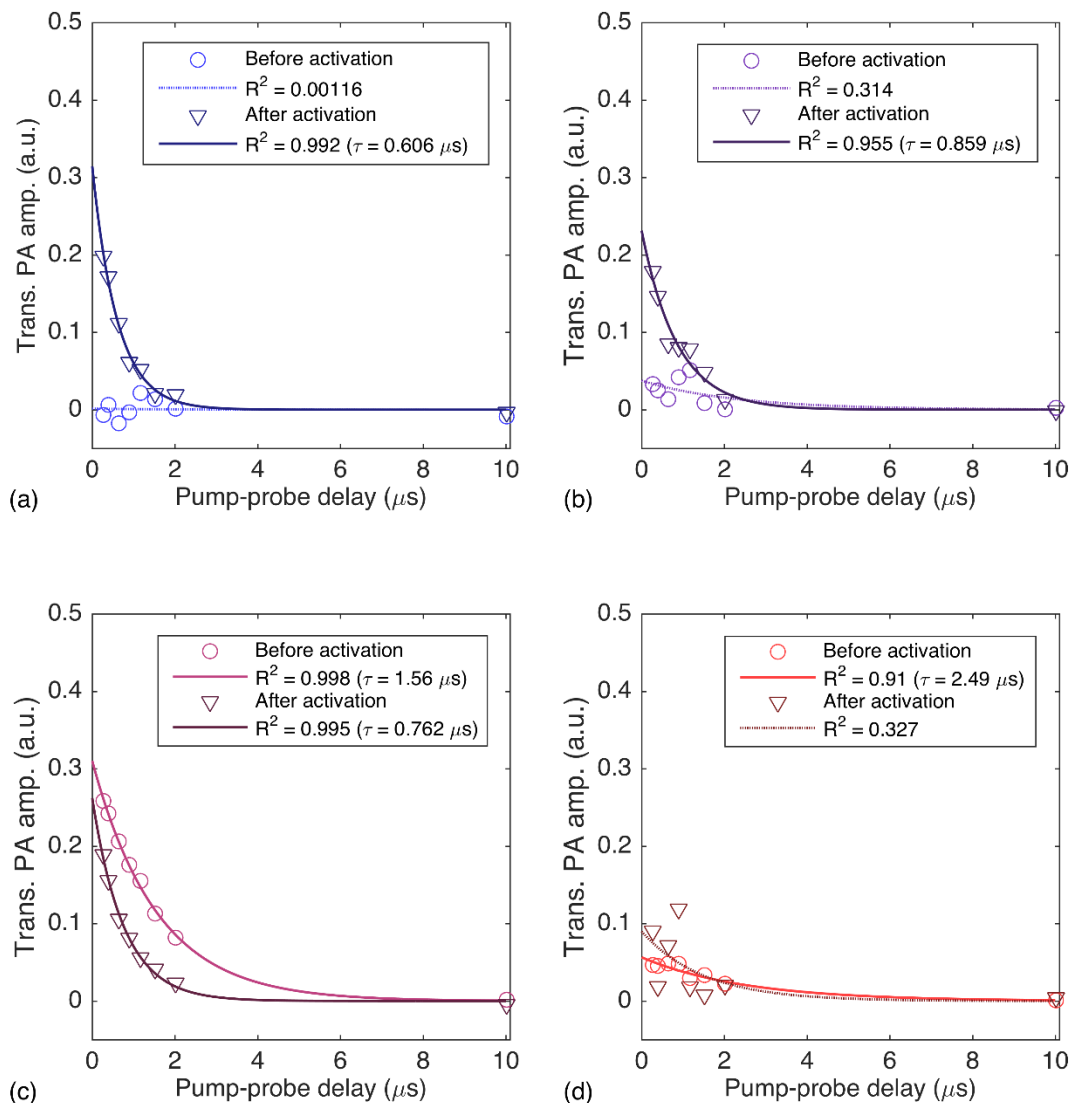
Figure 4.7. PALI signal from peptides is activated by MMP-2. Photoacoustic lifetime signal of the back of a plastic tube filled with E2R3 (a), E4R5 (b), E4 + R5 (c) and E4scrR5 (d) solutions at 50 μM before (○) and after (▽) incubation with MMP-2 (2 ng/μL) for one hour. The lifetimes were calculated by applying an exponential fit to the maximum of the envelope signal.

In vivo imaging of enzyme activity using an activatable probe requires high contrast differentiation between the inactive and active state of the probe. This condition is best

satisfied by linkerless probes due to their high dimerization and self-quenching efficiency before cleavage, and high separation efficiency and photoacoustic lifetime signal after cleavage. The length of the electrostatic zipper, if greater than zero, was not found to significantly impact dimerization and separation efficiency. After cleavage, the probe is expected to passively accumulate inside the tumor due to the enhanced permeability and retention effect, which causes tumor blood vessels to be leakier, more fragile and more disorganized in both structure and shape than healthy vessels[240]. Furthermore, the potential cell-associative and/or cell-penetrating capability of the polyarginine side chain would result in accumulation of half the probe in tumor cells as the other half is cleared away[234], which we suggest could result in time-dependent contrast enhancement.

4.3.5 Comparison of activation rate between peptide probes

The initial rate of activation (V_0) was 20 to 33 times higher in activatable probes incubated with the enzyme than in those incubated in distilled water (Figure 4.7). In addition, no significant difference was found between activatable probes and the negative controls in water ($P > 0.1$). These results suggest that most of the cleavage is due to enzymatic activity and not to spontaneous separation of the PLGLAG cleavage site. Moreover, for E4DR5 (with a non-functional D-amino acid recognition site) and E4scrR5 (with a non-functional scrambled recognition site) controls, V_0 was the same with and without enzyme ($P > 0.1$), indicating that the D:M change observed for these controls is likely due to incubation conditions and not to enzymatic action.

Apart from E0R0, the activation rate of linkerless probes – but not that of XE_mR_nX probes – was found to decrease with peptide length when incubated with MMP-2. This dependence may be explained by an increase in binding affinity of the zipper side chains with length that would delay their separation after probe cleavage, or by side chain impact on the local conformation and accessibility of the cleavage site. Additionally, V_0 decreased with linker size when comparing probes of the same length ($V_{0,E4R5} > V_{0,GE4R5G} > V_{0,XE4R5X}$), however we did not find the difference between probes to be statistically significant ($P > 0.05$).
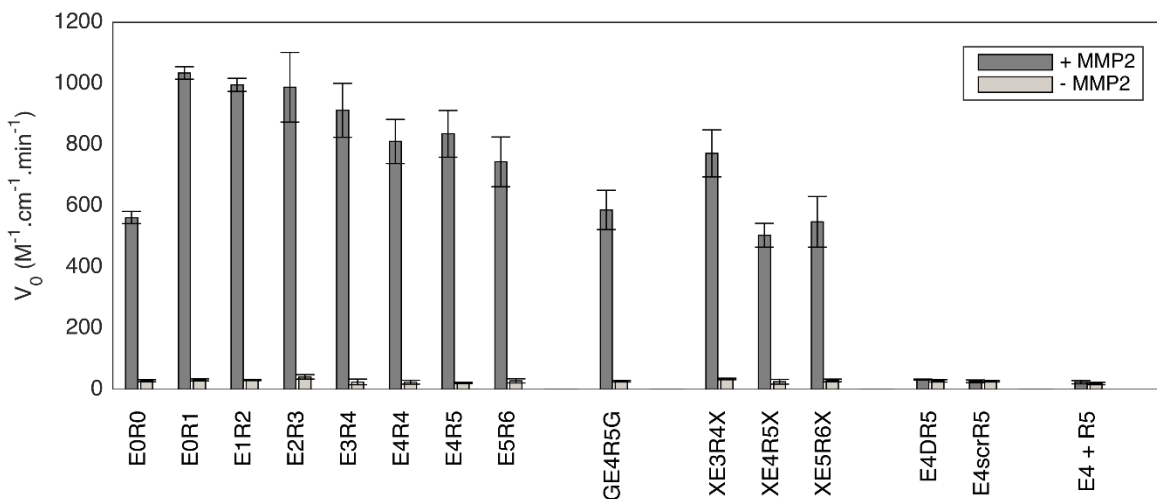
Figure 4.8. Peptide zipper length impacts activation rate. Mean ($\pm$ SEM) of the initial rate of activation of the peptide probes (10 μM) incubated with and without MMP-2 (0.2 ng/μL) for three hours.

In vitro testing of a smart probe's activation rate does not directly translate into in vivo probe behavior due to the complexity of live biological processes; but choosing a probe with an adequate activation rate and low non-specific activation provides some advantages. Because MMP enzymes can sometimes be found in the bloodstream[241], some

level of tumor-independent activation in the blood and accumulation into tissue or clearance organs may occur, which could lead to background noise and poor target/no-target contrast. Another level of complexity comes with probe specificity towards different enzymes, particularly those from the MMP family, which presents similar three-dimensional structures and substrate specificity[242]. One study comparing the sensitivity of the PLGLAG sequence to different enzymes found it to be mostly cleaved by MMP-2 and -9, and, to a lesser extent, by MMP-14[235]. Since several types of tumor secrete different enzymes at different times, the tumor enzymatic profile must be carefully considered when analyzing the performance of an enzyme-specific probe in vivo.

4.4 Conclusion

We have demonstrated the design and *in vitro* characterization of the first activatable photoacoustic lifetime probe. We have showed that a pair of covalently linked MB chromophores and the uncleaved probes dimerize and that their photoacoustic lifetime signal is suppressed due to self-quenching. The cleavage of the probe by MMP-2 results in probe separation and lifetime detection. We have also determined that probes without a flexible linker dimerize better before activation, which may translate into lower non-target signal *in vivo*, and less after cleavage, which would allow for a greater contrast. Finally, we have showed that the initial activation rate is decreasing with peptide length in linkerless probes. Other validation criteria for the probe may be tested in the future, such as cellular uptake, enzyme specificity, stability in the blood and toxicity to the cells. Moreover, the probe design can be adapted for testing a variety of biological and disease

molecular markers, both enzyme and DNA-based, by changing its recognition site sequence. Potential applications in the field of molecular imaging include enzyme-targeted drug research, cancer and inflammation-related disease diagnosis, and planning and monitoring of treatment efficacy.

Chapter 5: Perspectives and Future work

The uniting theme of this work is to advance new mechanisms to confer increased selectivity for molecular targeting agents. Each chapter presents a novel mechanism by which a molecular targeting agent, therapeutic and/or diagnostic in nature, can overcome the limitations of targeting a single epitope or physical property of the tumor. In all of the studies, the work builds on observations and innovations from previous works and provides important advances.

Mathematical modeling results from Chapter 2 provide valuable guidance for the design of multivalent molecular targeting agents. The impact of reduced monovalent affinity on selectivity has been observed in several studies, however to date there are few publications recognizing this as a systemic feature of the targeting agent, and no publications to date have systematically evaluated and optimized these multivalent ligand properties for increased selectivity. Yet it is critical to connect the modeling to experimental reality. Creating ligands of the optimal affinities proved to be the most challenging portion of this work. The non-modularity of the Fn scaffolds (and indeed the maleimide-PEO-SM of Chapter 3) was highly influential in the implementation of the molecular designs. Learning from this, it is crucial for further fusion engineering projects to conduct high-throughput screening of ligands in the context of which they will be used. Further biological AND gate work would benefit from display constructs that include a partner ligand and linker. One could also envision a multicolor sort that takes a library of diversified Fn fusions and isolates moderate affinity ligands to both targets simultaneously.

Additional considerations for the mathematical modeling work are properties of the biomarkers being targeted. We demonstrated the theoretical dependence of the selectivity on target expression, however there are many additional biomarker properties that have yet to be explored. Factors such as receptor internalization, diffusion, conformational shifts, heterogeneous cell surface distribution, and heterogeneous expression among a cell population can impact the selectivity and total uptake of a ligand. As genetic and proteomic profiling continue to improve the ability of clinicians to identify novel biomarkers for an individual patients, it is important to account for the unique biology of each of these biomarkers. Current biomarkers such as the ones used in this study have been extensively characterized, but the application of the biological AND gate is dependent on these properties that may not be known for many emerging targets.

For therapeutic applications, this dual-binding mode can be highly useful in improving drug specificity. Drug side effects can result from the lack of truly unique biomarkers, but also due to the high homology in active sites of enzymes and targets that are integral to a disease's phenotype. This lack of homolog selectivity is a difficult challenge in drug discovery and therapeutics. For a drug to be functional, it must not only have a strong affinity for only the target of interest, but also elicit a specific function. In Chapter 3, we have shown that there is no need to "reinvent the wheel" when it comes to improving a drug's potency and selectivity. We have shown that it is not only achievable, but robust and efficient to engineer a partner ligand that can improve the starting compounds. Similar to Chapter 2, when modifying both the CA and CXCR4 inhibitors with moieties amenable to chemical linkage, we saw similar decreases in binding strength of these

molecules (and also in potency for the CA inhibitor). This is a prime example of why it is so challenging to improve small molecule drug function purely by modifying the core chemical structure: not only do you have to survey a vast expanse of chemical space to find new selective interactions, you also need to mitigate the losses in favorable interactions incurred from modifying the structure. Using this PriSM concept, we can simplify the modification strategy to a less cumbersome flexible polymer linker, while utilizing high-throughput screening strategies for large protein-based libraries that access a much larger chemical space and can also access many more epitopes other than the active or allosteric site of interest that are more unique to the target homolog of interest.

As successful as the initial PriSM ligand engineering was, there were some limitations to the methodology. First, the conjugation chemistry was nonspecific and could functionalize any thiol on the yeast surface. This resulted in several clones isolated that had small molecule conjugation sites within the paratope of the binding partner. This could lead to the populations being constrained to only the epitope accessed by the small molecule, and overall limiting the ability to target any unique epitope on the target molecule. One solution is to use site-specific conjugation of small molecule to the C-terminal linker using either natural enzymatic targeting strategies (sortase, transglutaminase) or modified yeast to express non-natural amino acids containing bio-orthogonal labeling sites. These mechanisms of site-specific labeling would be a fruitful avenue of exploration to further optimize the applicability of this sorting strategy.

For the PALI probe project, there are a myriad of interesting biophysical properties of the probe that can be investigated. While it was demonstrated that the probe

can be cleaved and emit signal in the presence of targeted enzyme, there are limitations to the signal detection within tissue to be addressed. Currently, the ultrasound technology has not been able to detect concentrations of methylene blue in tissues below hundreds of micromolar, which is not a feasible dosing strategy. While this hurdle is currently being investigated by our colleagues, there are also several peptide and biodistribution properties that make the probe even more intriguing. For one, polyarginine peptides have been shown to internalize into cells, and this could greatly increase the residence time of the activated probe. Additionally, this could make cell surface anchored cleavage enzymes a valid target in addition to the soluble enzymes discussed. When this internalization is accounted for, the probe's biodistribution profile becomes uniquely interesting, creating a potential bimodal signal:noise profile during initial injection and distribution, and post injection/internalization when the majority of uncleaved probe has been cleared.

Each of these projects has the potential to be highly impactful in disease diagnostics and therapy, however each has their own hurdles to overcome. We have provided meaningful advances while also identifying important remaining challenges, including molecular modularity. Modularity is an important factor in developing multidomain and multimodal targeting agents, and it is critical to engineer these targeting agents within the construct that they will be used in the final application. With the emergence of many new biomarkers, and technologies to detect them, we will need these novel techniques to push the field of drug and molecular targeting forward.

Bibliography

1. Capelletto, E., Mariniello, A. & Novello, S. Targeted therapy in small cell lung cancer: A new era? (2017). doi:10.1016/j.lungcan.2017.03.004

2. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2016. *CA. Cancer J. Clin.* **66,** 271–289 (2016).

3. Yap, T. A., Carden, C. P. & Kaye, S. B. Beyond Chemotherapy: targeted therapies in ovarian cancer. *Nat. Rev.* **9,** 167–181 (2009).

4. Carter, P. J. & Lazar, G. A. Next generation antibody drugs: pursuit of the 'high-hanging fruit'. *Nat. Publ. Gr.* **17,** (2017).

5. Schechter, A. L. *et al.* The neu oncogene: An erb-B-related gene encoding a 185,000-Mrtumour antigen. *Nature* (1984). doi:10.1038/312513a0

6. Korea, S. *et al.* Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* **376,** 687–697 (2010).

7. Smith, I. *et al. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. www.thelancet.com* **369,** (2007).

8. Tamura, K. *et al.* 64Cu-DOTA-trastuzumab PET imaging in patients with HER2-positive breast cancer. *J. Nucl. Med.* **54,** 1869–75 (2013).

9. Dijkers, E. C. *et al.* Biodistribution of 89Zr-trastuzumab and PET imaging of HER2-positive lesions in patients with metastatic breast cancer. *Clin. Pharmacol. Ther.* **87,** 586–92 (2010).

10. Kwon, L. Y., Scollard, D. A. & Reilly, R. M. 64Cu-labeled trastuzumab Fab-PEG24-EGF radioimmunoconjugates bispecific for HER2 and EGFR: Pharmacokinetics, biodistribution, and tumor imaging by PET in comparison to monospecific agents. *Mol. Pharm.* **14,** 492–501 (2017).

11. Mortimer, J. E. *et al.* Functional Imaging of Human Epidermal Growth Factor Receptor 2-Positive Metastatic Breast Cancer Using 64Cu-DOTA-Trastuzumab PET. *J. Nucl. Med.* **55,** 23–9 (2014).

12. Cuzick, J. *et al.* Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. *Lancet* **381,** 1827–1834 (2013).

13. Robertson, J. F. *et al.* Comparison of the Short-Term Biological Effects of 7_-[9-(4,4,5,5,5- pentafluoropentylsulfinyl)-nonyl]estra-1,3,5, (10)-triene-3,17-diol

(Faslodex) versus Tamoxifen in Postmenopausal Women with Primary Breast Cancer 1. *CANCER Res.* **61,** 17 (2001).

14. Rocca, A., Maltoni, R., Bravaccini, S., Donati, C. & Andreis, D. Clinical utility of fulvestrant in the treatment of breast cancer: a report on the emerging clinical evidence. *Cancer Manag. Res.* 10–3083 (2018). doi:10.2147/CMAR.S137772

15. Narita, Y. Bevacizumab for glioblastoma. *Ther. Clin. Risk Manag.* 11–1759 (2015). doi:10.2147/TCRM.S58289

16. Harshman, L. C. & Srinivas, S. The bevacizumab experience in advanced renal cell carcinoma. *Onco. Targets. Ther.* 3–179 (2010). doi:10.2147/OTT.S8157

17. Yang, J. C. *et al. A Randomized Trial of Bevacizumab, an Anti-Vascular Endothelial Growth Factor Antibody, for Metastatic Renal Cancer. n engl j med* **5,** (2003).

18. Mitchell, P. *et al.* Ranibizumab (Lucentis) in neovascular age-related macular degeneration: evidence from clinical trials. doi:10.1136/bjo.2009.159160

19. Widakowich, C., De Castro, G., De Azambuja, E., Dinh, P. & Awada, A. Review: Side Effects of Approved Molecular Targeted Therapies in Solid Cancers LEARNING OBJECTIVES. doi:10.1634/theoncologist.12-12

20. Avery, R. L. *et al.* Intravitreal Bevacizumab (Avastin) in the Treatment of Proliferative Diabetic Retinopathy. *Ophthalmology* **113,** 1695–1705 (2006).

21. Taylor, P. C. & Feldmann, M. Anti-TNF biologic agents: still the therapy of choice for rheumatoid arthritis. *Nat. Rev. Rheumatol.* **5,** 578–82 (2009).

22. Cohen, B. L. & Sachar, D. B. Update on anti-tumor necrosis factor agents and other new drugs for inflammatory bowel disease. *BMJ* **357,** (2017).

23. Taxonera, C. *et al.* Adalimumab Maintenance Treatment in Ulcerative Colitis: Outcomes by Prior Anti-TNF Use and Efficacy of Dose Escalation A ´ ngel Ponferrada angelponmedicina@yahoo.es. *Dig. Dis. Sci.* **62,** 481–490 (2017).

24. Chul Park, S. & Tae Jeen, Y. Current and Emerging Biologics for Ulcerative Colitis. *Gut Liver* **9,** 18–27 (2015).

25. Langley, R. G. Effective and sustainable biologic treatment of psoriasis: What can we learn from new clinical data? *J. Eur. Acad. Dermatology Venereol.* **26,** 21–29 (2012).

26. Heem Wong, C., Wei Siah, K. & Lo, A. W. Estimation of clinical trial success rates and related parameters. *Biostatistics* **00,** 1–14 (2018).

27. Toczek, J., Meadows, J. L. & Sadeghi, M. M. Novel Molecular Imaging Approaches to Abdominal Aortic Aneurysm Risk Stratification Advances in Cardiovascular Imaging. (2016). doi:10.1161/CIRCIMAGING.115.003023

28. Weber, J., Haberkorn, U. & Mier, W. Cancer Stratification by Molecular Imaging. *Int. J. Mol. Sci* **16,** 4918–4946 (2015).

29. Tolmachev, V., Tolmachev, V., Stone-Elander, S. & Orlova, A. Radiolabelled receptor-tyrosine-kinase targeting drugs for patient stratification and monitoring of therapy response: prospects and pitfalls. *Lancet Oncol.* **11,** 992–1000 (2010).

30. Jauw, Y. W. S. *et al.* Immuno-positron emission tomography with zirconium-89-labeled monoclonal antibodies in oncology: What can we learn from initial clinical trials? *Front. Pharmacol.* **7,** 1–15 (2016).

31. Maurer, T. *et al.* Diagnostic Efficacy of 68 Gallium-PSMA Positron Emission Tomography Compared to Conventional Imaging for Lymph Node Staging of 130 Consecutive Patients with Intermediate to High Risk Prostate Cancer. **195,** 1436–1443 (2016).

32. Zanzonico, P. *et al.* Positron Emission Tomography (PET)-based Compartmental Modeling of 124 I-A33 Antibody: Quantitative Characterization of Patient-specific Tumor Targeting in Colorectal Cancer HHS Public Access. *Eur J Nucl Med Mol Imaging* **42,** 1700–1706 (2015).

33. Menke-Van Der Houven Van Oordt, C. W. *et al. Zr-cetuximab PET imaging in patients with advanced colorectal cancer. Oncotarget* **6,**

34. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1,** 755–768 (2006).

35. Mcmahon, C. *et al.* Yeast surface display platform for rapid discovery of conformationally selective nanobodies. (2018). doi:10.1038/s41594-018-0028-6

36. Ackerman, M. *et al.* Highly Avid Magnetic Bead Capture : An Efficient Selection Method for de novo Protein Engineering Utilizing Yeast Surface Display. (2009). doi:10.1021/bp.174

37. Liu, G. W. *et al.* Efficient Identification of Murine M2 Macrophage Peptide Targeting Ligands by Phage Display and Next-Generation Sequencing. *Bioconjug. Chem.* **26,** 1811–1817 (2015).

38. Clackson, T., Hoogenboom, H., Griffiths, A. D. & Winter, G. Making antibody fragments using phage display libraries. *Nature* **352,** 624–628 (1991).

39. Fu, B. *et al.* Identification and characterization of a novel phage display-derived peptide with affinity for human brain metastatic breast cancer. *Biotechnol. Lett.* **36,** 2291–301 (2014).

40. Davis, M. I. Comprehensive analysis of kinase inhibitor selectivity. (2011). doi:10.1038/nbt.1990

41. Hanks, S. K., Quinn, A. M. & Hunter, T. *The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic. Source: Science, New Series*

241, (1988).

42. *Signal Transduction by Receptors with Tyrosine Kinase Activity Review Axe1 Ullrich' and Joseph Schlessingert. Cell* **61,** (1990).

43. Rzepecki, A. K., Cheng, H., McLellan, B. N., Arbor, A. & York, N. Cutaneous toxicity as a predictive biomarker for clinical outcome in patients receiving anticancer therapy. *J. Am. Dermatology* **79,** 545–555 (2018).

44. Dienstmann, R., Braña, I., Rodon, J. & Tabernero, J. Toxicity as a biomarker of efficacy of molecular targeted therapies: focus on EGFR and VEGF inhibiting anticancer drugs. *Oncologist* **16,** 1729–40 (2011).

45. Hasinoff, B. B. & Patel, D. The lack of target specificity of small molecule anticancer kinase inhibitors is correlated with their ability to damage myocytes in vitro. (2010). doi:10.1016/j.taap.2010.08.026

46. Arora, A. & Scholar, E. M. Role of Tyrosine Kinase Inhibitors in Cancer Therapy. (2005). doi:10.1124/jpet.105.084145

47. Rosenzweig, S. A. Acquired Resistance to Drugs Targeting Tyrosine Kinases. doi:10.1016/bs.acr.2018.02.003

48. Roskoski, R. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol. Res.* **79,** 34–74 (2014).

49. Liu, R., Li, X. & Lam, K. S. Combinatorial Chemistry in Drug Discovery. doi:10.1016/j.cbpa.2017.03.017

50. Arencibia, J. M. *et al.* An allosteric inhibitor scaffold targeting the PIF-pocket of atypical PKCs. *ACS Chem. Biol.* acschembio.6b00827 (2017). doi:10.1021/acschembio.6b00827

51. Mishra, R. K., Shum, A. K., Platanias, L. C., Miller, R. J. & Schiltz, G. E. Discovery and characterization of novel small-molecule CXCR4 receptor agonists and antagonists. *Sci. Rep.* **6,** 30155 (2016).

52. Friedman, M. *et al.* Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. *J. Mol. Biol.* **376,** 1388–402 (2008).

53. Getmanova, E. *et al.* Antagonists to Human and Mouse Vascular Endothelial Growth Factor Receptor 2 Generated by Directed Protein Evolution In Vitro. *Chem. Biol.* **13,** 549–556 (2006).

54. Kruziki, M. A., Bhatnagar, S., Woldring, D. R., Duong, V. T. & Hackel, B. J. A 45-Amino-Acid Scaffold Mined from the PDB for High-Affinity Ligand Engineering. *Chem. Biol.* **22,** 946–956 (2015).

55. Woldring, D. R., Holec, P. V., Zhou, H. & Hackel, B. J. High-Throughput Ligand

Discovery Reveals a Sitewise Gradient of Diversity in Broadly Evolved Hydrophilic Fibronectin Domains. *PLoS One* **10,** e0138956 (2015).

56.   Dai, Z.-R. *et al.* A practical strategy to design and develop an isoform-specific fluorescent probe for a target enzyme: CYP1A1 as a case study †. (2017). doi:10.1039/c6sc03970g

57.   Pinard, M. A., Mahon, B. & McKenna, R. Probing the Surface of Human Carbonic Anhydrase for Clues towards the Design of Isoform Specific Inhibitors. *Biomed Res Int* **2015,** 453543 (2015).

58.   Ah-San Tang, A., Tiede, C., Hughes, D. J., McPherson, M. & Tomlinson, D. C. *Isolation of isoform-specific binding proteins (Affimers) by phage display using negative selection. Sci. Signal* **10,** (2017).

59.   Telli, M. L., Hunt, S. a., Carlson, R. W. & Guardino, A. E. Trastuzumab-related cardiotoxicity: Calling into question the concept of reversibility. *J. Clin. Oncol.* **25,** 3525–3533 (2007).

60.   Bria, E. *et al.* Cardiotoxicity and incidence of brain metastases after adjuvant trastuzumab for early breast cancer: the dark side of the moon? A meta-analysis of the randomized trials. *Breast Cancer Res. Treat.* **109,** 231–9 (2008).

61.   Shahabi, V. *et al. Gene expression profiling of whole blood in ipilimumab-treated patients for identification of potential biomarkers of immune-related gastrointestinal adverse events*. (2013). doi:10.1186/1479-5876-11-75

62.   Milling, L., Zhang, Y. & Irvine, D. J. Delivering safer immunotherapies for cancer. *Adv. Drug Deliv. Rev.* **114,** 79–101 (2017).

63.   Haanen, J. B. A. G. *et al.* Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. doi:10.1093/annonc/mdx225

64.   Fecher, L., Agarwala, S., Hodi, S. & Weber, J. Ipilimumab and its toxicities: A multidisciplinary approach. *Oncologist* **18,** 733–743 (2013).

65.   Page, M. J. *et al.* Non-invasive imaging and cellular tracking of pulmonary emboli by near-infrared fluorescence and positron-emission tomography. *Nat. Commun.* **6,** 8448 (2015).

66.   Foran, P. *et al.* Targeted contrast agents--an adjunct to whole-body imaging: current concepts. *Semin. Musculoskelet. Radiol.* **14,** 86–94 (2010).

67.   Murrell, D. H., Foster, P. J. & Chambers, A. F. Brain metastases from breast cancer: Lessons from experimental magnetic resonance imaging studies and clinical implications. *J. Mol. Med.* **92,** 5–12 (2014).

68.   Weissleder, R. & Pittet, M. J. Imaging in the era of molecular oncology. *Nature* **452,** 580–9 (2008).

69. Thompson, J., Lawrentschuk, N., Frydenberg, M., Thompson, L. & Stricker, P. The role of magnetic resonance imaging in the diagnosis and management of prostate cancer. *BJU Int.* **112 Suppl,** 6–20 (2013).

70. Aarts, F. *et al.* Pretargeted radioimmunoscintigraphy in patients with primary colorectal cancer using a bispecific anticarcinoembryonic antigen CEA X anti-di-diethylenetriaminepentaacetic acid F(ab')2 antibody. *Cancer* **116,** 1111–7 (2010).

71. Baum, R. P. *et al.* Molecular imaging of HER2-expressing malignant tumors in breast cancer patients using synthetic 111In- or 68Ga-labeled affibody molecules. *J. Nucl. Med.* **51,** 892–7 (2010).

72. Pollock, S. B. *et al.* Highly multiplexed and quantitative cell-surface protein profiling using genetically barcoded antibodies. doi:10.1073/pnas.1721899115

73. Zorniak, M. *et al.* Yeast display biopanning identifies human antibodies targeting glioblastoma stem-like cells OPEN. doi:10.1038/s41598-017-16066-1

74. Tucker-Burden, C. *et al.* Lectins identify glycan biomarkers on glioblastoma-derived cancer stem cells. *Stem Cells Dev.* **21,** 2374–86 (2012).

75. Mohammed, A., Biegert, G., Adamec, J. & Helikar, T. *Identification of potential tissue-specific cancer biomarkers and development of cancer versus normal genomic classifiers*. *Oncotarget* **8,** (2017).

76. Ellwanger, K. *et al.* Highly Specific and Effective Targeting of EGFRvIII-Positive Tumors with TandAb Antibodies. *Front. Oncol.* **7,** 1–17 (2017).

77. Hynes, N. E. & Lane, H. A. ERBB RECEPTORS AND CANCER: THE COMPLEXITY OF TARGETED INHIBITORS. (2005). doi:10.1038/nrc1609

78. Ross, J. S. *et al.* The HER-2 Receptor and Breast Cancer: Ten Years of Targeted Anti-HER-2 Therapy and Personalized Medicine. *Oncologist* **14,** 320–368 (2009).

79. Xia, W. *et al.* Combining lapatinib (GW572016), a small molecule inhibitor of ErbB1 and ErbB2 tyrosine kinases, with therapeutic anti-ErbB2 antibodies enhances apoptosis of ErbB2-overexpressing breast cancer cells. *Oncogene* **24,** 6213–6221 (2005).

80. Kodack, D. P. *et al.* Combined targeting of HER2 and VEGFR2 for effective treatment of HER2-amplified breast cancer brain metastases. *Proc. Natl. Acad. Sci.* **109,** E3119–E3127 (2012).

81. Arteaga, C. L. *et al.* Epidermal Growth Factor Receptors in Human Breast Carcinoma Cells : A Potential Selective Target for Transforming Growth Factor α - Pseudomonas Exotoxin 40 Fusion Protein. 4703–4709 (1994).

82. Ilic, I., Jankovic, S. & Ilic, M. Bevacizumab Combined with Chemotherapy Improves Survival for Patients with Metastatic Colorectal Cancer: Evidence from Meta Analysis. (2016). doi:10.1371/journal.pone.0161912

83. Chibaudel, B., Tournigand, C., Andre, T. & de Gramont, a. Therapeutic strategy in unresectable metastatic colorectal cancer. *Ther. Adv. Med. Oncol.* **4,** 75–89 (2012).

84. Reichert, J. M. Antibodies to watch in 2016. *MAbs* **8,** 197–204 (2016).

85. Wilson, R. A. M., Evans, T. R. J., Fraser, A. R. & Nibbs, R. J. B. Immune checkpoint inhibitors: new strategies to checkmate cancer. *Clin. Exp. Immunol.* **191,** 133–148 (2018).

86. Park, J. W., Chang, H. J., Kim, B. C., Yeo, H. Y. & Kim, D. Y. Clinical validity of tissue carcinoembryonic antigen expression as ancillary to serum carcinoembryonic antigen concentration in patients curatively resected for colorectal cancer. *Color. Dis.* **15,** 503–511 (2013).

87. Tiernan, J. P. *et al.* Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting. *Br. J. Cancer* **108,** (2013).

88. Saito, G. *et al.* Relation between Carcinoembryonic Antigen Levels in Colon Cancer Tissue and Serum Carcinoembryonic Antigen Levels at Initial Surgery and Recurrence. *Oncology* (2016). doi:10.1159/000447062

89. Guadagni, F. *et al.* Quantitative analysis of CEA expression in colorectal adenocarcinoma and serum: Lack of correlation. *Int. J. Cancer* **72,** 949–954 (1997).

90. WEILI YANG, YONGSHENG LUO, SHUANGCHENG HU, YING LI & QING LIU. Value of combined detection of serum carcino-embryonic antigen, carbohydrate antigen 19-9 and cyclooxygenase-2 in the diagnosis of colorectal cancer. *Oncol. Lett.* **16,** 1551–1556 (2018).

91. Zhou, Z., Wu, X., Kresak, A., Griswold, M. & Lu, Z.-R. Peptide targeted tripod macrocyclic Gd(III) chelates for cancer molecular MRI. *Biomaterials* **34,** 7683–7693 (2013).

92. Pu, F. *et al.* GRPR-targeted Protein Contrast Agents for Molecular Imaging of Receptor Expression in Cancers by MRI OPEN. (2015). doi:10.1038/srep16214

93. Pais, A. & Degani, H. Estrogen Receptor-Targeted Contrast Agents for Molecular Magnetic Resonance Imaging of Breast Cancer Hormonal Status. *Front. Oncol.* **6,** 100 (2016).

94. Caplan, M. R. & Rosca, E. V. Targeting drugs to combinations of receptors: a modeling analysis of potential specificity. *Ann. Biomed. Eng.* **33,** 1113–24 (2005).

95. Rosca, E. V, Stukel, J. M., Gillies, R. J., Vagner, J. & Caplan, M. R. Specificity and mobility of biomacromolecular, multivalent constructs for cellular targeting. *Biomacromolecules* **8,** 3830–3835 (2007).

96. Shewmake, T. a, Solis, F. J., Gillies, R. J. & Caplan, M. R. Effects of linker length

and flexibility on multivalent targeting. *Biomacromolecules* **9,** 3057–64 (2008).

97. Dubacheva, G. V, Curk, T., Auzély-velty, R., Frenkel, D. & Richter, R. P. Designing multivalent probes for tunable superselective targeting. (2015). doi:10.1073/pnas.1500622112

98. Wang, S. & Dormidontova, E. E. Nanoparticle targeting using multivalent ligands: computer modeling. *Soft Matter* **7,** 4435 (2011).

99. Martinez-Veracoechea, F. J. & Frenkel, D. Designing super selectivity in multivalent nano-particle binding. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 10963–8 (2011).

100. Xu, L. *et al.* Enhanced targeting with heterobivalent ligands. *Mol. Cancer Ther.* **8,** 2356–2365 (2009).

101. Xu, L. *et al.* Heterobivalent ligands target cell-surface receptor combinations in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21295–300 (2012).

102. Ding, L. *et al.* Small Sized EGFR1 and HER2 Specific Bifunctional Antibody for Targeted Cancer Therapy. *Theranostics* **5,** (2015).

103. Mazor, Y. *et al.* Enhanced tumor-targeting selectivity by modulating bispecific antibody binding affinity and format valence. *Sci. Rep.* **7,** 40098 (2017).

104. Mazor, Y. *et al.* Insights into the molecular basis of a bispecific antibody's target selectivity. *MAbs* **7,** 461–469 (2015).

105. Mazor, Y. *et al.* Improving target cell specificity using a novel monovalent bispecific IgG design. 37–41 (2015). doi:10.1080/19420862.2015.1007816

106. Hackel, B. J., Kapila, A. & Wittrup, K. D. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *J. Mol. Biol.* **381,** 1238–52 (2008).

107. Li, G. *et al.* Construction of a linker library with widely controllable flexibility for fusion protein design. *Appl. Microbiol. Biotechnol.* **100,** 215–225 (2016).

108. Thurber, G. M. & Dane Wittrup, K. A mechanistic compartmental model for total antibody uptake in tumors. *J. Theor. Biol.* **314,** 57–68 (2012).

109. Thurber, G. M., Schmidt, M. M. & Wittrup, K. D. Factors determining antibody distribution in tumors. *Trends Pharmacol. Sci.* **29,** 57–61 (2008).

110. Hackel, B. J., Neil, J. R., White, F. M. & Wittrup, K. D. Epidermal growth factor receptor downregulation by small heterodimeric binding proteins. *Protein Eng. Des. Sel.* **25,** 47–57 (2012).

111. Mamluk, R. *et al.* Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2. *MAbs* **2,** 199–208 (2010).

112. Tolcher, A. W. *et al.* Phase I and pharmacokinetic study of CT-322 (BMS-844203), a targeted Adnectin inhibitor of VEGFR-2 based on a domain of human fibronectin. *Clin. Cancer Res.* **17,** 363–71 (2011).

113. Natarajan, A., Hackel, B. J. & Gambhir, S. S. A Novel Engineered Anti-CD20 Tracer Enables Early Time PET Imaging in a Humanized Transgenic Mouse Model of B-cell Non-Hodgkins Lymphoma. *Clin. Cancer Res.* **19,** 6820–9 (2013).

114. Hackel, B. J., Kimura, R. H. & Gambhir, S. S. Use of 64 Cu-labeled Fibronectin Domain with EGFR-. **263,** 179–188 (2012).

115. Pirie, C. M., Hackel, B. J., Rosenblum, M. G. & Wittrup, K. D. Convergent potency of internalized gelonin immunotoxins across varied cell lines, antigens, and targeting moieties. *J. Biol. Chem.* **286,** 4165–4172 (2011).

116. Mack, K., Rüger, R., Fellermeier, S., Seifert, O. & Kontermann, R. E. Dual Targeting of Tumor Cells with Bispecific Single-Chain Fv-Immunoliposomes. *Antibodies* **1,** 199–214 (2012).

117. Pirie, C. M., Hackel, B. J., Rosenblum, M. G. & Wittrup, K. D. Convergent potency of internalized gelonin immunotoxins across varied cell lines, antigens, and targeting moieties. *J. Biol. Chem.* **286,** 4165–72 (2011).

118. VanAntwerp, J. J. & Wittrup, K. D. Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.* **16,** 31–7 (2000).

119. Cho, Y. K. & Shusta, E. V. Antibody library screens using detergent-solubilized mammalian cell lysates as antigen sources. *Protein Eng. Des. Sel.* **23,** 567–577 (2010).

120. Stern, L. A. *et al.* Geometry and expression enhance enrichment of functional yeast-displayed ligands via cell panning. *Biotechnol. Bioeng.* (2016). doi:10.1002/bit.26001

121. Cunningham, B. C. & Wells, J. a. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244,** 1081–5 (1989).

122. Rathi, P. C. *et al.* Predicting 'hot' and 'warm' spots for fragment binding. *J. Med. Chem.* acs.jmedchem.7b00366 (2017). doi:10.1021/acs.jmedchem.7b00366

123. Binz, H. K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23,** 1257–68 (2005).

124. Hogan Jr., J. C. Combinatorial chemistry in drug discovery. *Nat. Biotechnol.* **15,** 328–330 (1997).

125. Liu, R., Li, X. & Lam, K. S. Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **38,** 117–126 (2017).

126. Poreba, M. *et al.* Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates. *Cell Death Differ.* **21,** 1482–1492 (2014).

127. Fauman, E. B., Rai, B. K. & Huang, E. S. Structure-based druggability assessmentâ€"identifying suitable targets for small molecule therapeutics. *Curr. Opin. Chem. Biol.* **15,** 463–468 (2011).

128. Orlova, A., Feldwisch, J., Abrahmsén, L. & Tolmachev, V. Update: affibody molecules for molecular imaging and therapy for cancer. *Cancer Biother. Radiopharm.* **22,** 573–84 (2007).

129. Bhatt, A. *et al.* Structure–Activity Relationships of Benzenesulfonamide-Based Inhibitors towards Carbonic Anhydrase Isoform Specificity. *ChemBioChem* **18,** 213–222 (2017).

130. Li, Z. *et al.* Structure-Guided Design of Novel, Potent, and Selective Macrocyclic Plasma Kallikrein Inhibitors. *ACS Med. Chem. Lett.* (2017). doi:10.1021/acsmedchemlett.6b00384

131. Kim, J. W. & Cochran, J. R. Targeting ligand–receptor interactions for development of cancer therapeutics. *Curr. Opin. Chem. Biol.* **38,** 62–69 (2017).

132. Harvey, A. L. Natural products in drug discovery. *Drug Discov. Today* **13,**

133. Josan, J. S. *et al.* Cell-Specific Targeting by Heterobivalent Ligands. *Bioconjug. Chem.* **22,** 1270–1278 (2011).

134. Roskoski, R. Invited Review A historical overview of protein kinases and their targeted small molecule inhibitors. *Pharmacol. Res.* **100,** 1–23 (2015).

135. Hoelder, S., Clarke, P. A. & Workman, P. Discovery of small molecule cancer drugs: Successes, challenges and opportunities. (2012). doi:10.1016/j.molonc.2012.02.004

136. Bollag, G. *et al.* Vemurafenib: the first drug approved for BRAF-mutant cancer. (2012). doi:10.1038/nrd3847

137. Guo, W., Wisniewski, J. A. & Ji, H. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg. Med. Chem. Lett.* **24,** 2546–2554 (2014).

138. Smith, M. C. & Gestwicki, J. E. Features of Protein-Protein Interactions that Translate into Potent Inhibitors: Topology, Surface Area and Affinity. doi:10.1017/erm.2012.10

139. Wurz, R. P. *et al.* A "Click Chemistry Platform" for the Rapid Synthesis of Bispecific Molecules for Inducing Protein Degradation. *J. Med. Chem.* acs.jmedchem.6b01781 (2017). doi:10.1021/acs.jmedchem.6b01781

140. Ueda, S. *et al.* Identification of novel non-peptide CXCR4 antagonists by ligand-based design approach. (2008). doi:10.1016/j.bmcl.2008.05.092

141. Schulz, M. M. P. *et al.* Phenotype-based high-content chemical library screening identifies statins as inhibitors of in vivo lymphangiogenesis. *Proc. Natl. Acad. Sci.* **109,** E2665–E2674 (2012).

142. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. *Opportunities and challenges in phenotypic drug discovery: an industry perspective*. (2017). doi:10.1038/nrd.2017.111

143. Myers, S. H., Temps, C., Houston, D. R., Brunton, V. G. & Unciti-Broceta, A. Development of Potent Inhibitors of Receptor Tyrosine Kinases by Ligand-Based Drug Design and Target-Biased Phenotypic Screening. (2018). doi:10.1021/acs.jmedchem.7b01605

144. Avesar, J. *et al.* Rapid phenotypic antimicrobial susceptibility testing using nanoliter arrays. (2017). doi:10.1073/pnas.1703736114

145. Mccafferty, J. Phenotypic Directed Antibody Selection. *Chem. Biol.* **21,** 170–171 (2014).

146. Koide, A. *et al. Exploring the capacity of minimalist protein interfaces: interface energetics and affinity maturation to picomolar K D of a single-domain antibody with a flat paratope*.

147. Verdine, G. L. & Walensky, L. D. The challenge of drugging undruggable targets in cancer: Lessons learned from targeting BCL-2 family members. *Clin. Cancer Res.* **13,** 7264–7270 (2007).

148. Phanie Pé Rot, S., Sperandio, O., Miteva, M. A., Camproux, A.-C. & Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* **15,** 656–667 (2010).

149. Citation: Gao, M. & Skolnick, J. A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. *PLoS Comput Biol* **9,** 1003302 (2013).

150. Walter, P., Metzger, J., Thiel, C. & Helms, V. Predicting where Small Molecules Bind at Protein-Protein Interfaces. *PLoS One* **8,** 58583 (2013).

151. Koide, S., Koide, A. & Lipovšek, D. Target-Binding Proteins Based on the 10th Human Fibronectin Type III Domain (10 Fn3). (2012). doi:10.1016/B978-0-12-396962-0.00006-9

152. Bloom, L. & Calabro, V. FN3: a new protein scaffold reaches the clinic. *Drug Discov. Today* **14,**

153. Koide, a, Bailey, C. W., Huang, X. & Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* **284,** 1141–51 (1998).

154. Sriram, K. & Insel, P. A. GPCRs as targets for approved drugs: How many targets and how many drugs? *Mol. Pharmacol. Fast Forw.* (2018). doi:10.1124/mol.117.111062

155. Hanson, M. A. & Stevens, R. C. Discovery of New GPCR Biology: One Receptor Structure at a Time. *Structure* **17,** 8–14 (2009).

156. Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21,** 541–551 (2011).

157. Fredriksson, R. *et al. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints.* (2003).

158. Hauser, A. S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* **172,** 41–54.e19 (2018).

159. Carbonic anhydrases (CAs; also known as carbonate dehydratases EC 4.2.1. (2008). doi:10.1038/nrd2467

160. Supuran, C. T. Journal of Enzyme Inhibition and Medicinal Chemistry Structure-based drug discovery of carbonic anhydrase inhibitors. (2012). doi:10.3109/14756366.2012.672983

161. Gao, B. B. *et al.* Extracellular carbonic anhydrase mediates hemorrhagic retinal and cerebral vascular permeability through prekallikrein activation. *Nat. Med.* **13,** 181–188 (2007).

162. Kaur, I. P., Smitha, R., Aggarwal, D. & Kapil, M. *Acetazolamide: future perspective in topical glaucoma therapeutics.*

163. Mcdonald, P. C., Winum, J.-Y., Supuran, C. T. & Dedhar, S. *Recent Developments in Targeting Carbonic Anhydrase IX for Cancer Therapeutics. Oncotarget* **3,** (2012).

164. Alterio, V. *et al. Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX.*

165. S, P.-T. *et al.* Neutralization of tumor acidity improves antitumor responses to immunotherapeutic interventions. (2015). doi:10.1158/0008-5472.CAN-15-1743

166. Mora, M. J. *et al.* Characterization, dissolution and in vivo evaluation of solid acetazolamide complexes. *Carbohydr. Polym.* **98,** 380–390 (2013).

167. Krebs, H. A. *Inhibition of Carbonic Anhydrase by Sulphonamides.* **43,**

168. Dudutienė, V. *et al.* Discovery and Characterization of Novel Selective Inhibitors of Carbonic Anhydrase IX. (2014). doi:10.1021/jm501003k

169. Ltd, I. ; I. U. Intrinsic thermodynamics of sulfonamide inhibitor binding to human carbonic anhydrases I and II. *J. Enzyme Inhib. Med. Chem.* **30,** 204–211 (2015).

170. Gaspari, R. *et al.* Kinetic and Structural Insights into the Mechanism of Binding of Sulfonamides to Human Carbonic Anhydrase by Computational and Experimental Studies. (2015). doi:10.1021/acs.jmedchem.5b01643

171. Coleman, J. E. Metal Ion Dependent Binding of Sulphonamide to Carbonic Anhydrase. *Nature* **214,** 193–194 (1967).

172. Jo, M. & Jung, S. T. Engineering therapeutic antibodies targeting G-protein–coupled receptors. *Exp. Mol. Med.* 207 (2016). doi:10.1038/emm.2015.105

173. Orwenyo, J. *et al.* Systematic Synthesis and Binding Study of HIV V3 Glycopeptides Reveal the Fine Epitopes of Several Broadly Neutralizing Antibodies. *ACS Chem. Biol.* acschembio.7b00319 (2017). doi:10.1021/acschembio.7b00319

174. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. a. & Kim, P. S. Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **99,** 103–115 (1999).

175. Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33,** 17–27 (2012).

176. Corin, K., Baaske, P., Ravel, D. B., Song, J. & Brown, E. A Robust and Rapid Method of Producing Soluble, Stable, and Functional G-Protein Coupled Receptors. *PLoS One* **6,** 23036 (2011).

177. McCusker, E. C., Bane, S. E., O'Malley, M. A. & Robinson, A. S. Heterologous GPCR expression: A bottleneck to obtaining crystal structures. *Biotechnol. Prog.* **23,** 540–547 (2007).

178. Link, A. J., Skretas, G., Strauch, E.-M., Chari, N. S. & Georgiou, G. Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli. *Protein Sci.* **17,** 1857–1863 (2008).

179. Costa-Neto, C. M., Mizanur, R. M., Jaakola, V.-P., Milí, D. & Veprintsev, D. B. Large-scale production and protein engineering of G protein-coupled receptors for structural studies. (2015). doi:10.3389/fphar.2015.00066

180. Liang, Z. *et al.* CXCR4/CXCL12 axis promotes VEGF-mediated tumor angiogenesis through Akt signaling pathway. (2007). doi:10.1016/j.bbrc.2007.05.182

181. Folkins, C. *et al.* Glioma tumor stem-like cells promote tumor angiogenesis and vasculogenesis via vascular endothelial growth factor and stromal-derived factor 1. *Cancer Res.* **69,** 7243–51 (2009).

182. Levoye, A., Balabanian, K., Baleux, F., Bachelerie, F. & Lagane, B. CXCR7 heterodimerizes with CXCR4 and regulates CXCL12-mediated G protein signaling. *Blood* **113,** 6085–6093 (2009).

183. Polacheck, W. J., Zervantonakis, I. K. & Kamm, R. D. Tumor cell migration in complex microenvironments. *Cell. Mol. Life Sci.* 1335–1356 (2012). doi:10.1007/s00018-012-1115-1

184. Berg, C. *et al.* Inhibition of HIV fusion by small molecule agonists through efficacy-engineering of CXCR4. *ACS Chem. Biol.* acschembio.8b00061 (2018). doi:10.1021/acschembio.8b00061

185. Nimmagadda, S. *et al.* Molecular imaging of CXCR4 receptor expression in human cancer xenografts with [64Cu]AMD3100 positron emission tomography. *Cancer Res.* **70,** 3935–3944 (2010).

186. Ramsey, D. M. & Mcalpine, S. R. Halting metastasis through CXCR4 inhibition. *Bioorg. Med. Chem. Lett.* **23,** 20–25 (2013).

187. Vass, M. *et al.* Chemical Diversity in the G Protein-Coupled Receptor Superfamily. (2018). doi:10.1016/j.tips.2018.02.004

188. Wesley K Kroeze1,2,5, Maria F Sassano1,2,5, Xi-Ping Huang1,2,5, Katherine Lansu1, J. D. M. & Patrick M Giguère1, N. S. & B. L. R. PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.*

189. Coudrat, T., Christopoulos, A., Sexton, P. M. & Wootten, D. Structural features embedded in G protein-coupled receptor co-crystal structures are key to their success in virtual screening. (2017). doi:10.1371/journal.pone.0174719

190. Hubner, H. *et al.* Structure-guided development of heterodimer-selective GPCR ligands. *Nat. Commun.* **7,** 1–12 (2016).

191. Hackel, B. J. & Wittrup, K. D. Yeast Surface Display in Protein Engineering and Analysis. **2,** 621–649 (2008).

192. Schlapschy, M. *et al.* PASylation: A biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. *Protein Eng. Des. Sel.* **26,** 489–501 (2013).

193. Hackel, B. J. & Wittrup, K. D. The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin G binders from the fibronectin scaffold. *Protein Eng. Des. Sel.* **23,** 211–219 (2010).

194. Park, S.-H. *et al.* Isolation and Characterization of a Monobody with a Fibronectin Domain III Scaffold That Specifically Binds EphA2. *PLoS One* **10,** e0132976 (2015).

195. Mendler, C. T. *et al.* High contrast tumor imaging with radio-labeled antibody Fab fragments tailored for optimized pharmacokinetics via PASylation. *MAbs* **7,** 96–109 (2015).

196. Sippel, K. H. *et al.* Structural Biology and Crystallization Communications High-

resolution structure of human carbonic anhydrase II complexed with acetazolamide reveals insights into inhibitor drug design. 992 doi:10.1107/S1744309109036665

197.   Mungalpara, J. *et al.* Rational Design of Conformationally Constrained Cyclopentapeptide Antagonists for C - X - C Chemokine Receptor 4 (CXCR4). **4,** (2012).

198.   Kim, J. *et al.* Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. *J. Clin. Oncol.* **23,** 2744–2753 (2005).

199.   Wu, Q. *et al.* Extracellular calcium increases CXCR4 expression on bone marrow-derived cells and enhances pro-angiogenesis therapy. *J. Cell. Mol. Med.* **13,** 3764–3773 (2009).

200.   Lombardi, L. *et al.* Chemokine receptor CXCR4: Role in gastrointestinal cancer. *Crit. Rev. Oncol. / Hematol.* **88,** 696–705 (2013).

201.   Oishi, S. *et al.* Development of Novel CXC Chemokine Receptor 7 (CXCR7) Ligands: Selectivity Switch from CXCR4 Antagonists with a Cyclic Pentapeptide Scaffold. *J. Med. Chem.* **58,** 5218–5225 (2015).

202.   Gourni, E. *et al.* PET of CXCR4 Expression by a 68Ga-Labeled Highly Specific Targeted Contrast Agent. *J. Nucl. Med.* **52,** 1803–1810 (2011).

203.   De Simone, G. & Supuran, C. T. Carbonic anhydrase IX: Biochemical and crystallographic characterization of a novel antitumor target. *BBA - Proteins Proteomics* **1804,** 404–409 (2009).

204.   Fang, Y., Chu, T. H., Ackerman, M. E. & Griswold, K. E. Going native: Direct high throughput screening of secreted full-length IgG antibodies against cell membrane proteins. *MAbs* **9,** 1253–1261 (2017).

205.   Kimura, R. H. *et al.* Pharmacokinetically stabilized cystine knot peptides that bind alpha-v-beta-6 integrin with single-digit nanomolar affinities for detection of pancreatic cancer. *Clin. Cancer Res.* **18,** 839–49 (2012).

206.   Luker, K. E. *et al.* Scavenging of CXCL12 by CXCR7 promotes tumor growth and metastasis of CXCR4-positive breast cancer cells. *Oncogene* (2012).

207.   Thelen, M. & Thelen, S. CXCR7, CXCR4 and CXCL12: An eccentric trio? *J. Neuroimmunol.* **198,** 9–13 (2008).

208.   Wang, L. V & Hu, S. Photoacoustic Tomography: In Vivo Imaging from Organelles to Organs. doi:10.1126/science.1216210

209.   Liu, Y., Nie, L. & Chen, X. Photoacoustic Molecular Imaging: From Multiscale Biomedical Applications Towards Early-Stage Theranostics. *Trends Biotechnol.* **34,** 420–433 (2016).

210. Weber, J., Beard, P. C. & Bohndiek, S. E. Contrast agents for molecular photoacoustic imaging. *Nat. Methods* **13,** 639–650 (2016).

211. Nie, L. & Chen, X. Structural and functional photoacoustic molecular tomography aided by emerging contrast agents. *Chem. Soc. Rev.* **43,** 7132–7170 (2014).

212. Levi, J. *et al.* Design, Synthesis, and Imaging of an Activatable Photoacoustic Probe. *J. Am. Chem. Soc.* **132,** 11264–11269 (2010).

213. Levi, J. *et al.* Molecular photoacoustic imaging of follicular thyroid carcinoma. *Clin. Cancer Res.* **19,** 1494–502 (2013).

214. Jansen, K., van Soest, G. & van der Steen, A. F. W. Intravascular photoacoustic imaging: a new tool for vulnerable plaque identification. *Ultrasound Med. Biol.* **40,** 1037–1048 (2014).

215. Dragulescu-Andrasi, A., Kothapalli, S.-R., Tikhomirov, G. A., Rao, J. & Gambhir, S. S. Activatable Oligomerizable Imaging Agents for Photoacoustic Imaging of Furin-Like Activity in Living Subjects. *J. Am. Chem. Soc.* **135,** 11015–11022 (2013).

216. Yang, K. *et al.* Visualization of Protease Activity In Vivo Using an Activatable Photo-Acoustic Imaging Probe Based on CuS Nanoparticles. *Theranostics* **4,** 134–141 (2014).

217. Li, H., Zhang, P., Smaga, L. P., Hoffman, R. A. & Chan, J. Photoacoustic Probes for Ratiometric Imaging of Copper(II). *J. Am. Chem. Soc.* **137,** 15628–15631 (2015).

218. Pu, K. *et al.* Semiconducting polymer nanoparticles as photoacoustic molecular imaging probes in living mice. *Nat. Nanotechnol.* **9,** 233 (2014).

219. Miao, Q., Lyu, Y., Ding, D. & Pu, K. Semiconducting Oligomer Nanoparticles as an Activatable Photoacoustic Probe with Amplified Brightness for In Vivo Imaging of pH. *Adv. Mater.* **28,** 3662–3668 (2016).

220. Song, J. *et al.* "Smart" gold nanoparticles for photoacoustic imaging: an imaging contrast agent responsive to the cancer microenvironment and signal amplification via pH-induced aggregation. *Chem. Commun.* **52,** 8287–8290 (2016).

221. Ng, K. K. *et al.* Stimuli-Responsive Photoacoustic Nanoswitch for in Vivo Sensing Applications. *ACS Nano* **8,** 8363–8373 (2014).

222. Razansky, D. *et al.* Multispectral optoacoustic tomography of matrix metalloproteinase activity in vulnerable human carotid plaques. *Mol. imaging Biol. MIB Off. Publ. Acad. Mol. Imaging* **14,** 277–285 (2012).

223. Ennis, B. W. & Matrisian, L. M. Matrix degrading metalloproteinases. *J. Neurooncol.* **18,** 105–109 (1994).

224. Gialeli, C., Theocharis, A. D. & Karamanos, N. K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS J.* **278,** 16–27 (2011).

225. Overall, C. M. & Kleifeld, O. Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. **6,** 227–239 (2006).

226. Levi, J. *et al.* Design, synthesis, and imaging of an activatable photoacoustic probe. *J. Am. Chem. Soc.* **132,** 11264–9 (2010).

227. van Hoek, A., Vervoort, J. & Visser, A. J. W. G. A subnanosecond resolving spectrofluorimeter for the analysis of protein fluorescence kinetics. *J. Biochem. Biophys. Methods* **7,** 243–254 (1983).

228. Beechem, J. M. & Brand, L. Time-resolved fluorescence of proteins. *Annu. Rev. Biochem.* **54,** 43–71 (1985).

229. Willis, K. J., Szabo, A. G., Zuker, M., Ridgeway, J. M. & Alpert, B. Fluorescence Decay Kinetics of the Tryptophyl Residues of Myoglobin: Effect of Heme Ligation and Evidence for Discrete Lifetime Components. *Biochemistry* **29,** 5270–5275 (1990).

230. Dimitrow, E. *et al.* Spectral fluorescence lifetime detection and selective melanin imaging by multiphoton laser tomography for melanoma diagnosis. *Exp. Dermatol.* **18,** 509–515 (2009).

231. Zemp, A. F. and P. S. and W. S. and R. J. Lifetime-weighted photoacoustic imaging. *J. Opt.* **18,** 124001 (2016).

232. O'Connor, A. E., Gallagher, W. M. & Byrne, A. T. Porphyrin and Nonporphyrin Photosensitizers in Oncology: Preclinical and Clinical Advances in Photodynamic Therapy. *Photochem. Photobiol.* **85,** 1053–1074 (2009).

233. Morgounova, E., Shao, Q., Hackel, B. J., Thomas, D. D. & Ashkenazi, S. Photoacoustic lifetime contrast between methylene blue monomers and self-quenched dimers as a model for dual-labeled activatable probes. *J. Biomed. Opt.* **18,** 56004 (2013).

234. Jiang, T. *et al.* Tumor imaging by means of proteolytic activation of cell-penetrating peptides. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 17867–72 (2004).

235. van Duijnhoven, S. M. J., Robillard, M. S., Nicolay, K. & Grüll, H. Tumor targeting of MMP-2/9 activatable cell-penetrating imaging probes is caused by tumor-independent activation. *J. Nucl. Med.* **52,** 279–86 (2011).

236. Rood, M. T. M. *et al.* MMP-2/9-Specific Activatable Lifetime Imaging Agent. *Sensors (Basel).* **15,** 11076–11091 (2015).

237. Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains.

FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. **14,** 231–241 (2000).

238. Toth, M. & Fridman, R. Assessment of Gelatinases (MMP-2 and MMP-9 by Gelatin Zymography. *Methods Mol. Med.* **57,** 163–174 (2001).

239. Mishra, A., Behera, R. K., Behera, P. K., Mishra, B. K. & Behera, G. B. Cyanines during the 1990s: A Review. *Chem. Rev.* **100,** 1973–2012 (2000).

240. Iyer, A. K., Khaled, G., Fang, J. & Maeda, H. Exploiting the enhanced permeability and retention effect for tumor targeting. *Drug Discov. Today* **11,** 812–818 (2006).

241. Jonsson, A., Hjalmarsson, C., Falk, P. & Ivarsson, M.-L. Levels of matrix metalloproteinases differ in plasma and serum - aspects regarding analysis of biological markers in cancer. *Br. J. Cancer* **115,** 703–706 (2016).

242. Nagase, H., Visse, R. & Murphy, G. Structure and function of matrix metalloproteinases and TIMPs. *Cardiovasc. Res.* **69,** 562–573 (2006).

Appendix

A.1.1 Derivation of association rate enhancement via linker
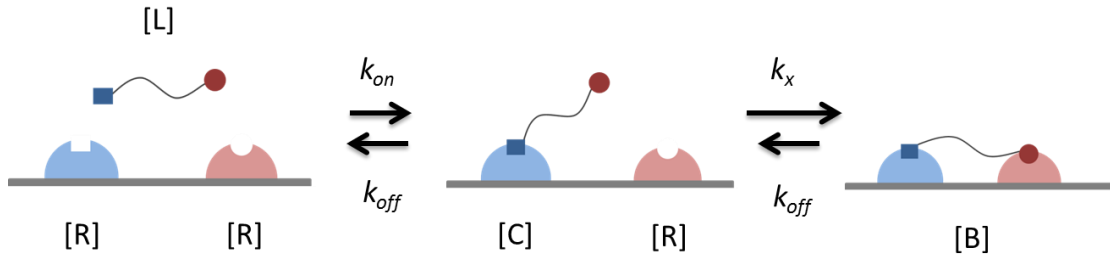
Figure A.1: Events leading to the cross-linking of two receptors by the AND gate. The change in K_x is due to the tethering of the molecule to the cell surface, increasing the effective receptor concentration for that binder.

To evaluate the effect of the linker length on the effective affinity, K_x, of the bivalent linker, the following schematic is used to describe the effect of tethering on the second binding event. A main assumption is that the linker does not affect the monovalent receptor:ligand interactions, i.e. the binding strength to the target is unchanged. However, the tethering of the ligand does result in an increase in the effective concentration of receptors that it can interact with (Fig A.2). To calculate this effective receptor concentration for a fully flexible linker of length L, the accessible volume is described by Equation A.1.1.
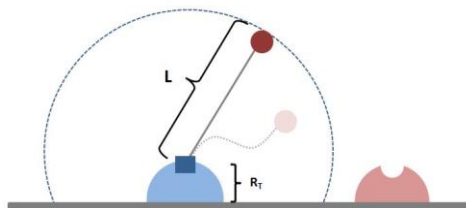
Figure A.2: Geometric model for linker effect on the second binding event's equilibrium constant, K_x.

Eqn (A.1.1) $V_{access} = \int_{-R_T}^{L} \pi(L^2 - x^2)\, dx = \frac{\pi}{3}(2L^3 + 3R_T L^2 - R_T)$

Within V_{access}, the number of accessible receptors for the second binding event is described by the area on the membrane reachable by L minus the area occupied by the first bound receptor

Eqn (A.1.2) $A_{access} = \pi L^2 - \pi(2R_T)^2$

The number of accessible receptors within that area is described by

Eqn (A.1.3) $N_{access} = \frac{N_T}{4\pi r_{cell}^2} A_{access} = \frac{N_T}{r_{cell}^2}\left(\frac{L^2}{4} - R_T^2\right)$

By dividing equation A.3 by A.1 and inserting the non-dimensional parameter η=L/R$_T$, the concentration in the area swept by the ligand, or R$_{access}$, is found to be

Eqn (A.1.4) $R_{access} = \frac{N_T}{\pi r_{cell}^2 R_T} \frac{\frac{1}{4}\eta^2 - 1}{\frac{2}{3}\eta^3 + \eta^2 - \frac{1}{3}}$

A.1.2 Calculation of ligand in vivo biodistribution

Table A.1.1 – Variables

L_n	Ligand in normal tissue
L_t	Ligand in tumor interstitium
L_i	Ligand inside cell
L_d	Ligand degraded in cell
L_p	Ligand in plasma
R_t	Receptor on cell surface
R_i	Receptor inside cell
C_t	1st receptor:ligand complex on cell
C_i	1st receptor:ligand complex inside cell
M_t	2nd receptor:ligand complex on cell
M_i	2nd receptor:ligand complex inside cell
B_t	Bivalent ligand:two receptor complex on cell
B_i	Bivalent ligand:two receptor complex inside cell

Table A.1.2 - Governing equations of ligand distribution

Eqn. A.1.5	$$\frac{dL_n}{dt} = \frac{2 * P * r_{cap}}{r_{krogh}^2}(L_p - \frac{L_n}{\epsilon}) - \frac{2 * P * k_{lymph} * r_{cap}}{r_{krogh}^2}(\frac{L_n}{\epsilon}) - \frac{6 * D * \epsilon}{r_{tumor}^2}(\frac{L_n}{\epsilon} - \frac{L_t}{\epsilon})$$
Eqn. A.1.6	$$\frac{dL_t}{dt} = \frac{2 * P * r_{cap}}{r_{krogh}^2}(L_p - \frac{L_t}{\epsilon}) - \frac{2 * P * k_{lymph} * r_{cap}}{r_{krogh}^2}(\frac{L_t}{\epsilon}) + \frac{6 * D * \epsilon}{r_{tumor}^2}(\frac{L_n}{\epsilon}$$ $$- \frac{L_t}{\epsilon}) - k_{on} * R_t * \frac{L_t}{\epsilon} + k_{off} * C_t - k_{on,M} * M_t * \frac{L_t}{\epsilon} + k_{off,M} * D_t$$
Eqn. A.1.7	$$\frac{dL_p}{dt} = -k_{clear} * L_p$$
Eqn. A.1.8	$$\frac{dL_i}{dt} = k_{off,I} * C_i - \frac{k_{on,I} * L_i * R_i}{\epsilon} - k_{deg,L} * L_i + k_{off,MI} * D_i - \frac{k_{on,MI} * L_i * M_I}{\epsilon}$$
Eqn. A.1.9	$$\frac{dL_d}{dt} = k_{deg,L} * L_i + k_{deg} * C_i - k_{elim} * L_D + k_{deg,M} * D_i$$
Eqn. A.1.10	$$\frac{dC_t}{dt} = k_{on} * L_t * \frac{R_t}{\epsilon} - k_{off} * C_i - k_{endo} * C_t - k_{x,M} * C_t * M_t + k_{off,M} * B_i$$
Eqn. A.1.11	$$\frac{dC_i}{dt} = k_{endo} * C_t - k_{rec} * C_i - k_{off,I} * C_i + k_{on,I} * L_i * \frac{R_i}{\epsilon} - k_{deg} * C_i - k_{x,MI} * C_i$$ $$* M_i + k_{off,MI} * B_i$$
Eqn. A.1.12	$$\frac{dR_t}{dt} = k_{syn} - k_{on} * L_t * {R_t}/{\epsilon} + k_{off} * C_t - k_{endo} * R_t + k_{rec} * R_i - k_{x,C} * D_t * R_t$$
Eqn. A.1.13	$$\frac{dR_i}{dt} = k_{endo} * R_t - k_{rec} * R_i - k_{deg} * R_i + k_{off,I} * C_i - k_{on,I} * L_i * \frac{R_i}{\epsilon} - k_{x,I} * D_i$$ $$* R_i + k_{off,I} * B_i$$
Eqn. A.1.14	$$\frac{dD_t}{dt} = k_{on,M} * L_t * \frac{M_t}{\epsilon} - k_{off,M} * D_t - k_{endo,M} * D_t - k_{x,C} * D_t * R_t + k_{off} * B_t$$ $$+ k_{rec,M} * D_i$$
Eqn. A.1.15	$$\frac{dD_i}{dt} = k_{endo,M} * D_t - k_{rec,M} * D_i - k_{off,MI} * D_i + k_{on,MI} * L_i * \frac{M_i}{\epsilon} - k_{deg,M} * D_i$$ $$- k_{x,MI} * C_i * R_i + k_{off,MI} * B_i$$
Eqn A.1.16	$$\frac{dM_t}{dt} = k_{syn,M} - k_{on,M} * L_t * {M_t}/{\epsilon} + k_{off,M} * D_t - k_{endo,M} * M_t + k_{rec,M} * M_i$$ $$- k_{x,M} * C_t * M_t + k_{off,M} * B_t$$

Eqn A.1.17	$$\frac{dM_i}{dt} = k_{endo} * M_t - k_{rec} * M_i - k_{deg} * M_i + k_{off,I} * D_i - k_{on,MI} * L_i * \frac{M_i}{\epsilon} - k_{x,I} * D_i * R_i + k_{off,I} * B_i$$
Eqn A.1.18	$$\frac{dB_i}{dt} = k_{x,M} * C_t * M_t - k_{off,M} * B_t + k_{x,C} * R_t * D_t - k_{off,M} * B_t + (k_{rec} + k_{rec,M}) * B_i - (k_{endo} + k_{endo}) * B_t$$
Eqn A.1.19	$$\frac{dB_i}{dt} = k_{x,MI} * C_i * M_i - k_{x,CI} * D_i * R_i - (k_{off,I} + k_{off,MI})B_I - (k_{rec} + k_{rec,M}) * B_i + (k_{endo} + k_{endo}) * B_t$$
Eqn. A.1.20	$$L_{tumor,total} = L_t + L_i + L_d + C_t + C_i + D_t + D_i + B_t + B_i$$
Eqn. A.1.21	$$Background = \begin{cases} L_p, L_p > L_n \\ L_n, L_n \geq L_p \end{cases}$$

The above equations were solved simultaneously using the Matlab ode15s Solver. To calculate constants within these equations, the following equations were used (adapted from Schmidt and Wittrup, *Mol. Cancer Ther.* 2009;8(10)).

Table A.1.3: Constitutive equations

Eqn. A.1.22	$$D = A * D_{small} + (1 - A) * D_{large}$$
Eqn. A.1.23	$$D_{large} = D_{free} * \left(\frac{D_{pore}}{D_{free}} (\lambda_{large}) \right)$$
Eqn. A.1.24	$$D_{small} = D_{free} * \left(\frac{D_{pore}}{D_{free}} (\lambda_{small}) \right)$$
Eqn. A.1.25	$$D_{free} = \frac{3 * 10^{-10}}{R_{mol} * 10^{-9}}$$
Eqn. A.1.26	$$R_{mol} = \left(\frac{MW}{1.32} \right)^{\frac{1}{3}} * (10^{-9})$$
Eqn. A.1.27	$$\lambda_{small} = R_{mol}/R_{small}$$
Eqn. A.1.28	$$\lambda_{large} = R_{mol}/R_{large}$$
Eqn. A.1.29	$$\lambda_{Psmall} = R_{mol}/R_{P,small}$$

Eqn. A.1.30	$$\lambda_{P,large} = R_{mol}/R_{P,large}$$
Eqn. A.1.31	$$\phi = \begin{cases} (1-\lambda)^2, \lambda < 1 \\ \\ 0, \lambda > 1 \end{cases}$$
Eqn. A.1.32	$$D_{pore}/D_{free} = \begin{cases} \dfrac{1 - 2.105*\lambda + 2.0865*\lambda^3 - 1.7068*\lambda^5 + 0.72603*\lambda^6}{1 - 0.78587*\lambda^5}, \quad \lambda < 0.6 \\ 0.5223 - 1.8375*\lambda + 2.1603*\lambda^2 - 0.8484*\lambda^3, \quad 0.6 \le \lambda \le 1 \\ 0, \lambda > 1 \end{cases}$$
Eqn. A.1.33	$$P_{small} = D_{free} * \left(\frac{D_{pore}}{D_{free}}(\lambda_{Psmall}) \right) * (\phi(\lambda_{Psmall}))$$
Eqn. A.1.34	$$P_{large} = D_{free} * \left(\frac{D_{pore}}{D_{free}}(\lambda_{Plarge}) \right) * (\phi(\lambda_{Plarge}))$$
Eqn. A.1.35	$$P = A_{perm} * P_{small} + B_{perm} * P_{large}$$
Eqn. A.1.36	$$\varepsilon = V_i * \left(A * \phi(\lambda_{small}) + (1-A) * \phi(\lambda_{large}) \right)$$
Eqn. A.1.37	$$k_{off} = K_d * k_{on}$$
Eqn. A.1.38	$$k_{rec} = \frac{k_{endo}}{\left(\frac{1}{sig} - 1\right) * \left(\frac{rho}{rho + 1}\right)}$$
Eqn. A.1.39	$$k_{deg} = \frac{k_{rec}}{rho}$$
Eqn. A.1.40	$$k_{syn} = \text{Surf} * \text{k}_{endo}/(1 + \text{rho})/(6.02 * 10^{23}) * (0.5)/(4/3 * \text{pi} * (8 * 10^{-6})^3 * 1000)$$
Eqn. A.1.41	$$k_{degL} = 0.01/15$$
Eqn. A.1.41	$$k_{recM} = \frac{k_{endoM}}{\left(\frac{1}{sigM} - 1\right) * \left(\frac{rhoM}{rhoM + 1}\right)}$$
Eqn. A.1.42	$$SurfM = Biv * SurfI, Biv = 1 \ for \ bivalent \ ligand, = 0 \ for \ monovalent$$
Eqn. A.1.43	$$k_{degM} = \frac{k_{recM}}{rhoM}$$
Eqn. A.1.44	$$k_{syn} = \text{SurfM} * \text{k}_{endoM}/(1 + \text{rhoM})/(6.02 * 10^{23}) * (0.5)/(4/3 * \text{pi} * (8 * 10^{-6})^3 * 1000)$$
Eqn. A.1.45	$$k_{xM} = k_{onM} * enhancement$$

Eqn. A.1.46	$$k_{xC} = k_{on} * enhancement$$
Eqn. A.1.47	$$k_{xMI} = k_{onMI} * enhancement$$
Eqn. A.1.48	$$k_{xCI} = k_{onCI} * enhancement$$
Eqn. A.1.49	$$ClNR = ClNR_0 - \delta * \frac{R_{mol}}{R_{mol} + \gamma}$$
Eqn. A.1.50	$$\phi_{Kdiff} = e^{-\alpha*R_{mol}}$$
Eqn. A.1.51	$$\phi_{Kconv} = e^{-\beta*R_{mol}}$$
Eqn. A.1.52	$$Pe = \frac{\phi_{Kconv} * v * MT}{\phi_{Kdiff} * D_{free}}$$
Eqn. A.1.53	$$\theta = \frac{\phi_{Kconv}}{1 - e^{-sig*Pe} + \phi_{Kconv} * e^{-sig*Pe}}$$
Eqn. A.1.54	$$ClR = GFR * \theta$$
Eqn. A.1.55	$$k_{clear} = \frac{ClR + ClNR}{V}$$
Eqn. A.1.56	$$R_{T,0} = \frac{k_{syn}}{k_{endo} * \left(1 + \frac{k_{rec}}{k_{deg}}\right)}$$
Eqn. A.1.57	$$R_{i,0} = k_{syn} * \frac{\left(1 + \frac{k_{rec}}{k_{deg}}\right)}{(k_{rec} + k_{deg})}$$
Eqn. A.1.58	$$M_{T,0} = \frac{k_{synM}}{k_{endoM} * \left(1 + \frac{k_{recM}}{k_{degM}}\right)}$$
Eqn. A.1.59	$$M_{i,0} = k_{synM} * \frac{\left(1 + \frac{k_{recM}}{k_{degM}}\right)}{(k_{recM} + k_{degM})}$$

Table A.1.4: Constants

Constant	Name	Value	Units
MW	Ligand molecular weight	22	kDa
Lp0	initial ligand plasma conc.	$25*10^{-9}$	M
V	Plasma Volume	0.002	L
r_{cap}	Capillary radius	$8*10^{\wedge}-6$	m
r_{Krogh}	Krogh radius	$50*10^{\wedge}-6$	m

r_{tumor}	Tumor radius	$2*10^{-3}$	m
R_{small}	Small diffusive pore size	$13.8*10^{-9}$	m
R_{large}	Large diffusive pore size	$1000*10^{-9}$	m
Aperm	Permeability of large pore in vessel	1760	m^{-1}
Bperm	Permeability of small pore in vessel	65	m^{-1}
RpSmall	small permeable pore size	$4.5*10^{-9}$	m
RpLarge	large permeable pore size	$500*10^{-9}$	m
A	Fraction of large diffusive pores	0.9	--
ClNR0	small molecule non-renal clearance	$(2*10^{-3})/3600$	[L/s]
α	convective clearance coefficient	$1.6*10^9$	--
β	diffusive clearance coefficient	$0.95*10^9$	--
δ	convective clearance coefficient	$(1.94*10^{-3})/3600$	--
γ	non-renal clearance coefficient	$0.20*10^{-9}$	--
sig	geometric factor	2	--
v	fluid velocity	10^{-5}	m/s
MT	membrane thickness	$100*10^{-9}$	m
GFR	glomular filtration rate	$(10*10^{-3})/3600$	mL/s
V_i	void volume	0.5	--
k_{lymph}	ratio of lymphatic to vascular permeability	30	--
k_{on}	association rate constant	$2*10^5$	$M^{-1}s^{-1}$
k_{endo}	endocytosis rate constant	$4.2*10^{-6}$	s^{-1}
rho	ratio of recycle rate constant over degradation rate constant	0.15	--
sig	percent of surface expressed receptors	0.90	--
k_{degL}	ligand degradation rate constant	0.01/15	--
k_{elim}	label elimination rate constant	$1*10^{-5}$	--
k_{onM}	association rate constant,	$2*10^5$	s^{-1}
k_{endo}	endocytosis rate constant	$2.1*10^{-6}$	s^{-1}
rhoM	ratio of recycle rate constant over degradation rate constant for monovalent	0.15	--
sigM	percent of surface expressed receptor-ligand complexes	0.90	--
enhancement	volume correction with tethering	171	--