ADVANCES IN THE DEVELOPMENT OF HIF-1ALPHA-ACTIVATED PROTEIN SWITCHES FOR ENZYME PRODRUG THERAPY

By R. Clay Wright

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March 19, 2014
Abstract

Prevailing approaches for developing cancer protein therapeutics focus on creating proteins that therapeutically modulate a cancer marker’s function. Such an approach limits the therapeutic mechanism to those that naturally arise from modulation of the cancer marker and precludes the use of cancer markers for which therapeutic modulation is not feasible. Furthermore, many potential protein therapies lack the desired cancer targeting. The ability to link recognition of any cancer marker with activation of any desired therapeutic function would enormously expand the number of possible protein therapeutics.

We have previously engineered a switchable prodrug-activating enzyme that selectively kills human cancer cells that accumulate the cancer marker hypoxia-inducible factor 1α (HIF-1α). This HIF-1α-activated enzyme switch (Haps59) was created by fusing the prodrug-converting enzyme yeast cytosine deaminase (yCD) and the CH1 domain of the p300 protein, which binds HIF-1α. Haps59 autonomously increases its ability to convert the prodrug 5-fluorcytosine (5FC) into the chemotherapeutic 5-fluorouracil (5FU) in a HIF-1α-dependent manner, rendering colon and breast cancer cells more susceptible to the prodrug. However, the difference in 5FC sensitivity between the presence and absence of HIF-1α was not as large as desired.

Using a variety of mutagenesis methods, followed by a two-tiered genetic selection for improved switches, we have identified new HIF-1α-activated enzymes that confer E. coli with modest increases in HIF-1α-dependent 5FC toxicity.
However, the current bottleneck in further translation of HIF-1α-activated protein switches is screening potential switch candidates in mammalian cells. To accommodate higher throughput, we explored the use of Flp recombinase-mediated isogenic integration. While initial results in this system are promising, these experiments also brought to light the disadvantageous promiscuous binding activity of the CH1 domain, which we further confirmed in *E. coli*. This promiscuous binding and subsequent off-target activation needs to be examined under normal physiological conditions to pinpoint off-target activity in these potential therapeutics. With relevant aberrant activators identified, further directed evolution can be used to improve the cancer specificity of HIF-1α-activated protein switches.

**Advisor:** Marc Ostermeier, Professor of Chemical and Biomolecular Engineering, Johns Hopkins University

**Readers:** Sharon Gerecht, Associate Professor of Chemical and Biomolecular Engineering, Johns Hopkins University

Craig A. Townsend, Alsoph H. Corwin Professor of Chemistry, Johns Hopkins University
Acknowledgments

About four or five years ago Rich Heins gave me a piece of advice while he was writing his thesis: “Set aside a day or two a month to write about everything you’ve done recently.” Seeing this as sage advice from a seasoned veteran, I scheduled a monthly recurring “writing day” on my calendar. Did I ever write on writing day? Only if writing, coincidentally, was at the top of the to-do list. Do I now wish I had written on writing day? Absolutely.

I garnered many other sage wisdoms and valuable skills from the other members of the Ostermeier lab over the years. For that and for making B26 feel like home, I’ll be forever appreciative. I would especially like to thank Chapman Wright for getting me started down this road and for being a constant sounding board during the early stages of this work. Also, Elad Firnberg and Brian Chaikind have been great friends and exceptionally helpful and encouraging. Elad, I couldn’t have randomly selected a better housemate. Arjun Khakhar, thanks for your ceaseless assistance with the linker libraries and RICP library. Your creativity and optimism are ever inspiring. To everyone else in the department, especially the Betenbaugh lab, Gerecht lab, and Wirtz lab, thank you for your generous advice and use of equipment and space. Without your help this wouldn’t have been possible. Of course, I should also mention all of the friends I have turned into brewers. Your friendship and creativity has always been a welcome escape.

I would like to thank Dr. Marc Ostermeier for providing me with boundless support and allowing me the freedom to pursue directions outside both of our expertise. You have been a truly exceptional mentor. I have learned more from you than you know. To my readers, Dr. Sharon Gerecht and Dr. Craig Townsend, and committee members, I am greatly appreciative of your mentorship, guidance, and patience.

Finally, I wouldn’t be here without my family, who have always supported my creativity and pursuit of knowledge. Thanks for your support and for assisting (or putting up) with all of my experiments over the years. Rachel, the past 3 years have been the best of my life, I can’t wait for March 29th. Thank you for your emotional support and for taking good care of me while I finished this. I promise to reciprocate when you are finishing your dissertation.
# Table of Contents

Abstract ........................................................................................................................................ ii

Acknowledgments ....................................................................................................................... iv

Table of Contents ....................................................................................................................... v

List of Figures ............................................................................................................................ ix

List of Tables ................................................................................................................................ x

Chapter 1: Introduction and background .................................................................................. 1

Cancer and chemotherapy ........................................................................................................ 1

Prodrug Therapy ......................................................................................................................... 2

Cancer marker-activated protein switches ................................................................................ 5

Prodrug-activating enzymes .................................................................................................... 8

Herpes simplex virus thymidine kinase ....................................................................................... 8

Cytosine deaminase and 5-fluorocytosine ............................................................................... 8

Mechanism of 5-fluorouracil toxicity ....................................................................................... 9

Yeast cytosine deaminase .......................................................................................................... 11

Structure .................................................................................................................................... 11

Mechanism of yeast cytosine deaminase ................................................................................... 12

Thermostabilization ................................................................................................................... 14

Random mutagenesis ................................................................................................................ 14

HIF-1α and the CH1 domain of the transcriptional coactivator of p300 ..................... 15

HIF-1α, oxygen regulation, and cancer ..................................................................................... 15

The CH1 domain of p300/CBP ............................................................................................... 18

Structure of the CH1/HIF-1α complex ..................................................................................... 18
Chapter 2: 1st generation switches ................................................................. 25
Library and selection .................................................................................. 25
Characterizing activity in E. coli .................................................................. 29
Methods – Dot toxicity assay ....................................................................... 30
Results – Dot toxicity assay ......................................................................... 31
Switching mechanism in E. coli ..................................................................... 32
Characterizing activity in cancer cell lines .................................................... 33
Switching mechanism in cancer cell lines ...................................................... 36

Chapter 3: 2nd generation switches .............................................................. 37
Introduction .................................................................................................. 37
Methods ....................................................................................................... 37
Library creation ............................................................................................ 37
Error-prone PCR .......................................................................................... 37
Linker mutagenesis ....................................................................................... 38
Two-tiered selections .................................................................................... 41
Negative selection ....................................................................................... 42
Positive selection ......................................................................................... 42
Screening ...................................................................................................... 42
Expression and western blots ....................................................................... 44
Results ......................................................................................................... 45
Random mutagenesis libraries ..................................................................... 45
Linker libraries ............................................................................................. 47
Mechanism of improved switches ................................................................. 49

Chapter 4: 1st generation Haps libraries of greater diversity ......................... 55
Creation of FI-RKO cells .............................................................................................................. 97

Chapter 6: The promiscuous CH1 domain .................................................................................. 99

Methods ......................................................................................................................................... 100

Expression of other CH1-binding peptides in *E. coli* ............................................................... 100

Immunoprecipitations of FLAG-Hap59 and E1A from FI-293 cell lysate ................................. 101

Culture, electroporation and immunoprecipitations from HPDE cells ....................................... 102

Results ........................................................................................................................................... 103

Aberrant switch activation in *E. coli* ......................................................................................... 103

Immunoprecipitations of E1A and Fhaps59 from FI-293 cells ..................................................... 105

Immunoprecipitations of FHap59 in HPDE cells ......................................................................... 106

Chapter 7: Conclusions and future directions .......................................................................... 109

A 3rd generation of Haps .............................................................................................................. 109

Isogenic integration in colon carcinoma cells .............................................................................. 111

Are CH1-based switches therapeutically relevant? ................................................................... 112

Evolving a cancer-specific CH1 domain ....................................................................................... 114

Conclusion ..................................................................................................................................... 115

References ..................................................................................................................................... 116

Curriculum Vitae .......................................................................................................................... 137
List of Figures

Figure 1.1 - Cartoon of HIF-1α-activated protein switches.................................................6

Figure 1.2 – Crystal structure of yCD dimer. .................................................................11

Figure 1.3 - The active site of yCD. .............................................................................12

Figure 1.4 – Regulation of HIF-1α signaling.................................................................17

Figure 1.5 – Structure of the HIF-1α CTAD and CH1 domain complex.........................19

Figure 2.1 - Schematic showing domain insertion method used to create the yCD-CH1 hybrid library. ..................................................................................................................26

Figure 2.2 – Schematic of the two-tiered selection.........................................................27

Figure 2.3 – Schematic representations of the sequences of Haps3 and Haps59............28

Figure 2.4 - Screening Haps59 and Haps3 in liquid culture.............................................29

Figure 2.5 – Toxicity of 5FC to E. coli cells expressing Haps59 and Haps3 as assessed by dot toxicity assay. ..................................................................................................................32

Figure 2.6 – Accumulation of Haps59 and yCD expressed in GIA39 coexpressing GST or GST-HIF-1α-CTAD..............................................................................................................33

Figure 2.7 - 5FC and 5FU sensitivity experiments in RKO and MCF7 cells.....................34

Figure 2.8 - Characterization of Haps59 in human cancer cells.....................................35

Figure 2.9 - Haps59 and HIF-1a interact in RKO cells....................................................36

Figure 3.1 – Strategies for creating improved HIF-1α-activate protein switches. ......46

Figure 3.2 – Screen of the best variants from random mutagenesis and linker libraries. .............................................................................................................................................48

Figure 3.3 - Characterization of Ehaps22 and Ehaps08 in E. coli.................................49
Figure 3.4 – Time course western blot of Ehaps08 and Ehaps22........................................50

Figure 3.5 – Effects of the addition of protease inhibitors on Haps59 degradation..52

Figure 3.6 – Western of cultures induced at 0.1 OD<sub>600</sub> and grown for 2 hours........52

Figure 3.7 – Structural model of Haps59 and Ehaps08 (E249K mutation).................53

Figure 4.1 - Schematic of the construction of libraries of random circular
permutations of CH1 that are randomly inserted into the yCD domain of FLAG-tagged yCD..........................................................63

Figure 4.2 - Details of primer design and method for creation of deletions and
duplications at the site of linearization.............................................................64

Figure 4.3 – Plasmid map of pSkunk-cpCH1 used to add random linkers to new
termini of circularly permuted CH1 domains.................................................66

Figure 4.4 – Representative images of cyclization and random digestion of the CH1
domain................................................................................................................71

Figure 4.5 – PCR screen of pCDNA-cpCH1 library...................................................71

Figure 4.6 – PCR screen of the insert region of pSkunk of surviving members of
positive selections on the RICP naïve library in GIA39-HIF............................74

Figure 4.7 - PCR screen of the insert region of pSkunk of P50 selected RICP library
members in GIA39 surviving negative selections..............................................75

Figure 4.8 – Dot toxicity screen of members RICP-P50-N350-P50............................76

Figure 4.9 - Characterization of RICP12 and RICP95 in E. coli. ..............................77

Figure 4.10 – Circular permutation sites in RICP12 and RICP95.............................78

Figure 4.11 – Screening of TICP clones in GIA39 with pGA or pGA-HIF................79

Figure 5.1 – Plasmid maps of pCAGGS-EGFP and pCAGGSneo-EGFP..................86
Figure 5.2 – Transient transfection 5FC toxicity assay in HeLa cells.........................94
Figure 5.3 – Transient transfection 5FC toxicity assay in 293 cells. .........................94
Figure 5.4 – Optimization and controls for 5FC toxicity in FI-293 cells....................96
Figure 5.5 – Characterization of FI-RKO clones. ......................................................98
Figure 6.1 – Effects of CITED2 and E1A expression on Haps59 as assessed by E. coli
dot toxicity assay........................................................................................................104
Figure 6.2 – Immunoprecipitation of FLAG-tagged yCD and Haps59 or E1A from FI-
293 cells.......................................................................................................................106
Figure 6.3 – Immunoprecipitation of FHaps59 from HPDE and 293 cells visualized
by silver staining ...........................................................................................................108
Figure 7.1 – Schematic of FLAG epitope removal using a type II s enzyme..............110
List of Tables

Table 1.1 - Known binding partners of the CH1 domain .................................................................21

Table 3.1 – List of ePCR library members having similar activity to Haps59. ...............47

Table 3.2 – List of linker library members with similar therapeutic windows to

Haps59. ........................................................................................................................................49

Table 3.3 – ePCR and Linker library statistics and selections........................................54

Table 4.1 – RICP and TICP library and selection statistics. ..............................................81
Chapter 1: Introduction and background

Cancer and chemotherapy

Based on the current incidence rates, cancer will affect two out of every five people over their lifetimes [1]. Cancer is not a single disease but rather a collection of diseases, with every incidence being unique. This singularity is a result of the way cancers come about. At its simplest, a cancer is a group of cells that has subverted their preprogrammed developmental path and prescribed function and gone rogue multiplying without regard for the surrounding normal cells. The development of cancer, in this way, is an evolutionary process. Cells under stress acquire mutations which will in rare cases confer positive growth phenotypes known as the “hallmarks of cancer” including proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis [2]. These hallmarks allow cancerous cells to outcompete normal cells and evade immune surveillance resulting in their malignant nature. In addition cancerous or neoplastic cells may develop a mutator phenotype, increasing their probability of acquiring an oncogenic mutation and gaining additional hallmark traits.

Some oncogenes or their associated hallmark phenotypes can be exploited for treatment with small molecule drugs or antibodies. Traditional chemotherapy exploits the rapid growth rate conferred by proliferative signaling and evasion of growth suppression. The ability to target specific oncogenes is becoming increasingly prevalent with the decreasing price of next generation sequencing and
other diagnostic techniques associated with personalized medicine. However, several frequently observed oncogenes are often considered undruggable due to their ubiquitous nature or structure.

The majority of cancers are still treated with surgery and/or radiation, followed by a regimen of chemotherapy. Most chemotherapeutic agents target the rapid division of cancer cells by inhibiting DNA synthesis or metabolic pathways. Discovered more than 50 years ago, most chemotherapeutics have indiscriminant toxicity, killing malignant cells as well as other fast growing cells including bone marrow cells, the digestive epithelium, skin cells, and hair follicle cells. This general toxicity leads to the gruesome side effects associated with chemotherapy as well as frequent infections due to immunosuppressive effects.

The ratio of the effectiveness of a therapeutic against disease tissue versus its non-target toxicity in normal tissue is referred to as the therapeutic window or index. To circumvent the negative effects and expand the therapeutic window of chemotherapeutics, methods of directly delivering these drugs to tumors have been developed. Chemotherapeutics can be loaded into particles designed to bind and be internalized by tumor cells, or to get nonspecifically stuck in tortuous tumor vasculature. However, the liver and spleen also often capture these particles resulting in off-target toxicity in these organs.

**Prodrug Therapy**

Another method for selective delivery of chemotherapeutics involves modifying the chemotherapeutic with a labile group rendering it non-toxic. This form of the chemotherapeutic, termed the “prodrug,” is the substrate of an enzyme,
which removes the labile group and releases the toxic chemotherapeutic. Therefore specific delivery of the enzyme to cancerous cells would allow the selective delivery of the toxic chemotherapeutic. To date, tumor-specific prodrug conversion has been accomplished in several ways. Initially prodrugs were designed such that the tumor environment or enzymes over-expressed within tumor cells would liberate the active compound. This approach was used in attempt to reduce the side effects of the extremely potent chemotherapeutic, methotrexate [3]. Prodrugs activated by proteases that are commonly over expressed in the tumor microenvironment have been slightly more successful [4, 5]. This approach however is limited to tumors overexpressing the effective protease and is easily evaded by metastases without developed microenvironments.

Soon after the initial development of prodrug therapies, antibody production technology rapidly advanced due to the development of hybridomas, which has generated several new approaches to cancer therapy. Antibodies targeting tumor cell-surface markers have allowed the immune system to target tumor cells for destruction. These antibodies may also be conjugated to chemotherapeutics facilitating more directed delivery. Antibodies have also been used to deliver, selectively to tumors, enzymes capable of activating prodrugs in Antibody-Directed Enzyme Prodrug Therapy (ADEPT) [6, 7, 8]. ADEPT calls for the antibody-enzyme conjugate to be delivered systemically and allowed to bind specifically to the tumor and unbound conjugate to be cleared from the body prior to delivery of the prodrug. This process requires precise knowledge of pharmacokinetics for maximum efficacy.
Inactivating molecules or antibodies can be used to enhance clearance of the unbound antibodies [9].

Gene-directed enzyme prodrug therapy (GDEPT) is a slightly more recent development, providing an improved therapeutic index by targeting expression of the prodrug-converting enzyme to tumor cells. By limiting delivery of the enzyme (or the DNA encoding it) specifically to cancerous cells, the enzyme vector and prodrug can be delivered simultaneously. The selectivity of GDEPT, as well as any enzyme prodrug therapy, is limited by the method used to target the delivery and/or expression of the enzyme specifically to cancer cells. For GDEPT, this can currently be achieved using tumor targeting nanoparticles and viruses. However as selective delivery methods have been developed, a tradeoff between selectivity and efficiency has been observed [10]. Highly infectious, replicative and non-replicative viruses have been modified to selectively infect tumor cells but this modification generally decreases the efficiency of the delivery [11, 12, 13, 14, 15, 16]. Similarly, tumor-targeting DNA nanoparticles have been developed [10]. In each case the delivery efficiency is decreased limiting the effectiveness of GDEPT.

Selectivity can also be achieved through transcriptional targeting, using tumor specific promoters. Transcriptional targeting has been verified to limit enzyme expression and associated toxicity to tumor cells in a phase 1 clinical trial [17]. The brief assessment of effectiveness in the study was inconclusive and limited by low enrollment. Additionally, the potential for off-target expression was limited by direct injection of the expression plasmid into the cancerous nodule. To achieve the maximum therapeutic index in a particular GDEPT system, a compromise must
be made by tuning selectivity and efficiency of delivery and transcription of the therapeutic enzyme.

**Cancer marker-activated protein switches**

As protein engineers we aimed to avoid the pitfalls of selective delivery by engineering an enzyme whose activity is regulated by a cancer marker. This enzyme would essentially act as a molecular switch, which is inactive or “off” in its default state and active or “on” only in the presence of a specific cancer marker. Such an enzyme-switch would allow systemic delivery of both the enzyme and prodrug, as the enzyme switch would be able to discriminate cancer marker expressing cells from non-cancerous cells. Additionally this technology may be paired with selective delivery and transcriptional regulation. The burden of discriminating cancer cells from normal cells being on the cancer-activated enzyme avoids the inefficiency of targeted delivery and allows distal tumors and metastatic cells to be effected as well as the primary tumor.

Previously, protein switches have been made that detect small molecules and activate the antibiotic resistance gene, beta-lactamase [18, 19, 20]. These switches have been engineered, using existing proteins with the desired input and output functions. To create a switch from these existing proteins, a gene library is made by permuting and fusing the coding sequences for the input and output proteins in as many different configurations as possible. Those fusions acting as switches—whose output function is regulated by the input signal—are selected from this library using directed evolution, which is based on the principle of natural selection. From this diverse library, fusions with the desired off-state are enriched by creating selective
This thesis revolves around the concept of a protein switch based on the prodrug converting enzyme or suicide gene, yeast cytosine deaminase (yCD), which converts the non-toxic prodrug 5-fluorocytosine (5FC) to the highly toxic chemotherapeutic 5-fluorouracil (5FU). The enzymatic activity of the yCD-domain of the switch is modulated by the cancer marker, hypoxia-inducible factor-1α (HIF-1α). In normal cells—where HIF-1α is absent—the switch’s cytosine deaminase domain is inactive. Only in the presence of HIF-1α is the enzymatic domain active and able to produce chemotherapeutic.

Figure 1.1 - Cartoon of HIF-1α-activated protein switches.
This thesis revolves around the concept of a protein switch based on the prodrug converting enzyme or suicide gene, yeast cytosine deaminase (yCD), which converts the non-toxic prodrug 5-fluorocytosine (5FC) to the highly toxic chemotherapeutic 5-fluorouracil (5FU). The enzymatic activity of the yCD-domain of the switch is modulated by the cancer marker, hypoxia-inducible factor-1α (HIF-1α). In normal cells—where HIF-1α is absent—the switch’s cytosine deaminase domain is inactive. Only in the presence of HIF-1α is the enzymatic domain active and able to produce chemotherapeutic.

pressure against the output function in the in the absence of the input signal (e.g. selecting against yCD activity in the absence of HIF-1α, for our HIF-1α-activated protein switches). Fusions with the desired on-state are then selected from this enriched library by generating selective pressure for the output function in the presence of the input function (e.g. selecting for yCD activity in the presence of HIF-1α). This scheme, of course, is for a directly acting switch. An inverse switch can also be envisioned.

Developing an enzyme switch first requires the selection of both appropriate output and input domains. For our cancer marker-activated protein switch we could select from any of the currently available enzyme-prodrug combinations for the output enzymatic domain. This broad set of enzymes mostly originates from microorganisms, as they are required to provide an activity exogenous to normal
human tissues. These enzymes are typically metabolic enzymes, producing essential metabolites. These essential metabolites can then be modified making anti-metabolites or chemotherapeutics, which inhibit necessary downstream metabolic processes resulting in toxicity. Several enzyme-prodrug pairs have been assessed in clinical trials [11, 12, 13, 17, 21].

For the input domain of our cancer marker-activated enzyme we could choose any cancer marker-binding protein. However, this first requires the selection of a cancer marker, as our switch would only be as cancer-specific as the chosen cancer marker. To create a broadly effective therapeutic switch, we desired a cancer marker that is present at high levels in a variety of cancers and is essentially absent in normal tissues. Based on these criteria, we selected Hypoxia-inducible factor 1α (HIF-1α) as the input signal for our cancer marker-activated enzyme. HIF-1α is nearly undetectable in normal tissue with a half-life of less than five minutes, yet it accumulates to high levels in many cancer types—including pancreatic, breast, prostate, and colorectal—causing increased invasiveness, altered metabolism, angiogenesis, metastasis, and drug resistance [22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]. As an input domain detecting HIF-1α, we chose one of HIF-1α’s natural binding partners, the CH1 domain of the human p300 protein, which is known to interact strongly with HIF-1α [33].
Prodrug-activating enzymes

Herpes simplex virus thymidine kinase

One of the most prevalent prodrug-converting enzymes is the thymidine kinase from human herpes simplex virus-1 (HSVtk). Although human tissues also express a thymidine kinase, HSVtk exhibits broad specificity allowing activation of several prodrugs, for which human thymidine kinase has low affinity [34]. As a kinase, HSVtk renders its toxic chemotherapeutic products charged. Without the assistance of gap-junctions, this charge prevents the active chemotherapeutic from diffusing into and exhibiting toxicity in neighboring cells.

This diffusive toxicity, known as the “bystander effect” is beneficial in the instance that not all tumor cells express the prodrug-converting enzyme. The bystander effect may also contribute to some systemic toxicity, but there is little evidence that this is of concern, especially considering that any GDEPT strategy will be less toxic than traditional chemotherapy. It is generally agreed that the bystander effect is beneficial for the treatment of solid tumors. For tumors of the nervous system, the bystander effect would cause significant damage to neighboring nerve cells. Treatment of gliomas using viral delivery of HSVtk followed by prodrug injections has been shown to be efficacious and safe [35].

Cytosine deaminase and 5-fluorocytosine

Another enzyme prodrug pair, cytosine deaminase and 5-fluorocytosine, exhibits a strong bystander effect. Cytosine deaminases catalyze the decomposition of cytosine (or 5-fluorocytosine) into uracil (or 5-fluorouracil) and ammonia in
bacteria and fungi. 5-fluorouracil (5FU, trade-name Efudex) is a highly potent chemotherapeutic, which exhibits toxic antimetabolite effects by both inhibiting necessary enzymes and being incorporated into DNA and RNA [36]. 5-fluorocytosine (5FC, Fluycytosine, and marketed under several brand names) is an anti-fungal compound with very low toxicity in humans, allowing doses from 150-200 mg/kg/day [37]. 5FC is typically given every 6 hours (q.i.d.) and reaches maximum serum concentrations of 50-100 μg/mL within 1-2 hours [38].

**Mechanism of 5-fluourouracil toxicity**

As mentioned above 5-fluorouracil exhibits several modes of toxicity. First however, it must be converted to active forms 5-fluorodeoxyuridine monophosphate (FdUMP), 5-fluorodeoxyuridine triphosphate (FdUTP), and 5-fluourouridine triphosphate (FUTP). 5FU is activated by two separate pathways. It is converted to 5-fluourouridine monophosphate (FUMP) by orotate phosphoribosyltransferase or uridine phosphorylase and uridine kinase, and is then phosphorylated to FUDP and FUTP, which can then be incorporated into RNA. Thymidine phosphorylase can also act on 5FU forming fluorodeoxyuridine (FUDR). FUDR is then phosphorylated by thymidine kinase to FdUMP, which inhibits thymidylate synthase, causing depletion of deoxythymidine triphosphate pools. FdUMP can also be formed from FUDP through action of ribonucleotide reductase and dephosphorylation. Phosphorylation of FdUMP or FdUDP to FdUTP allows incorporation into DNA also exerting toxicity [36].

Inhibition of thymidylate synthase shifts the balance of thymidine nucleotides to uridine nucleotides, which along with FdUTP can be incorporated
into DNA. These bases are subsequently excised by uracil-DNA-glycosylase, leaving apyrimidinic sites which due to lack of thymidine nucleotides cannot be repaired. This interference in both DNA synthesis and repair leads to significant DNA damage and cell death. Incorporation of FUTP into RNA also exhibits significant toxicity by inhibiting splicing and polyadenylation of mRNA and processing of rRNA and tRNA.

Cytosine deaminases from both the bacterium *Escherichia coli* and yeast *Saccharomyces cerevisiae* have been used in GDEPT systems. Yeast cytosine deaminase (yCD) has a 22-fold lower $K_m$ for 5FC than its bacterial counterpart, and is accordingly more effective [39]. Additionally yCD is a homodimer whereas bacterial cytosine deaminase is homohexameric. Although no cooperativity has been observed in either enzyme [40, 41] and enzyme-antibody fusions exhibit near-native activity in presumed monomeric states [7, 8, 42], such fusions (as our desired switch) may be destabilized relative to the native quaternary structure.

Another potential sticking point for any non-human enzyme for GDEPT is immunogenicity. Once again, as with the bystander effect, this is a double-edged sword. An enzyme, or its delivery method, cannot be so immunogenic that an effective dose will induce adverse effects. However, some immunogenicity may provide an adjuvant effect, priming the immune system to recognize and kill distant tumor cells. Phase 1 trials with cytosine deaminases from yeast and *E. coli* have not led to dose limiting toxicities or adverse events due to the immunogenicity of these enzymes, even in the context of viral delivery [13, 17]. In the case that significant immunogenicity is observed, transcriptional targeting can be used to limit this to the tumor site.
Yeast cytosine deaminase

Due to the favorable characteristics above as well as the other technical advantages below, we chose yCD and 5FC as the enzyme/prodrug pair for the initial cancer-activated enzyme switches.

Structure

The structure of yCD was initially solved in 2003 by Barry Stoddard’s group at the Fred Hutchinson Cancer Center [40]. This lab along with the group of Margaret Black at Washington State University has made significant contributions to the advancement and understanding of prodrug converting enzymes. yCD is a zinc metalloenzyme of 158 residues and 17.5 kDa. The crystal structure of yCD reveals a head-to-tail homodimer of mixed α/β topology. A central β-sheet is surrounded by alpha helices. Although there is no reported cooperativity in yCD, dimerization does occlude 1000 Å² of mostly hydrophobic surface area. There are several loops in the structure, however not one extends off the surface of the protein.

Figure 1.2 – Crystal structure of yCD dimer.
Two yCD molecules (in cyan and green) align with their N-termini (shown in blue) and C-termini (shown in red) on opposite sides of the dimer. Hydrophobic interactions between α-helices stabilize the dimer and shield these hydrophobic surfaces from solvent. The distance from α-carbon of the N- to C-terminus is marked in orange and measures 35.7 Å.
The entrance to the active site of yCD is located in the space between the kinked C-terminal helix and the β-sheet (bottom-left corner of Figure 1.2), and is occluded from solvent by a loop between the sheet and helix on the opposite side of the molecule. The active site involves residues from the helices, sheets and loops on this side of the molecule, as well as a catalytically active zinc ion, which is coordinated by His62, Cys91, Cys94, and a water molecule (Figure 1.3). This water molecule acts as a nucleophile in the deamination of the substrate. The only catalytically active residue is Glu64, which acts as a proton shuttle between the zinc-water complex and the substrate. The substrate is positioned by π-stacking interactions with His62, and is occluded from the solvent by Phe114 and Trp152.

**Figure 1.3 - The active site of yCD.**
Active site residues are colored per element in stick representation. Also in stick representation but colored orange, is the mechanism-based inhibitor 4-(R)-hydroxyl-3,4-dihydropyrimidine (DHP). The catalytic zinc is shown as a grey sphere.

**Mechanism of yeast cytosine deaminase**

The detailed mechanism of yeast cytosine deaminase was elucidated by ONIOM computational study [43], using the high resolution structures solved by the Stoddard group [40]. A diagram of the described mechanism can be found in [43].
With the substrate bound, be it cytosine or 5-fluorocytosine, the carboxyl group of Glu64 (which exists in its deprotonated form) is hydrogen bonded to both a water molecule coordinating zinc and the amino group of the substrate. This hydrogen bonding arrangement allows Glu64 to remove the shared proton from the zinc-coordinated water molecule. The now protonated carboxyl group rotates and is attacked by an electron pair shared by N² and C⁴ of cytosine transferring the proton to N³, and inviting attack from the hydroxyl group which remains coordinated by zinc on C⁴. Upon protonation and rotation of Glu64, the amine group of cytosine is hydrogen bonded to the other oxygen of the carboxyl group, which also shares a hydrogen bond with the remaining proton of the now C⁴-bound-and-zinc-coordinated hydroxyl group. Again the carboxyl group of Glu64 attacks this final proton of the hydroxyl group, and the remaining electron pair of the oxygen collapses onto C⁴ of the ring removing the amine group forming uracil. This amine now removes the proton from Glu64 forming ammonia, and returning Glu64 to its deprotonated state. Uracil is hypothetically removed from the active site as ammonia is exchanged for water, which is once again attacked by Glu64 leaving a hydroxyl group that forms a gem-diol intermediate with C⁴ of uracil. The original zinc-coordinated oxygen of uracil can then be exchanged for the hydroxyl through rearrangement of this intermediate. The zinc-coordinated oxygen abstracts the proton from the Glu64 carboxyl, which can finally shuttle one more proton from the hydroxyl, releasing water and returning the enzyme to its original state with zinc-coordinating water. The mechanism is presumed to be the same for 5FC, as C⁵ is uninvolved in the mechanism, and substrate preference is minimal [44].
**Thermostabilization**

Despite yCD’s ability to bind and deaminate 5FC, its use in EPT is limited by its thermolabile nature. The melting temperature of wild-type (WT) yCD is 52°C, which is not problematic in its native yeast host, but in *E. coli*, rodents, or humans the half-life of the wild-type enzyme is significantly shortened. To overcome this issue, yCD was thermostabilized computationally using RosettaDesign [45]. Three mutations (A23L/V108I/I140L) were identified in the core of the protein that stabilized yCD, increasing its melting temperature by 10°C and its half-life at 50°C more than 29-fold. The resulting “triple mutant” was shown to significantly improve the growth rate of an auxotrophic strain of bacteria in minimal medium without uracil at 37°C. As our therapeutic switches will need to function at 37° we have chosen to use this thermostable variant yCD-triple to construct cancer-marker activated enzymes. Additionally, as protein engineers we rely heavily on the rapid growth rate and ease of genetic manipulation of *E. coli*, which thrive at 37°C.

**Random mutagenesis**

This group has also attempted to improve the activity of yCD towards 5FC using random mutagenesis [44]. Eleven codons in and around the active site were mutated using synthetic oligonucleotides incorporating 21% non-native bases at each position within these chosen codons. After transformation, this library was immediately plated on medium devoid of uracil to select for active yCD variants. 34,000 transformants were obtained on a naïve plate, representing less than 1% of the maximum diversity of this library on the DNA-level. The 50 colonies surviving on the selection plate were struck out on decreasing levels of 5FC to measure the
5FC sensitivity each variant conferred. Three variants conferring increased 5FC sensitivity were identified: D92E, M93L, and I98L. These variants were then tested for their ability to increase the sensitivity to 5FC of stable pools of C6 Rat glioma cells. Only D92E was found to marginally increase toxicity of these cells relative to WT yCD, but to a lesser extent than yCD-triple. In combination D92E and the three mutations of yCD-triple counteracted each other decreasing sensitivity to a level similar to WT yCD. This negative epistatic effect was somewhat confirmed with kinetic assays on both cytosine and 5FC, with yCD-triple-D92E having a 3-fold decreased catalytic efficiency. However, the combined variant did demonstrate an increased substrate preference for 5FC. Although no improved variants resulted from this study it is possible that with improved mutagenesis technologies [46] and higher sampling rates improved variants can be found.

**HIF-1α and the CH1 domain of the transcriptional coactivator of p300**

As a cancer marker to activate the desired therapeutic enzyme switch we chose HIF-1α, for its prevalence across many cancer type and its tight regulation in normal tissue. As the input domain for the HIF-1α-activated protein switches, we chose the CH1 domain of the human p300 protein. In response to hypoxia, HIF-1α interacts with the transcriptional coactivator p300 through its CH1 domain to enhance transcription from promoters containing hypoxia response elements.

**HIF-1α, oxygen regulation, and cancer**

The work of Semenza—who discovered hypoxia inducible factor-1α in 1992—and others have established HIF-1α as a viable target for cancer therapy [47]...
as it has been shown to preferentially accumulate within tumors and metastases relative to surrounding normal tissue [24, 48]. As a regulator for suicide enzyme switches it is an exquisite target. In the presence of oxygen HIF-1α has a half-life of less than 5 minutes [49]. Due to this short half-life, HIF-1 is essentially undetectable in normal adult skin, bladder, thymus, lymph node, thyroid, liver, heart, colon, lung, ovary, uterus, brain, breast, stomach, and prostate [24] and has only been weakly detected in select cells of the adrenal gland, kidney, pancreas, tonsils, spleen, and testes [48].

In low oxygen tensions (<5%) or in the absence of functional regulation, HIF-1α is stabilized and translocated to the nucleus where it activates transcription. HIF-1α accumulation has been shown to be exponentially proportional to the level of hypoxia both in vitro and in vivo [50, 51]. HIF-1α maintains cellular oxygen homeostasis during development and wound healing through activating transcription of hypoxia-response genes. These genes, when expressed later in life, are frequently tied to neoplastic hallmarks such as glucose metabolism, angiogenesis, invasion, metastasis, genomic instability, and immortality [52].

HIF-1α degradation is regulated via prolyl-hydroxylation dependent ubiquitination. Under normoxic conditions, HIF-1α is hydroxylated at Pro402 and Pro564 by Prolyl Hydroxylase Domain proteins (PHDs). PHDs coordinate iron and catalyze hydroxylation via O₂ and α-ketoglutarate substrates, generating hydroxylated proline, succinate, and CO₂. The hydroxylated proline residues are then recognized by the Von Hippel Lindau tumor suppressor protein (VHL) which
Figure 1.4 – Regulation of HIF-1α signaling.

HIF-1α acts as the substrate recognition subunit of a ubiquitin ligase complex leading to ubiquitination and degradation of HIF-1α via the 26S proteasome [53]. Additionally, HIF-1α is hydroxylated, by Factor Inhibiting HIF (FIH-1) at asparagine residue 803 inhibiting transcriptional activation [54].

Under hypoxic conditions hydroxylation is inhibited due to lack of molecular oxygen and HIF-1α accumulates and forms a complex with HIF-1β, allowing the HIF-1 complex to be translocated to the nucleus. In the nucleus HIF-1 interacts with transcriptional coactivators p300 and CREB-binding protein (CBP) initiating transcription from hypoxia response elements (HRE), which contain the sequence 5’-RCGTG-3’. Additionally, HIF-1α accumulation can be stimulated by inhibiting PHD
activity using iron chelators (such as desferroxamine), CoCl₂, or
dimethyloxalylglycine (DMOG, a competitive inhibitor of α-ketoglutarate).

The CH1 domain of p300/CBP

To engineer HIF-1α regulated activity in yCD we needed to identify a HIF-1α
binding protein. This binding domain could either be a designed antibody mimetic
protein or an endogenous HIF-1α-binding domain. An endogenous binding domain
may minimize the immunogenicity of the desired therapeutic and does not require
significant development prior to switch engineering. Significant previous work has
established thorough biophysical and structural characterization of HIF-1α in
complex with the Cytidine/Histidine-rich-1 (CH1) domain of the human proteins
p300 and the very closely related CBP, we chose this domain to affect HIF-1α-
dependence upon yCD activity.

Structure of the CH1/HIF-1α complex

The interaction between HIF-1α and the highly conserved CH1 domain of
p300 and CBP has been well characterized [33, 55, 56]. The CH1 domain (residues
323-423 of p300, 339-439 of CBP) binds to the C-terminal activation domain
(CTAD) of HIF-1α (residues 786-826 minimally, 776-826 conservatively) with very
high affinity (7 nM) [33]. This interaction is facilitated by three zinc atoms
coordinated by HCCC motifs near the coils connecting the four α-helices of the CH1
domain. The four helices form a tetrahedral bundle, around which the CTAD of HIF-
1α is threaded. The normally unstructured CTAD undergoes local folding forming
Figure 1.5 – Structure of the HIF-1α CTAD and CH1 domain complex. (Left) Cartoon representations of HIF-1α CTAD (purple) and the CH1 domain of p300 (green). N- and C-terminal α-carbons are labeled blue and red, respectively. Zinc molecules are represented as grey spheres. Asp803 of HIF-1α is colored by element. (Right) Space filling model of the complex.

Three α-helices within a groove on the surface of the CH1 domain. This interaction shields an unusually large surface area of 3393 Å² [56], resulting in the appearance of this complex as a single domain.

Binding-induced structural changes are desirable in the effector domain of protein switches as they can induce allosteric activation or cellular accumulation. Such changes within the CH1 domain have been debated [55, 57], although—with the exception of NMR spectra—the disagreement between these authors appears mostly rhetorical. Dial et al. [55] conclude that the CH1 domain exists as a partially folded molten globule in the presence of 3 equivalents of zinc, and gains its more stable tertiary structure upon binding of HIF-1α. In this way, they describe the complex as cofolding as the HIF-1α CTAD is unstructured until it is bound by the CH1 domain. These conclusions are supported by CD and NMR spectroscopy, as well as tryptophan and ANS fluorescence. The CD and NMR experiments were repeated by DeGuzman who found similar results in the zinc saturating conditions of Dial et al.,
but additionally found that in the range of 1.5 to 3 equivalents of zinc the CH1 domain yields spectra indicative of a stable tertiary structure around which the unstructured HIF-1α CTAD can fold [57]. Thus at higher zinc concentrations some binding-induced folding may occur inducing activation of CH1-CD switches. However no measurements of the effective nuclear concentrations of zinc nor p300/CBP exist so one can only speculate as to the native state of the CH1 domain or therapeutic switches containing it.

**Binding partners of the CH1 domain**

A potential downfall of the CH1 domain lies in its binding promiscuity. As a portion of a transcriptional coactivator, this domain is known to bind many transcription factors which are likely to be less tightly regulated than HIF-1α. Many of these transcription factors are implicated in cancer, however. Table 1.1 lists the known binders of the CH1 domain along with their physiological functions and implications as cancer biomarkers. Discerning effective biomarkers from aberrant activators that contribute to off-target toxicity will be key to the further translation of CH1-domain-based switches.
<table>
<thead>
<tr>
<th>CH1 Binding Partner</th>
<th>NCBI Accession</th>
<th>Full name/Synonyms</th>
<th>Chromosome</th>
<th>Minimum interacting residues (Phosphorylated residue-affinity increase)</th>
<th>% Identity of interacting portion in Mice</th>
<th>Cellular Role</th>
<th>Correlation with cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIF-1α</td>
<td>NP_001521</td>
<td>Hypoxia inducible factor-1α</td>
<td>14q23.2</td>
<td>786-826 [56]</td>
<td>100%</td>
<td>Induces VEGF, EPO, NOS, glycolytic enzymes. [58]</td>
<td>Positive [59]</td>
</tr>
<tr>
<td>p53</td>
<td>NP_000537</td>
<td>Tumor suppressor p53</td>
<td>17p13.1</td>
<td>1-57 [60] (T18-2-4X; S15,T18,S20-10X) [61]</td>
<td>63%</td>
<td>DNA damage response, apoptosis/cell-cycle arrest/DNA repair in response to cellular stress signals [62]</td>
<td>Positive, Loss-of-function (LoF) implicated in 60% of malignancies [63, 64]</td>
</tr>
<tr>
<td>Tal1</td>
<td>NP_003180</td>
<td>T-cell acute lymphocytic leukemia 1</td>
<td>1p32</td>
<td>185-240 [65]</td>
<td>100%</td>
<td>Activation of endothelial cells in angiogenesis, hematopoiesis [66], osteoclast differentiation [67], erythroid differentiation [65]</td>
<td>Positive, overexpressed in T-cell acute lymphoblastic leukemia [68]</td>
</tr>
<tr>
<td>Cited2</td>
<td>NP_001161</td>
<td>CBP/p300 interacting transactivator 2, p35srj, MRG-1</td>
<td>6q23</td>
<td>220-269 [69], 215-270 [70]</td>
<td>100%</td>
<td>Negative regulator of hypoxic signaling (competes with HIF for CH1 binding) [71], cardiac development, neural development [72]</td>
<td>Positive, high levels are associated with tumor aggressiveness [73], elevated levels in breast cancers, potentially leads to osteotropism [74, 75], transformative [76]</td>
</tr>
<tr>
<td>Ets-1</td>
<td>NP_0013729</td>
<td>Protein C erythroblastocyst-1</td>
<td>11q23</td>
<td>29-138 (T38,S41-34X) [77]</td>
<td>100%</td>
<td>Ras/MAPK transcriptional effector</td>
<td>Positive, overexpressed, mis-activated</td>
</tr>
<tr>
<td>CH1 Binding Partner</td>
<td>NCBI Accession</td>
<td>Full name/ Synonyms</td>
<td>Chromosome</td>
<td>Minimum interacting residues (Phosphorylated residue-affinity increase)</td>
<td>% Identity of interacting portion in Mice</td>
<td>Cellular Role</td>
<td>Correlation with cancer</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------</td>
<td>---------------------</td>
<td>------------</td>
<td>-------------------------------------------------</td>
<td>----------------------------------------</td>
<td>--------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>p73</td>
<td>CAA72221</td>
<td>1p36</td>
<td>1-68 [78]</td>
<td>regulation of apoptosis, similar function to p53</td>
<td>93%</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>NFκB-p65 subunit</td>
<td>Q04206</td>
<td>11q13</td>
<td>286-551 [79], 314-550 [80]</td>
<td>inflammatory signaling [79]</td>
<td>76%</td>
<td>Positive [81]</td>
<td></td>
</tr>
<tr>
<td>Pit-1</td>
<td>P28069</td>
<td>Pituitary specific positive transcription factor 1 (POU1F1)</td>
<td>3p11</td>
<td>2-80 [82]</td>
<td>94%</td>
<td>Pituitary development [83], prolactin and growth hormone transcription [84], cell proliferation [85], anti-apoptosis</td>
<td>Positive, GoF (overexpression) in pituitary tumors [86] and breast carcinoma (MCF7 cells) [87]</td>
</tr>
<tr>
<td>HNF-4</td>
<td>P41235</td>
<td>hepatocyte nuclear factor 4</td>
<td>20q13</td>
<td>119-375 [88], 748-851 [93], 768-838 [94] binds in same groove but opposite orientation to HIF and CITED2</td>
<td>98%</td>
<td>liver enriched nuclear hormone receptor, metabolic regulator, [89, 90] binding transthyretin promoter [91]</td>
<td>Negative, LoF increases proliferation, dedifferentiation, and metastasis [92]</td>
</tr>
<tr>
<td>Stat-2</td>
<td>NP_938146</td>
<td>Signal transducer and activator of transcription -2</td>
<td>12q13</td>
<td>interferon-α signaling</td>
<td>38%</td>
<td>Positive, constitutively activated in many malignancies [95]</td>
<td></td>
</tr>
<tr>
<td>CH1 Binding Partner</td>
<td>NCBI Accession</td>
<td>Full name/Synonyms</td>
<td>Chromosome</td>
<td>Minimum interacting residues (Phosphorylated residue-affinity increase)</td>
<td>% Identity of interacting portion in Mice</td>
<td>Cellular Role</td>
<td>Correlation with cancer</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>GABPα</td>
<td>AFH41795</td>
<td>GA binding protein transcription factor alpha subunit 60kDa, E4 transcription factor-1 (E4TF-1)</td>
<td>21q21</td>
<td>35-121 [96], beta sheet rich</td>
<td>95%</td>
<td>embryogenesis, immune system development, cell cycle progression, neuromuscular junctions, protein synthesis, cellular respiration, viral pathogenicity</td>
<td></td>
</tr>
<tr>
<td>GATA-2</td>
<td>AAA35869.1</td>
<td>3q21</td>
<td></td>
<td>[97]</td>
<td>91%</td>
<td>Hematopoiesis/erythropoiesis, mast cell formation</td>
<td>Positive, LoF mutations and overexpression are linked to leukemia [98]</td>
</tr>
<tr>
<td>Skp2</td>
<td>Q13309</td>
<td>S-phase kinase associated protein 2</td>
<td>5p13</td>
<td>91-424 [99]</td>
<td>86%</td>
<td>Inhibits p53 dependent apoptosis, stimulates c-myc activity</td>
<td>Positive [100]</td>
</tr>
<tr>
<td>HPV E7</td>
<td></td>
<td>1-98 [101]</td>
<td></td>
<td></td>
<td></td>
<td>Papilloma virus transcriptional hijacking</td>
<td>Positive [102]</td>
</tr>
<tr>
<td>Epstein-Barr virus Antigen 3C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prothymosin α</td>
<td>P06454</td>
<td>[103]</td>
<td></td>
<td></td>
<td>99%</td>
<td>Immune function, chromatin remodeling</td>
<td></td>
</tr>
<tr>
<td>AIRE</td>
<td></td>
<td>[104]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH1 Binding Partner</td>
<td>NCBI Accession</td>
<td>Full name/Synonyms</td>
<td>Chromosome</td>
<td>Minimum interacting residues (Phosphorylated residue-affinity increase)</td>
<td>% Identity of interacting portion in Mice</td>
<td>Cellular Role</td>
<td>Correlation with cancer</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>C/ATF</td>
<td></td>
<td>CCAAT/enhancer binding proteins related to activating transcription factor</td>
<td></td>
<td>[105]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ad5 E1A</td>
<td></td>
<td></td>
<td>1-77 [101]</td>
<td></td>
<td></td>
<td>Adenovirus transcriptional hijacking</td>
<td>Positive, increases cell proliferation</td>
</tr>
</tbody>
</table>
Chapter 2: 1st generation switches

The first generation of HIF-1α-activated protein switches (Haps) was created by Dr. Chapman Wright [106], using techniques developed in the Ostermeier lab for making protein switches by domain insertion [18, 19, 20, 107]. The majority of methods, results, and figures in this section was originally detailed in [106]. As Dr. Chapman Wright performed the majority of this work, I will only detail methods that I developed and performed as well as those relevant to later chapters.

Library and selection

The first Haps libraries were made using DNA derived from constructs pDIM-C8-yCD and pDIM-C8-CH1. The parent plasmid pDIM-C8 contains the p15a and f1 origins, chloramphenicol acetyltransferase, and the tac promoter [108]. Plasmid pDIM-C8-yCD was made using the sequence of yCD from S. cerevisiae genomic DNA. This yeast coding sequence contains several rare codons in E. coli, which may have limited expression of the members of this library during selections. Additionally, Homo sapiens—the end user of these therapeutics—and yeast have very dissimilar codon usage tables, but it is generally accepted that codon optimization is much more important in E. coli due to its rapid growth. After initial cloning, the yCD-triple mutations (A23L/V108I/I140L) were incorporated by Quickchange mutagenesis (Stratagene). From this point forward yCD will always refer to yCD-triple. This DNA was then amplified in E. coli and randomly digested with dilute DNaseI purified and repaired as previously described [19, 20, 107].
pDIM-C8-CH1 was made using *E. coli* optimized coding sequence of amino acids 334-420 of p300, flanked on its 5’ end by a genetically encoded 2X(GSGGG) amino acid linker (with a BamHI site encoded in the first GS) and on the 3’ by 1X(GSGGG) and an additional BamHI site. Three CH1 inserts were made using this template: 1) 3mer, which encodes an N-terminal glycine and C-terminal GGS linker, 2) 5mer with an N-terminal glycine and C-terminal GGGGS, and 3) circularly permuted CH1. The 3mer and 5mer were made by PCR with linker encoding primers. Circular permuted CH1 was made by digesting pDIM-C8-CH1 with BamHI, isolating and cyclizing the CH1 domain [109], then randomly digesting the cyclic DNA with S1 nuclease [18].

The randomly cut pDIM-C8-yCD was ligated with a mixture of the three varieties of CH1 domain and *E. coli* were transformed with this ligation product to

---

Figure 2.1 - Schematic showing domain insertion method used to create the yCD-CH1 hybrid library.
The CH1 domain inserts (cpCH1, CH1-3-mer, and CH1-5-mer) were mixed in an equimolar ratio before they were used in the ligation mixture with singly cut pDIM-yCD plasmids.
create a naïve library of $9.6 \times 10^6$ members, approximately 25% of which contained inserts based on subsequent PCR screening. This library was subjected to a two-tiered selection, which is the basis for the similar selections described throughout this work. Initially a negative selection was used to remove γCD-CH1 fusions with γCD activity independent of HIF-1α. This was achieved by growing library members in strain GIA39, which lacks the native bacterial cytosine deaminase, in the absence of HIF-1α and the presence of 5FC, which is converted to its toxic form, 5FU by any active members. The negative selection was repeated twice with this library.

![Figure 2.2 - Schematic of the two-tiered selection.](image)

GIA39 cells are transformed with a naïve library of Haps variants. After being amplified on LB plates this libraries are plated on Off-state or Negative selection plates. Variants able to grow on this plate are inactive in the absence of HIF-1α-CTAD. DNA is isolated from these pooled variants and used to transform GIA39 cells containing pGA-HIF. Again after amplification the remaining fraction of the library is plated on Positive or On-state selection plates, which only allow variants containing active cytosine deaminases in the presence of HIF-1α-CTAD.
Initially at high stringency (75 μg/mL 5FC), followed by lower stringency (50 μg/mL 5FC). DNA was isolated from the remaining members and used to transform GIA39 containing plasmid pGA-HIF, which expresses the CTAD of HIF-1α under the control of the arabinose promoter.

The positive selection is based on the yCD activity of the active library members complementing the cytosine deaminase deficient strain. Thus in the absence of uracil and in the presence of an excess of cytosine, only members with an active yCD will be able to grow. The negatively selected members of the initial library were then subjected to positive selection at 25 μg/mL cytosine. Ninety-nine colonies formed on this positive selection plate and only 8 of these contained in-frame fusions of CH1 and yCD. Of these 8, one sequence was repeated twice and another 3 times. These repeated members, Haps3 and Haps59, were the only functional switches after screening. Haps3 contains the 3mer direct insertion CH1 domain inserted after the 8th amino acid of yCD (Figure 2.3). Haps59 contains a non-circularly permuted CH1 domain from the circular permutation library, which has no linker residues on its N-terminus and two glycine residues on its C-terminus.

**Figure 2.3 – Schematic representations of the sequences of Haps3 and Haps59**

Red represents yCD residues; blue represents residues of the CH1 domain; grey represents linker regions. Numbers above denote amino acid numbers in yCD, and letters below denote the amino acid sequence of the linkers in single-letter code.
The CH1 domain in Haps59 is also inserted after the 8th residue of yCD, but the 9th residue is deleted.

Characterizing activity in *E. coli*

Screening was initially done in liquid minimal media with and without arabinose (i.e. with and without HIF-1α) in the presence of increasing concentrations of 5FC (Figure 2.4). However, this method frequently yielded inconsistent results. It was often the case that one in three replicates would significantly outgrow the other two, and there was a consistent increase in growth observed at low levels of 5FC.

![Figure 2.4 - Screening Haps59 and Haps3 in liquid culture.](image)

**Figure 2.4 - Screening Haps59 and Haps3 in liquid culture.** Growth of GIA39 cells expressing MBP, yCD, or Haps59 and either GST or gstHIF-1a in minimal liquid media as a function of 5FC concentration. Media either omitted (open symbols) or contained (solid symbols) 0.15% arabinose to compare the effects of coexpressing GST or gstHIF-1a. (A) GIA39 cells expressing control proteins (MBP + GST, circles; MBP + gstHIF-1a, triangles; yCD + GST, diamonds; yCD + gstHIF-1a, squares). GIA39 cells expressing (B) Haps3 or (C) Haps59 with either GST (circles) or gstHIF-1a (triangles) demonstrate that gstHIF-1a increases the 5FC toxicity of cells expressing these switches. For all graphs, error bars, SD (N=6). Experiments with yCD (A) revealed that both GST and gstHIF-1a induction by arabinose caused a small but equal increase in 5FC toxicity. This effect likely arises from the added burden of high expression from the arabinose promoter to cells that are coping with near lethal levels of 5FC. In contrast, gstHIF-1a expression increased the 5FC toxicity of cells expressing Haps3 (B) and Haps59 (C) to a much greater extent than did GST expression. We attribute the increase in cell density observed at sublethal 5FC concentrations for cells expressing yCD, Haps3, or Haps59 to a stress response that allows growth to higher densities in the minimal media because the increase is not observed when MBP is expressed.
These results led us to explore other options for screening 5FC- and HIF-1α-dependent toxicity in *E. coli*. A common method for screening variants of antibiotic resistance genes and complementary genes to auxotrophic strains uses spots of serial dilutions of log phase cultures on plates of increasing concentrations of antibiotic or metabolite as in Foit et al. [110] Figure 2. A similar assay was devised and growth conditions and inducer concentrations were optimized to yield spots of equivalent density in the absence of 5FC for GIA39 cells coexpressing yCD, MBP, Haps3, or Haps59 and GST or GST-HIF-1α-CTAD. We tested this dot toxicity with positive and negative controls (yCD and MBP, respectively) as well as Haps3 and Haps59 (Figure 2.5). As shown here, a protein with no cytosine deaminase activity such as MBP will allow all dilutions to grow to high density. Cytosine deaminase and active switches will inhibit growth of lower dilutions with increasing 5FC. 5FC-dependent toxicity can be quantified as either the last surviving dilution at a particular 5FC concentration (as in Figure 3.3), or as the minimum inhibitory concentration (MIC) of 5FC for a particular dilution.

*Methods – Dot toxicity assay*

Fresh colonies were struck out from glycerol stocks onto LB agar plates containing 50 µg/mL chloramphenicol, 100 µg/mL ampicillin, and 0.2% glucose and grown overnight at 37°C. Colonies from the LB plates were inoculated into unselective liquid dropout medium (1X yeast nitrogen base, 1X yeast dropout media without uracil, 10 µg/mL uracil, 2% glucose, 100 µg/mL ampicillin, and 50 µg/mL chloramphenicol) and incubated overnight with constant shaking at 37°C. Overnight cultures were diluted 1/100 into fresh unselective liquid medium and incubated for
4-6 hours until \( \text{OD}_{600} \) was greater than 0.3. Cultures were diluted to 0.3 \( \text{OD}_{600} \) and serial dilutions were made such that the \( \text{OD}_{600} \) of the final dilution was less than 0.0003 (typically 8 3.33X-dilutions). One \( \mu \text{L} \) of each dilution was spotted by reverse pipetting on unselective solid media (2% select agar) supplemented with 1 mM IPTG, 0.05% (w/v) L-arabinose and 100 \( \mu \text{g/mL} \) ampicillin, containing increasing concentrations of 5-FC in omnitrays (Nunc) and incubated overnight at 37°C. For pouring omnitrays, 40 mL of medium was added to a 50 mL conical tube containing the appropriate amount 5FC, mixed gently for ~5 seconds, and poured into the omnitray, avoiding bubbles. If bubbles did make it into the plate they were moved to the edge of the plate with a sterile pipette tip, as oppose to running a flame across the plate to avoid 5FC degradation. After the plates solidified they were inverted and left slightly ajar allowing condensation to evaporate at room temperature for approximately 2 hours.

**Results – Dot toxicity assay**

This assay consistently demonstrated MBP and yCD containing cells were insensitive to coexpression of GST-HIF-1\( \alpha \)-CTAD, but cells expressing Haps3 and Hap59 demonstrated increased 5FC sensitivity in the presence of HIF-1\( \alpha \)-CTAD. Haps3 exhibited a decrease in MIC for the fourth 3.33X dilution from 300 \( \mu \text{M} \) 5FC to 200 \( \mu \text{M} \) with the addition of HIF-1\( \alpha \)-CTAD. Haps59 had a slightly larger window shifting from a MIC of >700 \( \mu \text{M} \) to nearly 300 \( \mu \text{M} \) in the presence of HIF-1\( \alpha \)-CTAD.
Figure 2.5 – Toxicity of 5FC to *E. coli* cells expressing Haps59 and Haps3 as assessed by dot toxicity assay.

Growth of GIA39 cells expressing MBP, yCD, Haps3, or Haps59 and GST (HIF-1α −) or GST-HIF-1α-CTAD (HIF-1α +).

Switching mechanism in *E. coli*

We next set out to identify the mechanism of these switches to understand the reason for the switching phenotype they conferred to *E. coli*. Previously our lab has elucidated two mechanisms for protein switches: allostery and stabilization. An allosteric switch is in an inactive conformation until binding of the effector molecule causes conformational activation of the switch. A switch—acting purely via stabilization—is active both in the presence and absence of its effector, but is also unstable in the absence of the effector and degraded rapidly. Thus, only in the presence of the effector will this switch accumulate in the cell. As there were significant difficulties in purifying Haps59 in the absence of GST-HIF-1α-CTAD, we suspected that it acted via the stabilization mechanism. To test this hypothesis lysates of cells expressing yCD or Haps59 and GST or GST-HIF-1α-CTAD were western blotted and probed with α-yCD antibodies. This blot shows that Haps59
does indeed accumulate to high levels (similar to those of yCD) only in the presence of GST-HIF-1α-CTAD.

![Figure 2.6 - Accumulation of Haps59 and yCD expressed in GIA39 coexpressing GST or GST-HIF-1α-CTAD.](image)

Cultures were grown in 25 mL unselective minimal media in 125 mL flasks until OD_{600} reached 0.2, when 1 mM IPTG was added to all cultures and 0.05% L-arabinose was added to half of the cultures. Cultures were grown for 12-16 hours after induction, at which point 6.8×10^8 cells were lysed using BugBuster™ Reagent (Novagen) following the manufacturers protocol, and adding 1/500 Protease Inhibitor Cocktail (Sigma). Lysates were separated and transferred to a PVDF membrane and probed with α-yCD antibodies.

**Characterizing activity in cancer cell lines**

As the end goal for these therapeutic switches is selective eradication of cancer cells, we next characterized the effects of these switches on RKO colorectal cancer cells and MCF7 breast cancer cells. These cell lines were chosen because they are known to accumulate HIF-1α in solid tumors but lack HIF-1α in normoxic culture. These cell lines were confirmed to be sensitive to 5FU at similar levels (Figure 2.7 A&D), RKO being slightly more sensitive than MCF7. Both cell lines were also insensitive to 5FC up to 2 mM. These cell lines were stabily transfected with an empty vector (EV), the same vector containing yCD, Haps3, or Haps59, and clones were picked that contained similar levels of transgene expression for characterization.
Figure 2.7 - 5FC and 5FU sensitivity experiments in RKO and MCF7 cells.

(A) Parental RKO cells treated with 5FU (solid blue diamonds) or 5FC (open blue circles). Each point represents the mean of six experiments performed on three separate days. Error bars, SD (N=6). (B) Positive and negative control experiments in stable RKO cell lines. The 5FC sensitivity of RKO cells stably expressing yCD (solid red symbols) or an empty vector control (open blue symbols) that were incubated in normoxic conditions in the absence (squares) or presence (circles) of 100 μM Co²⁺ or in 1% O₂ (diamonds). Error bars, SD (N=6). (C) Haps3-expressing RKO cells incubated in normoxic conditions show a small increase in sensitivity to 5FC in the presence of 100 μM Co²⁺ (solid red squares) over cells in the absence Co²⁺ (open blue squares). Error bars, SD (N=4). Controls from RKO cells expressing yCD or EV (B) are shown in gray for comparison. (D) Parental MCF7 cells treated with 5FU (solid blue diamonds) or 5FC (open blue circles). Each point represents the mean of nine experiments performed on three separate days. Error bars, SD (N=9). (E) Positive and negative control experiments in stable MCF7 cell lines. The 5FC sensitivity of MCF7 cells stably expressing yCD (solid red symbols) or an empty vector control (open blue symbols) that were incubated in normoxic conditions in the absence (squares) or presence (circles) of 100 μM Co²⁺ or in 1% O₂ (diamonds). Error bars, SD (N=9).

The HIF-1α-dependence of 5FC toxicity for these cell lines was tested by growing cells in increasing concentrations of 5FC and subjecting identical sets of plates to normoxia, hypoxia (1% O₂), and CoCl₂. In these cell lines both Haps3 and Haps59 were much more active than in *E. coli* in both the on- and off-states. Haps3 was so highly active that both its on- and off-states collapsed nearly onto yCD, almost eliminating its switching effect (switching ratio <2). Haps59 still exhibited a
therapeutic window in both cell lines, albeit narrower than in *E. coli* (Figure 2.8). We suspect this is due to expression of other CH1 binding partners.

**Figure 2.8 - Characterization of Haps59 in human cancer cells.**

(A–C) RKO colorectal cancer cells; (D–F) MCF7 breast cancer cells. (A and D) Co$^{2+}$ causes accumulation of HIF-1α in cells expressing Haps59, yCD, or the empty vector control as detected by Western blot with anti-HIF-1α antibodies. (B and E) Both Co$^{2+}$ (orange solid squares) and exposure to hypoxic growth conditions (red solid squares) increase the 5FC sensitivity of Haps59-expressing cells compared to Haps59-expressing cells not exposed to those conditions (open blue squares). The 5FC sensitivities of cells expressing yCD (solid diamonds) and EV control (solid triangle) are shown for comparison. Each point represents the mean from three different clones of each cell line, and each clone was tested in three separate experiments (error bars, SD, N=9). (C and F) Haps59 accumulates at higher levels in cells when HIF-1α is present. In these experiments a FLAG epitope was appended to the N terminus of Haps59 (FLAG-Haps59) for switch detection by Western blot. Cells were cultured under normoxic (N), hypoxic (H), or normoxic conditions with the addition of 100 μM cobalt chloride(Co$^{2+}$). Detection of β-actin served as a loading control.

In RKO cells expressing HIF-1α, Haps59 was very active, conferring an LC$_{50}$ of 65 μM 5FC in hypoxic conditions and 30 μM with cobalt. In normoxic conditions Haps59 was less active with an LC$_{50}$ of 325 μM. Thus Haps59 exhibited 5- to 11-fold switching in RKO cells. In MCF7 cells, Haps59 appears to be much less active in the presence of HIF-1α. The switching ratio in MCF7 cells is also lower (4- to 5-fold).
This may be partially due to expression differences. Additionally MCF7 cells are slightly less sensitive to 5FU, which may contribute to this effect.

**Switching mechanism in cancer cell lines**

Western blots on lysates of both RKO and MCF7 cells transiently transfected with either EV or FLAG-haps59 demonstrated that, as in *E. coli*, Haps59 is activated due to stabilization in the presence of HIF-1α (Figure 2.8 C & F). As further validation that the binding of HIF-1α is responsible for Haps59 stabilization and activation we performed co-immunoprecipitations using the same lysates as above to confirm this interaction (Figure 2.9). Indeed, HIF-1α and N-terminally FLAG-tagged Haps59 were coprecipitated using antibodies binding HIF-1α and the FLAG-tag of Haps59, verifying this their interaction. HIF-1α and yCD did not interact.

**Figure 2.9 - Haps59 and HIF-1α interact in RKO cells.**
In these experiments, a FLAG epitope was appended to the N terminus of Haps59 (FLAG-Haps59) for switch immunoprecipitation and detection. Western blot (WB) analysis of coimmunoprecipitation reactions with lysates of RKO cells expressing FLAG-Haps59, yCD, or an EV control show that HIF-1α is precipitated with anti-FLAG antibodies only in FLAG-Haps59-expressing cells (Top) and that FLAG-Haps59 is precipitated by anti-HIF-1α antibodies (Bottom). The far right lane (*) shows that coimmunoprecipitations of lysates of FLAG-Haps59-expressing cells using TATA-binding protein antibody [a negative control instead of the IP antibodies (Left)] did not result in detection of HIF-1α or FLAG-Haps59.
Chapter 3: 2nd generation switches

Introduction

Haps59 was a significant success as a proof of the therapeutic switch concept. However, the results in RKO and MCF7 cells left room for improvement as in the on-state Haps59 was several times less active than yCD and in the off-state there was still significant toxicity. In this study, we sought to create improved HIF-1α-activated protein switches via directed evolution, with the goal of expanding the therapeutic window of these switches and improving the chance of success in future translational studies.

We created three types of libraries in an attempt to identify switches with properties superior to Haps59 via positive and negative genetic selections. These three libraries were 1) a random mutagenesis library of Haps59 (Figure 3.1A), 2) a library in which the linker regions between the CH1 and yCD domains of Haps59 were varied in length and amino acid sequence (Figure 3.1B), and 3) a library in which all possible circularly permuted variants of the CH1 domain were inserted into all possible positions within yCD (described in Chapter 4).

Methods

Library creation

Error-prone PCR

Initially a PCR was performed using Phusion HF DNA polymerase and primers flanking genes encoding Haps59 and Haps3 to generate templates for the error-prone reaction. This template was purified using a Zymo DNA clean and
concentrate (DCC) 5 column (Zymo Research) and error prone PCR was performed as previously described [111], to achieve an average of approximately 3 nucleotide substitutions per 1000 bases. The resulting PCR product was purified using a Zymo DCC column and digested with NcoI and SpeI. The digested fragment was electrophoresed on an agarose gel and purified using Qiagen Gel Purification columns (Qiagen). This purified fragment was ligated into similarly digested pSkunk backbone using T4 DNA polymerase overnight at 16°C. DH5α cells were electroporated with the resulting library and plated on a 24.5 × 24.5 cm bioassay dish of LB agar containing 50 μg/mL streptomycin and 0.2% glucose. After overnight growth, this naïve library was recovered from the plate and plasmid DNA isolated by miniprep (Qiagen) for transformation into selection strains, as described below.

*Linker mutagenesis*

Oligonucleotides encoding 0, 1, 2, and 3 NNK degenerate codons and annealing to the fusion points between the N-terminal portion of yCD and the CH1 domain and the CH1 domain and the C-terminal portion of yCD were synthesized by IDT. Each N- and C-terminal linker pair of oligos was used to perform Kunkel Mutagenesis [112].

For this technique, single-stranded uracil-containing template DNA (ssdUDNA) must be purified from phage, which was produced by CJ236 (dut ung f1 ‘(Cmr), Coli Genetic Stock Center) containing pSkunk-Hap59 grown in the presence of R408 helper phage (Promega). This was done by first transforming pSkunk-haps59 into CJ236 and plating on LB Agar supplemented with 25 μg/mL
chloramphenicol and 50 μg/mL streptomycin. These plates were incubated overnight at 37°C. Although it was not done here, it has since been found that the initial cultures should be grown at 30°C in the presence of 125 μg/mL deoxythymidine to minimize detrimental mutations caused by the dut-ung phenotype [46]. A single colony from this plate was inoculated into 10 mL LB medium supplemented with 50 μg/mL streptomycin and 25 μg/mL chloramphenicol and incubated overnight at 37°C. In the morning, the density of this culture was measured and 2×10⁷ cells (using the heuristic 2×10⁸ cells/mL=1 OD₆₀₀) were inoculated into 2 mL TBG medium [113] containing 1×10⁸ plaque forming units (PFU) R408 helper phage (Promega) and supplemented with 50μg/mL streptomycin in an 18 mm test tube. This culture was incubated for 6 hours at 37°C, at which point the entire culture was centrifuged at 16k × g and the phage containing supernatant was moved to a new tube. A total of 150 μL of 2.5 M NaCl, 20% PEG-8000 solution was added to each mL of supernatant and mixed by briefly vortexing. This solution was incubated at 4°C overnight to precipitate the phage particles. In the morning, the precipitate was pelleted by centrifugation (21k × g, 10 min, 4°C). The pelleted phage was resuspended in 150 μL PBS. ssdUDNA was purified from the phage particles using the QIAprep Spin M13 kit (Qiagen) per the manufacturer’s protocol.

One μg of this DNA was used as a template for second strand synthesis primed by the phosphorylated random linker encoding oligos in a reaction containing T4 polymerase and T4 ligase. N- and C-terminal primer pair mixes were made and 4 molar equivalents (to the ~0.76 pmol of vector in 1 μg) were
phosphorylated in 20 µL of 1X T4 polynucleotide kinase buffer with 1mM ATP containing 5 units of T4 polynucleotide kinase. This reaction was incubated at 37°C for 1 hour and then heat inactivated at 65°C for 20 minutes. One µg of vector ssdUDNA was then added to the phosphorylated oligonucleotides and hybridized by heating to 94°C and slowly cooling over 30 min to 4°C in an Eppendorf MasterCycler thermocycler. At this point the reaction was diluted to a total volume of 100 µL containing 1X T4 DNA ligase buffer, 0.5 mg/mL BSA, 0.5 µM each dNTPs, 3 units T4 DNA polymerase, and 400 cohesive end units of T4 ligase. The reaction was incubated in a thermocycler with the following program: 5 minutes at 4°C, 5 minutes at 25°C, 2 hours at 37°C, 20 minutes at 25°C, 20 minutes at 75°C, hold at 4°C. The resulting double stranded products were purified using a DNA clean and concentrator kit (Zymo Genetics).

The purified products were electroporated into DH5α cells, which degrade the uracil-containing strand and synthesize a new complement strand now containing the mutations encoded by the oligonucleotides. Each library was transformed separately and plated on an individual 24.5 × 24.5 cm bioassay dish of LB agar containing 50 µg/mL streptomycin and 0.2% glucose. After overnight incubation at 37°C, the naïve libraries were recovered from these plates in minimal sweep medium (1X M9 salts, 2% glucose, 15% glycerol) and centrifuged at 3k × g for 10 minutes and decanted. The cell pellet was resuspended in one pellet volume minimal sweep medium and aliquotted for storage at -80°C. A rough estimate of mutants was made by screening colonies on dilution plates by PCR using primers annealing to the wild-type linkers, at a temperature that would give a product on the
wild-type construct. Plasmid DNA was isolated from a single library aliquot by miniprep (Qiagen) and used to transform selection strains.

Two-tiered selections

Each library was independently subjected to the two-tiered selection (Figure 2.2). Following the isolation of naïve library DNA, 25 ng was transformed into GIA39 cells for negative selections. Transformants were plated on LB agar containing 0.2% glucose and 50 μg/mL streptomycin in 24.5 × 24.5 cm bioassay dishes. After overnight incubation at 37°C, the resulting colonies were recovered from the plate with minimal sweep medium and the collected cells were centrifuged for 10 minutes at 3k × g. Supernatant was discarded and library members were resuspended in 1 cell pellet volume of fresh minimal sweep medium. Small aliquots were then stored at -80°C. A single aliquot was thawed and several dilutions were plated on unselective solid medium (1X yeast nitrogen base, 1X yeast dropout medium without uracil, 2% select agar, 5 μg/mL uracil, 2% glucose, and 50 μg/mL streptomycin, 1mM IPTG) to determine the concentration of viable cells. Based on this concentration, a volume corresponding to 600,000 cells from a freshly thawed aliquot was plated per 24.5 cm square bioassay dish in each selection, unless this is greater than 10 times the library size, in which case 10-times-the-maximum-theoretical-diversity-of-the-library cells were plated. Dilutions were plated on 10 cm petri dishes of selective and unselective media to confirm the number of cells plated and to calculate the percent of surviving library members.
**Negative selection**

Library members were plated on unselective medium supplemented with 500-1000 μM 5-fluorocytosine. After 18-24 hours of growth at 37°C, surviving members were recovered from the plate in minimal sweep medium by sweeping, centrifuged at 3k × g for 10 minutes, resuspended in 1 cell pellet volume of fresh minimal sweep medium, aliquoted, and stored at -80°C. If greater than 30% of the plated members survived the selection, the selection was repeated. If not, library plasmid DNA was prepped from an aliquot of the selected members and GIA39 cells harboring pGA-HIF (GIA39-HIF) were electroporated with 25 ng of this DNA and plated on LB agar containing 0.2% (w/v) glucose, 100 μg/mL carbenicillin, and 50 μg/mL streptomycin. Selected library members were collected, aliquoted, and stored as described above.

**Positive selection**

Negatively selected libraries in GIA39-HIF were plated on dropout medium without uracil (1X yeast nitrogen base, 1X yeast dropout medium without uracil, 2% select agar, 2% glucose, 50 μg/mL streptomycin, 100 μg/mL carbenicillin, 1mM IPTG, 0.05% (w/v) L-arabinose) supplemented with cytosine at levels between 500 ng/mL and 5 μg/mL. Selective plates were incubated at 37°C for 36 hours prior to picking the largest colonies for screening.

**Screening**

Colonies isolated from positive selection plates were struck out on unselective plates with and without L-arabinose and supplemented with 300 μM
5FC (as well as an LB plate for further propagation) as an initial assessment of switching. In cases where a large number of colonies survived the final tier of selection, these colonies were inoculated in 1 mL unselective minimal medium in a 96-well block, struck out on LB plates, and then incubated overnight at 37°C. In the morning the liquid cultures were diluted 1/100 into fresh medium and incubated for an additional 4-6 hours, before diluting and spotting (as described below) on unselective plates with and without L-arabinose and supplemented with 300 μM 5FC. Growth on plates with and without L-arabinose was then compared to assess switching.

Colonies from the LB plates of members with the largest ratio of growth in the absence versus presence of arabinose were incubated overnight in LB media containing 50 μg/mL streptomycin at 37°C. Plasmid DNA was isolated by miniprep and sequenced. These members were then subcloned into fresh pSKunk vector and confirmed by Sanger sequencing. Chemically competent [114] GIA39 containing pGA or pGA-HIF were transformed with the subcloned switches. Fresh colonies of these strains were inoculated into unselective liquid dropout medium (1X yeast nitrogen base, 1X yeast dropout medium without uracil, 10 μg/mL uracil, 2% glucose, 100 μg/mL ampicillin, and 50 μg/mL streptomycin) and incubated overnight with constant shaking at 37°C. Overnight cultures were diluted 1/100 into fresh unselective liquid medium and incubated for 4-6 hours until OD$_{600}$ was greater than 0.3. Cultures were diluted to 0.3 OD$_{600}$ and serial dilutions were made such that the OD$_{600}$ of the final dilution was less than 0.0003. One μL of each dilution was spotted by reverse pipetting on unselective solid media supplemented with 0.05%
(w/v) L-arabinose and 100 μg/mL ampicillin, containing increasing concentrations of 5-FC in omnitrays (Nunc) and grown overnight at 37°C. Members exhibiting a high ratio of growth in GIA39-pGA relative to GIA39-HIF were chosen for further study.

Expression and western blots

Single colonies of GIA39 containing switch variants in pSKunk and either pGA or pGA-HIF were inoculated in 5 mL of unselective liquid medium supplemented with an additional 5 μg/mL uracil at 37°C with shaking for 12 hours. This culture was used to innoculate 25 mL of high uracil unselective liquid medium in a 125 mL erlenmeyer flask to 0.02 OD₆₀₀. Cultures were grown to 0.1-0.3 OD₆₀₀ and induced with 0.05% (w/v) arabinose and 1mM IPTG final. Cultures were incubated for an additional 2 hours to overnight and 680 million cells were spun down in a 15 mL conical tube, decanted, and stored at -80°C. After freezing, 500 μL of Bug Buster Reagent containing (1/500) protease inhibitor cocktail (Sigma), r-Lysozyme and Benzonase nuclease (Novagen) was used to lyse the cells, per the manufacturers protocol. Protein concentrations of lysates were measured using DC protein assay (Bio-Rad).

Equal amounts of protein were electrophoresed on a 4-12% Bis-Tris NuPAGE gel (Life Technologies) and transfered via TransBlot SD to a PVDF membrane (Bio-Rad). Membranes were blocked with 5% milk (Bio-Rad) in tris-buffered saline + 0.5% tween-20 (TBST). Sheep polyclonal anti-yCD antibodies (Thermo) were diluted 1/500 into TBST + 5% milk and the blot was incubated overnight in this solution at 4°C with gentle shaking. In the morning, the blot was washed briefly in
TBST and prepared for the Snap ID protein detection system (Millipore) as recommended by the manufacturer. The blot was washed 3 times with 75 mL of TBST in the Snap ID and then incubated for 30 minutes at room temperature with a 1/6000 dilution of donkey anti-sheep secondary antibodies (Bethel Laboratories) in TBST+0.5% (w/v) milk. The blot was washed again as before, prior to development with Clarity Western ECL substrate in a Universal Hood II (Bio-Rad).

**Results**

**Random mutagenesis libraries**

Random mutagenesis is the most commonly used method by both nature and protein engineers for improving proteins. We chose to use error-prone PCR (ePCR), an extremely simple, classic technique, to identify beneficial mutations within Haps3 and Hap59. ePCR is limited to making single base changes at a time, limiting the sequence space of our search to ~30% of the total single amino acid mutations. This limited search may minimize deleterious mutations however, due to the conservative nature of the genetic code [115]. Regardless, this technique would allow us to look at mutations throughout the entire gene expanding upon the focused mutagenesis of Stolworthy et al. [44].

We hypothesized that mutations in Haps59 or Haps3 might be able to accentuate its switching property. ePCR libraries of both switch genes were constructed under conditions designed to achieve on average a single non-synonymous mutation per variant. A total of 1.2 million members from these
libraries—representing in theory all of the single mutants and 25% of double mutants—were subjected to two-tiered selections.

**Figure 3.1** - Strategies for creating improved HIF-1α-activate protein switches. Numbers above indicate the amino acid residues in yCD. Letters below indicate the 1-letter abbreviation amino acid sequence of the corresponding linker. (A) Schematic of the linker libraries created in Haps59. Lines indicate which combinations of N- and C-terminal linkers were combined to comprise the 18 libraries. (B) Schematic of random mutagenesis libraries of Haps59 and Haps3. * indicates point mutations.

No variants with improved switching windows were identified. However, Ehaps08, a variant of Haps59 with an Asp to Lys mutation in the C-terminal residue (Figure 3.3A), exhibited increased activity in both the on- and off-states, while retaining a moderate therapeutic window (Figure 3.3B&C). Five separate colonies were isolated with this mutation, which results from a single G to A transition. This mutation likely increases the thermodynamic or proteolytic stability of the switch as it significantly increased the expression level (Figure 3.4, Figure 3.6). Potentially this mutation may provide a better starting point for future directed evolution experiments.
Table 3.1 – List of ePCR library members having similar activity to Haps59.
Amino acid mutations are listed in 1-letter code as numbered in the parent protein.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Library</th>
<th>Mutations in N-yCD</th>
<th>Mutations in CH1</th>
<th>Mutations in C-yCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ehaps01</td>
<td>haps59</td>
<td></td>
<td>K28R</td>
<td></td>
</tr>
<tr>
<td>Ehaps02</td>
<td>haps59</td>
<td></td>
<td>N26D</td>
<td>T186A</td>
</tr>
<tr>
<td>Ehaps04</td>
<td>haps3</td>
<td></td>
<td>V54I</td>
<td>E249K</td>
</tr>
<tr>
<td>Ehaps07</td>
<td>haps59</td>
<td>M6I</td>
<td>S65G</td>
<td>E249K</td>
</tr>
<tr>
<td>Ehaps08</td>
<td>haps59</td>
<td></td>
<td></td>
<td>E249K</td>
</tr>
<tr>
<td>Ehaps09</td>
<td>haps59</td>
<td></td>
<td>A69V</td>
<td>E249K</td>
</tr>
<tr>
<td>Ehaps10</td>
<td>haps3</td>
<td></td>
<td>V54I</td>
<td></td>
</tr>
<tr>
<td>Ehaps11</td>
<td>haps59</td>
<td></td>
<td>H25R</td>
<td>E249K</td>
</tr>
<tr>
<td>Ehaps14</td>
<td>haps3</td>
<td></td>
<td>V122I</td>
<td>F205Y K206R</td>
</tr>
</tbody>
</table>

Linker libraries

Haps3 and Hap59 differed only in the linker regions between the CH1 and yCD domains. We hypothesized that the length and composition of the inter-domain linkers is important for optimizing switching activity. We created a total of 18 libraries designed to encompass all possible 1.14 billion unique combinations of N- and C-terminal linkers of 0, 1, 2, and 3 random amino acids as well as a C-terminal WT Haps59 Gly-Gly linker (Figure 3.1).

A total of 16.6 million of these variants were subjected to the two-tiered selections under conditions that should enrich for members with increased on-state activity and decreased off-state activity. The stringency of the positive selection was increased by decreasing the amount of cytosine in the medium, and the stringency of the negative selection was increased by increasing the concentration of 5FC. The majority of members surviving both selections had nearly identical activity to Haps59 as assessed in *E. coli* cells (Figure 3.2). These members predominantly have no N-terminal linker and a 2 amino acid C-terminal linker composed of bulky,
hydrophilic amino acids. Only a few members contained N-terminal linkers, which apparently are detrimental to switching in general (Table 3.2).

All identified switches possessed essentially equivalent activity to Haps59 in E. coli after subcloning except Ehaps22, which showed improved switching properties (Figure 3.2). Ehaps22 contains an N-terminal Lys-Phe linker and a C-terminal Arg-Cys linker (Figure 3.3A). The improved performance of Ehaps22 provided a significant increase in the toxicity of 5FC in the presence of HIF-1α, up to a level approaching the toxicity provided by yCD (Figure 3.3B&C). The toxicity of 5FC in the absence of HIF-1α also increased, but the increase was marginal.

We had hoped to identify switches in which the on-state activity was increased and the off-state activity was decreased. However, no isolates from the linker libraries derived from Haps59 exhibited both an improved off-state and wider therapeutic window. We hypothesize however that by tuning the expression level of Haps in cancer cell lines the therapeutic window can be shifted to decrease the toxicity of 5FC in the absence of HIF-1α accumulation.

**Figure 3.2 – Screen of the best variants from random mutagenesis and linker libraries.**
DNA from members having the highest ratio of growth in the presence versus absence of arabinose in initial screens was transformed into GIA39-pGA and GIA39-HIF for further analysis.
Table 3.2 – List of linker library members with similar therapeutic windows to Haps59.

<table>
<thead>
<tr>
<th>Clone</th>
<th>N-terminal Linker</th>
<th>C-terminal linker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ehaps03</td>
<td>-</td>
<td>Arg-Arg</td>
</tr>
<tr>
<td>Ehaps05</td>
<td>-</td>
<td>Cys-Lys</td>
</tr>
<tr>
<td>Ehaps06</td>
<td>-</td>
<td>Gly-Val</td>
</tr>
<tr>
<td>Ehaps07</td>
<td>-</td>
<td>Ser-Arg</td>
</tr>
<tr>
<td>Ehaps12</td>
<td>-</td>
<td>Gly-Gln</td>
</tr>
<tr>
<td>Ehaps13</td>
<td>-</td>
<td>Gly-Pro</td>
</tr>
<tr>
<td>Ehaps16</td>
<td>-</td>
<td>Lys-Lys</td>
</tr>
<tr>
<td>Ehaps18</td>
<td>-</td>
<td>Gly-Arg</td>
</tr>
<tr>
<td>Ehaps31</td>
<td>Cys</td>
<td>Gly-Gly-Arg</td>
</tr>
<tr>
<td>Ehaps21</td>
<td>Ala-Val</td>
<td>Gly-Arg</td>
</tr>
<tr>
<td>Ehaps22</td>
<td>Lys-Phe</td>
<td>Arg-Cys</td>
</tr>
<tr>
<td>Ehaps32</td>
<td>Gly-Ser-Arg</td>
<td>Ala-Leu</td>
</tr>
</tbody>
</table>

Figure 3.3 - Characterization of Ehaps22 and Ehaps08 in E. coli.

(A) Diagram of Ehaps22 and Ehaps08 sequences. Ehaps22 was selected from the linker library containing 2-amino acid N- and C-terminal linkers. Ehaps08 was selected from the random mutagenesis library of Haps59. Numbers above indicate the amino acid residues in yCD. Letters below indicate the 1-letter abbreviation amino acid sequence of the corresponding linker or mutation. (B) 5FC dot toxicity assays of cells expressing Ehaps22 or Ehaps08. Dilutions of equal density log phase cultures containing either pGA (-HIF, i.e. not expressing HIF-1α) or pGA-HIF (+HIF, i.e. expressing HIF-1α) were spotted on minimal media plates containing increasing concentrations of 5FC. (C) Quantification of dot toxicity assay. The highest dilution at which growth was observed for each culture is plotted against the concentration of 5FC in each plate.

Mechanism of improved switches

Previous studies with Haps59 in both E. coli and cancer cell lines indicated that it functioned by a mechanism involving increased accumulation in the presence
of HIF-1α [106]. We performed western blots on the lysates of E. coli cells expressing Ehaps22 and Ehaps08 (with and without HIF-1α-CTAD expression) to characterize the mechanism of these switches.

We found the accumulation of all switches to be very sensitive to culture and lysis conditions. Inconsistencies in the relative accumulation of switches between samples with and without HIF-1α-CTAD were seen from trial to trial. Induction later in log phase seemed to increase accumulation in GST-HIF-1α-CTAD expressing samples relative to those only expressing GST (from highest induction OD$_{600}$ to lowest: Figure 3.4 top right, top left, Figure 3.6 note that this blot was of cultures only grown for 2 hours as opposed to 6). For induction near the beginning of log phase (Figure 3.5 and Figure 3.6) there is no apparent difference in accumulation of Haps59, with or without HIF-1α. This may be due to differences in chaperone or

![Figure 3.4 - Time course western blot of Ehaps08 and Ehaps22.](image)

In (A) Cultures were grown to 0.2 OD$_{600}$ and induced. In (B) Cultures were induced at 0.35 OD$_{600}$. HIF – indicated coexpression of GST, and HIF + indicates coexpression of GST-HIF-1α-CTAD. Samples were taken at several timepoints normalized to cell density and lysed. Lysate was normalized to total protein prior to separation and western blot analysis. Haps59 and its variants are ~28 kDa. yCD is ~18 kDa. (C) Growth curve was made using the average OD$_{600}$ across all cultures. Error bars represent one standard deviation.
protease expression at these different points throughout the growth curve. The rate of IPTG and arabinose import or expression dynamics of the tac and arabinose promoters may also contribute to this phenomenon.

To assess whether the differences in Haps59 accumulation between these experiments and those performed previously [106] were due to ineffective protease inhibitors in the previous experiments, we lysed samples in the absence protease inhibitors and also heat-treated samples in the absence of protease inhibitors (Figure 3.5). There are no stark differences between yCD expressing samples or Haps59 expressing samples, or between those expressing GST versus GST-HIF-1α-CTAD. Based on this result, native E. coli proteases may not be very active in this lysis buffer. A more telling experiment would be to evaluate the protease susceptibility of Haps59 in the presence of GST or GST-HIF-1α-CTAD prior to lysis in a pulse-chase fashion, using chloramphenicol to inhibit further expression and observing degradation over time. These experiments demonstrated that Haps59 may be more stable in the absence of GST-HIF-1α-CTAD than originally thought, especially in the window a few hours after induction. Also counter to our previous results, the significantly increased levels of expression seen here may bode well for future attempts at purification and in vitro characterization of these switches.

Regardless of the somewhat conflicting results to those obtained previously [106], we can make some general conclusions about Ehaps08 and Ehaps22. In all western blots, except Figure 3.6, Ehaps08 accumulates to higher levels than Haps59. This may be the result of decreased protease susceptibility due
Figure 3.5 – Effects of the addition of protease inhibitors on Haps59 degradation. Cultures were grown to 0.1 OD$_{600}$ and induced. Three samples were taken at each time point. Two samples were lysed in Bugbuster without protease inhibitors (−PI). One of those samples was heat-treated at 37°C for 30 minutes (HT). The final sample was lysed with 1/500 protease inhibitors as usual (+PI). After lysis and treatment samples were spun down at 21k × g for 20 minutes to pellet precipitated proteins, before preparing western blots.

![Western blot image]

Figure 3.6 – Western of cultures induced at 0.1 OD$_{600}$ and grown for 2 hours. Cultures were induced in early log phase to express GST (HIF −) or GST-HIF-1α-CTAD (HIF +) and yCD, Haps59, Ehaps22, or Ehaps08. After 2 hours cultures were lysed and analyzed by western blot. Anti-yCD blot was stripped and reprobed with anti-GroEL as a loading control.

to C-terminal mutation, as *E. coli* is known to express many C-terminal proteases [116]. This hypothesis is supported by the abundance of degradation products in Figure 3.4.12 and 20 hour blots, as well as the fact that at early time points (Figure 3.6) expression of all switches in all conditions is stronger. This effect may also be generally increased stability and/or increased catalytic efficiency as there is a
potential for the amine group of the mutant lysine to form an additional salt bridge with the backbone carboxyl group of Ile156 which for the rotamer depicted in (Figure 3.7) are separated by 3.1 Å. Ehaps22 expression was generally similar to Haps59 so we hypothesize that the linkers in this variant allow the active-site-containing C-terminal portion of yCD to adopt a more catalytically active conformation in the presence of HIF-1α relative to Haps59.

![Wild-type yCD/Haps59](image1)

![yCD-E158K/Ehaps08](image2)

**Figure 3.7 – Structural model of Haps59 and Ehaps08 (E249K mutation).** (Top) Surface model of yCD-triple dimer crystal structure (PDB: 1YSB), which is the wild type in the case of our error-prone PCR libraries, and (Bottom) E158K mutation (E249K in Haps59) as modeled using Pymol [117] mutation wizard. Rotamer was chosen by intuition. Coloring is by element with nitrogen as blue, oxygen as red, sulfur as yellow, carbon as green, except for carbons of the C-terminal residues, which are colored white.
**Table 3.3 - ePCR and Linker library statistics and selections.**

Calculations were performed as described by Bosley and Ostermeier [155]. Percent mutants in ehaps libraries were calculated as the percent of clones with an insert. For linker libraries percent mutants was determined using wild-type specific primers in PCR to determine the percent wild type and any clone not displaying a wild-type band was considered a mutant. Survival and transformants were calculated by counting colonies on dilution plates (nm, not measured; nd, not done).

<table>
<thead>
<tr>
<th>Library</th>
<th>N-terminal linker length (AA)</th>
<th>C-terminal linker length (AA)</th>
<th>Theoretical Max Diversity</th>
<th>Transformers</th>
<th>% mutants</th>
<th>Degeneracy</th>
<th>Prob. Complete</th>
<th>Negative 1 (5FC, μg/mL)</th>
<th>Selection size</th>
<th>% survival</th>
<th>Negative Selection 2 (5FC, μg/mL)</th>
<th>Positive Selection (Cytosine, μg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ehaps3</td>
<td></td>
<td></td>
<td></td>
<td>9.32E+06</td>
<td>41.7%</td>
<td></td>
<td></td>
<td>125</td>
<td>6.00E+05</td>
<td>nm</td>
<td>125</td>
<td>0.625</td>
</tr>
<tr>
<td>ehaps59</td>
<td></td>
<td></td>
<td></td>
<td>3.00E+06</td>
<td>23.0%</td>
<td></td>
<td></td>
<td>125</td>
<td>6.00E+05</td>
<td>nm</td>
<td>125</td>
<td>0.625</td>
</tr>
<tr>
<td>L01haps59</td>
<td>0</td>
<td>1</td>
<td>3.20E+01</td>
<td>9.26E+06</td>
<td>26.7%</td>
<td>3.20E+01</td>
<td>1.00</td>
<td>125</td>
<td>1.00E+04</td>
<td>7.2%</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L02haps59</td>
<td>0</td>
<td>2</td>
<td>1.02E+03</td>
<td>2.00E+06</td>
<td>35.7%</td>
<td>1.02E+03</td>
<td>1.00</td>
<td>100</td>
<td>1.50E+04</td>
<td>59.0%</td>
<td>125</td>
<td>0.625, 5.05</td>
</tr>
<tr>
<td>L03haps59</td>
<td>0</td>
<td>3</td>
<td>3.28E+04</td>
<td>3.66E+06</td>
<td>43.0%</td>
<td>3.28E+04</td>
<td>1.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>9.9%</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L1Whaps59</td>
<td>1</td>
<td>GG</td>
<td>3.20E+01</td>
<td>4.00E+06</td>
<td>66.0%</td>
<td>3.20E+01</td>
<td>1.00</td>
<td>125</td>
<td>1.00E+04</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L10haps59</td>
<td>0</td>
<td>1</td>
<td>3.20E+01</td>
<td>8.85E+05</td>
<td>62.0%</td>
<td>3.20E+01</td>
<td>1.00</td>
<td>125</td>
<td>1.00E+04</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L11haps59</td>
<td>1</td>
<td>1</td>
<td>2.56E+02</td>
<td>1.00E+07</td>
<td>40.0%</td>
<td>2.56E+02</td>
<td>1.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>38.5%</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L12haps59</td>
<td>1</td>
<td>2</td>
<td>3.28E+04</td>
<td>1.97E+05</td>
<td>66.7%</td>
<td>3.22E+04</td>
<td>0.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>10.0%</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L13haps59</td>
<td>1</td>
<td>3</td>
<td>1.05E+06</td>
<td>9.16E+07</td>
<td>41.7%</td>
<td>1.05E+06</td>
<td>1.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>59.0%</td>
<td>125</td>
<td>0.625</td>
</tr>
<tr>
<td>L2Whaps59</td>
<td>2</td>
<td>GG</td>
<td>2.56E+02</td>
<td>1.78E+08</td>
<td>41.7%</td>
<td>2.56E+02</td>
<td>1.00</td>
<td>125</td>
<td>1.00E+05</td>
<td>81.2%</td>
<td>125</td>
<td>0.625</td>
</tr>
<tr>
<td>L20haps59</td>
<td>0</td>
<td>2</td>
<td>2.56E+02</td>
<td>2.22E+08</td>
<td>33.3%</td>
<td>2.56E+02</td>
<td>1.00</td>
<td>125</td>
<td>1.00E+05</td>
<td>34.6%</td>
<td>125</td>
<td>0.625</td>
</tr>
<tr>
<td>L21haps59</td>
<td>0</td>
<td>3</td>
<td>3.28E+04</td>
<td>2.58E+04</td>
<td>58.3%</td>
<td>2.58E+04</td>
<td>0.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L22haps59</td>
<td>0</td>
<td>2</td>
<td>6.55E+04</td>
<td>1.49E+07</td>
<td>73.3%</td>
<td>6.55E+04</td>
<td>1.00</td>
<td>125</td>
<td>6.00E+05</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L23haps59</td>
<td>0</td>
<td>3</td>
<td>3.36E+07</td>
<td>1.80E+07</td>
<td>61.5%</td>
<td>9.43E+06</td>
<td>0.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>14.9%</td>
<td>nd</td>
<td>5, 0, 0.625</td>
</tr>
<tr>
<td>L3Whaps59</td>
<td>3</td>
<td>GG</td>
<td>3.28E+04</td>
<td>7.53E+07</td>
<td>54.6%</td>
<td>3.28E+04</td>
<td>1.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L30haps59</td>
<td>0</td>
<td>3</td>
<td>3.28E+04</td>
<td>1.69E+08</td>
<td>66.7%</td>
<td>3.28E+04</td>
<td>1.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L31haps59</td>
<td>0</td>
<td>1</td>
<td>1.05E+06</td>
<td>1.05E+06</td>
<td>70.0%</td>
<td>1.05E+06</td>
<td>1.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>63.3%</td>
<td>125</td>
<td>0.625</td>
</tr>
<tr>
<td>L32haps59</td>
<td>0</td>
<td>2</td>
<td>3.36E+07</td>
<td>3.76E+05</td>
<td>60.0%</td>
<td>2.25E+05</td>
<td>0.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
<tr>
<td>L33haps59</td>
<td>0</td>
<td>3</td>
<td>1.07E+09</td>
<td>1.30E+09</td>
<td>83.3%</td>
<td>6.82E+08</td>
<td>0.00</td>
<td>125</td>
<td>5.00E+05</td>
<td>nm</td>
<td>nd</td>
<td>0.625</td>
</tr>
</tbody>
</table>
Chapter 4: 1st generation Haps libraries of greater diversity

Introduction

The initial library of yCD-CH1 fusions created by Dr. Chapman Wright was nearly devoid of circular permutations of the CH1 domain, due to significant bias of S1 nuclease towards inverted repeat sequences [18] in the linker region of the cyclized CH1 domain. As the best switch discovered in our lab to date contains a circularly permuted insert domain [20], we decided to create a new library containing a broad variety of circular permutations of the CH1 domain.

Cyclization and introduction of random breaks using DNasel

Previously our lab has successfully created several switches via random domain and circular permutation using the non-specific nuclease activity of DNase1 and S1 nuclease [18, 19, 106, 107, 118]. These methods require milligram quantities of starting DNA and several days. Additionally these methods generate insertion sites throughout the plasmid not just the gene of interest, and of those within the gene most are out of frame. These methods may have substantial bias towards certain regions of DNA and additionally can cause large deletions and insertions, which require extremely precise reaction conditions to minimize. These drawbacks result in many library members of no value, significantly increasing the selection or screening throughput required to isolate functional switches. However with time, patience, and luck switches can still be created using these random nuclease-based methods.
Creating a random circular permutation library is a significant feat of molecular biology. The gene to be circularly permuted must initially be cyclized by dilute ligation, which is non-trivial. Typically a linker encoding a restriction site is added of sufficient length to span the distance between the original N- and C-termini of the gene to allow sticky-end ligation, which is more efficient than blunt-end ligation (Figure 2.1). Still, one must optimize both ligase and DNA concentrations, to selectively create the monomeric circular form. Temperature and salt concentration can also be increased to slightly decrease the persistence length of DNA thus increasing the probability of cyclization. These adjustments are limited to the extent that T4 ligase will still function. However, increased temperature will also affect the annealing of short cohesive ends.

Focusing on the effect of DNA concentration, Dugaiczyk et al. describe the propensity of a solution of DNA fragments to form monomeric circles over linear concatemers as the ratio of \( j \), the effective concentration of one end of a linear piece of DNA in the “neighborhood or volume” of the opposite end, to \( i \) the effective concentration of total ends in solution [119].

\[
\frac{j}{i} - \frac{N_0}{M l b},
\]

where \( N_0 \) is Avagadro’s number, \( M \) is the molar concentration, \( l \) is the contour length and \( b \) is the random coil segment length of the DNA fragment to be cyclized. According to their results, monomeric circular DNA is only formed in appreciable amounts when \( j/i \geq 2 \). Due to the linear dependence of \( l \) on molecular weight, the
equation for \( j \) can be simplified using the constant values of \( l, b, \) and \( j \) for lambda phage DNA, yielding

\[
- \quad \frac{j}{i}.
\]

In the case of the CH1 domain with a 10mer linker (300 bp, \( MW=1.85\times10^5 \) Da), given the properties of Lambda phage DNA (\( MW_\lambda=3.08\times10^7 \) Da, \( j_\lambda=3.6\times10^{11} \) ends/mL) and a DNA concentration of 10 ng/\( \mu \)L, the ratio \( j/i = 11.87 \), therefore favoring the formation of monomeric circular products. Torsional forces also play a significant role in cyclization [120]. In general an extra half-turn (5.2 bp) added to a DNA segment with an integer multiple of turns, will decrease the yield of monomeric circles \( \sim25\)-fold. Neither a 10mer, nor the original 15mer CH1 domain is ideal for cyclization, however a cyclization reaction of several micrograms of DNA yields enough monomeric circular product to proceed.

Following cyclization and the extremely inefficient isolation of monomeric circular DNA, these circles are randomly digested with DNaseI at very low concentrations such that these circles are cut less than once on average. Again, this singly cut product must be isolated, which is often made difficult by the fact that short fragments of DNA and their corresponding circles do not have significantly different mobility’s through agarose gel. Regardless of these known impediments we pursued this route for quite some time, before devising an alternative.

Random insertion and circular permutation by multiplex PCR

At around the same time this library was completed, we realized a much simpler and more accurate method for generating circular permutation libraries.
This method was inspired by the development of a multiplex inverse PCR [121] method for generating random insertion sites within an acceptor gene. Due to the rapidly declining costs of DNA synthesis, and the difficulties associated with random nuclease-based methods these original methods have since been abandoned for simpler and more focused multiplex PCR methods [122].

Multiplex PCR methods are easily adaptable to nearly any gene sequence with some minimal optimization of PCR conditions. The outline of these methods is very similar. First, abutting primer pairs for each desired insertion site (every codon, all surface residue codons, etc.) are designed. The development of a Matlab® (Mathworks) script to design primers of high quality and equivalent melting temperatures has standardized and facilitated this process. These primers are synthesized and shipped in 96-well plates, which allows specific pairs to be mixed rapidly with a multichannel pipette. Additionally primers can be offset to create deletions or duplications (Figure 4.2). With primer mixes made it is simply a matter of aliquoting a PCR master mix containing the template plasmid to a 96-well PCR plate, adding the primer mixes, and starting the thermocycler. The PCR fragments can then be pooled, purified, and directly ligated, assuming the insert is phosphorylated.

For creating random insertion sites in a gene, the template for multiplex PCR is a plasmid containing the gene. To create circular permutations, a gene-dimer template is required (Figure 4.1). Select circular permutations have been made in the past using PCR on gene-dimer templates [123], so adaptation of this technique to multiplex PCR was a clear extension. The only potential concern was that—
because a pair of abutting primers anneals in each dimer—the outer primers would amplify the backbone of the template plasmid. However, the reaction is easily biased towards the circular permutant by keeping extension times minimal and the backbone of the vector significantly longer than the gene-dimer. Additionally the plasmid backbone could be cut or the gene-dimer could be excised prior to PCR.

Using multiplex PCR to generate circular permutations and random insertion sites, we made two yCD-CH1 fusion libraries. The first contained all possible circular permutations inserted at every codon within yCD and included up to 2 codon duplications and deletions at each site. We dubbed this library RICP, for random insertion and circular permutation. The second, targeted insertion and circular permutation (TICP) contained select permutations and insertions at surface-loop residues. Residues that are solvent accessible, flexible, loosely packed, and between secondary structure elements are most favorable for successful circular permutation [124]. TICP included deletions of up to 3 residues moving across each loop and additional linkers of 1, 2, or 3 random amino acids on the new termini of CH1 circular permutations.

**Methods**

*Cyclization and random digestion with DNaseI*

See [20] for additional discussion of these methods.

We began by cyclizing CH1 domains with 5mer, 10mer, and 15mer (1, 2, and 3x GSGGG) linkers (from here on these are referred to as simply 5mer, 10mer 15mer). The three constructs were created by PCR on pDIM-C8-CH1 template [106]
using appropriate primers to add BamHI sites and linker segments. These PCR products were purified using Zymo DCC columns, digested with BamHI, and purified again using Zymo DCC columns. These fragments were ligated under dilute conditions (2.5 ng/µL) with 400 NEB units of T4 ligase per 100 µL total volume at room temperature overnight (Figure 4.4B). This reaction was then purified by phenol/chloroform extraction and concentrated by ethanol precipitation [125]. To avoid losses in gel purification the concentrated products of ligation were digested directly with dilute DNasel.

To determine the optimal concentration of DNasel, a working stock of DNasel was made to 0.25 U/µL in 1 mM MnCl₂, 0.05 M Tris pH 7.5, 0.05 mg/mL BSA. Dilute ligation products were diluted in the same buffer to 25 ng/µL and 8 µL was aliquoted to several tubes. Several 2-fold dilutions of the DNasel working stock were made and 2 µL of each dilution was added to an individual tube of DNA, mixed, and incubated at room temperature for 8 minutes. The reactions were stopped by adding 2 µL of 100mM EDTA pH 8.5 and heating at 75°C for 10 minutes. The products are then separated and visualized on an agarose gel.

Five or 10 mU/µg was found to be the optimum ratio of DNasel to DNA depending on the yield of monomeric circles in the dilute ligation. This optimum ratio in general is where 30-50% of the desired singly cut product is formed from monomeric circles and results in a sharp band. This optimal reaction was then repeated 6 times, reactions were combined, and electrophoresed on an agarose gel. The band of singly cut DNA was isolated and purified using Gel Extraction columns (Qiagen). Finally this fragment was repaired in a total volume of 100 µL containing
1X T4 ligase buffer, 0.1 mg/mL BSA, 0.2 mM dNTPS, 1 U/µg-DNA T4 DNA polymerase, and 160 cohesive end U/µg-DNA T4DNA-ligase for 20 min and 12°C. The DNA was then purified and concentrated by phenol/chloroform extraction and ethanol precipitation [125].

Acceptor DNA was then prepared using DNasel in a similar fashion to digest plasmid pRCW00012. pRCW00012 was made adding an N-terminal FLAG-tag to human-optimized yCD coding sequence along with appropriate homology to pSKunk to allow CPEC [126]. CPEC was performed, and the products were transformed into DH5α. A single clone was verified by sequencing and several overnight cultures of this clone were prepared. DNA was isolated from these cultures by miniprep (Qiagen). The DNasel concentration for digesting this vector was optimized as above. Fourteen 50-µL reactions containing 0.6mU-DNasel/µg-DNA were performed. The reactions were purified by phenol/chloroform extraction and ethanol precipitation and then repaired in a reaction identical to that above, except enzyme concentrations were decreased 3-fold as linearized product was not isolated in this case. The linearized product was then isolated as above, and dephosphorylated by adding Antarctic phosphatase buffer to 1X and 20 U/µg of Antarctic phosphatase, incubating at 37°C for 1 hour, and then heat inactivating at 65°C for 10 minutes. This solution was used directly in ligation.

The circularly permuted CH1 (cpCH1) insert DNA and randomly cut acceptor DNA were then ligated in a 25 µL reaction containing 250 ng acceptor DNA, 140.2 ng insert DNA (1:7 vector to insert ratio), 1X T4 DNA ligase buffer, 7.5% PEG-8000, and 5 MU (cohesive end) T4 DNA ligase, overnight at room temperature. A similar
ligation was made using pCDNA3.1 that was previously digested with BamHI, repaired, and dephosphorylated as above. This ligation will allow for future propagation of the circular permutation library and cloning using type 2S restriction enzymes. The two ligation reactions were then purified using a Zymo DCC column, eluting twice with 6 µL H2O. Two electroporations of 5 µL of this DNA in 40 µL of DH10B competent cells were performed and plated on LB-agar supplemented with 50 µg/mL streptomycin and 0.2% glucose.

Development of the Random Insertion and Circular Permutation Primer Generator

This Matlab script was based largely on a previous script to generate primers for comprehensive codon mutagenesis using PFunkel [46]. This script was modified to generate forward and reverse primer sequences with 5’ ends at the start/end of each codon. Primer are 15-40 bases in length, have a melting temperature (T_M) above 60°C within 4°C for each pair, and have GC content between 40 and 60%. For yCD and CH1 these parameters yielded specific products in more than 99% of reactions with the conditions described below.

Random Insertion and Circular Permutation library

As a template for multiplex inverse PCR [122] generating random insertion sites in yCD, pRCW00012 (pSKunk-FLAG-yCD) was used. As templates for circular permutation of the CH1 domain, two copies of the CH1 domain were connected by genetically-encoded flexible linkers. As several linker variants may be used as template in each PCR, we rationally designed two linkers to span the distance between the original N- and C-termini. To determine the linker length, the distance
Figure 4.1 - Schematic of the construction of libraries of random circular permutations of CH1 that are randomly inserted into the yCD domain of FLAG-tagged yCD.

Multiplex PCR was used to create a library of all possible circular permutations of the CH1 domain randomly inserted into every position within yCD.

between the alpha carbons of the N- and C-terminal residues were measured using Pymol [117]. Depending on the secondary structure of the linker residues, one amino acid will span a distance of 1.5–3 Å. The distance between the N- and C-termini of the CH1 domain is 22.0 Å, so we chose a 10mer (2X GSGGG) and 15mer (3X GSGGG) linker to span the dimer. These dimers were cloned into pCDNA3.1 (Life Technologies) using CPEC [126] with primers designed in j5 [127]. These plasmids were named pRCW00017 (10mer) and pRCW00018 (15mer). DNA was isolated from overnight cultures of sequence verified clones, and the two plasmids
containing the 10mer and 15mer linker variants were mixed in equimolar quantities to be used as a template for each PCR reaction.

Abutting primers were designed using the RICP primer generator Matlab script and obtained from IDT. For each primer pair, a primer mix containing 10 µM each (forward and reverse) primer was made. Additionally one and two duplication and deletion mixes were made, by shifting the reverse primer ahead and back one and two codons.

Figure 4.2 - Details of primer design and method for creation of deletions and duplications at the site of linearization.
As an example, amplification using a set of four forward and four reverse primers is shown in the region around codons 19–31 of a gene. Different combinations of forward and reverse primers results in codon deletions or tandem duplications as indicated in the table.

Parallel PCR reactions were performed in a volume of 15 µL in 96-well PCR plates. Initially, a template mixture for the total number of reactions plus an additional 10% was made containing 4 µL/reaction water, 4 µL/reaction 5.5 M betaine, 0.6 µL/reaction DMSO, and 0.4 µL/reaction 10 ng/µL template (or mixture of templates). 9 µL was then aliquoted to each well. One µL of the 10 mM primer pair mix was then added to each well.
A thermocycler was then preheated to 98°C, and 10 µL of Phusion HF Master Mix (New England Biolabs, Ipswich, MA) was added to each well immediately before placing the plate in the thermocycler. After an initial 30 seconds at 98°C for the CH1 domain circular permutations or 3 minutes for the inverse PCR opening yCD, the following cycle was repeated 30 times: 30 seconds at 98°C, 10 seconds at 63°C, 15 seconds/kb for the CH1 domains and 30 seconds/kb for yCD inverse PCRs at 72°C. This was followed by incubation at 72°C for 5 min for CH1 domain PCRs and 10 minutes for yCD inverse PCRs. The success of each reaction was verified by electrophoresing 5 µL in an agarose gel. Twelve µL from each reaction (keeping CH1 domain inserts and yCD acceptor plasmids separate) were combined in a polypropylene tube and purified by phenol/chloroform extraction followed by ethanol precipitation [125].

These libraries of DNA were then electrophoresed in separate agarose gels and bands of the desired size were isolated using Purelink gel extraction kit (Life Technologies) according to the manufacturer’s instructions. The library of CH1 domain circular permutations was then phosphorylated using NEB’s Quick Blunting Kit, per the manufacturer’s instructions. It is important that T4 polymerase be included in this reaction to chew back 3’ ends leaving free 5’ ends, on which the kinase acts much more readily than blunt ends. Using T4 polynucleotide kinase alone, even in the presence of crowding factors, does not yield enough 5’ phosphorylated DNA for efficient ligation. The products of this reaction, after heat treatment at 70°C for 10 min, may be directly added to the ligation reaction.
Five hundred nanograms of acceptor yCD plasmid and a threefold molar excess of insert CH1 circular permutations were ligated in a 200 µL containing 7.5% PEG-8000, 1X T4 ligase buffer, and 10,000 cohesive end units of T4 DNA ligase. This reaction was cycled between 10 and 30°C every 30 s for 30 min [128], and then diluted 10-fold with water and purified by phenol/chloroform extraction and ethanol precipitation [125]. Two hundred nanograms of ligation product were electroporated into 40 µL of DH5α competent cells (Life Technologies) five times. The naïve library was amplified, and prepped as above.

**Targetted Insertion and Circular Permutation library**

Circular permutation of CH1 and random insertion sites in yCD were made using the same methods as above simply with a different set of primer mixes. Flexible surface residues were identified using Pymol [117]. Loop and turn residues were visualized as spheres and those making any contact with solvent (visualized as

![Figure 4.3](image)

**Figure 4.3 - Plasmid map of pSkunk-cpCH1 used to add random linkers to new termini of circularly permuted CH1 domains.**

pSkunk backbone was amplified by inverse PCR adding Bsal sites and random linker sequences, and circularly permuted CH1 domains were ligated in this backbone. Bsal sites cut at the red hashed line, leaving a 4 base 5’ overhang (in the random linker region) which can then be filled in with T4 DNA polymerase.
a mesh surface) were chosen as circular permutation or insertion sites. Up to 3 amino acid deletions were made by shifting forward primers ahead, relative to reverse primers. PCR products were purified as above.

To add random interdomain linkers we devised a strategy using the type 2S endonuclease BsaI. By performing inverse PCR on pSkunk with primers containing on their 5\' ends 0, 1, 2, or 3X 5\'-NNK-3\’ linkers (or 5\'-MNN-3\’ for the reverse primer) followed by a BsaI site and ligating the collection of circular permutations of the CH1 domain into this vector, we could then digest with BsaI and fill in the overhangs to recover circular permutations with random linkers. This procedure was performed and the resulting circular permutations of the CH1 domain with random linkers were ligated with the yCD targeted insertion vector PCRs as above and transformed, making a very small library of 2.4×10^4 members. We frequently observed incomplete cutting by BsaI, which contributed to this small library size. As this enzyme is frequently used in Golden-Gate cloning and no other incomplete digestion has been reported we are unsure of the cause of this.

Selections and screens

Selection and screening plates were made as described in Chapter 2, but with different levels of stringency. A variety of selections were performed on the RICP and TICP libraries. We developed a shorthand for these selections: N or P for negative and positive, followed by the concentration of 5FC (µM) or cytosine (µg/mL). Initially the RICP library was subjected to very stringent selections, a negative selection containing 1000 µM 5FC followed by a positive selection containing 5 µg/mL cytosine (or in shorthand N1000-P5). Second, a less stringent
two-tiered selection was performed at 500 µM 5FC and both 25 µg/mL cytosine and 5 µg/mL cytosine (N500-P25 and N500-P5). Third, we performed three-tiered selections on the RICP library. First this library was placed under positive selection at 5 µg/mL cytosine, after which pSkunk-RICP DNA from surviving members was purified by gel extraction and retransformed into GIA39. This remaining library was subjected to negative selection at 500 µM 5FC and DNA from the surviving members was used to transform GIA39-HIF cells. The remaining library members were once again plated on positive selection plates containing 5 µg/mL cytosine (P5-N500-P5). Finally, a less stringent three-tiered selection was performed of the following sequence: positive at 50 µg/mL cytosine, negative at 350 µg/mL, and again P50 (P50-N350-P50). The TICP library was subjected to a similar selection without the final positive selection as less than 100 members survived the N350 selection.

Once a series of selections had winnowed a library to under 1000 members these survivors were screened by PCR. Colonies were picked with a pipette tip and a short (5mm) heavy streak was made on an LB Agar plate containing appropriate antibiotics (in some cases, e.g. when several colonies were crowded together, a similar streak was made on a positive selection plate first, in some other cases colonies were inoculated into liquid unselective minimal medium and liquid LB medium for screening with and without arabinose and DNA preparation, respectively), then the tip was dipped and swizzled in 7.5 µL of GoTaq Green (Promega) PCR mix containing 1 µM each forward and reverse primers. The forward primer tac-seq-f and pSKunk-seq-r were routinely used. PCRs were initially heated to 94°C for 3 minutes, and then the following cycle was repeated 30 times: 94°C for
30 seconds, 58°C for 20 seconds, 72°C for 30 seconds. This was followed with a 5 minute incubation at 72°C, and products were cooled to 4°C indefinitely. With the PCRs in the thermocycler, secondary and tertiary streaks of each colony were extended from the initial heavy streaks and plates were incubated overnight at 37°C.

PCR products were directly loaded and separated in an agarose gel. DNA was isolated from those colonies with products around 1000 bp (the approximate size of a yCD-CH1 fusion), via miniprep (Qiagen). This DNA was sequenced (Genewiz) using the pSKunk-seq-r primer, and used to dilutely (~1 ng/100 µL cells) transform chemically competent [114] GIA39 cells harboring pGA and pGA-HIF separately. Colonies resulting from these transformations were incubated overnight in unselective minimal media and spotted on screening plates as described in Chapter 2.

Results

Due to high non-specific binding to mammalian proteins of anti-yCD antibodies, we appended a FLAG-tag to the N-terminus of Haps59 to facilitate characterization Haps59 in mammalian cell lines. FLAG-Haps59 and Haps59 exhibited similar HIF-1a-dependent accumulation levels in RKO cells [106], although the selective toxicity conferred by FLAG-Haps59 has not been tested in these cells. We did discover that the addition of the FLAG-tag to Haps59 abolishes switching activity in *E. coli* by increasing the 5FC toxicity of the off-state in the absence of coexpressed HIF-1a, up to the same level as in the presence of HIF-1a (Figure 4.9). Since the CH1 domain of Haps59 is located very near the N-terminus of
the protein, we speculate that the FLAG-tag stabilizes the CH1-domain resulting in increased accumulation of the switch (Figure 3.4B). In an effort to avoid future issues with characterization and increase the potential for being able to purify the resulting switches we decided to examine if a switch could be constructed starting with FLAG-yCD. We constructed a libraries similar in design to that in our previous work [106] in which random circular permutations of the CH1 gene were randomly inserted into the FLAG-yCD gene. Initially D Nasel was used to create random circular permutations and insertions sites. However, these libraries were abandoned due to poor quality and the development PCR methodologies [122], which allowed us to quickly generate much more focused and complete libraries.

DNasel library

Monomeric circular CH1 domains containing 5mer, 10mer, and 15mer linkers were made successfully and digested with DNAsel (Figure 4.4). Plasmid containing FLAG-tagged yCD (pRCW00012) was also digested with DNAsel, and the two were ligated together. Additionally, the circularly permuted CH1-domains (cpCH1) were ligated into pCDNA3.1 to allow this library of permutants to be subcloned for future use. Transformation resulted in a library of $6.3 \times 10^6$ members for the pRCW00012-cpCh1 ligation and $1.4 \times 10^7$ for the pCDNA-cpCH1 ligation. Transformants from the pCDNA-cpCH1 library were subjected to a colony PCR screen with primers annealing up and downstream of the insertion site to determine the quality of the circular permutants generated. Only 2 of 14 clones screened contained near-full-length CH1 domains. 10 contained deletions or
duplications of ~50 bases or more. While not entirely a loss, we were hesitant to move forward with this library, due to the large deletions in most members.

Figure 4.4 – Representative images of cyclization and random digestion of the CH1 domain. 
(A) 10mer CH1 domain after PCR and digestion with BamHI and a dilute (5ng/µL) ligation of this fragment. After 4 hours at room temperature the dilute ligation appears to have gone to completion, eliminating linear monomeric form. This reaction still yields mostly higher molecular weight bands but some monomeric circular product as indicated by the arrow. (B) Dilute ligations of 5mer, 10mer and 15mer CH1 domains. Monomeric circles are the lowest band. (C) DNasel digestion of 15mer CH1 domain dilute ligation reaction. Decreasing amounts of DNasel were used to digest equal amounts of dilute ligation products to give monomeric circularly permuted CH1 domains, indicated by the arrow.

Figure 4.5 – PCR screen of pCDNA-cpCH1 library. PCR was performed directly on colonies from transformation plates using primers that anneal ~50 bp 5’ and 3’ of the CH1 insertion site. Each reaction was run on an agarose gel and imaged. A full length cpCH1 domain with no deletions or duplications would be ~400 bp.

Random insertion and circular permutation library

Using multiplex PCR to generate random insertion sites and random circular permutations, we created an RICP library containing all 1.4 million theoretical
members and performed positive and negative selections at various stringencies (for a list of details of these selections see Table 4.1 at the end of this chapter). An initial very stringent negative selection containing 1000 μM 5FC (N1000) was performed followed by a similarly stringent positive selection with 5μg/mL cytosine (P5). Both tiers of this selection covered the entire library diversity. Eighty-four percent of this library survived the N1000 tier of this selection. Thirty-four colonies formed on the P5 selection plates. None of these colonies demonstrated arabinose-dependent 5FC sensitivity.

Due to these negative results, the extremely low survival rate in the P5 selection (less than 1 in 1-million), and apparent deletion of yCD in PCR screens, we hypothesized that the negative selection was too stringent and potentially induced stress resulting in the deletion of yCD or any yCD-CH1 fusions. The 34 survivors of the P5 selection then would have to result from reversion of the genomic bacterial cytosine deaminase or upregulation of some cryptic salvage pathway (most likely some other salvage pathway as none of the selected variants showed any sensitivity to 5FC in screening). This led us to attempt to find switches in less stringent selections.

Going back to the naïve library we performed an N500-P25 selection. This negative selection allowed 90% of this library to survive. Fifty-one colonies survived this positive selection. However, once again none of these colonies was sensitive to 5FC when screened. We struck out these selected variants on P25 selection media again to confirm their phenotype and the majority would not grow as single colonies. These results led us to seriously evaluate the stringency of both selections.
As previous negative selections had eliminated only 10-20% of the naïve RICP library we decided to instead perform an initial positive selection to eliminate a larger percentage of undesirable members up front, allowing us to carry fewer members through the later selections making them more robust. We transformed GIA39-HIF cells with the naïve library and tested various stringencies of positive selections. The percent survival across selections from 3 µg/mL cytosine to 50 µg/mL cytosine was unexpectedly very consistent. This result led us to examine the identity of several clones from each selection via PCR screen. Using primers that anneal outside of the yCD gene, we were able to determine the number of clones with and without a CH1 insert (Figure 4.6).

Assuming that members without CH1 inserts are more likely to have cytosine deaminase activity similar to WT yCD, we used the percentage of clones having WT yCD-sized bands as an approximation of the true positive rate. Counter-intuitively, the less stringent P25 and P50 selections resulted in a higher proportion of yCD-sized clones. We did observe that colonies grew on these plates within 18-24 hours whereas at P15 and lower, colonies required 30-36 hours to form. It is possible that at these lower levels, survivors are able to scavenge uracil from their dying neighbors or are able to garner uracil through some other mechanism, forming false positive colonies.

We proceeded, performing a P50 selection. Colonies formed on this selection plate were collected and DNA was isolated by miniprep from a small aliquot. This DNA was separated on an agarose gel and the band corresponding to supercoiled pSKunk-RICP was isolated purified and electroporated into GIA39 cells.
Figure 4.6 – PCR screen of the insert region of pSkunk of surviving members of positive selections on the RICP naïve library in GIA39-HIF.
6.5×10⁴ viable cells were plated on 10 cm petri dishes containing increasing stringencies of positive selection media. Total surviving cells were counted after 36 hours of growth at 37°C and % survival was calculated. 24 colonies (or 28 for P25 and P50) were picked from a small area of each plate and dipped into PCR mix containing primers annealing outside of the coding region of yCD. PCR products were then separated on agarose gels. yCD length bands (~700 bp) are indicated by arrow heads.
We similarly examined several negative selection conditions from 100 to 500 µM 5FC. At 100 µM (N100) there was not a significant difference from an unselective plate. At N200 there was noticeable heterogeneity in colony size, but not a significant difference in colony number. Significant cell death occurred at N300, allowing only 1.3% of cells to survive. At N400 toxicity was even more significant with 0.26% survival. Oddly, 3.7% of cells survived the N500 selection. It is possible that this result is within the range of experimental error. However, as we have observed odd growth patterns in the presence of 5FC in the past (Figure 2.4), we were concerned that at this selection level, the stress of higher selective pressure and the increased potential for mutations caused by 5FU could induce resistance genes and/or otherwise cause false negatives (or false positives in this negative selection).

![Figure 4.7 - PCR screen of the insert region of P50 selected RICP library members in GIA39 surviving negative selections.](image)

Significant drop in survival occurred at N300. Colonies were picked from a small area of each plate and dipped into PCR mix containing primers annealing outside of the coding region of yCD. PCR products were then separated on agarose gels. yCD length bands (~700 bp) are indicated grey lines.

We performed PCR screens on surviving members of these selection tests as well (Figure 4.7). At N400 and above survivors with significant deletions in yCD
were observed, which may be indicative of strong selection against yCD activity.

However with a sample size of 9, it is difficult to make any strong conclusions. Bands at approximately the same size as yCD were observed at all selection levels, although this does not necessitate significant cytosine deaminase activity as inactivating mutations, deletions, or duplications could exist in these variants. We decided to proceed with an N350 selection on the remaining members of the RICP library after P50 selection. This level still exerts strong selection pressure but would hopefully minimize the occurrence of false negatives seen in N500.

Survival in this selection was higher than in small-scale tests (6.8%), leaving ~1,500 unique members remaining. DNA was prepped from these members and used to transform GIA39-HIF cells once again. 15,000 viable cells from this RICP-P50-N350 library were plated on a P50 plate, and after 24 hours of growth there were 623 colonies on this plate. These colonies were struck out on LB Agar plates and subjected to PCR screen. Only 23% of these colonies contained yCD-CH1 fusions, the rest contained bands the size of yCD. The members containing yCD-CH1

Figure 4.8 – Dot toxicity screen of members RICP-P50-N350-P50. Two dilutions of log phase cultures of resulting members with GST or GST-HIF-1α-CTAD under coexpression were spotted on plates containing increasing concentrations of 5FC.
fusions were screened for arabinose dependent 5FC toxicity as outlined in Chapter 2. Very few members exhibited arabinose-dependent toxicity, and those that did were marginally different in the presence and absence of arabinose. Regardless, plasmid DNA was isolated from these members sequenced and retransformed into both pGA and pGA-HIF containing GIA39 cells. In sequencing these members we found several members containing CH1 domains in the backwards orientation. This was odd but plausible as there are no stop codons in the reverse sequence, and none of these fusions appear to be HIF-1α-responsive anyway. The transformants were then screened for HIF-1α-CTAD-dependent 5FC toxicity (Figure 4.8).

The results of this screen were not promising. A few members, as well as pRW019 (pSkunk-FLAG-Haps59), appeared to be less active in the presence of HIF-

![Figure 4.9 - Characterization of RICP12 and RICP95 in E. coli.](image)

(A) Dot toxicity assay of RICP12 and RICP95. (B) Diagram of RICP12 and RICP95 sequence. Numbers above indicate the amino acid residues of the corresponding wild-type proteins. Letters below indicate the 1-letter abbreviation amino acid linker sequence.
1α-CTAD than in its absence. Only RICP12 and RICP95 showed very slight switching, in the desired direction, and were repeated under more 5FC conditions (Figure 4.9A). Neither of these fusions conferred a significant switching phenotype to *E. coli*.

Both RICP12 and RICP95 had insertions very near the N-terminus of yCD (Figure 4.9B), generally consistent with Haps3 and Haps59. The circular permutation at the 87th residue of CH1 in RICP95 is towards the end of the last α-helix, which does not contact HIF-1α but does coordinate a zinc molecule via cysteines 81 and 84. The 34-31 circular permutation in RICP12 is in the long flexible loop between the first and second helices also relatively removed from HIF-1α-contact, but Cys34 also coordinates zinc.

![Circular permutation sites in RICP12 and RICP95.](image)

*Figure 4.10 – Circular permutation sites in RICP12 and RICP95.*
The CH1 domain (green) and HIF-1α-CTAD (purple) are shown in cartoon representation. Grey spheres represent zinc atoms. New N- and C-termini in circular permutations are shown as sticks. New termini (Cys34 and Val31) of RICP12 are on the left side of the molecule. Leu87 and Pro86 are in the front right.

Following this extensive yet unsuccessful search of the RICP library, we used a similar selection scheme to search the TICP library. This library was initially very
small (8.4×10^4 members), and was therefore narrowed down to a screen-able number of colonies after only P50 and N350 selections. Of the naïve library, 2.3% survived the initial positive selection, and 13.8% of the remaining members survived the N350 selection. These colonies were screened by PCR initially as above and isolated DNA from those containing inserts the approximate size of yCD-CH1 fusions, was sequenced and used to transform GIA39-pGA and GIA39-HIF cells.

Only 3 of the 16 members sequenced contained CH1 domains in the correct orientation. Five of 16 members were Haps59 in its original form without the N-terminal FLAG-tag. This was not included in the library and must be attributed to contamination. Despite concerns over this contamination, it does act as a quasi-control demonstrating that the selection is effective in enriching for switches. Once again none of the intended members of this library were responsive to HIF-1α or sensitive to 5FC.

![Figure 4.11 – Screening of TICP clones in GIA39 with pGA or pGA-HIF.](image)

Dilutions of log phase cultures harboring plasmids expressing TICP clones and GST (−HIF) or GST-HIF-1α-CTAD (+HIF) were spotted on plates of increasing concentrations of 5FC.
Conclusions

Based on our extensive examination of the RICP library and the lack of switches isolated we conclude that an N-terminal FLAG-tag in is a barrier to yCD-CH1 switch creation. It is possible however that the addition of linkers either after the FLAG-tag or between the CH1 domain and the subsequent fragments of yCD (as in the TICP library) could increase the chance of finding switches. Additionally these fusions may confer a switching phenotype upon removal of the FLAG-tag. It is also possible based on the selective accumulation of FLAG-Haps59 in the presence of HIF-1α in mammalian cell lines that RICP12 and RICP95 may exhibit a switching phenotype in these cell lines. The creation of a larger TICP library and further examination of this sequence space may be worthwhile. A library excluding the N-terminal FLAG-tag may be more fruitful, however.
Table 4.1 – RICP and TICP library and selection statistics.

<table>
<thead>
<tr>
<th>Library</th>
<th>Max Diversity</th>
<th>Transformants</th>
<th>% meaningful</th>
<th>Degeneracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RICP</td>
<td>1.42E+06</td>
<td>1.22E+08</td>
<td>82.6%</td>
<td>1.42E+06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selection</th>
<th>Selection size</th>
<th>% survival</th>
<th>Remaining members</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1000</td>
<td>1.00E+07</td>
<td>84.0%</td>
<td>1.19E+06</td>
</tr>
<tr>
<td>P5</td>
<td>6.00E+06</td>
<td>0.0%</td>
<td>34 colonies</td>
</tr>
<tr>
<td>N500</td>
<td>3.50E+06</td>
<td>90.0%</td>
<td>1.28E+06</td>
</tr>
<tr>
<td>P25</td>
<td>3.50E+06</td>
<td>0.0%</td>
<td>50 colonies</td>
</tr>
<tr>
<td>P5</td>
<td>2.50E+06</td>
<td>5.0%</td>
<td>7.10E+04</td>
</tr>
<tr>
<td>N500</td>
<td>5.00E+05</td>
<td>32.7%</td>
<td>2.32E+04</td>
</tr>
<tr>
<td>P5</td>
<td>4.00E+05</td>
<td>0.0%</td>
<td>24 colonies</td>
</tr>
<tr>
<td>P5</td>
<td>2.50E+06</td>
<td>5.0%</td>
<td>7.10E+04</td>
</tr>
<tr>
<td>N500</td>
<td>5.00E+05</td>
<td>32.7%</td>
<td>2.32E+04</td>
</tr>
<tr>
<td>N800</td>
<td>1.85E+05</td>
<td>77.0%</td>
<td>1.79E+04</td>
</tr>
<tr>
<td>P3</td>
<td>2.00E+05</td>
<td>0.1%</td>
<td>112 colonies</td>
</tr>
<tr>
<td>P50</td>
<td>2.00E+06</td>
<td>1.0%</td>
<td>2.00E+04</td>
</tr>
<tr>
<td>N350</td>
<td>2.50E+05</td>
<td>6.8%</td>
<td>9.66E+02</td>
</tr>
<tr>
<td>P50</td>
<td>15000</td>
<td>4.15%</td>
<td>6.23E+02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Library</th>
<th>Max Diversity</th>
<th>Transformants</th>
<th>% meaningful</th>
<th>Degeneracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>TICP</td>
<td>&gt;1E+12</td>
<td>8.40E+04</td>
<td>14.0%</td>
<td>8.40E+04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selection</th>
<th>Selection size</th>
<th>% survival</th>
<th>Remaining members</th>
</tr>
</thead>
<tbody>
<tr>
<td>P50</td>
<td>5.00E+05</td>
<td>2.30%</td>
<td>1.93E+01</td>
</tr>
<tr>
<td>N350</td>
<td>1.93E+04</td>
<td>13.8%</td>
<td>2.67E+02</td>
</tr>
</tbody>
</table>
Chapter 5: Switch characterization in human cell lines

As a lab that has worked almost exclusively with *E. coli*, we learned how painstakingly slow work in eukaryotes is from the work characterizing Haps59. We also learned that the activity of switches can differ significantly between *E. coli* and mammalian cells and also between different mammalian cell lines. Previously we characterized Haps59 in RKO and MCF7 cancer cell lines by random stable integration of an encoding vector. This approach requires significant time and resources, and it introduces additional variation in expression level through differences in both the number and location of integrations between clones. To avoid future difficulties and ultimately provide a better picture of how these switches may perform *in vivo*, we wished to develop a rapid, robust, and reliable method for screening switches in mammalian cells.

There are several possible methods for achieving this goal. Viral transduction is commonly known to be a rapid and reliable method for introducing exogenous genes into mammalian cell lines. We initially attempted to use this method to integrate Haps59 and Haps3 into cancer cells lines. However, we soon found that viral transduction can affect HIF-1α signaling [129, 130] and may introduce additional potential binding partners for the CH1 domain [131]. For these reasons we decided to avoid the use viral transduction as a screening tool.

High throughput transient transfection methods are another possibility. High throughput transfections require significant optimization of culture conditions and transfection reagent and conditions prior to assay optimization, but are frequently...
used to study protein function and even gene circuit function in mammalian cells [132]. In addition there are many commercially available products for such methods. Regardless of the reagent and conditions used there is the potential for high variability between samples since each transfection reaction must be made independently. It may be possible to bicistronically express a reporter protein by which this variability could be measured. Even with this matter of contention potentially controlled, limiting transfection efficiency, toxicity of transfection reagents, and the limited time window of expression (not to mention the dynamics within this window) make the development of a transient transfection screening method a daunting task.

Although the creation of stable lines is less than rapid, it can provide significant advantages in robustness and reliability, and in recent years tools have been developed to improve the speed and efficiency of this process. Namely, we were interested in evaluating the Flp-In™ system from Life Technologies. This technology originally developed by Schlake and Bode [133], uses Flp recombinase from *S. cerevisiae* to mediate recombination of a target gene into the genome.

In this system, a cell line with a Flp recombinase target (FRT) site-containing vector stably integrated at a single site in its genome is cotransfected with a Flp recombinase expression plasmid and a plasmid containing the gene to be integrated as well as a second FRT site. The Flp recombinase mediates recombination between the FRT sites of the vector containing the gene to be integrated and the site stably integrated in the genome. The selectable resistance marker on the vector is initially lacking a start codon, acting as a safeguard against non-specific integration. Only
after recombination is the start codon appended and the resistance marker expressed. Additionally this recombination event inactivates the zeocin resistance-LacZ fusion downstream of the FRT site in the genome. Thus one can screen for lack of β-galactosidase activity, to confirm specific integration.

In attempt to avoid the nonconformities introduced by random integration as well significantly increase screening throughput, we explored the use of Flp recombinase-mediated isogenic integration [133] by integrating our switches into commercially available Flp-In 293 cells. This method of integration should prove to be specific, efficient, amenable to high throughput, and directly comparable (without the need to control for multiple integrations). The only drawback to this system is the necessity of a cell line with the FRT-site-containing vector previously integrated at a single stable genomic locus.

Methods

High throughput transfection

To account for variable transfection efficiency across samples we created a semi-bicistronic expression plasmid using the self-cleaving 2A peptide from *Thosea asigna* virus (T2A) [134]. This peptide allows stoichiometric expression of two proteins from a single transcript (not truly bicistronic), leaving only a short peptide sequence on the C-terminus of the upstream protein and a single proline residue on the N-terminus of the downstream protein. Cleavage of this sequence is not perfectly efficient and varies from system to system. Since we have designed this plasmid, the 2A sequence from porcine teschovirus-1 has been deemed more
universally efficient [135]. We placed the reporter protein EGFP in the 5’ position, as the remnants of the T2A peptide are known to not interfere with its function. Additionally this only adds a proline residue to the N-terminus of the switch variant we wish to express, hopefully with minimal effect. We chose to place this coding sequence under transcriptional control of the hybrid cytolomegavirus enhancer/chicken β-actin promoter and intron/rabbit beta-globin intron and exon (CAG promoter) which shows much more consistent expression across cell types than the CMV promoter [136] and will hopefully prove more resistant to down-regulation [137, 138].

We found PCR across the CAG promoter quite difficult, likely due to secondary structure, and subsequently created our desired plasmid by digesting pCAGGS-EGFP (a kind gift from Dr. Nicholas Boylan) with BsrGI and blunt-cloning the T2A-yCD/Haps59/switch variant. The linear vector was blunted using NEB’s Quick Blunting Kit and treated with alkaline phosphatase. T2A-fusions were made by PCR on pSKunk vectors containing switches or controls with phosphorylated primers. The forward primer encoded the T2A peptide. We also created a version of these vectors containing a neomycin resistance gene and f1 origin from pCDNA3.1 (Life Technologies) to allow G418 selection for random integration into mammalian cells and Kunkel/PFunkel mutagenesis. Plasmids were propagated in DH5α E. coli and DNA was isolated from these cells for transfection with the endotoxin-free PureYield™ Plasmid Miniprep kit.
Figure 5.1 – Plasmid maps of pCAGGS-EGFP and pCAGGSneo-EGFP.
HeLa cells were obtained from ATCC and cultured EMEM supplemented with 10% FBS. We found X-tremeGene (Roche) to give relatively high efficiency with minimal toxicity at a ratio of 2 µL X-tremeGene per 1 µg DNA. 1500 cells were seeded per well of a 96-well plate, and the next morning they were transfected with X-tremeGene per the manufacturers protocol. Twenty-four hours after transfection media was changed to media containing 5FC and with or without 100 µM cobalt chloride. Plates were incubated for an additional 48 hours, prior to the addition of Alamar blue (1/10th dilution). Alamar blue was incubated with the cells for 90 min prior to reading GFP fluorescence emission at 509 nm after excitation at 488 nm with a cutoff filter at 595 nm, followed by an Alamar blue reading (585 nm emission, 565 excitation, 570 cutoff).

Isogenic integration in Flp-In 293 cells

FlpIn-293 cells (FI-293), pCDNA5/FRT/TO, and pOG44 were obtained from Life Technologies. FI-293 cells were maintained in DMEM high-glucose w/ L-glutamine supplemented with an additional 2mM L-glutamine (Gibco) and 100 µg/mL zeocin (Life Technologies). yCD, Haps59, and the switch variants created in this work were cloned into pCDNA5/FRT/TO, and DNA for transfections was prepped from overnight cultures of E. coli using PureYield Miniprep kit (Promega). A total of 2.7 µg of pOG44 and 300 ng of pCDNA5/FRT/TO-EV, -yCD, and -Haps59 were transfected into 100,000 FI-293 cells, plated eighteen hours prior in a single well of a 6-well plate using CalPhos Mammalian Transfection Kit (Clontech) following the manufacturers directions. Twelve hours after transfection the cells were washed with PBS and fresh media was added. Thirty-six hours after
transfection the cells were split 1 to 5 into a fresh 6-well plate of media containing 150 μg/mL hygromycin-B, instead of zeocin. Media was changed every 3-4 days until foci formed, in about 2 weeks. These stable pools were passaged 3 times, at a 1/10 dilution into 200 μg/mL hygromycin-B, before being used in toxicity assays.

5FC toxicity assay

The inner 60 wells of a 96 well plate were seeded with 1000 cells per well in 100 μL of media containing appropriate selection markers. The next day 50 μL of media was added to achieve the desired final concentrations of 5FC and 50 μM CoCl₂ (in +HIF samples). Cells were incubated for 4 days prior to assaying their viability/survival.

Cell survival/viability assay

Total DNA was measured using a SYBR green assay. Once untreated control wells were near confluence, cells were washed twice with PBS, and after removing the second wash lysing the cells by freezing for 1 hour to overnight at -80°C. Plates were then thawed and 200 μL of 0.075% SYBR green (Life Technologies) in sterile filtered deionized water was added to each well, pipetting up and down 7 times. Plates were incubated at 37°C for 4-6 hours to complete lysis and establish equilibrium binding of SYBR green and DNA. The fluorescent emission was measured at 535 nm after excitation at 485 nm with a 515 nm cutoff filter. Percent survival was calculated by normalizing the total DNA of surviving cells in each well between wells containing no 5FC (i.e. 100% survival) and wells containing no cells (i.e. 0% survival).
Lysis for western blots and immunoprecipitations

Cells were grown to 75% confluence in 10 cm dishes and media was changed to contain increasing concentrations of CoCl$_2$ (or dishes were placed in a hypoxia chamber) and incubated for 4 hours. After treatment the plates were incubated at 4°C for 30 minutes. From this point on all reagents were ice-cold. Cells were washed with PBS twice. After removing as much of the second wash as possible, 500 µL of RIPA buffer plus 1 µL protease inhibitors was added per 10 cm dish and cells were gently removed from the dish with a rubber cell scraper (also ice-cold). The solution of cells was transferred to a pre-chilled microcentrifuge tube and incubated at 4°C for 1 hour with vigorous shaking. The lysate was then centrifuged at 21k x g for 20 minutes and the supernatant was moved to a new pre-chilled tube. Total protein concentration of each lysate was measured using a DC protein assay (Bio-Rad) standardized to BSA.

HIF-1α western blots

Equal amounts of total protein were loaded on a Novex Nupage 4-12% Bis-Tris Gel (Life Technologies) and MOPS running buffer was used to separate high molecular weight proteins. The gel was electrophoresed per the manufacturers protocol. Proteins were transferred to PVDF membrane (Bio-Rad) using a TransBlot SD blotter (Bio-Rad) and a discontinuous buffering system (Anode buffer: 2X Transfer buffer (Life Technologies), 10% methanol, 1% Nupage Antioxidant (Life Technologies), Cathode buffer: 2X Transfer buffer, 10% methanol, 1% Nupage Antioxidant, 0.04% SDS) for 30 minutes at 15 volts. Membranes were blocked with 5% milk (Bio-Rad) in tris-buffered saline plus 0.5% tween-20 (TBST). Mouse
monoclonal anti-HIF-1α antibodies (BD) were diluted 1/1000 into TBST +5% milk and the blot was incubated overnight in this solution at 4°C with gentle shaking. In the morning, the blot was washed briefly in TBST and prepared for the Snap ID protein detection system (Millipore) as recommended by the manufacturer. The blot was washed 3 times with 75 mL of TBST in the Snap ID and then incubated for 30 minutes at room temperature with a 1/6000 dilution of goat anti-mouse secondary antibodies (Bio-Rad) in TBST plus 0.5% (w/v) milk. The blot was washed again as before, prior to development with Clarity Western ECL substrate in a Universal Hood II (Bio-Rad).

Creation of Flp-In-RKO cells

RKO cells were obtained from ATCC and the plasmid pFRT/LacZeo2 was obtained from Life Technologies. RKO cells were maintained initially in EMEM (Quality biologicals) supplemented with 10% FBS (Hyclone).

To create this line we initially linearized pFRT with ScaI as recommended by the manufacturer to prevent integration from disrupting other necessary elements. Two-hundred-and-fifty ng of the linearized plasmid and 2.75 µg of salmon sperm DNA was transfected using the CalPhos Kit (Clontech) into 400,000 cells plated 24 hours prior to transfection in a 6-well plate. A negative control transfection was done using 3 µg salmon sperm DNA. One hour prior to transfection media was changed to DMEM with 10% FBS. Eighteen hours after transfection cells were washed with PBS and media was changed back to EMEM plus 10% FBS. Forty hours after transfection cells were split 1 to 10 into 10 cm dishes in EMEM with 10% FBS and 250 µg/mL zeocin (Life Technologies). Media was changed every 3 days until no
cells remained on the negative control plate. If there were more dead cells than live cells on the plate at any point, or significant debris in the media the plates were washed with PBS. Once there were no surviving cells on the negative control plate, the zeocin concentration in the media was reduced to 200 µg/mL.

Either only a single colony of resistant cells formed and spread rapidly or several formed in close proximity preventing us from isolating clonal populations easily. Therefore, we cloned this line by limiting dilution. Cells were trypsinized and diluted to 1 cell per 200 µL and seeded to 96-well plates with 100 µL per well. One hundred µL of conditioned media was added the next day containing 200 µg/mL zeocin, for a final concentration of 100 µg/mL. One hundred µL of media was replenished every week. After 2 weeks the wells containing cells were marked and examined for single colonies of cells. Wells containing single colonies were further propagated, eventually passaging to 24 well plates. Once cultures in 24 well plates were confluent they were split to 2 wells of a 96 well plate as well as a 6 well plate for further propagation. The 96-well plate was used for an Alamar blue assay to measure cell number, and a β-galactosidase assay (Life Technologies) to measure the level of expression from pFRT. The ratio of β-galactosidase activity to cell number was nearly equivalent across all 15 clones tested. The 3 clones with the most healthy morphology and growth rate were chosen for further analysis.

These clones were split to T75 flasks as well as T25 flasks. Upon reaching confluence the T75 flask of cells was trypsinized and half of the cells were frozen. DNA was prepped from the other half using GeneElute Mammalian DNA Mini Kit (Sigma). This DNA was used as template for real-time quantitative PCR.
determination of pFRT copy number [139]. Primers detecting pFRT were designed using Primer3 [Untergasser 2012]. Control primers amplifying GPR15 and ZNF80 were designed by Hoebeeck et al. [139]. Ten ng of DNA isolated from each clone, 4 subsequent 2-fold dilutions, and a no template control were added to PCR reactions containing 1X iQ SYBR green supermix containing 250 nM each forward and reverse primer. PCR reactions were run in a Bio-Rad iQ cycler for 10 minutes at 95°C followed by 40 cycles of 15 seconds at 95°C and 60 seconds at 60°C. A melting curve analysis and agarose gel electrophoresis were used to confirm specificity of the PCRs. The copy number of pFRT relative to known single copy controls (GPR15 and ZFN80) was calculated using the comparative Ct method [139]. No clones appeared to have more than a single integration of pFRT. Three clones were chosen based for having the most similar morphologies to wild-type RKO cells and for having high, medium, and low levels of β-galactosidase activity in initial assays. These clones (FI-RKO -2, -10, and -13) were passaged regularly for 1 month and the β-galactosidase assay was repeated on this later passage as well as the original frozen stock to ensure pFRT is integrated in a stable location in the genomes of these clones.

Towards the end of this month of propagation each of the 3 chosen clones began to lose adherence and increase typical doubling time. We tested richer a richer medium formulation with these cells (DMEM, high glucose, sodium pyruvate, 4mM L-glutamine, 10% FBS), which returned them to their normal phenotype and morphology within 1 passage. We are not sure if this was caused by a faulty batch of EMEM or FBS or the added stress of LacZ-zeocin fusion expression, but from this
point forward we cultured these cells in this DMEM formulation. Since this change these cells have maintained wild-type RKO morphology and growth rate.

**Results**

**High throughput transfections**

Initially we attempted to optimize a method of high throughput transfection screening in HeLa cells, as they are prolific, well-characterized, tolerant of transfection, 5FU sensitive [141], and known to only accumulate HIF-1α in hypoxic or hypoxia mimetic conditions [142]. Unfortunately after significant optimization, we were unable to recreate our previous results in stable cell lines [106]. Oddly, 5FC appeared less toxic in hypoxia than normoxia in this assay (Figure 5.2). This may be a result of the significant inhibition of growth or Alamar blue fluorescence due to CoCl₂ after the stress of transfection. Cells need to be actively dividing and replicating DNA for 5FU to exert full toxicity, thus if CoCl₂ inhibits cell division 5FC will appear less toxic. It is also possible that the significant stress of CoCl₂ and transfection causes the cells to lose the plasmid more rapidly, which only adds to the confounding selection against transfected cells of 5FC. By the end point of these assays, GFP expression was below detectable levels, preventing us from measuring loss of plasmid.

We found total DNA quantification to be a more reliable method of survival/viability measurement between hypoxic and normoxic conditions than Alamar Blue (Life Technologies). Alamar Blue is nontoxic allowing viability to be quantified at several time points and also is compatible with simultaneous
Figure 5.2 - Transient transfection 5FC toxicity assay in HeLa cells.

pCAGGS-EGFP-T2A vectors containing no insert (EV), yCD, or Haps59 were used to transfect HeLa cells. % Survival (A) is calculated from relative fluorescence values (B). Error bars represent SEM of 5 wells.

Figure 5.3 - Transient transfection 5FC toxicity assay in 293 cells.

pCAGGS-EGFP-T2A vectors containing no insert (EV), yCD, or Haps59 were used to transfect 293 cells. % Survival (A) is calculated from relative fluorescence values (B). Error bars represent SEM of 3 wells.
fluorescent quantification of GFP. However, hypoxia and CoCl$_2$ were found to
decrease the signal from this reagent 10-20%. As the measured spectral change of
Alamar Blue is caused by cellular metabolism the hypoxia associated change to
glycolysis is likely responsible for this decrease, but media pH may also contribute.
This effect may fall out after normalization, but we found total DNA measurement to
be more consistent between samples.

We also attempted high throughput transfection assays in RKO cells and 293
cells (Figure 5.3) with similarly dismal results. An assay of this type introduces far
too many confounding factors to even be of use in narrowing down a panel of
switches. In the end it would better to make several pools of stable clones
containing pCAGGS constructs that can be mixed to equivalent expression levels
based on fluorescence. This would remove several confounding factors, place
significantly less stress on the cells during the assay, and allow a longer time course,
resulting in a broader window across a narrower range of 5FC. Fortunately,
however the Flp-In system solves all of these problems.

**Screening in Flp-In 293 cells**

We integrated Haps59, yCD and empty vector (EV) controls as well as new
switches resulting from this work into the Flp-In$^{TM}$-293 cell line (Flp-In 293 cells).
Three passages (1:10 dilution) after foci formed on the selection plates, stable pools
showed no detectable β-galactosidase activity and were zeocin sensitive, allowing us
to assume they are isogenic. The control isogenic stable pools were then assayed for
HIF-1α-dependent 5FC toxicity. HIF-1α was induced with 50 μM CoCl$_2$, which was
shown to induce HIF-1α while minimizing toxicity (Figure 5.4A).
The negative control EV cell line survived in both the presence and absence of CoCl₂ up to the highest levels of 5FC tested (512 μM). There was a less than 5% decrease in survival rate due to CoCl₂ alone. In contrast, cell lines expressing yCD had an EC₅₀ of ~25 μM regardless of the addition of CoCl₂. These results were expected and are similar to our results with RKO and MCF7 cells [106]. However, FI-293 cells expressing Haps59 lacked a switching phenotype, exhibiting a level of 5FC toxicity nearly equivalent to that of yCD expressing cells in both the presence and absence of CoCl₂ (Figure 5.4C). This lack of switching may be partially due to higher expression levels, as 293 cells are commonly known to express high levels of...
recombinant proteins. In addition, Haps59 may be activated by another binding partner of the promiscuous CH1 domain which is present in 293 cells.

The aberrant activation of Haps59 in FI-293 cells means that these cells have limited use for screening CH1-derived switches identified in E. coli. Moreover, the aberrant activation is potentially problematic for two additional reasons. First, it may indicate that domains other than the HIF-1α C-terminal activation domain (CTAD) are capable of activating Haps59. Second, the aberrant activation of Haps59 in FI-293 cells indicates that delivery of a Haps59 gene via viral gene delivery may be problematic, as many viral proteins are known to interact with the CH1 domain [131].

Creation of FI-RKO cells

As the switching properties of Haps59 were best demonstrated in RKO cells [106], we set out to create a Flp-In compatible version of these cells to facilitate rapid integration and testing of new switches. We created this cell line as outlined by Life Technologies and show here that this cell line contains a single integration of the pFRT/LacZeo2 vector at stable locus. As Southern blots proved difficult without the ability to use radiolabelled probes we used quantitative PCR to measure the copy number of pFRT within the genome as described by Hoebeeck et al. [139] (Figure 5.5A). All of the selected clones appeared to have a single integration. Three clones were chosen with different levels of β-galactosidase activity in the initial screening, for further propagation and characterization. Since these clones have a single integration of pFRT, the level of β-galactosidase activity
Figure 5.5 – Characterization of FI-RKO clones.
(A) Copy number of pFRT relative to ZFN80 and GPR15 in 3 FI-RKO clones, as calculated by quantitative PCR on genomic DNA using the comparative Ct method. Error bars represent standard deviation between 2 technical replicates. (B) β-galactosidase activity relative to total protein in FI-RKO variants at early and late passage.

likely represents the level of transcriptional activity from region of the genome in which pFRT is integrated. We hypothesized that these different clones could be used to examine dosage effects of our switches. Upon further examination (Figure 5.5B), the difference in expression across these clones is minimal and variable over time. Order of magnitude changes would be required to examine dosage effects, which would be more easily achieved using synthetic promoters [143]. This experiment did however confirm stable expression from integrated pFRT over several passages, demonstrating at least preliminarily that pFRT is integrated in a “safe-harbor” within the genome of each of these clones.
Chapter 6: The promiscuous CH1 domain

The CH1 domain, being part of a transcriptional coactivator, binds to many different transcription factors assisting in recruitment of transcriptional machinery and enhancing transcriptional activity. The majority of the known binders of the CH1 domain are oncogenes but unfortunately they are not all as tightly regulated as HIF. Potentially, these binders could also activate Haps59. Thus, Haps59’s ability to be activated only in cancer cells is uncertain. The CH1 domain is known to interact with many human and viral proteins including p53 [60, 63, 64], CITED2 [69, 73, 74, 75], NF-κB-p65 [79, 81], Stat-2 [93, 95], Pit-1 [82, 86, 87], HPV E7 [101], adenovirus E1A [101], and HNF-4 [88, 92]. As we observed significant differences in the behavior of Haps59 between RKO, MCF7, and 293 cells, we were interested in examining the effect these other interactors may have on Haps59.

As an initial study we tested whether CH1-interacting domains from proteins other than HIF-1α can activate Haps59. We replaced the HIF-1α-CTAD in plasmid pGA-HIF with the CH1-interacting fragments of CITED2, p53, and the adenovirus E1A protein and tested these constructs’ ability to confer increased 5FC toxicity to E. coli cells expressing Haps59. CITED2 and p53 were chosen as representative endogenous interactors with CH1. CITED2 is of interest due to its role as a negative regulator of HIF-1α [69]. E1A was chosen as a representative viral protein. E1A is of particular interest as it is a potential cause of undesired activation of Haps59 in FI-293 since 293 cells were originally created by transformation of sheared adenovirus DNA [144] and have been shown to express E1A [145].
We have also put some effort into identifying potential aberrant activators in normal cells by attempting to isolate the binding partners of Haps59 using co-immunoprecipitation. We performed these experiments in the HPDE cell line, which was made by transforming human pancreatic duct epithelial cells with a fragment of the HPV-16 genome containing the E6 and E7 proteins [146]. These cells were subsequently shown to exhibit essentially normal phenotype and genotype [147]. Upon identifying the undesirable binders of Haps59 in normal cells we could then use directed evolution to abolish this binding activity in future generations of HIF-1α-activated protein switches.

Methods

Expression of other CH1-binding peptides in E. coli

The coding sequence of the CH1-binding fragments of CITED2 (residues 220-269) and p53 (residues 1-57) were synthesized using overlapping oligonucleotides [148]. The coding sequence of residues 1-77 of E1A was amplified from genomic DNA of 293 cells by PCR. Homologous regions of the 3’ portion of GST and downstream flanking vector were added to the CITED2 and E1A fragments by PCR. p53 was appended, in-frame, on the 5’ end of GST in a similar fashion. Primers for adding these homologous flanking regions were designed using j5 [127]. These fragments were incorporated into pGA backbone using circular polymerase extension cloning [126], and products were transformed into DH5α E. coli. DNA from sequence verified clones was then used to transform GIA39 E. coli. Chemically competent cells of GIA39 pGA-CITED2 and GIA39 pG-E1A were made [114] and
transformed with pSkunk-Haps59. Fresh colonies of these transformations as well as GIA39 pGA and GIA39 pGA-HIF containing pSkunk-Haps59 were used to perform dot toxicity assays as described in Chapter 2.

**Immunoprecipitations of FLAG-Hap59 and E1A from FI-293 cell lysate**

FI-293 cells containing pCDNA5-FLAG-Haps59 (FHaps59) or pCDNA5-FLAG-yCD (FyCD) were created, grown, and lysed as detailed in the previous chapter, except these cells were not treated with CoCl₂, and in some cases cells were lysed with non-denaturing lysis buffer (20 mM Tris-HCl pH 8, 137 mM NaCl, 1% Triton X-100, 2 mM EDTA, 1/500 protease inhibitors). In later experiments, 10 µM zinc acetate was added to lysis buffer. All immunoprecipitation procedures were performed at 4°C. Lysates were diluted to 1-2.5 mg/mL total protein in 200 µL total volume and 1-5 µg of antibody was added. These solutions were incubated at 4°C overnight with constant shaking. Following this incubation, 25 µL of Protein G magnetic beads (NEB) were added and incubated overnight (or in some cases 4 hours) at 4°C with constant shaking. Tubes were briefly centrifuged to collect solution and then placed in a Magna GrIP Rack (Millipore) to precipitate the beads. Beads were washed with three times with 500 µL lysis buffer, and then finally resuspended in 40 µL 1X LDS loading buffer with 10mM DTT (Life Technologies). Samples were incubated at 80°C for 10 minutes and placed in the magnetic rack. Samples were moved to a new tube and 10-20 µL was separated on a Novex Nupage 4-12% Bis-Tris Gel (Life Technologies) using MES running buffer per the manufacturers protocol. Proteins were transferred as above except using a continuous buffering system containing 2X Transfer buffer (Life Technologies), 10%
methanol, 1% Nupage Antioxidant (Life Technologies). Membranes were blocked with 5% milk (Bio-Rad) in tris-buffered saline plus 0.5% tween-20 (TBST) for 1 hour at room temperature. All primary antibodies were diluted into TBST plus 5% milk and incubated with the blot overnight at 4°C with gentle shaking. Sheep anti-yCD antibodies (Thermo) were diluted 1/500 as were rabbit anti-E1A (Sigma). Mouse anti-E1A (Sigma), mouse anti-FLAG-M2 (Sigma), and rabbit anti-FLAG antibodies (Sigma) were diluted 1/1000. In the morning, the blot was washed briefly in TBST and prepared for the Snap ID protein detection system (Millipore) as recommended by the manufacturer. The blot was washed 3 times with 75 mL of TBST in the Snap ID and then incubated for 30 minutes at room temperature with a 1/6000 dilution of goat anti-mouse (Bio-Rad), goat anti-rabbit (Bio-Rad), or donkey anti-sheep secondary antibodies (Bethel Labs) in TBST plus 0.5% (w/v) milk. The blot was washed again as before, prior to development with Clarity Western ECL substrate (Bio-Rad) in a Universal Hood II (Bio-Rad).

**Culture, electroporation and immunoprecipitations from HPDE cells**

HPDE cells were provided by Dr. James R. Eshleman and cultured in keratinocyte serum free media (Gibco). Cells were passaged using 0.05% trypsin (Gibco), which was neutralized with an equal volume of defined trypsin inhibitor (Gibco).

HPDE cells were found to transfect poorly with lipid based reagents. Instead a Gene-Pulser II with Capacitance Extender II was used to electroporate these cells. Cells that were ~75% confluent were trypsinized, neutralized, and centrifuged at 100 × g for 5 minutes. Cells were then washed with PBS and resuspended at a
concentration of $2.5 \times 10^6$ per mL in PBS. Plasmid DNA (pCDNA-EYFP or pCDNA-FHaps59) was added to a final concentration of 10 µg/mL, and 400 µL was transferred to a 0.4 cm-gap electroporation cuvette. Cells were electroporated at 200 V and 950 µF and diluted into growth medium. Cells were incubated for 36 hours at 37°C prior to lysis as described above.

Five hundred µg of each lysate in 1 mL was added to an initial volume of 100µL anti-FLAG-M2 agarose beads (Sigma) that had been washed with lysis buffer. This solution was incubated overnight at 4°C with gentle shaking. Beads were then spun down at 3000 × g for 2 minutes and washed three times with 500 µL lysis buffer. Complexes were eluted from the beads in 60 µL 0.1 M glycine pH 2.5 for 5 minutes. Eluate was removed from the beads after centrifugation and neutralized with 6 µL 0.5 M Tris pH 8.5, 1.5 M NaCl. Eluate was mixed with 4X LDS loading dye to 1X with 10 mM DTT and electrophoresed as above. Western blots were performed as above. Silver staining was done using Bio-Rad’s silver staining kit.

Analogous experiments were repeated with 293 cells.

**Results**

**Aberrant switch activation in *E. coli***

We fused the known CH1-binding fragments of CITED2, p53, and E1A to GST—in similar fashion to the studies originally confirming the interaction of these fragments with the CH1 domain—and tested the ability of these fragments to activate Haps59 and confer 5FC-dependent toxicity in *E. coli*. We found that CITED2 activates Haps59 to a similar extent as HIF-1α and potentially even more so. p53
Figure 6.1 – Effects of CITED2 and E1A expression on Haps59 as assessed by E. coli dot toxicity assay.

(A) GIA39 E. coli expressing Haps59 and GST, GST-HIF-1α-CTAD, GST fused to the CH1-interacting fragment of CITED2, or p53 were spotted on plates containing increasing levels of 5FC. (B) After initial experiments in 293 cells it came to our attention that E1A also interacts with the CH1 domain. The CH1-interacting fragment of E1A was also cloned into pGA and tested for its ability to activate Haps59.

and E1A however do not (Figure 6.1). It is possible that this effect is due to the lower affinities of E1A and unphosphorylated p53 for the CH1 domain.

HIF-1α and the CH1 domain bind with an affinity of 7 nM [33]. In performing its function as a negative regulator of HIF-1α, CITED2 outcompetes HIF-1α for binding of the CH1 domain with an IC50 of 3.7 nM [149]. The fragment of E1A expressed here has a Kd of 220 nM for the CH1 domain [101]. The p53 fragment has an affinity of 1.13 µM [60]. Upon phosphorylation at Ser15, Thr18 and Ser20 p53-CH1 affinity increases greater than 10-fold [61, 150]. It is possible that mimicking phosphorylation by mutating these residues to aspartic acid would increase activation of Haps59.
We were unable to find evidence in *E. coli* that E1A is responsible for the aberrant activation seen in FI-293 cells. It is possible that E1A levels in 293 cells are much higher (relative to Haps59) than levels achieved in *E. coli*. It is also possible that the GST-fusion interferes with the interaction between E1A and the CH1 domain.

**Immunoprecipitations of E1A and Fhaps59 from FI-293 cells**

We also attempted to detect an E1A-Haps59 interaction in FI-293 cells. We lysed FI-293 cells stably expressing Fhaps59 or FyCD and performed immunoprecipitations under several conditions using a variety of anti-E1A and anti-FLAG antibodies. Western blots on the precipitated proteins were used to detect interacting species.

We were unable to coprecipitate E1A and FHaps59 from lysates of FI-293 cells further confirming our results in *E. coli* (Figure 6.2). Interaction between these two proteins was not observed under any conditions, whether using RIPA or non-denaturing lysis buffer and monoclonal mouse anti-E1A, polyclonal sheep anti-E1A, monoclonal mouse anti-FLAG, monoclonal rabbit anti-FLAG, or polyclonal sheep anti-yCD antibodies at various concentrations. Zinc was also added to the lysis buffers to no effect. Although we could not find any evidence of interaction between E1A and FHaps59, this does not prove that E1A is not responsible for activation Haps59 in FI-293 cells in the absence of CoCl₂. It may be possible to identify the aberrant activator or activators in these cells using protein mass spectrometry or an antibody microarray, but this would be more worthwhile in a physiologically relevant system.
Figure 6.2 – Immunoprecipitation of FLAG-tagged yCD and Haps59 or E1A from FI-293 cells.
FI-293 cells stably expressing FyCD or FHaps59 were lysed in the presence of zinc and antibodies against the FLAG epitope or E1A were added. These antibodies and bound proteins were precipitated from solution using Protein G magnetic beads. All proteins were then eluted from these beads and western blotted for the presence of E1A or the FLAG-epitope. (A) Immunoprecipitations were done with mouse anti-FLAG and mouse anti-E1A antibodies. (B) Immunoprecipitations were done with rabbit anti-FLAG and mouse anti-E1A antibodies, except for * and which was done with mouse anti-FLAG. * and ** were eluted with glycine instead of lysis buffer. The mouse anti-E1A antibodies recognize 2 isoforms of E1A.

Immunoprecipitations of FHap59 in HPDE cells

There have been several cell lines developed, which are reported to have near normal genotype and phenotype. This is especially true in recent years with the discovery of hTERT immortalization. Although cells in 2-D culture may not be exactly physiologically relevant, we proposed these cell lines may provide a more
therapeutically relevant platform for identifying proteins that interact with Haps59. As an initial study we attempted to isolate such proteins using a human pancreatic duct epithelial cell line, which was immortalized by transformation with the genes encoding HPV E6 and E7. After these experiments were completed we discovered that the CH1 domain and HPV E6 are known to interact. HPV E6 may outcompete other interactions, making this cell line a poor choice for these studies. If similar studies are pursued in the future, an hTERT-immortalized cell line or even primary cells would be more ideal.

After significant optimization HPDE cells were transfected via electroporation with an efficiency of almost 50% but still poorly expressed FHaps59, which is barely visible in western blots (Figure 6.3A). We were able to immunoprecipitate FHaps59 from these cells; however, due to its low expression we were unable to discriminate any specific binding partners. There is no discernable band for the known binding partner HPV E6, which is ~19 kDa. Viewed optimistically, the low levels of Haps59 accumulation in HPDE cells may be due to the lack of binding partners in these cells. Regardless, the only visible difference between the eluate samples is the FHaps59 band and increased antibody fragments (25 and 50 kDa) in the EYFP sample (Figure 6.3B). These antibody fragments are due to bead carry-over after elution. In future experiments, FyCD would be a much better control allowing discrimination between yCD-binding proteins and CH1-binding proteins.
Figure 6.3 – Immunoprecipitation of FHaps59 from HPDE and 293 cells visualized by silver staining.

(A) Lysate of HPDE cells transiently expressing EYFP or FHaps59 were mixed with anti-FLAG antibody conjugated beads to precipitate FHaps59 and any interacting proteins. The precipitated proteins and lysate were electrophoresed and gels were silver stained. FHaps59 is indicated by the arrowheads, and its identity was confirmed by western blot. (B) Analagous experiments were performed in 293 cells. All lanes are from the same gel, just rearranged to reflect the order of samples in A.

The 293 cells had higher expression of FHaps59 than HPDE cells (Figure 6.3 B). Large-scale precipitations and purifications might be used to identify the binders of FHaps59, especially in cells able to express FHaps59 to higher levels. Additionally, preclearing of lysate, blocking of the beads, and more stringent washing could improve the sensitivity of detection. The challenge is finding a near normal cell line that transfects well, or making stable cell lines expressing high levels of FHaps59 while retaining otherwise normal phenotype and genotype.
Chapter 7: Conclusions and future directions

In directing evolution there are almost always an inconceivable number of directions forward; the problem lies in picking the most fruitful path. In the case of these HIF-1α-activated protein switches this decision is certainly challenging. While there are many directions forward for potentially improving the therapeutic window of these switches, it is still uncertain how these switches will function in a therapeutic context. To move forward in the most rational manner, we must identify the challenges ahead on the path to the clinic.

A 3rd generation of Haps

Improving the switching activity beyond that of Ehaps22 may not be of immediate priority to the translation of these therapeutics, but throughout this work several prospects for improvement have been identified that deserve mention. Although the error prone PCR library we generated was relatively fruitless, it was also a relatively narrow search. PFunkel, a new technique for performing random or targeted mutagenesis of whole codons was developed soon after this library was created [46]. This method allows for comprehensive codon mutagenesis—mutating every codon in a gene to every other possible codon. Such a library would cover the additional 60% of mutations missed by our error-prone PCR library. Additionally, only a single cycle of mutation and selection was performed using error-prone PCR. This cycle could be repeated (with either error-prone PCR or PFunkel) rapidly generating more diversity and potentially realizing more well behaved protein switches. It may also be advantageous to perform several rounds of PFunkel
Figure 7.1 – Schematic of FLAG epitope removal using a type IIΔ enzyme. The FLAG tag can be removed from current RICP and TICP libraries by PCR subcloning with a forward primer encoding a type IIΔ restriction site that cuts at the first codon of yCD and a reverse primer that anneals at some restriction site beyond the yCD coding sequence. To retain directionality these PCR products would first be digested with the type IIΔ enzyme and blunted. Then they would be digested with the type II enzyme and ligated with an appropriately prepared vector.

mutagenesis and positive selection on yCD alone, allowing yCD to reach its optimum in *E. coli*. This optimized variant may be much more amenable to switch creation.

The RICP library and linker libraries were interrogated quite thoroughly with the exception of a few (notably 2/3, 3/2, and 3/3 N-/C-terminal linker libraries). Only a minuscule amount of the TICP library was examined, however. Based on the lack of switches resulting from the RICP library it may be best to remove the FLAG epitope from the current and any future TICP libraries. The FLAG tag could be removed from any of the existing libraries by PCR using a primer annealing to the FLAG tag sequence and encoding on its 5’ end a type IIΔ restriction site with a long distance between its recognition site and cut site such as AcuI, Bpml, BpuEl, BsgI (all 16 bp), or Mmel (20 bp) (Figure 7.1).
It may also be possible to increase the number of switches found while retaining the FLAG tag by adding a linker between the FLAG tag and the N-terminus of yCD. This could be approached in the same manner as the linker libraries, making a library of all possible linkers between the FLAG tag and Haps59 and selecting for the member with the best switching. However, it may only require a short flexible linker, and several could be design and screened rapidly.

**Isogenic integration in colon carcinoma cells**

The development of FI-RKO cells will, hopefully, significantly expedite the translation of these switches. This cell line should provide an ideal platform for further screening of the several switches isolated throughout this work. Toxicity assays in this cell line will provide necessary quantification of HIF-1α- and 5FC-dependent toxicity. FI-RKO cells will allow switch candidates developed in *E. coli* to be rapidly tested in this cancer cell line, while simultaneously establishing cells for future xenograft studies in mice.

This cell line could potentially be improved, for *in vitro* assay use, by integrating an oxygen-stable variant of HIF-1α under the control of an inducible promoter. This expression cassette could simply be added to pCDNA5. Cell lines could also be created expressing a truncated (CH1-binding defective) version of HIF-1α, allowing near complete isolation of HIF-1α’s effect on toxicity. These assays still disregard physiological relevance, which stands as a weighty potential stumbling block just ahead of delivery, dose, and schedule.

Dosage could be explored somewhat in FI-RKO cells, by integrating several promoter of various strengths controlling expression of the same switch. This
experiment could significantly increase our knowledge of the mechanism of these switches. Based on the current stability model, if a switch is expressed at high levels overloading the cell’s degradation machinery, it may no longer exhibit a switching phenotype, whereas lower expression levels may expand the therapeutic window of the same switch. Having this basic understanding of the optimum protein dose may be useful in guiding delivery vector selection or design.

**Are CH1-based switches therapeutically relevant?**

Or is promiscuous binding a limiting factor? This is the principal question moving forward. The answer will guide future work on this project and shed significant light on questions raised throughout this work. Unfortunately, getting to that answer is not a simple task.

Testing the cancer specificity of CH1-based switches in a therapeutically relevant *in vivo* system will initially require optimization of one or many systemic gene delivery vehicles as well as 5FC dose and schedule. Fortunately, many previous trials using yCD/5FC have been performed in mice, aiding in this optimization. We propose that these vehicles be tested in tumor bearing mice for their ability to deliver Haps59, or Ehaps22, while simultaneously measuring off-target activation of these switches. Overall delivery would be measured using a standard luciferase reporter, while spatially resolved switch activation (cytosine deaminase activity) would be measured using $[^{19}\text{F}]$-labeled 5FC via magnetic resonance [151, 152, 153]. Tissues with significant off-target activity could be harvested and Haps-binding proteins could be isolated by co-immunoprecipitation and identified by protein mass spectrometry. As the majority of known CH1-binding partners are identical in
mice and human (Table 1.1), this experiment would hopefully allow us to identify aberrant activators due to similar levels of homology.

This complex experiment would require significant resources but would provide a near-complete picture of how these therapeutics will function in a relevant system. This would also allow us to identify areas of off-target activity and potentially identify the responsible aberrantly activating molecule. Switches can then hopefully be evolved in such a manner as to abolish this binding activity while retaining affinity for HIF-1α and other cancer-specific binders.

It may be that the CH1 domain can exquisitely identify cancer cells, but it is also possible that the list of known CH1 binding partners is so rich in oncogenes because cancer systems are studied more frequently than normal physiology. A simpler (and perhaps the ideal preliminary) route to identifying the binding partners of Haps59 in normal cells types may be to use several different hTERT-immortalized cell lines. hTERT cell lines from many different tissues are now commercially available. Stable pools expressing FHaps59 and FyCD should be created as many of these cells are poorly transfected (similar to most primary cells). Immunoprecipitations in these stable lines can be optimized more readily compared with our experiments in HPDE cells, as transfections will not need to be repeated for each test. With an optimized immunoprecipitation the proteins interacting with FHaps59 but not FyCD, as demonstrated in silver-stained polyacrylamide gels, can be excised from these gels, and identified by protein mass spectrometry.

An alternative way to identify sites of off-target activity and associated aberrant activators, while still working in a physiologically relevant system, may be
to perform immunohistochemistry and immunoprecipitations on normal tissue biopsies. GST-CH1 fusions purified from *E. coli* have been shown to interact with HIF-1α *in vitro* [154], and it may be possible to use this fusion or similar fusions (e.g. GST-FHaps59, HRP-CH1, MBP-CH1) for immunohistochemistry and immunoprecipitation. However, the high concentrations of CH1-fusions required to outcompete endogenous CH1 binders may limit hinder these experiments.

Initially, immunohistochemistry on sections of biopsies from a variety of tissues could be used to identify tissues with a high concentration of CH1-binding partners. The identified biopsies could then be lysed, incubated with GST-CH1, and bound to a glutathione column, which binds GST. However, it may be better to use a different tag for immunoprecipitation as many human tissues express GSTs of their own. Regardless of this decision, the proteins interacting with the CH1 domain could possibly be eluted by chelating the CH1-coordinated zinc, causing the CH1 domain to unfold [55]. The eluted interactors could then be identified by protein mass spectrometry. This approach could be tested initially and optimized using lysates of mammalian cells over-expressing CITED2 or a lower affinity CH1-binding partner.

**Evolving a cancer-specific CH1 domain**

With the binding partners of the CH1 domain identified, we could then abolish the affinity of the CH1 domain and existing switches for their binding partners in normal tissues using directed evolution. Initially, the fragments of these proteins that both interact with the CH1-domain and express well in *E. coli* can be identified, by coexpressing a variety of fragments of these proteins (or potentially the whole protein) along with Haps59 and measuring the 5FC-sensitivity conferred
by each using the above described dot toxicity assay. Once Haps-binding fragments are identified, they can be expressed (likely one at a time) during negative selections on a library of Haps variants. This library would contain all possible single, double, and triple, codon mutations within the CH1 (except those amino acids coordinating zinc) and could be made rapidly using PFunkel [46]. Positive selections, with HIF-1α being expressed, would then be performed on the remaining members to ensure that the mutations do not abolish HIF-1α binding. Several rounds of mutation and selection will likely be needed to completely abolish binding of the aberrant activators while retaining HIF-1α binding.

Conclusion

This work has modestly improved the HIF-1α-dependence of 5FC toxicity conferred to *E. coli* by HIF-1α-activated protein switches. In addition, we have developed several novel techniques for switch creation and improved our knowledge of the design principles of yCD-CH1 switch creation, which will facilitate the development of future switches. Hopefully, in the near future the improvements demonstrated in *E. coli* will be confirmed in the newly created FI-RKO cells. This cell line will also aid in the translation of future switches. While we strongly believe in the cancer therapeutic switch concept, more research is necessary to confirm the cancer specificity of CH1 domain-based switches preceding further translation. We have made some attempts to characterize the cancer-specificity of these switches, and based on the knowledge gained from these attempts, we have laid out a more direct route forward.


Curriculum Vitae

Robert Clayton Duclos Wright was born to Steven and Barbara Wright on August 12, 1985 in Lynchburg, VA. After graduating from North Mecklenburg High School in 2004, he attended North Carolina State University. He graduated summa cum laude with his Bachelor of Science degree in Chemical and Biomolecular Engineering in 2008. In the fall of 2008, Clay began doctoral studies in the Department of Chemical and Biomolecular Engineering at the Johns Hopkins University. In 2010, he was awarded a Graduate Research Fellowship from the National Science Foundation.