

**APPROACHES TO IMPROVE EXPRESSION AND SPECIFICITY  
OF AN ANTIBODY PROBE AGAINST FIBRONECTIN**

A Thesis  
Presented to  
The Academic Faculty

By

Leandro Moretti

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science in  
Bioengineering

Georgia Institute of Technology  
December 2016

**COPYRIGHT© 2016 BY LEANDRO MORETTI**

**APPROACHES TO IMPROVE EXPRESSION AND SPECIFICITY  
OF AN ANTIBODY PROBE AGAINST FIBRONECTIN**

Approved by:

Dr. Thomas Barker, Advisor  
School of Biomedical Engineering  
*Georgia Institute of Technology*

Dr. Andrés Garcia  
School of Mechanical Engineering  
*Georgia Institute of Technology*

Dr. Susan Thomas  
School of Mechanical Engineering  
*Georgia Institute of Technology*

Date Approved: 16 August 2016

## ACKNOWLEDGEMENTS

I would like to acknowledge the numerous people who welcomed and supported me during my short time at Georgia Tech. First and foremost, I would like to thank my advisor Dr. Thomas Barker, for choosing me as his next student. His guidance and optimism made even the longest working hours enjoyable, and I look forward to continue this collaboration in the future. I would also like to thank Dr. Andrés García and Dr. Susan Thomas for agreeing to be part of my committee without hesitation. Furthermore, I would like to thank Dr. Anton Bryksin, molecular biologist extraordinaire, for his technical advice and support without which this project could not come to fruition.

I am thankful for Laura Paige and Shannon Sullivan, for helping during all of my program changes in just one year. I am particularly grateful to the fellow graduate students in the Barker lab: John Nicosia for always being a friendly and helpful presence, Haylee Bachman for being a cheerful research companion, Vicky Stefanelli for being a role model for seriousness and dedication, and Dwight Chambers for becoming my mentor and going above and beyond the collaboration expected out of colleagues. They all contributed to a productive but light hearted work environment, a rare finding from which I benefitted greatly and will miss moving forward.

I need to also thank Mark Colasurdo, my roommate, who started this adventure in Atlanta with me. As we supported each other through classes and other tough times, I truly enjoyed the time spent together. I would also like to thank my best friends Alec Willard and Matteo Diddi, who keep cheering me on even if separated by miles of land and ocean. I am especially grateful for my girlfriend, Amber Goad, who has been my moral support

throughout this time, helping me overcome bumps and larger obstacles along the road with kindness, tenderness and unconditional love.

Most importantly, I would like to acknowledge my family, who never doubted my ability to succeed. I have been fortunate to have parents willing to make several sacrifices in order for me to achieve the best possible future, even if at the cost of proximity. Even if they have given me love and support in their own ways, I would not have all of my accomplishments without them.

In the end, I would like to dedicate this work to my grandparents who are not among us anymore but would have celebrated this accomplishment of mine.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
LIST OF ABBREVIATIONS.....	viii
SUMMARY .....	x
<u>CHAPTER</u>	
1 LITERATURE REVIEW .....	1
IDIOPATHIC PULMONARY FIBROSIS.....	1
RELEVANCY OF FIBRONECTIN CONFORMATION.....	4
PHAGE DISPLAY TECHNOLOGY.....	9
ISSUES WITH PROTEIN EXPRESSION IN E. COLI.....	12
2 RECOVERING THE H5 SEQUENCE .....	18
EXPERIMENTAL SETUP.....	19
RESULTS AND DISCUSSION.....	20
CONCLUSION.....	25
3 OPTIMIZING THE H5 SEQUENCE.....	26
CODE IMPLEMENTATION.....	27
RESULTS AND DISCUSSION.....	28
CONCLUSION.....	29
4 ALTERING SPECIFICITY BY INTRODUCING VARIATION IN THE H5 SEQUENCE.....	31
EXPERIMENTAL SETUP.....	31
RESULTS AND DISCUSSION.....	33
CONCLUSION.....	39
5 CONCLUSION AND FUTURE DIRECTIONS.....	40
APPENDIX A: H5 SEQUENCES.....	43
APPENDIX B: CODON OPTIMIZATION CODE .....	45
REFERENCES .....	52

## LIST OF TABLES

<b>Table 1:</b> Primers used for PCR reactions and sequencing of H5 and pIT2 vector.. .....	20
---	----

## LIST OF FIGURES

<b>Figure 1:</b> Tomography scan showing classic features of IPF..	3
<b>Figure 2:</b> Modular structure of fibronectin subunit	5
<b>Figure 3:</b> Steered molecular dynamics of simulation of early stages of FNIII10 unfolding under force.....	7
<b>Figure 4:</b> Schematic representation of different antibody formats.	10
<b>Figure 5:</b> Map of pIT2 vector (Lifesciences).....	19
<b>Figure 6:</b> Amplified H5 stock products of different size.....	22
<b>Figure 7:</b> Amplified samples of various H5 scFv stocks.	23
<b>Figure 8:</b> Amplified samples of the anti-Ubiquitin scFv and the H5 DNA sequences....	24
<b>Figure 9:</b> Products of the control PCR reactions (Phusion and Taq)	34
<b>Figure 10:</b> H5 mutants digestion products.....	35
<b>Figure 11:</b> pIT2 digestion products.....	36
<b>Figure 12:</b> Ligation of pIT2 and H5 mutants products.....	37
<b>Figure 13:</b> Transformed DH5 $\alpha$ serial dilutions on LB agar plates..	38

## LIST OF ABBREVIATIONS

10III	fibronectin 10th type III repeat
9III	fibronectin 10th type III repeat
cfu	colony forming unit
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleoside triphosphate
dTTP	deoxythymidine triphosphate
ECM	extra cellular matrix
ELISA	enzyme-linked immunosorbent assay
ePCR	error prone polymerase chain reaction
Fc	fixed chain domain of antibodies
Fn	fibronectin
IPF	idiopathic pulmonary fibrosis
LB	lysogeny broth
mRNA	messenger ribonucleic acid
PCR	polymerase chain reaction
RGD	arginine, glycine, aspartic acid
scFv	single chain variable fragment
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
SPR	surface plasmon resonance

TGF $\beta$

transforming growth factor beta

tRNA

transfer ribonucleic acid

Ub

ubiquitin

## SUMMARY

Phage display is a convenient method to select proteins of interest based on binding affinity. The Barker lab recently used this technology to discover an antibody fragment (scFv), termed H5, capable of selectively binding to the integrin binding domain of fibronectin (Fn) under a strained configuration, typically due to forces exerted on the extracellular matrix (ECM) by cells.

Large scale production of H5 scFv is hindered by non-optimized codons and suspected protein toxicity, since the *E. coli* strain producing the protein grew poorly at normal temperatures. Moreover, the transfected vector containing H5, pIT2, appeared to be degraded in consecutively selected bacterial cultures, as suggested by the isolation of significantly shorter than expected plasmids.

After significant effort, reconstruction of the pIT2 vector sequence was accomplished from multiple colonies and cross referencing with the provided anti-ubiquitin scFv antibody fragment, and the complete DNA sequence of H5 was recovered. The sequence was codon optimized for expression in an *E. coli* strain engineered for high levels of scFv production. Furthermore, the H5 DNA was recombined with error prone PCR, to generate a library of random mutants, which was transformed into a stable, *E. coli* strain. This library will be panned on the targets again through phage display in order to find an improved version of H5, defined by more selective binding to the extended conformation of the integrin binding domain of Fn.

# **CHAPTER 1**

## **LITERATURE REVIEW**

### **IDIOPATHIC PULMONARY FIBROSIS**

Idiopathic pulmonary fibrosis (IPF) is the most common form of interstitial lung disease. Incidence of IPF in the US is deemed to be 20 to 60 cases per a hundred thousand people, to which 14000 to 34000 more cases are diagnosed each year (Álvarez, Levine, & Rojas, 2015). The disease typically affects older people, preferentially male (1.5 to 1 ratio with females), with 66 being the median age of diagnosis (King Jr, Pardo, & Selman, 2011). IPF is a scarring, progressive pathology with a 5-year survival rate of 20% (Kreuter et al., 2015). Early symptoms of the disease include breathlessness upon exertion and dry cough that interferes with the patient's daily activities (Meltzer & Noble, 2008). Given the patients age, these traits are often attributed to aging or other syndromes, thus delaying the diagnosis. Clinically, IPF is characterized by inspiratory crackles during auscultation, often along with clubbing. Patients also suffer from decreased forced vital capacity and forced expiratory volume, coupled with impaired gas exchange, which can be demonstrated by measuring the diffusion capacity (Meltzer & Noble, 2008).

Guidelines for diagnosis from the American Thoracic and European Respiratory Societies (ATS/ERS) recommend a combined clinical exam, computer tomography and lung biopsy. Typical patterns of fibrosis visible through high resolution computer tomography consist of evident bilateral, peripheral and subpleural reticular infiltrates. Often, the presence of advanced fibrosis is indicated by honeycomb changes and traction bronchiectasis (Meltzer & Noble, 2008). Lung biopsy is still considered the gold standard for IPF diagnosis, because it allows detection of histopathological lesions and structures

like fibroblastic foci, which are pale-staining whirls of loose extracellular matrix molecules (mostly collagen), interspersed with numerous fibroblasts (Meltzer & Noble, 2008). However, the high degree of spatial heterogeneity of fibrotic tissue, as shown in Figure 1, hinders the overall usefulness of biopsies when not assisted by imaging techniques. The current Clinical Practice Guideline on management of IPF recommends several pharmacological treatments such as pirfenidone, a downregulator of key profibrotic factors, and nintedanib, an inhibitor for several tyrosine kinases. While the latter two compounds showed most promise in recent clinical trials, none of them represent a cure for IPF since patients continue to degenerate despite such treatments (Kreuter et al., 2015). Because today IPF is still an incurable disease outside of a lung transplant, it is necessary to extensively investigate the mechanisms behind the etiology and progression of IPF, in order to identify targets for therapy in the near future.



**Figure 1:** Tomography scan showing classic features of IPF. Arrowheads illustrate honeycomb structural changes, while arrow shows traction bronchiectasis. (Meltzer & Noble, 2008).

A major pathological process in IPF is the deposition of provisional matrix, made up by fibrinogen and fibronectin (Fn), in the alveolar spaces following lung injury. Evidence suggests that this provisional matrix could stimulate epithelial to mesenchymal transition even without TGF $\beta$  (King Jr et al., 2011). Persistence of such matrix leads to the differentiation of fibroblasts to myofibroblasts which organize in fibroblastic foci and release excessive amounts of ECM (King Jr et al., 2011). Fibroblasts that have differentiated down a contractile, myofibroblastic pathway are known to exhibit significant contractile force. Such cell-derived forces are capable of stressing the surrounding ECM, leading to increased microenvironmental stiffness, including Fn strain (Brown, Fiore,

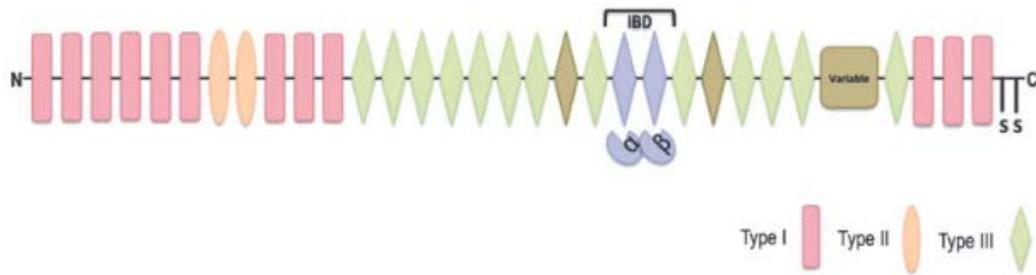
Sulchek, & Barker, 2013). Myofibroblast presence and the related stress increase could become a significant IPF marker to facilitate diagnosis.

Our lab and others have shown that fibrotic lung ECM is significantly stiffer compared to normal ECM, using atomic force spectroscopy (Brown et al., 2013). Moreover, recent investigations have suggested that reduced lung compliance is associated with increased integrin  $\alpha5\beta1$  mediated Fn deposition and basement membrane matrix assembly (Bennett et al., 2013) Thus, it is hypothesized that the mechanical changes to the Fn matrix may be a molecular signature for progression of the disease.

The severity of IPF and the need for a quicker diagnosis warrant further research on developing potentially non-invasive probes to detect biophysical changes of Fn and disease progression.

### **RELEVANCY OF FIBRONECTIN CONFORMATION**

The ECM of various tissues have different composition and organization, despite the same basic components being water, polysaccharides and proteins. Fibrous ECM proteins form three-dimensional scaffolds that support cells and influence tissue morphology (Frantz, Stewart, & Weaver, 2010). Rigidity of the ECM scaffold acts as a signal to the surrounding cells through to mechanotransduction (Mochitate, Pawelek, & Grinnell, 1991). Among ECM fibular proteins, Fn is of particular interest because of its presence in the provisional matrix of tissues undergoing regenerations, its effects on cell behavior through specific integrin interactions, its high degree of spatial flexibility, and unfolding response to force.



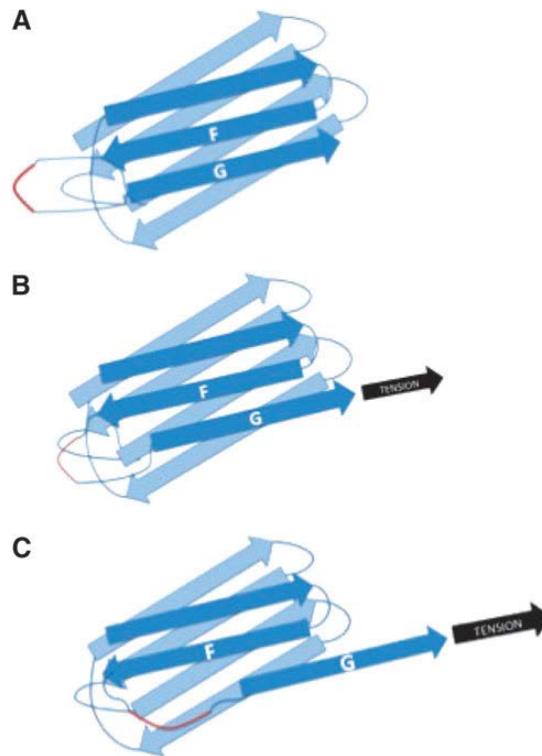
**Figure 2:** Modular structure of fibronectin subunit (Bachman, Nicosia, Dysart, & Barker, 2015).

Fn is a glycoprotein formed by two identical subunits weighing 220 KDa each and are covalently linked by a couple of disulfide bonds towards the C-termini. A single subunit is composed by three repeated types of domain: Type I, Type II and Type III (Bachman et al., 2015). While the former two types are set in a rigid conformation because of internal disulfide bonds, the latter type completely lacks this feature. Specifically, Type III domains are made up by antiparallel  $\beta$  sheets connected by flexible loops and held together just by hydrogen bonds (Gao et al., 2003; Main, Harvey, Baron, Boyd, & Campbell, 1992). Consequently, Type III repeats are highly susceptible to unfolding due to force (Krammer, Craig, Thomas, Schulten, & Vogel, 2002).

10III was ascertained to be the Type III repeat with the lowest threshold for deformation with both steered molecular dynamics and atomic force microscopy (“The Mechanical Hierarchies of Fibronectin Observed with Single-molecule AFM,” n.d.), implying that it will be one of the first domains to unfold under force.

Fn 10th Type III (10III) repeat possess a canonical arginine, glycine, aspartic acid, serine (RGDS) peptide sequence, which acts as a cell binding site, while the 9th Type III repeat contain a synergistic site that provides domain recognition and mediates cell adhesion and cytoskeletal organization (Dufour et al., 1988; Pierschbacher & Ruoslahti,

1984). The proximity of these two sites has been the object of several studies because of the importance of physical coupling between extracellular and integrin selectivity. Indeed, RGD peptides by themselves are not enough to select specific integrin binding, since they are ubiquitous in other ECM proteins and ligate several kinds of integrin (Pierschbacher & Ruoslahti, 1984). For example, integrin  $\alpha 5\beta 1$  binding requires RGD and the aforementioned synergy sequence (PSHRN) on the 9th Type III repeat (Mardon & Grant, 1994) to be about 32 Å apart. However, when a force of just 10 pN is applied, the 10th repeat partially unfolds and the RGD domain is removed from the synergy site, so that the distance between the two increases to 55 Å (Krammer et al., 2002). This 23 Å change greatly reduces  $\alpha 5\beta 1$  binding (Stabenfeldt, Brown, & Barker, 2010). Fibroblasts lower bound for mechanical force exerted is in the tens of nN (Tan et al., 2003). Conversely, integrins that do not coordinate binding with the synergy site, such as  $\alpha v\beta 3$ , are not affected by the change in distance (Brown, Rowe, & Barker, 2010).



**Figure 3:** Steered molecular dynamics of simulation of early stages of 10III unfolding under force. RGD loop is shown in red between the F and G strands. As the G strand of 10III (shown in (A)) is pulled (B), the RGD loop is brought closer to the bulk of the module (C) (Bachman et al., 2015).

Thanks to its binding sites for integrins, ECM proteins, growth factors and other Fn molecules, Fn is associated with coordinated tissue behaviors, including morphogenesis, wound healing (Pankov & Yamada, 2002), and even pathologies like IPF *in vivo*. This led to hypothesize that the force sensitivity of Fn could be considered a mechanism of mechanosensitive control of ligand recognition (Krammer, Lu, Isralewitz, Schulten, & Vogel, 1999). To investigate this mechanism, recombinant 9III and 10III were expressed as single Fn fragments for *in vivo* integrin targeting experiments. Furthermore, a leucine to proline modified version, Fn 9\*10, was developed to provide greater stability to RGD and the synergy site (Krammer et al., 2002). This new fragment offered more clustering of  $\alpha 5$  integrins than compared with wild type Fn. This and other assays suggest that integrin

engagement selectivity affects disparaged cells responses. On the other hand, recombinant 9III and 10III were expressed with the addition of short flexible linkers, which are mainly made up by a varying of consecutive glycine residues, in order to provide tensile character and increase distance between the synergistic sites (Grant, Spitzfaden, Altroff, Campbell, & Mardon, 1997). Our lab has successfully employed a similar Fn fragment, dubbed 4G whose linker is comprised by 4 glycine residues. Fn conformation affects the ability of cells to migrate on or assemble a Fn matrix, when only  $\alpha v$  integrins are expressed, while those two functions are recovered upon the reintroduction of  $\alpha 5$  subunits ( $\beta 1$  as second integrin subunit in both cases) (Zhang et al., 1993). The action of  $\alpha 5 \beta 1$  in Zhang et al. assay mirrors the effects on lung compliance in IPF by Fn and the same integrin in wild type cells. As previously stated, Fn plays a major role in wound healing and the provisional matrix. It is theorized that the integrin specificity mechanism would play a role in that too. While the Fn matrix mimicking constructs by the Hocking group (Roy & Hocking, 2012) showed improved wound repair, the contribution of integrins  $\alpha 5 \beta 1$  and  $\alpha v \beta 3$  could not be conclusively discerned. Also, integrin specific Fn like constructs on polyacrylamide (PAA) gels of varying stiffness managed to mask the stiffness induced epithelial-to-mesenchymal transition, as long as the  $\alpha 5$  and  $\alpha 3$  subunits were engaged.

In the end, Fn conformation has far reaching impact beyond the mechanical properties as a substrate or ECM component thanks to its integrin specificity switch. Nonetheless, new probes are needed to fully understand the complexities of Fn and its effects on physiological process and pathologies. Specifically, while sufficient *in vitro* evidence for the unfolding of the Fn domains exists, a reliable method to image conformational changes in more complex tissues or *in vivo* is needed.

## PHAGE DISPLAY TECHNOLOGY

Phage display technology was first reported on by Smith in 1985 (Smith, 1985) when the technique was employed only for the selection of simple peptides. As the technology was refined, application such as pathogen detection (Petrenko & Vodyanoy, 2003), enzyme activity enhancement by randomly altering the active site (Clark & March, 2006; Fernandez-Gacio, Uguen, & Fastrez, 2003), but phage display rose to prominence because of its simplicity in monovalent antibody fragment (Fab) production compared to monoclonal antibody production in hybridomas (Ascione et al., 2005).

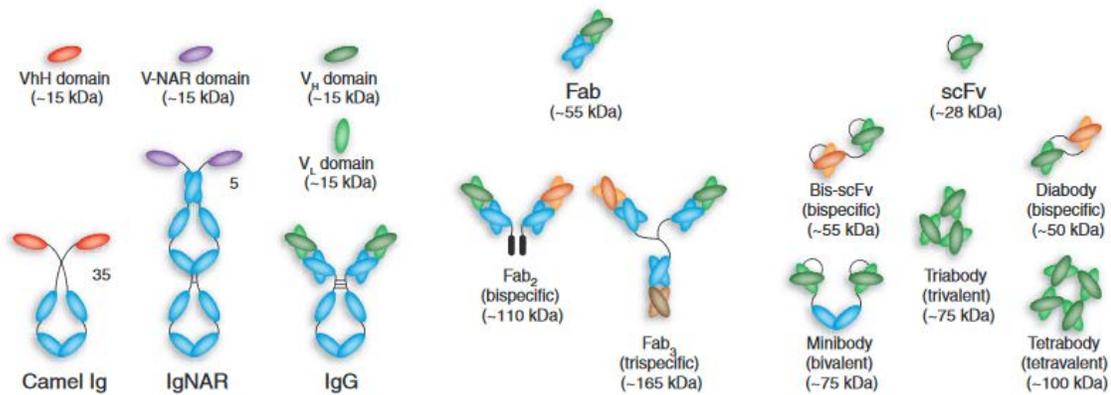
Arguably, the most lucrative product of this approach is Adalimumab (marketed as Humira and approved only in 2008), a full size antibody, which is a rheumatoid arthritis drug that was derived through phage display. This drug acts by binding TNF $\alpha$ , preventing the interaction with its receptor, thus blocking the inflammatory response of autoimmune diseases (Humira, 2016).

Intact antibodies like immunoglobulins gamma (IgG) are bivalent, with high functional affinity and high retention times. The presence of the fixed chain (Fc) domain, however, can lead to undesirable effects in certain applications such as poor contrast in imaging or the triggering of cytokine release *in vitro* and especially *in vivo*. These issues were solved by engineering monovalent fragments. Single chain antibody fragments (scFv), Fab, single variable V<sub>H</sub> and V<sub>L</sub> domains all belong to the above category (Holliger & Hudson, 2005).

The advantages of scFv over intact antibodies do not end with the lack of the Fc segment. Because this antibody format is made up by the V<sub>H</sub> and V<sub>L</sub> domains joined by a

flexible linker portion, it enables targeting of cryptic epitopes because of their reduced size (Ward, Güssow, Griffiths, Jones, & Winter, 1989). For example, several viruses have their target receptor binding domain inside narrow cavities of their surface antigens, making them inaccessible to intact antibodies. Despite not being the ideal antibody fragments for *in vivo* tumor targeting due to their rapid clearance and poor retention due to the monovalent binding capacity and size, scFv provide high specificity to target traditionally inaccessible binding sites, such as G protein receptors, enzyme active sites and cryptic proteins sites (Holliger & Hudson, 2005).

Thanks to the discovery of high affinity single V-like domain in camelids and sharks (Dooley & Flajnik, 2005; Genst et al., 2004), *in vivo* viable human single domains have been designed and made into libraries after addressing the issues with stability, solubility, resistance to aggregation (Jespers, Schon, James, Veprintsev, & Winter, 2004).



**Figure 4:** Schematic representation of different antibody formats, showing intact ‘classic’ IgG molecules alongside camelid VhH-Ig and shark Ig-NAR immunoglobulins. Camelid VhH-Ig and shark Ig-NARs are unusual immunoglobulin -like structures comprising a homodimeric pair of two chains of V-like and C-like domains (neither has a light chain), in which the displayed V domains bind the target independently. Shark Ig-NARs comprise a homodimer of one variable domain (V-NAR) and five C-like constant domains (C-NAR). A variety of antibody fragments are depicted, including Fab, scFv, single-domain V H , VhH and V-NAR and multimeric formats, such as minibodies, bis-scFv, diabodies, triabodies, tetrabodies and chemically conjugated Fab’ multimers (sizes given in kilodaltons are approximate) (Holliger & Hudson, 2005).

Phage display is considered ideal for selection of scFv for other characteristics as well (Packer & Liu, 2015). Firstly, it offers a covalent linkage between genotype and phenotype of the screened peptide, allows library sizes of  $10^{10}$  (Clackson, Hoogenboom, Griffiths, & Winter, 1991), focuses on binding interactions, and the fragments can be usually expressed in the periplasmic space of a bacteria like *E. coli* with little to no difficulty. The technique is also accessible once phage libraries are created or purchased. Typically, antibody repertoires are fabricated by ligating the  $V_H$  and  $V_L$  PCR products, then cloned into a display vector in order to fuse the protein of interest with a phage coat protein, usually gene III (gIII). Certain phage displays formats, denominated phagemid, have a vector that does not encode all the genes necessary for phage replication ensuing infection. Thus, a helper phage is used to infect *E. coli* cells that have been transformed with the incomplete vector (Hammers & Stanley, 2014). Once produced from the libraries, phages are incubated with antigen, oversampling the antibody repertoires. After incubation with antigen and stringent washes, phages are eluted by incubation with trypsin protease. This step will elute phage by cutting the c-myc tag between antibody fragment and the phage gIII protein. It will also remove background infectivity originating from gene III protein of the trypsin sensitive KM13 helper phage, since the gIII protein originating from the repertoire phagemid is trypsin-resistant. The eluted phage is then used to infect *E. coli* bacteria and titres are determined by plating of dilution series. For subsequent rounds of selections, colonies from the first round are scraped from agar plates and phages are produced in liquid culture, PEG purified and selected by binding to antigen. After three rounds of selection, individual clones are isolated, grown overnight and soluble fragments

are produced in a 96-well format. Finally, antigen-specific clones are identified by ELISA (C. M. Y. Lee, Iorno, Sierro, & Christ, 2007).

Following a similar phage display protocol, the Barker lab has successfully developed high binding affinity and highly selective scFv with potential for several applications in the study of integrin binding to Fn and IPF mechanisms. The 4G and 9\*10 described earlier were used as epitopes to model the molecular scale conformational spacing between the PHSRN and RGD sites. By performing phage display and selection using Tomlinson I + J antibody libraries, the isolated scFv antibodies demonstrated specificity in binding to the 4G and 9\*10 fragments in an ELISA assay, which represent a “stretched” and a “relaxed” conformations respectively. After surface plasmon resonance (SPR) analysis, the  $K_D$  for the stretched state was found to be 16 nM, while the  $K_D$  for the relaxed state was 107 nM, making the selected scFv, dubbed H5, bind preferentially to the stretched configuration of the Fn fragment, as intended (unpublished data).

### **ISSUES WITH PROTEIN EXPRESSION IN E. COLI**

*E. coli* has been used as a host for recombinant protein production, including immunoglobulin fragments, because of its advantages over mammalian cells. This Gram negative bacteria can produce the target protein in large quantities, grows much faster than other hosts and transformation with plasmids is simple and requires small amounts of DNA (Verma, Boleti, & George, 1998). For the above reasons, *E. coli* enables inexpensive protein production with a quick turnaround time as well. There exist categories of proteins that cannot be synthesized in bacteria, such as glycosylated proteins, thus excluding full immunoglobulins from this expression system.

Recombinant antibody fragments expressed in the cytoplasm form insoluble inclusion bodies, mainly because of the relatively reductive environment there. Aggregates are often caused by interactions with misfolded proteins. After harvesting the target protein, several downstream processing strategies are available to promote the correct refolding of the fragment, which involve forming the proper disulfide bonds, if any (Verma et al., 1998). Since scFv contain 2 disulfide bonds, a large investment of time and effort would become necessary to reform the correct combination of sulfide bonds. The more disulfide bonds are present in a protein, the higher the number of possible combinations. For two of such bonds, there would be three combinations, implying that if the bonds formed randomly, only one third of the refolded product would possess the correct tertiary structure (Jaenicke, 1991). An alternative approach to this ordeal involves using a leader sequence to direct the antibody fragment into *E. coli* periplasmic space. The space between the inner and outer membranes of Gram negative bacteria is more oxidizing compared to the cytoplasm, contains chaperonin equivalents and disulfide isomerases (Skerra & Pluckthun, 1988). The pelB leader sequence at the N terminus, which is cleaved by signal peptidase in the periplasm, has been used successfully for decades for this purpose (Ferenci & Silhavy, 1987; Lei, Lin, Wang, Callaway, & Wilcox, 1987). Other signal peptides are available for the same application and are reviewed elsewhere (Choi & Lee, 2004; Mergulhão, Summers, & Monteiro, 2005). Osmotic shocks or cell wall permeabilization avoids the need for cell lysis while allowing secretion of the product of interest in the culture medium (Shokri, Sandén, & Larsson, n.d.).

Even when feasible, optimizing antibody fragment production in *E. coli* is no small feat. Choosing the most appropriate strain, promoter and vector are decisions to be made

before any expression attempt. Arguably, the most common strain used for protein expression is BL21 (DE3), since it lacks Lon protease (Gottesman, 1996) and outer membrane protease OmpT (Grodberg & Dunn, 1988), both of which would cause degradation of the foreign protein being expressed. Moreover, plasmid loss is prevented because of the hsdSB mutation in the parental strain. Other derivative strains from BL21 offer additional features, such as T7 RNA polymerase under the prophage  $\lambda$ DE3, which favor certain applications (Rosano & Ceccarelli, 2014).

The role of promoters is to control the expression of the protein of interest. Good promoter attributes include low basal expression, strong expression post induction, transferable to other *E. coli* strains and rely on a convenient induction mechanism. Moreover, it should be independent from the growth medium (Terpe, 2006). The lac promoter was among the first studied and used as the basis of subsequent iterations, like the tac and trc. Specifically, lac promoter is weak even after induction and “leaky” (relatively high levels of basal expression). While these traits make it hardly usable for large scale and high level expression, they represent an advantage for the production of proteins that are toxic to the host, as is often the case with scFv. While the above promoters are induced by isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), they are susceptible to the metabolic state, manifested by the cytoplasmic level of cAMP. In order to solve this issues, the T7 RNA polymerase system was devised (Terpe, 2006). This polymerase transcribes DNA sequences up to five times faster than the native RNA polymerase and is expressed under a different promoter in the BL21 DE3 strain, which enables strong induction of T7 with IPTG despite the presence of glucose and independently of cAMP levels (Grossman, Kawasaki, Punreddy, & Osburne, 1998).

The vector choice impacts multiple aspects of the expression system. First, it contains the replicon, an origin of replication and its cis-acting control elements, which controls the copy number (Solar, Giraldo, Ruiz-Echevarría, Espinosa, & Díaz-Orejas, 1998). While it is intuitive to correlate a high plasmid number with high levels of protein production, high copy number plasmids risk causing a metabolic burden on the host. This adverse effect leads to plasmid instability and decrease in protein producing bacteria (Bentley, Mirjalili, Andersen, Davis, & Kompala, 1990; Birnbaum & Bailey, 1991). Vectors with wild type ColE1 origin oscillate between 15 and 20 copies per cell (C. Lee, Kim, Shin, & Hwang, 2006), while vectors under ColE1 derivatives range from 15-60 to 500-700 copies (Bolivar et al., 1977; Minton, 1984). Additionally, vectors define the selection marker to keep plasmid-free cells from growing. Antibiotic resistance genes are thus included in the vector backbone. Ampicillin is one of the most common antibiotic used, despite potential for complete depletion in the growth medium due to continuous secretion of  $\beta$ -lactamase (Korpimäki, Kurittu, & Karp, 2003). To obviate this issue, and to allow selection for two different plasmids at once, several other markers are available, such as chloramphenicol and kanamycin. Alternatively, semi-synthetic penicillins have been developed in order to be used in the medium in place of ampicillin. Carbenicillin is one of these compounds, boasting improved stability to temperature, pH conditions and enzymatic degradations (“Ampicillin vs\_ Carbenicillin.pdf,” n.d.).

Careful planning of an expression system tailored for the protein of interest is no guarantee for high level expression. Issues related to the target protein alone can hinder or even prevent production unless specific measures are taken, since no or low level expression of recombinant protein is likely due to protein toxicity. Unwanted interactions

with homeostasis and normal proliferation of the host slows down the growth rate, leading to a lower final cell density (Dong, Nilsson, & Kurland, 1995). If the deleterious effects are noticeable even before induction, it could prove beneficial to switch to a promoter that offers tighter regulations, such as the araBAD or T7 in conjunction with a host expressing T7 lysozyme, a natural inhibitor of T7 RNA polymerase. If changing promoter is not an option, using media supplemented with glucose as a source of carbon can be effective for the lac family of promoters, in order to avoid the inducer (Studier, 2005). If the toxic effects are noticeable only after induction, tunable promoters should be used or a new host like C41(DE3) or C43(DE3) should be considered. Both strains lead to higher level expression of proteins considered highly toxic in BL21 DE3, however C43(DE3) offers better result and favors plasmid stability (lack of plasmid degradation) after induction when compared to its cognate strain (Dumon-Seignovert, Cariot, & Vuillard, 2004). Culture temperature can also play a factor, since lowering it from 37°C reduces the potential for inclusion bodies and aggregation of the target protein and can reduce the plasmid copy number (Saida, Uzan, Odaert, & Bontems, 2006).

If host growth were unaffected, but the protein production was still lackluster, toxicity can be excluded and the investigation should focus on the DNA sequence codon bias. Theoretically, satisfactory protein expression could be easily achieved by codon optimizing the insert. However, there are two opposing school of thoughts on the approach to follow: one focuses on imitating native highly expressed genes, according to the Codon Adoption Index (CAI), with some consideration for mRNA secondary structure (Sharp & Li, 1987); the other prefers usage of rare codons that correspond to higher charging tRNA levels during host starvation (Elf, Nilsson, Tenson, & Ehrenberg, 2003; Welch et al., 2009).

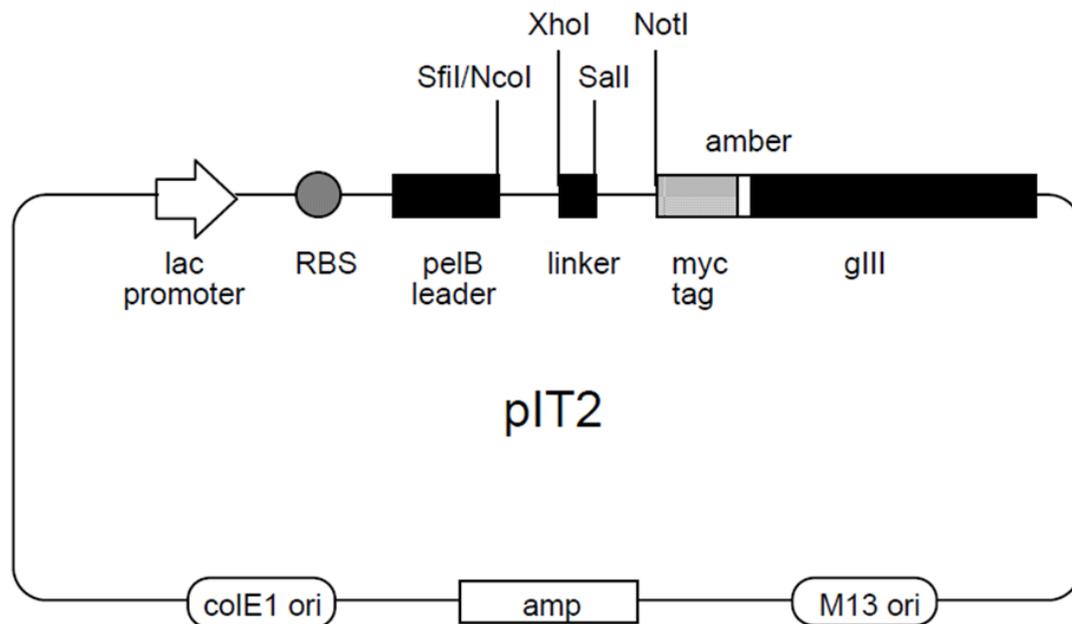
The strategies outlined above informed the methods and approaches presented in this work. Evidence suggests that large scale production of H5 scFv is hindered by non-optimized codons and suspected protein toxicity, since the E. coli strain producing the protein grew poorly at typical temperatures (37 °C). Moreover, the transfected vector containing H5, pIT2, appeared to be degraded in consecutively selected bacterial cultures, as suggested by the isolation of significantly shorter than expect plasmids. The following chapter covers the process that, after significant effort, accomplished the reconstruction of the pIT2 vector sequence and the complete DNA sequence of H5 as well.

## **CHAPTER 2**

### **RECOVERING THE H5 SEQUENCE**

The Barker lab isolated the H5 scFv antibody through phage display from the Tomlinson I + J libraries, developed by Greg Winter's group and distributed by Lifesciences. Despite providing the primers to sequence the scFv inserts, the exact sequence of the pIT2 vector was not made available besides the map reported in Figure XX. This piece of information was used to infer the DNA sequence of pIT2 in order to devise better primers and formulate a starting consensus sequence after the provided sequencing primers did not provide satisfactory results. The newly designed primers allowed amplification of the insert DNA of both H5 and the anti-ubiquitin scFv, provided with the libraries as a positive control during the phage display process. The PCR products of both scFv were sequenced, thus enabling the reconstruction of H5.

Secondly, the hypothesized pIT2 sequence was investigated and confirmed by designing seven sequencing primers pairs that framed roughly 700 bp sections with 100 bp overlaps. Those sequencing results increased our confidence in the pIT2 and H5 consensus sequence.



**Figure 5:** Map of pIT2 vector (Lifesciences). The pelB leader enables secretion of the scFv into the periplasmic space, thus diminishing the aggregation and misfolding typical of a bacterial expression system.

### EXPERIMENTAL SETUP

HB2151 *E. coli* cells expected to contain H5 and the pIT2 vector that were miniprepred by using a spin column kit (Qiagen) attempt to retrieve the H5 sequence. All PCR reactions contained the following mixture: 10  $\mu\text{L}$  of 5X GC buffer (NEB), 1  $\mu\text{L}$  of dNTP (10 mM solution, NEB), 1  $\mu\text{L}$  of DMSO (NEB), 0.6  $\mu\text{L}$  of Phusion polymerase (NEB), 7.5  $\mu\text{L}$  of each primer (2  $\mu\text{M}$  solution in EB buffer, primers purchased from IDT). All reactions included 5  $\mu\text{L}$  of the DNA template and were topped off with DI water to 50  $\mu\text{L}$ . The template solution was prepared from and diluted to 20 ng/  $\mu\text{L}$ .

The reaction mixtures were cycled on a BioRad Dyad thermocycler according to the following protocol. Denaturation took place at 98  $^{\circ}\text{C}$  for 10 seconds, followed by annealing at 75  $^{\circ}\text{C}$  for 30 seconds and extension at 72  $^{\circ}\text{C}$  for 5 minutes, repeated for 30

times. The cycles were preceded by holding the samples at 98 °C at the start for 30 seconds and a final extension step at 72°C for 5 minutes.

**Table 1:** Primers used for PCR reactions and sequencing of H5 and pIT2 vector. All primers were purchased from IDT.

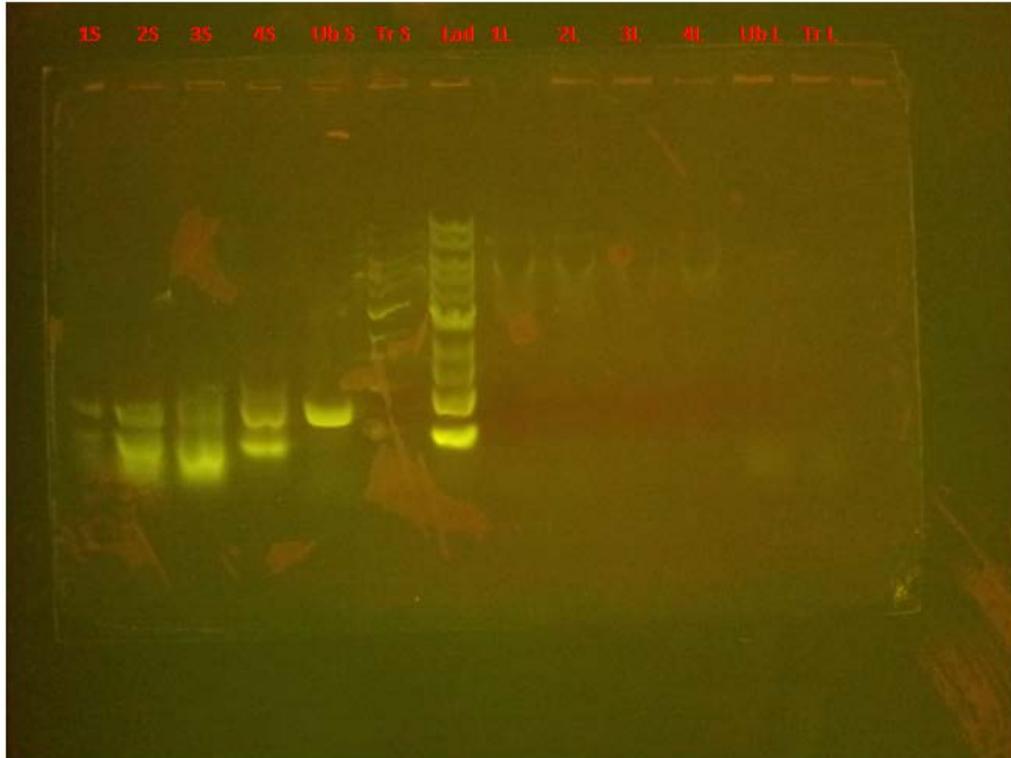
Name	Description	Sequence
118	118SEQ/pIT2/pHEN	CTATGCGGCCCCATTCA
119	119SEQ/pIT2/link seq	CGACCCGCCACCGCCGCTG
133	133SP/pIT2/LMB3	CAGGAAACAGCTATGAC
141	141SP/pIT2H5/VI	CGAAAACCTGCTGATAATTTTGACTACTG
362	362 SEQ/pSB1C3/BB_VF2	CCACCTGACGTCTAAGAAAC
363	363 SEQ/pSB1C3/BB_VR	GTATTACCGCCTTTGAGTGA
665	665 pIT2 + H5 sequencing primer 1327-2081 Forward	GTAGCACC GCCTACATACCT
666	666 pIT2 + H5 sequencing primer 1327-2081 Reverse	CACAACATACGAGCCGGAAG
667	667 pIT2 + H5 sequencing primer 2944-3672 Forward	AAGGTGGAAATCAAACGGGC
668	668 pIT2 + H5 sequencing primer 2944-3672 Reverse	TTGAGGCAGGTCAGACGATT
669	669 pIT2 + H5 sequencing primer 2442-3180 Forward	GTTCAACCACCTCCAGAGACA
670	670 pIT2 + H5 sequencing primer 2442-3180 Reverse	ACCAGTACAAACCACAACGC
671	671 pIT2 + H5 sequencing primer 3161-3929 Forward	GCGTTGTGGTTTGTACTGGT
672	672 pIT2 + H5 sequencing primer 3161-3929 Reverse	GCAGCACCGTAATCAGTAGC
673	673 pIT2 + H5 sequencing primer 1390-2096 Forward	AAGTCGTGTCTTACCGGGTT
674	674 pIT2 + H5 sequencing primer 1390-2096 Reverse	CCGCTCACAATCCACACAA
675	675 pIT2 + H5 sequencing primer 2654-3388 Forward	AGTCTCCATCCTCCCTGTCT
676	676 pIT2 + H5 sequencing primer 2654-3388 Reverse	GGTTTTGCTCAGTACCAGGC
677	677 pIT2 + H5 sequencing primer 733-1511 Forward	GAGGCGGATAAAGTTGCAGG
678	678 pIT2 + H5 sequencing primer 733-1511 Reverse	CTCAGTTCGGTGATAGGTCGT
679	H5_gblock_Fwd	CTATGACCATGATTACGCCAAGC
680	H5_gblock_Bck	GTGATGGTGATGATGATGTGC

## RESULTS AND DISCUSSION

Twenty colonies from a bacteria stock expected to contain H5 and its plasmid were grown and picked on a M9 minimal medium plate. Picked colonies were expanded at 37 °C then had their plasmid DNA harvested and sequenced by using the recommended Lifesciences primers, 363 and 362. Bacterial stock struggled to grow even at the theoretically optimal temperature of 37 °C, thus suggesting cytotoxic effects on cells at 37

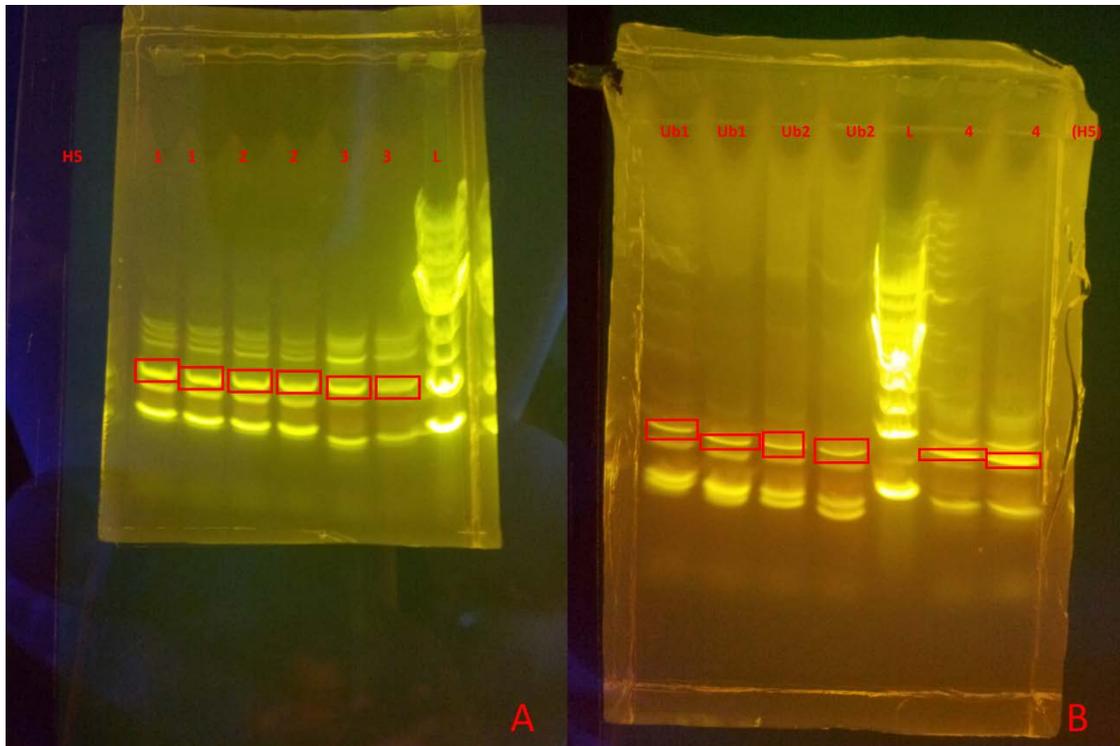
°C of the scFv. All of the reads were made up of short (few hundred bases), with extremely low Q16 and Q20 scores. Because of such results, a detailed pIT2 sequence was hypothesized based on Figure 5 and used to design different primers. Additionally, two new stocks of H5 bacteria were identified, one prepared on May 2012 and another on July 2012. Four cell cultures in SOB were grown at 30 °C to avoid most cytotoxic effects. The cultures were dubbed 1 and 2 from the July stock, while 3 and 4 from the May stock.

Culture tubes 1 and 3 had media supplemented with 1% glucose. Samples were amplified after being miniprepperd, by also using a primer derived from the hypothesized pIT2, 119. As additional control, TG1 cells expressing a scFv from the Tomlinson I + J library targeting ubiquitin (termed Ub), which was expected to share most of the H5 sequence and pIT2. As a negative control, ready to be infected by phage TG1-Tr cells were also used. By using two couples of primers, largely different band sizes were expected, as detailed in Figure 6.



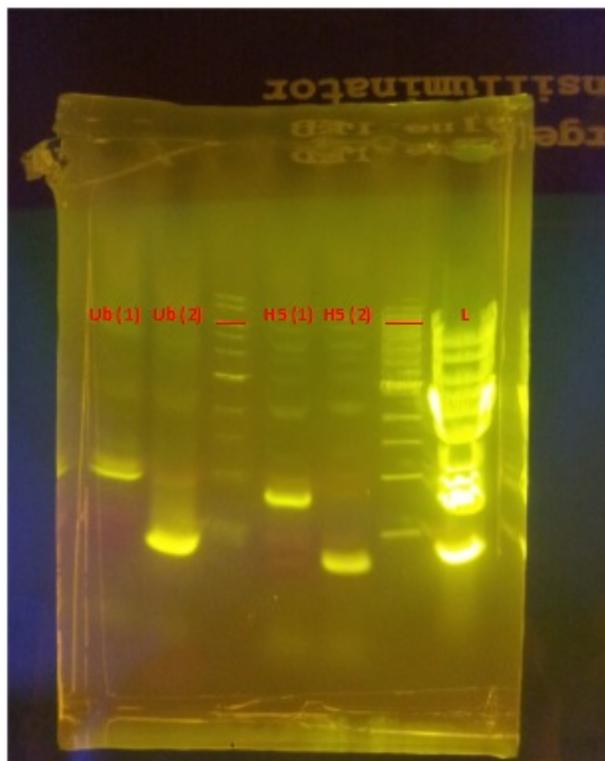
**Figure 6:** Primers 363 and 119 were used in the first 6 out of 12 reactions. Expected product about 1Kb, labeled as S(hort). Primers 363 and 362 were used for the remaining 6 reactions. Product should be 3.5 Kb, thus L(ong). H5 colonies were numbered from 1 to 4, the anti-Ubiquitin TG1 strain was labeled Ub and the TG1 Tr strain was labeled Tr, Ladder used was 1Kb by NEB

The larger size band (from primers 363, 362) was not visible after running the 0.8% agarose gel. Thus, the four cultures described above were used in larger total volumes (100  $\mu$ L) of PCR reactions with primers 119 and 363. While the presence of glucose in media had little effect on cell growth and plasmid sequence, the time at which the bacterial stock was frozen appears to change results drastically. After excising the 1 kb band (Figure 7), purifying the DNA in agarose with a spin column kit (Qiagen), the products were sequenced. Unlike samples 1 and 2, samples 3 and 4 from May 2012 returned low quality reads, that did not match the ones from 1 and 2 nor parts of the hypothesized pIT2.



**Figure 7:** A. The 1kb bright and well defined band (red boxes) was cut out, the agarose dissolved and the three samples were sequenced separately. Notice potentially empty vector/construct underneath samples 1, 2 bands of interest. B. Based on the previous successful extraction, analogous bands were excised. This time the bands of interest appear to be less than 1 Kb. The bright and narrow band (red boxes) was cut out, the agarose dissolved and the three samples were sequenced separately.

The poor sequencing results from 3 and 4 cultures raised suspicion about the possibility of *E. coli* altering the DNA sequence of the plasmid in order to avoid cytotoxic effects from the scFv being expressed, as described earlier in Chapter 1, thus invalidating the whole May stock for H5 recovery purposes. A new PCR amplification set was prepared with culture 1 along with the anti-ubiquitin control, which should slightly differ in DNA sequence but not in size (number of base pairs). Half of the amplifications would use the primer couple 133, 118, which should flank the whole H5 insert, while the couple 141, 118 would cover only one hypervariable region of the scFv, according to the hypothesized vector map.



**Figure 8:** Amplified samples of the anti-Ubiquitin scFv and the H5 DNA sequences. For columns (1), primers 133 - 118 size were used and the product had an expected size of 940 bp. For columns (2), primers 141 - 118 were used and expected product size was about 500 bp. Ladder used is 1 kb from NEB.

Since the sizes for each amplification set matched the expectations, the samples were excised, cleared from agarose then sequenced. Long, high quality reads shared the general consensus developed from the earlier experiments, shown in Appendix A.

Lastly, several primers (665 to 678) were designed using Primer3 in order to have each PCR amplified segment not much longer than 700 bp and roughly 100 bases of overlap for redundancy. The resulting products were purified with a spin column kit, then sequenced. While the entirety of pIT2 could not be sequenced, we are confident about the 3000 bp portion obtained, which encompasses the most relevant parts of pIT2 (upstream and downstream the H5 insert).

## CONCLUSION

Recovery of the H5 sequence was a non-trivial task because of the lack of information on the vector used, pIT2, and the suggested primers unreliability. Furthermore, inconsistent in PCR amplification results while using the same set of primers suggested heterogeneity in our H5 stocks. A significant portion of the bacteria contained the empty pIT2 alone. This is likely due to metabolic processes during the growing phase due to the cytotoxicity of the scFv at 37 °C. Despite all these issues, the H5 and pIT2 sequences were used as the starting point for the codon optimization and the library creation strategies detailed in the following chapters. As an additional consequence of this work, all of the bacterial cultures involving pIT2 and H5 derivatives were grown 30 °C to avoid the insert deletion issues mentioned above.

## **CHAPTER 3**

### **OPTIMIZING THE H5 SEQUENCE**

Achieving high level of protein expression is a fundamental endeavor of every research or industrial project in synthetic biology. More and more often polypeptide and protein production rely on synthetic DNA because of the increase in speed and convenience of DNA manipulation tools (Newcomb, Carlson, & Aldrich, 2007). Given the redundancy of mRNA codons mapped to amino acids, there exist several genes that translate to the same protein even within the same organism. Thus, usual strategies to improve protein expression focus on mimicking the codon bias of natural highly expressed genes in the host of choice. This led to the widespread adoption of the CAI (Sharp & Li, 1987). However, while there are several reports of such strategy working out, it is worth noticing that only the successful cases are reported in the literature, and those differ greatly from case to case in many regards, making it difficult to extrapolate trends that correlate with good levels of expression.

Welch and colleagues attempted to shed some light on the relation between protein expression and DNA (mRNA) sequences with a model relying on partial least square regression that correlated sequence properties with actual expression of a scFv and a DNA polymerase in BL21 DE3 pLysS strain of *E. coli* (Welch et al., 2009). They concluded that while GC bias, runs of C and G, internal Shine-Dalgarno motifs, overrepresentation of codon pairs and RNaseE cleavage sites do not correlate with expression levels, while a distributed property like codon usage is strongly correlated (Welch et al., 2009). A small fraction scFv variants tested showed expression levels well below the model prediction, while also showing complex secondary mRNA structure in the first 15 codons. This

suggests that deleterious 5' region motifs or mRNA secondary structure should still be accounted for best results (Kudla, Murray, Tollervey, & Plotkin, 2009).

This work is particularly relevant because the model developed indicates significant differences in expression levels for genes designed according to the CAI or by focusing on rare codons. It is suggested that the biochemical reason for the success of the latter strategy lies in the charging levels of the tRNA isoacceptor (aminoacyl-tRNA) cognate to the uncommon codons, defined according to the CAI (Elf et al., 2003).

Elf et al. work suggests that in cases of amino acid starvation, the rate of codon reading (mRNA translation) is affected only by the rates of supply and demand for the charged tRNA. Indeed, in the vast majority of cases in which different tRNA read synonymous codons, at least one isoacceptor tRNA remained at a high charging level even when the rate of cognate amino acid supply approached zero. Thus, this theory formulates that charging levels of tRNA isoacceptors will change vastly based on amino acid limitations, with the further ramification that reading of so defined rare codons takes place at a higher rate when there is almost zero supply of the cognate amino acid.

The insights from Welch et al. model and experimental evidence were used to automatize the codon optimization for genes expressing scFv, polymerases or a generic protein to a lesser degree. The routine was written in Matlab (Mathworks) and is reported in the Appendix B section.

## **CODE IMPLEMENTATION**

Matlab was used because of the convenience of the prebuilt functions for DNA sequence manipulation. The main function allows the user to browse for a file in FASTA

format, open it and checks if it contain an amino acid or nucleotide sequence. In the latter case, the sequence is translated to amino acids. Then the codon optimizer function is called within a for loop, once for every amino acid making up the sequence. Given the modularity of the optimizer function, it is possible to change the call line in the main function and codon optimize for a specific category of protein, in this case scFv.

The function reverse-translating amino acids is made up by a switch statement, with one case for each amino acid and the stop codon mark. Every time this function is called, a random number between 0 and 1 is drawn. Each case statement is structured with consecutive if statements, checking if the random number is within an amino acid specific set of intervals between 0 and 1. Those intervals were elaborated from the experimental data on scFv mutants expressed in the work from Welch et al., which are reported in the Supplemental Material (Welch et al., 2009). The variants that showed the highest expression levels were tallied in order to build codon frequency tables that reflected the theory of rare codons affecting translation rate. This data is also available for the DNA polymerase variants investigated and for a combination of the most highly expressed variants (Welch et al., 2009).

After all amino acids are reverse-translated, the main function writes a new FASTA file with the optimized sequence.

## **RESULTS AND DISCUSSION**

After recovering the H5 and pIT2 sequences as described in Chapter 2, the scFv domains and linker parts of the larger H5 sequence were used as input for the Matlab routine, as shown in Appendix A. This was done because the His-tag and the myc-tag were

beyond the chosen restriction enzyme sites, SphI towards the 5' end and NotI towards the 3' end (Appendix A). The trimmed H5 sequence was used as the input for the Matlab routine five times. Those optimized sequences, while expressing the same amino acids, differ in the use of synonymous codon because of the different random number being drawn every time the optimizer function was called. The algorithm was implemented this way because of the nature of the data available, frequency tables. While there exists a sequence that would lead to the best expression among those five, each of them is significantly correlated to lead to higher levels of expression as compared to DNA using traditional strategies (Welch et al., 2009).

The five sequences were inspected to avoid strong secondary mRNA structure in the first 45 bases (15 codons) or other distributed deleterious motifs like four consecutive identical bases (such as 4 G in a row). The chosen sequence is reported in Appendix A. In order to make such sequence useful in the library creation procedure, portions of pIT2 to include the two tags (at the 3' end) and the SphI restriction site were added. This DNA sequenced was then synthesized and purchased in gBlock format from IDT.

## **CONCLUSION**

A short Matlab routine was written to codon optimize DNA sequences for expression in BL21 DE3 pLysS. This method relied on the experimental data from Welch et al. that was combined in frequency tables. While their work was focused on scFv and DNA polymerase, this routine can be adapted to any protein category as long as a codon frequency table is provided because of the built-in modularity.

A shorter version of the optimized H5 sequence was used as input in for this code and was codon optimized. While this program sufficed in its first version for the purposes of its work, it could certainly be improved by automating tasks relevant to the determining the best DNA sequence for expression, like checking for certain deleterious mRNA motifs. Moreover, other studies based on the original work by Welch et al. attempted to improve the accuracy of the model predictions by adding more parameters, beyond the codon translation under starvation speed patterns (Fernandes & Vinga, 2016). For example, Fernandes and Vinga also included free energy of the growing polypeptide in their model. While such improvements would yield better predictions on the expressions level of the target protein and could be adapted in this code algorithm, the potentially marginal improvement in actual, *in vitro* expression might not warrant the trade off in computational speed.

## **CHAPTER 4**

### **ALTERING SPECIFICITY BY INTRODUCING VARIATION IN THE H5 SEQUENCE**

The methods employed to generate potentially beneficial mutations in the H5 antibody fragments are described in this section. First, the BL21 DE3 pLysS optimized DNA sequence was amplified and mutated through ePCR. Secondly, the mutated inserts were ligated into pIT2 using restriction enzymes. Lastly, the new vectors were transformed in competent DH5 $\alpha$  High Efficiency cells, thus creating stable library ready to undergo phage display.

#### **EXPERIMENTAL SETUP**

To achieve a library of mutated clones of H5, the optimized DNA sequence in gBlock format was used as the template in ePCR. First, a 50  $\mu$ L volume control run was set up in order to evaluate efficiency of Taq polymerase (from NEB), following the protocol provided by the manufacturer. The amplified products were compared with the amplification results from a standard PCR run using Phusion polymerase (NEB), also following the provided protocol. Both trial runs were cycled on a BioRad Dyad at 94  $^{\circ}$ C for 1 minute, then at 45  $^{\circ}$ C for 1 minute then at 72  $^{\circ}$ C for 1 more minute. This cycle was repeated 30 times, while no hotstart procedure or final extension step were included (Cadwell & Joyce, 1992). Both products were run on a thick 0.8% agarose gel, excised, purified then sequenced (Macrogen). For the mutagenic reactions, the protocol from Caldwell and Joyce (Cadwell & Joyce, 1992) was adopted with minor variations in order to limit bias in base change. Specifically, twelve 50  $\mu$ L reactions each contained 2.5 units

of Taq polymerase, 10  $\mu$ L of 10X Standard Tag (Magnesium free) buffer (NEB), 20 fmoles of H5 template and 30 pmoles of each primer. To increase errors and base substitutions, 7 mM of magnesium chloride and 0.5 mM manganese chloride were added, while the inclusion of 0.2 mM of dATP, dGTP and 2 mM of dTTP and dCTP was aimed to minimize the base substitution bias. All of the above aliquots were cycled on a BioRad Dyad as described above.

The mutated products were purified with a PCR spin column kit by Qiagen, eluted in EB buffer (Qiagen), then underwent double digestion to expose the “sticky ends” necessary for the ligation into the pIT2 plasmid. 10  $\mu$ g of ePCR purified product was incubated for 1 hour at 37 °C along with 40 units of SphI-HF, 40 units of NotI-HF in Cutsmart buffer (restriction enzymes and buffers from NEB), in a total reaction volume of 100  $\mu$ L. The digested dsDNA was purified with a spin column kit as above.

The full pIT2 plasmid containing the past H5 insert was underwent double digestion as well. After harvesting the entire plasmid from pre-existing H5 stocks grown in SOB media via miniprep, 1  $\mu$ g of the vector was added to 10 units of units of SphI-HF, 10 units of NotI-HF in Cutsmart buffer (NEB), in a total reaction volume of 100  $\mu$ L, for an incubation time of 90 minutes at 37 °C. In order to verify the completion of the digestion, the products were run on a thick 0.8% agarose gel, excised and purified with the Qiagen kit as above.

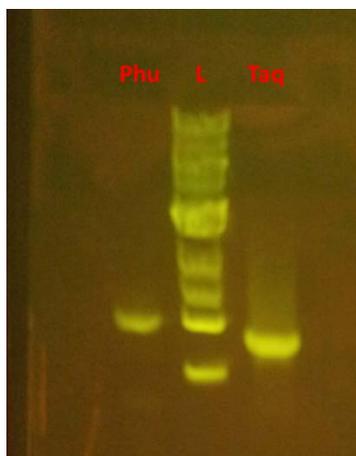
The ligation of the mutated inserts in the open pIT2 vector took place immediately after. The protocol adopted was partially adapted from You and Zhang (You & Percival Zhang, 2012). To ensure high efficiency, 1  $\mu$ g of mutated inserts and 0.5  $\mu$ g of the empty plasmid were incubated with 1200 units of T4 ligase (NEB) in T4 buffer (NEB) for a total

reaction volume of 600  $\mu\text{L}$ . The reagents were added on ice, then the vial was kept at 4  $^{\circ}\text{C}$  for 24 hours. After the reaction completed, the newly formed library was purified with a PCR purification kit as above.

Purified plasmids were then transformed into DH5 $\alpha$  High Efficiency competent cells from NEB, along with the pUC19 control vector, following the provided protocol: four aliquots of 50  $\mu\text{L}$  were thawed on ice, then mixed with 5  $\mu\text{L}$  of the purified library (45 ng of cDNA) each. The four mixtures were kept on ice for 30 minutes, heat shocked at 42  $^{\circ}\text{C}$  for 30 seconds, placed on ice again for 5 minutes then added to 950  $\mu\text{L}$  of SOC medium and incubated at 37  $^{\circ}\text{C}$  for 1 hour while shaking at 250 rpm. The four aliquots were pooled then serial dilutions in SOC were used to seed LB agar plates in duplicate (100  $\mu\text{L}$  of culture per plate) in order to estimate the library size (Daugherty, Chen, Iverson, & Georgiou, 2000). LB medium was supplemented with 100  $\mu\text{g}/\text{mL}$  of carbenicillin and 1% glucose. The remaining starting culture from the transformation was grown in 100 mL of the same LB media to create a large library stock. After 24 hours of growth at 30  $^{\circ}\text{C}$ , 20 colonies were picked at random from the plates to be grown in LB medium supplemented as above, minipreped and sequenced to estimate the error rate of the library (Daugherty et al., 2000).

## **RESULTS AND DISCUSSION**

The products from the regular PCR using Taq and Phusion polymerase respectively yielded bands of comparable sizes when compared on the agarose gel, as shown in Figure 9.

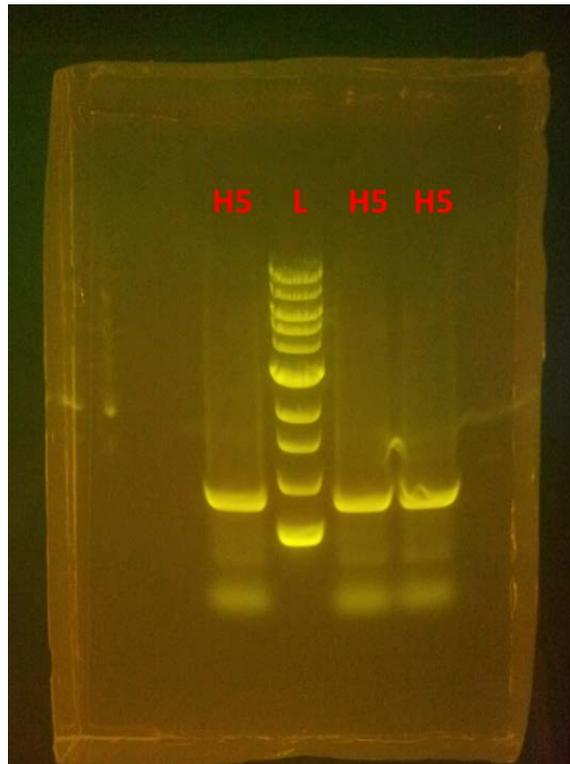


**Figure 9:** Products of the control PCR reactions. Left column contains products from Phusion amplification, the right products from Taq polymerase and the central column is 1 kb ladder from NEB. The same volume of DNA was loaded in each of the wells (4  $\mu$ L of sample and 1  $\mu$ L of 6x loading dye). The apparent difference in product size can be explained by the not uniform run of the dye (not shown) which implies agarose gel artifacts.

The sequencing reaction confirmed that the two products were indeed the H5 sequence including the restriction sites but without the histidine tag section. Using a NanoDrop spectrophotometer after spin column purification and elution in 50  $\mu$ L of EB buffer, the final sample concentrations of the test runs were 25 ng/ $\mu$ L for the Phusion and 104 ng/ $\mu$ L for the Taq polymerase. These results suggested that Taq polymerase could provide a significant amplification of the gBlock used as starting material, even though the two yields are not directly comparable due to difference in starting template amounts and different definition of enzyme units (NEB website, accessed 20 June 2016). The amount of amplified DNA obtained from the test Taq PCR was used to plan the amount of reactions and reagents needed to create a sizeable library, as explained in the next paragraph.

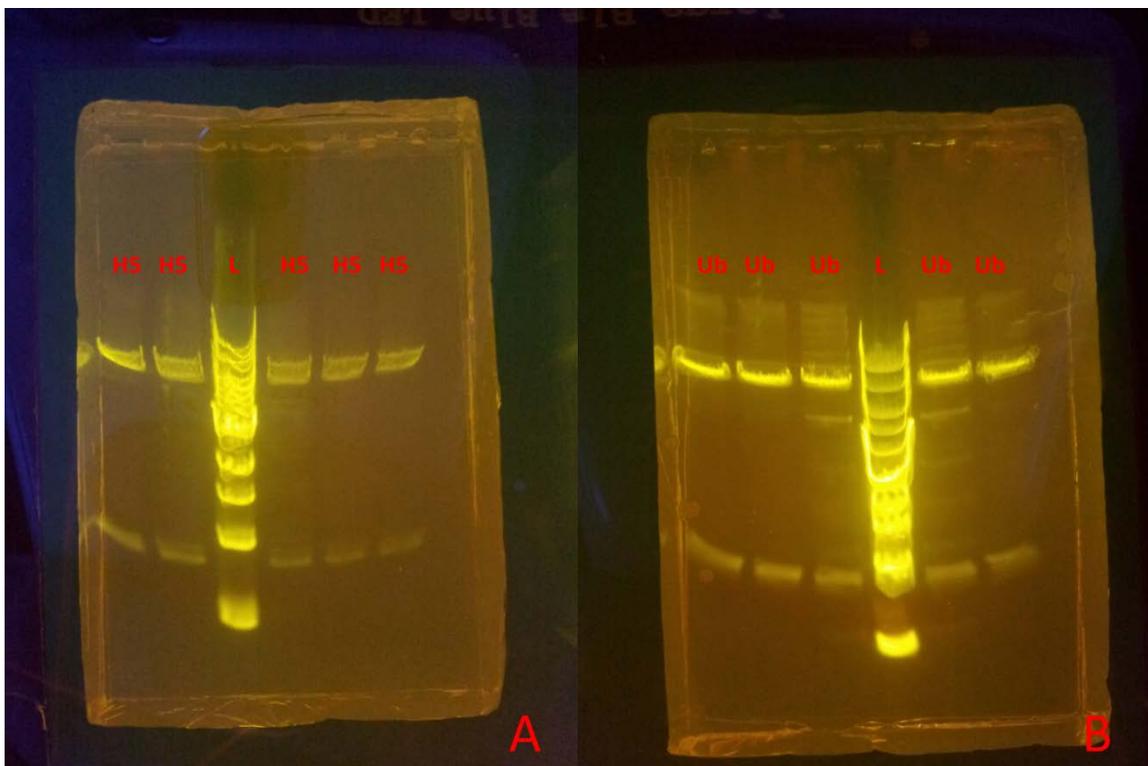
The target library size was chosen to be  $10^5$  clones, equivalent to  $10^5$  colonies forming units (cfu) on an agar plate. Despite the theoretical efficiency of the competent cell line, a conservative  $10^5$  cfu/ $\mu$ g of DNA efficiency was assumed, thus the required

amount of ligated plasmids was 1  $\mu\text{g}$ . The molecular weights of the pIT2 vector with insert, empty pIT2 and H5 cloned insert were calculated to be  $3.24 \times 10^6$ ,  $2.71 \times 10^6$  and  $5.36 \times 10^5$  g/mol respectively, based on the sequences of the starting samples. Consequently,  $3.08 \times 10^{-13}$  moles of ligated pIT2 + insert clones would be needed to create a library of such a size. Assuming a ligation efficiency of 10% despite the long reaction time and low temperature used, as well as a 10X stoichiometric excess of inserts in the ligation mixtures to drive the pIT2+insert ligation instead of other combinations,  $3.08 \times 10^{-12}$  moles of vector and  $3.08 \times 10^{-11}$  moles H5 clones would be required. These amount translate to roughly 8.3  $\mu\text{g}$  and 16.5  $\mu\text{g}$  of dsDNA respectively. Furthermore, about a four-fold factor was included to account for human errors and product loss during the several purification steps, leading to the needed amounts of 30 and 60  $\mu\text{g}$  of pIT2 vectors and H5 clones.



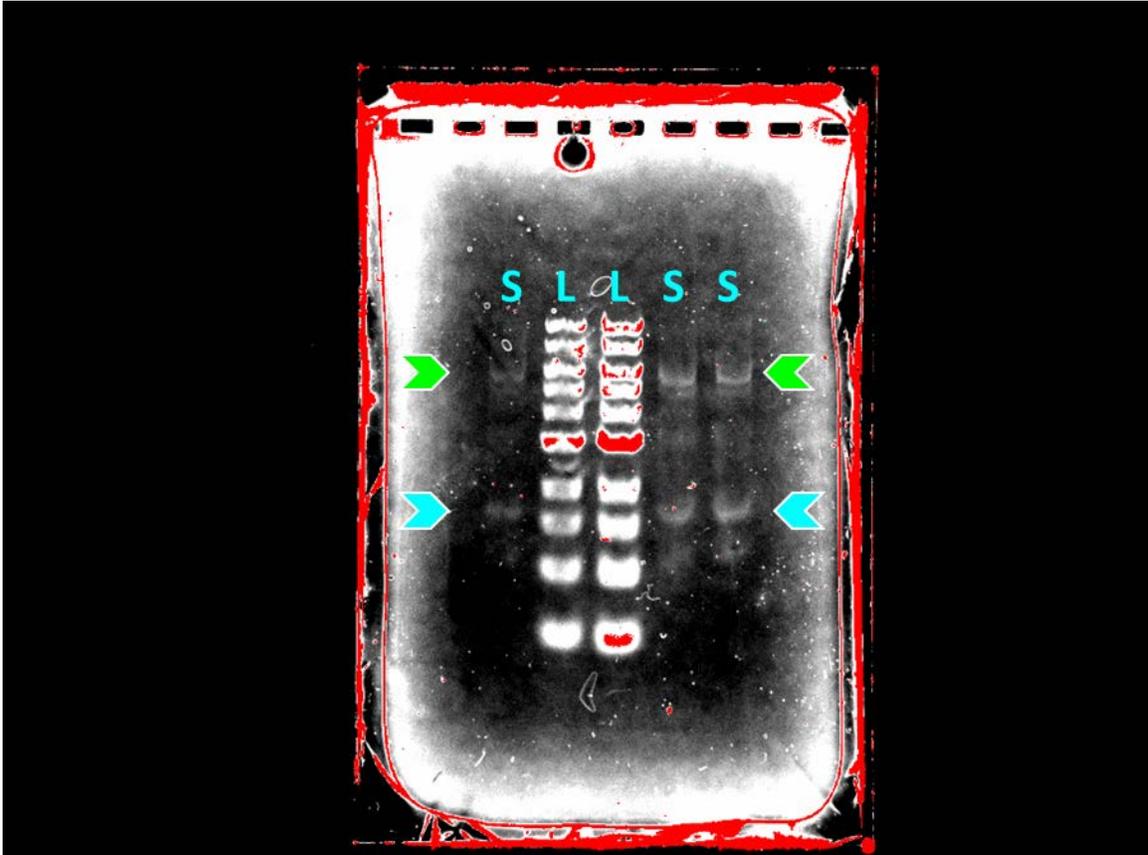
**Figure 10:** agarose gel containing aliquots from the same digestion reaction in three lanes. The size of the main band in each H5 lane matches the expected size and the short segments trimmed at the 5' and 3' ends end pooled closest to the bottom of gel (faint band), suggesting that the intended digestion succeeded.

The High Fidelity versions of the two restriction enzymes were employed in this protocol for two reasons: they have maximum efficiency in the same Cutsmart buffer and show limited deleterious star activity. These advantages saved time, effort and contributed to quality digested products. The digested pIT2 products run on agarose gel confirm the size of the empty pIT2 vector along with the H5 insert being excised, as shown by the bands below the 1 kb line in Figure 11.



**Figure 11:** A. shows the digested pIT2 plasmid from identical reaction vials. The plasmids were first harvested from a stock colonies confirmed for expressing the original H5. The band about 1 kb represents the H5 insert, now removed. B. shows the digested pIT2 plasmid from identical reaction vials. The plasmids were first harvested from a stock colonies expressing the anti-ubiquitin scFv used as control in phage screening. The band about 1 kb represents such insert, now removed.

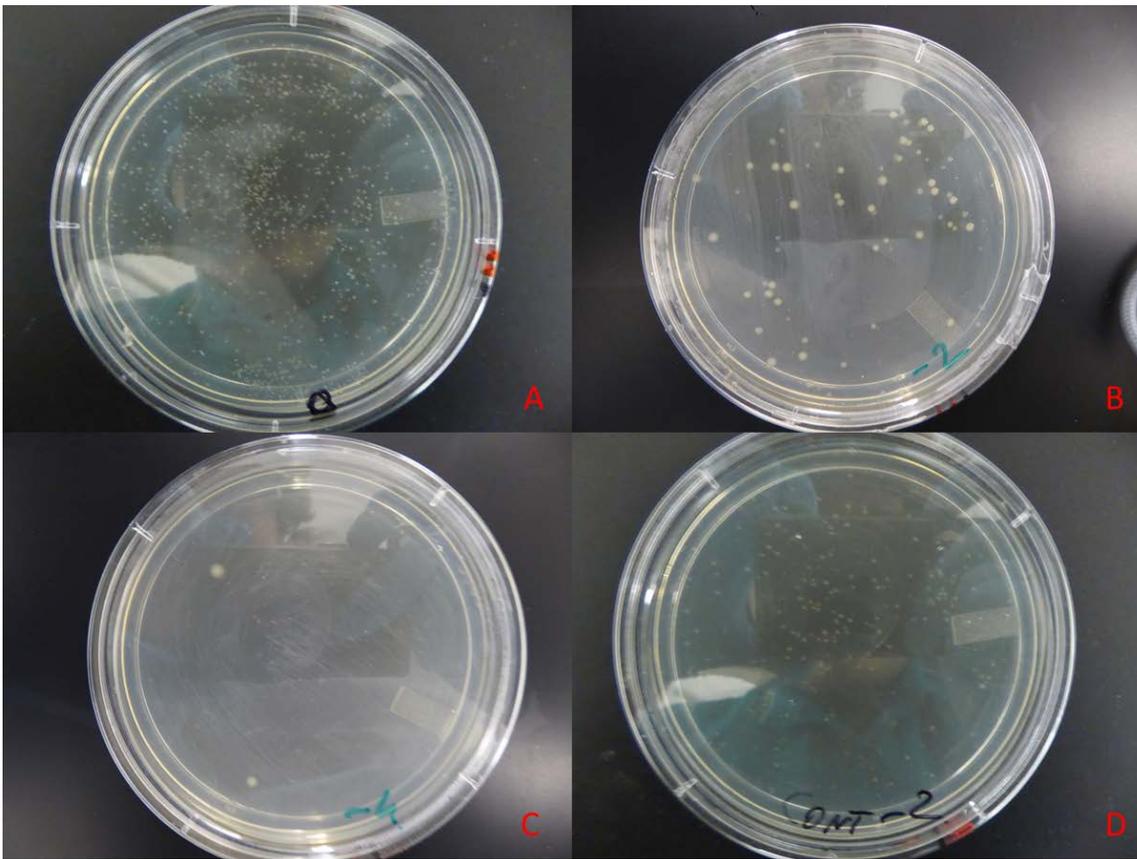
The ligated products were also run on an agarose gel to obtain a visual confirmation that the main reaction result was the pIT2 plus insert instead of other useless combinations like multiple joined vectors or multiple inserts together.



**Figure 12:** Aliquots from the same ligation mix were run on 0.8 % agarose gel to verify the formation of the desired products. The green arrowheads indicate two close bands between 5 and 6 kb, the largest of which represent the correct pIT2 + H5 insert product, while the other represents empty ligated pIT2. Cyan arrowheads indicate the presence of products made up of several joined H5 mutants. Far from obtaining 100% correct product formation, this gel suggests that the colonies surviving selection in carbenicillin (given by pIT2) may or may not express an H5 mutant.

Agar plates seeded after the DH5 $\alpha$  transformation were used to estimate the library. As shown in Figure 13, colonies grew even on the 10<sup>-4</sup> dilution plates. Taking the average between the duplicates, it can be safely assumed that the final library size is greater or equal than 10<sup>4</sup> individual mutants (Daugherty et al., 2000). Most of the 20 single colonies that

were expanded, minipreped and sequenced show a comparable plasmid size to the original H5. Sequencing results, assembled in Geneious 9.0 (Biomatters Ltd.) showed an average mutation rate of 0.358%, as compared to the rate of 0.66% from the source protocol (Cadwell & Joyce, 1992). While the adapted protocol used here appears to be less effective than its source, a lower mutation rate is not necessarily a negative for this application. Indeed, a larger average mutation rate could have led to several mutant clones that completely lose functionality, despite their  $V_H$  and  $V_L$  sequence, because of related amino acid mutations in the linker region or the portions conferring structure to the scFv



**Figure 13:** Pictures of transformed DH5 $\alpha$  serial dilutions on LB agar plates supplemented with 1% glucose and 100  $\mu\text{g}/\text{mL}$  carbenicillin, after 24 hours incubation at 30  $^{\circ}\text{C}$ . A. no dilution plate from pooled cell aliquots. B. 100-fold dilution plate. C. 10000-fold dilution plate, with two colonies. D. 100-fold dilution of control transformation with pUC19 plasmid. B and C were imaged after 36 hours for picture quality purposes.

## CONCLUSION

The codon optimized H5 was successfully used as a template for an ePCR reaction aimed to introduce point mutations in the scFv sequence. The mutated fragments were inserted in the original pIT2 vector then transformed in DH5 $\alpha$  to ensure plasmid stability over time. The library size is estimated to be greater or equal to 10<sup>4</sup> clones, each with an average error rate of 0.358%. The library is now ready to be screened through phage display in order to identify an improved version of the H5 probe, used to discern between strained and native states of Fn. Screening for a probe for another cryptic ECM protein site is not precluded either, given the stochastic nature of the point mutations in the library.

## CHAPTER 5

### CONCLUSION AND FUTURE DIRECTIONS

After the creation of the library with a diversity of  $10^4$ , each copy varying by 3 nucleotides on average (for an average rate of 0.358%), the next steps involve rounds of selection in order to distinguish at least 30 clones that show promise in binding the stretched Fn conformation fragment, 4G. Afterwards, clone candidates will be sequenced and their affinity and selectivity for the 4G fragment will be quantified. Here follow the detailed procedures for panning and affinity estimation.

Aliquots of BL21 DE3 pLysS library will be infected by the helper bacteriophage KM13, which is a trypsin sensitive derivative of M13 while growing at OD of in LB media. The resulting phages will be enriched by subsequent (usually three) rounds of screening against the target molecule, in this case the 9<sup>th</sup> and 10<sup>th</sup> type III repeats of Fn. The antigen can be biotinylated or physisorbed to the substrate (C. M. Y. Lee et al., 2007) used, whether an immuno-tube or simple tissue culture plastic. Given the reliability of the 4G fragment as a model for strained Fn, it will be used to screen phage clones, without need of biotinylation. After stringent washes, the unbound phages will be removed from subsequent screens, then the bound phages will be eluted with trypsin, collected and used to infect fresh E. coli. Finally, a blinded double round of ELISA will be used to identify the promising phage clones. One 96 well plate will be coated with the 4G fragment, while the second will be coated with 9\*10, which represents the same Fn type III repeats in their canonical configuration. The clones selected will be the ones showing qualitatively strong binding to 4G and weak to no binding on 9\*10.

The selected phages will be sequenced and their binding affinity will be quantified through SPR assays. Specifically, a self-assembled monolayer of alkanethiols comprised of 20% -COOH and 80% -OH functional groups will be assembled on a BIACORE gold sensor chip surface, by incubation of 1mM stock alkanethiol solution on the gold chip overnight at room temperature. SPR experiments will be performed on a BIACORE 2000 instrument. The two Fn fragments will be immobilized on the chip using EDC/sulfo-NHS chemistry, while several dilutions of the scFv fragments will be prepared in binding buffer. SPR sensograms will be processed with Scrubber software for double referencing, and curves will be fit to experimental data using a 2-site heterogeneous surface model using a third party software.

The clone with the largest improvement in binding affinity for Fn and binding specificity over the current H5 probe will be transformed for large expression into BL21 DE3 pLysS strain. The current values for H5 are 16 nM for the strained state, while the  $K_D$  for the relaxed state was 107 nM. Once a system for large scFv is established, competitive binding assays will be performed with soluble Fn fragments displaying only either the 9<sup>th</sup> (FnIII6- 9) or 10<sup>th</sup> (FnIII10- 14) type III repeats, which should indicate that the new H5 epitope is located within Fn 9III repeat as well (unpublished data).

Since the improvements achieved thus far would overcome most developmental hurdles for delivering and using the antibody for non-invasive imaging of tissue fibrosis, the data suggest that the Fn integrin binding mechano-switch is a pathophysiologically relevant signature (unpublished data). This system could then provide insights into mechanisms of tissue fibrosis.

Additionally, recent reports suggests that  $\alpha_v$  integrins on myofibroblasts are implicated for fibrogenesis in a broad range of fibrotic diseases (Henderson et al., 2013), and that pharmacological blockade of  $\alpha_v$  integrins ameliorates liver and lung fibrosis. Past standard cell attachment experiments to Fn-adsorbed substrates, which presented multiple conformations of Fn due to surface-mediated unfolding, have confirmed the previous H5 clone selectively inhibits Fn-integrin  $\alpha_v\beta_3$ , but not  $\alpha_5\beta_1$  interactions with Fn (unpublished data). Given that our past H5 antibody version demonstrated such effectiveness in vitro, the new H5 would be ready for in vivo studies, bar functionalization to improve circulation time and organ targeting.

Because of all of the above reasons, the new H5 shows even greater potential for not only imaging and diagnostic application in fibrosis, but also for targeting therapeutic moieties on Fn, thus becoming a prime candidate for future theranostic assays in animal models.

## APPENDIX A: H5 SEQUENCES

**Legend.** Cyan: linker region. Red: restriction sites for SphI and NotI respectively. Yellow: HIS-tag. Green: myc-tag. All capitalized nucleotides were used as input for the codon optimization routine described in Chapter 3 and reported in Appendix B.

<b>Recovered H5 sequence</b>
<p>ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTACTCGCGGCCAGCCGGCCATG                      GCCGAGGTGCAGCTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTCCCTGAGAC                      TCTCCTGTGCAGCCTCTGGATTACCTTTAGCAGCTATGCCATGAGCTGGGTCCGCCAGGCTC                      CAGGGAAGGGGCTGGAGTGGGTCTCAGATATTTATGATGGTGGTGTACAAATTACGCAGAC                      TCCGTGAAGGGCCGGTTCACCACCTCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATG                      AACAGCCTGAGAGCCGAGGACACGGCCGTATATACTGTGCGAAAACGCTGATAATTTTGAC                      TACTGGGGCCAGGGAACCCTGGTCACCGTCTCGAGCGGTGGAGGCGGTTTCAGGCGGAGGTG                      GCAGCGGCGGTGGCGGGTCCACGGACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCA                      TCTGTAGGAGACAGAGTCAACATCACTTGCCGGGCAAGTCAGAGCATTAGCAGCTATTTAAAT                      TGGTATCAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCACTTTGCAA                      AGTGGGGTCCCATCAAGGTTCAAGTGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGC                      AGTCTGCAACCTGAAGATTTTGCAACTTACTACTGTCAACAGGCTAATAGTGCTCCTACTACGT                      TCGGCCAAGGGACCAAGGTGGAATCAAACGGgcgccgacatcatcatcaccatcaggggcccagag                      acaaaaactcatctcagaagaggatctgaatggggcccatag</p>
<b>Codon optimized portion</b>
<p>ATGAAGTACCTGCTGCCGACCGCAGCCGCCGGTCTGCTGTTATTGGCAGCCCAGCCAGCCAT                      GGCCGAGGTTCAACTGCTGGAGAGCGGCGGCGGCCTCGTGCAACCAGGTGGTTCGCTGCGT                      CTGTCCTGTGCCGCTCCGTTTTACGTTACGACAGCTATGCCATGAGCTGGGTGCGTCAAGCC                      CCGGGCAAGGGTCTGGAGTGGGTTAGCGACATCTACGATGGTGGTGGTACCAATTATGCAGA                      CAGCGTTAAGGGCCGCTTCACGACGTCCCGTGACAATAGCAAAAACACCCTGTATTTGCAGAT                      GAATTCGCTGCGCGCCGAGGATACCGCGTCTATACTGCGCCAAGACGGCAGACAACCTCG                      ACTACTGGGGTCAGGGCACCCCTCGTTACGGTTAGCTCGGGCGGTGGCGGTAGCGGTGGTGG                      TGGCTCCGGTGGTGGCGGCTCACGGATATCCAGATGACGCAGTCGCCTAGCAGCCTGAGCG</p>

CAAGCGTGGGTGACCGTGTTACCATTACCTGTCGTGCCTCTCAATCGATCTCTTCGTATCTGAA  
TTGGTACCAGCAGAAGCCGGGTAAGGCACCGAAACTGCTGATTTACGCCGCCAGCACGCTGC  
AGAGCGGCGTCCGAGCCGTTTCAGCGGTTCTGGTAGCGGCACGGACTTCACTCTGACCATCA  
GCAGCCTGCAACCGGAGGATTTGCCACCTACTACTGCCAGCAGGCCAATAGCGCCCCGACCA  
CGTTTGGTCAAGGTACCAAGGTTGAAATTAAGCGTG

**gBlock used for ePCR**

ctatgacatgattacgccaagcttgc<sup>atgca</sup>aaattctatttcaaggagacagtcataATGAAGTACCTGCTGCCGA  
CCGCAGCCGCCGGTCTGCTGTTATTGGCAGCCCAGCCAGCCATGGCCGAGGTTCAACTGCTG  
GAGAGCGGCGGCGGCCTCGTGCAACCAGGTGGTTCGCTGCGTCTGTCTGTGCCGCCTCCGG  
TTTTACGTTACAGCAGCTATGCCATGAGCTGGGTGCGTCAAGCCCCGGGCAAGGGTCTGGAGT  
GGGTTAGCGACATCTACGATGGTGGTGGTACCAATTATGCAGACAGCGTTAAGGGCCGCTTC  
ACGACGTCCCGTGACAATAGCAAAAACACCCTGTATTTGCAGATGAATTCGCTGCGCGCCGAG  
GATACCGCCGTCTATTA<sup>CTGCGCCAAGACGGCAGACA</sup>ACTTCGACTACTGGGGTCAGGGCAC  
CCTCGTTACGGTTA<sup>GCTCGGGCGGTGGCGGTAGCGGTGGTGGTGGCTCCGGTGGTGGCGGC</sup>  
<sup>TC</sup>CACGGATATCCAGATGACGCAGTCGCCTAGCAGCCTGAGCGCAAGCGTGGGTGACCGTGT  
TACCATTACCTGTCGTGCCTCTCAATCGATCTCTTCGTATCTGAATTGGTACCAGCAGAAGCCG  
GGTAAGGCACCGAAACTGCTGATTTACGCCGCCAGCACGCTGCAGAGCGGCGTCCGAGCCG  
TTTCAGCGGTTCTGGTAGCGGCACGGACTTCACTCTGACCATCAGCAGCCTGCAACCGGAGG  
ATTTTGCCACCTACTACTGCCAGCAGGCCAATAGCGCCCCGACCACGTTTGGTCAAGGTACCA  
AGGTTGAAATTAAGCGT<sup>GCggccgc</sup><sup>atcatcatcaccatcac</sup>g



```
new_title=strcat(optimized,title);  
fastawrite(new_title, newseq);
```

```

function [dna] = optimscFv(aa)
%Function reads an amminoacid and returns EColi optimized codon

% Codons are chosen based on frequency table from Welch et al., PLOS
one 2009
% Various amminoacids are sorted through thanks to a case statement
% Random number generator is reseeded every time the function is
called.
% Fnc takes in a string and returns a three letter string

%Leandro Moretti, June 10th 2016

%seeding random number generator and getting value
rng('shuffle');
n=rand;

%Beginning switch statment, one case per amminoacid
switch aa
%%
%alanine
    case {'A', 'a'}
        if n<=0.24
            dna='GCA';
        else if n<=43
            dna='GCC';
        else if n<=55
            dna='GCT';
        else
            dna='GCG';
        end
    end
end
%%
%arginine
    case {'R', 'r'}
        if n<=0.03
            dna='AGA';
        else if n<=0.38
            dna='CGC';
        else
            dna='CGT';
        end
    end
end
%%
%asparagine
    case {'N', 'n'}
        if n<=0.47
            dna='AAT';
        else
            dna='AAC';
        end
    end
end
%%
%aspartic acid
    case {'D', 'd'}
        if n<=0.46

```

```

        dna='GAT';
    else
        dna='GAC';
    end

%%
%cysteine
    case {'C','c'}
        if n<=0.42
            dna='TGT';
        else
            dna='TGC';
        end

%%
%glutamine
    case {'Q','q'}
        if n<=0.45
            dna='CAA';
        else
            dna='CAG';
        end

%%
%glutamic acids
    case {'E','e'}
        if n<=0.43
            dna='GAA';
        else
            dna='GAG';
        end

%%
%glycine
    case {'G','g'}
        if n<=0.39
            dna='GGC';
        else
            dna='GGT';
        end

%%
%histidine
    case {'H','h'}
        if n<=0.38
            dna='CAT';
        else
            dna='CAC';
        end

%%
%isoleucine
    case {'I','i'}
        if n<=0.49
            dna='ATT';
        else

```

```

        dna='ATC';
    end

%%
%leucine
    case {'L','l'}
        if n<=0.03
            dna='CTC';
        else if n<=0.05
            dna='CTT';
        else if n<=0.08
            dna='TTA';
        else if n<=0.22
            dna='TTG';
        else
            dna='CTG';
        end
    end
end

%%
%lysine
    case {'K','k'}
        if n<=0.49
            dna='AAG';
        else
            dna='AAA';
        end
    end

%%
%methionine
    case {'M','m'}
        dna='ATG';
    end

%%
%phenylalanine
    case {'F','f'}
        if n<=0.45
            dna='TTT';
        else
            dna='TTC';
        end
    end

%%
%proline
    case {'P','p'}
        if n<=0.1
            dna='CCA';
        else if n<=0.19
            dna='CCT';
        else
            dna='CCG';
        end
    end
end

```

```

%%
%serine
    case {'S', 's'}
        if n<=0.02
            dna='TCA';
        else if n<=0.12
            dna='TCT';
        else if n<=0.17
            dna='TCG';
        else if n<=0.3
            dna='TCC';
        else
            dna='AGC';
        end
    end
end

%%
%threonine
    case {'T', 't'}
        if n<=0.1
            dna='ACT';
        else if n<=0.43
            dna='ACG';
        else
            dna='ACC';
        end
    end
end

%%
%tryptophane
    case {'W', 'w'}
        dna='TGG';
    end

%%
%tyrosine
    case {'Y', 'y'}
        if n<=0.42
            dna='TAT';
        else
            dna='TAC';
        end
    end

%%
%valine
    case {'V', 'v'}
        if n<=0.03
            dna='GTA';
        else if n<=0.31
            dna='GTC';
        else if n<=0.65
            dna='GTG';
        else
            dna='GTT';
        end
    end
end

```

```
                end
            end

%%
%stop codon
    case {'*'}
        dna='TAG';
        %always Amber codon as stop codon

%%
%end of switch statement
    otherwise
        fprintf('undefined amino acid letter/n try again');
        dna='XXX';
end

end

end
```

## REFERENCES

- Álvarez, D., Levine, M., & Rojas, M. (2015). Regenerative medicine in the treatment of idiopathic pulmonary fibrosis: current position. *Stem Cells and Cloning: Advances and Applications*, 8, 61–65. <http://doi.org/10.2147/SCCAA.S49801>
- Ampicillin vs\_ Carbenicillin.pdf. (n.d.). Retrieved from [http://cellgro.com/media/upload/file/productinfosheets/new/Ampicillin%20vs\\_%20Carbenicillin.pdf](http://cellgro.com/media/upload/file/productinfosheets/new/Ampicillin%20vs_%20Carbenicillin.pdf)
- Ascione, A., Flego, M., Zamboni, S., De Cinti, E., Dupuis, M. L., & Cianfriglia, M. (2005). Application of a synthetic phage antibody library (ETH-2) for the isolation of single chain fragment variable (scFv) human antibodies to the pathogenic isoform of the hamster prion protein (HaPrPsc). *Hybridoma* (2005), 24(3), 127–132. <http://doi.org/10.1089/hyb.2005.24.127>
- Bachman, H., Nicosia, J., Dysart, M., & Barker, T. H. (2015). Utilizing Fibronectin Integrin-Binding Specificity to Control Cellular Responses. *Advances in Wound Care*, 4(8), 501–511. <http://doi.org/10.1089/wound.2014.0621>
- Bennett, K. M., Afanador, M. D., Lal, C. V., Xu, H., Persad, E., Legan, S. K., ... Schwarz, M. A. (2013). Ephrin-B2 Reverse Signaling Increases  $\alpha 5 \beta 1$  Integrin-Mediated Fibronectin Deposition and Reduces Distal Lung Compliance. *American Journal of Respiratory Cell and Molecular Biology*, 49(4), 680–687. <http://doi.org/10.1165/rcmb.2013-0002OC>
- Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., & Kompala, D. S. (1990). Plasmid-encoded protein: The principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnology and Bioengineering*, 35(7), 668–681. <http://doi.org/10.1002/bit.260350704>
- Birnbaum, S., & Bailey, J. E. (1991). Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli. *Biotechnology and Bioengineering*, 37(8), 736–745. <http://doi.org/10.1002/bit.260370808>
- Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., ... Falkow, S. (1977). Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene*, 2(2), 95–113. [http://doi.org/10.1016/0378-1119\(77\)90000-2](http://doi.org/10.1016/0378-1119(77)90000-2)
- Brown, A. C., Fiore, V. F., Sulchek, T. A., & Barker, T. H. (2013). Physical and chemical microenvironmental cues orthogonally control the degree and duration of fibrosis-associated epithelial-to-mesenchymal transitions. *The Journal of Pathology*, 229(1), 25–35. <http://doi.org/10.1002/path.4114>
- Brown, A. C., Rowe, J. A., & Barker, T. H. (2010). Guiding Epithelial Cell Phenotypes with Engineered Integrin-Specific Recombinant Fibronectin Fragments. *Tissue Engineering Part A*, 17(1–2), 139–150. <http://doi.org/10.1089/ten.tea.2010.0199>
- Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *Genome Research*, 2(1), 28–33. <http://doi.org/10.1101/gr.2.1.28>
- Choi, J. H., & Lee, S. Y. (2004). Secretory and extracellular production of recombinant proteins using Escherichia coli. *Applied Microbiology and Biotechnology*, 64(5), 625–635. <http://doi.org/10.1007/s00253-004-1559-9>
- Clackson, T., Hoogenboom, H. R., Griffiths, A. D., & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature*, 352(6336), 624–628. <http://doi.org/10.1038/352624a0>

- Clark, J. R., & March, J. B. (2006). Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials. *Trends in Biotechnology*, 24(5), 212–218. <http://doi.org/10.1016/j.tibtech.2006.03.003>
- Daugherty, P. S., Chen, G., Iverson, B. L., & Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proceedings of the National Academy of Sciences*, 97(5), 2029–2034. <http://doi.org/10.1073/pnas.030527597>
- Dong, H., Nilsson, L., & Kurland, C. G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, 177(6), 1497–1504.
- Dooley, H., & Flajnik, M. F. (2005). Shark immunity bites back: affinity maturation and memory response in the nurse shark, *Ginglymostoma cirratum*. *European Journal of Immunology*, 35(3), 936–945. <http://doi.org/10.1002/eji.200425760>
- Dufour, S., Duband, J. L., Humphries, M. J., Obara, M., Yamada, K. M., & Thiery, J. P. (1988). Attachment, spreading and locomotion of avian neural crest cells are mediated by multiple adhesion sites on fibronectin molecules. *The EMBO Journal*, 7(9), 2661–2671.
- Dumon-Seignovert, L., Cariot, G., & Vuillard, L. (2004). The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*, 37(1), 203–206. <http://doi.org/10.1016/j.pep.2004.04.025>
- Elf, J., Nilsson, D., Tenson, T., & Ehrenberg, M. (2003). Selective Charging of tRNA Isoacceptors Explains Patterns of Codon Usage. *Science*, 300(5626), 1718–1722. <http://doi.org/10.1126/science.1083811>
- Ferenci, T., & Silhavy, T. J. (1987). Sequence information required for protein translocation from the cytoplasm. *Journal of Bacteriology*, 169(12), 5339–5342.
- Fernandes, A., & Vinga, S. (2016). Improving Protein Expression Prediction Using Extra Features and Ensemble Averaging. *PLoS One*, 11(3), e0150369.
- Fernandez-Gacio, A., Uguen, M., & Fastrez, J. (2003). Phage display as a tool for the directed evolution of enzymes. *Trends in Biotechnology*, 21(9), 408–414. [http://doi.org/10.1016/S0167-7799\(03\)00194-X](http://doi.org/10.1016/S0167-7799(03)00194-X)
- Frantz, C., Stewart, K. M., & Weaver, V. M. (2010). The extracellular matrix at a glance. *J Cell Sci*, 123(24), 4195–4200. <http://doi.org/10.1242/jcs.023820>
- Gao, M., Craig, D., Lequin, O., Campbell, I. D., Vogel, V., & Schulten, K. (2003). Structure and functional significance of mechanically unfolded fibronectin type III1 intermediates. *Proceedings of the National Academy of Sciences*, 100(25), 14784–14789. <http://doi.org/10.1073/pnas.2334390100>
- Genst, E. D., Handelberg, F., Meirhaeghe, A. V., Vynck, S., Loris, R., Wyns, L., & Muyldermans, S. (2004). Chemical Basis for the Affinity Maturation of a Camel Single Domain Antibody. *Journal of Biological Chemistry*, 279(51), 53593–53601. <http://doi.org/10.1074/jbc.M407843200>
- Gottesman, S. (1996). Proteases and Their Targets in *Escherichia coli*. *Annual Review of Genetics*, 30(1), 465–506. <http://doi.org/10.1146/annurev.genet.30.1.465>
- Grant, R. P., Spitzfaden, C., Altroff, H., Campbell, I. D., & Mardon, H. J. (1997). Structural requirements for biological activity of the ninth and tenth FIII domains of human fibronectin. *Journal of Biological Chemistry*, 272(10), 6159–6166.

- Grodberg, J., & Dunn, J. J. (1988). Grodberg, J. & Dunn, J.J. OmpT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* 170, 1245–1258. *Journal of Bacteriology*, 170(3), 1245–53.
- Grossman, T. H., Kawasaki, E. S., Punreddy, S. R., & Osburne, M. S. (1998). Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene*, 209(1–2), 95–103. [http://doi.org/10.1016/S0378-1119\(98\)00020-1](http://doi.org/10.1016/S0378-1119(98)00020-1)
- Hammers, C. M., & Stanley, J. R. (2014). Antibody Phage Display: Technique and Applications. *Journal of Investigative Dermatology*, 134(2), 1–5. <http://doi.org/10.1038/jid.2013.521>
- Henderson, N. C., Arnold, T. D., Katamura, Y., Giacomini, M. M., Rodriguez, J. D., McCarty, J. H., ... Sheppard, D. (2013). Targeting of  $\alpha$ v integrin identifies a core molecular pathway that regulates fibrosis in several organs. *Nature Medicine*, 19(12), 1617–1624. <http://doi.org/10.1038/nm.3282>
- Holliger, P., & Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnology*, 23(9), 1126–1136. <http://doi.org/10.1038/nbt1142>
- Humira. (2016). Humira (AbbVie Inc.): FDA Package Insert. Retrieved June 26, 2016, from <http://druginserts.com/lib/rx/meds/humira-2/>
- Jaenicke, R. (1991). Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry*, 30(13), 3147–3161. <http://doi.org/10.1021/bi00227a001>
- King Jr, T. E., Pardo, A., & Selman, M. (2011). Idiopathic pulmonary fibrosis. *The Lancet*, 378(9807), 1949–1961. [http://doi.org/10.1016/S0140-6736\(11\)60052-4](http://doi.org/10.1016/S0140-6736(11)60052-4)
- Korpimäki, T., Kurittu, J., & Karp, M. (2003). Surprisingly fast disappearance of  $\beta$ -lactam selection pressure in cultivation as detected with novel biosensing approaches. *Journal of Microbiological Methods*, 53(1), 37–42. [http://doi.org/10.1016/S0167-7012\(02\)00213-0](http://doi.org/10.1016/S0167-7012(02)00213-0)
- Krammer, A., Craig, D., Thomas, W. E., Schulten, K., & Vogel, V. (2002). A structural model for force regulated integrin binding to fibronectin's RGD-synergy site. *Matrix Biology*, 21(2), 139–147. [http://doi.org/10.1016/S0945-053X\(01\)00197-4](http://doi.org/10.1016/S0945-053X(01)00197-4)
- Krammer, A., Lu, H., Isralewitz, B., Schulten, K., & Vogel, V. (1999). Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. *Proceedings of the National Academy of Sciences*, 96(4), 1351–1356. <http://doi.org/10.1073/pnas.96.4.1351>
- Kreuter, M., Bonella, F., Wijzenbeek, M., Maher, T. M., Spagnolo, P., Kreuter, M., ... Spagnolo, P. (2015). Pharmacological Treatment of Idiopathic Pulmonary Fibrosis: Current Approaches, Unsolved Issues, and Future Perspectives, Pharmacological Treatment of Idiopathic Pulmonary Fibrosis: Current Approaches, Unsolved Issues, and Future Perspectives. *BioMed Research International*, *BioMed Research International*, 2015, 2015, e329481. <http://doi.org/10.1155/2015/329481>
- Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *Science*, 324(5924), 255–258.
- Lee, C., Kim, J., Shin, S. G., & Hwang, S. (2006). Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. *Journal of Biotechnology*, 123(3), 273–280. <http://doi.org/10.1016/j.jbiotec.2005.11.014>
- Lee, C. M. Y., Iorno, N., Sierro, F., & Christ, D. (2007). Selection of human antibody fragments by phage display. *Nature Protocols*, 2(11), 3001–3008. <http://doi.org/10.1038/nprot.2007.448>

- Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J., & Wilcox, G. (1987). Characterization of the *Erwinia carotovora* pelB gene and its product pectate lyase. *Journal of Bacteriology*, *169*(9), 4379–4383.
- Main, A. L., Harvey, T. S., Baron, M., Boyd, J., & Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions. *Cell*, *71*(4), 671–678. [http://doi.org/10.1016/0092-8674\(92\)90600-H](http://doi.org/10.1016/0092-8674(92)90600-H)
- Mardon, H. J., & Grant, K. E. (1994). The role of the ninth and tenth type III domains of human fibronectin in cell adhesion. *FEBS Letters*, *340*(3), 197–201. [http://doi.org/10.1016/0014-5793\(94\)80137-1](http://doi.org/10.1016/0014-5793(94)80137-1)
- Meltzer, E. B., & Noble, P. W. (2008). Idiopathic pulmonary fibrosis. *Orphanet Journal of Rare Diseases*, *3*, 8. <http://doi.org/10.1186/1750-1172-3-8>
- Mergulhão, F. J. M., Summers, D. K., & Monteiro, G. A. (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances*, *23*(3), 177–202. <http://doi.org/10.1016/j.biotechadv.2004.11.003>
- Minton, N. P. (1984). Improved plasmid vectors for the isolation of translational lac gene fusions. *Gene*, *31*(1–3), 269–273. [http://doi.org/10.1016/0378-1119\(84\)90220-8](http://doi.org/10.1016/0378-1119(84)90220-8)
- Mochitate, K., Pawelek, P., & Grinnell, F. (1991). Stress relaxation of contracted collagen gels: Disruption of actin filament bundles, release of cell surface fibronectin, and down-regulation of DNA and protein synthesis. *Experimental Cell Research*, *193*(1), 198–207. [http://doi.org/10.1016/0014-4827\(91\)90556-A](http://doi.org/10.1016/0014-4827(91)90556-A)
- Newcomb, J., Carlson, R., & Aldrich, S. (2007). Genome synthesis and design futures: implications for the US economy. *Bio Economic Research Associates*.
- Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics*, *16*(7), 379–394. <http://doi.org/10.1038/nrg3927>
- Pankov, R., & Yamada, K. M. (2002). Fibronectin at a glance. *Journal of Cell Science*, *115*(20), 3861–3863. <http://doi.org/10.1242/jcs.00059>
- Petrenko, V. A., & Vodyanoy, V. J. (2003). Phage display for detection of biological threat agents. *Journal of Microbiological Methods*, *53*(2), 253–262.
- Pierschbacher, M. D., & Ruoslahti, E. (1984). Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature*, *309*(5963), 30–33.
- Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Microbiotechnology, Ecotoxicology and Bioremediation*, *5*, 172. <http://doi.org/10.3389/fmicb.2014.00172>
- Roy, D. C., & Hocking, D. C. (2012). Recombinant Fibronectin Matrix Mimetics Specify Integrin Adhesion and Extracellular Matrix Assembly. *Tissue Engineering Part A*, *19*(3–4), 558–570. <http://doi.org/10.1089/ten.tea.2012.0257>
- Saida, F., Uzan, M., Odaert, B., & Bontems, F. (2006). Expression of Highly Toxic Genes in *E. coli*: Special Strategies and Genetic Tools. *Current Protein and Peptide Science*, *7*(1), 47–56.
- Sharp, P. M., & Li, W.-H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, *15*(3), 1281–1295. <http://doi.org/10.1093/nar/15.3.1281>

- Shokri, A., Sandén, A., & Larsson, G. (n.d.). Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. *Applied Microbiology and Biotechnology*, 60(6), 654–664. <http://doi.org/10.1007/s00253-002-1156-8>
- Skerra, A., & Pluckthun, A. (1988). Assembly of a Functional Immunoglobulin Fv Fragment in *Escherichia Coli*. *Science*, 240(4855), 1038.
- Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705), 1315–1317. <http://doi.org/10.1126/science.4001944>
- Solar, G. del, Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., & Díaz-Orejas, R. (1998). Replication and Control of Circular Bacterial Plasmids. *Microbiology and Molecular Biology Reviews*, 62(2), 434–464.
- Stabenfeldt, S. E., Brown, A. C., & Barker, T. H. (2010). Engineering ECM Complexity into Biomaterials for Directing Cell Fate. In K. Roy (Ed.), *Biomaterials as Stem Cell Niche* (pp. 1–18). Springer Berlin Heidelberg. Retrieved from [http://link.springer.com/chapter/10.1007/8415\\_2010\\_1](http://link.springer.com/chapter/10.1007/8415_2010_1)
- Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1), 207–234. <http://doi.org/10.1016/j.pep.2005.01.016>
- Tan, J. L., Tien, J., Pirone, D. M., Gray, D. S., Bhadriraju, K., & Chen, C. S. (2003). Cells lying on a bed of microneedles: An approach to isolate mechanical force. *Proceedings of the National Academy of Sciences*, 100(4), 1484–1489. <http://doi.org/10.1073/pnas.0235407100>
- Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72(2), 211–222. <http://doi.org/10.1007/s00253-006-0465-8>
- The Mechanical Hierarchies of Fibronectin Observed with Single-molecule AFM. (n.d.). Retrieved June 30, 2016, from <http://www.sciencedirect.com/science/article/pii/S0022283602003066>
- Verma, R., Boleti, E., & George, A. J. T. (1998). Antibody engineering: Comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of Immunological Methods*, 216(1–2), 165–181. [http://doi.org/10.1016/S0022-1759\(98\)00077-5](http://doi.org/10.1016/S0022-1759(98)00077-5)
- Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T., & Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature*, 341(6242), 544–546. <http://doi.org/10.1038/341544a0>
- Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., & Gustafsson, C. (2009). Design Parameters to Control Synthetic Gene Expression in *Escherichia coli*. *PLOS ONE*, 4(9), e7002. <http://doi.org/10.1371/journal.pone.0007002>
- You, C., & Percival Zhang, Y.-H. (2012). Easy preparation of a large-size random gene mutagenesis library in *Escherichia coli*. *Analytical Biochemistry*, 428(1), 7–12. <http://doi.org/10.1016/j.ab.2012.05.022>
- Zhang, Z., Morla, A. O., Vuori, K., Bauer, J. S., Juliano, R. L., & Ruoslahti, E. (1993). The alpha v beta 1 integrin functions as a fibronectin receptor but does not support fibronectin matrix assembly and cell migration on fibronectin. *The Journal of Cell Biology*, 122(1), 235–242. <http://doi.org/10.1083/jcb.122.1.235>