1-1-2013

Improving Protein Therapeutics Through Quantitative Molecular Engineering Approaches and A Cell-Based Oral Delivery Platform

Ting Wun Ng
University of Pennsylvania, daphnnee@gmail.com

Improving Protein Therapeutics Through Quantitative Molecular Engineering Approaches and A Cell-Based Oral Delivery Platform

Abstract

Proteins, with their ability to perform a variety of highly specific biological functions, have emerged as an important class of therapeutics. However, to fully harness their therapeutic potential, proteins often need to be optimized by molecular engineering; therapeutic efficacy can be improved by modulating protein properties such as binding affinity/specificity, half-life, bioavailability, and immunogenicity. In this work, we first present an introductory example in which a mechanistic mathematical model was used to improve target selection for directed evolution of an aglycosylated Fc domain of an antibody to enhance phagocytosis of tumor cells. Several aspects of directed evolution experimental methods were then optimized. A model-guided ligation strategy was developed to maximize ligation yield in DNA library construction, and this design tool is freely available through a web server. Streamlined protocols for mRNA display and ribosome display, which are powerful in vitro selection methods, were also created to allow robust selection of a variety of therapeutic proteins, including monomeric Fc domains, designed ankyrin repeat proteins, a single-chain insulin analog (SCI-57), and leptin. Anti-ICAM-1 scFv antibody fragments were also optimized for ribosome display by grafting complementarity determining regions onto a stable human framework. In addition to engineering the proteins themselves, effective delivery systems are essential for maximizing the therapeutic benefit of these proteins in a clinical setting. We therefore also developed a novel oral delivery platform based on the food-grade bacterium Lactococcus lactis. SCI-57, leptin, and SCI-57-leptin fusion proteins have been successfully secreted from this host in vitro and preliminary studies in a diabetic mouse model show reduced glucose levels after oral administration of L. lactis secreting SCI-57. We then further improved the secretion potential of this host through directed evolution of a L. lactis signal peptide. In summary, our studies have provided important advances to the field of protein engineering through the development of mechanistic mathematical models, streamlined experimental methodologies, and polypeptides with improved properties. This work has also opened up the possibility of systemic delivery of protein therapeutics using live microorganisms.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Bioengineering

First Advisor
Casim A. Sarkar

Keywords
mathematical model, oral protein delivery, protein engineering

Subject Categories
Biomedical

IMPROVING PROTEIN THERAPEUTICS THROUGH

QUANTITATIVE MOLECULAR ENGINEERING APPROACHES

AND A CELL-BASED ORAL DELIVERY PLATFORM

Daphne T. W. Ng

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

2013

_____

Casim A. Sarkar, Ph.D., Assistant Professor of Bioengineering

Supervisor of Dissertation

_____

Daniel A. Hammer, Ph.D., Professor of Bioengineering

Graduate Group Chairperson

Dissertation Committee:

Mark Goulian, Ph.D., Professor of Biology (Committee Chair)

Arjun Raj, Ph.D., Assistant Professor of Bioengineering

Andrew Tsourkas, Ph.D., Associate Professor of Bioengineering

IMPROVING PROTEIN THERAPEUTICS THROUGH QUANTITATIVE
MOLECULAR ENGINEERING APPROACHES AND A CELL-BASED ORAL
DELIVERY PLATFORM

COPYRIGHT

2013

Ting Wun Ng

To my parents

This work would not have been possible without their love, support and occasional reminders :)

ACKNOWLEDGMENTS

I feel truly blessed with many amazing people throughout my PhD journey, who have given me much needed support and encouragement in one form or the other. I am also truly grateful for the opportunity to explore my scientific curiosity in the intellectual environment provided by the University of Pennsylvania over the past few years.

First, I am most fortunate for having the opportunity to learn from my advisor and mentor, Dr. Casim Sarkar. I am extremely grateful for his patience and support as he witnessed my growth as a scientist. I find myself being continually inspired as his enthusiasm and passion in the subject matter is highly contagious. His encouragement on creativity and hard work has consistently pushed me to reach further than what I thought I could accomplish.

I am also indebted to my thesis committee, Dr. Mark Goulian, Dr. Andrew Tsourkas and Dr. Arjun Raj for their guidance and helpful suggestions throughout my dissertation. Their generous support and encouragement both during and outside of formal committee meetings are very much appreciated.

Some of the most interesting work would not have been possible without wonderful collaborators whom I had the pleasure to work with. I would like to thank Dr. George Georgiou, Sang Taek and Will for the opportunity in aglycosylated IgG modeling; Dr. Ali Naji and Ming for the help and expertise in diabetic mice experiments;

and Dr. Vladimir Muzykantov and Colin for the discussions regarding the anti-ICAM-1 scFv project.

My dissertation was truly enjoyable thanks to my incredible and talented colleagues in the Sarkar lab: Najaf, Pam, Santhosh and Ellen. I greatly appreciate their support, help and all the lively discussions. They have never failed to amaze me with their scientific vigor and generosity.

To all my friends and family who have walked alongside me, I am forever grateful for your support, inspiration and encouragement. Thank you for listening to my tears and laughter, and for sharing a part of your lives with me. To my parents, thank you for showering me with love and never-ending support. Although far away, your patience, support and faith have motivated me every day and shaped me throughout my academic career.

Last but not least, I must thank my best friend and my love, Hubert, for your company, support, encouragement, patience, love, and for keeping me happy and sane.

ABSTRACT

IMPROVING PROTEIN THERAPEUTICS THROUGH QUANTITATIVE MOLECULAR ENGINEERING APPROACHES AND A CELL-BASED ORAL DELIVERY PLATFORM

Daphne T. W. Ng

Casim A. Sarkar, Ph.D.

Proteins, with their ability to perform a variety of highly specific biological functions, have emerged as an important class of therapeutics. However, to fully harness their therapeutic potential, proteins often need to be optimized by molecular engineering; therapeutic efficacy can be improved by modulating protein properties such as binding affinity/specificity, half-life, bioavailability, and immunogenicity. In this work, we first present an introductory example in which a mechanistic mathematical model was used to improve target selection for directed evolution of an aglycosylated Fc domain of an antibody to enhance phagocytosis of tumor cells. Several aspects of directed evolution experimental methods were then optimized. A model-guided ligation strategy was developed to maximize ligation yield in DNA library construction, and this design tool is freely available through a web server. Streamlined protocols for mRNA display and ribosome display, which are powerful *in vitro* selection methods, were also created to allow robust selection of a variety of therapeutic proteins, including monomeric Fc domains, designed ankyrin repeat proteins, a single-chain insulin analog (SCI-57), and

leptin. Anti-ICAM-1 scFv antibody fragments were also optimized for ribosome display by grafting complementarity determining regions onto a stable human framework. In addition to engineering the proteins themselves, effective delivery systems are essential for maximizing the therapeutic benefit of these proteins in a clinical setting. We therefore also developed a novel oral delivery platform based on the food-grade bacterium *Lactococcus lactis*. SCI-57, leptin, and SCI-57-leptin fusion proteins have been successfully secreted from this host *in vitro* and preliminary studies in a diabetic mouse model show reduced glucose levels after oral administration of *L. lactis* secreting SCI-57. We then further improved the secretion potential of this host through directed evolution of a *L. lactis* signal peptide. In summary, our studies have provided important advances to the field of protein engineering through the development of mechanistic mathematical models, streamlined experimental methodologies, and polypeptides with improved properties. This work has also opened up the possibility of systemic delivery of protein therapeutics using live microorganisms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Protein therapeutics

Proteins, encoded by 25,000–40,000 different genes in the human genome[1,2], comprise one of the most diverse classes of biological macromolecules. They perform a variety of important biochemical tasks in the body, including catalysis and cellular signaling, and they also serve mechanical and structural roles. There is increasing interest in harnessing the potential of proteins to develop new therapeutics, which have traditionally been dominated by small molecules[3]. Protein therapeutics can offer a range of highly specific and complex functions that cannot be achieved by small molecules[4], thus providing novel treatments to alleviate diseases. Proteins with catalytic activity can replace or augment natural enzymes or they can provide entirely novel functions; proteins with specific binding properties can block the functions of their targets or they can be used to deliver a cytotoxic or effector payload to specific cells[4]. Today, there are over 200 approved protein therapeutics in the market[5], with many more promising ones in development. Among the approved therapeutics are blood factors, thrombolytics and anticoagulants, hormones, growth factors, interferons and interleukins, vaccines, and monoclonal antibody-based products to treat a variety of diseases such as hemophilia, myocardial infarction, diabetes, anemia, hepatitis, rheumatoid arthritis, and cancer[5].

Since the approval of the first recombinant protein therapeutic human insulin in 1982[6], advances in recombinant protein technology have largely fueled the popularity of protein therapeutics in the past few decades. It is now possible to circumvent the inefficient and costly purification of proteins from natural sources and to produce highly pure proteins efficiently and in large quantities in the laboratory. Recombinant proteins can be produced from a variety of host systems including bacteria, yeast, insect cells, mammalian cells, and transgenic animals and plants, each of which offers differing advantages in production cost and processing (e.g. post-translational modifications)[4]. Perhaps more importantly, another major advantage that recombinant protein technology offers is the ability to modify or create novel protein properties[7], which is often necessary in the development and optimization of protein therapeutics for specific clinical applications.

1.2 Improving protein therapeutics through molecular engineering

Paralleling the increasing popularity of protein-based therapeutics, the field of protein engineering has rapidly emerged since the 1980s[7]. Protein engineering has greatly expanded the repertoire of proteins available for therapeutic purposes. The majority of protein drugs consist of monoclonal antibodies[8], in which the Fab region enables high-affinity target binding and the Fc region modulates immunological effector functions and significantly extends circulatory half-life *in vivo*[9]. Monoclonal antibodies have been traditionally produced from hybridomas, but recombinant DNA technology now allows

chimerization and humanization of monoclonal antibodies to reduce immunogenicity and also enables display libraries to be constructed and screened in order to further improve affinities and specificities[8]. In addition, natural ligands and receptors represent another important class of protein therapeutics; these proteins are utilized for their innate ability to interact with high affinity and specificity with targets in the body, but they are often further engineered to create custom therapeutics with optimized properties[10]. A new generation of potential therapeutics includes engineered protein scaffolds, which are typically derived from stable, soluble, and easily expressed monomeric proteins that can overcome the limitations posed by antibodies or natural proteins[11]. More than 50 such alternative protein scaffolds have been proposed, including Affibodies[12], Monobodies[13], and designed ankyrin repeat proteins (DARPins)[14], most of which have been engineered to interact with different molecules of interest.

Additional engineering efforts are motivated by limitations in current protein therapeutics, including their high production costs, large sizes with limited tissue penetration and localization, inability to cross epithelial and endothelial structures, and rapid degradation by proteases and clearance from the body[7]. Drug potency can be improved by increasing affinity and specificity for protein binders or catalytic activity for enzymes and also by augmenting half-life and bioavailability by reducing clearance and decreasing degradation[9]. In addition, immunogenicity can be alleviated by "humanizing" drugs[15], and production costs can be reduced in some cases by eliminating glycosylation and/or disulfide bonds to allow production in lower organisms and simplify downstream

purification. Furthermore, the host itself can be engineered to increase protein expression levels.

Protein engineering efforts have relied on two complementary approaches, rational design and directed evolution. In rational design, detailed knowledge of protein structure and function are used to "design" new proteins or to improve the properties of existing proteins. Computational algorithms are often used to aid rational design, by sampling the amino acid sequence in protein structure-function space. The "design cycle" often includes iterating between molecular modeling and experimental construction and analysis of the designed protein until the desired properties are achieved[16]. This hypothesis-driven manipulation of proteins is direct and inexpensive, and it has achieved success when structural knowledge of the protein is available[17]. However, the need for detailed structural information and computational complexity in predicting the combinatorial effects of multiple mutations has prevented the rational design of many interesting proteins. Directed evolution, on the other hand, works by mimicking the natural evolutionary process involving repeated cycles of mutation and selection in order to evolve molecules with desired properties. This method is indirect and more labor-intensive, and the combinatorial nature of the sequence space can give an astronomically large number of possible polypeptide sequences[18]. However, given that minimal structural information is required and powerful screening and selection techniques are now available, directed evolution has become a popular and effective protein engineering approach.

1.3 Directed evolution

Directed evolution consists of iterative cycles of diversity generation and selection to evolve proteins with new and improved properties (Fig. 1.1). The starting template is usually chosen based on its "proximity" to the desired function[19], and a library of mutants is constructed "around" the parent molecule to sample the surrounding sequence space. Alternatively, a naïve synthetic library built on a stable scaffold can be used as the starting material. To generate molecular diversity, different randomization methods can be employed. Error-prone PCR is frequently used to introduce random point mutations, as point mutations can often act alone or synergistically to give rise to new, unexpected functions. Taq polymerase with mutagenic buffer[20,21] or nucleotide triphosphate analogs[22], or other error-prone DNA polymerases can be used. This method can sample a broad spectrum of sequence space, but sampling cannot be done exhaustively as the number of possible mutations far exceeds the upper limit of any current screening or selection capabilities. Saturation mutagenesis, on the other hand, can efficiently sample a highly focused region of sequence space, often a "hot spot" which is typically around the active site of an enzyme or binding site of a protein. A saturation mutagenesis library is usually constructed using degenerate oligonucleotides[23] or primers synthesized with trinucleotide phosphoramidites[24]. A third method, DNA shuffling, allows *in vitro* homologous recombination of selected mutant genes[25,26]. This method can be used together with error-prone PCR and/or saturation mutagenesis to recombine beneficial mutations from different parents and allow the further separation of beneficial mutational combinations from neutral or deleterious ones.

5
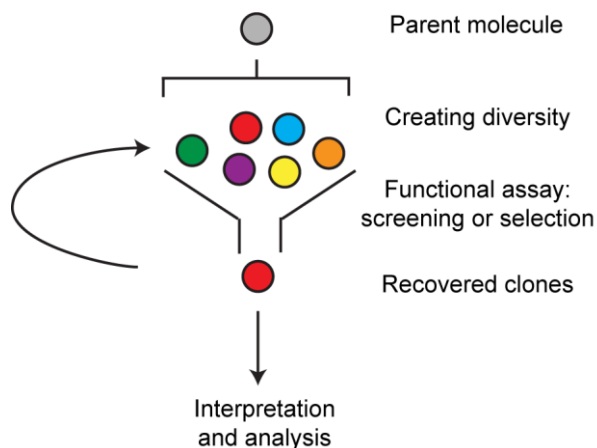
Figure 1.1. Schematic overview of directed evolution. Directed evolution starts with a parent molecule with functions to be optimized, or a naïve molecule for evolving new properties. Libraries of new molecules are generated by incorporating mutations at various sites, which are then subjected to functional assays to screen or select for mutants with desired properties. Central to the functional assays in all directed evolution methodologies is the ability to couple phenotype (selected property) to genotype (nucleotide sequence) in order to determine the sequences of improved variants. The best clones are used for the next iteration of mutation and selection or they are subjected to analysis once the engineering objective is met[19].

With the library of mutants in hand, a functional assay is needed to find the subset of mutants that contain the properties of interest. Screening methodologies involve direct functional evaluation of each individual clone, usually with low to medium throughput (10^2 to 10^6 clones/round). Screening is straightforward and can utilize almost any biochemical or biophysical detection method[27]. Selection methods, on the other hand, allow much higher numbers of variants (10^6 to $> 10^{12}$ clones/round) to be filtered, but carefully designed assays are needed to select, retain, and amplify clones with the desired properties while minimizing false positives arising from the indirect nature of most

selections. Nevertheless, the advantage of sampling a much larger fraction of sequence space might justify the efforts in setting up such selections[27].

In the case of directed protein evolution, the available methodologies ensure physical linkage of the displayed phenotype (desired protein property) and the genotype (cDNA or mRNA encoding the variant) to allow recovery and identification of mutant sequences after screening or selection. Depending on the protein of interest and the properties to be engineered, *in vivo* display techniques (e.g., phage display[28] and yeast surface display[29]) or *in vitro* display techniques (e.g., mRNA display[30] and ribosome display[31]) can be employed. We focused our investigation on two powerful *in vitro* display technologies, ribosome display and mRNA display, as their cell-free nature enables selections on the largest possible libraries to date (up to 10^{14} members). Ribosome display involves *in vitro* translation of mRNA sequences lacking stop codons to generate ternary complexes in which each mRNA and its corresponding protein are noncovalently associated *via* a stalled ribosome[32]. mRNA display uses an mRNA-DNA-puromycin fusion as its template for *in vitro* translation, which results in a covalent puromycin linkage between the mRNA and the corresponding nascent peptide[33]. Unlike *in vivo* display methods, which rely on a transformation step to incorporate the library into display formats[34], ribosome display and mRNA display are performed entirely *in vitro*, so the theoretical limit to library size is the amount of translation mixture that can be physically accommodated in a tube. Another advantage is that *in vitro* display allows easy introduction of additional diversity between generations because the PCR step, often used to add diversity, is inherent to these methods[35].

Finally, in all directed evolution methods, screened or selected clones with desired properties are obtained and their sequences are analyzed to identify beneficial mutations. These variants can either be used as the starting template for the next round of evolution or further analyzed once the engineering objective is fulfilled[19].

1.4 The role of mathematical modeling in directed evolution

The success of directed evolution experiments is directly impacted by the ability to construct and propagate a high-quality library of variants. Since each library member represents a unique combination of properties, any loss due to insufficient sampling or inefficient experimental procedures could be detrimental to the identification of improved variants through a series of screenings or selections. Furthermore, each round of directed evolution is extremely labor intensive, so the higher the enrichment ratio per round, the faster the engineering objective can be fulfilled. Once clones with beneficial properties have been identified, performing additional experiments to trace the source of these properties to specific mutations can provide tremendous mechanistic insight into the sequence-structure-function relationship of the protein and aid further engineering efforts. Due to the combinatorial nature of mutants and complexity of selection assays, statistical and quantitative approaches are very useful in guiding and optimizing directed evolution experiments and in evaluating the multitude of mutational effects both during and after the selection process.

Various computational tools have enabled construction of highly guided libraries[36], and statistical analyses have allowed estimation of the completeness and diversity of library assembled using various library construction methods[37,38]. Not only can quantitative approaches allow higher quality libraries to be constructed to maximize variants containing beneficial mutations, but knowledge of the frequency and location of mutations within the ensemble of variants can help guide sampling and selection procedures[39–46]. Mathematical models have also been developed to optimize screening and selection parameters in order to maximize enrichment of desired properties[47,48]. Depending on the choice of assay, the optimal experimental conditions are often non-intuitive, and conventional wisdom can misguide the selection efforts[48]. In many cases, the optimal experimental parameters can be identified through simple mathematical analyses without the need for extensive experimental trial-and-error, which can then be incorporated into suitable screening and selection assays[47–49]. In addition, data-driven models that employ multivariate statistical techniques to model protein sequence-activity relationships (ProSAR) can guide the evolutionary process[50]. ProSAR classifies individual mutations as beneficial, neutral, or deleterious, which informs the design of subsequent libraries[51–56]. Finally, mechanistic models have been used to relate biochemical properties to desired protein functions to explain experimental observations and provide insights into future engineering efforts[57–60]. Molecular-level kinetic models can also be incorporated into systems-level pharmacokinetic/pharmacodynamic models to investigate physiological effects arising from changes in molecular properties, providing a more holistic approach in the development of new protein therapeutics[61].

1.5 The significance and bottlenecks for oral delivery

Although the tools for protein engineering have grown at a rapid rate, the methods for delivering protein drugs have remained similar despite much research. For decades, proteins have been administered by subcutaneous, intramuscular, or intravenous injections[62]. These invasive methods lead to poor patient compliance and complications at injection sites, such as pain, irritation, infection, and tissue damage[63]. While many life-saving drugs have been administered through injections, other potential therapeutics for a variety of chronic or non-lethal diseases are not entering the clinic due to barriers in administration and dosage requirement. Depending on the injection type, medical personnel might need to administer the drug, which becomes impractical for short-acting therapeutics that require multiple injections per day. Injection also brings about a social stigma and poor patient acceptance, and missed doses can have deleterious health outcomes[62]. Thus, non-invasive approaches such as pulmonary, transdermal, or oral delivery are more preferable. Oral delivery, in particular, has been the route of choice for the majority of small-molecule drugs. This mode of administration is easy to use, widely accepted, and can prevent the problems and side effects associated with injections[63]. Patient compliance is also significantly improved, which often leads to improved health outcomes[64], especially for chronic illnesses. Moreover, for some proteins (e.g., insulin), oral delivery offers an additional advantage of more accurately recapitulating normal physiology. After absorption in the intestine, orally delivered insulin reaches the portal system, more closely approximating what occurs in a non-diabetic individual[65].

10

Despite the many advantages of oral delivery, this route of administration is not yet feasible for most proteins due to low bioavailability[66]. The first reason is poor stability due to low pH of the gastric fluid and enzymatic degradation at multiple sites along the gastrointestinal tract[67]. Protein drugs administered orally are subjected to pre-systemic degradation, first encountering pepsin in the stomach, then pancreatic proteases such as trypsin, α-chymotrypsin, and elastase in the intestinal lumen, and proteolytic enzymes associated with enterocytes[68]. The second reason for low bioavailability is due to poor penetration of the intestinal wall to reach systemic circulation[67]. The intestine is lined with closely packed epithelial cells, which form tight junctions that keep out most macromolecules (e.g., proteins) to prevent absorption of unwanted and potentially harmful molecules[68] (Fig. 1.2). The transcellular route of absorption, which requires large hydrophilic therapeutic proteins to cross by passive diffusion, is extremely difficult. The alternative, paracellular route through tight junctions, is also not feasible for molecules with the physical dimensions of proteins[69]. By comparison, small-molecule drugs, usually with molecular weight less than 500 Da, can engage both pathways[68].
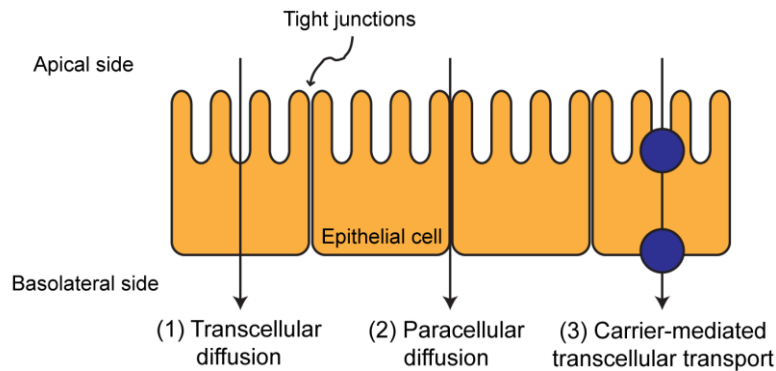
Figure 1.2. Pathways of intestinal drug absorption. The intestine is tightly packed with intestinal epithelial cells, which form tight junctions. (1) Transcellular diffusion involves diffusion through both the apical and

basolateral plasma membranes of the epithelial cells, which is difficult for large, highly charged molecules. (2) Paracellular diffusion involves diffusion through the aqueous intercellular space between adjacent epitheial cells. However, a group of interacting transmembrane and cytosolic proteins (tight junctions) is selectively permeable to only small, hydrophilic molecules. (3) Carrier-mediated transcellular transport actively transports specifically recognized molecules across the epithelial cells.

To increase oral bioavailability, various strategies have been developed. Approaches involving permeation enhancers or protease inhibitors as additives could be effective, but their long-term use remains questionable because of possible absorption of unwanted molecules and disturbance of the digestion of nutritive proteins[70]. Another strategy involves encapsulating protein drugs in microparticles, nanoparticles, or liposomes to protect the therapeutics from enzymatic digestion. However, there are concerns regarding the stability of proteins during the processing, release, and storage of these formulations as well as the fate of synthetic particles[70]. It is also necessary to consider the use of a delivery carrier system to cross the intestinal barrier. A class of cell-penetrating peptides, such as TAT, penetratin, and oligoarginine, have been used to internalize different biomolecules into cells[71–73]. But such peptide-cell interactions are non-specific and may cause undesired toxicity[74]. Harnessing endogenous intestinal transport systems, a potentially safer and more practical strategy, has also been used to increase intestinal absorption of peptides and proteins. For example, the transferrin receptor and vitamin B_{12} receptor have been utilized to transport transferrin- and vitamin B_{12}-conjugated drugs respectively[75,76]; however, such approaches might interfere with normal absorption of these nutrients. Therefore, further improvements in oral delivery

strategies are necessary to elicit maximum benefit from the current and next generations of protein therapeutics.

1.6 Significance of research

Development of new protein therapeutics requires parallel efforts in many different areas, and this work has contributed in three major directions. First, our molecular engineering efforts have brought about important improvements in display technologies and in the properties of several different classes of proteins that are commonly used as therapeutic modalities. Two central streamlined display platforms, mRNA display and ribosome display, have been developed for monoclonal antibodies and Fc domains, natural ligands, and alternative scaffolds such as DARPins. Many important protein properties were examined, including aglycosylation to remove glycan heterogeneity and enable facile engineering of novel effector functions, disulfide bond removal for easier protein production, fusion to the Fc domain of an antibody to increase half-life and bioavailability, antibody humanization to reduce immunogenicity, utilization of single-chain analogs of heterodimeric proteins for increased stability, and engineered signal peptides for improved protein secretion. Two of the engineering strategies resulted in improvements over existing molecules: (1) grafting the complementarity determining regions (CDRs) of an anti-ICAM-1 single-chain variable fragment (scFv) into a stable framework improved its biophysical properties and (2) directed evolution of a signal peptide enhanced protein secretion from *Lactococcus lactis*.

Second, our mathematical modeling efforts can be used in a broad spectrum of applications to guide directed evolution experiments. For library construction, an experimentally validated, easily parameterized thermodynamic model was developed to optimize ligation reactions and provide general insights. A free, web-based ligation calculator was subsequently developed to allow facile and customized optimization of ligation reactions by the scientific community. In addition, mechanistic models were developed to explain and guide two different directed evolution experiments. The relationship between mRNA folding and translation/secretion efficiencies was quantitatively explained in the first model, which was successfully used to explain and quantify the effects of silent mutations in signal peptides. Our second model provides mechanistic insight into how IgG Fc receptor affinity relates to effector functions in phagocytic cells. The identification of a dominant allele that has largely been overlooked in triggering phagocytosis will be vital to future Fc engineering efforts.

Third, our novel cell-based delivery platform using *L. lactis* offers an alternative approach in addressing some of the bottlenecks in oral delivery. This food-grade bacterium was successfully used to secrete bioactive proteins both *in vitro* and *in vivo*. One of the major hurdles in oral delivery of proteins, proteolytic degradation, can be attenuated since recombinant protein is locally produced by the bacteria in the intestine. Ongoing work to conjugate protein therapeutics to a viable transcytosis partner may overcome the other major problem of poor absorption. This work opens up the possibility of systemic delivery of protein therapeutics via oral administration using live microorganisms.

1.7 Chapter layout

We start by presenting in Chapter 2 an introductory example of an application in which directed evolution was used to engineer a useful protein therapeutic. The Fc domain of an IgG antibody was aglycosylated and evolved to achieve increased affinity to activating Fcγ receptors for an enhanced immune response. We developed a mathematical model to help explain how receptor affinities and contradictory signaling relate to improved antibody dependent cell-mediated phagocytosis (ADCP). We have also provided guidelines for future engineering of Fc domains with enhanced effector functions by identifying a new target allele for improving ADCP.

The following three chapters describe our optimizations and selections in various directed evolution applications. In Chapter 3, initial library construction for generation of diversity in directed evolution experiments was examined, and a model-guided ligation strategy is shown to be beneficial in optimizing assembly of DNA libraries. We demonstrated that display vectors should be redesigned to utilize Type IIS restriction sites to improve ligation yield, and we developed an easily parameterized thermodynamic model and accompanying web-based ligation calculator to predict product distributions. General insights and guidelines were provided for optimizing ligation reactions for both *in vivo* and *in vitro* display methodologies. A high-quality library would benefit the most from an effective selection procedure, and we present a streamlined mRNA display protocol in Chapter 4. We tested the procedure by performing affinity selection against

Her2 using binary DARPin libraries. An enrichment factor of up to two orders of magnitude per selection round was achieved. Display of monomeric IgG Fc domains using this new protocol is also discussed. In Chapter 5, other scaffold and therapeutic proteins, such as ankyrin repeat proteins, a single-chain insulin analog, leptin, and scFv against ICAM-1 were displayed using an updated version of ribosome display. CDR grafting into stable human frameworks, which is traditionally used to humanize antibodies to reduce immunogenicity for therapeutic purposes, was shown to be an effective strategy for engineering antibodies with higher thermodynamic stability and more efficient folding.

Although advances in biomolecular engineering are critical to improving the potencies of therapeutic proteins, an effective oral delivery system is also necessary to maximize the therapeutic benefit for proteins such as insulin. We propose a novel oral delivery platform utilizing the food-grade bacterium *L. lactis* in Chapters 6 and 7. In Chapter 6, *L. lactis* was used to successfully secrete a single-chain insulin, leptin, and insulin-leptin fusion proteins. Preliminary *in vivo* experiments in which insulin-secreting bacteria have been orally delivered into diabetic mice also show promising results. In Chapter 7, we further exploit the secretion capability of *L. lactis* by performing directed evolution on its major signal peptide. Saturation mutagenesis was used to generate diversity and the subsequent signal peptides were screened for enhanced secretion. Engineered signal peptides were identified that can replace wild-type signal peptides in many biotechnological and clinical applications of *L. lactis*.

1.8 References

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

2. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).

3. Pavlou, A. K. & Reichert, J. M. Recombinant protein therapeutics—success rates, market trends and values to 2010. *Nat. Biotechnol.* **22**, 1513–1519 (2004).

4. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7**, 21–39 (2008).

5. Walsh, G. Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* **28**, 917–924 (2010).

6. Goeddel, D. V. *et al.* Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 106–110 (1979).

7. Carter, P. J. Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp. Cell Res.* **317**, 1261–1269 (2011).

8. Carter, P. J. Potent antibody therapeutics by design. *Nat. Rev. Immunol.* **6**, 343–357 (2006).

9. Caravella, J. & Lugovskoy, A. Design of next-generation protein therapeutics. *Curr. Opin. Chem. Biol.* **14**, 520–528 (2010).

10. Kariolis, M. S., Kapur, S. & Cochran, J. R. Beyond antibodies: using biological principles to guide the development of next-generation protein therapeutics. *Curr. Opin. Biotechnol.* **24**, 1–6 (2013). doi:10.1016/j.copbio.2013.03. 017

11. Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* **13**, 245–255 (2009).

12. Nygren, P.-A. Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.* **275**, 2668–2676 (2008).

13. Koide, A. & Koide, S. Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. *Methods Mol. Biol.* **352**, 95–109 (2007).

14.   Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPins: a new generation of protein therapeutics. *Drug Discov. Today* **13**, 695–701 (2008).

15.   Chirino, A. J., Ary, M. L. & Marshall, S. A. Minimizing the immunogenicity of protein therapeutics. *Drug Discov. Today* **9**, 82–90 (2004).

16.   Hellinga, H. Rational protein design: combining theory and experiment. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10015–10017 (1997).

17.   Marshall, S. A., Lazar, G. A., Chirino, A. J. & Desjarlais, J. R. Rational design and engineering of therapeutic proteins. *Drug Discov. Today* **8**, 212–221 (2003).

18.   Jäckel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **37**, 153–173 (2008).

19.   Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

20.   Leung, D. W., Chen, E. & Goeddel, D. V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**, 11–15 (1989).

21.   Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33 (1992).

22.   Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **255**, 589–603 (1996).

23.   Telenius, H. *et al.* Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725 (1992).

24.   Virnekäs, B. *et al.* Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* **22**, 5600–5607 (1994).

25.   Stemmer, W. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).

26.   Zhao, H., Giver, L., Shao, Z., Affholter, J. & Arnold, F. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).

27. Goldsmith, M. & Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **22**, 406–412 (2012).

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

29. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

30. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297–12302 (1997).

31. Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4937–4942 (1997).

32. Dreier, B. & Plückthun, A. Ribosome Display: a technology for selecting and evolving proteins from large libraries. *Methods Mol. Biol.* **687**, 283–306 (2011).

33. Keefe, A. D. Protein selection using mRNA display. *Curr. Protoc. Mol. Biol.*, 24.5.1–24.5.34 (2001).

34. Sergeeva, A., Kolonin, M. G., Molldrem, J. J., Pasqualini, R. & Arap, W. Display technologies: application for the discovery of drug and gene delivery agents. *Adv. Drug Deliv. Rev.* **58**, 1622–1654 (2006).

35. Lipovsek, D. & Plückthun, A. *In-vitro* protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67 (2004).

36. Patrick, W. M. & Firth, A. E. Strategies and computational tools for improving randomized protein libraries. *Biomol. Eng.* **22**, 105–112 (2005).

37. Patrick, W. M., Firth, A. E. & Blackburn, J. M. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng. Des. Sel.* **16**, 451–457 (2003).

38. Bosley, A. D. & Ostermeier, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng.* **22**, 57–61 (2005).

39. Wong, T. S., Roccatano, D., Zacharias, M. & Schwaneberg, U. A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* **355**, 858–871 (2006).

40. Rasila, T. S., Pajunen, M. I. & Savilahti, H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal. Biochem.* **388**, 71–80 (2009).

41. Wang, D., Zhao, C., Cheng, R. & Sun, F. Estimation of the mutation rate during error-prone polymerase chain reaction. *J. Comput. Biol.* **7**, 143–158 (2000).

42. Moore, G. L. & Maranas, C. D. Modeling DNA mutation and recombination for directed evolution experiments. *J. Theor. Biol.* **205**, 483–503 (2000).

43. Sun, F. Modeling DNA shuffling. *J. Comput. Biol.* **6**, 77–90 (1999).

44. Moore, G. L., Maranas, C. D., Lutz, S. & Benkovic, S. J. Predicting crossover generation in DNA shuffling. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3226–3231 (2001).

45. Moore, G. L. & Maranas, C. D. Predicting out-of-sequence reassembly in DNA shuffling. *J. Theor. Biol.* **219**, 9–17 (2002).

46. Maheshri, N. & Schaffer, D. V. Computational and experimental analysis of DNA shuffling. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3071–3076 (2003).

47. Boder, E. T. & Wittrup, K. D. Optimal screening of surface-displayed polypeptide libraries. *Biotechnol. Prog.* **14**, 55–62 (1998).

48. Zahnd, C., Sarkar, C. A. & Plückthun, A. Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution. *Protein Eng. Des. Sel.* **23**, 175–184 (2010).

49. Wittrup, K. D. Protein engineering by cell-surface display. *Curr. Opin. Biotechnol.* **12**, 395–399 (2001).

50. Fox, R. J. & Huisman, G. W. Enzyme optimization: moving from blind evolution to statistical exploration of sequence-function space. *Trends Biotechnol.* **26**, 132–138 (2008).

51. Larsson, A.-K., Emrén, L. O., Bardsley, W. G. & Mannervik, B. Directed enzyme evolution guided by multidimensional analysis of substrate-activity space. *Protein Eng. Des. Sel.* **17**, 49–55 (2004).

52. Fox, R. J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338–344 (2007).

53. Reddy, S. T. *et al.* Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat. Biotechnol.* **28**, 965–969 (2010).

54. Liao, J. *et al.* Engineering proteinase K using machine learning and synthetic genes. *BMC Biotechnol.* **7**:16 (2007).

55. Li, Y. *et al.* A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* **25**, 1051–1056 (2007).

56. Barak, Y., Nov, Y., Ackerley, D. F. & Matin, A. Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J.* **2**, 171–179 (2008).

57. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).

58. Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465–471 (2009).

59. Walkiewicz, K. *et al.* Small changes in enzyme function can lead to surprisingly large fitness effects during adaptive evolution of antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21408–21413 (2012).

60. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592 (2005).

61. Rao, B. M., Lauffenburger, D. A. & Wittrup, K. D. Integrating cell-level kinetic modeling into the design of engineered protein therapeutics. *Nat. Biotechnol.* **23**, 191–194 (2005).

62. Cleland, J. L., Daugherty, A. & Mrsny, R. Emerging protein delivery methods. *Curr. Opin. Biotechnol.* **12**, 212–219 (2001).

63. Jain, K. K. Drug Delivery Systems - An Overview. *Drug Delivery Systems* 1–50 (2008).

64. Roter, D., Hall, J., Merisca, R. & Nordstrom, B. Effectiveness of interventions to improve patient compliance: a meta-analysis. *Med. Care* **36**, 1138–1161 (1998).

65. Gordon Still, J. Development of oral insulin: progress and current status. *Diabetes Metab. Res. Rev.* **18**, S29–S37 (2002).

66. Mahato, R. I., Narang, A. S., Thoma, L. & Miller, D. D. Emerging trends in oral delivery of peptide and protein drugs. *Crit. Rev. Ther. Drug Carrier Syst.* **20**, 153–214 (2003).

67. Morishita, M. & Peppas, N. A. Is the oral route possible for peptide and protein drug delivery? *Drug Discov. Today* **11**, 905–910 (2006).

68. Hamman, J. H., Enslin, G. M. & Kotzé, A. F. Oral delivery of peptide drugs: barriers and developments. *BioDrugs* **19**, 165–177 (2005).

69. Muller, G. Oral Delivery of Protein Drugs: Driver for Personalized Medicine? *Curr. Issues Mol. Biol.* **13**, 13–24 (2011).

70. Park, K., Kwon, I. C. & Park, K. Oral protein delivery: Current status and future prospect. *React. Funct. Polym.* **71**, 280–287 (2011).

71. Thorén, P. E. . *et al.* Uptake of analogs of penetratin, Tat(48–60) and oligoarginine in live cells. *Biochem. Biophys. Res. Commun.* **307**, 100–107 (2003).

72. Liang, J. F. & Yang, V. C. Insulin-cell penetrating peptide hybrids with improved intestinal absorption efficiency. *Biochem. Biophys. Res. Commun.* **335**, 734–738 (2005).

73. Kamei, N. *et al.* Usefulness of cell-penetrating peptides to improve intestinal insulin absorption. *J. Control Release* **132**, 21–25 (2008).

74. Zorko, M. & Langel, U. Cell-penetrating peptides: mechanism and kinetics of cargo delivery. *Adv. Drug Deliv. Rev.* **57**, 529–545 (2005).

75. Kavimandan, N. J., Losi, E. & Peppas, N. A. Novel delivery system based on complexation hydrogels as delivery vehicles for insulin-transferrin conjugates. *Biomaterials* **27**, 3846–3854 (2006).

76. Chalasani, K. B., Russell-Jones, G. J., Yandrapu, S. K., Diwan, P. V & Jain, S. K. A novel vitamin B12-nanosphere conjugate carrier system for peroral delivery of insulin. *J. Control Release* **117**, 421–429 (2007).

CHAPTER 2

Mathematical modeling improves target identification for enhancing phagocytosis with engineered, aglycosylated IgGs

(Adapted from Jung S. T. et al. (2013) Effective phagocytosis of low Her2 tumor cell lines with engineered, aglycosylated IgG displaying high FcγRIIa affinity and selectivity, *ACS Chemical Biology*, Volume 8, Issue 2, 368-375.)

2.1 Introduction

Antibodies trigger a wide array of responses in leukocytes by linking cellular targets on pathogens with Fc γ receptors (FcγRs) expressed on the surface of macrophages, granulocytes, dendritic cells, natural killer (NK) cells, and B cells. In humans, the binding of immune complexes (ICs) to activating receptors FcγRI, FcγRIIa, and FcγRIIIa initiates signaling through intracellular immunoreceptor tyrosine-based activation motifs (ITAMs) to effect potent antibody-dependent cell-mediated cytotoxicity (ADCC), antibody-dependent cell-mediated phagocytosis (ADCP), degranulation, T cell proliferation, and cytokine release responses[1]. Activation is moderated by the binding of ICs to the sole inhibitory receptor, FcγRIIb, inducing immunoreceptor tyrosine-based inhibitory motif (ITIM) signaling that leads to anti-inflammatory responses[2]. The affinity of IgG

antibodies for activating and inhibitory FcγRs depends on subclass, glycosylation state, and other more subtle effects such as Fc allotype[3]. IgG1 is the most commonly used isotype for therapeutic applications and engages all of the activating receptors, as well as the inhibitory FcγRIIb[3]. There is evidence that the ratio of binding affinities of IgG1 in solution for activating and inhibitory FcγRs (A/I ratio) plays a significant role in determining the extent of effector functions displayed by immune cells[4–6]. However, since signaling by FcγR receptors requires interactions with multivalent ICs, other parameters such as avidity of the ICs and expression level and cell surface distribution of the FcγRs on effector cells likely impact the nature of the response in addition to the A/I ratio.

There are two common polymorphisms in each of the FcγRIIIa and FcγRIIa subclasses that also affect the binding affinity to IgG1 antibodies[3]. Engagement of FcγRIIIa is considered important for the recruitment of NK effector cells that do not express the inhibitory receptor FcγRIIb[7]. Lymphoma patients expressing the high affinity FcγRIIIa allotype, FcγRIIIa-V158, show improved clinical outcomes when treated with anti-CD20 (rituximab) compared to patients homozygous for the low affinity FcγRIIIa-F158 allele[8,9]. Amino acid substitutions and engineering of the glycan appended to N297 of the Fc have been employed successfully to enhance affinity toward both FcγRIIIa alleles[10,11]. MGAH22, an Fc engineered Her2-specific antibody that shows improved clearance of medium and low Her2-expressing cancer cells, is under phase I clinical evaluation[12]. Likewise, the defucosylated antibody GA201 (anti-EGFR) that exhibits

about 50-fold higher FcγRIIIa-V158 affinity is in phase I/II studies for high EGFR expressing solid tumors[13].

Expression of the high affinity FcγRIIa-H131 allele correlates with improved clinical response rates following treatment with trastuzumab, rituximab, or cetuximab[9,14,15]. Conversely, individuals homozygous for FcγRIIa-R131 have an increased risk of bacterial infection and autoimmunity due to the low affinity interaction with IgGs impairing immune complex clearance[16–19]. Because the low affinity FcγRIIa-R131 allele occurs at a frequency between 0.47–0.57 in African Americans, 0.45–0.53 in Caucasians, and 0.19–0.45 in Asians, the engineering of antibodies that can enhance responses in patients bearing the FcγRIIa-R131 allele is of significant mechanistic and clinical interest[19–21]. Engagement of FcγRIIa is particularly important for phagocytosis of tumor cells by macrophages, which were recently shown to comprise almost 20% of infiltrating leukocytes in breast cancer biopsies[22].

Attempts to engineering IgG1 antibodies with FcγRIIa selectivity have led either to increased affinity for both FcγRIIa and FcγRIIb and/or relatively modest improvements in the A/I ratio (< 6-fold change compared to wild-type)[10,23,24]. Two predominant issues complicate the engineering of antibodies that display optimal FcγRIIa/FcγRIIb selectivity and FcγRIIa affinity for enhanced phagocytic potential: First, FcγRIIa shares 96% amino acid identity with FcγRIIb, and therefore the generation of Fc ligands that discriminate between the two is challenging[7]. Second, because of the avidity effects involved in the interaction of antibody-decorated tumors with effector cells it is

not easy to determine *a priori* whether high affinity to FcγRIIa, or a higher A/I ratio, or both are needed to enhance phagocytosis.

Typically, the interaction of antibodies with cognate FcγRs is critically dependent on N-linked glycosylation at residue N297 in the Fc[25]. Aglycosylated antibodies show nearly complete loss of FcγR binding and effector functions[24,26,27]. Contrary to previous findings[28], recent analysis of aglycosylated, bacterially expressed human Fc domains suggested that the absence of the glycan increases the conformational flexibility around the CH2-CH3 hinge and that the "closed conformations" (reduced CH2 distances) observed in prior aglycosylated crystal structures are likely due to crystal packing effects[29]. Screening of very large combinatorial libraries by bacterial display led to the isolation of the aglycosylated Fc5 variants, which has high affinity toward FcγRI[26] and displays a CH2 distance intermediate between those of fully glycosylated and aglycosylated Fc domains. Collectively, these findings suggest that the high flexibility of aglycosylated Fcs might allow for the identification of amino acid substitutions that stabilize conformers not accessible to glycosylated antibodies. So far, aglycosylated antibodies have been used for applications where inflammatory engagement of FcγRs or C1q complement proteins is undesirable[30]. Yet the generation of aglycosylated antibodies with unique affinities and selectivities may open the possibility for a new class of therapeutics with intriguing biological properties. We hypothesized that such substitutions might confer unusually high FcγRIIa affinity and selectivity with respect to binding to the highly homologous FcγRIIb for improved phagocytic potentiation.

The presence of both activating and inhibitory FcγRs allows effector cells to generate, suppress, and fine tune their responses based on the occupancy of these receptors[31]. Given the multitude of receptor complexes that can arise during engagement of effector cells, it is difficult to intuit how these receptor signals, often conflicting in nature, result in a specific cellular response. Mathematical modeling can provide a quantitative framework for better understanding this nuanced regulation[32]. We developed a mathematical model to link biomolecular interactions at the cell surface (IgG binding to Fcγ receptor subunits) to cellular response (ADCP). Experimental results with our novel Fc variants, which have altered receptor-binding properties and varied cellular responses, were used for model parameterization and validation. The model provides insights into the molecular determinants of strong ADCP and suggests general design rules for further improving the efficacy of Fc variants that elicit this cellular response.

This work has been performed in close collaboration with George Georgiou's laboratory at the University of Texas at Austin. All of the experimental work has been performed by our collaborators and all of the computational work has been performed by us. Using a new library screening system, an engineered trastuzumab variant displaying > 160-fold increased affinity for the FcγRIIa-R131 allele and a > 25 increase in A/I ratio compared to clinical grade glycosylated trastuzumab (Herceptin) has been isolated. This engineered trastuzumab was shown to elicit markedly increased ADCP with human macrophages as effector cells relative to Herceptin for both +3 (SKOV-3) and +2 (MDA-MB-453) Her2 expressing cell lines[33]. A mathematical model has been developed to provide design criteria for further improving the efficacy of Fc variants by elucidating a

27

mechanism that links their FcγRIIa/FcγRIIb binding to activation of phagocytosis. To the best of our knowledge, this is the first report of an aglycosylated antibody that binds to FcγRs with affinity and selectivity that far exceed those of clinically used glycosylated antibodies or of any engineered glycosylated mutants reported thus far, engineered either through mutagenesis programs or via glycoengineering[10,11,23,34].

2.2 Experimental methods and results

2.2.1 Development of a bacterial display system for IgG having a free Fc domain

To discover new aglycosylated Fc mutants, a bacterial display system has been developed in which full length IgG1 antibodies are tethered to the inner membrane of *E. coli*, leaving the Fc domain free to interact with fluorescently labeled FcγRs. Full length IgG1 display system was used to accurately capture the conformational and steric effects present when antibodies interact with Fc binding ligands (Fig. 2.1). For this purpose, the light chain (L chain) was fused at the N-terminal with the PelB leader peptide and the first 6 amino acids of the *E. coli* inner membrane lipoprotein NlpA, which anchor it to the periplasmic side of the bacterial inner membrane[35]. To avoid steric constrains that might arise if both L chains of an IgG molecule are membrane-anchored, the plasmid was designed to express NlpA anchored L chain (V_L-C_κ) as well as unanchored L chain from the same promoter. Heavy chain polypeptides secreted into the periplasm assemble either with (i) two unanchored L chains, (ii) a membrane-anchored and an unanchored L chain, or (iii) two membrane-anchored L chains. The unanchored IgG in the periplasm is

preferentially released from the bacteria following spheroplasting, whereas anchored IgG comprising one membrane-anchored and one unanchored L chain (or two membrane-anchored L chains) is retained on the surface of the spheroplasts. The Fc domain in the membrane-anchored IgG is free to interact with exogenously added, fluorescently labeled FcγRs for fluorescence activated cell sorting analysis (FACS) and library screening. To validate the IgG display strategy, the H and L chains of AglycoT-Fc5 (aglycosylated trastuzumab with E382V/M428I[26]) and AglycoT-Fc2a (aglycosylated trastuzumab with S298G/T299A[24]) were cloned separately and expressed in *E. coli*. Under optimized culture conditions, AglycoT-Fc5 expressing spheroplasts displayed high fluorescence following labeling with FcγRI-FITC, whereas spheroplasts expressing AglycoT-Fc2a could be labeled selectively with FcγRIIa-GST/anti-GST-FITC[36]. In both instances, a large increase in fluorescent signal was observed relative to the negative control, indicating that inner membrane-anchored IgG display is well suited for the engineering of aglycosylated antibody Fc variants that bind to other Fc receptors (FcγRI, FcγRIIa, FcγRIIb, FcγRIIIa/b, or FcRn).
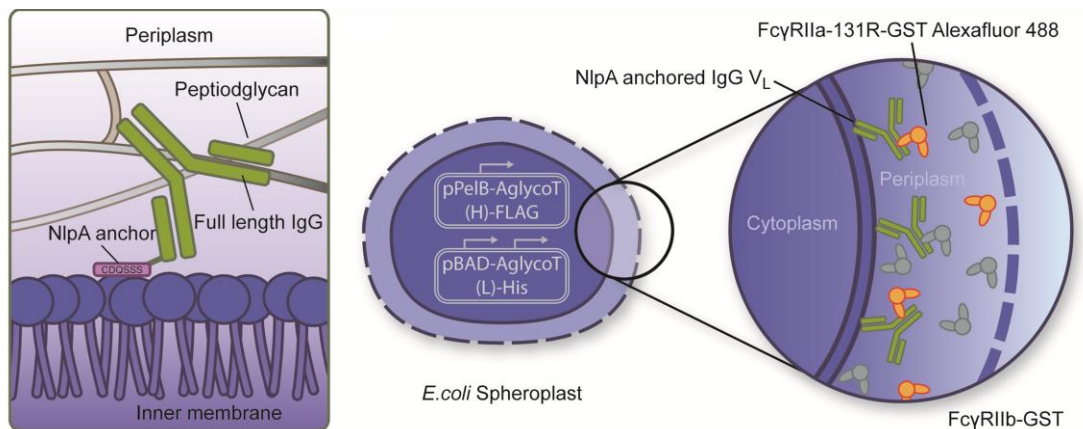
Figure 2.1. Schematic of a covalently anchored, full-length IgG1 display system for Fc engineering. Schematic diagram showing the display of full length IgG1 by NlpA tethering to inner membrane via a single membrane-anchored light chain. Soluble fluorescent FcγRIIa-R131-GST and non-fluorescent FcγRIIb-GST can then compete for binding to full length library variants on the surface of bacterial spheroplasts before FACS sorting.

2.2.2 Isolation and characterization of aglycosylated IgG variants displaying high affinity and binding selectivity toward the FcγRIIa-R131 allele

An IgG that combined mutations from both Fc5 and Fc2a (E382V/M428I/ S298G/T299A) was constructed and shown to display binding affinity to both FcγRI and FcγRIIa/FcγRIIb comparable to that of Fc5 and Fc2a, respectively, when expressed as individual variants[36]. SPR analysis of the purified AglycoT-Fc5-2a revealed that binding to FcγRI was significantly reduced relative to Herceptin (11-fold decrease or a 3-fold decrease relative to AglycoT-Fc5). Interestingly, this antibody showed a small (40%) but reproducible increase in the selectivity for FcγRIIa-R131 allele relative to FcγRIIb ($A_{IIa-R131}$/I ratio) when compared to Herceptin and an even more significant increase in $A_{IIa-R131}$/I relative to AglycoT-Fc2a (Table 2.1). Therefore, AglycoT-Fc5-2a was used as a template for random mutagenesis and the isolation of variants that exhibit improved binding to the low affinity FcγRIIa-R131 allele and simultaneously increased $A_{IIa-R131}$/I ratios.

Table 2.1. SPR analysis showing kinetic on and off rates and affinity for glycosylated herceptin and aglycosylated trastuzumab Fc variants binding to FcγRIIa-R131, FcγRIIa-H131, FcγRIIb and FcγRI.

Variant	FcγR	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_D (µM)	A/I ratio	K_D relative to Herceptin
Herceptin	FcγRI	1.8 x 10^5	2.7 x 10^{-4}	0.0015 [a]		1
	FcγRIIa-H131	7.2 x 10^5	8.5 x 10^{-2}	0.12	11	1
	FcγRIIa-R131	1.2 x 10^5	3.7 x 10^{-2}	0.31	4.2	1
	FcγRIIb	2.8 x 10^4	3.8 x 10^{-2}	1.3		1
AglycoT-Fc5-2a	FcγRI	1.5 x 10^5	2.3 x 10^{-3}	0.016		0.096
	FcγRIIa-H131	3.1 x 10^5	1.2 x 10^{-1}	0.37	4.3	0.32
	FcγRIIa-R131	1.3 x 10^5	3.6 x 10^{-2}	0.27	5.9	1.1
	FcγRIIb	5.8 x 10^4	9.1 x 10^{-2}	1.6		0.81
AglycoT-Fc1001	FcγRI	4.3 x 10^5	3.3 x 10^{-2}	0.077		0.020
	FcγRIIa-H131	2.6 x 10^5	5.4 x 10^{-2}	0.20	2.4	0.60
	FcγRIIa-R131	2.6 x 10^5	3.9 x 10^{-3}	0.015	31	21
	FcγRIIb	2.9 x 10^4	1.3 x 10^{-2}	0.47		2.8
AglycoT-Fc1002	FcγRI	N/A	N/A	N/A		N/A
	FcγRIIa-H131	1.5 x 10^5	2.8 x 10^{-2}	0.19	10	0.63
	FcγRIIa-R131	1.8 x 10^5	4.0 x 10^{-2}	0.22	8.6	1.4
	FcγRIIb	2.0 x 10^4	3.9 x 10^{-2}	1.9		0.68
AglycoT-Fc1003	FcγRI	N/A	N/A	N/A		N/A
	FcγRIIa-H131	3.3 x 10^5	5.9 x 10^{-2}	0.18	5.6	0.67
	FcγRIIa-R131	2.6 x 10^5	3.2 x 10^{-2}	0.12	8.3	2.6
	FcγRIIb	3.8 x 10^4	3.8 x 10^{-2}	1		1.3
AglycoT-Fc1004	FcγRI	4.0 x 10^5	2.6 x 10^{-2}	0.064		0.024
	FcγRIIa-H131	5.3 x 10^5	1.1 x 10^{-2}	0.021	9.5	5.7
	FcγRIIa-R131	3.2 x 10^5	6.2 x 10^{-4}	0.0019	105	163
	FcγRIIb	2.5 x 10^4	4.9 x 10^{-3}	0.2		6.5

[a] Affinity reported in a previous study[26], using the same method.

A library of random mutations in the Fc domain of AglycoT-Fc5-2a was created by error prone PCR to yield 4.8×10^9 transformants. Cells were incubated with a progressively increasing excess of unlabeled FcγRIIb-GST (extracellular domain of FcγRIIb fused with glutathione-S-transferase (GST) from *Schistosoma japonicum*), over highly homologous fluorescent FcγRIIa-R131-GST-Alexa488 conjugate to help enrich variants displaying improved FcγRIIa-R131/FcγRIIb selectivity. After five rounds of sorting, four individual clones showing markedly improved fluorescence upon incubation with 20 nM FcγRIIa-R131-GST-Alexa-488 and 100 nM of FcγRIIb-GST relative to spheroplasted cells expressing AglycoT-Fc5-2a were isolated (Fig. 2.2A). Cells expressing the clone termed AglycoT-Fc1001 (S298G, T299A, H268P, E294K, N361S and E382V, M428L (Fig. 2.2B)) exhibited the highest fluorescence under these conditions (Fig. 2.2D). Interestingly, the best performing variant in ADCP assays, AglycoT-Fc1004 (S298G, T299A, N390D, E382V, M428L (Fig. 2.2C)) showed the lowest fluorescent signal (Fig. 2.2D). Expression differences between the Fc1001 and Fc1004 variants are likely to be responsible for the qualitative difference between FACS signal and FcγRIIa binding; this can be addressed by two color sorting that also monitors antibody expression. Note that in AglycoT-Fc5-2a, residue M428 had been mutated to Ile, which was further changed to a Leu in AglycoT-Fc1004 and the three other selected clones pointing to the functional significance of this residue for effector FcγR binding (Fig. 2.2A).
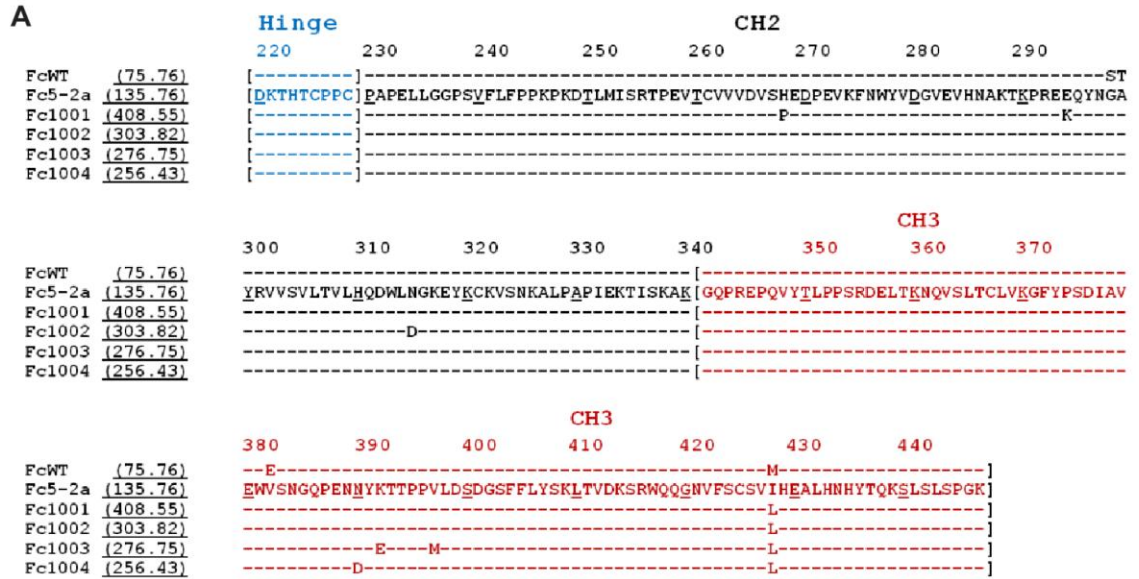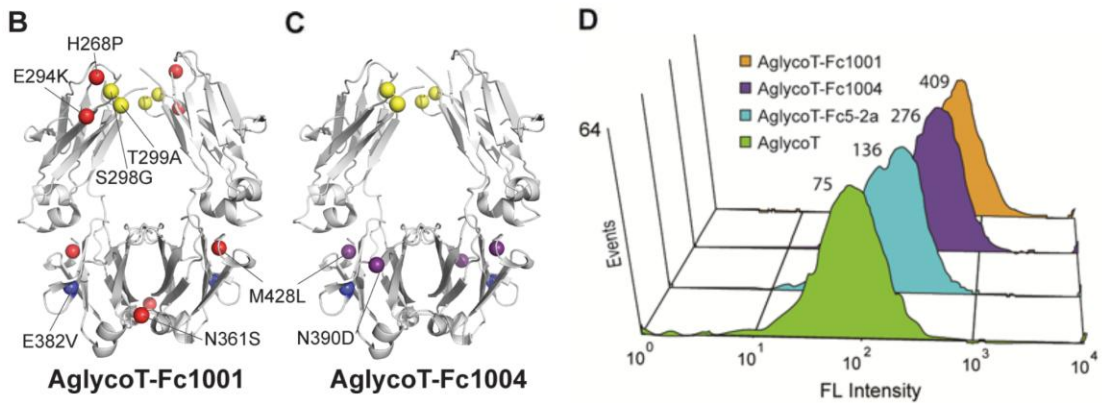
Figure 2.2. Isolated Fc variants. (A) Sequences and corresponding FACS signals for isolated Fc variants. Mutations isolated in (B) AglycoT-FC1001 and (C) AglycoT-Fc1004 are shown on the 3D structure of an aglycosylated IgG1 Fc (PDB code: 3S7G). Yellow = AglycoT-Fc2a base mutations, blue = AglycoT-Fc5 base mutations, red = AglycoT-Fc1001 mutations and purple = AglycoT-Fc1004. (D) Fluorescent histogram of variant binding to 30 nM FcγRIIa as detected by secondary goat anti-GST-FITC diluted 1:200 from a 1 mg/ml stock.

The four IgG variants corresponding to the sequences encoded by the highest fluorescence clones isolated by FACS were expressed in HEK293F cells to limit

endotoxin contamination and purified. All IgGs contained the S298G/T299A mutations that block glycosylation at N297 and thus do not bind Concanavalin A. Binding of dimeric purified GST fusions of FcγRIIa-H131, FcγRIIa-R131, and FcγRIIb to each IgG variant was determined by SPR analysis (Table 2.1). A bivalent binding model was used to fit the sensorgram curves and obtain monomeric binding constants in accordance with earlier studies[37]. This methodology resulted in equilibrium dissociation values, k_{on} and k_{off} rate constants for the binding of Herceptin and the various mutant Fcs to FcγRs that were in excellent agreement with the recent comprehensive binding data of all human IgG subclasses to effector FcγRs[3]. AglycoT-Fc1004 IgG displayed the highest affinity for both FcγRIIa polymorphisms: a remarkable 163-fold improvement in K_D for the low affinity FcγRIIa-R131 relative to Herceptin and a more modest 5.7-fold affinity improvement for the high affinity FcγRIIa-H131 allele. The large increase in affinity toward FcγRIIa-R131 is consistent with the use of this receptor polymorphism for FACS screening. While AglycoT-Fc1004 showed somewhat higher affinity toward the inhibitory FcγRIIb, the much greater increase in FcγRIIa-R131 affinity resulted in 25-fold improvement in $A_{IIa-R131}$/I selectivity compared to Herceptin and a 16-fold increase in selectivity relative to its parental variant, AglycoT-Fc5-2a. AglycoT-Fc1001 displayed 21-fold higher FcγRIIa-R131 affinity than Herceptin and an $A_{IIa-R131}$/I of 7.5 (5.3-fold increase relative to AglycoFc5-2a). The two other IgGs isolated from the screening displayed only marginal increases in FcγRIIa-R131 binding and lower affinity toward the FcγRIIa-H131 allele.

34

The binding kinetics of the two highest affinity clones AglycoT-Fc1004 and AglycoT-Fc1001 and their parental AglcycoT-Fc5-2a toward FcγRI were evaluated by SPR, and the data were fit with a 1:1 Langmuir binding model[37] (Table 2.1). AglycoT-Fc5-2a displayed a 10-fold decrease in FcγRI binding compared to Herceptin, which was further reduced in AglycoT-Fc1004 and AglycoT-Fc1001 to almost a 50-fold lower level. Evidently, the S298G/T299A mutations that conferred binding to FcγRIIa and FcγRIIb to aglycosylated IgGs suppressed the ability of the Fc5 amino acid substitutions to bind FcγRI. AglycoT-Fc1004 has only two amino acid substitutions in the CH3 domain (N390D, M428L) yet has 5-fold lower K_D for FcγRI relative to its parental AglycoT-Fc5-2a template. Our data show that the reason for the decreased FcγRI affinity of AglycoT-Fc1004 results from the N390D mutation and the subtle change I428L (compared to I428M in the parental antibody Fc5-2a). Both of these mutations are in the CH3 domain, which is distal to the FcγR binding epitope, underscoring the significance of conformational flexibility in these interactions. This finding is consistent with the recent structural data that indicates in the absence of glycosylation the Fc domain displays a high degree of conformational flexibility and our hypothesis that mutations within the CH3 domain can stabilize particular conformers with unusual FcγR binding properties[29]. In particular, M428L is located near the hinge region of the CH2 and CH3 domains where this mutation likely influences the local conformation and favorably impacts the binding of antibodies to FcRn[38].

The isolated IgGs displayed no binding to either of the two FcγRIIIa polymorphisms (F158 and V158) as determined by ELISA[36]. Interestingly, both

35

AglycoT-Fc1001 and AglycoT-Fc1004 showed higher binding to the neonatal Fc receptor, FcRn, at endosomal pH 6.0 but not at physiological pH[36]. This was most likely the consequence of the M428L mutation, which had been shown earlier to improve binding affinity for FcRn at pH 6.0 by 11-fold and thus confer extended IgG serum persistence and pharmacokinetics in animal models[38].

2.2.2 Enhanced ADCP of low and medium Her2$^+$ cell lines

Surface expression of Her2 on SKBR-3, SKOV-3, and MDA-MB-453 Her2$^+$ cells was evaluated by FACS following incubation with Herceptin and a fluorescently labeled secondary antibody. SKBR-3 cells showed a 2-fold higher FACS signal relative to SKOV-3, which in turn had a 2-fold greater Her2 binding signal compared to that of MDA-MB-453 cells (Fig. 2.3A). These results are consistent with the total Her2 expression in these cell lines as determined by Western blotting[39].
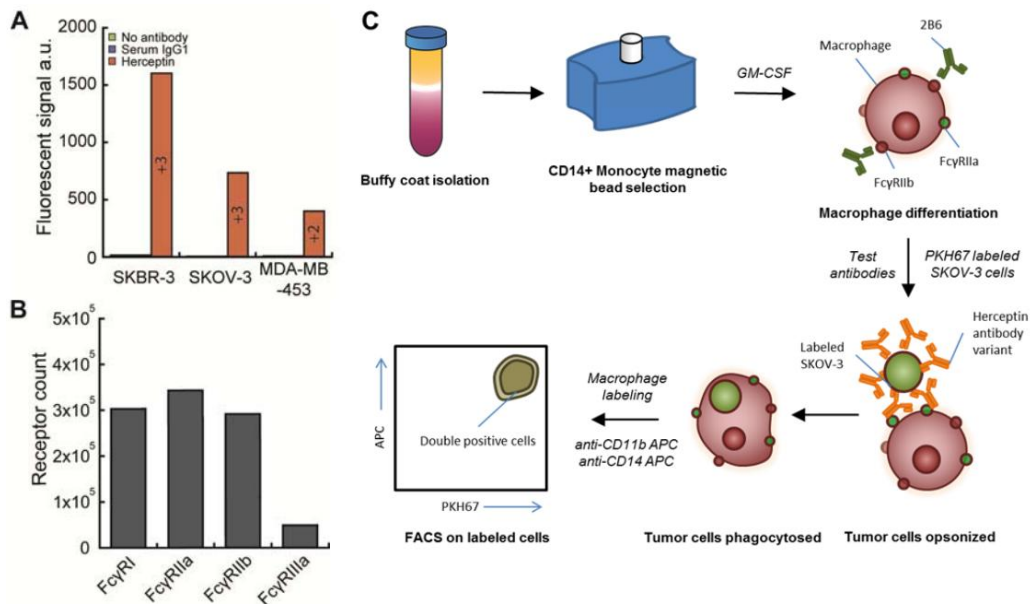
Figure 2.3. Expression level of Her2 and FcγR on tumor cell lines and macrophages for ADCP assay. (A) Her2 expression level on tumor cell lines used for ADCP was confirmed by labeling with 10 μg/ml Herceptin or IgG1 pooled from human serum followed by fluorescent donkey anti-human IgG (H+L) FITC Fab at a 1:50 dilution. Bars are labeled with the immunohistochemical staining category assigned to each tumor cell line[33]. (B) FcγR counts on macrophages were determined using a fluorescent Quantum Simply Cellular bead assay. Macrophages were labeled with 20 μg/ml anti-FcγRI-FITC, 10 μg/ml anti-FcγRIIa-FITC, 1 μg/ml 2B6-N287D-FITC and 20 μg/ml anti-FcγRIII-FITC. (C) Schematic diagram showing the experimental workflow of the ADCP analysis.

$CD14^+$ $CD11b^+$ macrophages were prepared from pooled PBMCs (peripheral blood mononuclear cells) following cultivation with GM-CSF cytokine[23]. The surface densities of FcγRI, FcγRIIa, FcγRIIb, and FcγRIII on macrophages were determined by FACS. We observed high FcγRI, FcγRIIa, and FcγRIIb levels in macrophages prepared at different times and from different pooled PBMC fractions (Fig. 2.3B). Notably, the level of FcγRI (CD64) was comparable to that of FcγRIIa and FcγRIIb, in contrast to an earlier report that showed low level FcγRI expression in human macrophages[23].

ADCP was determined by incubating PKH67 labeled, IgG opsonized tumor cells with macrophages at a low (1:5 MDA-MB-453 or 1:10 SKOV-3) ratio of target to effector cells (Fig. 2.3C). The numbers of $CD11b^+$ $CD14^+$ macrophages that stained with PKH67 (arising from associated/ingested tumor cells) were determined by FACS[23] and fluorescence microscopy. As a positive control we used glycosylated trastuzumab with the mutation G236A (GlycoT-G236A), which has been reported to confer the highest $A_{IIa-R131}/I$ and $A_{IIa-H131}/I$ ratios reported in the literature (5.5 and 5.7, respectively, relative to wild-type IgG) and to mediate increased macrophage mediated ADCP of $EpCAM^+$

LS180 cells[23]. We found that AglycoT-Fc1004 internalized approximately 75% more tumor cells ($P < 0.01$) relative to Herceptin and 40% more relative to its glycosylated, Fc engineered GlycoT-G236A variant (Fig. 2.4). Equally importantly, AglycoT-Fc1004 displayed the same high level of ADCP with both the SKOV-3 and with the even lower Her2 expressing MDA-MB-453 cell line. These two cell lines correspond to approximately +3 and +2 Her2$^+$ tumor cells, respectively, which have been shown to be recalcitrant to Herceptin treatment[40,41]. Consequently, AglycoT-Fc1004 may provide the opportunity to treat patients having tumors with Her2 moderate or low expression that do not respond well to Herceptin. We expect binding to the low affinity FcγRIIa-R131 polymorphism to be particularly desirable since therapeutic responses in patients that are homozygotic or heterozygotic for this allele show poorer responses to anti-neoplastic antibodies[9,14,15]. On the other hand, AglycoT-Fc1001 could not induce phagocytosis of tumor cells above background for either cell line. Likewise, the addition of 2B6, an antagonistic FcγRIIb antibody, had no significant effect on ADCP with any of the engineered antibodies tested or with Herceptin, suggesting that tumor cell surface-bound ICs likely displace 2B6 from FcγRIIb allowing ITIM mediated signaling to occur.
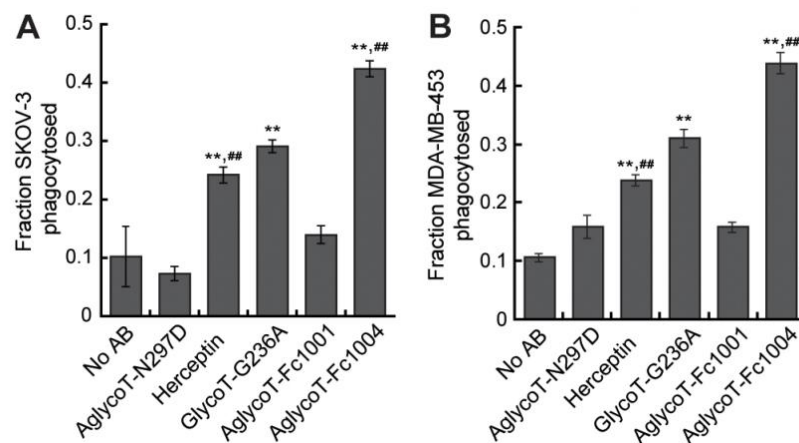
Figure 2.4. ADCP mediated by macrophages with various Fc variants. Human monocyte derived macrophages were used with (A) +3 Her2[+] SKOV-3 ovarian cancer and (B) +2 Her2[+] MDA-MB-453 breast cancer tumor cell lines as target cells. [**] Welch's *t* test *P* < 0.01 for samples compared to No Ab negative control, [##] Welch's *t* test *P* < 0.01 for samples relative to G236A positive control.

2.3 Computational methods

2.3.1 Parameters and schematic of the mathematical model

A mathematical model was developed to better understand the interaction between Her2-expressing cells (SKOV-3 or MDA-MD-453) and macrophages, and to predict the relative level of ADCP based on the K_D values of the Fc variants to the different FcγRIIs (Table 2.1). Model nomenclature and parameter values are provided in Table 2.2.

Table 2.2. Parameters used to generate FcγRIIa/b activation model.

Parameter	Description	Value	Reference
Physical parameters			
SK_dia	Diameter of SKOV-3 cell	10 μm	Zhang et al.[42]
MD_dia	Diameter of MDA-MB-453 cell	10 μm	Rhodes et al., Zhang et al.[33,42]
Mac_dia	Diameter of macrophage	21 μm	Krombach et al.[43]
Cell_gap	Gap distance between SKOV-3/MDA-MB-453 cell and macrophage	12 nm	Pease et al., Sondermann et al. [44,45]
Contact_area	Contact area of FcγRII receptors on macrophage	104.7 μm^2	Kuo and Lauffenburger, Heiple et al., Tolentino et al.[46–48]

Expression level parameters

SK_HER2	HER2 expression level on SKOV-3	7.36×10^5	Costantini et al., Xu et al.[49,50]
MD_HER2	HER2 expression level on MDA-MB-453	4.00×10^5	Costantini et al., Xu et al.[49,50]
Mac_IIaH	Number of FcγRIIa-H131 on macrophage	171271	This work
Mac_IIaR	Number of FcγRIIa-R131 on macrophage	171271	This work
Mac_IIb	Number of FcγRIIb on macrophage	291150	This work
$[IgG]_s$	Free serum IgG concentration	10 μM	This work
$[L_0]_{SK}$	SKOV-3 Her2-bound IgG effective concentration	342.2 μM	This work
$[L_0]_{MD}$	MDA-MB-453 Her2-bound IgG effective concentration	176.2 μM	This work

Affinity parameters

K_{diff}	Equilibrium constant for partitioning of FcγRII receptors in/out of contact area on macrophage	16.64	This work
K_{cross}	Equilibrium dissociation constant for the crosslinking of FcγRII receptors	$2500/\mu m^2$	This work
$K_{D\_IIa\_IgG}$	Equilibrium dissociation constant between serum IgG and FcγRIIa-H131/FcγRIIa-R131	0.72 μM	Maenaka et al.[51]
$K_{D\_IIb\_IgG}$	Equilibrium dissociation constant between serum IgG and FcγRIIb	2.4 μM	Maenaka et al.[51]
$K_{D\_IIa-H131}$	Equilibrium dissociation constant between Fc variants and FcγRIIa-H131	(Table 2.1)	This work
$K_{D\_IIa-R131}$	Equilibrium dissociation constant between Fc variants and FcγRIIa-R131	(Table 2.1)	This work
K_{D_IIb}	Equilibrium dissociation constant between Fc variants and FcγRIIb	(Table 2.1)	This work

The interaction between Her2-bound IgG on a tumor cell and FcγRs on a macrophage occurs in a "contact" region where the two cells are in close proximity, but

this interaction is not possible outside of this region because of physical constraints imparted by the curvature of the two cells (Fig. 2.5). The contact area was estimated to be 1/3 of the surface area of the smaller SKOV-3/MDA-MD-453 cells based on geometric considerations. A lower bound of 1/10 of the surface area has been calculated for a non-deforming bead[46], but the actual contact area is significantly higher because macrophages deform and spread around IgG-bound cells[47,48].
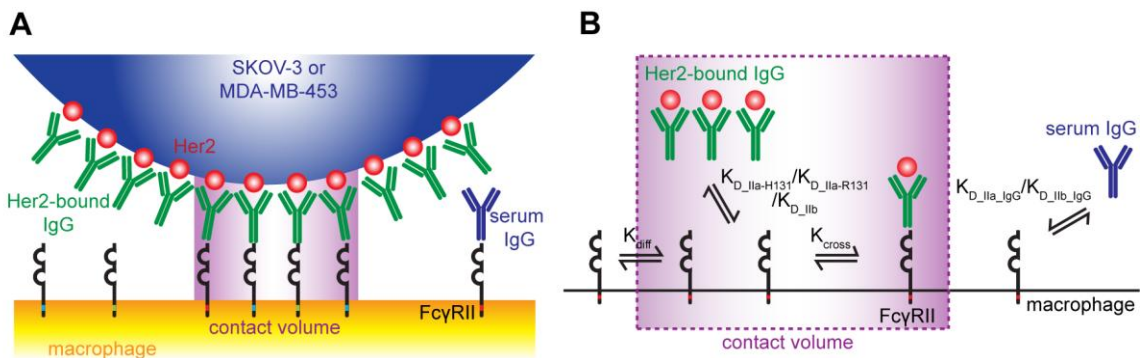
Figure 2.5. Schematic of the mathematical model. (A) Overall schematic of the various species considered in the mathematical model. Anti-Her2 IgG molecules with different Fc domains saturably bind the Her2 receptors on SKOV-3 or MDA-MB-453 cells. The different Fc domains have various binding affinities to FcγRIIa-H131, FcγRIIa-R131, and FcγRIIb on macrophages in the contact region. Serum IgG is also included in the model and is able to bind to FcγRII receptors both inside and outside of the contact region. (B) Schematic of binding interactions and their respective affinity parameters considered in the model.

2.3.2 Interactions between species in the model

We considered three types of FcγRs on macrophages: FcγRIIa-H131, FcγRIIa-R131, and FcγRIIb. The number of FcγRIIa and FcγRIIb were experimentally quantified (Fig. 2.3B), and the FcγRIIa-H131 and FcγIIa-R131 variants were assumed to exist in 50/50

proportions[20]. FcγRIIs partitioned between the contact area and the free (non-contact) area with equilibrium constant K_{diff}:

$$FcγRIIx_{free} \xleftarrow{\quad} ^{Kdiff} \xrightarrow{\quad} FcγRIIx_{contact}$$
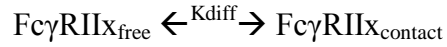
where K_{diff} = [FcγRIIx$_{free}$] / [FcγRIIx$_{contact}$] = (macrophage surface area – contact _area) / (contact_area) and FcγRIIx = FcγRIIa-H131, FcγRIIa-R131 or FcγRIIb.

All Her2 receptors on SKOV-3 and MDA-MD-453 cells were considered to be evenly distributed and receptor numbers were calculated from experimental quantification of relative expression levels (Fig. 2.3A) and total absolute values from literature[49,50]. The effective concentration of cell-bound IgG ([L_0]) was calculated in an "effective contact volume", defined as the product of contact_area and cell_gap. These cell-bound IgGs were free to interact with FcγRIIs in the contact area (FcγRIIx$_{contact}$) with ligand depletion:

$$L + FcγRIIx_{contact} \xleftarrow{\quad} ^{KD\_x} \xrightarrow{\quad} FcγRIIx_{cell\text{-}bound}$$
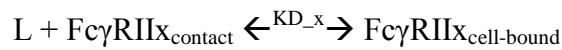
where K_{D_x} = $K_{D\_IIa\text{-}H131}$, $K_{D\_IIa\text{-}R131}$, or K_{D_IIb} from SPR data for different Fc variants (Table 2.1).

To mimic conditions in our *in vitro* experiments, as well as in normal physiology, 10 μM serum IgG ([IgG]$_s$) was included in the system to compete with Her2-bound IgG for binding with Fc receptors. Serum IgG was assumed to be in excess and could bind FcγRIIs anywhere on the macrophage surface without ligand depletion:

$$IgG_s + FcγRIIx_{free} \xleftarrow{\quad} ^{KD\_x\_IgG} \xrightarrow{\quad} FcγRIIx_{free\ I\ bound}$$

42

$$\text{IgG}_s + \text{Fc}\gamma\text{RIIx}_{\text{contact}} \xleftarrow{\text{KD\_x\_IgG}} \text{Fc}\gamma\text{RIIx}_{\text{contact I bound}}$$
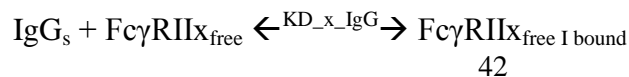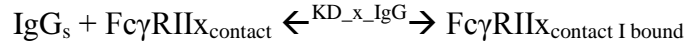
where $K_{D\_x\_IgG} = K_{D\_IIa\_IgG}$ or $K_{D\_IIb\_IgG}$.

Since receptor crosslinking leads to cell activation or inhibition, we assumed that dimers represent the minimal signaling unit (and serve as proxies for any higher-order receptor clusters). Any IgG-bound FcγRIIs were allowed to dimerize with equilibrium dissociation constant K_{cross}. K_{cross} was chosen to be 2500 #/μm^2, which maximizes the difference in the number of crosslinked receptors with and without MDA-MB-453 cells (chosen for this signal optimization since they express fewer Her2 molecules than SKOV-3 cells and therefore have lower signals). All possible combinations of dimers were allowed between the three FcγRII subunits with the same crosslinking constant, whether occupied by serum IgG of Her2-bound IgG. However, geometric constraints limited receptors in the contact area to only crosslink with those in the contact area, while receptors outside of the contact area could only crosslink with those outside:

$$\text{Fc}\gamma\text{RIIx}_{\text{cell-bound}} + \text{Fc}\gamma\text{RIIx}_{\text{cell-bound}} \xleftarrow{\text{Kcross}} \text{Fc}\gamma\text{RIIx} - \text{Fc}\gamma\text{RIIx}$$

$$\text{Fc}\gamma\text{RIIx}_{\text{contact I bound}} + \text{Fc}\gamma\text{RIIx}_{\text{contact I bound}} \xleftarrow{\text{Kcross}} \text{Fc}\gamma\text{RIIx} - \text{Fc}\gamma\text{RIIx}$$

$$\text{Fc}\gamma\text{RIIx}_{\text{free I bound}} + \text{Fc}\gamma\text{RIIx}_{\text{free I bound}} \xleftarrow{\text{Kcross}} \text{Fc}\gamma\text{RIIx} - \text{Fc}\gamma\text{RIIx}$$
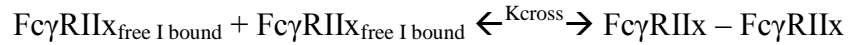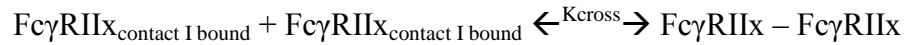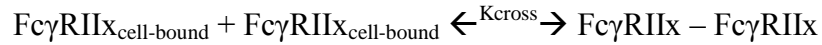
No discrimination was made between crosslinked receptors in and out of the contact area because all of them could lead to activating or inhibitory signal. However, local concentration effects made the density of dimers (and potentially higher-order clusters) much higher in the contact area.

The diffusion, binding and crosslinking reactions above yield the following system of equations (x = IIa-H131, IIa-R131 or IIb):

$$[Fc\gamma RIIx]_{free} = K_{diff} [Fc\gamma RIIx]_{contact} \qquad (1, 2, 3)$$

$$[([L_0]_{SK/MD} - [Fc\gamma RIIx]_{cell\text{-}bound})/(N_{av}*cell\_gap)]*[Fc\gamma RIIx]_{contact}$$
$$= K_{D\_x} [Fc\gamma RIIx]_{cell\text{-}bound} \qquad (4, 5, 6)$$

$$[IgG]_s [Fc\gamma RIIx]_{free} = K_{D\_x\_IgG} [Fc\gamma RIIx]_{free\ I\ bound} \qquad (7, 8, 9)$$

$$[IgG]_s [Fc\gamma RIIx]_{contact} = K_{D\_x\_IgG} [Fc\gamma RIIx]_{contact\ I\ bound} \qquad (10, 11, 12)$$

$$[Fc\gamma RIIx]_{cell\text{-}bound} [Fc\gamma RIIx]_{cell\text{-}bound} = K_{cross} [Fc\gamma RIIx : Fc\gamma RIIx] \qquad (13\ to\ 18)$$

$$[Fc\gamma RIIx]_{contact\ I\ bound} [Fc\gamma RIIx]_{cell\text{-}bound} = K_{cross} [Fc\gamma RIIx : Fc\gamma RIIx] \qquad (19\ to\ 27)$$

$$[Fc\gamma RIIx]_{free\ I\ bound} [Fc\gamma RIIx]_{free\ I\ bound} = K_{cross} [Fc\gamma RIIx : Fc\gamma RIIx] \qquad (28\ to\ 33)$$

where $[Fc\gamma RIIx]$ is in $\#/\mu m^2$. Conservation of mass gives:

all $[Fc\gamma RIIx]$ in free area * (macrophage total surface area – contact_area)
$$+ \text{all } [Fc\gamma RIIx] \text{ in contact area} * contact\_area = Mac\_IIx \qquad (34, 35, 36)$$

This system of equations was solved in Matlab using *fsolve* to obtain the 36 unknowns.

2.3.3 Activating and inhibitory receptors signaling to ADCP

Finally, the resulting distribution of FcγRII dimers was correlated to the experimental output of ADCP. The relative contribution of activating/inhibiting homodimers to this cellular response is not known and, although there is evidence that heterodimers (FcγRIIa crosslinked with FcγRIIb) do form[2], it is not known whether they activate or inhibit. Therefore, we did not assign any *a priori* functions or signaling weights to these species, but rather allowed their contributions to be determined by the model by assigning an "intrinsic signaling potency" to each subunit. The signaling potency for FcγRIIb was fixed at -1 (negative for inhibitory) and the signaling potencies for FcγRIIa-H131 and FcγRIIa-R131 were allowed to vary freely. We then assumed that the signaling potency of each dimeric species (used as a proxy for higher-order receptor clusters and phagocytic propensity) as the sum of potencies of the two receptor subunits comprising the dimer. The overall response was calculated from:

$$\text{Phagocytosis} \propto \Sigma \text{ (Fc}\gamma\text{RIIx : Fc}\gamma\text{RIIx * signaling potency of the respective dimer)}$$

The level of phagocytosis was then compared to ADCP experimental data for only two Fc variants to obtain the intrinsic signaling potencies for FcγRIIa-H131 and FcγRIIa-R131 relative to the signaling potency for FcγRIIb (which was -1; negative for inhibition). Fitted potency values are presented in Table 2.3.

Table 2.3. Intrinsic signaling potencies of FcγRIIa-H131, FcγRIIa-R131, and FcγRIIb.

ADCP data considered	Intrinsic signaling potency		
	FcγRIIa-H131	FcγRIIa-R131	FcγRIIb [a]
Fitting Herceptin and AglycoT-Fc1001 only	2.8	0.4	-1

Fitting Herceptin and AglycoT-Fc1004 only	2.0	0.3	-1
Fitting AglycoT-Fc1001 and AglycoT-Fc1004 only	11.3	0.1	-1

[a] The signaling potency of FcγRIIb was always held fixed at -1, since the goal of each fit was to determine the *relative* potencies among the three receptor subunits.

2.4 Results

2.4.1 Antigen-bound IgG, but not monomeric IgG, can elicit phagocytic response

To gain insights into the molecular determinants triggering ADCP, especially the puzzling lack of phagocytosis of Her2 expressing tumor cells opsonized with the AglycoT-Fc1001 antibody, we developed a quantitative model to relate the formation of IgG:FcγRIIa and IgG:FcγRIIb ICs to phagocytosis. FcγRs are free to diffuse into and out of a "contact" region, where they can interact with Her2-bound IgGs that coat the adjacent tumor cell (Fig. 2.5). IgG-bound FcγRs can then crosslink and generate an activating or inhibitory signal, with ADCP as the output. Monomeric IgG, which is present in serum and interstitial fluid at concentrations sufficient to saturate all FcγRII subunits, can compete with antigen-bound IgG for FcγR binding. Our simulations suggest that antigen-bound IgG, but not monomeric IgG, can elicit a significant phagocytic response. This is because the clustered presentation of antigen-bound IgGs in the contact region results in an increased local concentration of Fc domains that leads to the displacement of monomeric IgG and promotes a sufficiently high level of FcγR cross-

links to elicit effector responses. This result is in complete agreement with *in vivo* observations that serum IgG alone, in the absence of ICs, cannot elicit FcγR signaling[1].

2.4.2 Mathematical modeling accurately captures ADCP results

To link biomolecular interactions at the cell surface to ADCP efficacy, we assigned an "intrinsic signaling potency" parameter to each receptor subunit and assumed that the signaling potency of each dimeric IgG:FcγR (used as a proxy for higher-order IgG:FcγR clusters) was equal to the sum of its constituents. The net signal generated by all FcγR complexes on the macrophage surface constituted the level of ADCP response[3]. With the exception of the intrinsic signaling potencies for FcγRIIa-H131 and FcγRIIa-R131 (relative to FcγRIIb), all other parameters were obtained from experimental data, literature values, or simulations (Table 2.2). To parameterize the two relative signaling potencies for FcγRIIa-H131 and FcγRIIa-R131, the model was fit only to ADCP data for Herceptin and AglycoT-Fc1001 using SKOV-3 and MDA-MB-453 cells; this model was then used to predict the expected ADCP response of AglycoT-Fc1004 in both cell types (Fig. 2.6A and B). Similarly, parameterization of the model using the data for any two antibodies within the Herceptin, AglycoT-Fc1001, and AglycoT-Fc1004 set accurately predicted the ADCP response for the third antibody (Fig. 2.6C−F). Notably, the trend in magnitudes of the relative intrinsic signaling potencies was always the same: FcγRIIa-H131 > FcγRIIb > FcγRIIa-R131 (Table 2.3).
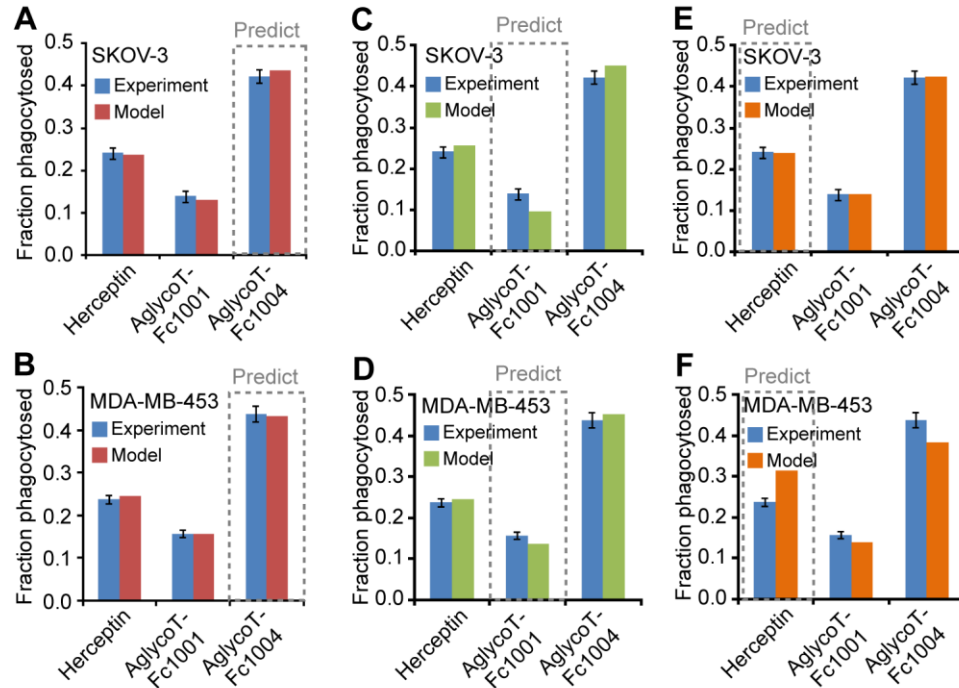
Figure 2.6. Validation of the mathematical model with ADCP data. (A, B) ADCP data for Herceptin and AglycoT-Fc1001 were used to fit intrinsic signaling potencies of the receptor subunits. This model was then directly used to predict ADCP responses for AglycoT-Fc1004 with both SKOV-3 cells (A) and MDA-MB-453 cells (B). Data from any two of the Fc variants can be selected and the ADCP response for the third variant can be accurately predicted. ADCP data for (C, D) Herceptin and AglycoT-Fc1004 or (E, F) AglycoT-Fc1001 and AglycoT-Fc1004 were used to obtain intrinsic signaling potency values in the model, which was then used to predict ADCP responses for AglycoT-Fc1001 with SKOV-3 cells (C) and MDA-MB-453 cells (D), or Herceptin with SKOV-3 cells (E) and MDA-MB-453 cells (F).

Using the relative signaling potencies determined by the first fit (using experimental data for Herceptin and AglycoT-Fc1001 only), ADCP responses of the other Fc variants were predicted for both SKOV-3 and MDA-MB-453 cells (Fig. 2.7). Since ADCP data for those Fc variants were not quantified, quantitative comparisons between model and experimental results were not possible. Nevertheless, qualitatively,

the very low ADCP response exhibited by AglycoT-Fc5-2a has been observed experimentally (G. Georgiou, personal communication). The predicted phagocytosis level for the two other variants, AglycoT-Fc1002 and AglycoT-Fc1003, did not exceed that for wild-type Herceptin, which is in agreement with our initial hypothesis after SPR analysis when those variants were not chosen for further characterization. AglycoT-Fc1004 exhibits the highest predicted ADCP level, also in agreement with experimental results (Fig. 2.4).
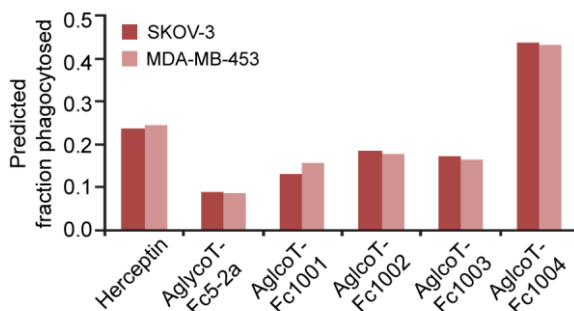
Figure 2.7. Predicted phagocytic responses of other Fc variants. Based on $K_{D\_IIa-H131}$, $K_{D\_IIa-R131}$, and K_{D_IIb} values from SPR analysis (Table 2.1), phagocytic response was predicted by the mathematical model for both SKOV-3 cells (dark red bar) and MDA-MB-453 cells (light red bar). Intrinsic signaling potencies for FcγRIIa-H131, FcγRIIa-R131 and FcγRIIb were the same as in Fig. 2.6A, B (2.8, 0.4, and -1 respectively).

2.4.3 Activation signaling and ADCP response is dominated by FcγRIIa-H131 affinity

The model provides a framework for understanding the molecular basis for enhanced ADCP response. From SPR analysis (Table 2.1), the most potent variant, AglycoT-Fc1004, has the highest affinity for all three FcγRII subunits, but it has a lower $A_{IIa-H131}$/I

ratio than Herceptin. To determine whether absolute affinity or A/I ratio plays a more important factor in physiological response, the model was used to plot predicted phagocytic response for various $K_{D\_IIa-H131}$, $K_{D\_IIa-R131}$, and K_{D_IIb} values (Fig. 2.8). In Fig. 2.8A, K_{D_IIb} was held constant at 1 µM, which is the same order of magnitude as all experimentally determined K_{D_IIb} values. The IgG variants examined in this work were overlaid on this heat map on the basis of their $K_{D\_IIa-H131}$ and $K_{D\_IIa-R131}$ values to show their approximate expected ADCP responses (see Fig. 2.7 for actual predictions). The simulations show that enhanced ADCP most dominantly correlates with increased affinity to FcγRIIa-H131, with minimal dependence on FcγRIIa-R131 affinity. This is due to the fact that the intrinsic signaling potency of FcγRIIa-H131 is predicted to be significantly stronger than that of FcγRIIa-R131, which is, in turn, weaker than FcγRIIb (Table 2.3). This prediction is also consistent with clinical observations that FcγRIIa-R131 homozygous individuals are more susceptible to bacterial infections and autoimmune diseases[16–18], perhaps due to inefficient generation of activation signals in the absence of FcγRIIa-H131 subunits.
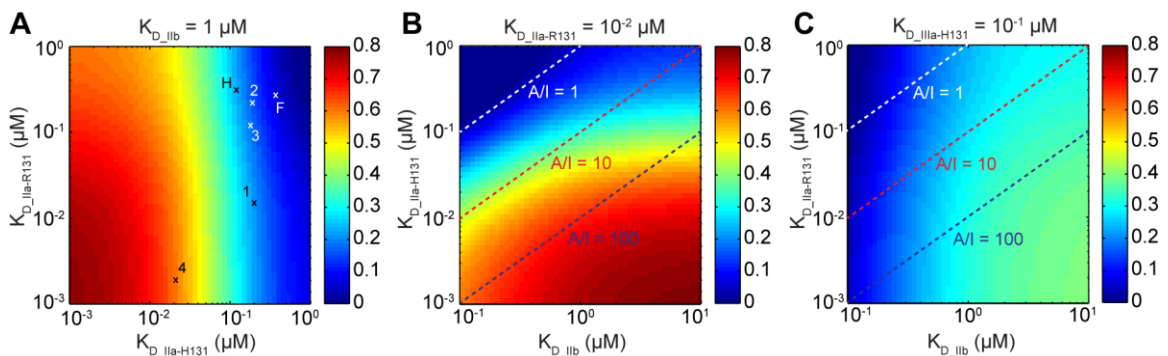
Figure 2.8. Mathematical modeling provides insights into design rules governing phagocytic potency of Fc variants. Predicted fraction of SKOV-3 cells phagocytosed for a range of $K_{D\_IIa-H131}$, $K_{D\_IIa-R131}$, and K_{D_IIb}

values. (A) K_{D_IIb} was held constant at 1 μM, and $K_{D\_IIa-H131}$ and $K_{D\_IIa-R131}$ values were varied. Phagocytic responses for Herceptin (H), AglycoT-Fc5-2a (F), AglycoT-Fc1001 (1), AglycoT-Fc1002 (2), AglycoT-Fc1003 (3), and AglycoT-Fc1004 (4) were predicted on the basis of their experimentally determined equilibrium binding constants to FcγRIIa-H131 and FcγRIIa-R131 (see Table 2.1). (B) $K_{D\_IIa-R131}$ was held constant at 0.01 μM and K_{D_IIb} and $K_{D\_IIa-H131}$ values were varied: fixed $A_{IIa-H131}/I$ (i.e., $K_{D\_IIb}/K_{D\_IIa-H131}$) ratios of 1 (white dash), 10 (red dash), and 100 (blue dash) are shown. (F) $K_{D\_IIa-H131}$ was held constant at 0.1 μM, and K_{D_IIb} and $K_{D\_IIa-R131}$ values were valued. The same ratios of $A_{IIa-R131}/I$ affinity as in panel B are shown.

2.4.4 Implications for future Fc engineering

We further used our mathematical model to elucidate design rules for future engineering of Fc variants. It has been suggested that effector response can be predicted by the A/I ratio, defined as the ratio of IgG Fc affinity for FcγRIIa to that for FcγRIIb[6]. To investigate the utility of A/I ratio in predicting phagocytosis, lines of constant $A_{IIa-H131}/I$ or $A_{IIa-R131}/I$ were plotted for fixed $K_{D\_IIa-R131}$ and $K_{D\_IIa-H131}$, respectively in Figure 2.8 (i.e., $K_{D\_IIb}/K_{D\_IIa-H131}$ in Fig. 2.8B and $K_{D\_IIb}/K_{D\_IIa-R131}$ in Fig. 2.8C). Even though a higher A/I ratio tends to elicit a higher effector response, phagocytic activity can vary significantly even for a fixed A/I ratio, indicating that absolute receptor affinities also play a significant, if not dominant, role in shaping cellular response. Interestingly, for a fixed A/I ratio, increasing FcγRIIa-H131 affinity generally leads to increased phagocytosis (Fig. 2.8B), while increasing FcγRIIa-R131 affinity leads to decreased phagocytosis (Fig. 2.8C). This can be explained by the relative intrinsic signaling potencies of the receptor subunits (Table 2.3). At a fixed $A_{IIa-H131}/I$, increasing the affinity for FcγRIIa-H131 (potency > 1) would also increase the affinity for FcγRIIb

(potency = −1), but the net effect would be increased activation due to the differential potencies. By contrast, at a fixed $A_{IIa\text{-}R131}/I$, increasing affinities for both subunits would reduce ADCP response because FcγRIIb is more potent at signaling than FcγRIIa-R131 (potency < 1). Our findings agree with recent crystal structure data that suggest the FcγRIIa-R131 polymorphism adopts a conformation that maybe less favorable for multimerization than FcyRIIa-H131, therefore limiting the proximity of ITAM domains required to transduce activation signals[52]. The model suggests that efforts to improve the design of IgG Fc variants for enahnaced ADCP should first focus on increasing affinity for FcγRIIa-H131 (Fig. 2.8A and B) and, if not possible or insufficient, then on decreasing affinity for FcγRIIb (Fig. 2.8B and C). If binding to the low affinity FcγRIIa-R131 allele is required, then both improved FcγRIIa-R131 affinity and A/I ratio are necessary for more potent ADCP (Fig. 2.8C). While FcγRIIa and IIb have been established to be by far the most important receptors for phagocytosis by tumor cells[23,53], the inclusion of FcγRI and FcγRIIIa in the model could be used to determine design criteria for glycosylated antibodies or future aglycosylated variants with more complex FcγR mediated effector function.

2.5 Discussion

In this work, a bacterial display system has been developed to display full length IgG1 antibodies tethered to the inner membrane of *E. coli*. Full length IgG, instead of only the human Fc domain and hinge, was displayed because earlier work showed that binding of

aglycosylated Fc2a mutant, which had been engineered to bind FcγRIIa/FcγRIIb[24], was impaired when displayed in the homodimeric Fc display system[26]. Since aglycosylated antibodies are displayed, a more complex display system capable of glycan processing is not necessary. The selected antibodies can also be more readily transferred to expression in large quantities by *E. coli* following successful selections. This system has the additional advantage of utilizing FACS, a high-throughput screening method allowing up to 10^8 clones/round to be screened. This elegant display system was combined with random mutagenesis to generate a diversity of sequences. Ultimately, just two point mutations were able to engender up to 5.7-fold and 160-fold increases in affinity for the FcγRIIa-H131 and -R131 alleles, respectively, as well as a 75% increase in tumor cell phagocytosis by macrophages (Fig. 2.2A). This general theme of library construction to generate diversity, followed by screening or selection to achieve desired protein properties, will be seen frequently in the next few chapters.

The generation of highly FcγR selective Fc domains might allow for directed recruitment of certain immune effector cells, such as macrophages as shown here (because AglycoT-Fc1004 does not engage FcγRIIIa and has 50-fold lower affinity for FcγRI). So far, aglycosylated antibodies have been used for applications where inflammatory engagement of FcγRs or C1q complement proteins is undesirable[30]. The generation of aglycosylated antibodies with unique affinities and selectivities may open the possibility for a new class of therapeutics with intriguing biological properties. Fc engineered aglycosylated IgGs offer additional advantages by bypassing the need for N-linked glycosylation; eliminating glycan heterogeneity and broadening the range of

expression systems that can be used to produce large quantities of antibody. In several recent clinical trials, no evidence of aglycosylated antibody immunogenicity has been reported[30]. However, at this point it is unclear whether mutations to an aglycosylated structure will induce immunogenicity either through residue change or the generation of non-native conformational epitopes. The results from these studies will be reported in the future.

The present work has focused on improving phagocytosis using FcγRIIa-R131, instead of FcγRIIa-H131, especially since patients homozygous for FcγRIIa-R131 have an increased risk of bacterial infection and autoimmunity[16–19]. The screens were designed and set up to only enrich variants displaying improved FcγRIIa-R131/FcγRIIb selectivity (or $A_{IIa-R131}/I$ ratio), by using a progressively increasing excess of unlabeled FcγRIIb over fluorescent FcγRIIa-R131 conjugate. Interestingly, our mathematical model suggests that the increased phagocytic response from the best mutant, AglycoT-Fc1004, is predominantly attributed to increased affinity to FcγRIIa-H131. While AglycoT-Fc1004 exhibited a 5.7-fold enhancement in affinity to FcγRIIa-H131 and slight decrease in $A_{IIa-H131}/I$ ratio (0.88), leading to a 75% increase in ADCP, these properties only emerged by coincidence during enrichment for $A_{IIa-R131}/I$. By comparison, AglycoT-Fc1004 displayed a 163-fold increase in FcγRIIa-R131 affinity, with a 25-fold improvement in $A_{IIa-R131}/I$ due to the direct screen. This favorably suggests that a future screen with the same display system set up instead to use fluorescent FcγRIIa-H131 conjugate would likely achieve a significantly greater improvement in FcγRIIa-H131 affinity and $A_{IIa-H131}/I$ selectivity. Since our model predicts that these two effects alone are dominant in

modulating ADCP (Fig. 2.8A, B and Table 2.3), this new screening strategy should yield variants with ADCP responses greater than that elicited by AglycoT-Fc1004. In addition, such Fc variants could be used much more broadly in treatment, since 74% of African Americans and 81% of Caucasian Americans are FcγRIIa-H131 homozygous or FcγRIIa-H131/R131 heterozygous[20].

Here, a mechanistic model has been used after directed evolution experiments to successfully link the biochemical properties of a clone to its consequent cellular response. Specifically, the model predicts ADCP from binding affinities to multiple FcγRII receptors. This mathematical framework can be further generalized to include other FcγRs to model additional effector functions. The initial goal in constructing the model was to quantitatively capture the diverse, non-intuitive behaviors of the selected mutants, but it also provided unexpected biological insights, leading to the identification of a new, more potent target for future engineering of Fc domains.

2.6 References

1. Nimmerjahn, F. & Ravetch, J. V. Antibody-mediated modulation of immune responses. *Immunological Rev.* **236**, 265–275 (2010).

2. Smith, K. G. C. & Clatworthy, M. R. FcγRIIB in autoimmunity and infection: evolutionary and therapeutic implications. *Nat. Rev. Immunol.* **10**, 328–343 (2010).

3. Bruhns, P. *et al.* Specificity and affinity of human Fcγ receptors and their polymorphic variants for human IgG subclasses. *Blood* **113**, 3716–3725 (2009).

4. Lazar, G. A. *et al.* Engineered antibody Fc variants with enhanced effector function. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4005–4010 (2006).

5.   Stavenhagen, J. B. *et al.* Fc optimization of therapeutic antibodies enhances their ability to kill tumor cells *in vitro* and controls tumor expansion *in vivo* via low-affinity activating Fcγ receptors. *Cancer Res.* **67**, 8882–8890 (2007).

6.   Nimmerjahn, F. & Ravetch, J. V. Divergent immunoglobulin G subclass activity through selective Fc receptor binding. *Science* **310**, 1510–1512 (2005).

7.   Rankin, C. T. *et al.* CD32B, the human inhibitory Fc-γ receptor IIB, as a target for monoclonal antibody therapy of B-cell lymphoma. *Blood* **108**, 2384–2391 (2006).

8.   Cartron, G. *et al.* Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor FcγRIIIa gene. *Blood* **99**, 754–758 (2002).

9.   Weng, W.-K. & Levy, R. Two immunoglobulin G fragment C receptor polymorphisms independently predict response to rituximab in patients with follicular lymphoma. *J. Clin. Oncol.* **21**, 3940–3947 (2003).

10.   Shields, R. L. *et al.* High resolution mapping of the binding site on human IgG1 for FcγRI, FcγRII, FcγRIII, and FcRn and design of IgG1 variants with improved binding to the FcγR. *J. Biol. Chem.* **276**, 6591–6604 (2001).

11.   Umana, P., Jean-Mairet, J., Moudry, R., Amstutz, H. & Bailey, J. E. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. *Nat. Biotechnol.* **17**, 176–180 (1999).

12.   Nordstrom, J. *et al.* Anti-tumor activity and toxicokinetics analysis of MGAH22, an anti-HER2 monoclonal antibody with enhanced Fcγ receptor binding properties. *Breast Cancer Res.* **13**, R123 (2011).

13.   Paz-Ares, L. G. *et al.* Phase I pharmacokinetic and pharmacodynamic dose-escalation study of RG7160 (GA201), the first glycoengineered monoclonal antibody against the epidermal growth factor receptor, in patients with advanced solid tumors. *J. Clin. Oncol.* **29**, 3783–3790 (2011).

14.   Zhang, W. *et al.* FCGR2A and FCGR3A polymorphisms associated with clinical outcome of epidermal growth factor receptor–expressing metastatic colorectal cancer patients treated with single-agent cetuximab. *J. Clin. Oncol.* **25**, 3712–3718 (2007).

15.   Tamura, K. *et al.* FcγR2A and 3A polymorphisms predict clinical outcome of trastuzumab in both neoadjuvant and metastatic settings in patients with HER2-positive breast cancer. *Ann. Oncol.* **22**, 1302–1307 (2011).

16. Yee, A. M. F., Phan, H. M., Zuniga, R., Salmon, J. E. & Musher, D. M. Association between FcγRIIa-R131 allotype and bacteremic pneumococcal pneumonia. *Clin. Infect. Dis.* **30**, 25–28 (2000).

17. Sanders, L. A. *et al.* Human immunoglobulin G (IgG) Fc receptor IIA (CD32) polymorphism and IgG2-mediated bacterial phagocytosis by neutrophils. *Infect. Immun.* **63**, 73–81 (1995).

18. Shashidharamurthy, R. *et al.* Dynamics of the interaction of human IgG subtype immune complexes with cells expressing R and H allelic forms of a low-affinity Fcγ receptor CD32A. *J. Immunol.* **183**, 8216–8224 (2009).

19. Karassa, F. B., Trikalinos, T. A. & Ioannidis, J. P. A. Role of the Fcγ receptor IIa polymorphism in susceptibility to systemic lupus erythematosus and lupus nephritis: A meta-analysis. *Arthritis Rheum.* **46**, 1563–1571 (2002).

20. Reilly, A. F. *et al.* Genetic diversity in human Fc receptor II for immunoglobulin G: Fcγ receptor IIA ligand-binding polymorphism. *Clin. Diagn. Lab. Immunol.* **1**, 640–644 (1994).

21. Sullivan, K. E. *et al.* Analysis of polymorphisms affecting immune complex handling in systemic lupus erythematosus. *Rheumatology* **42**, 446–452 (2003).

22. Ruffell, B. *et al.* Leukocyte composition of human breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2796–2801 (2012).

23. Richards, J. O. *et al.* Optimization of antibody binding to FcγRIIa enhances macrophage phagocytosis of tumor cells. *Mol. Cancer Ther.* **7**, 2517–2527 (2008).

24. Sazinsky, S. L. *et al.* Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20167–20172 (2008).

25. Jefferis, R. Recombinanat antibody therapeutics: the impact of glycosylation on mechanisms of action. *Trends Pharmacol. Sci.* **30**, 356–362 (2009).

26. Jung, S. T. *et al.* Aglycosylated IgG variants expressed in bacteria that selectively bind FcγRI potentiate tumor cell killing by monocyte-dendritic cells. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 604–609 (2010).

27. Simmons, L. C. *et al.* Expression of full-length immunoglobulins in *Escherichia coli*: rapid and efficient production of aglycosylated antibodies. *J. Immunol. Methods* **263**, 133–147 (2002).

28. Krapp, S., Mimura, Y., Jefferis, R., Huber, R. & Sondermann, P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J. Mol. Biol.* **325**, 979–989 (2003).

29. Borrok, M. J., Jung, S. T., Kang, T. H., Monzingo, A. F. & Georgiou, G. Revisiting the role of glycosylation in the structure of human IgG Fc. *ACS Chem. Biol.* **7**, 1596–1602 (2012).

30. Jung, S. T., Kang, T. H., Kelton, W. & Georgiou, G. Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. *Curr. Opin. Biotechnol.* **22**, 858–867 (2011).

31. Sibéril, S., Dutertre, C.-A., Fridman, W.-H. & Teillaud, J.-L. FcγR: The key to optimize therapeutic antibodies? *Crit. Rev. Oncol. Hematol.* **62**, 26–33 (2007).

32. Goldstein, B., Faeder, J. R. & Hlavacek, W. S. Mathematical and computational models of immune-receptor signalling. *Nat. Rev. Immunol.* **4**, 445–456 (2004).

33. Rhodes, A., Jasani, B., Anderson, E., Dodson, A. R. & Balaton, A. J. Evaluation of HER-2/neu immunohistochemical assay sensitivity and scoring on formalin-fixed and paraffin-processed cell lines and breast tumors: a comparative study involving results from laboratories in 21 countries. *Am. J. Clin. Pathol.* **118**, 408–417 (2002).

34. Stavenhagen, J. B. *et al.* Enhancing the potency of therapeutic monoclonal antibodies via Fc optimization. *Adv. Enzyme Regul.* **48**, 152–164 (2008).

35. Harvey, B. R. *et al.* Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9193–9198 (2004).

36. Jung, S. T. *et al.* Effective phagocytosis of low Her2 tumor cell lines with engineered, aglycosylated IgG displaying high FcγRIIa affinity and selectivity. *ACS Chem. Biol.* **8**, 368–375 (2013).

37. Pabbisetty, K. B. *et al.* Kinetic analysis of the binding of monomeric and dimeric ephrins to Eph receptors: Correlation to function in a growth cone collapse assay. *Protein Sci.* **16**, 355–361 (2007).

38. Hinton, P. R. *et al.* An engineered human IgG1 antibody with longer serum half-life. *J. Immunol.* **176**, 346–356 (2006).

39. Moasser, M. M., Basso, A., Averbuch, S. D. & Rosen, N. The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER2-overexpressing tumor cells. *Cancer Res.* **61**, 7184–7188 (2001).

40. Reim, F. *et al.* Immunoselection of breast and ovarian cancer cells with trastuzumab and natural killer cells: selective escape of CD44high/CD24low/ HER2low breast cancer stem cells. *Cancer Res.* **69**, 8058–8066 (2009).

41. Narayan, M. *et al.* Trastuzumab-induced HER reprogramming in "resistant" breast carcinoma cells. *Cancer Res.* **69**, 2191–2194 (2009).

42. Zhang, X. *et al.* Follicle-stimulating hormone peptide can facilitate paclitaxel nanoparticles to target ovarian carcinoma in vivo. *Cancer Res.* **69**, 6506–6514 (2009).

43. Krombach, F. *et al.* Cell size of alveolar macrophages: an interspecies comparison. *Environ. Health Perspect.* **105**, 1261–1263 (1997).

44. Pease, L. F., Elliott, J. T., Tsai, D.-H., Zachariah, M. R. & Tarlov, M. J. Determination of protein aggregation with differential mobility analysis: application to IgG antibody. *Biotechnol. Bioeng.* **101**, 1214–1222 (2008).

45. Sondermann, P., Huber, R. & Jacob, U. Crystal structure of the soluble form of the human Fcγ-receptor IIb: a new member of the immunoglobulin superfamily at 1.7 Å resolution. *EMBO J.* **18**, 1095–1103 (1999).

46. Kuo, S. C. & Lauffenburger, D. A. Relationship between receptor/ligand binding affinity and adhesion strength. *Biophys. J.* **65**, 2191–2200 (1993).

47. Heiple, J. M., Wright, S. D., Allen, N. S. & Silverstein, S. C. Macrophages form circular zones of very close apposition to IgG-coated surfaces. *Cell Motil. Cytoskeleton* **15**, 260–270 (1990).

48. Tolentino, T. P. *et al.* Measuring diffusion and binding kinetics by contact area FRAP. *Biophys. J.* **95**, 920–930 (2008).

49. Costantini, D. L., Bateman, K., McLarty, K., Vallis, K. a & Reilly, R. M. Trastuzumab-resistant breast cancer cells remain sensitive to the auger electron-emitting radiotherapeutic agent ^{111}In-NLS-trastuzumab and are radiosensitized by methotrexate. *J. Nucl. Med.* **49**, 1498–1505 (2008).

50. Xu, F. J. *et al.* Radioiodinated antibody targeting of the Her-2/neu oncoprotein. *Nucl. Med. Biol.* **24**, 451–459 (1997).

51. Maenaka, K., Van der Merwe, P. a, Stuart, D. I., Jones, E. Y. & Sondermann, P. The human low affinity Fcγ receptors IIa, IIb, and III bind IgG with fast kinetics and distinct thermodynamic properties. *J. Biol. Chem.* **276**, 44898–44904 (2001).

52. Ramsland, P. A. *et al.* Structural basis for FcγRIIa recognition of human IgG and formation of inflammatory signaling complexes. *J. Immunol.* **187**, 3208–3217 (2011).

53. Indik, Z., Kelly, C., Chien, P., Levinson, A. I. & Schreiber, A. D. Human FcγRII, in the absence of other Fc gamma receptors, mediates a phagocytic signal. *J. Clin. Invest.* **88**, 1766–1771 (1991).

CHAPTER 3

Model-guided ligation strategy for optimal assembly of DNA libraries

Since the success of a directed evolution experiment depends heavily on library construction, we begin our discussion on directed evolution by first investigating ways to improve DNA library assembly by optimizing the ligation reaction.

3.1 Introduction

DNA ligation is indispensible in molecular biology research. The ligation reaction, usually catalyzed by a DNA ligase, consists of joining DNA molecule ends by creating a phosphodiester bond between the 3' hydroxyl and the 5' phosphate of adjacent nucleotides[1]. Maximizing ligation efficiency is important in systems where the vector or insert molecule is scarce or where high fidelity of ligated product is required. A prime example in which poor ligation efficiency can jeopardize success is in directed evolution experiments, because the quality and size of the protein library greatly depend on the extent and accuracy of DNA ligation during initial library construction as well as between selection rounds[2].

However, methods for optimizing the ligation reaction are extremely limited and typically rely on extensive trial and error through experimentation. The plethora of ligation guidelines[a] and web calculators[b] do not provide any optimization capabilities or information beyond basic recommendations and unit conversions. Most current ligation calculators perform calculations for the traditional circular vector-insert ligations, allowing input of mass concentrations, vector:insert molar ratios and fragment lengths while outputting the volumes of vector and insert stock solution to be used in the ligation reaction[3]. There is no precise control over the extent of product formation because different experiments utilize ends with different properties (e.g. palindromic vs. non-palindromic) and sequences of overhang bases, as well as different lengths of insert and vector. Also, since ligation and transformation are often coupled, the troubleshooting space is exponentially increased. Furthermore, optimization requires extensive work and has to be repeated anytime the system is varied. Thus, a more generic but easily altered strategy is needed to improve DNA ligation, and mathematical modeling provides a promising approach.

While mathematical models for ligation were developed decades ago[4–7], these models are either too specific to be applied in different experimental systems or too general to offer significant improvement over the universal guidelines. In addition, by

---

[a] https://www.neb.com/tools-and-resources/troubleshooting-guides/troubleshooting-tips-for-ligation-reactions
http://www.promega.com/~/media/Files/Resources/Protocols/Product%20Information%20Sheets/G/T4%20DNA%20Ligase%20Blue%20White%20Cloning%20Qualified%20Protocol.pdf

[b] http://www.insilico.uni-duesseldorf.de/Lig_Input.html
http://www.molecularstation.com/calculators/laboratory-research/ligation-insert-vector-calculator/
http://django.gibthon.org/tools/ligcalc/

assuming irreversibility in the ligation reactions, the models mentioned above can only provide distribution of various species with respect to products but not reactants. Advances in molecular biology, specifically the extensive use of polymerase chain reaction (PCR) and the availability of Type IIS restriction enzymes[8], necessitates the development of new models that accommodate ligation reactions with linear fragments as well as non-palindromic ends.

Most commercially available enzymes for restriction digest belong to the Type IIP (palindromic) family, which recognize, and cleave within, symmetric DNA sequences[9,10]. Using palindromic ends generated by these enzymes complicates downstream intermolecular ligations because palindromic ends in two identical molecules can anneal, thus competing with intermolecular ligations that form the product of interest. To avoid such byproducts, Type IIS restriction enzymes can be used. Type IIS enzymes cleave at asymmetric target sites, some away from their recognition sites[9], allowing custom design of sticky ends. The advantages of rationally designing the sequences of the sticky ends include eliminating undesired intramolecular or intermolecular ligations that lead to unwanted species, increasing ligation fidelity by forcing directional cloning, and optimizing ligation efficiency by matching the free energy of annealing at sticky ends, among others[8]. In addition, it is possible to generate multiple different sticky ends in a single digestion reaction with one Type IIS enzyme, streamlining restriction digest protocols. Several Type IIS enzymes (e.g. BbsI, BsaI, BsmBI and SapI) are suitable candidates for routine cloning in replacement of Type IIP enzymes because: (1) they do not require two copies of their target sequences before

cleavage, (2) they have high fidelity in the number of bases between recognition and cutting sites, and (3) they cleave relatively close to the recognition sites[11]. In this study, we used BbsI as a model Type IIS restriction enzyme for comparison of ligation events with Type IIP enzymes NcoI and HindIII.

Here, we redesigned the original ribosome display vector[12] with two BbsI cutting sites adjacent to the NcoI or HindIII cutting sites that flank the insert and then used this second-generation display vector (pRDV2) to directly compare ligation strategies. We tested ligation of the sequence-of-interest, Off7[12], to either doubly digested pRDV2 (circular ligation) or PCR-generated and digested fragments of the upstream and downstream ends of pRDV2 (linear ligation). We demonstrate that circular and linear ligation following digestion with BbsI give up to 15- and 2.6-fold increases, respectively, in the yield of desired product compared with digestion with NcoI and HindIII. Based on Type IIS sticky ends, an easily parameterized thermodynamic model was developed to rationally guide ligation optimization. Our model offers significant advancement over previous models because not only does our model provide thermodynamic insight and general guidance in ligation optimization, but it also allows custom characterization (e.g., different ligase, sticky ends, and ligation temperature) if needed. The only parameters in the model are the free energies of ligation, one for each sticky end. To demonstrate the utility of our model, we first experimentally determined ΔG_1 and ΔG_2 for two sticky ends, CGAA and GCTT, and then computationally predicted the relative abundance of ligation products from three-piece linear ligation reactions and transformation yields from circular ligation reactions. All these predictions were successfully validated

64

experimentally. Finally, to demonstrate how our model may be useful in directed evolution applications, we performed simulations that provide guidelines for optimizing ligations for both *in vivo* and *in vitro* display technologies.

We have further expanded this thermodynamic model for general use in ligation by incorporating an estimator of free energy for sticky ends of any sequence and length (including mismatches) and by porting the entire model online (https://alliance.seas.upenn.edu/~caslab/ligcal) for free and easy access by the scientific community. Our web server estimates the free energy of ligation for specific sticky end sequences from DNA thermodynamics, while also providing free energies of possible mismatches, to rationally guide the design of highly specific sticky end sequences to be generated from Type IIS restriction digests. Once specific sequences are identified, the estimated free energies can be used directly, or they can be determined from a simple two-piece linear ligation experiment, which has the additional advantage of making no assumptions about the ligation conditions (e.g., ligase type, ligation temperature, and salt concentration). With the sticky end free energies, the distribution of ligation products can be predicted in two-piece linear ligation, three-piece linear ligation, or vector-insert circular ligation. This information can be used to estimate library diversity, library size, or the relative number of colonies after transformation; it can also help in deciding between purification schemes or determining if there is enough product for the next step. Moreover, our ligation calculator can predict product formation over a range of input concentrations and molar ratios of reactants to allow custom optimization of the reaction, whether it be to maximize library diversity or to maximize reaction efficiency. Finally, a

ligation worksheet is available through the web site to facilitate conversion between mass concentrations and molar concentrations, as well as to formulate the ligation reaction. Overall, this customizable, model-guided strategy and web-based ligation calculator provides a means to rapidly identify the globally optimum conditions for multi-parametric ligation reactions in order to maximize formation of the desired product and, thus, library size.

3.2 Materials and methods

3.2.1 DNA manipulations

T4 DNA ligase (#M0202L) was used in all ligation reactions with the supplied $1\times$ T4 ligase reaction buffer (New England Biolabs; NEB). Ligation reactions were carried out with 200 cohesive end units per 10 μl reaction at 25°C for 3 h, followed by heat inactivation at 65°C for 10 min. All restriction and nicking enzymes were purchased from NEB. Ten-fold overdigestion was used in all digestion and nicking reactions by adjusting time and units of enzymes used. Phusion high-fidelity DNA polymerase (NEB) was used in all PCR reactions as recommended by the manufacturer. Qiagen PCR purification or gel extraction kit was used to purify or isolate DNA between the above steps following the standard procedure. DNA concentrations were determined by absorbance readings at 260 nm using a spectrophotometer (NanoDrop 1000, Thermo Scientific).

3.2.2 Plasmid construction

Primers used for DNA amplification (Integrated DNA Technologies) are listed in Table 3.1. The initial template for further plasmid engineering was pRDVstops:Off7[13], which is a modified version of the original ribosome display vector pRDV[12] that contains the off7 ankyrin gene. The first BbsI site was introduced by amplifying pRDVstops:Off7 with primers 3/4, digesting with BseYI and NcoI, and ligating into similarly cut pRDVstops-Off7. The ligation mixture was transformed into chemically competent *E. coli* XL1-Blue cells[14]. The resulting plasmid pRDV1b:Off7 was isolated with a Qiagen miniprep kit following the standard procedure. The second BbsI site was introduced by amplifying pRDVstops:Off7 with primers 5/6 and cloning into pRDV1b:Off7 via the HindIII and NotI sites, giving rise to pRDV2:Off7 (Fig. 3.1). To replace off7 with β-lactamase to obtain pRDV2:Blac, the insert from NcoI- and HindIII-digested pRDV-Blac was ligated into the vector from similarly digested pRDV2:Off7. All clones were confirmed by DNA sequencing.

Table 3.1. Primers used in this study.

Primer Number	Primer Name	Nucleotide Sequence (5' → 3') [a]
1	T7B_noBsaI	ATACGAAATTAATACGACTCACTATAGGGACACCACAACGG
2	tolAk	CCGCACACCAGTAAGGTGTGCGGTTTCAGTTGCCGCTTTCTTTCT
3	BbsIsite1_BseYI_f	GTTCGTGCACACAGCCCAGC
4	BbsIsite1_NcoI_r	TCCGCCATGGATATATCTCCTTCTTTCGGTGTCTTCAAATTATTTCTATTG
5	BbsIsite2_HindIII_f	GCAAAAGCTTGGGTCTTCCGGTGGCCAGAAGCAAGC

6	BbsIsite2_NotI_r	TCATCTGCCTCTGCGGCCGC
7	pRDV_BbsI_f	AATAATTTGAAGACACCGAAAGAAG
8	pRDV_BbsI_r	GGCCACCGGAAGACCCAAGC
9	pRDV_NcoI_f	AGAAGGAGATATATCCATGG
10	pRDV_HindIII_r	GGCCACCAGATCCAAGCTT

[a] Restriction sites are underlined.
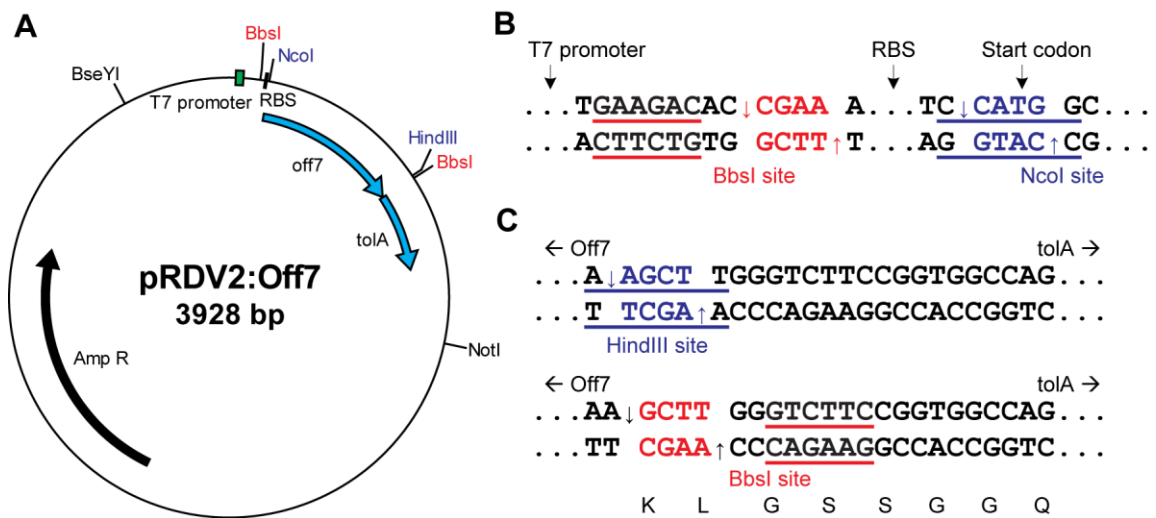
Figure 3.1. Schematic representation of pRDV2 vector design and NcoI, HindIII, and BbsI cutting sites. (A) Representative map of pRDV2:Off7 showing T7 promoter, ribosome binding site (RBS), off7 gene, tolA spacer, and the various restriction enzyme cutting sites. Primer 1 is a forward primer that anneals at the beginning of the T7 promoter and primer 2 is a reverse primer that anneals within tolA. Primers 3–10 anneal at their respective positions near restriction sites as indicated by the names of the primers (Table 3.1). (B) Partial DNA sequence of NcoI and BbsI restriction sites upstream of off7. The recognition sequences are underlined and the cutting sites are indicated by arrows. (C) Partial DNA sequence of HindIII (top) and BbsI (bottom) restriction sites downstream of off7. The two sequences are identical but are annotated separately for clarity. Amino acid translations in frame with Off7 are shown below the sequence.

3.2.3 Linear ligation

To obtain the three DNA fragments for linear ligation – the T7 region, off7 and the tolA region – pRDV2-Off7 was amplified by PCR with primers 1/4, 7/8 and 5/2, respectively. The PCR-amplified T7 region was digested with NcoI or BbsI, while off7 was digested with NcoI/HindIII or BbsI and tolA region was digested with HindIII or BbsI. Two- or three-piece linear ligation reactions were carried out as described above with the molar ratio and concentration of each species as specified in the experiment. Ligation products were resolved and visualized using gel electrophoresis. Agarose gels (1.5% in 1× TAE buffer) were stained with ethidium bromide and imaged using an ultraviolet transilluminator and CCD camera. Band intensities were quantified with Kodak molecular imaging software v4.5.0 using exACTGene 1 kb plus and 50 bp mini DNA ladders (Fisher Scientific) as standards. For each image, the relative intensities of each band were obtained by averaging quantification at three different aperture and exposure settings.

3.2.4 Circular ligation

pRDV2:Blac was digested with NcoI/HindIII or BbsI and then gel purified to obtain doubly digested linear pRDV2 vector. Circular ligation was carried out by ligating this vector to digested off7 insert (prepared as described above) at specific molar ratios and concentrations. Gel electrophoresis (1% agarose gels) was used to resolve and visualize ligation products. Ligation PCR was carried out by amplifying 75 ng ligation product

with primers 1/2 for 10 cycles. Transformations were performed using XL1-Blue cells as described above and the resulting colonies were counted manually.

3.2.5 Thermodynamic model: linear ligation

To guide ligation optimization more rationally, a mathematical model was developed that predicts product distributions based on input DNA concentrations and free energies of the ligation events. Nomenclature used in the thermodynamic model is provided in Table 3.2.

Table 3.2. Nomenclature used in the thermodynamic model.

Symbol	Description
[species]	Concentration of DNA species (in nM)
[species$_n$]	Concentration of nicked DNA species (in nM)
K_{D1}	Equilibrium dissociation constant for annealing of sticky end CGAA (in nM)
K_{D2}	Equilibrium dissociation constant for annealing of sticky end GCTT (in nM)
K_{lig1}	Equilibrium dissociation constant for the sealing of DNA phosphodiester bonds by T4 DNA ligase for sticky end CGAA (unitless)
K_{lig2}	Equilibrium dissociation constant for the sealing of DNA phosphodiester bonds by T4 DNA ligase for sticky end GCTT (unitless)
K_1	$K_{D1} \times K_{lig1}$. Lumped equilibrium dissociation constant for overall ligation of sticky end CGAA (in nM)
K_2	$K_{D2} \times K_{lig2}$. Lumped equilibrium dissociation constant for overall ligation of sticky end GCTT (in nM)
J	The Jacobson-Stockmayer factor[15]. Effective concentration of one end of a linear DNA molecule with respect to the other (in ends/ml)
[j]	Effective concentration of one end of a linear DNA molecule with respect to the other (in nM)
K_{Dcir1}	Equilibrium dissociation constant for the circularization annealing of sticky end CGAA (unitless)
K_{Dcir2}	Equilibrium dissociation constant for the circularization annealing of sticky end GCTT (unitless)
K_{cir1}	Lumped equilibrium dissociation constant for the overall circularization and ligation of sticky end CGAA (unitless)

K_{cir2}	Lumped equilibrium dissociation constant for the overall circularization and ligation of sticky end GCTT (unitless)

A thermodynamic model of linear ligation was developed by considering ligation to be a two-step reaction: annealing of sticky ends and sealing of the phosphodiester backbone. The thermodynamic reactions for 3-piece linear ligations are as follow:

$$A + B \xleftarrow{KD1}\rightarrow AB_n \xleftarrow{Klig1}\rightarrow AB$$

$$B + C \xleftarrow{KD2}\rightarrow BC_n \xleftarrow{Klig2}\rightarrow BC$$

$$AB + C \xleftarrow{KD2}\rightarrow ABC_{2n} \xleftarrow{Klig2}\rightarrow ABC_2$$

$$BC + A \xleftarrow{KD1}\rightarrow ABC_{1n} \xleftarrow{Klig1}\rightarrow ABC_1$$
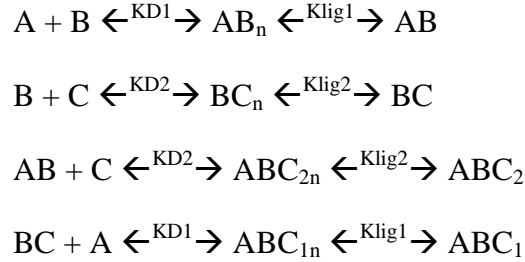
where K_{D1} and K_{D2} are the dissociation constants for annealing of sticky ends CGAA and GCTT, respectively, and K_{lig1} and K_{lig2} are equilibrium dissociation constants for the sealing of DNA phosphodiester bonds by T4 DNA ligase. We assumed that products sealed by ligase at one site could not re-nick at the other site (e.g., ABC_1 could not be converted into ABC_{2n}).

$$K_{D1} = [A][B]/[AB_n] \quad \text{or} \quad [BC][A]/[ABC_{1n}]$$

$$K_{D2} = [B][C]/[BC_n] \quad \text{or} \quad [AB][C]/[ABC_{2n}]$$

$$K_{lig1} = [AB_n]/[AB] \quad \text{or} \quad [ABC_{1n}]/[ABC_1]$$

$$K_{lig2} = [BC_n]/[BC] \quad \text{or} \quad [ABC_{2n}]/[ABC_2]$$

At equilibrium, the first reaction gives:

$$[A][B] - (K_{D1})(K_{lig1})[AB] = 0$$

and the mass balance of species A gives:

$$[A] + [AB](1+K_{lig1}) + [ABC_1](1+K_{lig1}) + [ABC_2](1+K_{lig2}) = [A]_t$$

To reduce the number of parameters, K_{D1} and K_{lig1} are multiplied to give the lumped parameter K_1 and, similarly, K_{D2} and K_{lig2} are lumped into parameter K_2. The ligation equilibrium constants K_{lig1} and K_{lig2} can be estimated by the standard free energy of ligation (ΔG^o) of -6.3 kcal/mol[16] and can thus be eliminated in the mass balance since ($1 + K_{ligx}$) ≈ 1 for $K_{ligx} << 1$ (where x = 1 or 2). The values of K_1 and K_2 were independently determined by two-piece linear ligation (described below). Applying these simplifications gives a system of seven equations:

Thermodynamic equilibrium equations:

$$[A][B] - K_1[AB] = 0 \tag{1}$$

$$[B][C] - K_2[BC] = 0 \tag{2}$$

$$[AB][C] - K_2[ABC_2] = 0 \tag{3}$$

$$[BC][A] - K_1[ABC_1] = 0 \tag{4}$$

Mass balance equations:

$$[A] + [AB] + [ABC_1] + [ABC_2] = [A]_t \tag{5}$$

$$[B] + [AB] + [BC] + [ABC_1] + [ABC_2] = [B]_t \tag{6}$$

$$[C] + [BC] + [ABC_1] + [ABC_2] = [C]_t \tag{7}$$

Based on input DNA concentrations $[A]_t$, $[B]_t$ and $[C]_t$, this system of equations was solved in Matlab using *fsolve* with tolerance value of 10^{-8} to obtain the unknowns $[A]$, $[B]$, $[C]$, $[AB]$, $[BC]$, and $[ABC]$. Molar ratios of the species were obtained with reference to product $[ABC]$ and mass ratios were calculated using the respective base pairs of the species. Percent incorporation (defined as full-length product over total amount of insert added) was calculated using $[ABC]/[B]_t \times 100\%$ and percent efficiency (defined as full-length product over total amount of reactants added) was calculated using

$[ABC]/([A]_t + [B]_t + [C]_t) \times 100\%$, where the subscript t represents the total amount of that species. For experimental determination, the mass fraction of each species was quantified on an agarose gel by densitometry and converted to absolute mass by multiplying each value by the total mass of the reactants. The absolute mass of correct product was then converted to moles and divided by the corresponding moles of insert or reactants for the given experimental conditions.

3.2.6 Thermodynamic model: circular ligation

The thermodynamic reactions for circular ligations are as follow:

$$V + I \xleftarrow{KD1}\rightarrow VI_{1n} \xleftarrow{Klig1}\rightarrow VI_1$$

$$V + I \xleftarrow{KD2}\rightarrow VI_{2n} \xleftarrow{Klig2}\rightarrow VI_2$$

$$VI_1 \xleftarrow{KDcir2}\rightarrow VIc_{2n} \xleftarrow{Klig2}\rightarrow VIc_2$$

$$VI_2 \xleftarrow{KDcir1}\rightarrow VIc_{1n} \xleftarrow{Klig1}\rightarrow VIc_1$$

$$VI_1 + I \xleftarrow{KD2}\rightarrow IVI_{2n} \xleftarrow{Klig2}\rightarrow IVI_2$$

$$VI_2 + I \xleftarrow{KD1}\rightarrow IVI_{1n} \xleftarrow{Klig1}\rightarrow IVI_1$$

$$VI_1 + V \xleftarrow{KD2}\rightarrow VIV_{2n} \xleftarrow{Klig2}\rightarrow VIV_2$$

$$VI_2 + V \xleftarrow{KD1}\rightarrow VIV_{1n} \xleftarrow{Klig1}\rightarrow VIV_1$$

$$IVI + V \xleftarrow{KD1}\rightarrow VIVI_{1an} \xleftarrow{Klig1}\rightarrow VIVI_{1a}$$

$$IVI + V \xleftarrow{KD2}\rightarrow VIVI_{2an} \xleftarrow{Klig2}\rightarrow VIVI_{2a}$$

$$VIV + I \xleftarrow{KD1}\rightarrow VIVI_{1bn} \xleftarrow{Klig1}\rightarrow VIVI_{1b}$$

$$VIV + I \xleftarrow{KD2}\rightarrow VIVI_{2bn} \xleftarrow{Klig2}\rightarrow VIVI_{2b}$$

$$VIVI_1 \xleftarrow{KDcir2}\rightarrow VIVIc_{2n} \xleftarrow{Klig2}\rightarrow VIVIc_2$$

73

$$VIVI_2 \xleftarrow{\text{KDcir1}} \rightarrow VIVIc_{1n} \xleftarrow{\text{Klig1}} \rightarrow VIVIc_1$$

where K_{D1}, K_{D2}, K_{lig1}, and K_{lig2} are equilibrium constants as defined above. K_{Dcir1} and K_{Dcir2} are equilibrium dissociation constants for circularization.

$$K_{Dcir1} = [VI_2]/[VIc_{1n}] \qquad \text{or} \qquad [VIVI_2]/[VIVIc_{1n}]$$

$$K_{Dcir2} = [VI_1]/[VIc_{2n}] \qquad \text{or} \qquad [VIVI_1]/[VIVIc_{2n}]$$

Analogous to what was done in the linear ligation model, K_{Dcir1} and K_{lig1} are multiplied to give the lumped parameter K_{cir1} and K_{Dcir2} and K_{lig2} are lumped into parameter K_{cir2}. To obtain K_{cir1} and K_{cir2}, the Jacobson-Stockmayer factor j was first calculated:

$$j = j_\lambda \, (MW_\lambda \, / \, MW)^{3/2}$$

where $j_\lambda = 3.6 \times 10^{11}$ ends/ml and $MW_\lambda = 30.8 \times 10^6$ g/mol[4]. By treating the annealing of V and I at one end followed by circularization at the other end as the binding of a bipartite ligand to a bipartite receptor, the circularization constants could be calculated as:

$$K_{cir1} = K_1/[j] \qquad \text{and} \qquad K_{cir2} = K_2/[j]$$

where [j] (mol/volume) is the effective concentration converted from j (ends/ml). Applying the same simplifications regarding K_{lig1} and K_{lig2}, the thermodynamic equilibrium equations for circular ligation can be written as follow:

$$[V][I] - K_1[VI_1] = 0 \tag{8}$$

$$[V][I] - K_2[VI_2] = 0 \tag{9}$$

$$[VI_1] - K_{cir2\_VI}[VIc_2] = 0 \tag{10}$$

$$[VI_2] - K_{cir1\_VI}[VIc_1] = 0 \tag{11}$$

$$[VI_1][I] - K_2[IVI_2] = 0 \tag{12}$$

$$[VI_2][I] - K_1[IVI_1] = 0 \tag{13}$$

$$[VI_1][V] - K_2[VIV_2] = 0 \tag{14}$$

$$[VI_2][V] - K_1[VIV_1] = 0 \qquad (15)$$

$$[IVI][V] - K_1[VIVI_{1a}] = 0 \qquad (16)$$

$$[IVI][V] - K_2[VIVI_{2a}] = 0 \qquad (17)$$

$$[VIV][I] - K_1[VIVI_{1b}] = 0 \qquad (18)$$

$$[VIV][I] - K_2[VIVI_{2b}] = 0 \qquad (19)$$

$$[VIVI_1] - K_{cir2\_VIVI}[VIVIc_2] = 0 \qquad (20)$$

$$[VIVI_2] - K_{cir1\_VIVI}[VIVIc_1] = 0 \qquad (21)$$

$$[V] + [VI] + [VIc] + [IVI] + 2([VIV] + [VIVI] + [VIVIc]) = [V]_t \qquad (22)$$

$$[I] + [VI] + [VIc] + 2[IVI] + [VIV] + 2([VIVI] + [VIVIc]) = [I]_t \qquad (23)$$

where $[VI] = [VI_1] + [VI_2]$, $[VIc] = [VIc_1] + [VIc_2]$, $[IVI] = [IVI_1] + [IVI_2]$, $[VIV] = [VIV_1] + [VIV_2]$, $[VIVI_1] = [VIVI_{1a}] + [VIVI_{1b}]$, $[VIVI_2] = [VIVI_{2a}] + [VIVI_{2b}]$, $[VIVI] = [VIVI_1] + [VIVI_2]$ and $[VIVIc] = [VIVIc_1] + [VIVIc_2]$.

This system of equations was solved in Matlab using *fsolve* with tolerance value of 10^{-8} to obtain the unknowns $[VIc]$ and $[VIVIc]$. Molar ratios of the species were obtained with reference to product $[VIc]$ at 1:1 I:V molar ratio and mass ratios were calculated using the respective base pairs of the species. Percent incorporation (defined as correctly ligated circular vector–insert product over total amount of insert added) was calculated using $[VIc]/[I]_t \times 100\%$, percent efficiency (defined as correctly ligated circular product over total amount of vector and insert added) was calculated using $[VIc]/([V]_t + [I]_t) \times 100\%$, and percent accuracy (defined as correctly ligated circular product over total circular product formed) was calculated using $[VIc]/([VIc] + [VIVIc]) \times 100\%$. Experimental values were calculated by dividing colony number by moles of insert or reactants at a fixed vector concentration (8 nM or 3 nM); these ratios were then divided

75

by the sum of all such experimental ratios and multiplied by the sum of all corresponding simulated values at the same vector concentration. This enabled direct comparison of experimental results and model simulations.

3.2.7 Training of thermodynamic model to determine ΔG_1 and ΔG_2

Two-piece linear ligation was carried out to experimentally determine ΔG_1 and ΔG_2 for the sticky ends CGAA and GCTT, respectively. To determine ΔG_1, BbsI-digested T7 (A) and off7 (B) were ligated at molar ratios of 1:1, 2:1, and 1:2 with 21, 32, and 53 nM off7. The product mass ratios for the above nine conditions were quantified from agarose gels (Table 3.3). Similar to three-piece ligation, the thermodynamic equilibrium equations for two-piece ligation can be written as follows:

$$[A][B] - K_1[AB] = 0 \tag{24}$$

while the mass balance equations are

$$[A] + [AB] = [A]_t \tag{25}$$

and $\quad [B] + [AB] = [B]_t \tag{26}$

This system of equations was solved with a range of ΔG_1 in Matlab using *fsolve* with tolerance value of 10^{-8} and compared to experimentally determined results. Least squares regression of ΔG_1 was carried out to two decimal places by minimizing the sum of the squared residuals. Similarly, to determine ΔG_2, BbsI-digested off7 (B) and tolA (C) were ligated at molar ratios of 1:1, 1:2, and 2:1 with 21, 32, and 53 nM off7 and the product mass ratios were quantified (Table 3.3). The system of equations can be written as follows:

76

$$[B][C] - K_2[BC] = 0 \tag{27}$$

$$[B] + [BC] = [B]_t \tag{28}$$

$$[C] + [BC] = [C]_t \tag{29}$$

Table 3.3. Experimentally determined mass ratios for training two-piece ligations.

Species	Concentration of off7 (nM)		
	21	**32**	**53**
Species	**Mass ratios of species**		
(T7:off7 = 1:1)			
T7	0.22	0.21	0.15
off7	0.49	0.32	0.44
T7-off7	1.00	1.00	1.00
(T7:off7 = 2:1)			
T7	0.20	0.33	0.27
off7	0.22	0.28	0.23
T7-off7	1.00	1.00	1.00
(T7:off7 = 1:2)			
T7	0.00	0.18	0.08
off7	0.97	0.64	0.64
T7-off7	1.00	1.00	1.00
(off7:tolA = 1:1)			
tolA	0.18	0.31	0.42
off7	0.16	0.27	0.20
off7-tolA	1.00	1.00	1.00
(off7:tolA = 1:2)			
tolA	0.53	0.88	1.16
off7	0.16	0.31	0.24
off7-tolA	1.00	1.00	1.00
(off7:tolA = 2:1)			

tolA	0.00	0.00	0.10
off7	0.53	0.32	0.24
off7-tolA	1.00	1.00	1.00

3.2.8 Correlation test

Correlation tests were conducted to determine whether model predictions were statistically significant in explaining the experimental results. For three-piece linear ligation, a correlation test was performed for each molar ratio (10:1:10, 2:1:2, 1:1:1, 1:2:1, and 1:10:1). The pairwise linear correlation coefficient for the mass ratios of the six species in the model and the corresponding mass ratios in three independent experiments was obtained by the *corr* function in Matlab's statistics toolbox. This value was then compared to 1000 correlation coefficients that were determined by randomly permuting the mass ratios of species in the model. The *p*-value was defined as the fraction of random correlation coefficients greater than the actual correlation coefficient. For vector-insert circular ligation, a correlation test was performed for each vector concentration (0.24, 3.02, and 6.04 nM) using the same method but by comparing molar ratios for the three different input insert:vector ratios in the model and the corresponding colony numbers in three independent experiments.

3.2.9 Ribosome display with pRDVstops-Off7 and pRDV2-Off7

The two ribosome display vectors, pRDVstops:Off7 and pRDV2:Off7, were PCR amplified with primers 1/2 and the products were used directly in *in vitro* transcription as

previously described[17]. Ribosome display selection particles were generated using the PURExpress *in vitro* protein synthesis kit (NEB) using mRNA as a template. For the translation reaction, 10 µl solution A, 7.5 µl solution B, 0.5 µl RNasin Plus, 2 µl off7 mRNA (1.4×10^{13} molecules) and 5 µl H_2O were mixed and incubated at $37^{\circ}C$ for 30 min. Translation was stopped with 420 µl ice-cold WBT buffer (50 mM Tris-acetate pH 7.5, 150 mM NaCl, 50 mM magnesium acetate, 0.05% Tween-20[17]) with 0.5% casein, and centrifuged at 20,000g for 5 min at $4^{\circ}C$. 100 µl supernatant was used in each well for binding experiment. Binding was performed using Maxisorp plates (NUNC, Thermo Fisher Scientific) prepared as follows: plates were coated with 100 µl 66 nM NeutrAvidin (Thermo Fisher Scientific) in Tris-buffered saline (TBS) overnight at $4^{\circ}C$ and then blocked with 0.5% casein (Sigma-Aldrich) in TBS at room temperature for 1 hr with shaking. For positive controls, wells were incubated with 1.83 µl 27.4 µM biotinylated maltose binding protein (MBP) in 200 µl TBS with 0.5% casein to immobilize the off7 target. Binding was performed at $4^{\circ}C$ for 1 hr with shaking. The plate was washed with WBT with 0.5% casein (three immediate washes and three 15-minute washes). Reverse transcription was performed using AffinityScript reverse transcriptase (Agilent Technologies) with reverse primers 10 and 8 for pRDVstops-Off7 and pRDV2-Off7, respectively. A modified *in situ* reverse transcription protocol[18] was used: 13.5 µl solution X (0.25 µl reverse primer, 12.75 µl H_2O and 0.5 µl RNasin Plus) was added to each well and the reaction was incubated at $65^{\circ}C$ for 5 min and placed on ice for 1 min. Then, 6.5 µl solution Y (2 µl 10x AffinityScript buffer, 2 µl 0.1 M DTT, 2 µl 5mM dNTPs and 0.5 µl AffinityScript reverse transcriptase) was added to each well and the reaction was incubated at $45^{\circ}C$ for 1 hr followed by $70^{\circ}C$ for 15 min and $4^{\circ}C$ for 7 min.

For PCR amplification, 2 μl of each reaction was taken and combined with primers 9/10 and 7/8 for pRDVstops:Off7 and pRDV2:Off7, respectively, for 25 cycles. Gel electrophoresis (1% agarose) was used to visualize the product from the ribosome display round.

3.3 Software development

3.3.1 Web server availability and implementation

The ligation calculator is freely available at https://alliance.seas.upenn.edu/~caslab/ligcal. The user interface is entirely web-based and all entry forms were generated by the standard HyperText Markup Language 5 (HTML5) with Cascading Style Sheets 2.1 (CSS 2.1). All underlying Common Gateway Interface (CGI) scripts were written in Python 2.7 (http://www.python.org) with two open-source scientific computing packages NumPy (http://www.numpy.org) and SciPy (http://www.scipy.org)[19], as well as the plotting library matplotlib (http://matplotlib.org)[20]. The interface has been tested on several web browsers, including Chrome 26, Firefox 16, Internet Explorer 10, and Safari 5.1.9.

3.3.2 Sticky end free energy calculator

Estimation of a sticky end free energy from DNA thermodynamics is performed by summing 1) the stacking energy arising from the two bases adjacent to the two nick

sites[21,22], 2) the hydrogen-bonding energy between complementary sticky ends[23], and 3) the energy for formation of the phosphodiester bonds[16]. If desired, the default stacking and phosphodiester bond energies in the online calculator can be changed or omitted entirely. Hydrogen bonding energies are computed using the parameters and methods for nearest-neighbor DNA thermodynamics[23], with the assumption that temperature and salt dependence act in opposite directions to the free energy and would thus negate the effect of one another[23]. Free energies for mismatched sticky ends are estimated similarly but from tabulated values of nearest-neighbor parameters for one or two mismatches in two base pair sequences. For one out of two mismatches, parameter values are obtained from literature for G•T[24], G•A[25], C•A[26], C•T[27], A•A[28], C•C[28], G•G[28], and T•T[28] mismatches. Since there are no literature values for the nearest-neighbor free energy change for two consecutive mismatches (except GG/TT, GT/TG and TG/GT[24]), we estimated their values by summing the free energy contributions from each of the two base pairs (Table 3.4). This assumes that each base pair contributes independently to the overall free energy, which is less accurate than the nearest-neighbor calculation in which neighboring bases contribute to the overall free energy, but is nonetheless a good approximation (Table 3.6). The sum of the squared residuals between calculated and literature values (for complete matches[23] and single mismatches[24–28]) was then minimized using Solver in Microsoft Excel to obtain the energetic contributions of the individual base pairs. The resulting parameters are shown in Table 3.5 and the residuals of this global fit are shown in Table 3.6. Free energies for mismatched sticky ends with a one nucleotide gap are estimated in the same manner except that there are no contributions from DNA base

stacking. Free energies for two sticky ends of any length as well as any possible

mismatch combinations can be estimated by this method.

Table 3.4. Calculated free energy contributions from individual base pairs.

Base pairs	Free energy (kcal/mol)	
AA	1.14	
AT	-0.51	(not used)
AC	1.50	
AG	0.75	
TA	-0.35	(not used)
TT	1.09	
TC	1.38	
TG	0.60	
CA	1.50	
CT	1.38	
CC	1.63	
CG	-0.87	(not used)
GA	0.75	
GT	0.49	
GC	-0.93	(not used)
GG	0.44	

Table 3.5. Nearest-neighbor free energy matrix used in the sticky end free energy calculator. [a]

5->3/3->5	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AA	2.28	0.69	2.64	1.89	0.61	-1.00	0.88	0.14	2.64	1.33	3.01	2.26	1.89	0.74	2.26	1.50
AT	0.61	2.23	2.52	1.63	-0.88	0.69	0.73	0.07	0.77	2.59	2.88	2.00	0.02	1.84	2.13	1.25
AC	2.64	2.52	2.77	0.17	0.77	0.64	1.33	-1.44	3.01	2.88	3.14	0.47	2.26	2.13	2.38	-0.52
AG	1.89	1.74	0.43	1.58	0.02	0.71	-1.28	-0.13	2.26	2.10	0.79	1.94	1.51	1.35	0.11	1.19
TA	0.69	-0.58	0.92	0.42	2.23	0.68	2.59	1.84	2.52	0.97	2.88	2.13	1.74	0.43	2.10	1.35
TT	-1.00	0.68	0.75	0.34	0.69	2.18	2.47	1.58	0.64	2.47	2.76	1.87	0.71	1.69	1.98	0.74
TC	1.33	0.97	1.05	-1.30	2.59	2.47	2.72	0.45	2.88	2.75	3.01	0.62	2.10	1.97	2.23	0.08
TG	0.74	0.43	-1.45	0.44	1.84	1.69	-0.12	1.53	2.13	1.97	0.62	1.81	1.35	0.52	-0.47	1.03
CA	2.64	0.92	3.01	2.26	2.52	0.75	2.88	2.13	2.77	1.05	3.14	2.38	0.43	-1.45	0.75	0.03
CT	0.88	2.59	2.88	2.00	0.73	2.47	2.76	1.87	1.33	2.72	3.01	2.13	-1.28	-0.12	0.40	-0.32
CC	3.01	2.88	3.14	0.81	2.88	2.76	3.01	0.98	3.14	3.01	3.26	0.79	0.79	0.62	0.70	-1.84
CG	2.26	2.10	0.75	1.94	2.13	1.98	0.40	1.82	2.38	2.23	0.70	2.07	0.11	-0.47	-2.17	-0.11
GA	1.89	0.42	2.26	1.50	1.63	0.34	2.00	1.24	0.17	-1.30	0.81	-0.25	1.58	0.44	1.94	1.19
GT	0.14	1.84	2.13	1.24	0.07	1.58	1.87	1.15	-1.44	0.45	0.98	-0.59	-0.13	1.53	1.82	0.93
GC	2.26	2.13	2.38	-0.25	2.00	1.87	2.13	-0.59	0.47	0.62	0.79	-2.24	1.94	1.81	2.07	-1.11
GG	1.50	1.35	0.03	1.19	1.25	0.74	-0.32	0.93	-0.52	0.08	-1.84	-1.11	1.19	1.03	-0.11	0.87

[a] Values in yellow represent interaction between complementary bases[23], values in dark blue represent G•T mismatches[24], values in green represent G•A mismatches[25], values in red represent A•C mismatches[26], values in purple represent C•T mismatches[27], values in light blue represent A•A, G•G, C•C or T•T mismatches[28], and values in white represent complete mismatch values calculated using data in Table 3.4.

Table 3.6. Residuals between calculated and literature values after fitting. [a]

5->3/3->5	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AA		0.01			0.00	0.02	0.01	0.01		0.03				0.11		
AT	0.00					0.02	0.01	0.02	0.01	0.05			0.05			
AC				0.00	0.05	0.05	0.04	0.00				0.01				0.12
AG			0.03		0.05	0.39	0.01	0.00			0.02				0.05	
TA	0.01	0.01	0.06	0.00		0.00			0.00					0.03		
TT	0.02	0.00	0.08	0.04	0.01				0.05				0.39			0.12
TC	0.03	0.00	0.05	0.00				0.09				0.03				0.17
TG	0.11	0.03	0.05	0.12			0.12					0.01		0.45	0.04	
CA		0.06				0.08				0.05			0.03	0.05	0.01	0.02
CT	0.01			0.02					0.04				0.01	0.12	0.01	0.00
CC			0.06					0.28				0.01	0.02	0.01	0.00	0.00
CG			0.01					0.01			0.00		0.05	0.04	0.18	0.10
GA		0.00			0.04				0.00	0.00	0.06	0.00		0.12		
GT	0.01			0.01				0.03	0.00	0.09	0.28	0.02	0.00			
GC			0.00					0.02	0.01	0.03	0.01	0.14				0.38
GG		0.02			0.12	0.00		0.12	0.17	0.00	0.38			0.10		

[a] Color representations are identical to those in Table 3.5.

Alternatively, a sticky end free energy can be estimated from up to ten trials in a linear ligation experiment using two DNA fragments containing the sticky end of interest as discussed in Section 3.2.7. At equilibrium, the two-piece linear ligation reaction and the mass balances yield three equations (24), (25) and (26) which can be analytically solved to give:

$$[AB] = 0.5 \times (K_D + [A]_t + [B]_t - \{(K_D + [A]_t + [B]_t)^2 - 4[A]_t[B]_t)\}^{1/2})$$

where [AB] is the concentration of the ligation product at equilibrium and $[A]_t$ and $[B]_t$ are the input concentrations of DNA fragments A and B, respectively. [A] and [B], which represent the concentrations of reactants still present at equilibrium, can be easily calculated by subtracting [AB] from $[A]_t$ and $[B]_t$, respectively. For a range of free energies (and, thus, K_D values), the computed [A], [B], and [AB] values are compared to those in all of the experimental trials to obtain a sum of squared residuals between model and experimental results. The free energy that yields the best global fit of species concentrations (i.e., the smallest sum of squared residuals) is the experimentally determined free energy of the sticky end. The number of base pairs is used to convert output mass ratio into molar ratio and the number of overhang bases is used to determine the free energy search space. This range of free energies has been estimated from DNA thermodynamics for all possible combinations of 2, 3, 4, and 5 base overhangs with random flanking bases, with both sources of stacking energy[21,22]. Alternatively, the free energy range can be explicitly specified.

3.3.3 Linear and circular ligation calculator

The mathematical models and experimental validations for both linear and circular ligations were previously described in Section 3.2.5 and 3.2.6 respectively. For two-piece linear ligation, three-piece linear ligation, and circular ligation, systems of 3, 7, and 16 equations are solved in Python using *scipy.optimize.fsolve* to obtain the 3, 7, and 16 unknowns, respectively. Ligation recommendations consist of computed heatmaps for incorporation, efficiency, and accuracy (circular ligation only) for molar ratios ranging from 1/8 to 8 and library concentrations ranging from 10% to 100% of the concentration of the valuable species. Incorporation is defined as the concentration of full-length product divided by the input concentration of the valuable species; efficiency is defined as the concentration of full-length product divided by the total input DNA concentration; and accuracy is defined as the percentage of circular vector-insert product over the sum of all circular products.

3.3.4 Ligation worksheet

To facilitate conversion between mass concentrations and molar concentrations and to simplify formulation of the final ligation reaction, a ligation workbook has been developed in Microsoft Excel and the file is available for download from the website. The first worksheet in the file contains a simple calculator to convert the mass

concentration of a sample in stock or in a reaction to a molar concentration[c]. The second worksheet contains a calculator for formulating two- or three-piece linear ligation reactions, allowing custom input of available mass concentrations, numbers of base pairs, targeted molar concentrations for each DNA species, ligase volume, and total reaction volume. The last worksheet contains a calculator for circular vector-insert ligations, which functions similarly to the linear ligation worksheet except that vector:insert molar ratio is specified.

3.4 Results

3.4.1 Optimal design of Type IIS target sites

We have designed a new ribosome display vector, pRDV2, which utilizes Type IIS restriction sites as an alternative to the predominant Type IIP restriction sites (Fig. 3.1A). The ribosome display vector pRDV was chosen as the vector template, and off7, a gene evolved from ribosome display to bind maltose-binding protein, was chosen to be the model insert[12]. The transcription/translation cassette of pRDV-Off7 consists of the T7 promoter, ribosome-binding site, off7 gene and tolA spacer (for more details on ribosome display, see Chapter 5). The off7 gene is usually flanked by the restriction sites of Type IIP restriction enzymes, such as NcoI and HindIII. In a directed evolution experiment, these two sites would be used for assembling the initial library into pRDV, reattaching the T7 promoter and tolA regions to the library insert between selection rounds, as well

---

[c] http://www.invitrogen.com/site/us/en/home/References/Ambion-Tech-Support/rna-tools-and-calculators/dna-and-rna-molecular-weights-and-conversions.html

as analyzing selection results by transformation and subsequent sequencing of individual clones[29]. Each of the above steps requires high ligation efficiency and fidelity to preserve the diversity of the library as the success of any selections depends heavily on the quality of the construction, propagation and analysis of the library. We thus picked a directed evolution application as the motivation of our study, although our results are generally applicable in any context where ligation efficiency is important.

NcoI and HindIII cut double-stranded DNA with four-base palindromic overhangs CATG and AGCT, respectively (Fig. 3.1B and C). Identical palindromic ends in different DNA molecules, usually caused by digestion with the same restriction enzyme, can anneal by forming complementary base pairs. However, a molecule can also anneal to another copy of itself in reverse orientation as palindromic ends are self-complementary. This leads to the loss of molecules and unwanted homodimer byproducts which is an inherent limitation of any Type IIP restriction enzyme. Fortunately, the two ends CATG and AGCT cannot anneal to each other for more than 1 base, which generally falls outside of the mismatch tolerance of T4 DNA ligase[30,31]. Thus, NcoI and HindIII represent two good palindromic restriction sites for drawing comparisons to Type IIS sites.

Since the Type IIS restriction enzyme BbsI cuts at sites 2 and 6 bp away from the recognition sequence, it allows us to custom design the sequence of each cutting site. CGAA and GCTT (Fig. 3.1B and C) were chosen for several reasons. First, they are non-palindromic and eliminate homodimer formation. Second, they cannot recombine with each other within mismatch tolerance of T4 DNA ligase. In contrast, the use of sub-

optimal cutting sequences ACTA and GCTT, which have two out of four correctly positioned complementary bases between sense and anti-sense overhangs, gave a very high misannealing of the two sites. The selection of CGAA and GCTT overhangs provides a simple solution to the above two problems, as the two sites are identical in overall base composition (thus providing similar annealing temperatures) but have opposite sequence orientations (thus minimizing spurious annealing). Third, the first BbsI site was inserted between the T7 promoter and RBS so as not to perturb transcription and translation, while the second site was inserted at the beginning of tolA preserving in-frame glycines and serines to maintain the flexibility of the spacer (Fig. 3.1C, bottom). Single-clone ribosome display was carried out using pRDV2:Off7 to demonstrate that insertion of the BbsI sites had no adverse effect on selection (Fig. 3.2). Lastly, the BbsI sites were positioned next to the original NcoI and HindIII sites to generate similar fragment sizes, which facilitated direct comparison of the methods.
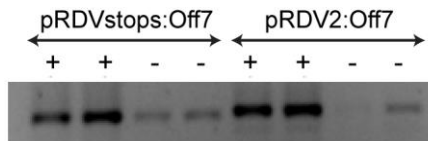
Figure 3.2. Results from one round of ribosome display with pRDVstops:Off7 and pRDV2:Off7 as templates. A "+" denotes a well with MBP immobilized and a "−" denotes a well without MBP. Each condition was tested in duplicate.

3.4.2 Vector-insert ligation of Type IIS digested fragments yields more full-length product

We first performed ligation of off7 and pRDV2 with NcoI/HindIII cuts or with two BbsI cuts. A constant amount of vector with different molar ratios of insert:vector were ligated and the products were resolved and visualized with agarose gel electrophoresis. Figure 3.3A shows ligation of off7 and pRDV2 digested by NcoI and HindIII. Even without insert, many high molecular weight products were formed through vector–vector annealing events. As the amount of insert increased, high molecular weight products remained and additional lower molecular weight bands were visible. The presence of circular species and the exponential combination of linear and circular products make determination of the identity of these bands extremely difficult (the sizes of products containing up to four ligated molecules are provided in Table 3.7). pRDV2:Off7 nicked at a single BsmI site was included to mimic the circular ligation product of interest since both are of identical length and should have similar migration properties in the agarose gel, but no ligation products are visible at this distance in the gel. In contrast, ligation of off7 and pRDV2 cut by BbsI shows a faint product band at the same distance as the nicked pRDV2-Off7, and this band is strongest for a 1:1 insert:vector ratio (Fig. 3.3B). In the BbsI ligation, excess insert at 0.5 kb (5:1 and 3:1) or excess doubly digested vector at 3–4 kb (1:3 and no insert) was clearly present, indicating that neither insert–insert nor vector–vector annealing events occurred with our designed non-palindromic sticky ends. The sizes of the product bands are provided in Table 3.7.

89
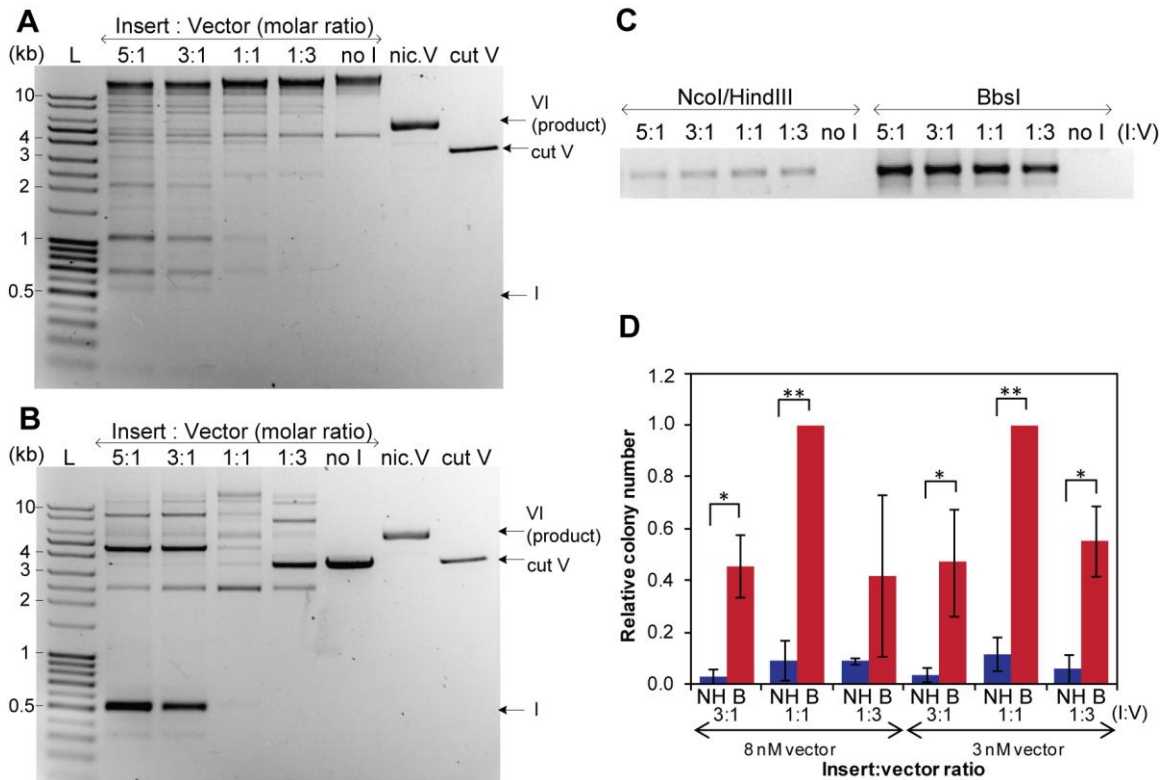
Figure 3.3. Comparison of NcoI/HindIII- and BbsI-digested products in vector-insert ligation (V: digested pRDV2 vector; I: digested off7 insert). Ligation of (A) NcoI/HindIII-digested or (B) BbsI-digested products resolved in a 1% agarose gel. Amount of vector ligated was fixed at 200 ng (~8 nM) for each lane with the corresponding amount of insert based on the molar ratio shown. Nic.V represents Nb.BsmI-nicked pRDV2:Off7 which should migrate at approximately the same rate as the ligation product VIc (c denotes circular product). Cut V represents NcoI/HindIII- or BbsI-digested pRDV2 backbone (in A and B, respectively) used for ligation reactions. (C) Vector (75 ng ≈ 8 nM) ligated with the corresponding amount of insert was PCR amplified with primers 1/2 for 10 cycles. Products at 902 bp (full-length product) are shown. (D) Vector (8 or 3 nM) ligated with the corresponding amount of insert was transformed into XL1-Blue cells. Shown here is the number of colonies relative to the number obtained for 1:1 BbsI ligation of the same vector concentration. With 8 nM vector, vector-only transformations gave 2–10 colonies with NcoI/HindIII digestion and 1–11 colonies with BbsI digestion; with 3 nM vector, vector-only transformation gave 0–2 colonies with NcoI/HindIII digestion and 0–2 colonies with BbsI digestion. NH: sites from NcoI/HindIII digestion (blue), B: sites from BbsI digestion (red). The data represent the mean ± SD of two independent experiments, $*P < 0.05$ and $**P < 0.01$. A one-tailed t-test was used to examine statistical significance.

90

Table 3.7. Possible identities and sizes of species formed from insert-vector and three-piece linear ligation of fragments generated by NcoI/HindIII or BbsI restriction digests.

Insert-vector circular ligation			Three-piece linear ligation		
Number of based pairs				**Number of base pairs**	
Species	**NcoI/HindIII**	**BbsI**	**Species**	**NcoI/HindIII**	**BbsI**
V	3428	3408	A	106	87
I	500	520	B	500	520
VI	3928	3928	C	296	295
VI (circular)	3928 [a]	3928 [a]	AB	606	607
VV	6856	– [b]	BC	796	815
VV (circular)	6856 [a]	– [b]	AA	212	– [b]
II	1000	– [b]	CC	592	– [b]
II (circular)	1000 [a]	– [b]	BB	1000	– [b]
IVI [c]	4428	4448	BB (circular)	1000 [a]	– [b]
VIV [c]	7356	7336	ABC	902	902
VVV	10,284	– [b]	ABB	1106	– [b]
III	1500	– [b]	BBC	1296	– [b]
VIVI [c]	7856	7856	BBB	1500	– [b]
VIVI [c] (circular)	7856 [a]	7856 [a]	ABBA	1212	– [b]
VIII [d]	4928	– [b]	CBBC	1595	– [b]
VIII [d] (circular)	4928 [a]	– [b]	ABBB	1606	– [b]
VVVI [d]	10,784	– [b]	BBBC	1796	– [b]
VVVI [d] (circular)	10,784 [a]	– [b]	BBBB	2000	– [b]
VVVV	13,712	– [b]	BBBB (circular)	2000 [a]	– [b]
VVVV (circular)	13,712 [a]	– [b]	5th order or higher [e]	varies	– [b]
IIII	2000	– [b]			
IIII (circular)	2000 [a]	– [b]			
5th order or higher [e]	varies	varies			

[a] The migration distance of circular DNA molecules varies depending on the conformation and cannot easily be determined on an agarose gel.

[b] These species should not form due to incompatible ends created by BbsI digestion.

[c] The orientation shown is only for BbsI digestion; other combinations are possible for NcoI/HindIII digestion (i.e. VVI, IVI, IIV, etc.)

For comparison of the relative amount of correctly ligated product, each ligation mixture was subjected to 10 cycles of PCR amplification with outer primers that anneal to insert-flanking regions of pRDV2 (Fig. 3.3C). This procedure should selectively amplify circular vector-insert products and not linear vector-insert products or other undesired products (vector-vector, insert-insert, etc.). Across all molar ratios, ligation of BbsI-digested fragments gave more amplified product than NcoI/HindIII-digested products. Approximate quantification of the gel indicated a 5- to 10-fold increase in BbsI products over NcoI/HindIII products. In addition, the ligation mixture was transformed and the number of bacterial colonies were counted (Fig. 3.3D). To minimize any variability in transformation efficiencies across experiments, colony numbers were normalized to the result for the 1:1 insert:vector ligation reaction at the same vector concentration. At both vector concentrations, BbsI-digested products gave 4- to 15-fold more colonies than NcoI/HindIII-digested products at all molar ratios examined. All these results suggest that it is preferable to use Type IIS restriction enzymes in generating fragments for ligation to yield more full-length product.

3.4.3 Three-piece linear ligation of Type IIS digested fragments yields more full-length product

In order to gain more understanding of the ligation reaction, a minimal system consisting of the linear ligation of three DNA fragments was developed. This scheme is also particularly useful for applications in which transformation is not needed and linear ligation products can be directly used (e.g., *in vitro* directed evolution). A constant amount of digested off7 was ligated to digested T7 and tolA at the specified molar ratios, and the products were resolved and visualized on an agarose gel (Fig. 3.4A). At all molar ratios, more full-length ligated product (at 902 bp) was observed for BbsI-digested fragments than NcoI/HindIII-digested fragments. Quantification of the gel confirmed this observation, indicating a 1.5- to 2.6-fold increase in full-length species. It should be noted that the smallest increase (1.5-fold) was achieved using the suboptimal 1:2:1 A:B:C ratio, and more significant improvements can be realized when using ratios in which A and C are in excess (discussed below). A closer look at ligation following NcoI and HindIII digestion (Fig. 3.4B) shows that the loss of full-length product was due to stoichiometry-dependent byproduct formation. With excess T7 and tolA (10:1:10 and 2:1:2), T7-T7 and tolA-tolA homodimers are the major products, but the desired product is also maximized since off7 is less likely to homodimerize when T7 and tolA are in such abundance. In contrast, with excess off7 (1:2:1 and 1:10:1), many high molecular weight products could be seen, likely due to the undesired formation of off7 multimers, which can further circularize or anneal to T7 or tolA fragments to generate further species diversity. Indeed, a ligation reaction containing only NcoI/HindIII-digested off7 showed many high molecular weight bands. A closer look at ligation following BbsI digestion

(Fig. 3.4C) showed minimal byproducts: the only major species other than reactants or full- length product were T7-off7 and off7-tolA, both of which were intermediate ligation products. More full-length products were observed with higher molar excesses of T7 and tolA, as there was no risk of byproduct formation. Again, this minimal ligation system clearly shows that BbsI-digested fragments are preferable to NcoI/HindIII-digested fragments for maximizing generation of the desired linear product in a ligation reaction.
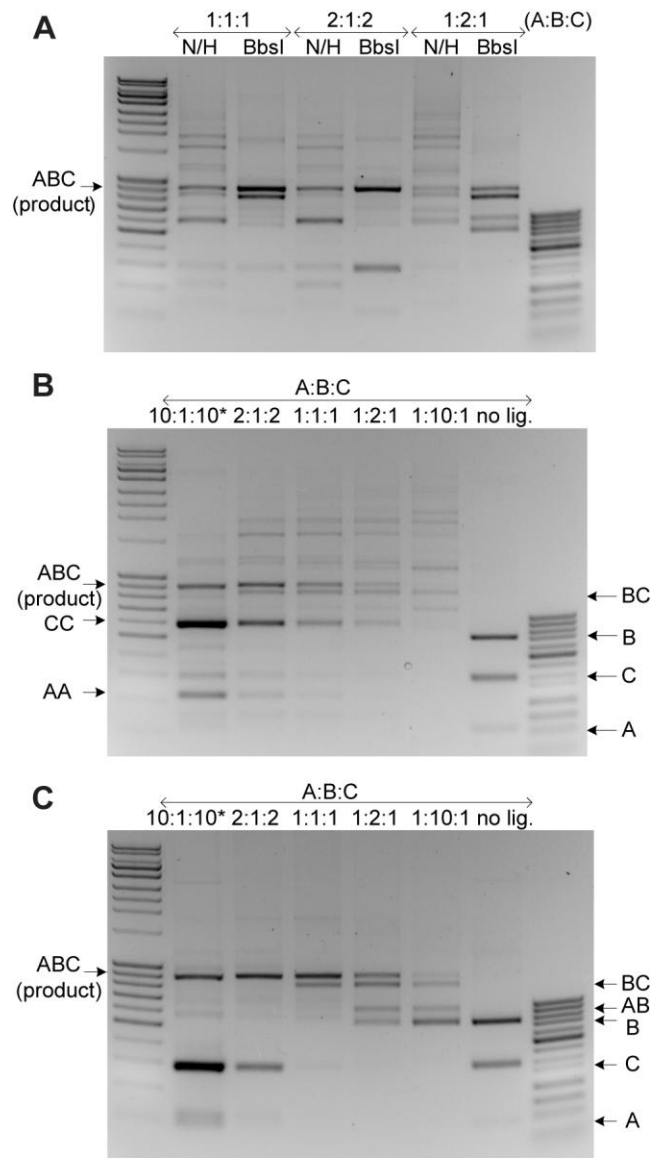
Figure 3.4. Representative 1.5% agarose gels showing NcoI/HindIII- and BbsI- digested products in three-piece linear ligation (A: T7; B: off7; C: tolA). (A) Comparison of linear ligation band patterns between NcoI/HindIII-(N/H) and BbsI-digested reactants at three molar ratios of A: B: C. In all ligation reactions, 100 ng (~54 nM) of B was used with corresponding amounts of A and C. The arrow shows full-length ABC product at 902 bp. Linear ligation products of (B) NcoI/HindIII-digested and (C) BbsI-digested product at different molar ratios. In all ligation reactions, 100 ng (~24 nM) of B was used (except for *: 50 ng of B was used in 10:1:10). Arrows in (B) indicate positions of selected species; locations of higher order species can be found in Table 3.7. Arrows in (C) indicate positions of all of the species. No lig.: 1:1:1 A:B:C reaction without ligase.

3.4.4 Development and training of a computational model for Type IIS ligation

Viewing ligation as a two-step process, a thermodynamic model was developed for both the minimal three-piece ligation and the more complicated vector-insert ligation for fragments generated by Type IIS restriction enzymes (Fig. 3.5). The first step is the annealing of two sticky ends by hydrogen bonding across complementary bases and/or stacking between adjacent bases[22], and the second step is the sealing of the DNA backbone by the formation of phosphodiester bonds[32]. We modeled each ligation reaction as a reversible equilibrium process with the free energy defined by the identity of the sticky ends involved in that particular reaction. This thermodynamic model was used to estimate output from input when ligation had proceeded for a sufficiently long time to reach equilibrium. Assuming that both sealing events after an annealing event occur in rapid succession and that only ligated products, but not annealed and nicked products, are stable enough to proceed to the next ligation reaction, a system of thermodynamic equations was written.

95
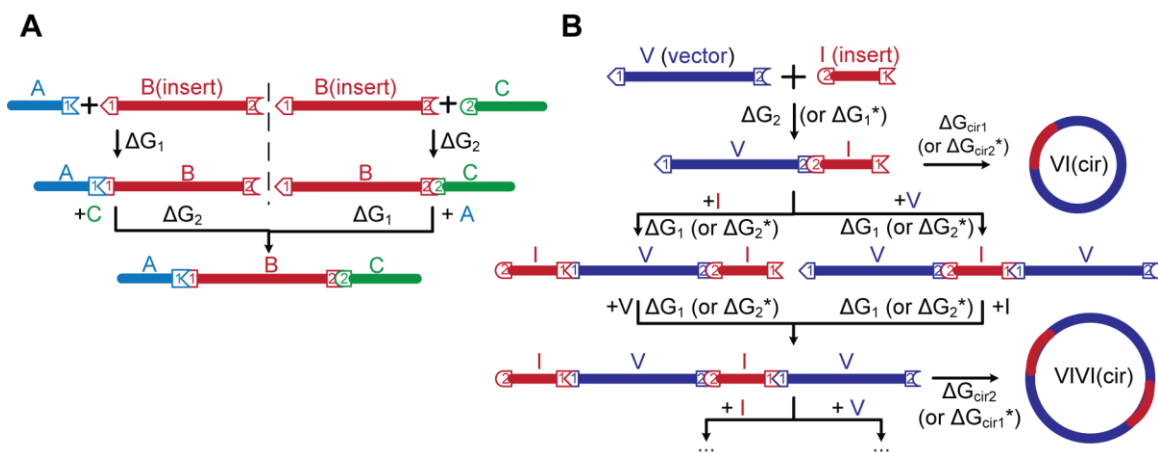
Figure 3.5. Schematic of the thermodynamic model proposed in this study. Refer to Table 3.2 for nomenclature used. (A) Model of three-piece linear ligation based on two lumped free energy parameters: annealing and sealing at site 1 (CGAA) and site 2 (GCTT) were represented by ΔG_1 and ΔG_2, respectively. All possible species are shown in the diagram. A: T7, B: off7, C: tolA. (B) Model of vector-insert circular ligation based on ΔG_1, ΔG_2, ΔG_{cir1} and ΔG_{cir2}, where the latter two were calculated from the three ΔG_1, ΔG_2 and j. For clarity, only one orientation is shown for VI, IVI, VIV and IVIV, along with their corresponding free energies. Only products containing up to four reactant molecules were included in the model. *: free energy parameters used for the alternative orientations (i.e. when V and I anneal in a different configuration than that shown). Cir: circular product.

Although free energies of annealing for specific sequences can be estimated using available calculators, experimentally determined free energies are more accurate since the specific ligation conditions (e.g., temperature and salt concentration) are intrinsically captured in the values. In our model, these free energies of annealing can easily be lumped with the free energies of sealing to yield one ΔG parameter per sticky end. Thus, we carried out two-piece linear ligation at different molar ratios to experimentally determine ΔG_1 and ΔG_2 for the sticky ends CGAA and GCTT, respectively. ΔG_1 was

determined to be -12.00 kcal/mol and ΔG_2 was determined to be -12.23 kcal/mol. These two values were used for all subsequent simulations.

Figure 3.5A shows the three-piece linear ligation model for non-palindromic sticky ends. All possible species were considered. We assumed no homodimer formation, no cross-reactivity between sites 1 and 2, and no blunt-end ligation. Figure 3.5B shows the vector-insert ligation model for non-palindromic sticky ends. Species containing up to four fragments were explicitly included and any higher order molecules were assumed to be insignificant. This assumption is valid in the range of concentrations and molar ratios that were most commonly used, and accounting for higher-order species would have a minimal effect on the optimizations that were examined in this work. The probability of circularization is reflected by the Jacobson–Stockmayer factor j which represents the effective concentration of one end of a linear molecule with respect to the other, which is dependent on length of the linear molecule but independent of DNA concentration[15]. If the competing DNA concentration is above j, linear multimers will be favored; conversely, circularization will be favored if the DNA concentration is below j[4]. ΔG_{cir1} and ΔG_{cir2} were calculated by dividing the corresponding K values by this effective concentration factor.

3.4.5 Thermodynamic model accurately captures ligation product formation

To test the validity of the model, we compared its predictions to experimental results. First, three-piece linear ligations with five different molar ratios were simulated by

solving the system of equations to obtain equilibrium molar concentrations for the reactants, intermediates and products in each ligation (Fig. 3.6A–E). For each reactant molar ratio, product molar concentrations were converted into product molar ratios by normalizing to full-length product and further converted to mass ratio based on the number of base pairs. This was done to facilitate direct comparison with experiments in which mass ratios were determined by quantifying product bands in agarose gels. Across all reactant molar ratios, the model predictions were in good agreement with the experimental results. The only major difference between predictions and experiments was for 1:10:1, where predictions generated mass ratios much higher than those obtained by experiments (Fig. 3.6E). This was likely due to inaccuracy from saturating gel signals and further error amplification from the small amount of full-length product compared with intermediates and reactants. Correlation tests conducted within each molar ratio gave $P < 0.05$ for 10:1:10, 1:1:1 and 1:2:1, and $P < 0.01$ for 2:1:2 and 1:10:1, showing that predictions and experimental results were indeed significantly correlated.
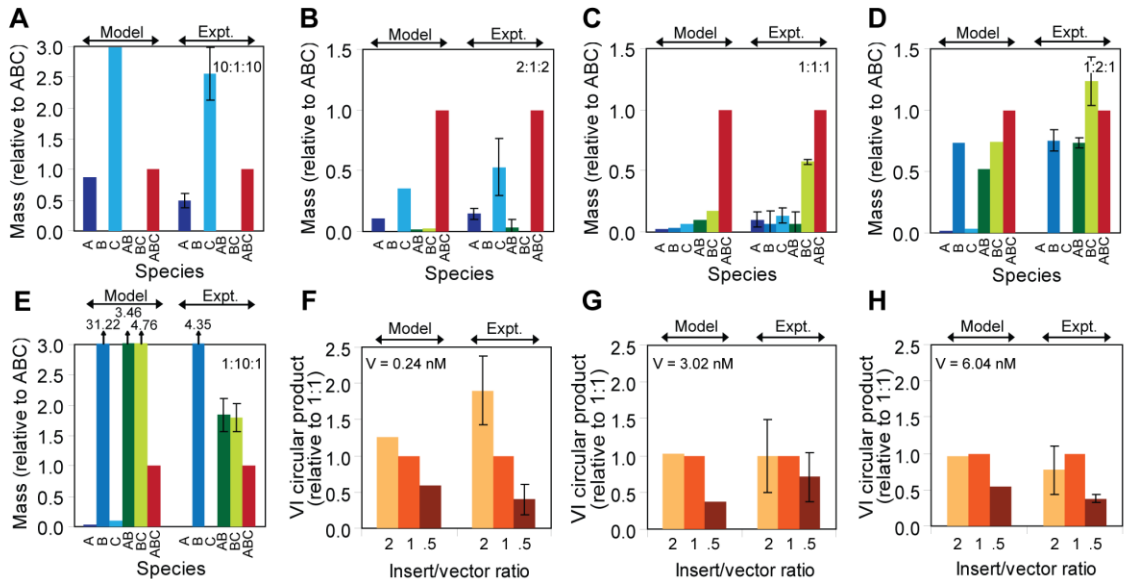
Figure 3.6. Validation of thermodynamic model with experimental results. Model predictions and experimental results for three-piece linear ligation of BbsI-digested products with A: B: C (molar ratios) of (A) 10:1:10, (B) 2:1:2, (C) 1:1:1, (D) 1:2:1 and (E) 1:10:1. Values shown are mass ratios normalized to the mass of the desired product, ABC. ΔG_1 = -12.00 kcal/mol, ΔG_2 = -12.23 kcal/mol, and B = 24 nM were used in all ligation simulations (except in (A), in which B = 12 nM was used). In all experiments, 100 ng (~24 nM) B was used (except in (A), in which 50 ng (~12 nM) B was used). Data represent the mean ± SD of three independent experiments (one of which is shown in Fig. 3.4C). Corresponding comparisons in molar ratios are given in Fig. 3.7. Model predictions and experimental results for vector-insert circular ligation of BbsI-digested products with vector concentrations of (F) 0.24 nM, (G) 3.02 nM and (H) 6.04 nM. Values shown are molar ratios of VI circular product (model) or colony numbers (experiment) normalized to the corresponding value for insert/vector = 1. ΔG_1 = -12.00 kcal/mol and ΔG_2 = -12.23 kcal/mol. In all experiments, 25 ng of vector was ligated to insert using the specified molar ratio. Within each vector concentration, relative colony numbers were obtained by subtracting the background (colony number for corresponding vector-only ligation) and dividing this specific colony number by the specific colony number for I/V = 1. Vector-only transformations gave 0 colonies with 0.24 nM vector, 0–2 colonies with 3.02 nM vector, and 0–7 colonies with 6.04 nM vector.

For model validation above, we chose to plot the species using relative mass because quantification of the product bands in the agarose gels directly yields this quantity. Any error in experimental procedure or DNA quantification would artificially magnify errors in molar ratios. This is particularly problematic when considering lower molecular weight species, since small absolute errors in mass quantification lead to large absolute differences in corresponding moles, which in turn could dramatically, yet artificially, alter a molar ratio (e.g., compare relative amounts of species A in Fig. 3.6C and Fig. 3.7C). Nevertheless, with this caveat in mind, the use of molar quantities can provide additional insight into ligation efficiency under different conditions, so we also present the data in molar units in Figure 3.7.
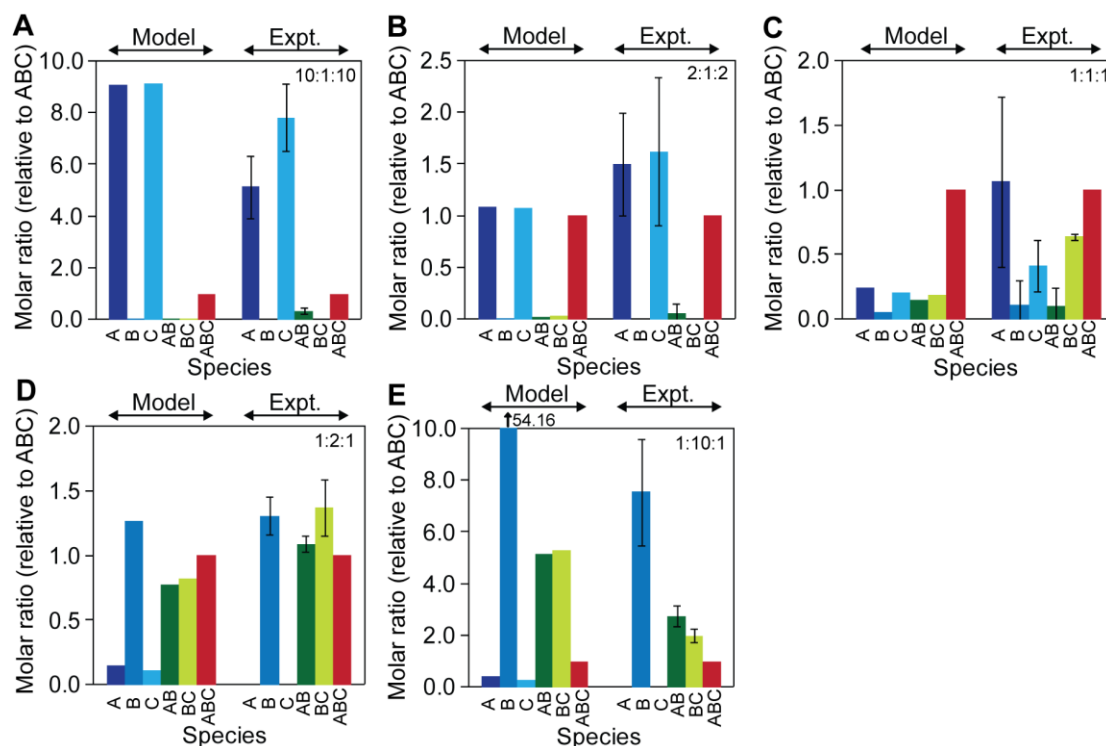
Figure 3.7. Comparison of thermodynamic model and experimental results (same data as in Fig. 3.6A–E but presented here in molar quantities). Model predictions and experimental results for three-piece linear ligation of BbsI-digested products with A:B:C (molar ratios) of (A) 10:1:10, (B) 2:1:2, (C) 1:1:1, (D) 1:2:1, and (E) 1:10:1. Values shown are molar ratios normalized to the number of moles of the desired product, ABC. ΔG_1 = -12.00 kcal/mol, ΔG_2 = -12.23 kcal/mol, and B = 24 nM were used in all ligation simulations and experiments (except in (A), in which B = 12 nM was used). Data represent the mean ± SD of three independent experiments.

We then proceeded to test the predictions of the vector-insert ligation model. Ligations at three different molar insert:vector ratios were simulated at low, medium and high vector concentrations. Within each vector concentration, all correctly ligated circular products were normalized to the 1:1 insert:vector yield (Fig. 3.6F–H). At low concentration, a 2:1 ratio was predicted to yield more functional product because this

ligation is limited by the availability of reactants. At high concentration, a 2:1 ratio was predicted to have a slightly adverse effect because free insert would compete with circularization reaction by forming more linear insert-vector-insert products. These predictions were experimentally tested by bacterial transformation, assuming the number of colonies corresponds primarily to circular vector-insert products. The experimental trends again match the computational predictions. Correlation tests conducted within each vector concentration gave $P < 0.05$ for 3.02 and 6.04 nM vector and $P < 0.01$ for 0.24 nM vector, again statistically confirming that our model predictions are significant.

3.4.6 Optimal conditions to preserve insert in linear ligation

To further explore a specific application of the linear ligation model, we investigated model predictions for applications where the insert is the valuable entity. For example, in *in vitro* directed evolution methods such as ribosome display[33] and mRNA display[34], the amount of insert ligated would directly influence library quality and diversity. For such applications, linear ligation is preferable to vector-insert ligation because no molecules would be lost in byproduct formation. The product could also be directly used, without the need for PCR amplification or transformation, allowing better quantification of reactions. First, we simulated ligations at different insert concentrations and across a range of reactant molar ratios based on our two previously characterized cutting sites, CGAA and GCTT. Percent incorporation of insert into full-length product (Fig. 3.8A) and ligation efficiency defined as product divided by reactants (Fig. 3.8B) were determined both by simulations and by experiments. Across all molar ratios, increasing

101

reactant concentrations led to an increase in overall incorporation of insert. At each concentration, increasing the molar ratios of the upstream and downstream pieces also monotonically increased incorporation of insert, in agreement with experimental results. Since there are no byproducts, any increase in reactants (insert or either end) will shift the equilibrium in each ligation reactions towards product formation. Ligation efficiency, however, shows a clear optimum at different ends/insert ratios in both simulations and experiments. At sufficiently high concentrations, 1:1:1 gives optimal ligation efficiency. As concentration decreases, an excess of ends is needed to achieve maximum ligation efficiency. Since the amount of insert is fixed, increasing the concentration of the ends is needed to compensate for the unfavorable equilibrium caused by the lower overall reactant concentration, though the compensation is never enough to reach the same ligation efficiency.
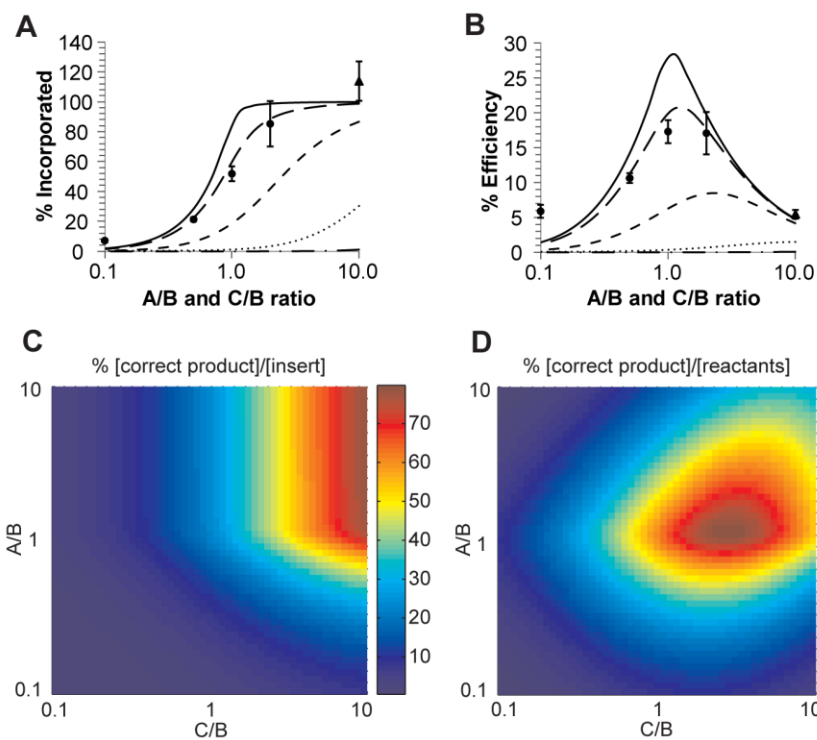
Figure 3.8. Simulations of three-piece linear ligation with experimental comparison. Percent incorporation is defined as product/insert ($[ABC]/[B]_t \times 100\%$) and percent efficiency is defined as product/reactants ($[ABC]/([A]_t + [B]_t + [C]_t) \times 100\%$). (A) Percent incorporation and (B) percent efficiency of linear ligation with 100 nM (solid line), 10 nM (long dashed line), 1 nM (short dashed line), 0.1 nM (dotted line), 0.01 nM (dashed and dotted line) of B (off7/insert) and specified amounts of A (T7) and C (tolA), all digested by BbsI. $\Delta G_1 = -12.00$ kcal/mol and $\Delta G_2 = -12.23$ kcal/mol. The overlaid experimental results, quantified from Fig. 3.4C, show ligation reactions with 24 nM of B in A:B:C reactant stoichiometry of 1:10:1, 1:2:1, 1:1:1 and 1:2:1 (solid circles) and 12 nM of B in an A:B:C reactant stoichiometry of 10:1:10 (solid triangle). Data represent the mean ± SD of three independent experiments. (C) Percent incorporation and (D) percent efficiency of linear ligation with 10 nM of B (off7/insert) and specified amounts of A (T7) and C (tolA), all digested by BbsI. $\Delta G_1 = -14.00$ kcal/mol and $\Delta G_2 = -10.00$ kcal/mol.

We also investigated the effect of different ΔG_1 and ΔG_2 values on optimal incorporation and efficiency while allowing concentrations of the two ends to vary (Fig. 3.8C, D). Since the linear ligation model is symmetric (the lengths of the end fragments have no effect on molar concentration or ligation behavior), we perturbed ΔG_1 and ΔG_2 only once in separate directions. ΔG_1 was set to be more negative (-14.00 kcal/mol; equilibrium more favorable towards product) and ΔG_2 was set to be less negative (-10.00 kcal/mol; equilibrium less favorable towards product). An intermediate insert concentration was chosen (10 nM: long dash in Fig. 3.8A and B). For a less negative ΔG value (ΔG_2), a higher molar excess of that end (C) is needed to achieve high insert incorporation (Fig. 3.8C) as well as efficiency (Fig. 3.8B). However, even using an excess of C cannot compensate for the less favorable equilibrium to achieve the same incorporation and efficiency values. If the insert concentration is increased (e.g., 100 nM: solid line in Fig. 3.8A and B), the shift in molar excess of C to achieve optimal incorporation and efficiency is smaller and the optimal values are also.

103

Based on these simulation results, it is recommended that, for high incorporation of insert and maximal ligation efficiency, the two sticky end sites should be designed to have ΔG values that are as similar and as negative as possible (i.e. match the number of G/C and A/T base-pairs on the two sticky ends with no more than two A/T base-pairs per sticky end). In three-piece linear ligation reactions in which the middle piece (insert) needs to be maximally incorporated, the upstream and downstream pieces should be used in as great excess as possible.

3.4.7 Optimal conditions to preserve insert in vector-insert ligation

Vector-insert ligation represents the more common type of ligation, so we also investigated model predictions for this setup, again where the insert is the valuable entity. These insights may be useful for *in vivo* directed evolution methods that require circular display vectors, including phage display[35], bacterial display[36] and yeast display[37]. Again, we simulated ligations across a range of vector concentrations and insert:vector molar ratios based on ΔG_1 and ΔG_2 values for our previously characterized cutting sites, CGAA and GCTT. We examined percent incorporation of insert into circular vector-insert molecules (Fig. 3.9A), ligation efficiency defined as product concentration divided by the sum of the reactant concentrations (Fig. 3.9B) and accuracy defined as the concentration of circular vector-insert molecules divided by the sum of all circular product concentrations (Fig. 3.9C). At a fixed vector concentration, increasing the insert:vector ratio decreases incorporation and efficiency as insert-vector-insert is preferentially formed over circular insert-vector molecules. On the other hand, even though low

insert:vector ratio gives minimal byproduct, the insert concentration is too low for sufficient product formation to achieve high efficiency. Incorporation in this regime remains high because the input, total insert, is also low. At a fixed insert:vector ratio, optimal incorporation and efficiency lie at an intermediate concentration (which is around the K value, where $K = \exp(\Delta G/RT)$) because a low concentration shifts the equilibrium to the reactant side and a high concentration promotes the formation of vector-insert-vector products. Circular product shows high accuracy at medium-to-low vector concentrations. At high vector concentrations, the accuracy remains good for very low or high insert:vector ratios as those ratio are not preferable in forming circular vector-insert-vector-insert products. Only at high vector and insert concentrations with approximately an equal molar ratio does the accuracy drop. To further highlight the utility of this model in experimental design, we overlaid our experimental results from Fig. 3.3D on these heat maps. The model not only captures the experimental trends, but it also indicates how to rationally change the experimental conditions to further improve a given ligation reaction.
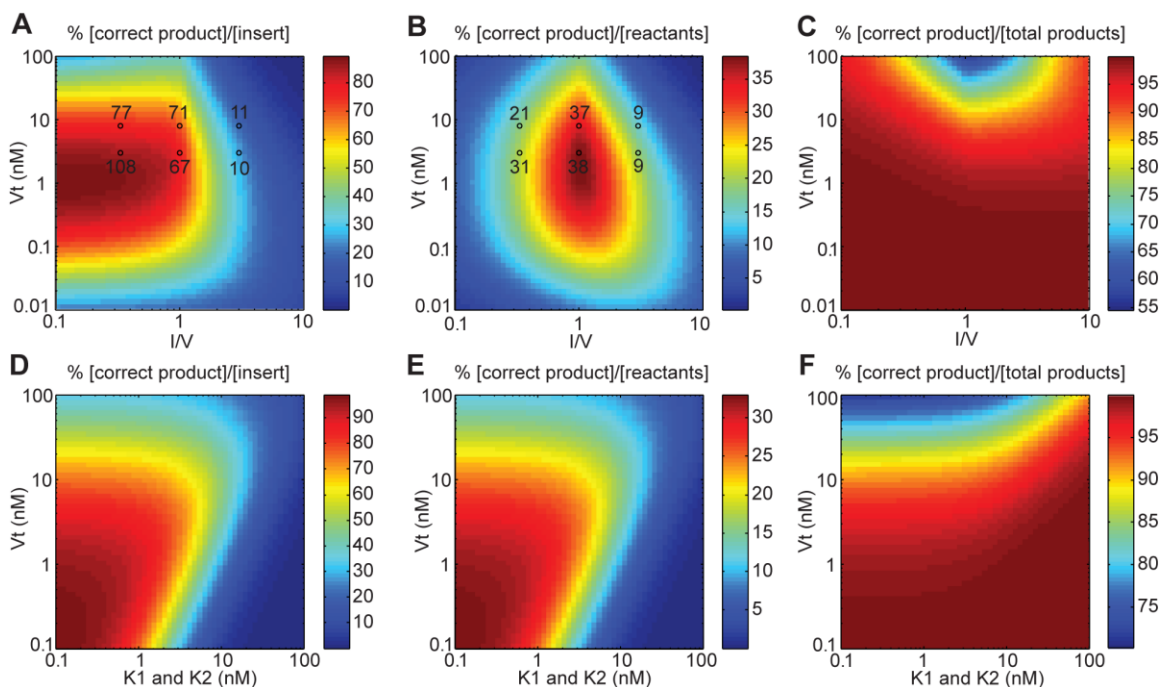
Figure 3.9. Simulations of vector-insert circular ligation with experimental comparison. Percent incorporation is defined as correct product/insert ($[VIc]/[I]_t \times 100\%$), percent efficiency is defined as correct product/reactants ($[VIc]/([V]_t + [I]_t) \times 100\%$), and percent accuracy is defined as correct product/total product ($[VIc]/([VIc] + [VIVIc]) \times 100\%$). (A) Percent incorporation, (B) percent efficiency and (C) percent accuracy of circular ligation of the specified amount of BbsI-digested vector and insert. $\Delta G_1 = -12.00$ kcal/mol and $\Delta G_2 = -12.23$ kcal/mol. Experimental values from Fig. 3.3D are shown on the heat maps next to the open circles that correspond to the tested reaction conditions. (D) Percent incorporation, (E) percent efficiency and (F) percent accuracy of circular ligation for the specified amount of BbsI-digested vector and I:V = 1:2. For reference, K values of 0.1, 1, 10 and 100 nM correspond to ΔG values of -13.63, -12.27, -10.91 and -9.54 kcal/mol, respectively.

Finally, we investigated the effect of different ΔG_1 and ΔG_2 values in optimal insert incorporation, efficiency and accuracy of ligation by fixing the insert:vector ratio at 1:2 and allowing the total vector concentration to vary. This ratio was chosen because it strikes a good balance between incorporation and efficiency, based on the analysis above.

Incorporation and efficiency show the same trends (Fig. 3.9D and E, respectively). At different K_1 and K_2 values, the optimal incorporation and efficiency appears to be at a vector concentration very close to the K values. At fixed K_1 and K_2 values of 1 nM (corresponding ΔG of -12.27 kcal/mol which is close to our experimental value), the trends shown in Fig. 3.9A and B could be seen. As expected, accuracy drops at high vector concentration (Fig. 3.9F). This effect is slightly counteracted at unfavorable equilibrium (high K value) as the competing reactions to form circular vector-insert-vector-insert are not as strong. Combining the two simulation results, it is recommended that for high insert incorporation, maximal ligation efficiency and best accuracy, the insert:vector ratio should be 1 or slightly less and the vector concentration should be kept close to the K values.

3.4.8 Ligation calculator: free energy estimation by DNA thermodynamics

To facilitate more widespread use of our thermodynamic model, we have developed a web-based ligation calculator. The only parameter required by the thermodynamic model is the lumped equilibrium dissociation constant for ligation, which can be calculated from the free energies of annealing and ligation for each sticky end. The first section of the web ligation calculator estimates free energy of annealing based on the unique overhang and flanking bases of each sticky end. An added advantage is that the free energy of sticky end misannealing (to itself or to another sticky end) can also be quantified, allowing quick and quantitative evaluation of the performance of each sticky end, which can help guide rational design of these sequences.

An example of this free energy calculation is shown in Figure 3.10A. There are three components to the free energy. First, hydrogen bonding interactions between complementary bases drive the annealing of complementary sticky ends. The free energy change due to these interactions can be estimated by oligonucleotide nearest-neighbor thermodynamics, which assumes that the stability of each base pair depends on the identity and orientation of neighboring base pairs[23]. Values for all 16 such interactions have been tabulated[23], and the hydrogen-bonding free energy between two complementary 4-base overhang sticky ends can be estimated by the addition of three nearest-neighbor free energy values. The second component comes from stacking between adjacent bases across the nick sites, and information about the flanking bases and the type of overhang (5' or 3') are used to compute this free energy. Since the input flanking bases are strictly one base upstream and downstream of the sticky ends in the upper strand, the type of overhang then dictates the locations of the nick sites and the appropriate bases that participate in stacking. The free energies of base stacking for all 16 combinations have also been determined[21,22], although there are some discrepancies between the two studies. Using tabulated values from Florián et al.[21] in the overall free energy calculation, we estimated the free energy values for sticky ends (with parenthetical flanking bases) (C)CGAA(A) and (A)GCTT(G) to be -11.77 and -12.02 kcal/mol, respectively; these predictions closely match our experimentally determined values of -12.00 and -12.23 kcal/mol (Section 3.4.4). We therefore set the base stacking interactions from Florián et al.[21] as the default, but the options to use values from Protozanova et al.[22] or to ignore base stacking contributions are also available. The third and final component of the sticky-end free energy comes from repair of the

phosphodiester bond at the nick site. Assuming that products sealed by ligase at one site cannot re-nick at the other site, this free energy can be lumped into the free energy contributions from hydrogen bonds and base stacking to give the overall free energy of sticky end. The default value is -6.3 kcal/mol[16], but any custom value can be specified by the user. Using all of the default parameters, the free energies of randomly generated 4-base overhang sticky ends (both 5' and 3' overhangs) with their complementary sticky ends were calculated (Fig. 3.10B). The values are approximately normally distributed with a mean of -11.63 kcal/mol, a standard deviation of 1.03 kcal/mol, and a range of -15.15 to -8.74 kcal/mol. The subset of palindromic 4-base overhangs is shown in red.
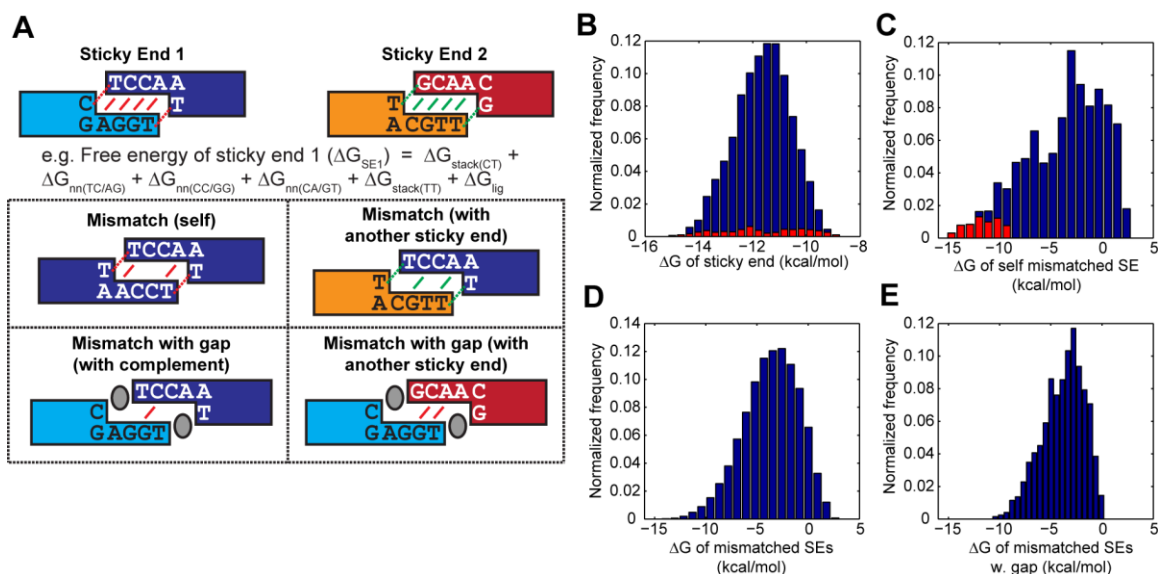
Figure 3.10. Schematic and distribution of values obtained from the sticky end free energy calculator. (A) Two sticky ends – (C)TCCA(A) and (T)GCAA(C) – with 5' overhangs are shown in both correct and mismatch orientations, with complementary bases and stacking interactions indicated by color lines. The lumped free energy of sticky end (C)TCCA(A) (ΔG_{SE1}) is calculated as the sum of two stacking energies (ΔG_{stack}), complementary base pairing interactions by nearest-neighbor thermodynamics (ΔG_{nn}), and the energies for phosphodiester bond formation (ΔG_{lig}). The lumped free energy for mismatches are calculated similarly exception that mismatches with a one nucleotide gap do not include stacking energy

contributions. Distributions of free energies of 100,000 randomly generated 4 bp overhang sticky ends are given for (B) complementary sticky ends, (C) self-misannealing sticky ends, (D) two different sticky ends that misanneal, and (E) two different sticky ends that misanneal with a one nucleotide gap. Distributions of free energies for the subset of palindromic sticky ends are shown in red in (B) and (C).

Since Type IIS restriction enzymes can, in principle, generate all possible 4-base sticky ends, our web calculator also computes free energies for sticky end mismatches to quantify the formation of unwanted byproducts. Four types of mismatches are calculated (Fig. 3.10A). The first two scenarios − mismatch between sticky end and itself (and complementary sticky end and itself) and mismatch between two different sticky ends (all 4 possible combinations) − form the majority of unwanted byproducts and are the main competitors to correct product formation. Mismatch energies are calculated in the same way as described above, except with an expanded set of nearest-neighbor parameters which include mismatches[24–28]. Figure 3.10C shows the distribution of free energies for self-annealing of 4-base overhang sticky ends, with palindromic sticky ends indicated in red. Compared to the free energies for sticky ends annealing to their complementary bases (Fig. 3.10B), this distribution is negatively skewed with a median value of -3.07 kcal/mol and a range from -15.15 to 2.84 kcal/mol. Due to the nature of palindromic sticky ends, the free energies for self-annealing and complementary base annealing are identical, which explains why palindromic sticky ends generated by Type IIP restriction enzymes result in significant ligation byproducts and are therefore not preferred for maximizing ligation efficiencies. The distribution of free energies between two randomly generated 4-base overhangs (Fig. 3.10D) shows significant spread to the right (with a

110

range of -14.88 to 2.88 kcal/mol) and a slight negative skew compared to Figure 3.10B. The last two types of mismatches – self-misannealing sticky ends and misannealing with another sticky ends with a one nucleotide gap – were included since there is evidence that DNA ligase can seal one-nucleotide gaps[38,39], although this occurs with lower efficiency and should therefore be a secondary consideration in the evaluation of sticky ends. Gap mismatches between two random sticky ends results in a distribution of free energies similar to that of mismatches without gaps, but with less spread (from -10.71 to 0.22 kcal/mol) due to the lack of stacking energies and one less nearest-neighbor interaction term. Overall, most non-palindromic sticky ends have mismatch free energies significantly less negative than their free energies for accurate interactions with their complementary ends, thus providing flexibility in rationally customizing sticky end sequences with high specificity.

3.4.9 Ligation calculator: sticky end free energy estimation by experiments

Alternatively, sticky end free energies can be experimentally determined through a series of simple ligation reactions containing two DNA fragments with the sticky end sequence of interest as described during model development. This method, though more time-consuming, can be more accurate since it enables fitting of free energies based on the specific ligation reaction conditions routinely employed by the user; it is especially recommended for applications involving general-use display vectors, since the sticky ends are typically constant for all applications involving these vectors. Using user-specified input molar concentrations of the two DNA fragments (up to ten independent

trials can be inputted), the calculator globally fits the free energy for a specific sticky end from the output mass ratios of the remaining reactants and full-length product, which can be obtained by quantifying band intensities on a DNA agarose gel. We previously performed nine trials for the sequences (C)CGAA(A) and (A)GCTT(G) using this method and obtained the free energy values of -12.00 and -12.23 kcal/mol (Section 3.4.4), which closely match the predictions from DNA thermodynamics described above (-11.77 and -12.02 kcal/mol).

3.4.10 Linear ligation calculator

With the free energies of up to two sticky ends in hand, the distribution of linear ligation products can be predicted (Fig. 3.11A). A major assumption in the calculation is that the chosen sticky ends are non-palindromic and do not form major byproducts. This is reasonable since the free energies of complementary base-paring (Fig. 3.10B) are, on average, ~8 kcal/mol lower than those for mismatch annealing (Fig. 3.10C, D and E); thus, byproducts can typically be safely neglected since each 1 kcal/mol increase in free energy corresponds to a 5.41-fold decrease in equilibrium affinity (equivalently, a 7.38×10^6-fold decrease in equilibrium affinity for a 8 kcal/mol increase in free energy).
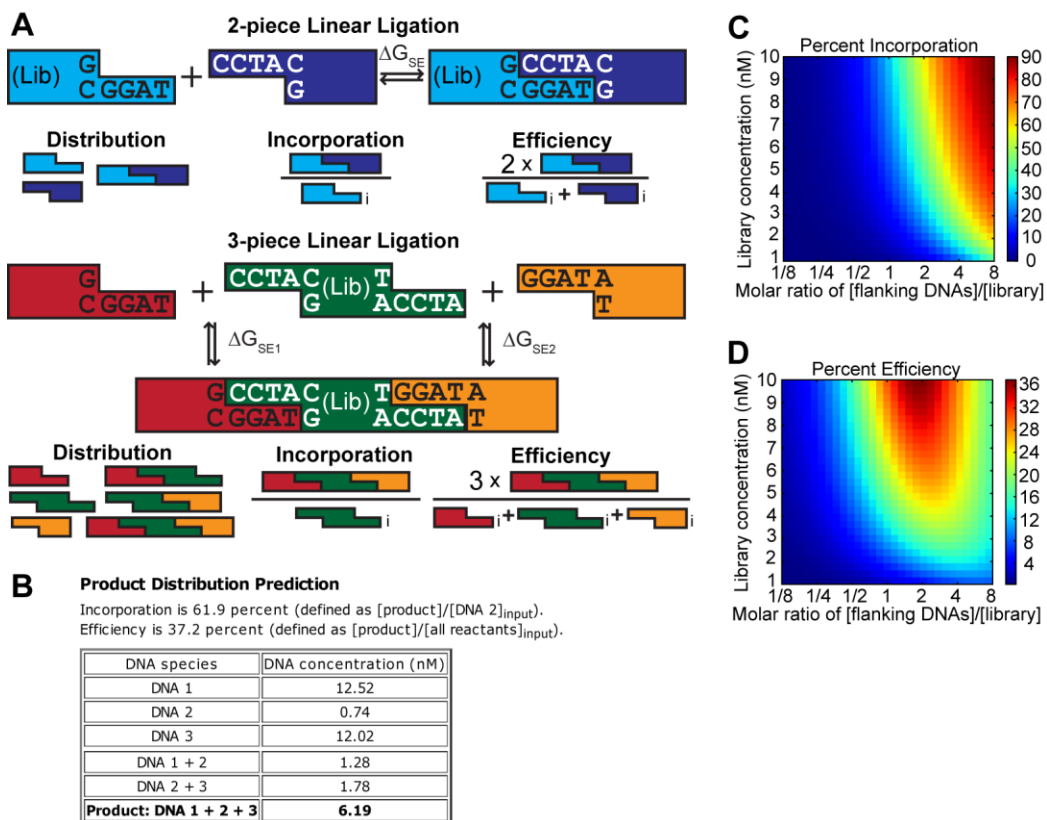
Figure 3.11. Schematic and representative output from the linear ligation calculator. (A) Schematics are given for two-piece and three-piece ligations as well as for the definitions of product distribution, incorporation, and efficiency. The free energy values used in this calculation (ΔG_{SE1} and ΔG_{SE2}) are obtained from the sticky end free energy calculator (Fig. 3.10). (B) Screenshot of linear ligation prediction output showing product distributions of a three-piece ligation reaction with 20, 10, and 20 nM of DNA 1, 2, and 3, respectively. Representative outputs from the linear ligation recommendation calculator showing (C) incorporation and (D) efficiency of ligation across a range of library concentrations and molar ratios for a three-piece ligation reaction with 10 nM library construct (DNA 2). ΔG_{SE1} ((G)CCTA(C)) of -11.10 kcal/mol and ΔG_{SE2} ((T)GGAT(A)) of -11.32 kcal/mol were used in all simulations.

The first part of the linear ligation calculator predicts product concentrations, incorporation, and efficiency from input DNA concentrations for two or three linear DNA fragments with one or two sticky end free energies, respectively (Fig. 3.11A). The

user can specify which DNA species is "valuable" (e.g., the fragment containing diversity in a DNA library assembly application) for calculation of percent incorporation. Again, the molar DNA concentrations can be converted from mass concentrations and volumes with the ligation excel worksheet. A representative output for three-piece DNA ligation with 10 nM valuable entity (DNA 2), 2:1 molar ratios for the other DNA fragments (DNA 1 and DNA 3), and the sticky end sequences (G)CCTA(C) and (T)GGAT(A) is shown in Fig. 3.11B. Product distribution reveals that most of DNA 2 (61.9%) is incorporated into full-length product, which is driven by the excess of the other two reactants (approximately half of DNA 1 and DNA 3 remain unreacted). The overall ligation efficiency (37.2%) indicates the percentage of all input reactants (DNA 1, 2 and 3) that end up in full-length products; the remainder are either reactants or intermediate products (DNA 1 + 2 or DNA 2 + 3).

The linear ligation calculator can also predict how incorporation and efficiency vary over a range of input concentrations and molar ratios. This feature is particularly useful for determining the optimal ligation setup for a specific application. Representative outputs of incorporation and efficiency for a three-piece linear ligation with 10 nM DNA 2 (valuable entity) and the sticky ends (G)CCTA(C) and (T)GGAT(A) are shown (Fig. 3.11C and D). For a given molar ratio, increasing reactant concentrations enhances incorporation and efficiency because the product is thermodynamically favored. For a given concentration, augmenting the molar ratio monotonically increases incorporation of the valuable entity, while ligation efficiency shows an optimum at intermediate molar ratios. At high molar ratios, incorporation becomes more favorable because any increase

in reactants (in the absence of byproducts) will shift the equilibrium towards product formation, but the overall efficiency drops after an optimal value since increased product formation cannot compensate for the excess reactants used. If Figure 3.11D represented a routine ligation reaction where the objective is often to maximize efficiency, the optimal reaction conditions would be to use a DNA 2 concentration close to the highest concentration available (10 nM) and ~2:1 molar ratios of [DNA 1]:[DNA 2] and [DNA 3]:[DNA 2]. For more stringent applications such as DNA library assembly in which library incorporation needs to be maximized without unnecessarily wasting other DNA fragments, the optimal setup based on Figure 3.11D would be to use the maximum concentration of the library fragment and ~4:1 molar ratios of the other fragments; using still higher molar ratios would only marginally increase incorporation but efficiency would markedly decline. The chosen molar concentrations and molar ratios can be directly entered into the ligation worksheet to determine the final ligation reaction mix based on available concentrations and volumes of DNA fragments. The calculator assumes that the valuable species is limiting and that excess amounts of the other fragments are available; if this is not the case, then a lower concentration of the valuable species should be entered into the calculator such that availability of the other DNA fragments is no longer an issue. If needed, the user can iterate between the ligation calculator and the worksheet to determine a final reaction scheme that is experimentally feasible.

3.4.11 Circular ligation calculator

Analogous to the linear ligation calculator, the circular ligation calculator predicts product distributions and offers ligation recommendations for insert-vector assembly. Using input molar concentrations of vector and insert, two sticky-end free energy values, and the number of base pairs of the full length vector-insert product, concentrations are calculated for all products containing up to four ligated molecules (any higher molecular weight molecules are assumed to be negligible) (Fig 3.12A). The probability of ligated vector-insert molecules (and vector-insert-vector-insert molecules) forming circular products depends on the effective concentration of the sticky end in the molecule itself compared to the concentrations of free vector and insert; this effective concentration is a function of the length of the ligated molecule, which is why the number of base pairs in the product is a necessary input[4,15]. Since circular ligation and transformation are typically performed in series, transformation efficiency (via colony counting) can serve as an experimental proxy for ligation efficiency. For DNA library applications, percent incorporation can be useful in estimating final library size and diversity. Accuracy gives the amount of correct vector-insert product formation compared to that of vector-insert-vector-insert circular product, although this value is likely a conservative estimate since the transformation efficiency of vector-insert-vector-insert circular products will typically be less than that for the desired product since transformation efficiency decreases with increasing plasmid size[40].
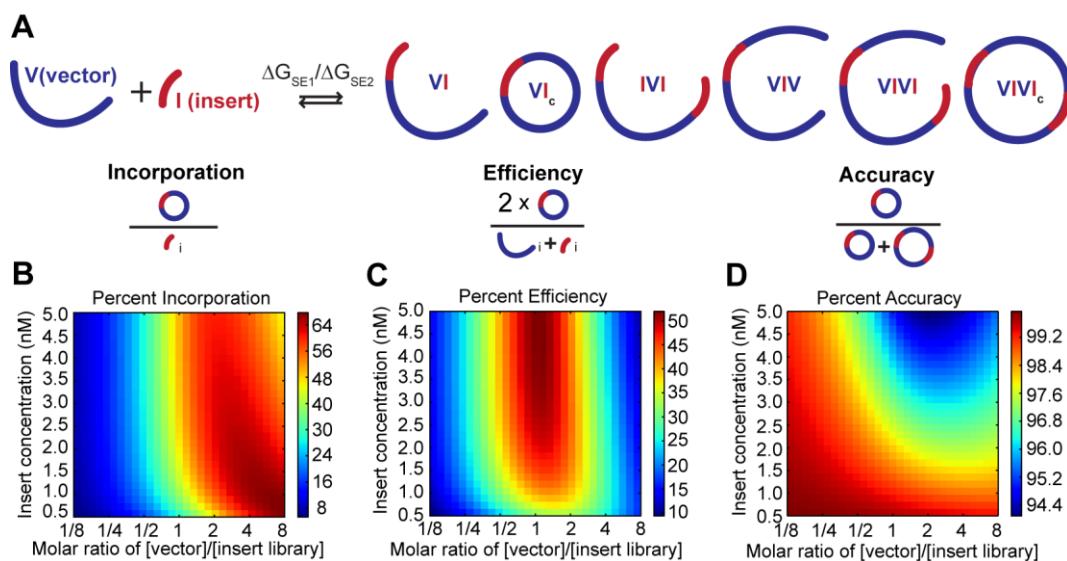
Figure 3.12. Schematic and representative output from the circular ligation calculator. (A) Schematics are given for a circular vector-insert ligation as well as for the definitions of incorporation, efficiency, and accuracy. Representative outputs from the circular ligation recommendation calculator showing (B) incorporation, (C) efficiency, and (D) accuracy of ligation across a range of insert concentrations and vector:insert molar ratios for a circular ligation with 5 nM library construct (insert). In all simulations, the total number of base pairs of vector and insert was 3500, ΔG_{SE1} was -11.10 kcal/mol, and ΔG_{SE2} was -11.32 kcal/mol.

The circular ligation calculator can also predict how incorporation (Fig. 3.12B), efficiency (Fig. 3.12C), and accuracy (Fig. 3.12D) vary over a range of insert concentrations and vector:insert molar ratios. At low vector:insert ratios, incorporation and efficiency are low because insert-vector-insert is preferentially formed over circular insert-vector products, but accuracy remains high since unbalanced molar ratios do not favor circular insert-vector-insert-vector molecules. Conversely, high vector:insert ratios result in low efficiency because vector-insert-vector products are preferentially formed and the effect on incorporation is mixed as the benefit of having excess vector to drive

product formation can be offset by an increase in vector-insert-vector byproduct. Maximum incorporation occurs when the vector concentration is around the value of the average equilibrium dissociation constant. In Fig. 3.12B, in which the equilibrium dissociation constants for the two sticky ends are 7.21 nM and 4.98 nM, maximal incorporation occurs in a region of concentration space that includes 3.5 nM insert with a 2:1 molar ratio (7 nM vector), 1.5 nM insert with a 4:1 ratio (6 nM vector), and 0.75 nM with an 8:1 ratio (6 nM vector). Maximum efficiency, on the other hand, is achieved at ~1:1 molar vector:insert ratio and when the insert concentration is approximately the value of the average equilibrium dissociation constant, which is around 5 nM in Fig. 3.12C. Accuracy only drops at high insert concentrations and approximately equimolar amounts of vector, although the absolute accuracy remains very high in all of concentration space and should therefore not be a major factor in optimizing circular ligation reactions. As an example optimization exercise, for normal cloning application where efficiency should be maximized, 5 nM insert with 1:1 vector:insert would be optimal; for library assembly where the balance between high incorporation and efficiency is important, 3.5 nM insert with 2:1 ratio would be preferable (Fig. 3.12B, C). Finally, the selected insert concentration and molar ratio can again be directly entered into the ligation worksheet to determine the final ligation reaction mix based on the availability of the amount of vector and insert.

3.5 Discussion

We have investigated the optimal ligation strategy for applications in both *in vitro* and *in vivo* directed evolution. We have reported up to 15- and 2.6-fold increases in desired products for circular and linear ligation reactions, respectively, when Type IIS restriction enzymes were used instead of Type IIP enzymes to generate DNA fragments, indicating that even order-of-magnitude improvements in library size can be achieved by this strategy. Furthermore, our simulation results can help to more narrowly focus optimization efforts and our web-based ligation calculator can enable easy determination of maximum ligation yields and library size using Type IIS restriction enzymes without extensive experimental testing.

To maximize ligation efficiency, next-generation display vectors should contain recognition sites of Type IIS restriction enzymes and cutting sites that are non-palindromic and have a matching number of G/C and A/T base-pairs. As demonstrated in this work, a convenient pair of non-palindromic cutting sites would have the same base compositions but would not exhibit any complementary to one another, thus eliminating undesired annealing. For linear ligations, the maximum concentration of digested reactants should be used, with an excess of upstream and downstream pieces to drive the equilibrium towards ligated products. For circular ligations, a concentration comparable to the K values of the sticky ends should be used with an insert:vector ratio between 1:2 and 1:1. These recommendations are different from the common ligation guidelines suggesting that insert:vector molar ratios greater than one be used and that the concentration of both should be kept within a certain range[6], which are not applicable to

fragments generated by Type IIS restriction enzymes or in directed evolution applications where the insert library is scarce. Our recommendations and even our cutting sites should be generally applicable to any ligation with T4 DNA ligase where the incorporation of insert is essential. Other custom cutting sites with four-base overhangs are not expected to deviate greatly from our results because our two sites CGAA and GCTT both contain 50% GC content and represent moderate ΔG values.

We presented two methods for estimating the free energy of ligation: 1) using DNA thermodynamics, and 2) by experimental measurement. The two methods gave similar results for the set of experimental conditions that we examined. If a specific application requires conditions sufficiently different from those examined in this work (e.g. different temperature, salt concentration or ligase), our model could still be used to capture ligation behavior as long as the ΔG values are properly parameterized (e.g. using the simple experimental procedure outlined herein). Importantly, once this characterization has been performed for a given pair of overhangs, the model could be applied for any system that uses these annealing sequences. Again, in directed evolution experiments, a common display vector is typically used for many applications, so the trained model could be repeatedly applied to all experimental systems using a redesigned display vector with Type IIS restriction sites.

Our model is not only practically useful for ligation applications, but also provides thermodynamic insight into the annealing of sticky ends. The ligation reaction is rarely analyzed quantitatively as there is no thermodynamic information readily available in the literature. We have estimated the free energy of ligation by obtaining free energy

values of DNA annealing from the nearest-neighbor method[23,41]. However, most tabulated values are optimized at $37^{\circ}C$ and 1 M NaCl, and salt and temperature corrections are only accurate to a certain extent and have not been systematically investigated for ligation. Empirical observations are available for specific ligation reactions but are not easily generalizable[42]. In contrast, considerable work has been done to characterize the free energy of sealing by ligase[16,43]. From the ΔG values obtained from model training, we could calculate the free energy of sticky-end annealing since we know the free energy contributed by ligase sealing. In principle, if a variety of sticky ends were investigated at different temperatures, free energy values for annealing could be obtained, providing insight into DNA thermodynamics.

In summary, for applications where ligation efficiency is crucial for experimental success, we propose that vectors be redesigned to exploit the desirable properties of commercially available Type IIS restriction enzymes. We have developed and validated a thermodynamic model of ligation to provide specific experimental strategies for *in vitro* and *in vivo* directed evolution applications, and our web-based calculator can readily be adapted to other ligation conditions. Last but not least, our model can be used to parse the complex thermodynamics of sticky-end annealing, further contributing to our understanding of this common molecular biology technique.

3.6 References

1. Sambrook, J. & Russell, D. W. Plasmids and their usefulness in molecular cloning. *Molecular Cloning: A Laboratory Manual* 1.157–1.159 (2001).

2. Tobias, A. V. Preparing libraries in *Escherichia coli*. *Directed evolution library creation: methods and protocols* 11–16 (2003).

3. Muller, P. Y., Studer, E. & Miserez, A. R. Molecular biocomputing suite: a word processor add-in for the analysis and manipulation of nucleic acid and protein sequence data. *BioTechniques* **31**, 1306–1313 (2001).

4. Dugaiczyk, A., Boyer, H. W. & Goodman, H. M. Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. *J. Mol. Biol.* **96**, 171–184 (1975).

5. Legerski, R. J. & Robberson, D. L. Analysis and optimization of recombinant DNA joining reactions. *J. Mol. Biol.* **181**, 297–312 (1985).

6. Revie, D., Smith, D. W. & Yee, T. W. Kinetic analysis for optimization of DNA ligation reactions. *Nucleic Acids Res.* **16**, 10301–10321 (1988).

7. Dardel, F. Computer simulation of DNA ligation: determination of initial DNA concentrations favouring the formation of recombinant molecules. *Nucleic Acids Res.* **16**, 1767–1778 (1988).

8. Szybalski, W., Kim, S. C., Hasan, N. & Podhajska, A. J. Class-IIS restriction enzymes–a review. *Gene* **100**, 13–26 (1991).

9. Roberts, R. J. *et al.* A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**, 1805–1812 (2003).

10. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **38**, D234–236 (2010).

11. Bath, A. J., Milsom, S. E., Gormley, N. A. & Halford, S. E. Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.* **277**, 4024–4033 (2002).

12. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004).

13. Barendt, P. A., Shah, N. A., Barendt, G. A. & Sarkar, C. A. Broad-specificity mRNA-rRNA complementarity in efficient protein translation. *PLoS Genet.* **8**, e1002598 (2012).

14. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).

15. Jacobson, H. & Stockmayer, W. H. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* **18**, 1600–1606 (1950).

16. Dickson, K. S., Burns, C. M. & Richardson, J. P. Determination of the free-energy change for repair of a DNA phosphodiester bond. *J. Biol. Chem.* **275**, 15828–15831 (2000).

17. Dreier, B. & Plückthun, A. Ribosome display: a technology for selecting and evolving proteins from large libraries. *Methods Mol. Biol.* **687**, 283–306 (2011).

18. He, M. & Taussig, M. J. Eukaryotic ribosome display with *in situ* DNA recovery. *Nat. Methods* **4**, 281–289 (2007).

19. Jones, E., Oliphant, T., Peterson, P. *et al.* SciPy: Open source scientific tools for Python. (2001).at <http://www.scipy.org/>

20. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

21. Florián, J., Šponer, J. & Warshel, A. Thermodynamic parameters for stacking and hydrogen bonding of nucleic acid bases in aqueous solution: ab initio/Langevin dipoles study. *J. Phys. Chem. B* **103**, 884–892 (1999).

22. Protozanova, E., Yakovchuk, P. & Frank-Kamenetskii, M. D. Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* **342**, 775–785 (2004).

23. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460–1465 (1998).

24. Allawi, H. T. & SantaLucia, J. Thermodynamics of internal G•T mismatches in DNA. *Biochemistry* **36**, 10581–10594 (1997).

25. Allawi, H. T. & SantaLucia, J. Nearest neighbor thermodynamic parameters for internal G•A mismatches in DNA. *Biochemistry* **37**, 2170–2179 (1998).

26.   Allawi, H. T. & SantaLucia, J. Nearest-neighbor thermodynamics of internal A•C mismatches in DNA: sequence dependence and pH effects. *Biochemistry* **37**, 9435–9444 (1998).

27.   Allawi, H. T. & SantaLucia, J. Thermodynamics of internal C•T mismatches in DNA. *Nucleic Acids Res.* **26**, 2694–2701 (1998).

28.   Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J. Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A•A, C•C, G•G, and T•T Mismatches. *Biochemistry* **38**, 3468–3477 (1999).

29.   Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nat. Methods* **4**, 269–279 (2007).

30.   Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077–1080 (1988).

31.   Harada, K. & Orgel, L. E. Unexpected substrate specificity of T4 DNA ligase revealed by *in vitro* selection. *Nucleic Acids Res.* **21**, 2287–2291 (1993).

32.   Cherepanov, A. V. & de Vries, S. Kinetics and thermodynamics of nick sealing by T4 DNA ligase. *Eur. J. Biochem.* **270**, 4315–4325 (2003).

33.   Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4937–4942 (1997).

34.   Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297–12302 (1997).

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

36.   Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10444–10448 (1993).

37.   Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

38.   Goffin, C., Bailly, V. & Verly, W. G. Nicks 3' or 5' to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase. *Nucleic Acids Res.* **15**, 8755–8771 (1987).

39. Gu, J. *et al.* XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.* **26**, 1010–1023 (2007).

40. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).

41. SantaLucia, J., Allawi, H. T. & Seneviratne, P. A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562 (1996).

42. Ferretti, L. & Sgaramella, V. Temperature dependence of the joining by T4 DNA ligase of termini produced by type II restriction endonucleases. *Nucleic Acids Res.* **9**, 85–93 (1981).

43. Sugino, A. *et al.* Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base-paired ends. *J. Biol. Chem.* **252**, 3987–3894 (1977).

CHAPTER 4

Streamlined protocol for mRNA display and the display of monomeric IgG Fc

4.1 Introduction

In vitro directed evolution techniques, such as ribosome display[1] and mRNA display[2], are powerful tools for protein engineering, capable of handling libraries containing up to ~10^{14} members. The high library diversity is possible because these *in vitro* display technologies eliminated the need for transformation to generate libraries[3]. Moreover, PCR-based mutagenesis is extremely convenient as they are inherent to the display procedure[3]. Ribosome display involves *in vitro* translation of mRNA sequences lacking stop codons to generate ternary complexes in which each mRNA and its corresponding protein are noncovalently associated *via* a stalled ribosome. This requires that any selections carried out with these ternary complexes be performed under conditions that maintain the integrity of the ribosome[4]. By contrast, mRNA display uses an mRNA-DNA-puromycin fusion as its template for *in vitro* translation, which results in a covalent puromycin linkage between the mRNA and the corresponding nascent peptide[5]. This creates a highly stable selection particle, which is particularly useful for performing *in vitro* selections in harsh environments that are not compatible with ribosome display.

Since its introduction in 1997, mRNA display has been used in a wide variety of applications, such as the investigation of protein-protein interactions[6] and the development of peptides[7], enzymes[8], scFvs[9], and novel binders based on alternative scaffolds[10]. Despite the potential utility of mRNA display in a plethora of protein engineering applications, the technically demanding nature of this method has precluded its widespread use. Currently established protocols for mRNA display take ~4–7 days per round, which includes several lengthy purification steps. We therefore sought to simplify and streamline the mRNA display procedure to make it more broadly accessible to the scientific community. The time required for 1 round is reduced to 2 days with our protocol, thus simplifying and accelerating mRNA display experiments.

We then displayed a human monomeric IgG1 Fc (mFc) using this streamlined protocol to investigate its binding to the neonatal Fc receptor (FcRn). The display of mFc was chosen as a proof of principle to illustrate possible variations and special features of mRNA display, since this is an ideally suited platform for future engineering of mFc properties, which we explain below. mFc also represents a very important fusion partner to therapeutic proteins, so engineered versions, tailored for specific applications, should have great potential in the clinic.

As its name suggests, the neonatal Fc receptor is involved in the transport of IgG antibodies from a mother to her child, across the placenta and the intestine[11]. In both cases, FcRn binds the Fc portion of IgG with high affinity at acidic pH (< 6.5) but not at a physiological pH (7.4)[12,13]. In adult humans, FcRn is expressed in several cell types, including the vascular endothelium, antigen-presenting cells (APCs), and the intestine[11].

It has been shown that FcRn present on the vascular endothelium mediates IgG homeostasis by rescuing IgGs that have been passively internalized by pinocytosis and recycling them back to the cell surface, thus prolonging the serum half-life of IgGs[14]. On the other hand, FcRn expressed on intestinal epithelial cells[15] can transport IgG across the lumen by receptor-mediated transcytosis[16]. The unique pH-dependent binding of Fc and FcRn, which in turn modulates *in vivo* IgG levels or mediates transcytosis, has made this interaction the subject of many engineering efforts[17–22].

Recently, Ying et al.[23] developed three human monomeric IgG1 Fcs – mFc.1, mFc.23, and mFc.67 – that are highly soluble and retain pH-dependent binding to human FcRn. Eliminating the need for the dimeric form of Fc for receptor binding opens up many possibilities for developing much smaller therapeutic antibodies and for fusing the Fc domain to other active proteins to prolong half-life or facilitate transcytosis. In our proof-of-principle studies, we chose mFc.67 as the starting template due to its high thermal stability and high affinity to FcRn at pH 6.0[23]. Based on previous mutational studies on the IgG1 Fc fragment, we also included two sets of mutations that could potentially improve mFc.67 binding to FcRn: M252L/T256F (LF) double mutations[18] and M252Y/S254T/T256E (YTE) triple mutations[24].

The mRNA display platform is ideally suited for displaying monomeric Fcs against FcRn. We previously discussed evolving the Fc domain for selective FcγRII binding by displaying full length IgG in a bacterial display system (Chapter 2). However, mRNA display can be used for the smaller and simpler monomeric Fc domain since this moiety is sufficient for FcRn binding. Our streamlined mRNA display method uses a

bacterial *in vitro* translation system, which is potentially problematic for studying glycosylated proteins or domains, but notably, glycosylation state does not affect the interactions between Fc and FcRn[25]. mRNA display also tolerates a wider range of pH values compared to ribosome display, due to the covalent nature of the mRNA-puromycin-protein complexes; thus, pH stability selections and/or pH-dependent binding selections (e.g., binding at pH 6.0 and elution at pH 7.4 for Fc/FcRn) can be carried out more easily and robustly. Moreover, if future engineering of trafficking processes such as transcytosis are to be performed on mammalian cells, the minimal, covalently linked selection particles offer the greatest chance of being internalized and trafficked intact. In this chapter, we have successfully demonstrated binding of mFc.67, mFc.67(LF), and mFc.67(YTE) against FcRn using our streamlined mRNA display protocol.

4.2 Materials and methods

4.2.1 Materials

Oligonucleotides were purchased from Integrated DNA Technologies (IDT; Coralville, IA) unless otherwise specified (Table 4.1). DNA purification was performed using agarose gel electrophoresis with SYBR Safe (Invitrogen, Carlsbad, CA) and the QIAquick gel extraction kit (Qiagen, Valencia, CA). Restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase (used for all ligations), and Phusion DNA polymerase (used for all PCRs) were purchased from New England Biolabs (Ipswich, MA) and used as recommended by the manufacturer. DNA and RNA concentrations

were determined by absorbance readings at 260 nm using a NanoDrop 1000

spectrophotometer (Thermo Scientific, Waltham, MA).

Table 4.1. Plasmids used in this study.

Plasmid	Description	Source
pRDV2:Off7	pRDV2 containing the FLAG-tagged Off7 gene; Amp^r	Chapter 3
pRDV2:H10-2-G3	pRDV2 containing the H10-2-G3 gene; Amp^r	This work
pRDV2:Off7m	pRDV2 containing the FLAG-tagged Off7 gene modified with silent mutations to eliminate repeating nucleotides; Amp^r	This work
PIGG-P14	Eukaryotic expression vector with the hinge region, CH2 and CH3 domain of IgG Fc	Courtesy of Dr. Don Siegel
pRDV2:Hinge-mFc_E	pRDV2 containing monomeric Fc with the hinge region; Amp^r	This work
pRDV2:mFc.67	pRDV2 containing the mFc.67 gene; Amp^r	This work
pRDV2:mFc.67_intB	pRDV2 containing the mFc.67 gene (with internal BsaI silent mutation); Amp^r	This work
pRDV2:mFc.67_intB_LF	pRDV2 containing the mFc.67 gene with M252L/T256F double mutations (and with internal BsaI silent mutation); Amp^r	This work
pRDV2:mFc.67_intB_YTE	pRDV2 containing the mFc.67 gene with M252Y/S254T/T256E triple mutations (and with internal BsaI silent mutation); Amp^r	This work

4.2.2 Plasmid construction

Plasmids used in this work are listed in Table 4.2. To create DNA constructs suitable for

in vitro transcription, the genes encoding H10-2-G3 and Off7 were first cloned into

pRDV2 (from Chapter 3). The cDNA encoding H10-2-G3[26] was synthesized by IDT and

PCR amplified with primers 1/2. The resulting product was digested with BamHI and

HindIII, and ligated into similarly digested pRDV2:Off7. The ligation mixture was

transformed into chemically competent *E. coli* XL1-Blue[27]. Individual colonies were

selected for standard liquid culturing and plasmid extraction procedures using a Qiagen

miniprep kit. The purified plasmid, pRDV2:H10-2-G3, was sequence verified. The cDNA encoding Off7[28] was modified with silent mutations to eliminate repeating nucleotides in order to minimize nonspecific amplification products in PCR. This gene was synthesized by IDT and PCR amplified with primers 3/4. The resulting product was BamHI/HindIII digested and ligated into pRDV2 as described above to obtain pRDV2:Off7m, which was also purified by miniprep and sequence verified.

Table 4.2. Oligonucleotides used in this study.

Number	Name	Nucleotide Sequence (5' → 3') [a]
1	H10_BamHI_f	GACAAAGGATCCGACCTGGGTAAAAAACTGCTG
2	H10_HindIII_r	AGACCCAAGCTTTTGCAGGATTTCAGCCAG
3	pRDVFLAG_BamHI_f	GATGACGATGACAAAGGATCC
4	pRDV_BbsI_r	GGCCACCGGAAGACCCAAGC
5	T7_no_BsaI	ATACGAAATTAATACGACTCACTATAGGGACACCACAACGG
6	DNA-puromycin linker	AAAAAAAAAAAAAAAAAAAAAAAAAAACC-puromycin
7	GS_splint	TTTTTTTTTTGGCCACCGGAA
8	mFc.67_f	AGACACCGAAAGAAGGAGATATATCCATGGCACCTGAACTCCTG
9	mFc.67GGG_f	ACCAGGTCAGCCTGAGCTGCCTGG
10	mFc.67GGG_r	CCAGGCAGCTCAGGCTGACCTGGT
11	pRDV_BbsI_f	AATAATTTGAAGACACCGAAAGAAG
12	mFc.67BsaImut_f	GTGCAAGGTGTCCAACAAAG
13	mFc.67BsaImut_r	CTTTGTTGGACACCTTGCAC
14	LF_f	GACACCCTCCTGATCTCCCGGTTTCCTGAGGTC
15	LF_r	GACCTCAGGAAACCGGGAGATCAGGAGGGTGTC
16	YTE_f	GACACCCTCTATATCACCCGGGAACCTGAGGTC
17	YTE_r	GACCTCAGGTTCCCGGGTGATATAGAGGGTGTC
18	pRDV_Off7stops_r	TTGCAGGATTTCAGCCAGG

[a] Restriction sites are underlined. Mutation sites are double underlined.

To construct mFc.67 and its variants, pRDV2:Hinge-mFc_E was first constructed from PIGG-P14 (courtesy of Dr. Don Siegel, University of Pennsylvania). The hinge region, CH2, and CH3 domains were each PCR amplified and assembled to eliminate introns from the original construct in PIGG-P14; this product was cloned into pRDV2 to give pRDV2:Hinge-mFc_E. pRDV2:mFc.67 was then constructed by overlap extension PCR with 1) a PCR amplicon of pRDV2:Hinge-mFc_E using primers 8/10, and 2) a gene fragment of mFc.67 (from residue 60 onwards[23]), synthesized by IDT, and PCR amplified with primers 9/4. The assembled product was amplified with primers 11/4, digested with BbsI, and ligated into similarly digested pRDV2:Off7. To remove the internal BsaI site in mFc.67 to facilitate cloning in future selection experiments, pRDV2:mFc.67 was PCR amplified with two separate primer sets (11/13 and 12/4) to create the two fragments assembled in overlap extension PCR using primers 11/4. The final product was digested with BbsI and ligated into pRDV2 to obtain pRDV2:mFc.67_intB. For incorporation of the M252L/T256F (LF) mutations[18] or the M252Y/S254T/T256E (YTE) mutations[24], pRDV2:mFc.67_intB was PCR amplified with primers 11/15 (LF_piece1) and 14/4 (LF_piece2) or 11/17 (YTE_piece1) and 16/4 (YTE_piece2), respectively. LF_piece1 and LF_piece2 were assembled in overlap extension PCR with primers 11/4, and YTE_piece1 and YTE_piece2 were assembled with the same primers. The PCR products were digested with BbsI and ligated into pRDV2 to give pRDV2:mFc.67_intB_LF and pRDV2:mFc.67_intB_YTE, respectively.

4.2.3 *In vitro* transcription

Primers 5 and 4 were used to amplify plasmids pRDV2:H10-2-G3, pRDV2:Off7m, pRDV2:Off7, pRDV2:mFc.67_intB, pRDV2:mFc.67_intB_LF and pRDV2:mFc.67_intB_YTE. The H10-2-G3 and Off7 products were gel purified and eluted with 50 µl EB (Qiagen), while the products of mFc.67 and its variants were used directly for *in vitro* transcription. Gel-purified products (46 µl) or PCR products (22.5 µl) were used directly for transcription as described previously[29]. Lithium chloride (LiCl) precipitation was used to purify the newly transcribed mRNA as described[29] and the product was reconstituted in 20 µl RNase-free water.

4.2.4 Ligation of mRNA to DNA-puromycin linker

A DNA-puromycin linker[5] (oligonucleotide 6 in Table 4.2) was purchased from Biosearch Technologies (Novato, CA) and an oligonucleotide splint (oligonucleotide 7 in Table 4.2) was also used. The DNA-puromycin linker was phosphorylated according to Keefe et al.[5] and desalted using illustra ProbeQuant G-50 Micro Columns (GE Healthcare, Piscataway, NJ). The ligation was adapted from Keefe et al.[5] to include: 1 nmol mRNA, 1 nmol 5′ phosphate DNA linker, 1 nmol GS_splint, water up to 86 µl, and 2 µl RNasin Plus RNase inhibitor (Promega, Madison, WI). These components were heated at 95°C for 2 min. Then, 10 µl 10× T4 DNA ligase buffer was added. The reaction was vortexed and then cooled on ice for 10 min. The tubes were removed from ice for 5 min, at which time 4 µl high-concentration T4 DNA ligase (2000 U/µl) was added. The

reaction was incubated at room temperature for 50 min, and then 400 µl 7 M urea was added and the tube was lightly vortexed.

4.2.5 Purification of mRNA-DNA-puromycin

Each ligation reaction was loaded onto a 100 kDa cutoff Amicon Ultra-0.5 ml centrifugal filter unit (Millipore, Billerica, MA) and ultrafiltration was performed at 2000g for 15 min. After addition of 7 M urea (200 µl), the device was centrifuged again. RNase-free water (500 µl) was then added, followed by an additional spin. This washing step with water was repeated twice more to remove essentially all of the urea.

4.2.6 PURExpress reaction

The PURExpress *in vitro* protein synthesis kit (New England Biolabs; NEB) was used for translation reactions. For Her2 selection experiments, the master mix included 10 µl Solution A, 7.5 µl Solution B, 0.5 µl RNasin Plus, and 5 µl RNase-free water. The master mix was then split into 4 reaction tubes with 5.5 µl each, and 60 pmol total mRNA was added to each reaction (ribosome content of PURExpress is 15 pmol per reaction). In the first round, the input in the 4 reactions contained the mRNAs encoding H10-2-G3 and Off7m in molar ratios of 1:10, 1:100, 1:1,000, and 1:10,000, with the mRNA encoding Off7m in molar excess in all cases. In each subsequent round, the input DNA from the previous round was transcribed and ~60 pmol total mRNA was added to each translation reaction. Translation was performed for 30 min at 37$^{\text{o}}$C and then the reaction was kept at

134

room temperature for another 10 min. In the first round, 1 μl of this translation reaction was saved and kept on ice for direct RT-PCR. Then, 100 μl ice-cold TBS-cas [TBS (50 mM Tris-HCl pH 7.4, 150 mM NaCl)[4] with 0.5% casein (Sigma-Aldrich, St. Louis, MO)] was added to the remaining translation reaction and the solution was kept on ice.

For FcRn selection experiments, PURExpress master mix included 10 μl Solution A, 7.5 μl Solution B, 1 μl Disulfide bond enhancer 1 (NEB), 1 μl Disulfide bond enhancer 2 (NEB), 0.5 μl RNasin Plus, and 3 μl RNase-free water. mRNA encoding mFc.67 (or the LF or YTE variants) was added to the master mix to achieve a 1:1 molar ratio of mRNA:ribosomes. Translation was performed as described above and then each reaction (one per well) was stopped by adding 100 μl of one of four ice-cold panning buffers: TBST-cas-6 [TBS-6 (50 mM Tris-HCl pH 6.0, 150 mM NaCl) with 0.5% casein and 0.05% Tween 20 (Bio-Rad, Hercules, CA)], TBST-bsa-6 [TBS-6 with 0.5% bovine serum albumin (BSA; Roche, Indianapolis, IN) with 0.05% Tween 20], TBST-cas (TBS-cas with 0.05% Tween 20), or TBST-bsa.

4.2.7 Affinity selection

Affinity selection for Her2 was performed on NUNC Maxisorp plates (Thermo Fisher Scientific, Rochester, NY). Each well was coated with 1.4×10^{13} Her2 receptors (Sino Biological, Beijing, China) in 100 μl TBS for at least 16 h at 4°C, washed 3 times with 300 μl TBS, blocked with casein (300 μl TBSC added) at room temperature for 1 h with shaking, and then washed 3 times with TBS and once with TBS-cas. The translation

reaction in TBS-cas was added to each well and panned for 1 h at 4°C with shaking to allow binding. The wells were subjected to 3 quick washes with TBS-cas and 3 additional 5-min washes prior to reverse transcription.

For FcRn selection, target immobilization was carried out as described above for Her2, except that 1.2×10^{13} mouse FcRn receptors (R&D Systems #6775-FC, Minneapolis, MN) were present in each well. For negative controls, 100 μl TBS without FcRn was added to each well at 4°C for 16 h. Wells were washed 3 times with 300 μl TBS, blocked with casein or BSA (300 μl TBS-cas or TBS-bsa) at room temperature for 1 h, and washed 3 times with TBS and once with the appropriate panning buffer. The translation reaction (100 μl) was added to each well and panned at 4°C for various times, as specified in the experiment. The wells were washed 6 times with TBST-cas-6, TBST-bsa-6, TBST-cas, or TBST-bsa at 4°C or 37°C, with washing times as specified.

4.2.8 Reverse transcription

For Her2 selection, reverse transcription was performed *in situ* using AffinityScript reverse transcriptase (Agilent Technologies, Santa Clara, CA) and reverse primer 4. A reverse transcription protocol adapted from He and Taussig[30] was followed. Briefly, Solution 1 (12.75 μl water, 0.25 μl 100 μM primer 4, and 0.5 μl RNasin Plus) was pipetted into the wells (or into 1 μl saved translation reaction in the first round), incubated at 70°C for 5 min, and allowed to cool at room temperature for 10 min. Solution 2 [2 μl 10× AffinityScript buffer, 2 μl 0.1 M dithiothreitol, 2 μl dNTPs (5 mM

each), and 0.5 µl AffinityScript reverse transcriptase] was then added and the reaction was incubated at 25°C for 10 min, 50°C for 1 h, and then heat-inactivated at 70°C for 15 min. Although not necessary for highly oversampled libraries such as the ones used in this study, this reaction could be scaled up to recover rare sequences that may be immobilized higher on the well walls after selection.

For FcRn selection, pH-based elution was performed before reverse transcription. To elute the mRNA-puromycin-mFc complexes, 200 µl ice-cold TBS (pH 7.4) containing 50 µg/ml *S. cerevisiae* RNA (Sigma-Aldrich) was added and incubated for 1 min with shaking. The eluate was then added to 400 µl lysis buffer and purified with the High Pure RNA isolation kit (Roche) according to the manufacturer's instructions. The purified mRNA was eluted in 30 µl RNase-free water, denatured at 70°C for 10 min, and reverse transcribed as described[29], except that AffinityScript reverse transcriptase and reverse primer 4 were used.

4.2.9 PCR to verify enrichment

After each round of Her2 selection, the products of reverse transcription (2 µl) were amplified by PCR with primers 5/4 for 35 cycles in a 50 µl reaction (30 cycles for saved translation reaction without selection in the first round). Each PCR product was subjected to gel purification by excising a rectangle encompassing both full-length DARPin product bands and this purified DNA was then reamplified with T7_no_BsaI and pRDV_BbsI_r to obtain enough product for transcription for the next round. For

analytical gel electrophoresis, PCR products were visualized in agarose gels with ethidium bromide staining and imaged using an ultraviolet transilluminator and CCD camera. After FcRn selection, the products of reverse transcription (2 µl) were PCR amplified with primers 5/4 for 23 or 25 cycles in a 20 µl reaction and visualized similarly.

4.2.10 Determining the effect of Tween 20

PURExpress was used to translate the mRNA encoding Off7 (from pRDV2:Off7), with 5 pmol mRNA and ribosomes for each reaction condition. The reaction was translated at 37°C for 30 min, and allowed to incubate at room temperature for 10 min. Then, 500 µl ice-cold TBST-cas or TBS was added to the master translation reaction and loaded onto a 100 kDa cutoff Amicon Ultra-0.5 ml centrifugal filter unit. Ultrafiltration was performed at 2000g for 15 min. After addition of 500 µl TBST-cas or TBS, the device was centrifuged again. This washing step was repeated twice more. *In situ* reverse transcription was carried out as described above with primers 7 or 4, and 2 µl product was PCR amplified for 30 cycles with primers 11/7 or 11/4, respectively.

4.3 Results

4.3.1 Streamlined mRNA display protocol

A traditional mRNA display schematic is given in Figure 4.1[2,5,31], with our modifications highlighted in red. First, we redesigned the DNA construct with a different 5′ end (Fig.

4.1a). The 5′ untranslated region (5′ UTR), taken from the modified ribosome display vector pRDV2 (Chapter 3), contains a T7 promoter and a Shine-Dalgarno ribosome binding site (RBS). This 5′ UTR drives strong transcription by T7 RNA polymerase, as well as efficient translation by *Escherichia coli* ribosomes, and has been shown to work in both crude extract[1] and a cell-free reconstituted system[32]. Since the minimal translation system used in this work (discussed below) contains reconstituted components from the *E. coli* translational machinery, it is necessary to replace the tobacco mosaic virus translational enhancer commonly used with the eukaryotic rabbit reticulocyte lysate in mRNA display with the Shin-Dalgarno RBS to allow efficient translation with the prokaryotic *E. coli* components[33,34].
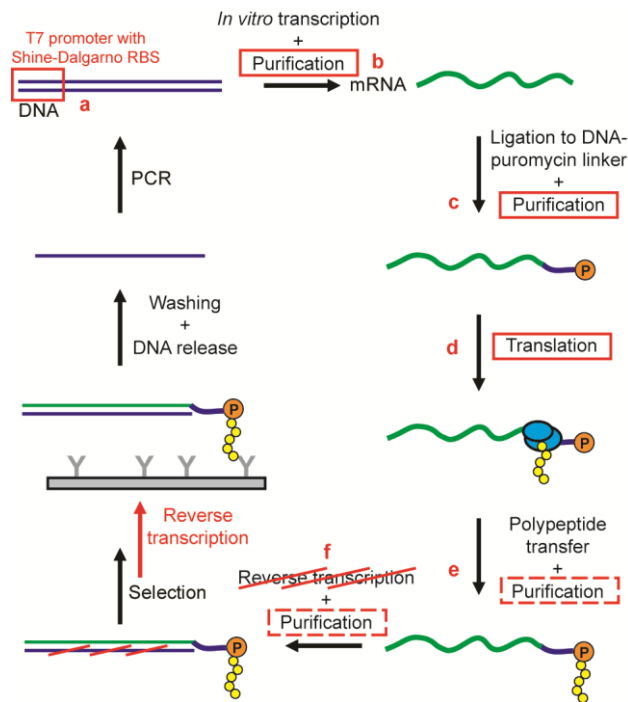
Figure 4.1. Changes in the streamlined mRNA display protocol. A traditional mRNA display protocol is shown[2,5,31], with our modifications highlighted in red: (a) 5' UTR with T7 promoter and Shine-Dalgarno RBS; (b) mRNA purification with LiCl precipitation instead of lengthy PAGE purification; (c) mRNA-

DNA-puromycin purification with ultrafiltration instead of lengthy PAGE purification; (d) *in vitro* translation with minimal, reconstituted translation system (PURExpress) instead of crude lysate to achieve higher library diversity with less nuclease and protease activity, which, in turn, can (e) eliminate need for purification of selection particles; (f) postpone reverse transcription, thus eliminating an additional purification step as well as avoiding a high-temperature incubation step of the displayed polypeptide prior to selection.

After *in vitro* transcription, mRNA is traditionally purified using denaturing PAGE (Fig. 4.1b)[2,5]. While effective at separating full-length RNA transcripts from truncated RNA and PCR primers, this method requires lengthy steps and manual excision from the gel. Instead, we use lithium chloride (LiCl) precipitation to isolate transcribed mRNA, a technique that has been used to purify mRNA prior to in vitro translation[4]. LiCl selectively precipitates RNA from unwanted DNA[35], smaller RNA, unincorporated NTPs, and proteins[36], thus enabling rapid purification while minimizing the number of processing steps during which detrimental nucleases and proteases could be introduced.

We have also replaced another PAGE purification step after ligating mRNA to the DNA-puromycin linker (Fig. 4.1c). Instead, we utilize the repertoire of commercially available ultrafiltration devices to quickly purify the mRNA-DNA-puromycin molecules from the DNA splint and unligated DNA-puromycin linkers, which can interfere with the subsequent translation reaction if not removed[5]. Although ultrafiltration cannot separate mRNA-DNA-puromycin molecules from unligated mRNA molecules, this limitation is also encountered in PAGE purification with templates longer than 500 nucleotides[5]. Furthermore, any mRNA molecules lacking puromycin will not be able to covalently attach to their corresponding polypeptide sequences and therefore will not be erroneously

enriched during the selection step. We have successfully purified the desired product using Amicon Ultra-0.5 ml (Millipore) or Vivaspin 500 (Sartorius) ultrafiltration columns, which are available in a range of molecular weight cutoffs (30−300 kDa) that enable their use across a broad range of construct sizes. In the current proof-of-principle studies, a 100 kDa ultrafiltation device can effectively remove the splint and unligated linkers (9.5 kDa), while retaining the mRNA-DNA-puromycin molecules (170−210 kDa) with a yield of 85−90%.

In vitro translation reactions for mRNA display and ribosome display have typically relied on crude cell lysates for the necessary translational machinery. However, cell lysates are not ideal for reproducibly achieving high-complexity libraries because they contain endogenous RNases and proteases and time-intensive optimization procedures must be applied to correct for the inherent batch-to-batch variability. These limitations can be overcome by using a minimal *E. coli*-based translation system reconstituted from purified components (PURExpress, New England Biolabs) (Fig 4.1d). The fully defined PURExpress system offers better reproducibility, lack of nucleases and proteases, and a high concentration of ribosomes. These characteristics maximize the number of intact selection particles, increasing the likelihood of a successful selection. The resulting mRNA display particles can be used directly in selection experiments without additional purification steps (Fig. 4.1e).

Selection is carried out after translation and followed by reverse transcription (RT). We delay RT until after selection because the displayed proteins may be adversely affected by the elevated temperatures required during the RT process, including initial

thermal denaturation of the RNA template and high-temperature incubation with the reverse transcriptase (Fig. 4.1f). Postponing RT also eliminates an additional purification step, as RT products can be used directly in subsequent PCR amplification but require a gel filtration step if RT is done prior to selection, further streamlining the procedure. However, since it is possible that delaying the RT step may allow for enrichment of RNA aptamers in addition to polypeptide binders in some affinity selections, an additional selection round can be performed with RT prior to selection to specifically isolate the polypeptide binders. Since the polypeptide binders should be substantially enriched at this point, the adverse effects of RT can be better tolerated than in initial rounds in which binders are scarce. Alternatively, DNA primers that are complementary to conserved regions of the mRNA can be added prior to selection to destabilize any RNA aptamers in the pool.

4.3.2 Affinity selection against Her2 using binary libraries

We performed a proof-of-principle selection to test our new streamlined mRNA display procedure. Two designed ankyrin repeat proteins (DARPins), Off7m[28], a maltose-binding protein (MBP) binder, and H10-2-G3[26], a Her2 binder, were assembled in mRNA display format and their mRNAs were combined at 4 different molar ratios (1:10 to 1:10,000 H10-2-G3:Off7). These binary libraries were subjected to selection against immobilized Her2. After 2−3 rounds of our streamlined mRNA display protocol, the Her2-binding DARPin was selectively enriched, as evidenced by agarose gel electrophoresis visualization of the RT-PCR products (Fig. 4.2). Since RT-PCR of the original 1:100

mRNA mixture shows no discernible signal for H10-2-G3 but the original 1:10 mixture shows a faint yet clear lower band, we infer that the enrichment factor in these selection experiments is as high as 2 orders of magnitude. On the basis of the gel, we estimate an average enrichment of approximately 50- to 100-fold per selection cycle.
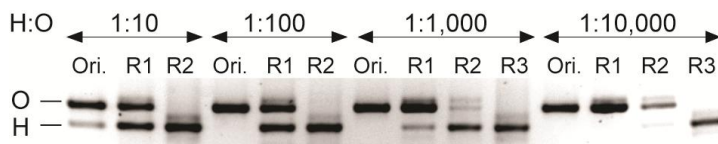
Figure 4.2. Model DARPin selections using our streamlined mRNA display protocol. Four binary libraries containing mRNAs encoding H10-2-G3 (H = a Her2 binder) and Off7m (O = an MBP binder) at different molar ratios (1:10, 1:100, 1:1,000, and 1:10,000) were subjected to 2−3 rounds of selections against immobilized Her2. The original library (Ori.) and selective enrichment of H10-2-G3 in the RT-PCR product after each round of mRNA display (R1, R2, and R3) is visualized by agarose gel electrophoresis. Each selection round took 2 days in the laboratory.

4.3.3 Variations of the mRNA display protocol

Several variations of the mRNA display protocol were investigated. The presence of Tween 20 in the reaction during the panning and washing steps (added with the TBS and blocking agent) has a pronounced effect on the resulting DNA products. Tween 20 is routinely included in ELISA and display technologies because this detergent can help saturate vacant sites and minimize nonspecific background binding[37], and this has been seen in mRNA display as well[9,38]. However, our preliminary studies using mRNA display to present Off7 to immobilized MBP showed that RT-PCR products contain almost no full-length DNA product, but many non-specific, low molecular weight bands, although there were indications of target-specific enrichment since these product bands were

stronger in the positive wells than in the negative ones (Fig. 4.3A). To more directly see this effect, RT-PCR was performed directly on the translation reaction, diluted in TBS with or without Tween 20, without an actual selection step (no panning, washing, or elution). As shown in Figure 4.3B, the addition of Tween 20 after translation adversely affected the yield of the correct RT-PCR product, instead resulting in amplification of lower molecular weight bands. Denaturing the mRNA at 70°C before RT, as part of the standard procedure, seemed to exacerbate this effect. In the absence of Tween 20, product bands of the correct size can be clearly seen (Fig. 4.3C). The appearance of 2 lower molecular weight species below the correct product band in Figure 4.3C have been identified by DNA sequencing as Off7 DARPin with only two of the three internal repeats and one of the three internal repeats, respectively. These byproducts may have arisen during the RT step if secondary structure in the mRNA allowed mispriming at the repeat junctions since the nucleotide sequences of the repeats are similar. Thus for our selection experiment against Her2 in the previous section, we constructed an Off7m, an Off7 variant with rationally introduced silent mutations to remove these repeating nucleotide motifs, which successfully eliminated the lower molecular weight byproducts. As we will discuss later, the effect of Tween 20 in a given selection strategy is likely to be protein specific; therefore, its inclusion in mRNA display protocols should be tested in pilot studies before starting selection experiments with new libraries and/or targets.
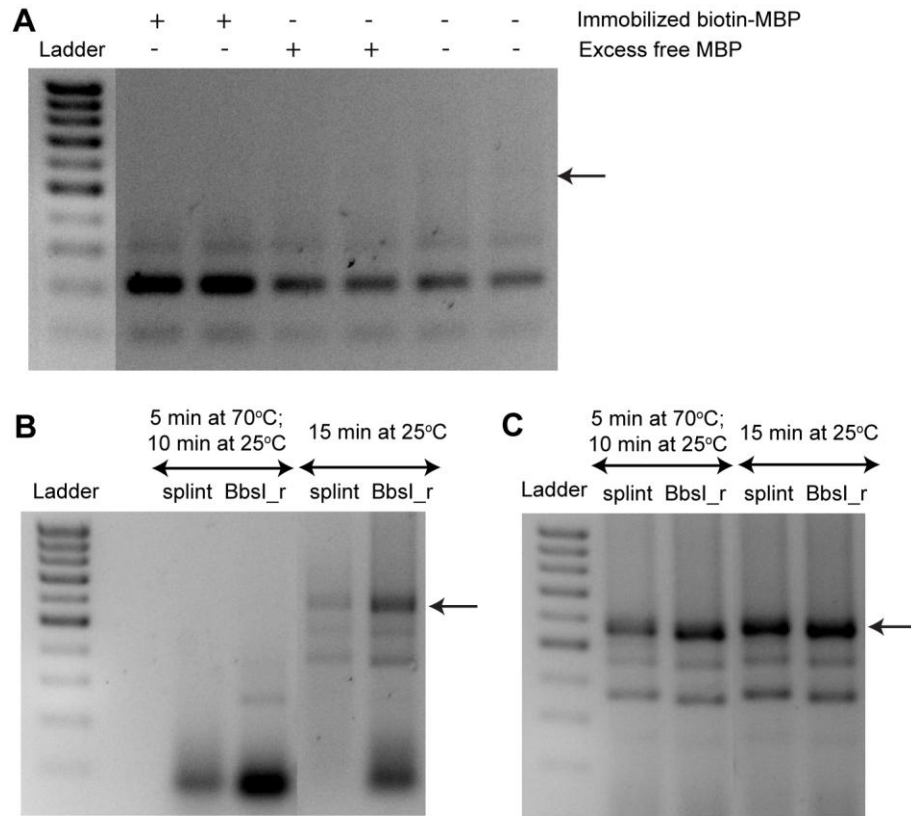
Figure 4.3. The effect of Tween 20 on Off7 RT-PCR products. (A) One mRNA display selection round with mRNA encoding Off7 (from pRDV-Off7) against immobilized MBP. TBST-cas was used as both the panning and washing buffers. RT and PCR were performed with reverse primer 18. The arrow indicates the size of the full-length product band (534 bp). (B, C) mRNA encoding Off7 (from pRDV2-Off7) was translated, diluted in TBS (B) with or (C) without 0.05% Tween 20, and purified. The products were reverse transcribed with GS_splint (splint; primer 7) or pRDV_BbsI_r (BbsI_r; primer 4). Products were either denatured at 70°C for 5 min, followed by a slow cool down at room temperature for 10 min, and then reverse transcribed as normal or they were incubated at room temperature for 15 min where the denaturation step was eliminated. exACTGene 1 kb plus DNA ladder (Fisher) was used. The arrow on each panel indicates the size of the full-length product band (556 bp).

Another variable in the protocol is the choice of blocking agent. Casein has been shown to be more effective than other commonly used blocking agents such as BSA and

gelatin[39], so it was chosen for the Her2 selection experiments described above. We have specifically observed that casein is able to prevent non-specific binding of smaller proteins better than BSA, likely due to the size heterogeneity of its component proteins, resulting in more effective blocking of wells on the microtiter plate (P. Barendt, personal communication). However, BSA blocking seemed to show a more uniform background signal in the display of mFc.67 (Fig. 4.4A), which might be due to secondary BSA recruitment from FcRn-BSA interactions[40]. These observations indicate that the choice of an optimal blocking agent will benefit from some experimental work as well.

mRNA recovery from *in situ* RT-PCR, as demonstrated in our streamlined protocol, can be altered to reflect selection needs. The *in situ* method simplifies the recovery process and avoids losses by bypassing the need for elution and purification of mRNA, since RT is performed directly on the display complexes[30]. However, *in situ* RT-PCR amplifies all mRNA present after selection, which can result in a higher signal-to-background ratio in some cases. Additionally, some selection experiments may incorporate specific elution protocols to recover only functional complexes (e.g., specific proteolysis, chemical cleavage, chelating agents, or pH change), which would prevent *in situ* RT-PCR. For example, the display of mFc.67 benefited from pH-based elution since the binding of Fc to FcRn is pH dependent. In the following section, we show that such a specific elution step is compatible with our streamlined mRNA display protocol.

4.3.4 Selection against FcRn with mFc.67 and its variants

mFc.67, mFc.67(LF) and mFc.67(YTE) was displayed in mRNA display format and selected against immobilized FcRn. All three proteins exhibited specific, pH-dependent binding to FcRn (Fig. 4.4A). The mRNA-puromycin-protein selection particles were panned against immobilized FcRn at pH 6.0, and subsequent washing at pH 6.0 was performed to remove unbound selection particles. Elution was performed by incubating the bound complexes for 1 minute in a buffer at pH 7.4, which should dramatically decrease the affinity of mFc to FcRn and thereby allow selective recovery of specifically bound selection particles. Strong specific signals can be seen at pH 6.0 when comparing all three mFc variants panned with or without immobilized FcRn (Fig. 4.4A), indicating that binding is indeed mediated by FcRn. When panning was done at pH 7.4, no specific signals were observed, further confirming that mFc binding to FcRn is pH dependent. Casein seems to lower the background signal to a greater extent than BSA, with the exception of mFc.67 incubated at pH 6.0 without FcRn and mFc.67 variants incubated at pH 7.4 without FcRn. Overall, mFc.67(LF) seemed to perform slightly better in the pH-dependent binding to FcRn, though a selection experiment with a ternary library of all three complexes could confirm this hypothesis.
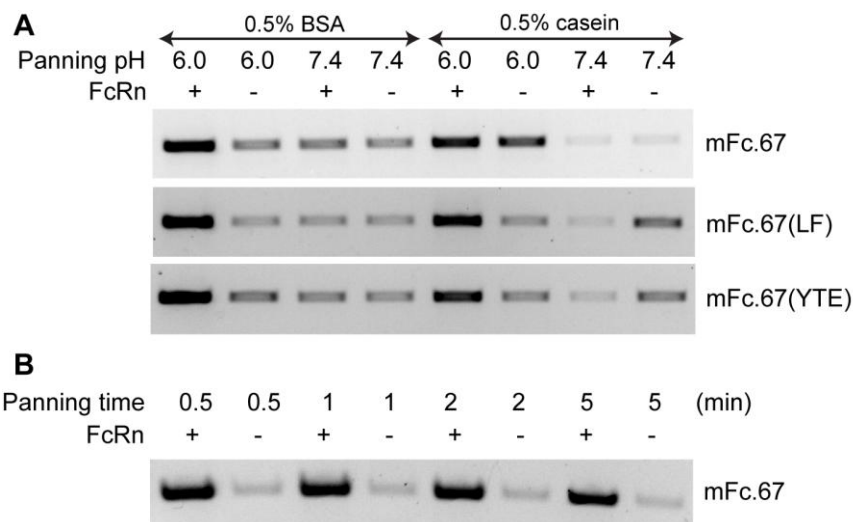
Figure 4.4. mRNA display with selection of mFc.67, mFc.67(LF), and mFc.67(YTE) against immobilized FcRn receptors. (A) Wells with (+) or without (-) immobilized FcRn were first blocked with 0.5% BSA or 0.5% casein, as specified. Monomeric Fc variants were panned against FcRn in TBS with 0.5% BSA/casein and 0.05% Tween 20 at pH 6.0 or pH 7.4 for 1 h at 4°C with shaking. Each well was subjected to three immediate washes and three 30-sec washes at 4°C, and then eluted in pH 7.4 TBS for 1 min. RT and 23 cycles of PCR were used to amplify the selected products. (B) mFc.67 was panned against FcRn in TBST-bsa-6 for 0.5, 1, 2, or 5 min at 4°C, as indicated. Each well was then subjected to three 2-min washes and three 10-min washes at 37°C. Products after RT and 25 cycles of PCR are shown.

We further altered the procedures to examine the impact of on-rate and off-rate conditions for mFc.67 selection, by using progressively shorter panning times at low temperature and longer washing times at high temperature, respectively. Several iterations were carried out, and the most stringent conditions are shown in Figure 4.4B. mFc.67 complexes were panned against FcRn at 4°C for 30 sec, 1 min, 2 min, and 5 min (from 1 hr previously), while the complexes were washed for 10 min at 37°C (from 30 sec at 4°C previously). These substantially more stringent selection conditions did not

seem to affect the signal-to-background ratio, suggesting that the affinity of mFc.67 at pH 6.0 might be higher than previously reported (K_D ~100 nM)[23].

4.4 Discussion

We have developed a simplified, streamlined protocol for mRNA display that reduces the time for a single selection round from ∼4−7 days to 2 days (Table 4.3). In addition to a number of simplifications in purification steps, the procedure postpones the RT step until after selection, which has the extra benefit of protecting displayed polypeptides from thermal denaturation. Wherever possible, complex or poorly regulated steps in the procedure have been replaced with simple, robust alternatives. Importantly, in applying these procedural simplifications, we have not compromised the achievable library size (∼10^{14}) or the enrichment capability of mRNA display. Furthermore, any of our streamlined steps can be readily reverted to the older procedures if there are additional constraints specific to a given experiment. Whether the entire protocol or only selected individual steps are adopted, this new protocol should enable faster and easier *in vitro* directed evolution experiments.

Table 4.3. Step-by-step comparison of two traditional mRNA display procedures and our proposed streamlined method.

Keefe procedure[5]	Cotten el al. procedure[31]	Proposed procedure
In vitro transcription (3 h)	*In vitro* transcription (7 h to overnight)	*In vitro* transcription (3 h)
Denaturing PAGE purification of mRNA (overnight + 3 h)	DNase digestion to remove cDNA from mRNA (3.5 h)	LiCl-based purification of mRNA [a,b] (2.5 h)

Phosphorylation of DNA-puromycin linker (2.5 h)	Conjugation of mRNA with DNA-puromycin linker (1 h)	Phosphorylation of DNA-puromycin linker (2.5 h)
Splinted ligation of DNA-puromycin linker and mRNA (1 h)		Splinted ligation of DNA-puromycin linker and mRNA (1 h)
Denaturaing PAGE purification of ligated mRNA (overnight + 3 h)	LiCl-based purification of ligated mRNA (4 h)	Ultrafiltration of ligated mRNA [c] (1.5 h)
In vitro translation with rabbit reticulocyte lysate (1.5 h)	*In vitro* translation with rabbit reticulocyte lysate (1 day)	*In vitro* translation with PURExpress system (1 h)
Oligo(dT) purification (1.5 h)	Oligo(dT) purification (3 h)	
Ni-NTA purification (3 h)	Reverse transcription (RT) (1 h)	
Reverse transcription (RT) (1 h)	Buffer exchange after RT (0.5 h)	
Buffer exchange after RT (0.5 h)	Anti-FLAG purification (4.5 h)	
Selection (varies)	Selection (varies)	Selection (varies)
PCR amplification (3 h)	PCR amplification (3 h)	PCR amplification [d] (3 h)
~ 4 days	**~ 7 days**	**~ 2 days**

[a] An RNA purification kit could be used to further shorten the time required for this step, but the LiCl precipitation procedure is significantly cheaper and also lends greater flexibility to the protocol by providing an additional potential stopping point that does not unnecessarily introduce an extra freeze-thaw cycle for the mRNA.

[b] Potential stopping point: storage of mRNA at -20°C at an intermediate step during LiCl-based purification.

[c] Potential stopping point: storage at -80°C after ultrafiltration of ligated mRNA.

[d] Potential stopping point: storage of PCR product at 4°C or -20°C after PCR amplification.

150

The dramatic, adverse effect of Tween 20 on RT-PCR of Off7 selected using mRNA display is surprising, since the experiment was carried out after purification of translation complexes and only trace amounts of Tween 20 must have been present. Tween 20 has been routinely used in RT-PCR in our lab and in many ribosome display protocols without showing any adverse effects, and this detergent has even been used in an RT buffer[41]. However, this effect is protein specific because mRNA display of mFc.67 yields no non-specific bands in the presence of Tween 20; in fact, the detergent actually enhanced the signal-to-background ratio for the display of mFc.67. Off7, due to its repeat nature, is prone to secondary structure artifacts during RT and, based on our findings, the presence of Tween 20 appears to promote or stabilize these artifacts, although it is still unknown how heat denaturation further exacerbates this effect. Interestingly, ribosome display of Off7 against MBP with Tween 20 as an additive shows no such artifacts, likely because the bound ribosomes prevent formation of extensive mRNA secondary structure, although the true molecular mechanisms cannot be established without further experiments.

With some alterations to the standard streamlined mRNA display protocol, including addition of Tween 20, usage of a different blocking agent, and specific elution, monomeric Fc variants were successfully displayed and exhibited specific, pH-dependent binding to FcRn receptors. It remains to be tested whether these mFcs bind to other Fc receptors, for example the FcγRs examined in Chapter 2. Though glycosylation is not necessary for Fc-FcRn interactions[25], the aglycosylated mFc form will likely need to acquire additional mutations before it will bind to FcγRs (Chapter 2). It will be

interesting to see if the point mutations introduced from stabilization of the monomeric and aglycosylated forms will have additive effects in enabling monomeric, aglycosylated mFc to bind to both FcRn and FcγRs. If so, the functions of mFc fragments can be extended beyond serum persistence and transcytosis to include antibody-dependent cellular phagocytosis and possibly other effector responses.

For this particular mFc variant (mFc.67), mRNA display may not be capable of improving binding properties without using unrealistically short panning times or long washing times. However, this new streamlined mRNA display strategy opens up many other potential selection schemes; for example, cell-based mRNA display can be used to select for mFc mutants with better transcytosis properties across cultured monolayers of intestinal epithelial cells. Nevertheless, the newly constructed LF and YTE mutants may already have therapeutic utility.

4.5 References

1.   Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4937–4942 (1997).

2.   Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297–12302 (1997).

3.   Lipovsek, D. & Plückthun, A. *In-vitro* protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67 (2004).

4.   Dreier, B. & Plückthun, A. Ribosome display: a technology for selecting and evolving proteins from large libraries. *Methods Mol. Biol.* **687**, 283–306 (2011).

5. Keefe, A. D. Protein selection using mRNA display. *Curr. Protoc. Mol. Biol.* **24**, 24.5.1–24.5.34 (2001).

6. Miyamoto-Sato, E. *et al.* A comprehensive resource of interacting protein regions for refining human transcription factor networks. *PloS ONE* **5**, e9289 (2010).

7. Kobayashi, T., Kakui, M., Shibui, T. & Kitano, Y. *In vitro* selection of a peptide inhibitor of human IL-6 using mRNA display. *Mol. Biotechnol.* **48**, 147–155 (2011).

8. Seelig, B. mRNA display for the selection and evolution of enzymes from *in vitro*-translated protein libraries. *Nat. Protoc.* **6**, 540–552 (2011).

9. Fukuda, I. *et al. In vitro* evolution of single-chain antibodies using mRNA display. *Nucleic Acids Res.* **34**, e127 (2006).

10. Liao, H.-I. *et al.* mRNA display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein. *J. Biol. Chem.* **284**, 17512–17520 (2009).

11. Roopenian, D. C. & Akilesh, S. FcRn: the neonatal Fc receptor comes of age. *Nat. Rev. Immunol.* **7**, 715–725 (2007).

12. Rodewald, R. pH-dependent binding of immunoglobulins to intestinal cells of the neonatal rat. *J. Cell Biol.* **71**, 666–670 (1976).

13. Simister, N. & Mostov, K. An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184–187 (1989).

14. Tesar, D. B. & Björkman, P. J. An intracellular traffic jam: Fc receptor-mediated transport of immunoglobulin G. *Curr. Opin. Struct. Biol.* **20**, 226–233 (2010).

15. Israel, E. J. *et al.* Expression of the neonatal Fc receptor, FcRn, on human intestinal epithelial cells. *Immunology* **92**, 69–74 (1997).

16. Dickinson, B. L. *et al.* Bidirectional FcRn-dependent IgG transport in a polarized human intestinal epithelial cell line. *J. Clin. Invest.* **104**, 903–911 (1999).

17. Vaughn, D. E. *et al.* Identification of critical IgG binding epitopes on the neonatal Fc receptor. *J. Mol. Biol.* **274**, 597–607 (1997).

18. Ghetie, V. *et al*. Increasing the serum persistence of an IgG fragment by random mutagenesis. *Nat. Biotechnol.* **15**, 637–640 (1997).

19. Dall'Acqua, W. F. *et al.* Increasing the affinity of a human IgG1 for the neonatal Fc receptor: biological consequences. *J. Immunol.* **169**, 5171–5180 (2002).

20. Kamei, D. T. *et al.* Quantitative methods for developing Fc mutants with extended half-lives. *Biotechnol. Bioeng.* **92**, 748–760 (2005).

21. Vaccaro, C., Zhou, J., Ober, R. J. & Ward, E. S. Engineering the Fc region of immunoglobulin G to modulate *in vivo* antibody levels. *Nat. Biotechnol.* **23**, 1283–1288 (2005).

22. Mezo, A. R. *et al.* Reduction of IgG in nonhuman primates by a peptide antagonist of the neonatal Fc receptor FcRn. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2337–2342 (2008).

23. Ying, T., Chen, W., Gong, R., Feng, Y. & Dimitrov, D. S. Soluble monomeric IgG1 Fc. *J. Biol. Chem.* **287**, 19399–19408 (2012).

24. Dall'Acqua, W. F., Kiener, P. A. & Wu, H. Properties of human IgG1s engineered for enhanced binding to the neonatal Fc receptor (FcRn). *J. Biol. Chem.* **281**, 23514–23524 (2006).

25. Jefferis, R. Glycosylation of recombinant antibody therapeutics. *Biotechnol. Prog.* **21**, 11–16 (2005).

26. Zahnd, C. *et al.* A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *J. Mol. Biol.* **369**, 1015–1028 (2007).

27. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).

28. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004).

29. Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: selecting and evolving proteins *in vitro* that specifically bind to a target. *Nat. Methods* **4**, 269–279 (2007).

30. He, M. & Taussig, M. Eukaryotic ribosome display with in situ DNA recovery. *Nat. Methods* **4**, 281–289 (2007).

31. Cotten, S. W., Zou, J., Valencia, C. A. & Liu, R. Selection of proteins with desired properties from natural proteome libraries using mRNA display. *Nat. Protoc.* **6**, 1163–1182 (2011).

32. Ueda, T., Kanamori, T. & Ohashi, H. Ribosome display with the PURE technology. *Methods Mol. Biol.* **607**, 219–225 (2010).

33. Ivanov, I. G. *et al.* Efficiency of the 5'-terminal sequence (Ω) of tobacco mosaic virus RNA for the initiation of eukaryotic gene translation in *Escherichia coli*. *Eur. J. Biochem.* **209**, 151–156 (1992).

34. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).

35. Kondo, T., Mukai, M. & Kondo, Y. Rapid isolation of plasmid DNA by LiCl-ethidium bromide treatment and gel filtration. *Anal. Biochem.* **198**, 30–35 (1991).

36. Cathala, G. *et al.* A method for isolation of intact, translationally active ribonucleic acid. *DNA* **2**, 329–335 (1983).

37. Steinitz, M. Quantitation of the blocking effect of tween 20 and bovine serum albumin in ELISA microwells. *Anal. Biochem.* **282**, 232–238 (2000).

38. Baggio, R. *et al.* Identification of epitope-like consensus motifs using mRNA display. *J. Mol. Recognit.* **15**, 126–134 (2002).

39. Vogt, R. F., Phillips, D. L., Henderson, L. O., Whitfield, W. & Spierto, F. W. Quantitative differences among various proteins as blocking agents for ELISA microtiter plates. *J. Immunol. Methods* **101**, 43–50 (1987).

40. Chaudhury, C., Brooks, C. L., Carter, D. C., Robinson, J. M. & Anderson, C. L. Albumin binding to FcRn: distinct from the FcRn-IgG interaction. *Biochemistry* **45**, 4983–4990 (2006).

41. Jokela, P., Joki-Korpela, P., Maaronen, M., Glumoff, V. & Hyypiä, T. Detection of human picornaviruses by multiplex reverse transcription-PCR and liquid hybridization. *J. Clin. Microbiol.* **43**, 1239–1245 (2005).

CHAPTER 5

Ribosome display of scaffold proteins and agonists

5.1 Introduction

Ribosome display is another powerful *in vitro* directed evolution technique[1]. It shares many characteristics with mRNA display[2], which we described in detail in the previous chapter. Like mRNA display, ribosome display is performed entirely *in vitro*, which enables the display of large libraries (up to 10^{14} members) and allows for easy additional mutagenesis[3]. Unlike mRNA display, ribosome display does not covalently link the mRNA and displayed protein; instead, the mRNA and its corresponding *in vitro* translated protein are noncovalently associated *via* a stalled ribosome[4]. A side-by-side graphical comparison of the two display methods is shown in Figure 5.1. First, the transcription/translation cassette is constructed by attaching the T7 promoter and ribosome binding site to the 5' end of the DNA library and an unstructured spacer to the 3' end. After PCR amplification, *in vitro* transcription yields mRNAs which contain 5' and 3' stem loops to confer additional stability[4]. The mRNA library, which lacks stop codons, is then translated, resulting in nascent polypeptides that are not released from the ribosomes. The last amino acid of each translated protein is thus connected to peptidyl-tRNA within the ribosome, the unstructured spacer fills the ribosomal tunnel, and the protein of interest folds outside of the ribosome[4]. Without release of the polypeptide

chain and dissociation of the ribosome, the mRNA also remains attached. This ternary mRNA-ribosome-protein complex is then used for selection[3]. The mRNAs encoding the selected proteins are recovered by reverse transcription (RT) and amplified by PCR[4]. The DNA can then be used in the next round of selection or it can be sequenced and analyzed.
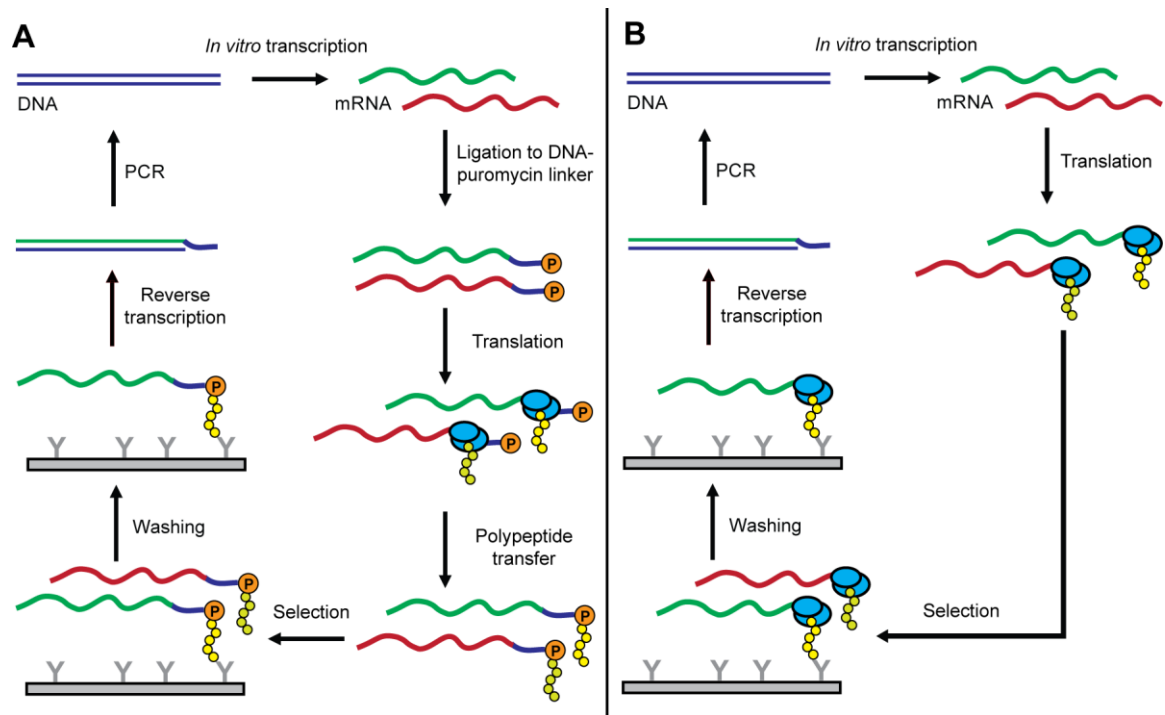
Figure 5.1. Comparison of ribosome display and mRNA display. (A) The mRNA display protocol, as described in Chapter 4. The selection particle is an mRNA-protein complex that is covalently connected by a puromycin linker. (B) In ribosome display, a DNA library is transcribed into an mRNA library by *in vitro* transcription. The mRNA library, which lacks stop codons, is translated and the nascent polypeptides remain attached since the ribosomes stall. The ternary mRNA-ribosome-protein complexes are used for selection. After washing to remove unwanted complexes, the mRNAs encoding the selected complexes are reversed transcribed into cDNAs and further amplified by PCR. This DNA can be used in the next round of selection or it can be sequenced and analyzed.

The ribosome display method is slightly simpler to perform than even our streamlined mRNA display protocol (Chapter 4) because no ligation to the DNA-puromycin linker is needed before translation. However, a notable drawback is that the mRNA-ribosome-protein ternary complex is inherently less stable, even though stabilization strategies (e.g., the addition of Mg^{2+}) have enabled complexes to be maintained intact for more than 10 days[5]. Ribosome display has been used in many protein engineering applications, including the development of peptides[6], enzymes[7], single chain variable fragments (scFvs)[8], and novel binders based on alternative scaffolds[9]. In this work, we examined some of the challenges of ribosome display and developed a robust display protocol for a variety of interesting scaffolds and therapeutic proteins.

We first investigated a new display strategy involving an emerging alternative protein scaffold, the designed ankyrin repeat protein (DARPin)[10]. Ankyrin repeat proteins comprise a prominent repeat protein family that occur in many species and naturally mediate protein-protein interactions in the cytoplasm, nucleus, membrane and extracellular space[11]. They usually consist of 4–6 stacks of 33 amino acids repeats, forming an elongated protein domain with a continuous hydrophobic core[12]. The diversity of functions exhibited by these repeat proteins have led to the hypothesis that such proteins can be engineered to generate novel binding specificities[13]. Using sequence and structural consensus analysis, libraries of ankyrin repeat proteins have been generated by fixing consensus residues and randomizing the protruding β-turn and the following α-helix within each internal repeat, and this hydrophobic core is sealed with N- and C-

158

terminal capping repeats (N-cap and C-cap)[13]. These scaffolds exhibit favorable biophysical properties, since they are soluble, very stable, and well-expressed[14]. DARPin libraries designated N2C and N3C (consisting of an N-cap, two or three internal repeats, respectively, and a C-cap) have been generated and used in ribosome display selections to enrich for binders against maltose binding protein (MBP). Through these selections, a high-affinity binder, Off7 (K_D of 4.4 nM)[9], was recovered. We used this high-affinity interaction between MBP and Off7 to test and develop a robust ribosome display protocol that incorporates advances from our own work and various other groups[15,16] into the standard procedure[17]. Furthermore, to examine the minimal binding module as well as the interaction between the internal repeats and the N- and C-caps, we progressively truncated Off7 and then assessed MBP binding using ribosome display. This reverse engineering approach can help elucidate the molecular determinants of stable folding and high affinity binding in DARPins.

We then used the same ribosome display strategy to display two important agonists: insulin, which, for decades, has been indispensible in the treatment of diabetes, and leptin, which has shown promise in treating obesity[18], lipodystrophy[19], hypothalamic amenorrhea[20], and diabetes[21]. Human insulin normally consists of separate A and B chains connected by three disulfide bonds, two inter-chain and one intra-chain[22]. However, since genotype and phenotype must be coupled for applications involving ribosome display, a single-chain insulin analog was used for these studies. Of the characterized single-chain analogs[23], SCI-57[24] was chosen because it retains the folding properties and the biological activity of wild-type human insulin, with additional

advantages of enhanced thermodynamic stability and reduced aggregation[24]. On the other hand, human leptin, which is naturally a single-chain protein, could be expressed in ribosome display without modifications. Since both proteins contain disulfide bonds, *in vitro* translation of these proteins using a reconstituted translation system from *E. coli* (PURExpress) could be problematic in achieving correct folding. Since previous studies have shown that the addition of chaperones or disulfide isomerases can help alleviate this problem[1,15,25], we examined the specific binding of SCI-57 and leptin to their cognate receptors using our ribosome display protocol with or without such supplements. We further investigated the effect of removing disulfide bonds in leptin, since this could increase production yield in lower organisms and simplify downstream purification. Our results provide insights into the stabilities of SCI-57 and leptin. These two therapeutic proteins will be discussed in more detail in the next chapter, which focuses on their secretion from *Lactococcus lactis*.

Next, we use a forward engineering approach to increase the stability of an scFv. Many scFvs have been expressed and evolved using ribosome display[1,8,26]; however, the success of such experiments can depend strongly on the stability of the protein displayed. Here, we studied two scFvs of interest: R6.5[27,28] and YN1 (Colin Greineder, unpublished results), which bind to human and mouse intercellular adhesion molecule-1 (ICAM-1), respectively. ICAM-1 is a transmembrane glycoprotein that is constitutively present on endothelial cells. It is upregulated by proinflammatory cytokines and participates in the firm arrest and transmigration of leukocytes from blood vessels to tissues[29]. Anti-ICAM-1 antibodies have great potential in various anti-inflammatory treatments; however,

clinical trials of the R6.5 antibody for treatment of refractory rheumatoid arthritis[30], renal transplant[31], and acute ischemic stroke[32] showed side effects and limited efficacy. The fusion of an anti-ICAM-1 scFv with the extracellular domain of thrombomodulin, an endothelial protective membrane protein, may allow for endothelial targeting and show better clinical outcomes. In fact, using this strategy, promising results have been seen in the treatment of acute thrombosis, inflammation, and lung injury (C. Greineder, personal communication). In collaboration with Dr. Vladimir Muzykantov and Dr. Colin Greineder (University of Pennsylvania), we thus sought to engineer a more stable anti-ICAM-1 scFv which would maintain binding to ICAM-1. Inspired by the established method of complementarity determining region (CDR) grafting onto closely related human antibody frameworks to reduce immunogenicity, we grafted the CDRs of R6.5 and YN1 onto more stable human frameworks in order to improve antibody stability and folding efficiency[33,34]. The human combinatorial antibody library (HuCAL) consensus frameworks[35] were used since scFvs with HuCAL frameworks have been shown to be solubly expressed in *E. coli*[35] and have been successfully used to select for novel binders with ribosome display[8], which is the platform of choice for further evolution of these stability-engineered scFvs. In addition, by improving stability through humanization, these antibodies and their progeny may proceed clinically with less additional modifications.

5.2 Materials and methods

5.2.1 Plasmid construction

For all DNA manipulation techniques and materials, please refer to Sections 4.2.1 and 4.2.2. Plasmids used in this work are listed in Table 5.1 and primers are listed in Table 5.2. To clone the off7 gene and subsets of its repeats into pRDV, various fragments from the off7 gene were PCR amplified from pRDV:Off7[9] (courtesy of Dr. Andreas Plückthun, University of Zurich). pRDV:Off7 was PCR amplified with primers 1/6, 1/5, 1/4, 3/6, 2/6, and 2/5 to obtain full off7 without the FLAG tag or spacer, off7 fragment comprising the N-cap and the 3 internal repeats, off7 fragment comprising the N-cap and 2 internal repeats, off7 fragment comprising 2 internal repeats and the C-cap, off7 fragment comprising the 3 internal repeats and the C-cap, and off7 fragment with only the 3 internal repeats, respectively. These PCR products were gel purified, digested with NcoI and EcoRI, and ligated into similarly digested pRDV, yielding pRDV:Off7-FS, pRDV:Off7NI3, pRDV:Off7NI2, pRDV:Off7I2C, pRDV:Off7I3C, and pRDV:Off7I3, respectively.

Table 5.1. Plasmids used in this study.

Plasmids	Characteristics	Sources
pRDV:Off7	pRDV containing the FLAG-tagged off7 gene; Amp[r]	Binz et al.[9]
pRDV:Off7-FS	pRDV containing the off7 gene without FLAG tag or extra spacer; Amp[r]	This work
pRDV:Off7NI3	pRDV containing the N-cap and all 3 internal repeats of the off7 gene; Amp[r]	This work
pRDV:Off7NI2	pRDV containing the N-cap and first 2 internal repeats of the off7 gene; Amp[r]	This work
pRDV:Off7I2C	pRDV containing the last 2 internal repeats and the C-cap of the off7 gene; Amp[r]	This work

pRDV:Off7I3C	pRDV containing all 3 internal repeats and the C-cap of the off7 gene; Amp[r]	This work
pRDV:Off7I3	pRDV containing all 3 internal repeats of the off7 gene; Amp[r]	This work
pRDV:SCI-57	pRDV containing the SCI-57 gene; Amp[r]	This work
pRDV:hLep	pRDV containing the human leptin gene; Amp[r]	This work
pRDV:hLepC96SC146S	pRDV containing the human leptin gene with C96S and C146S double mutations; Amp[r]	This work
pCR4:R6.5	pCR4 containing the R6.5 gene; Amp[r]	Courtesy of Dr. Vladimir Muzykantov and Dr. Colin Greineder
pCR4:YN1	pCR4 containing the YN1 gene; Amp[r]	Courtesy of Dr. Vladimir Muzykantov and Dr. Colin Greineder
pRDV2:R6.5	pRDV2 containing the R6.5 gene; Amp[r]	This work
pRDV2:YN1	pRDV2 containing the YN1 gene; Amp[r]	This work
pRDV2:hR6.5	pRDV2 containing the R6.5 gene grafted onto a human $V_{H3}/V_{\kappa2}$ framework; Amp[r]	This work
pRDV2:hYN1	pRDV2 containing the YN1 gene grafted onto a human $V_{H3}/V_{\kappa4}$ framework; Amp[r]	This work

Table 5.2. Primers used in this study.

Number	Name	Nucleotide Sequence (5' → 3') [a]
1	pRDV_Off7N_NcoI_f	GTCGATCCCCATGGGATCCGACCTGGGTAG
2	pRDV_Off7I1_NcoI_f	GGTCATGACCATGGCTGACGTTAATGCTGCTGACAATACT
3	pRDV_Off7I2_NcoI_f	ATTGACGCCCATGGCTGACGTTGACGCTTCTGACGTTTTT
4	pRDV_Off7I2_EcoRI_r	TACTGGATGAATTCACCGTTCTTCAGCAGAACTTCAACGATTTCCAGGTGACCCC
5	pRDV_Off7I3_EcoRI_r	ACGTACCTGAATTCACCGTGCTTCAGCAGAACTTCAACGATTTCCAGGTA
6	pRDV_Off7C_EcoRI_r	TCGGATATGAATTCATTAAGCTTTTGCAGGATTTC
7	SCI-57gene_f	ATATATCCATGGGCTTCGTTAACCAGCACCTGTGCGGTTCTGACCTGGTTGAAGCTCTGTACCTGGTTTGCGGTGAACGTGGTTTCTTCTACACCGACCCGACCGGTGGTGGTCCGCGTCGTGGTATCGTTGAACAGTGCTGCCACTCTATCTGCTCTCTGTACCAGCTGGAAAACTACTGCAACGAATTCGGATCTGGT

163

8	pRDV_NcoI_f	AGAAGGAGATATATCCATGG
9	pRDV_EcoRI_r	TGGCCACCAGATCCGAATTC
10	hLep_NcoI_f	TTGTCGACCCATGGTGCCCATCCAAAAAGTC
11	hLep_EcoRI_r	ATTTCGTTGAATTCGCACCCAGGGCTGAG
12	hLep_C96S_f	TCTAAGAGCAGCCACTTGC
13	hLep_C96S_r	GCAAGTGGCTGCTCTTAGA
14	hLep_C146S_r	ATTTCGTTGAATTCGCTCCCAGGGCTGAGGTC
15	pRDV_BbsI_f	AATAATTTGAAGACACCGAAAGAAG
16	pRDV_BbsI_r	GGCCACCGGAAGACCCAAGC
17	T7B_noBsaI	ATACGAAATTAATACGACTCACTATAGGGACACCACAACGG
18	tolAk	CCGCACACCAGTAAGGTGTGCGGTTTCAGTTGCCGCTTTCTTTCT

The gene encoding the single-chain insulin analog SCI-57 was constructed from primers 7/9 by extension PCR. Primers 8/9 were then used to amplify this SCI-57 template and the resulting product was cut with NcoI and EcoRI for ligation into the similarly cut pRDV, giving rise to pRDV:SCI-57. To construct pRDV:hLep, human leptin cDNA (GenBank accession no. BC069452; Open Biosystems, Waltham, MA) was PCR amplified using primers 10/11, digested with NcoI and EcoRI, and ligated into similarly cut pRDV. pRDV:hLepC96SC146S was constructed by overlap extension PCR of two pieces: the fragment from amplifying pRDV:hLep with primers 10/13 and the fragment from amplifying pRDV:hLep with primers 12/14. The resulting product was digested with NcoI and EcoRI and ligated into similarly cut pRDV.

pCR4:R6.5 and pCR4:YN1, which contained the original anti-ICAM-1 scFvs, were provided by Dr. Vladimir Muzykantov and Dr. Colin Greineder (University of

Pennsylvania). R6.5 and YN1 constructs were digested from the two pCR4 plasmids with BbsI and ligated into BbsI-digested pRDV2 (Chapter 3). For pRDV2:hR6.5 and pRDV2:hYN1, the grafted hR6.5 and hYN1 genes were synthesized in two mini-gene fragments for each construct (Integrated DNA Technologies, Coralville, IA). Overlap extension PCR was used to assemble the full-length constructs and amplification was performed with primers 15/16. The resulting products were digested with BbsI and ligated into similarly digested pRDV2.

5.2.2 CDR grafting onto stable frameworks

The background and approach for grafting CDRs onto stable antibody frameworks has been described[33]. The V_H and V_L sequences of R6.5 and YN1 were aligned with the consensus V_H and V_L domains derived from the Human Combinatorial Antibody Library (HuCAL®)[33,35,36]. Based on stability and similarity in framework residues, both R6.5 and YN1 V_H CDRs were grafted onto the human consensus V_{H3} domain; R6.5 and YN1 V_L CDRs were grafted onto the human consensus Vκ2 and Vκ4 domains, respectively. Based on Ewert et al.[33], color coding in Figure 5.2 highlights positions important for antigen contacts, V_L/V_H heterodimer contacts (yellow to red indicates less important to more important), and upper core residues (purple) with conserved central core residues (blue) that affect CDR conformation[33]. All colored positions contain residues from the original R6.5 and YN1 sequences, and all uncolored positions contain residues from the human consensus frameworks. For easy visualization, the actual amino acids are also color-coded based on residue type. The original HuCAL frameworks, unmodified R6.5

165

and YN1, and the grafted hR6.5 and hYN1 are shown in separate lines in Figure 5.2.
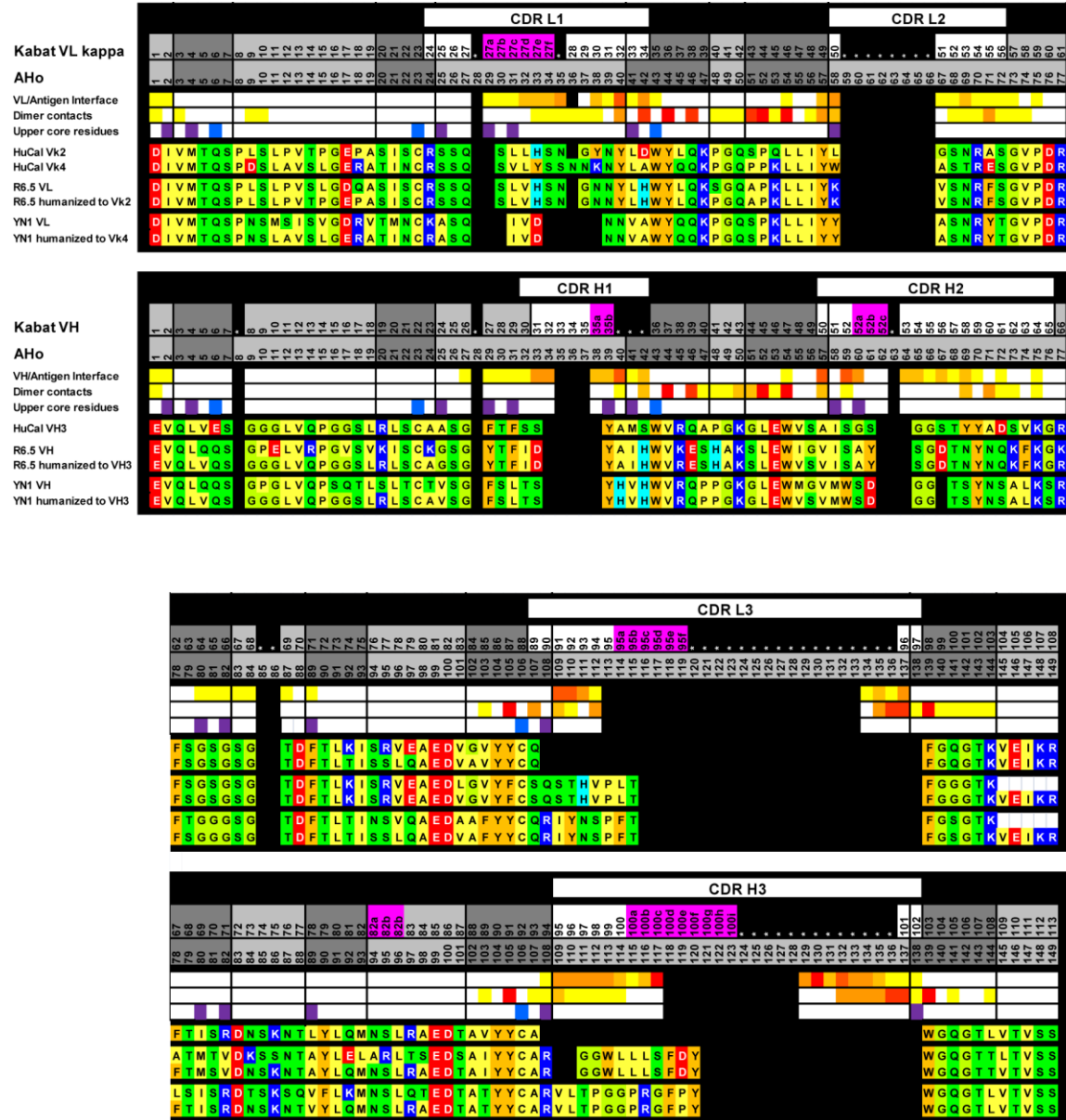
Figure 5.2. CDR grafting information for hR6.5 and hYN1. The color scheme, residue numbering system, and worksheet were adopted from Ewert et al.[33] Antigen contacts and V_L/V_H heterodimer interfaces were color-coded from yellow to red to reflect the average contribution of residue positions to the contact interface (white: 0%, yellow: 0–20%, light orange: 20–40%, orange: 40–60%, orange red: 60–80%, red: > 80%). Upper core residue positions are purple and central core residues are blue. R6.5 and YN1 residues before and after grafting are shown together with the human framework onto which they are grafted. The

actual amino acid residues are also color coded: orange: F, Y, W; yellow: I, L, V, P, A, M, C; light green: G; Green: S, T, N, Q; light blue: H; dark blue: R, K, red: D, E.

5.2.3 Reagents for ribosome display

For materials common to both ribosome and mRNA display, please refer to Section 4.2 for manufacturer's information. Insulin receptors were manufactured by R&D Systems (Minneapolis, MN) and leptin receptors (binding domain) were manufactured by Prospec (#CYT-508; Rehovot, Israel). Human insulin solution and leptin were from Sigma-Aldrich (#I9278; St. Louise, MO) and Peprotech (#300-27; Rocky Hill, NJ), respectively. Biotinylated human and mouse ICAM-1 were provided by our collaborator, Dr. Colin Greineder.

Non-biotinylated and biotinylated MBP were produced in *E. coli* using expression plasmids pAT224[9] (AY327139) and pAT224/pBirA[37], respectively. To make non-biotinylated MBP, XL1-Blue cells were transformed with pAT224, inoculated into 5 ml Luria Broth (LB) with 100 µg/ml ampicillin (Amp), and grown overnight at 37°C with shaking. Overnight cultures were diluted 1:25 into 100 ml pre-warmed LB with Amp and grown to an OD_{600} of 0.6. Expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was grown for an additional 4–5 h at 37°C with shaking. Cells were harvested by centrifugation at 4000g for 20 min and frozen at -20°C overnight. Non-biotinylated MBP was then purified via its His_6 tag with the QIAexpress Ni-NTA Fast Start Kit (Qiagen) and analyzed by SDS-PAGE. Biotinylated MBP was produced as described but with modifications[37]. XL1-Blue cells were

transformed with both pAT224 and pBirA, inoculated into 5 ml TYH media (20 g tryptone, 10 g yeast extract, 11 g HEPES, 5 g NaCl, 1 g $MgSO_4$, adjusted to pH 7.2–7.4 with KOH) with 75 μg/ml Amp and 10 μg/ml chloramphenicol (Cm), and grown overnight at 37^oC with shaking. Overnight cultures were diluted 1:100 into 250 ml of prewarmed TYH media with 75 μg/ml Amp and 0.5% glucose and grown at 37^oC with shaking to an OD_{600} of 0.6–0.7. Expression of biotinylated MBP was induced by addition of both 0.2 mM biotin solution (prepared by the addition of 6.1 mg d-biotin to warm 10 mM bicine buffer at pH 8.3 and sterile filtered through a 0.22-μm filter) and 1.5 mM IPTG, and the culture was grown for an additional 3 h. Cells were harvested by centrifugation at 5000g for 20 min and frozen at -80^oC overnight. Purification and analysis of biotinylated MBP was performed as described above.

5.2.4 Ribosome display

Primers 17 and 18 were first used to amplify the gene constructs within the pRDV and pRDV2 plasmids, and the resulting products were used directly for *in vitro* transcription as described previously[17]. The mRNA was purified as described[17] and reconstituted in 30 μl RNase-free water. *In vitro* translation was carried out with the PURExpress protein synthesis kit (New England Biolabs; NEB) in 25 μl according to manufacturer's instructions. The PURExpress master mix was supplemented with 0.5 μl RNasin Plus, and when appropriate, 0.5 μM protein disulfide isomerase (PDI) or 1 μl disulfide bond enhancer 1 and 1 μl disulfide bond enhancer 2 (NEB). The mRNA was added to the master mix in a 1:1 ratio with ribosomes. Translation was performed at 37^oC for 30–40

min, and immediately put on ice. Ice-cold washing buffer (100 µl) with Tween 20 and either casein or bovine serum albumin (BSA) [50 mM Tris-HCl (pH 7.5 at 4°C), 150 mM NaCl, 50 mM MgAc, 0.05% Tween 20 with either 0.5% casein or 0.5% BSA; WBT-cas/WBT-bsa, respectively] was added to the stopped translation for each well and kept on ice.

To prepare for selection, wells on a NUNC MaxiSorp plate (Thermo Fisher Scientific, Rochester, NY) were coated either with neutravidin (for biotinylated targets) or with the non-biotinylated targets directly at 4°C for at least 16 h without shaking. For Off7-MBP and anti-ICAM-1 experiments, 100 µl of 66 nM neutravidin (Pierce, Rockford, IL) was allowed to adsorb overnight; the neutravidin-coated wells were then washed three times with 300 µl TBS [50 mM Tris-HCl (pH 7.4) at 4°C, 150 mM NaCl], blocked with 300 µl TBS-cas or TBS-bsa [TBS with 0.5% casein or 0.5% BSA] at room temperature for 1 h with shaking, and washed again three times with 300 µl TBS. For positive wells, 2×10^{13} to 3×10^{13} molecules of biotinylated MBP or ICAM-1 in 200 µl TBS-cas or TBS-bsa were added to neutravidin-coated and casein/BSA-blocked wells; for negative wells, only buffer (200 µl TBS-cas or TBS-bsa) was added. Wells were incubated at 4°C for 1 hr with shaking. For insulin and leptin experiments, each well was directly coated with 2×10^{12} insulin receptors or 6×10^{12} leptin receptors at 4°C for at least 16 h without shaking. The wells were washed 3 times with 300 µl TBS and blocked with 300 µl TBS-cas at room temperature for 1 h with shaking.

All antigen-coated and casein/BSA-blocked wells were then washed 3 times with TBS and once with WBT-cas or WBT-bsa. The translation reaction in WBT-cas or WBT-

bsa was added to each well and panned for 1 h at 4°C with shaking. The wells were then subjected to various washes with WBT-cas or WBT-bsa (or WBT-cas with 0.25 M KCl for leptin binding experiments with KCl) prior to reverse transcription: 10 immediate washes for Off7-MBP; 3 immediate and 6 15-min washes for SCI-57-insulin receptor; 3 immediate and 5 5-min washes for leptin-leptin receptor; 8 immediate washes for leptin (C96SC146S)-leptin receptor; 3 immediate washes and 3 2-min washes for R6.5-human ICAM-1; and 7 immediate washes for hYN1-mouse ICAM-1.

Reverse transcription (RT) was performed *in situ* as described in Section 4.2.8 for all experiments except for SCI-57-insulin receptor. For SCI-57-insulin receptor selection, elution was performed by adding 200 µl elution buffer (TBS with 50 mM EDTA) with 50 µg/ml *S. cerevisiae* RNA and incubated at 4°C for 10 min with shaking. The eluted mRNA was then purified using the Roche High Pure RNA isolation kit and reverse transcribed as described in Section 4.2.8. Reverse primer 9 was used for all experiments with pRDV and reverse primer 16 was used for all experiments with pRDV2 (except for hYN1 experiments, in which reverse primer 18 was used).

The products from RT were amplified by PCR with inner primers 8/9 for experiments with pRDV and primers 15/16 for experiments with pRDV2. The PCR products were subjected to analytical gel electrophoresis with ethidium bromide and visualized using an ultraviolet transilluminator and CCD camera.

5.3 Results

5.3.1 The ribosome display protocol

Through the display of various molecules, many of which are illustrated in this chapter, we have improved the ribosome display protocol. Chronologically, the display of SCI-57 was examined first, followed by the Off7 fragments, human leptin, leptin variants, and finally R6.5 and YN1 scFvs. Changes were made to the standard protocol over time, most noticeably the switches from pRDV to pRDV2 and from mRNA elution to *in situ* reverse transcription. Selection targets were immobilized directly to the plastic surfaces of the wells when biotinylated targets were not available. Insulin and leptin receptors were adsorbed directly on the Maxisorp plate overnight at 4°C, and successfully exhibited specific binding to SCI-57 and leptin with ribosome display, respectively. However, ribosome display experiments with directly adsorbed nonbiotinylated MBP and ICAM-1 showed a worse signal-to-background ratio than with biotinylated MBP and ICAM-1 immobilized onto neutravidin-coated wells, suggesting that this latter method is preferred, if feasible[4,17]. The advantages of immobilizing biotinylated targets include robustness to stringent washing and ability to maintain the structure of the target[4]. Our current ribosome display protocol now incorporates modifications based on our own findings as well as advances from other laboratories, and therefore differs from the protocol established by the Plückthun laboratory[4,17] in a few main areas.

First, our display vector is pRDV2, which is pRDV[9] redesigned with restriction sites for the Type IIS enzyme BbsI (Chapter 3). Using pRDV2, desired ligation products can be increased up to 15- and 2.6-fold following circular and linear ligations,

respectively[38]. Maximizing the amount of desired ligation product is essential in optimizing DNA library construction since library diversity is directly affected by ligation efficiency. The pair of inner primers used during a round of ribosome display is now replaced by pRDV_BbsI_f (primer 15) and pRDV_BbsI_r (primer 16), which contain the BbsI recognition sites for use with pRDV2.

For *in vitro* translation of mRNA, crude extract from *E. coli* has been replaced with the PURExpress protein translation system (NEB), which is a minimal *E. coli*-based translation system reconstituted from purified components[15,25] (PURExpress was also used in our streamlined mRNA display procedure in Chapter 4). This cell-free protein synthesis system contains minimal nucleases and other inhibitory factors and allows for greater control of the translation reaction. A variation of the PURExpress system without release factors is also commercially available (NEB) and this can aid stalling of the ribosomes at the last codon and further stabilize the mRNA-ribosome-protein ternary complexes[15]. Molecular chaperones or isomerases can also be included in defined quantities to facilitate protein folding, and reaction conditions can be adjusted to allow for oxidative folding[25]. Some of these variations were tested and the results will be presented below.

An alternative to EDTA elution and purification of mRNA in the original protocol is *in situ* reverse transcription. As discussed in Section 4.3.3, this method involves performing RT directly on the display complexes in the wells, thus simplifying the recovery process and avoiding losses during elution and purification[16]. All experiments except the SCI-57-insulin receptor selection were carried out with *in situ* RT.

5.3.2 Binding of Off7 fragments to MBP

Wild-type Off7 has an N3C structure, with 3 internal repeats sandwiched between an N-cap and a C-cap. It has previously been evolved to low nanomolar affinity to MBP by ribosome display[9]; in the same work, a high-affinity binder with N2C structure was also obtained[9]. While both N3C and N2C ankyrin repeat proteins are extremely stable[14], we investigated the effect that removing repeats from Off7 would have on its ability to be displayed and bind MBP. It was hypothesized that the capping repeats, which seal the hydrophobic core[13], would be necessary to form a stable and well-folded ankyrin repeat protein capable of binding MBP.

Various fragments of Off7 were created: Off7-FS (a simplified form of Off7 with its FLAG tag and spacer removed), Off7NI3 (off7-FS with the C-cap removed), Off7NI2 (Off7-FS with the C-cap and the last internal repeat removed), Off7I2C (Off7-FS with the N-cap and the first internal repeat removed), Off7I3C (Off7-FS with N-cap removed), and Off7I3 (Off7-FS with N-cap and C-cap removed). Ribosome display of these Off7 fragments against biotinylated MBP revealed specific binding of Off7, Off7-FS, Off7NI3, and Off7I3C to immobilized MBP, but not to neutravidin-coated, MBP-free wells (Fig. 5.3). To further demonstrate specific binding, samples in which excess non-biotinylated MBP was added to during the panning step to compete for Off7 binding showed markedly lower levels of recovered complexes. By contrast, Off7NI2, Off7I2C, and Off7I3 showed increased signals in MBP wells compared to neutravidin-coated wells, but the signal intensities remained unchanged when excess MBP was added as a

competitor (Fig. 5.3). Our results indicate that at least 4 contiguous repeats of Off7 are required for binding MBP in ribosome display format. The lack of specific binding for Off7NI2, Off7I2C, and Off7I3 is likely due to disruptions to the binding interface or instability of the truncated protein structures.
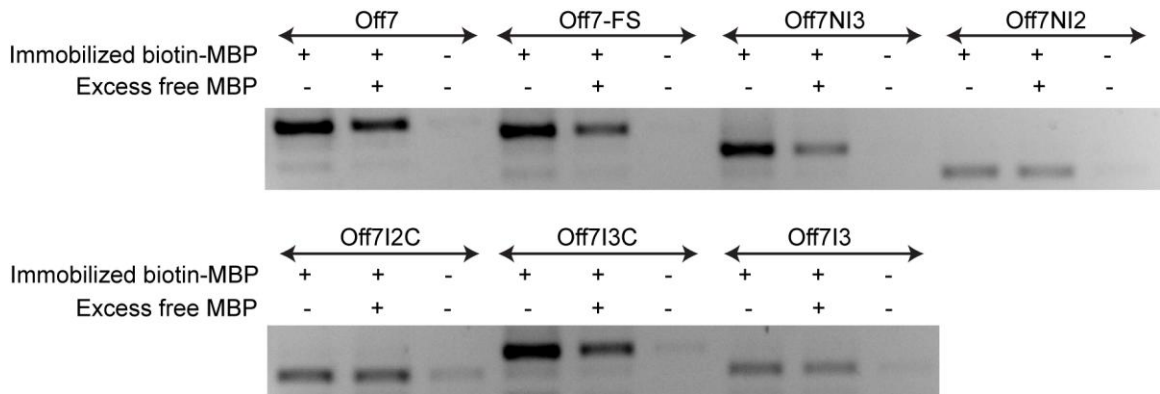
Figure 5.3. Ribosome display of Off7 and Off7 fragments against MBP. Biotinylated MBP was immobilized on neutravidin-coated wells prior to selection. To compete for specific Off7 binding to immobilized MBP, 3 μM non-biotinylated MBP was included during the panning step. Products after RT and 30 cycles of PCR are shown. The different sizes of the bands for each Off7 variant (as indicated above the gel) reflect different truncations of the wild-type Off7.

5.3.3 Ribosome display of SCI-57, human leptin, and leptin variants

The binding of SCI-57 and leptin to their cognate receptors was also investigated with ribosome display. SCI-57, like native insulin, requires three disulfide bonds for correct folding and biological activity[39]. Therefore, protein disulfide isomerase (PDI) was added to the translation reaction to enhance folding and the formation of disulfide bonds, which has previously been shown to have a strong beneficial effect in ribosome display[1]. With

PDI, ribosome display of SCI-57 suggested weak, but specific, binding to immobilized insulin receptors compared to background controls (Fig. 5.4A).
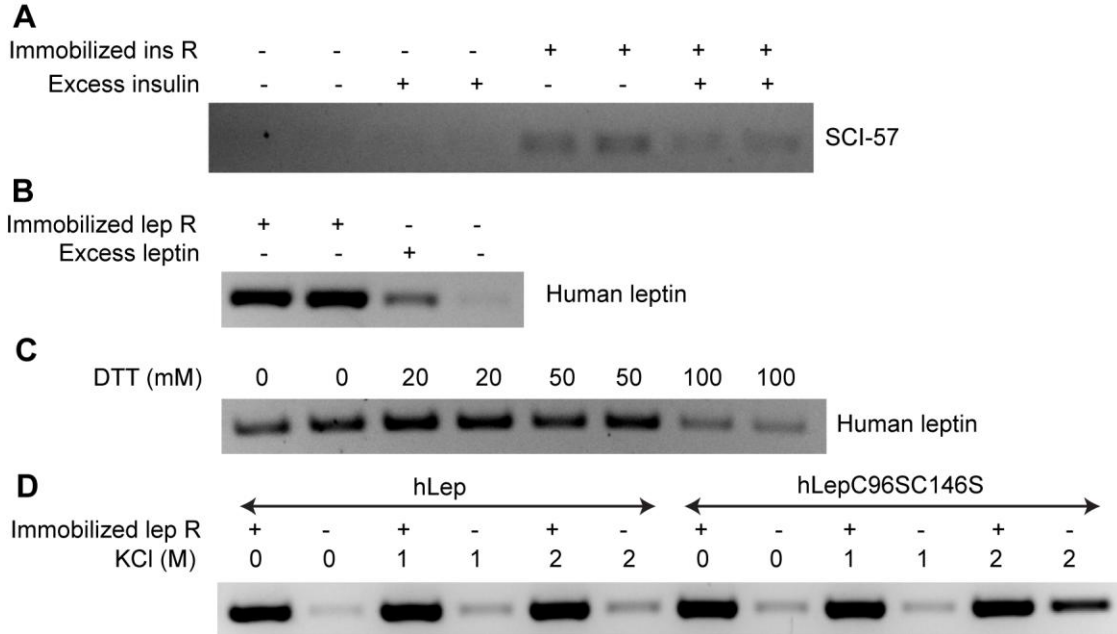
Figure 5.4. Ribosome display of SCI-57 and leptin against insulin and leptin receptors, respectively. (A) Insulin receptors were directly adsorbed onto wells and blocked with casein. To compete for SCI-57 binding, 30 μM excess insulin was included during the panning step. TBS-EDTA elution was used to isolate the mRNA. Products after RT and 19 cycles of PCR are shown. (B, C, D) Leptin receptors were directly adsorbed onto wells and blocked with casein. (B) To compete for leptin binding, 0.5 μM excess leptin was included during the panning step. Products after RT and 27 cycles of PCR are shown. (C) Various concentrations of DTT were included in the translation reaction with PURExpress. Translation was carried out as usual. Products after RT and 27 cycles of PCR are shown. (D) Various concentrations of KCl were included in WBT-cas during the panning step to assess hydrophobicity of the displayed protein. Both human leptin (hLep) and human leptin with C96S and C146S mutations (hLepC96SC146S) were tested. Products after RT and 25 cycles of PCR are shown.

Similarly, ribosome display of human leptin showed specific binding to immobilized leptin receptors and the level of recovered complexes diminished when free

leptin was added during the panning step (Fig. 5.4B). However, no isomerases or chaperones were used in this experiment. Native human leptin contains one intrachain disulfide bond between C96 and C146 (numbering starts after the signal peptide), but there are conflicting reports concerning the importance of this disulfide bond for biological activity[40–43]. To further investigate if the disulfide bond is required for binding to the leptin receptor, various concentrations of DTT were added to the translation reaction to disrupt disulfide bond formation. Only at the highest DTT concentration (100 mM) was a reduction in leptin-leptin receptor binding observed (Fig. 5.4C). We more stringently tested the importance of this disulfide bond by constructing a double mutant of leptin (C96S and C146S), which eliminated the disulfide bond completely. Ribosome display of leptin and leptin(C96S/C146S) showed no significant difference in leptin receptor binding under normal conditions (Fig. 5.4D). However, in high salt conditions (2 M KCl), the disulfide bond-free leptin variant is "stickier" than wild type, as assessed by the signals in the selection wells without leptin receptors. Since non-specific hydrophobic interactions are amplified under high salt concentration[44], our results suggest that the disulfide bond-free leptin variant is inherently less stable than wild-type leptin. Based on these results, the disulfide bond in leptin does not appear to be necessary for specific binding under normal conditions, but its presence enhances the stability of the molecule, which may be important both physiologically and biotechnologically.

176

5.3.4 CDR grafting of YN1 engenders specific binding of scFv to mouse ICAM-1

Full-length R6.5 and YN1 IgG antibodies show specific binding to human ICAM-1 or mouse ICAM-1, respectively, in ELISA experiments (C. Greineder, personal communication), demonstrating that the parental antibodies are capable of ICAM-1 binding and that the biotinylated human ICAM-1 and mouse ICAM-1 are functional. R6.5 and YN1 scFvs are in a V_H-linker-V_L orientation, where the linker consists of the 16 amino acids AGGGGSGGGGSGGGGS. These scFvs have been successfully expressed in insect cells (C. Greineder, personal communication). Surprisingly, when tested in ribosome display, no specific binding to human or mouse ICAM-1 can be observed with either scFv (Fig. 5.5A, B). Different variations to the display protocol were tested, including the addition of disulfide bond enhancers, different blocking agents, varying concentration of receptors and ternary complexes, and different elution methods, but all without success. Since full-length R6.5 and YN1 produced in insect cells are clearly functional, we hypothesized that the apparent lack of binding in ribosome display stems from the poor stability or folding of R6.5 and YN1 scFvs during *in vitro* synthesis. We therefore decided to graft the CDRs of these antibodies onto more stable scFv frameworks.

177
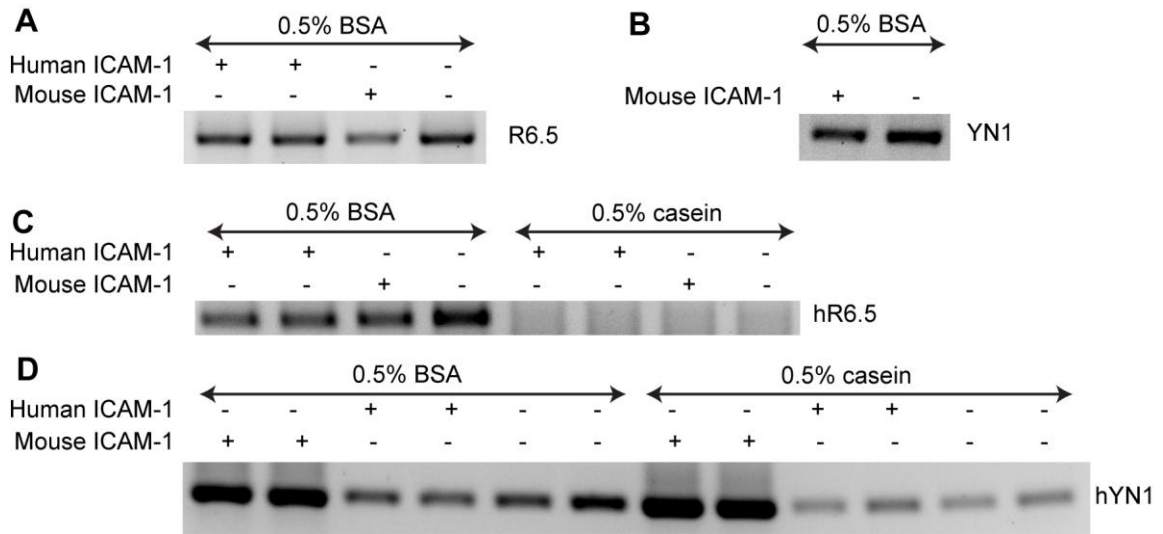
Figure 5.5. Ribosome display of anti-ICAM-1 scFvs against ICAM-1. Biotinylated human or mouse ICAM-1 was immobilized on neutravidin-coated wells and blocked with (A, B) BSA or (C, D) BSA or casein. The binding of (A) R6.5, (B) YN1, and (C) hR6.5 scFv with ICAM-1 was shown by ribosome display after 23 cycles of PCR. (D) The binding of the stabilized and humanized hYN1 with ICAM-1 was shown by ribosome display. Reverse primer 18 was used for RT and products after 25 cycles of PCR are shown.

Consensus human heavy- and light-chain frameworks from the HuCal library[35] were chosen. V_{H3} was chosen as the new heavy-chain framework for both R6.5 and YN1 because it is most commonly seen in nature[35], has exceptional stability[33,45], and was also the most frequently selected heavy-chain framework in previous ribosome display experiments[8]. $V_{\kappa2}$ and $V_{\kappa4}$ were chosen as the light-chain frameworks for R6.5 and YN1, respectively, since they most closely resemble their original light-chain frameworks, and because V_{κ} domains are more stable than V_{λ}[33]. The heavy and light chains were connected in the same V_H-linker-V_L orientation, but this time with the 20-amino-acid linker AGGGSGGGGSGGGGSGGGGS since this linker has previously been used to

connect scFvs for selection in ribosome display[8]. Since the new frameworks were chosen from human consensus sequences, the stability-enhanced scFvs hR6.5 and hYN1 were also simultaneously humanized. Although ribosome display of hR6.5 showed no specific binding to human ICAM-1 (Fig. 5.5C), ribosome display of hYN1 showed strong and specific binding to mouse ICAM-1 (Fig. 5.5D). Moreover, the choice of blocking agent did not affect binding, although casein blocking showed a lower background than BSA blocking. For RT, tolAk (primer 18) was used as the reverse primer instead of the usual inner reverse primer pRDV_BbsI_r (primer 16) since preliminary RT experiments showed the strongest specific bands when tolAk was used. Our CDR grafting strategy has enabled successful display of hYN1 scFv in ribosome display format, as assessed by specific mouse ICAM-1 binding.

5.4 Discussion

The Off7 truncation study revealed the importance of shielding the hydrophobic core of ankyrin repeats with the capping repeats. The removal of both N- and C-terminal capping repeats (Off7I3) showed no specific MBP binding (Fig. 5.3). Based on the co-crystal structure[9], the Off7-MBP interaction interface involves residues from the three internal repeats, but not the capping repeats (the single amino acid mutation in the N-cap is not responsible for binding)[9]. Therefore, the lack of Off7I3-MBP binding cannot be attributed to disruptions of the binding interface, but instead can be explained by the instability of the ankyrin repeat protein without the two capping repeats. With the

removal of either the N- or C-cap (Off7I3C or Off7NI3), no noticeable effect in binding to MBP was observed. However, our experimental conditions might not be stringent enough to explore any destabilization effect of a missing N- or C-cap, particularly since the tethered ribosome enhances the solubility of the displayed protein[44]. Nonetheless, our results underscore the important stabilizing effect of the hydrophilic caps in sealing the hydrophobic core of ankyrin repeat proteins[13].

On the other hand, the lack of specific binding by Off7NI2 and Off7I2C can be explained by disruptions to the Off7-MBP binding interface. Eight of the 12 potential interaction residues are located within the first and second internal repeats, while 10 of the 12 potential interaction residues are located within the second and third internal repeats[9]. It would be interesting to parse the energetic contributions of residues in each internal repeat by comparing MBP binding of Off7 with alanine substitutions for the interaction residues in the different internal repeats with ribosome display. Moreover, additional randomized repeats could be added to these fragments and selected for MBP binding to investigate if this step-wise approach could provide more stable and higher-affinity binders than wild-type Off7.

Ribosome display of SCI-57 showed specific binding to the insulin receptor (Fig. 5.4A), although the signal-to-background ratio could be further improved. One improvement could be replacing PDI with other chaperones or isomerases to achieve a translation reaction condition for optimal disulfide bonds formation. In particular, the disulfide bond enhancers from NEB could be tested, since they were specifically designed to supplement the PURExpress translation system to enhance proper folding of

disulfide-bonded proteins. Another improvement would be to optimize the buffer pH for better insulin folding. Previous work indicates that a basic pH (between 9.5 and 11) facilitates proper folding of insulin[46,47], and our work in engineering *Lactococcus lactis* to secrete SCI-57 (Chapter 6) also suggests that environmental pH plays an important role in obtaining bioactive molecules.

The role of the intrachain disulfide bond on the biological activity of leptin remains unclear. While Imagawa et al.[43] reported that eliminating the disulfide bond has no effect on human leptin signaling in mice, Müller et al.[40] and Giese et al.[41] demonstrated the opposite is true in rat adipocytes *in vitro* and in mice with mouse leptin, respectively. Our results suggest the disulfide bond is dispensable for specific binding of human leptin to leptin receptors, but the stability of the leptin double mutant (C96S/C146S) is likely to be lower. It would be interesting to investigate if the importance of the disulfide bond is species-specific. Future work could include engineering leptin(C96S/C146S) to increase its stability. A ribosome display screen could be set up under high salt conditions, where ternary complexes would first be incubated with agarose beads displaying different hydrophobic ligands to remove poorly folded proteins[44]. The remaining unbound complexes could then be selected for binding against leptin, achieving the dual engineering objective of increasing stability while maintaining (or enhancing) affinity.

The apparent lack of binding of hR6.5 to human ICAM-1 may be due to several reasons. It is possible that the V_{H3}-$V_{\kappa2}$ framework is less stable than the V_{H3}-$V_{\kappa4}$ framework used for hYN1; however, the ribosome display experiment carried out by

Hanes et al.[8] selected 12 useful binders with the V_{H3}-$V_{\kappa2}$ framework, but only 1 with the V_{H3}-$V_{\kappa4}$ framework. V_{H3}-$V_{\kappa2}$ is thus clearly able to produce scFvs that are functional and stable enough for ribosome display. Since full-length R6.5 IgG exhibited specific binding to human ICAM-1, one possibility is that the V_{H3}-$V_{\kappa2}$ framework cannot position the R6.5 CDRs in a conformation necessary for human ICAM-1 binding. Although the consensus human $V_{\kappa2}$ most closely resembles the wild-type R6.5 murine V_{κ} framework, other V_{κ} frameworks could be tested. Point mutations could be introduced into the framework of hR6.5 and screened for human ICAM-1 binding using ribosome display.

In summary, we have refined the ribosome display protocol, and we have demonstrated that Off7 fragments, SCI-57, leptin, and hYN1 can be successfully displayed in this format to achieve specific binding to their respective targets. Using ribosome display, we have elucidated some of the stability determinants of these molecules, including the importance of capping repeats in DARPins, the role of disulfide bonds in agonist folding, and the choice of antibody framework for functional display *in vitro*. Furthermore, these proteins are ready to be further engineered by constructing DNA libraries from these starting templates and employing *in vitro* evolution to quickly identify mutants with new and improved properties.

5.5 References

1.   Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4937–4942 (1997).

2. Roberts, R. & Szostak, J. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297–12302 (1997).

3. Lipovsek, D. & Plückthun, A. *In-vitro* protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67 (2004).

4. Dreier, B. & Plückthun, A. Ribosome display: a technology for selecting and evolving proteins from large libraries. *Methods Mol. Biol.* **687**, 283–306 (2011).

5. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Plückthun, A. Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 75–80 (2001).

6. Lamla, T. & Erdmann, V. A. searching sequence space for high-affinity binding peptides using ribosome display. *J. Mol. Biol.* **329**, 381–388 (2003).

7. Amstutz, P. *et al. In vitro* selection for catalytic activity with ribosome display. *J. Am. Chem. Soc.* **124**, 9396–9403 (2002).

8. Hanes, J., Schaffitzel, C., Knappik, a & Plückthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287–1292 (2000).

9. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004).

10. Binz, H. K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23**, 1257–1268 (2005).

11. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins Struct. Funct. Bioinforma.* **17**, 363–374 (1993).

12. Sedgwick, S. G. & Smerdon, S. J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316 (1999).

13. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).

14. Kohl, A. *et al.* Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1700–1705 (2003).

15. Ueda, T., Kanamori, T. & Ohashi, H. Ribosome display with the PURE technology. *Methods Mol. Biol.* **607**, 219–225 (2010).

16. He, M. & Taussig, M. Eukaryotic ribosome display with *in situ* DNA recovery. *Nat. Methods* **4**, 281–289 (2007).

17. Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nat. Methods* **4**, 269–279 (2007).

18. Dardeno, T. A. *et al.* Leptin in human physiology and therapeutics. *Front. Neuroendocrinol.* **31**, 377–393 (2010).

19. Oral, E. A. *et al.* Leptin-replacement therapy for lipodystrophy. *N. Engl. J. Med.* **346**, 570–578 (2002).

20. Welt, C. W. *et al.* Recombinant human leptin in women with hypothalamic amenorrhea. *N. Engl. J. Med.* **351**, 987–997 (2004).

21. Naito, M. *et al.* Therapeutic impact of leptin on diabetes, diabetic complications, and longevity in insulin-deficient diabetic mice. *Diabetes* **60**, 2265-2273 (2011).

22. Hua, Q. Insulin: a small protein with a long journey. *Protein Cell* **1**, 537–551 (2010).

23. Rajpal, G., Liu, M., Zhang, Y. & Arvan, P. Single-chain insulins as receptor agonists. *Mol. Endocrinol.* **23**, 679–688 (2009).

24. Hua, Q. *et al.* Design of an active ultrastable single-chain insulin analog: synthesis, structure, and therapeutic implications. *J. Biol. Chem.* **283**, 14703–14716 (2008).

25. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299–304 (2005).

26. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. & Plückthun, a Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14130–14135 (1998).

27. Cosimi, A. *et al*. *In vivo* effects of monoclonal antibody to ICAM-1 (CD54) in nonhuman primates with renal allografts. *J. Immunol.* **1**, 4604–4612 (1990).

28. Jedrzejas, M. J., Miglietta, J., Griffin, J. A. & Luo, M. Structure of a monoclonal anti-ICAM-1 antibody R6.5 Fab fragment at 2.8 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **51**, 380–385 (1995).

29. Lawson, C. & Wolf, S. ICAM-1 signaling in endothelial cells. *Pharmacol. Rep.* **61**, 22–32 (2009).

30. Kavanaugh, A. F. *et al.* Treatment of refractory rheumatoid arthritis with a monoclonal antibody to intercellular adhesion molecule 1. *Arthritis Rheum.* **37**, 992–999 (1994).

31. Salmela, K. *et al.* A randomized multicenter trial of the anti-ICAM-1 monoclonal antibody (enlimomab) for the prevention of acute rejection and delayed onset of graft function in cadaveric renal transplantation. *Transplantation* **67**, 729–736 (1999).

32. Furuya, K. *et al.* Examination of several potential mechanisms for the negative outcome in a clinical stroke trial of enlimomab, a murine anti-human intercellular adhesion molecule-1 antibody: a bedside-to-bench study. *Stroke* **32**, 2665–2674 (2001).

33. Ewert, S., Honegger, A. & Plückthun, A. Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. *Methods* **34**, 184–99 (2004).

34. Jung, S. & Plückthun, A. Improving *in vivo* folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng.* **10**, 959–66 (1997).

35. Knappik, A. *et al.* Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86 (2000).

36. Ewert, S., Cambillau, C., Conrath, K. & Plückthun, A. Biophysical properties of camelid V_{HH} domains compared to those of human V_H3 domains. *Biochemistry* **41**, 3628–3636 (2002).

37. Cull, M. & Schatz, P. Biotinylation of proteins *in vivo* and *in vitro* using small peptide tags. *Meth. Enzymol.* **326**, 430–440 (2000).

38. Ng, D. T. W. & Sarkar, C. A. Model-guided ligation strategy for optimal assembly of DNA libraries. *Protein Eng. Des. Sel.* **25**, 669–678 (2012).

39. Murray-Rust, J., McLeod, A., Blundell, T. & Wood, S. Structure and evolution of insulins: implications for receptor binding. *BioEssays* **14**, 325–331 (1992).

40. Müller, G., Ertl, J., Gerl, M. & Preibisch, G. leptin impairs metabolic actions of insulin in isolated rat adipocytes. *J. Biol. Chem.* **272** , 10585–10593 (1997).

41. Giese, K. *et al.* Reduction of food intake and weight gain by the ob protein requires a specific secondary structure and is reversible. *Mol. Med.* **2**, 50–58 (1996).

42. Boute, N. *et al.* The formation of an intrachain disulfide bond in the leptin protein is necessary for efficient leptin secretion. *Biochimie* **86**, 351–356 (2004).

43. Imagawa, K. *et al.* Structure-function studies of human leptin. *J. Biol. Chem.* **273**, 35245–35249 (1998).

44. Matsuura, T. & Plückthun, A. Selection based on the folding properties of proteins with ribosome display. *FEBS Lett.* **539**, 24–28 (2003).

45. Wörn, A. & Plückthun, A. An intrinsically stable antibody scFv fragment can tolerate the loss of both disulfide bonds and fold correctly. *FEBS Lett.* **427**, 357–361 (1998).

46. Hua, Q. *et al.* Mechanism of insulin chain combination: asymmetric roles of a-chain α-helices in disulfide pairing. *J. Biol. Chem.* **277**, 43443–43453 (2002).

47. Weiss, M. A. Proinsulin and the Genetics of Diabetes Mellitus. *J. Biol. Chem.* **284**, 19159–19163 (2009).

CHAPTER 6

A novel cell-based platform for oral delivery of insulin and leptin analogs

Although the advances in biomolecular engineering described in previous chapters are critical to improving the potencies of therapeutic proteins, effective delivery systems are also essential to maximize the therapeutic benefit of these proteins in a clinical setting.

6.1 Introduction

Oral administration is one of the most convenient ways of delivering drugs, but therapeutic proteins often must be administered by an invasive method such as intravenous or subcutaneous injection[1]. Oral delivery of protein drugs is not generally feasible due to poor stability during passage through the gastrointestinal tract and low permeability across the intestinal wall, resulting in insufficient bioavailability. To increase oral bioavailability of proteins, various encapsulation strategies have been developed to protect the polypeptides from enzymatic digestion[2], but most have encountered roadblocks that prevent them from advancing to a clinical setting[3]. Thus, better oral delivery systems are needed and one promising option is the microorganism *Lactococcus lactis*.

L. lactis is a Gram-positive bacterium widely used in the food industry for production of fermented products such as buttermilk and cheese and is therefore routinely consumed in these foods. *L. lactis* has a safe association with humans and has been proposed for use as a probiotic[4]. There has been increasing interest in the use of *L. lactis* as a mucosal delivery vehicle because it can survive passage through the stomach acid and contact with bile[5] and it can be engineered to express and secrete targeting molecules and adjuvants[6]. In this host, antigens and DNA have been introduced for mucosal vaccine delivery, single-chain variable fragments (scFvs) for anti-infectives, and allergens for allergy prevention[7]. To address potential safety concerns of using live *L. lactis* in humans, the thymidylate synthase gene can be removed from the host genome, rendering the auxotrophic bacteria dependent on thymidine or thymine for survival and thus biologically contained[8]. In addition, since the recombinant protein is still locally produced when the bacteria reach the intestine, proteolytic degradation is attenuated. Recently, a study involving the use of interleukin-10-secreting *L. lactis* to treat Crohn's disease has passed phase I clinical trials, supporting the notion that this live microorganism is a viable platform for oral protein delivery[9,10]. Here, we evaluated whether *Lactococcus lactis* would be capable of producing a bioactive insulin analog, human leptin, and insulin-leptin fusion proteins *in vitro*, and we further examined the *in vivo* performance of insulin-secreting *L. lactis*. Our work would open the door for using this host as a vehicle for oral delivery of protein therapeutics.

In addition to clear benefits in the ease of administration and improved patient compliance, oral delivery of insulin also most accurately recapitulates normal

physiological delivery. After absorption in the intestine, orally delivered insulin reaches the portal system, more closely approximating what occurs in a non-diabetic individual[11]. Insulin is normally secreted by pancreatic β-cells in the form of a single-chain precursor, proinsulin, which is subsequently cleaved into separate A (21 residues) and B (30 residues) chains[12]. However, this heterodimeric protein with three disulfide bonds (two interchain and one intrachain) would be challenging to synthesize using *L. lactis*. We therefore chose to heterologously express a single-chain insulin analog, SCI-57, which contains four substitutions in the A and B chains (one of which is present in Novalog® and another in Humalog®, rapid-acting insulin analogs already in clinical use by injection) and a 6-residue linker (GGGPRR) connecting these two chains, resulting in a single polypeptide 57 amino acids in length[13]. SCI-57 not only resembles the folding and biological activity of wild-type insulin, but it also has enhanced thermodynamic stability and reduced aggregation, and allows for simpler single-chain synthesis, making it an attractive insulin analog for oral delivery[13,14].

To further demonstrate the feasibility of using *L. lactis* as a general platform for oral delivery of protein drugs, we expressed another promising therapeutic protein, leptin, using the same system. Largely produced in white adipose tissue in healthy individuals, leptin acts mainly in the brain to control food intake and energy metabolism[15]. Leptin also targets non-hypothalamic tissues such as hematopoietic cells, T cells, adipocytes, skeletal muscles and hepatocytes[16]. Since its discovery two decades ago, leptin therapy has shown promises in treating a variety of diseases, including obesity[17], lipodystrophy[18], hypothalamic amenorrhea[19], and diabetes[20]. In fact, leptin/insulin co-therapy in the

treatment of diabetes has recently been shown to have advantages over insulin monotherapy by providing equivalent glycemic control without the increase in body fat and up-regulation of cholesterologenic and lipogenic transcription factors and enzymes[21]. We further investigated if fusions of SCI-57 and leptin could be expressed with both proteins retaining their biological functions. Using leptin as a fusion partner has the additional advantage of facilitating intestinal absorption of secreted SCI-57, since receptor-bound leptin is absorbed by duodenal enterocytes and released on the basolateral side towards the bloodstream through clathrin-mediated endocytosis and subsequent intracellular transit and exocytosis[22,23]. Since fusion proteins have long been shown to be able to cross the intestinal endothelial layer through receptor-mediated transcytosis[24], leptin represents a potential transcytosis enhancer as well as a diabetic therapeutic when delivered together with SCI-57.

We inducibly expressed SCI-57, leptin, and SCI-57-leptin fusions in *L. lactis* strain NZ9000 using the NICE system, which is based on a two-component signaling system involved in the biosynthesis of the bacteriocin nisin[25]. Briefly, upon nisin-binding, NisK, a histidine kinase, autophosphorylates and transfers its phosphate group to NisR, which when activated induces transcription of a gene of interest under the control of promoter P*nisA*[26,27]. Previous studies of this system using a reporter gene show a linear dose–response curve for increasing amounts of nisin[28]. For *in vivo* delivery of *L. lactis* secreting SCI-57, the cultures were pre-induced with nisin before oral administration in order to achieve immediate therapeutic effect and also to eliminate the need for ingestion

of nisin. Protein expression has been shown to proceed for 10 hours after an one-hour nisin pulse-induction[29].

We demonstrate that, upon nisin induction, *L. lactis* can secrete SCI-57, leptin, and SCI-57-leptin fusions if the genes are fused to the usp45 signal peptide[30], although bacterial growth rate depends on the timing of nisin induction. We further demonstrate that secreted SCI-57 and leptin are biologically active by assaying for insulin receptor signaling in an adipocyte cell line and leptin signaling in leptin receptor transfected cells, respectively. Interestingly, total expression does not correlate with bioactivity on cells, suggesting that proper folding of the polypeptide is a bottleneck in maximizing functional yield. We identify culture pH as an important regulator of bioactivity of SCI-57 and show that a neutral to slightly alkaline pH can significantly enhance the fraction of functional secreted SCI-57. Finally, we demonstrate reduction in blood glucose levels in a diabetic mouse model after oral administration of *L. lactis* secreting SCI-57, highlighting the potential of this novel cell-based platform for oral delivery of therapeutic proteins.

6.2 Materials and methods

6.2.1 Bacterial strains, plasmids, and growth conditions

Bacterial strains and plasmids used in this work are listed in Table 6.1. *Escherichia coli* was grown in TY medium (8 g tryptone, 5 g yeast extract, and 5 g sodium chloride per liter) at 37°C with shaking and *L. lactis* was grown in M17 medium (Oxoid, Hamshire, UK) containing 0.5% glucose (GM17) at 30°C statically (*i.e.*, without shaking). For *in*

vivo testing, *L. lactis* was resuspended in M9 medium (Amresco, Solon, OH) containing 0.5% glucose buffered with 50 mM CO_3^{2-}/HCO_3^- (BGM9). For buffering BGM9, 1M Na_2CO_3 and 1M $NaHCO_3$ were mixed at a molar ratio of 1:19 and added at a final concentration of 50 mM. Solid media were prepared by adding agar (15 g/l) to the corresponding broth. Chloramphenicol (Cm) was used at a final concentration of 10 µg/ml when culturing *E. coli* EC1000 or *L. lactis* NZ9000.

Table 6.1. Bacterial strains and plasmids used.

	Characteristics	Sources
Strains		
E. coli EC1000	RepA$^+$ MC1000, KM^r, carrying a single copy of the pWV01 *repA* gene in *glgB*	Leenhouts, et al.[31]
L. lactis NZ9000	*L. lactis* MG1363 (*nisRK* genes on the chromosome)	Kuipers et al.[27]
Plasmids		
pRDV:SCI-57	pRDV containing the SCI-57 gene; Amp^r	Chapter 5
pNZPnisA:CYTO-LLO	Modified pNZ8048 containing P*nisA* promoter with His-tagged *hly* gene; Cm^r	Bahey-El-Din et al.[32]
pNZPnisA:SCI-His	Modified pNZ8048 containing P*nisA* promoter with RGS-His-tagged *SCI-57* gene; Cm^r	This work
pNZPnisA:usp-SCI-His	Modified pNZ8048 containing P*nisA* promoter with RGS-His-tagged *SCI-57* gene; fusion with *usp45* signal peptide; Cm^r	This work
pNZPnisA:usp-SCI	Modified pNZ8048 containing P*nisA* promoter with *SCI-57* gene without tag; fusion with *usp45* signal peptide; Cm^r	This work
pNZPnisA:usp-His-Lep	Modified pNZ8048 containing P*nisA* promoter with RGS-His-tagged *Leptin* gene; fusion with *usp45* signal peptide; Cm^r	This work
pNZPnisA:usp-Lep	Modified pNZ8048 containing *PnisA* promoter with *Leptin* gene without tag; fusion with *usp45* signal peptide Cm^r	This work
pNZPnisA:usp-His-SCI-(linker)-Lep	Modified pNZ8048 containing *PnisA* promoter with RGS-His-tagged *SCI-57* and *Leptin* gene connected by a linker (one of the six); fusion with *usp45* signal peptide Cm^r	This work
pNZPnisA:usp-SCI-(linker)-	Modified pNZ8048 containing *PnisA* promoter with	This work

Lep	*SCI-57* and *Leptin* gene connected by a linker (one of the six) without tag; fusion with *usp45* signal peptide Cm^{r}	
pNZPthyA:usp-SCI	Modified pNZ8048 containing P*thyA* promoter with *SCI-57* gene without tag; fusion with *usp45* signal peptide; Cm^{r}	This work
pNZP23:usp-SCI	Modified pNZ8048 containing P23 promoter with *SCI-57* gene without tag; fusion with *usp45* signal peptide; Cm^{r}	This work
pNZP1:usp-SCI	Modified pNZ8048 containing P1 promoter with *SCI-57* gene without tag; fusion with *usp45* signal peptide; Cm^{r}	This work

6.2.2 DNA manipulations and transformations

Plasmid DNA from *E. coli* was isolated with a Qiagen miniprep kit (Qiagen, Valencia, CA) following the standard procedure; plasmid DNA from *L. lactis* was isolated with the same protocol except for an additional incubation of the cells with 4 mg/ml lysozyme (USB Affymetrix, Cleveland, OH) in P1 buffer at 37°C for 30 min. Phusion high-fidelity DNA polymerase (NEB, Ipswich, MA) was used in all PCR reactions as recommended by the manufacturer. Restriction enzymes and T4 DNA ligase were purchased from NEB. DNA was transformed into *E. coli* as described previously[33]. DNA was transformed into *L. lactis* by electroporation as described in the manufacturer's manual (MoBiTec, Göttingen, Germany).

Primers used for DNA amplification are listed in Table 6.2. SCI-57 was PCR amplified from pRDV:SCI-57 using primers 1/3 and 2/3 in a preliminary step for building full constructs with and without usp45 signal peptide, respectively. The resulting PCR products were further amplified with primers 4/9 and 2/9, respectively, to add the usp45 signal peptide and RGS-His tag or only the RGS-His tag. The P*nisA* promoter

region was PCR amplified using primers 7/8, with 8 introducing an EcoRI site at the end of P*nisA*. The resulting product was fused to the SCI-57-RGS-His gene product with or without usp45 signal peptide by assembly PCR. The assembled products were gel-purified and sequentially digested with KpnI and BglII. The digested products were then ligated into the similarly digested pNZPnisA:CYTO-LLO (plasmid courtesy of Dr. Cormac Gahan, University College Cork) using T4 DNA ligase to give pNZPnisA:usp-SCI-His (Fig. 6.1A) and pNZPnisA:SCI-His. To remove the RGS-His tag from pNZPnisA:usp-SCI-His to obtain pNZPnisA:usp-SCI, primers 5/6 were used to PCR amplify from pRDV:SCI-57. The resulting SCI-57 gene product without the RGS-His tag was sequentially digested using KpnI and BspEI and the product was ligated into the similarly digested pNZPnisA:usp-SCI-His vector, giving rise to pNZPnisA:usp-SCI.

Table 6.2. Primers used in this study.

Primer Number	Primer Name	Nucleotide Sequence (5' → 3') [a]
1	SCI57_usp_f	TCCGGAGTTTACGCTTTCGTTAACCAGCAC
2	SCI57_nousp_f	CACTCAAAGAATTCATGTTCGTTAACCAGCAC
3	SCI57_rgshis_r	TGGTGGTGATGGTGGGATCCTCTGTTGCAGTAGTTTTCCA
4	usp45_f	GCACTCAAAGAATTCATGAAAAAAAAGATTATCTCAGCTATTTTAATGTCTACAGTGATACTTTCTGCTGCAGCCCCGTTGTCCGGAGTTTACGCT
5	uspSCI57_f	CCCCGTTG<u>TCCGGA</u>GTTTACGCTTTCGTTAACCAGCAC
6	stopSCI57_KpnI_r	GAACTAGT<u>GGTACC</u>TCATTAGTTGCAGTAGTTTTCC
7	PnisA_BglII_f	TACAGCTCCA<u>AGATCT</u>AGTC
8	PnisA_EcoRI_r	CATG<u>AATTC</u>TTTGAGTGCCTCCTTATA
9	rgshis_KpnI_r	GAACTAGT<u>GGTACC</u>TCATTAATGATGGTGGTGATGGTGG
10	rgshis_lep_BspEI_f	AC<u>TCCGGA</u>GTTTACGCTAGAGGATCCCACCATCACCACCATCATGTGCCCATCCAAAAAG
11	lep_BspEI_f	CCCCGTTG<u>TCCGGA</u>GTTTACGCTGTGCCCATCCAAAAAG
12	lep_KpnI_r	GAACTAGT<u>GGTACC</u>TTATCAGCACCCAGGGCTG

13	GSshort_f	ACGGTGGCAGTGGTGTGCCCATCCAAAAAGTCCA
14	GSshort_r	ACACCACTGCCACCGTTGCAGTAGTTTTCCAGCT
15	GSmedium_f	CAGGCGGTAGTGGTGGCAGTGTGCCCATCCAAAAAGTCCA
16	GSmedium_r	CCACCACTACCGCCTGCACCGTTGCAGTAGTTTTCCAGCT
17	GSlong_f	GCGGCGGTGCAGGAGGCGGTAGTGGTGTGCCCATCCAAAAAGTCCA
18	GSlong_r	CCTCCTGCACCGCCGCTTCCGCCACCGTTGCAGTAGTTTTCCAGCT
19	helical_f	GCAGCTAAAGAAGCAGCCGCTAAAGCTGTGCCCATCCAAAAAGTCCA
20	helical_r	GGCTGCTTCTTTAGCTGCAGCTTCTGCGTTGCAGTAGTTTTCCAGCT
21	TCres_f	CCAGATGGTGACATTGATGGAAGTGTGCCCATCCAAAAAGTCCA
22	TCres_r	ATCAATGTCACCATCTGGACTACCGTTGCAGTAGTTTTCCAGCT
23	TCresThrom_f	CTTGTTCCACGTGGTTCTGGAAGTGTGCCCATCCAAAAAGTCCA
24	TCresThrom_r	AGAACCACGTGGAACAAGACTACCGTTGCAGTAGTTTTCCAGCT
25	PthyA_BglII_f	AGCTCC<u>AGATCT</u>AGAAAAAGCAGCTGTTGAATTAG
26	PthyA_EcoRI_r	TTTCAT<u>GAATTC</u>GAAATTTTCCTATCTTTTTTAATTCCCTTC
27	P23_BglII_f	AGCTCC<u>AGATCT</u>GATGGGGATAAAAGTGACCCG
28	P23_EcoRI_r	TTTCAT<u>GAATTC</u>ATATTTGGCCTCCCTTTTTAATTTAATTC
29	P1_BglII_f	AGCTCC<u>AGATCT</u>GATTAAAGTCATCTTACCTCTTTTATTAG
30	P1_asse_r	ATTTCTAGTGGGAAACCGTTGTGGTCTCCCGATACTTGTATTATAAC ATATCTACAAAGG
31	P1_asse_end_EcoRI _r	TTCAT<u>GAATTC</u>GTATATCTCCTTTCTAAAGTTAAACAAAATTATTTC TAGTGGGAAACCG

[a] Restriction sites are underlined.
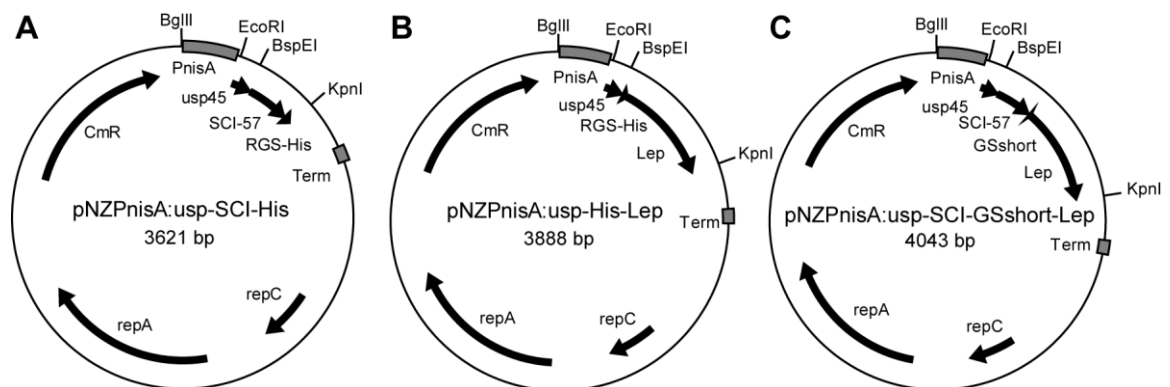
Figure 6.1. Schematic representation of the various constructs used in this study. Representative maps of (A) pNZPnisA:usp-SCI-His, (B) pNZPnisA:usp-His-Lep and (C) pNZPnisA:usp-SCI-GSshort-Lep showing promoter PnisA, usp45 signal peptide, SCI-57 gene, human leptin gene (Lep), GS short (GGSG) linker and RGS-His tag in the modified pNZ8048 backbone. All plasmids were maintained in *E. coli* EC1000 and subsequently transformed into electrocompetent *L. lactis* NZ9000.

To construct pNZPnisA:usp-His-Lep (Fig. 6.1B) and pNZPnisA:usp-Lep, human leptin cDNA (GenBank accession no. BC069452; Open Biosystems, Waltham, MA) was PCR amplified using primers 10/12 and 11/12 respectively. The resulting products were digested with KpnI and BspEI and the products were ligated into similarly digested pNZPnisA:usp-SCI. For SCI-57-leptin fusion plasmids, six different linkers were constructed: GGSG (GS short), GAGGSGGS (GS medium), GGGSGGAGGGSG (GS long), AEAAAKEAAAKA (helical), GSPDGDIDGS (trypsin/chymotrypsin resistant), and GSLVPRGSGS (trypsin/chymotrypsin resistant and thrombin cleavable). pNZPnisA:usp-SCI-(linker)-Lep (Fig. 6.1C) was constructed by first PCR amplifying the two pieces for assembly PCR, 1) pNZPnisA:SCI amplified with primers 5/14, 5/16, 5/18, 5/20, 5/22 and 5/24 for linker-fused SCI-57, and 2) leptin cDNA amplified with primers 13/12, 15/12, 17/12, 19/12, 21/12 and 23/12 for linker-fused leptin. The two pieces for each unique linker were then assembled in assembly PCR reactions, digested with KpnI and BspEI, and ligated into similarly digested pNZPnisA:usp-SCI.

To construct pNZPthyA:usp-SCI and pNZP23:usp-SCI, the promoter from the *L. lactis thyA* gene P*thyA*, and promoter P23, were isolated from NZ9000 by colony PCR with primers 25/26 and 27/28, respectively. To construct pNZP1:usp-SCI, the promoter

P1 was amplified from NZ9000 by colony PCR with primers 29/30, followed by a second PCR with primers 29/31. The resulting products were digested by EcoRI and BglII and ligated into similarly digested pNZPnisA:usp-SCI. All ligation mixtures were transformed into chemically competent *E. coli* EC1000 (strain courtesy of Dr. Jan Kok, University of Groningen). After confirming the constructs by DNA sequencing, the plasmids were transformed into electrocompetent *L. lactis* NZ9000.

6.2.3 Growth curve determination

Overnight cultures of *L. lactis* NZ9000(pNZPnisA:usp-SCI-His or pNZPnisA:usp-His-Lep) were diluted 1:25 into fresh GM17Cm medium. Nisin (Sigma, St. Louis, MO) at various concentrations was added at indicated times. Cultures were grown statically at 30°C up to 4 h after the latest induction point. Growth curves were determined by taking measurements of the optical density at 600 nm (OD_{600}) on a plate reader (Infinite M200, Tecan, Männedorf, Switzerland). The growth curves were fitted in Matlab (MathWorks, Natick, MA) using a logistic equation, $K/(1 + e^{-r(t-l)})$, where r is the growth rate, and the doubling times were calculated using $\ln(2)/r$.

6.2.4 Nisin induction and detection of secreted SCI-57 and leptin

Overnight cultures of *L. lactis* NZ9000(pNZPnisA:usp-SCI-His, pNZPnisA:SCI-His, pNZPnisA:usp-His-Lep, pNZPnisA:usp-SCI, or pNZPthyA:usp-SCI) were diluted 1:25 into fresh GM17Cm medium. For buffering with sodium phosphates, 1 M NaH_2PO_4 and

1 M Na$_2$HPO$_4$ were mixed at a molar ratio of 1:19 and added at a final concentration of 50 mM to achieve the desired pH. Cultures were grown to mid-log phase (OD$_{600}$ ≈ 0.4–0.5) for 1.5 to 3.5 h and induced with 1 or 10 ng/ml nisin. Cultures were grown for a total of 7.5 h. When investigating the effect of pH modulation, 2% or 10% culture volume of 5 N NaOH was added. At specified time points, aliquots were taken and OD$_{600}$ and pH were measured. Cells were removed from the supernatant by a 10-min centrifugation at 4°C and 5,000g. The supernatant was then passed through a 0.22-μm filter (Millipore, Billerica, MA) to remove any cells, and concentrated 20-fold in a 3-kDa cutoff filter (Millipore). 15.6 μl of concentrated or unconcentrated supernatant was mixed with reducing agent and lithium dodecyl sulfate (LDS) sample buffer for analysis by SDS-PAGE in a 12% NuPAGE® Bis-Tris gel as described by the manufacturer (Invitrogen, Carlsbad, CA). Proteins were then transferred to a nitrocellulose membrane (Invitrogen). RGS-His-tagged proteins were analyzed by Western blotting with the RGS-His antibody (Qiagen, #34610) or insulin antibody (Santa Cruz Biotechnology, Dallas, TX, sc-9168) and then IRDye800-conjugated goat-anti-mouse immunoglobulin G secondary antibody (Rockland, Gilbertsville, PA, #610-131-121) or goat-anti-rabbit immunoglobulin G secondary antibody (Rockland, #611–132–122). The blot was then scanned on an Odyssey infrared imager (LI-COR Biosciences, Lincoln, NE).

6.2.5 Bioactivity of secreted SCI-57 on 3T3-L1 cells

The murine 3T3-L1 preadipocyte cell line (courtesy of Dr. Christopher Chen, University of Pennsylvania) was maintained in Dulbecco's modified Eagle's medium (DMEM;

Invitrogen) supplemented with 10% calf serum (HyClone, Logan, UT), penicillin (100 U/ml), and streptomycin (100 μg/ml), with a change of medium every 3 days. The cells were differentiated using a slight modification of a published method[34]. Briefly, 3T3-L1 preadipocytes were allowed to grow for 2 days post-confluency and were then differentiated by addition of the same medium containing isobutylmethylxanthine (500 μM), dexamethasone (1 μM), and insulin (1 μg/ml) for 2 days and then medium containing only insulin additive for 3 additional days. The medium was then changed every 3 days until the cells contained large oil locules characteristic of fully differentiated adipocytes, typically around 9–12 days.

3T3-L1 preadipocytes were seeded on a six-well plate and differentiated as described above. Fully differentiated 3T3-L1 adipocytes were serum starved overnight with 0.5% calf serum in DMEM. Supernatant of *L. lactis* NZ9000(pNZPnisA:usp-SCI-His, pNZPnisA:usp-SCI or pNZPnisA:usp-SCI-(linker)-Lep) was prepared as described above. 100 μl of concentrated supernatant was added along with 2 ml DMEM to the serum-starved 3T3-L1 cells. Dilution into DMEM also ensured that all signaling assays were performed at near-neutral pH. After a 15-min incubation at 37°C, cells were washed once in PBS and lysed in cell extraction buffer (Invitrogen) supplemented with a protease inhibitor cocktail (Sigma, #P8340), phosphatase inhibitor cocktails (Sigma, #P0044 and #P5726), and phenylmethanesulfonylfluoride (Amresco, Solon, OH). Cell debris was removed by centrifugation at 14,000g and 4°C for 15 min. Cell lysate (20–50 μg, but constant for a given experiment) was mixed with reducing agent and LDS sample buffer for analysis by SDS-PAGE in a 4–12% Bis-Tris gel (Invitrogen). Proteins were then

transferred to a nitrocellulose membrane. Phosphorylated Akt (p-Akt) was quantified by Western blotting, as described above, using p-Akt(Ser473) or total Akt primary antibodies (Cell Signaling Technology, Danvers, MA, #4051 and #9272, respectively) and then IRDye800-conjugated goat-anti-mouse or goat-anti-rabbit immunoglobulin G secondary antibody (same as above). The proteins were quantified by their relative intensities from the IR800 channel on the infrared imager.

6.2.6 Bioactivity of secreted leptin on transfected HEK293 cells

Human embryonic kidney (HEK293) cells (Clontech #630931, Mountain View, CA) were maintained in DMEM supplemented with 10% fetal bovine serum (HyClone), penicillin (100 U/ml), and streptomycin (100 μg/ml). Transient transfection was carried out as described[35] but with modifications. Cells were seeded on a six-well plate at a density of 10^4 cells/ml with 2 ml per well. When the cells were around 50–80% confluent, they were transfected by adding FuGENE® 6 (Roche, Indianapolis, IN) and 700 ng leptin receptor (pcDNA3:hOBRb; courtesy of Dr. Christian Bjorbaek, Harvard Medical School), 300 ng signal transducer and activator of transcription-3 (STAT3)-activated firefly luciferase reporter (4xM67 pTATA TK-Luc; Addgene #8688, Cambridge, MA), and 0.01 ng renilla luciferase reporter (pcDNA-RLuc8; courtesy of Dr. Andrew Tsourkas, University of Pennsylvania) at a 3:1 FuGENE (μl):total DNA (μg) ratio. After 24 h, the media was changed into DMEM with 1% FBS containing either commercial leptin (Peprotech, Rocky Hill, NJ) or 50 μl of unconcentrated supernatant from *L. lactis* NZ9000(pNZPnisA:usp-Lep-His, pNZPnisA:usp-Lep, pNZPnisA:usp-His-

SCI-(linker)-Lep, or pNZPnisA:usp-SCI-(linker)-Lep) culture, prepared as described above. After a 24-h incubation at 37°C, cells were washed once in PBS and lysed in 500 μl passive lysis buffer (Promega, Madison, WI) for 15 min at room temperature. After centrifugation at 20,000g for 2 min, the supernatants were assayed with a dual-luciferase reporter assay (Promega) as described by the manufacturer using the luciferase feature on the Tecan plate reader (no attenuation; 0 ms settle time; 5000 ms integration time).

6.2.7 *In vivo* administration of *L. lactis* to diabetic mice

Streptozotocin was injected intraperitoneally into 5 C57BL/6 mice (male, 10–12 weeks old; The Jackson Laboratory; courtesy of Dr. Ali Naji, University of Pennsylvania) to induce diabetes one week before the inoculation experiments. Food and fluid were supplied *ad libitum* and tail blood glucose level was monitored daily with a glucometer (One Touch Ultra II). All animal procedures were carried out in accordance with Institutional Animal Care and Use Committee guidelines (IACUC protocol #803168).

Overnight cultures of *L. lactis* NZ9000(pNZPnisA:usp-SCI or pNZPthyA:usp-SCI) were diluted 1:25 into fresh GM17Cm medium and grown at 30°C until mid-log phase. The nisin-inducible culture was induced with 10 ng/ml nisin for 1 h. Both cultures were harvested by centrifugation at 5000g for 10 min at room temperature, and then resuspended in BGM9 media to a final concentration of 2×10^{10} to 2×10^{11} CFU/ml. Intragastric inoculation was performed by administering 500 μl *L. lactis* by oral gavage, using two mice for each culture. Tail blood glucose was measured before *L. lactis*

administration, 1 h, 2 h, 3 h, and 24 h after inoculation. A control mouse received oral gavage of 1 U insulin solution or no treatment.

6.3 Results

6.3.1 Effect of nisin addition on *L. lactis* growth rate

The growth rate of *L. lactis* NZ9000(pNZPnisA:usp-SCI-His) as a function of inducer concentration and time of induction was monitored by taking OD_{600} measurements every 30 min for 7 h. When added at 1 h, nisin affected cell growth in a dose-dependent manner, with the largest inhibitory effect occurring at the highest inducer concentration (10 ng/ml, Fig. 6.2A). Lower OD_{600} readings were observed as early as the first time point after nisin addition (within 30 min) and this attenuated signal persisted until the last time point (7 h), indicating that nisin acted immediately and continuously on the culture until saturation. However, when the culture was induced at 2.5 h, the extent to which nisin adversely affected growth rate and saturated culture density was noticeably mitigated (Fig. 6.2B). We therefore tested a wider range of inducer concentrations (0.001, 0.01, 0.1, 0.5, 1, 5, or 10 ng/ml nisin) and induction start times (0.5, 1, 1.5, 2, 2.5, or 3 h) and we quantified the doubling time for each combination of these two variables (Fig. 6.2C). Low nisin concentrations (≤ 0.1 ng/ml) had a minimal effect on cell growth at all induction times. High nisin concentrations (≥ 0.5 ng/ml) resulted in significantly longer doubling times at early induction start times, but there was a clear inverse correlation between induction start time and culture doubling time. The effect of nisin addition on

the growth rate of *L. lactis* NZ9000(pNZPnisA:usp-His-Lep) followed a similar trend, but the effect of nisin inhibition at each induction start time was greater for nisin concentrations ≥ 0.1 ng/ml, resulting in lower final culture densities (Fig. 6.2D, E) .
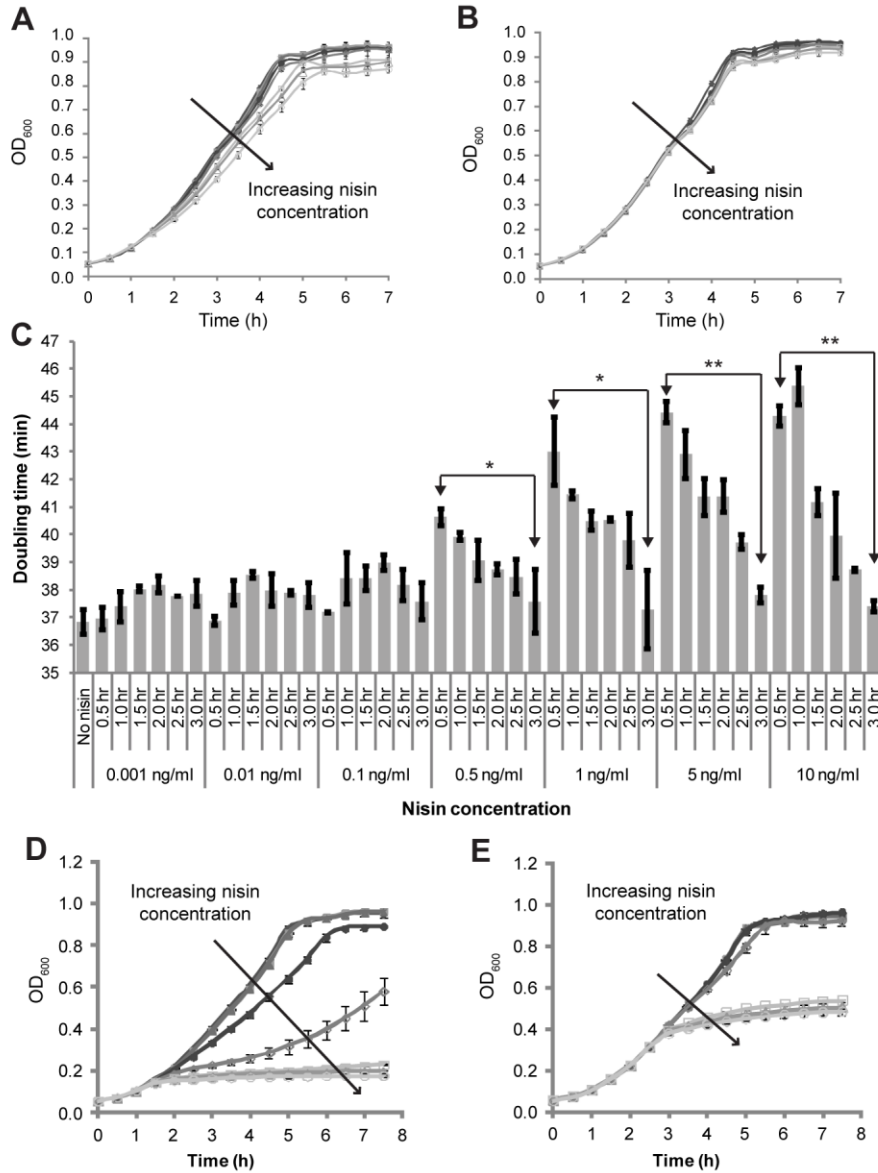
**Figure 6.2.** Effect of nisin concentration and induction time on growth rate. Various nisin concentrations (♦: no nisin; ■: 0.001 ng/ml; ▲: 0.01 ng/ml; ●: 0.1 ng/ml; ◊: 0.5 ng/ml; □: 1 ng/ml; Δ: 5 ng/ml; ○: 10 ng/ml) were added at (A) 1 h and (B) 2.5 h after 1:25 dilution of an overnight culture of

NZ9000(pNZPnisA:usp-SCI-His). Cultures were grown statically at 30°C for a total of 7 h. Growth was determined by measuring the optical density at 600 nm over time. (C) The growth curves were fitted to a logistic equation $K/(1 + e^{-r(t-l)})$, where r is the growth rate, and the doubling times were calculated from $\ln(2)/r$. *$P < 0.05$ and **$P < 0.01$ for statistical comparison of doubling times after 0.5 and 3 h induction at each nisin concentration using a one-tailed Student's t-test. (D, E) Effect of nisin concentration on growth rate of NZ9000(pNZPnisA:usp-His-Lep). Various nisin concentrations (same symbols as A, B) were added at (D) 1 h and (E) 2.5 h after dilution of overnight culture.

6.3.2 Secretion of SCI-57 and leptin by *L. lactis* into supernatant

L. lactis NZ9000(pNZPnisA:usp-SCI-His or pNZPnisA:SCI-His) were grown and induced with 0, 1, or 10 ng/ml nisin at 2.5 h. The 2.5-h induction start time was initially chosen because it seemed to balance growth rate (cultures are minimally affected even at 10 ng/ml nisin) and total induction time (SCI-57 expression can still proceed for several hours in a log-phase culture, Fig. 6.2). After induction for 4 additional hours, cells were removed by centrifugation and secreted SCI-57 was detected via the C-terminal RGS-His tag using Western blotting. Bands of ≈7–8 kDa (expected: 7.5 kDa) were detected in cultures induced with 1 or 10 ng/ml nisin, while no signal was detected in the nisin-free control (Fig. 6.3A). This confirms that there is no detectable leaky expression from the P*nisA* promoter and that nisin is necessary for inducing SCI-57 expression. In addition, no signal was detected in the supernatant from *L. lactis* harboring pNZPnisA:SCI-His (without usp45 signal peptide), whether or not nisin was added, indicating that the usp45 signal peptide is necessary for secretion of the downstream protein, which is in agreement with previous studies[36,37]. The molecular weight of RGS-His-tagged SCI-57 with the usp45 signal peptide is 10.3 kDa. All of our observed bands are below the 10 kDa mark,

indicating that the usp45 signal peptide has been cleaved, as expected, from SCI-57 before or during secretion into the supernatant (Fig. 6.3A). Since acidification of the medium by *L. lactis* may impact cell growth and/or protein secretion, we also tested the effect of adding sodium hydroxide or sodium phosphates to GM17 media. There was no improvement in growth rate and a slight decrease in secreted SCI-57 (Fig. 6.3A).
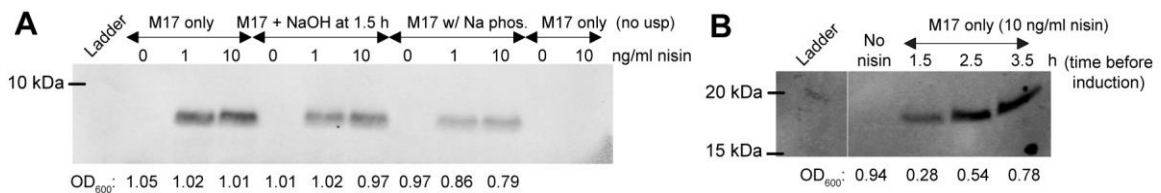
Figure 6.3. Western blot detection of SCI-57 and leptin in *L. lactis* supernatant. (A) Cultures of NZ9000(pNZPnisA:usp-SCI-His or pNZPnisA:SCI-His) were grown in GM17Cm media with or without 50 mM sodium phosphates for 2.5 h, and then induced with 1 or 10 ng/ml nisin for an additional 4 h. For modulating pH with NaOH, 10% culture volume of 5 N NaOH was added at 1.5 h after induction. Anti-RGS-His antibody was used to detect SCI-57 secreted into the supernatant. The molecular weight of RGS-His-tagged SCI-57, after cleavage of usp45 signal peptide, is 7.5 kDa. (B) NZ9000(pNZPnisA:usp-His-Lep) cultures were induced with 10 ng/ml nisin at 1.5, 2.5 or 3.5 h after dilution of overnight cultures and grown for a total of 7.5 h. Supernatants were concentrated 20-fold. The molecular weight of RGS-His-tagged leptin is 17 kDa.

Similarly, *L. lactis* NZ9000(pNZPnisA:usp-His-Lep) was grown and induced with 10 ng/ml nisin at 1.5, 2.5 or 3.5 h to investigate the effect of induction time on protein expression. Western blotting of concentrated supernatant via the N-terminal RGS-His tag showed bands between 15 and 20 kDa in nisin-induced cultures, suggesting the successful expression of RGS-His-tagged leptin (after cleavage of usp45 signal peptide) with an expected size of 17 kDa (Fig. 6.3B). Delaying the induction start time from 1.5 h

to 3.5 h resulted in an increase in culture density, which is in agreement with Fig. 6.2. The increase in leptin secretion, however, is modest, possibly due to the fact that the increased growth rate is offset by the reduced amount of time for protein synthesis before the culture reaches stationary phase.

6.3.3 Bioactivity of secreted SCI-57 on adipocytes

To investigate if secreted SCI-57 is properly folded and biologically active, its ability to functionally signal was tested on differentiated 3T3-L1 adipocytes. Insulin signals by binding and activating cell-surface insulin receptors, which phosphorylates adapter proteins such as the insulin receptor substrate (IRS) family, which then recruit and activate downstream effector molecules. One such effector protein that is required for insulin signaling is phosphatidylinositol 3-kinase (PI 3-kinase), which phosphorylates Akt[38,39]. In our assay, we added conditioned *L. lactis* medium to differentiated 3T3-L1 adipocytes and used p-Akt as a metric of insulin signaling.

Serum-starved, fully differentiated 3T3-L1 adipocytes were incubated with 1 nM commercial insulin solution or 100 µl 20-fold-concentrated supernatant for 15 min at 37°C and p-Akt was detected by Western blotting of cell lysates. Both commercial insulin and supernatant from induced NZ9000(pNZPnisA:usp-SCI-His) treated with NaOH showed a strong p-Akt signal (Fig. 6.4A,B). Importantly, no signals could be seen in supernatants from uninduced NZ9000(pNZPnisA:usp-SCI-His) treated with NaOH or from induced cells without buffering or with sodium phosphate buffering, even though

these latter cultures clearly secrete the full-length polypeptide (Fig. 6.3A). These results indicate that functional folding of secreted SCI-57 is an important bottleneck in obtaining bioactive product from *L. lactis*, but this bottleneck is dependent on the pH and buffer conditions in the medium. To estimate the concentration of active RGS-His-tagged SCI-57 in the supernatant, we quantified p-Akt signals from the Western blots of three independent experiments and found the signal of SCI-57 to be roughly equivalent to 1 nM commercial insulin (Fig. 6.4B). Since 100 μl of 20-fold-concentrated supernatant was added to 2 ml DMEM, the original concentration of functional SCI-57 secreted by NZ9000(pNZPnisA:uspSCI-57his) was approximately 1 nM.
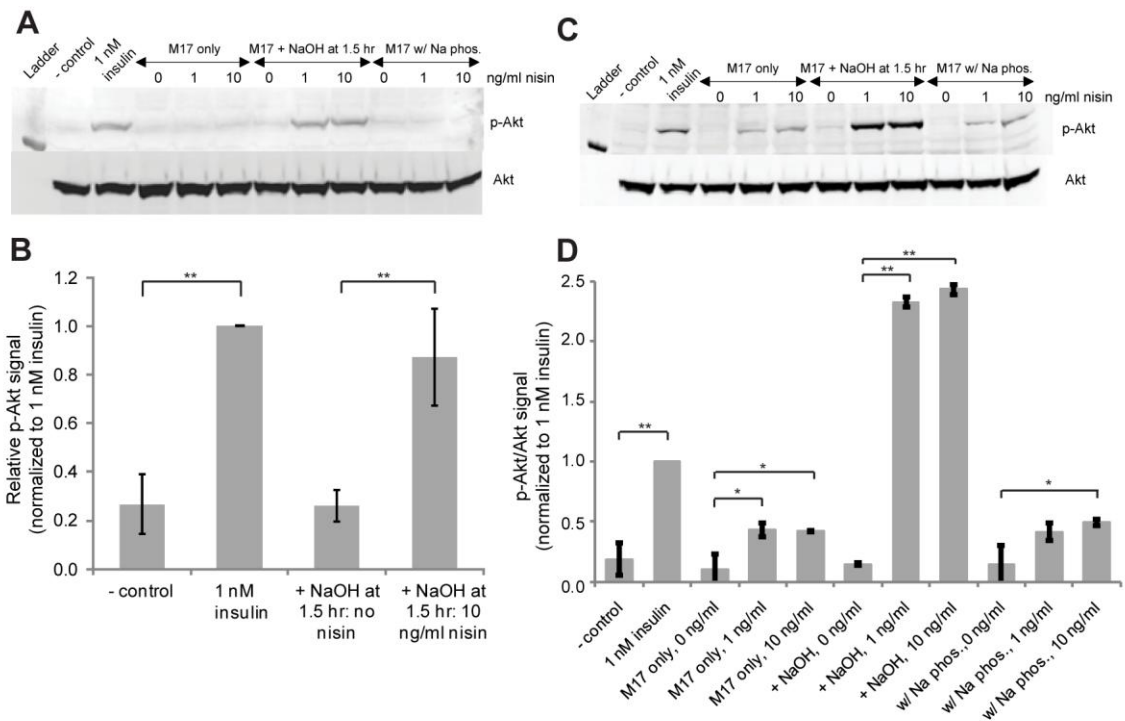
Figure 6.4. *In vitro* biological acitivity of secreted SCI-57. (A) Representative blot of p-Akt signaling. Supernatant of NZ9000(pNZPnisA:usp-SCI-His) cultures were prepared as described, and concentrated supernatant (100 μl) was added to 2 ml DMEM on fully differentiated, serum-starved 3T3-L1 adipocytes.

Cell lysates were blotted for p-Akt(Ser473). (B) Quantification of p-Akt signaling from three independent experiments. Relative intensities are shown by normalizing to the positive control (1 nM commercial insulin). $^{**}P < 0.01$. (C) Biological activity of secreted SCI-57 without the RGS-His tag. Representative Western blot showing relative p-Akt signals on differentiated 3T3-L1 adipocytes stimulated by supernatant from NZ9000 [pNZPnisA:usp-SCI (no RGS-His tag)] cultures. (D) Quantification of p-Akt signals from two independent experiments. Notation for the last nine samples is: medium composition, nisin concentration. $^{*}P < 0.05$ and $^{**}P < 0.01$.

In addition, we investigated whether the addition of the RGS-His tag, which is convenient for detection and assay development but not desirable for eventual *in vivo* applications, has any effect on the bioactivity of SCI-57. The p-Akt signaling experiments on 3T3-L1 adipocytes were repeated with conditioned media from NZ9000(pNZPnisA:usp-SCI) cultures. Interestingly, not only was p-Akt signaling much stronger for cultures treated with NaOH but it was now also detectable using untreated cultures or cultures with phosphate buffer (Fig. 6.4C,D). This suggests that the RGS-His tag adversely affects insulin receptor binding and/or SCI-57 folding.

6.3.4 Bioactivity of secreted leptin and SCI-57-leptin fusions

Leptin binds to the long-form leptin receptor (OBRb) on the cell surface, which in turn results in activation of the JAK/STAT signaling pathway, leading to phosphorylation and activation of downstream STAT3[40]. In our assay, HEK293 cells were co-transfected with leptin receptor (OBRb), STAT3-activated firefly luciferase (FLuc), and constitutively expressed renilla luciferase (RLuc). HEK293 cells were incubated with 1 nM or 10 nM commercial leptin or 50 µl 20-fold-concentrated supernatant for 24 h at 37°C. The

specific activity of leptin can be quantified by dividing the FLuc signal by the RLuc signal to normalize for transfection efficiency. Supernatants from the two induced NZ9000(pNZPnisA:usp-His-Lep and pNZPnisA:usp-Lep) cultures showed at least a 3-fold increase in FLuc/RLuc signal compared to uninduced controls, indicating that secreted leptin exhibited biological activity (Fig. 6.5A). Again, the RGS-His tag seems to adversely affect leptin folding and/or receptor binding. Based on the controls with 0.1 nM and 1 nM commercial leptin, the concentrations of RGS-His-tagged leptin and untagged leptin secreted by *L. lactis* were approximately 0.2 nM and 2 nM, respectively.
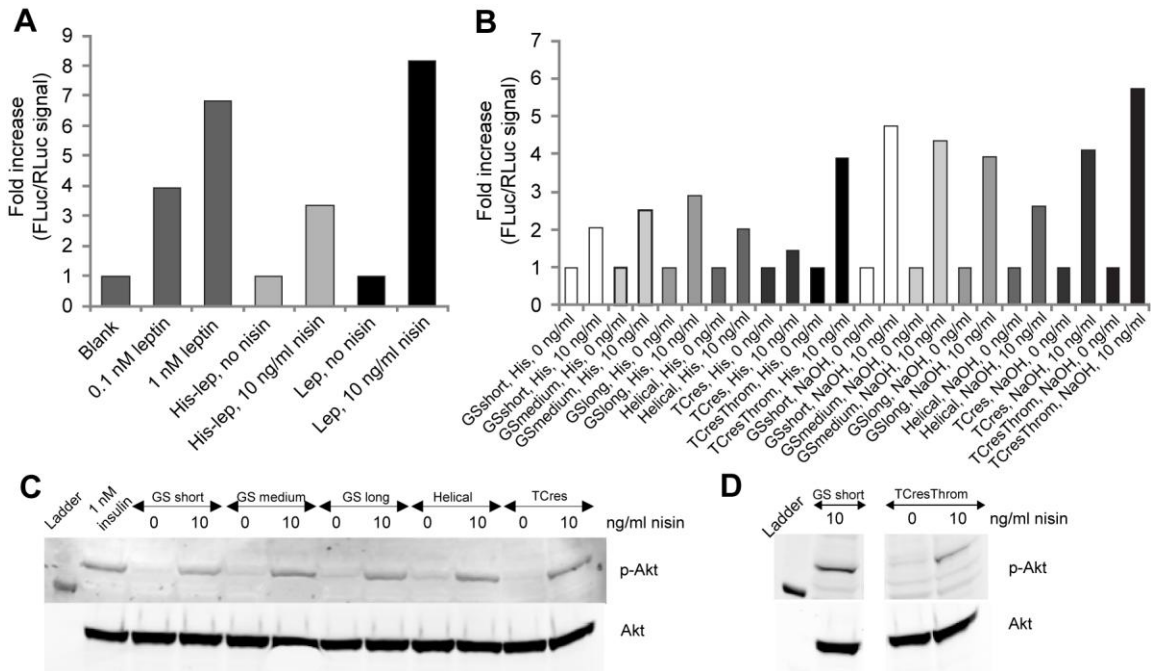
Figure 6.5. *In vitro* biological activity of secreted leptin and SCI-57-leptin fusions. (A) Commercial leptin or 50 µl 20-fold concentrated supernatant of NZ9000(pNZPnisA:usp-His-Lep or pNZPnisA:usp-Lep) cultures induced at 2.5 h for an additional 4 h were added to 2 ml DMEM on transfected HEK293 cells. Leptin bioactivity was determined by its downstream STAT3 transcriptional activity of firefly luciferase (FLuc). Results were normalized to renilla luciferase (RLuc) signal to account for differences in transfection efficiency. (B) Leptin bioactivity of secreted SCI-57-leptin fusion proteins was determined by

adding 50 μl unbuffered concentrated supernatant of NZ9000(pNZPnisA:usp-His-SCI-(linker)-Lep) or NaOH buffered concentrated supernatant of NZ9000(pNZPnisA:usp-SCI-(linker)-Lep) to transfected HEK293 cells. Notation for the samples is: linker type, RGSHis-tagged proteins or NaOH buffered culture, nisin concentration. (C, D) Insulin bioactivity of secreted SCI-57-leptin fusion proteins was determined by adding 100 μl NaOH buffered concentrated supernatant of NZ9000(pNZPnisA:usp-SCI-(linker)-Lep) to stimulate differentiated 3T3-L1 adipocytes.

Next, we created 6 different SCI-57-leptin fusion proteins to identify a suitable amino acid linker for the covalent attachment of SCI-57 and leptin. This peptide linkage is preferable to chemical conjugation because the entire fusion protein can then be synthesized by *L. lactis*. The C-terminus of SCI-57 was fused to the N-terminus of leptin because leptin contains two cysteine residues that form an essential disulfide bond between the middle of the protein and the C-terminus[41]. It has been shown that C-terminal fusions of leptin adversely affect production and secretion of the protein[42], but C-terminal fusions of SCI-57 in ribosome display format showed no apparent adverse effects in receptor binding (Chapter 5). We tested 3 glycine-serine flexible linkers GGSG (GS short), GAGGSGGS (GS medium), and GGGSGGAGGGSG (GS long), as well as an α-helical linker AEAAAKEAAAKA[43] (helical), a flexible linker with resistance to intestinal enzymes trypsin and chymotrypsin GSPDGDIDGS[44] (TCres), and a thrombin-cleavable linker with trypsin/chymotrypsin resistance GSLVPRGSGS[44,45] (TCresThrom), for their abilities in retaining bioactivity of the two fusion partners, SCI-57 and leptin. Leptin bioactivity from NZ9000(pNZPnisA:usp-His-SCI-(linker)-Lep or pNZPnisA:usp-SCI-(linker)-Lep) culture supernatants was assayed using transfected HEK293 cells as described above. RGS-His-tagged SCI-57-leptin fusions gave between 2.0- and 3.9-fold

increases over uninduced controls, while untagged fusions with NaOH buffering gave slightly higher signals of between 2.6 and 5.7-fold increase (Fig. 6.5B). The leptin in the SCI-57-leptin fusion proteins was clearly functional, and its biological activity did not seem to be negatively affected by NaOH buffering (10% culture volume of 5 N NaOH was added 1.5 h after induction to cultures secreting untagged fusion proteins). Similarly, insulin bioactivity from NaOH-treated NZ9000(pNZPnisA:usp-SCI-(linker)-Lep) cultures was assayed on differentiated 3T3-L1 adipocytes (Fig. 6.5C, D), and all 6 SCI-57-leptin fusion proteins showed a specific and similar p-Akt signal over uninduced control. Based on these *in vitro* bioassays, all 6 linkers were able to retain similar functional activity of both SCI-57 and leptin. Compared to the positive controls using commercial insulin and leptin, the concentrations of secreted untagged SCI-57-leptin fusion proteins were about 1 nM.

6.3.5 Effect of pH modulation on functional SCI-57 expression

The previous results (Figs. 6.3 and 6.4) indicate that the biological activity of SCI-57 is more heavily affected by the medium pH and buffering conditions than by the overall protein secretion level. We therefore investigated the insulin biological activity of various conditioned *L. lactis* media, each subject to a different pH profile, on 3T3-L1 adipocytes. Two modes of pH modulation were tested: (1) continuous control, in which 2% culture volume of 5 N NaOH was added hourly for 5 h; or (2) a single pulse, in which 10% culture volume of 5 N NaOH was added once at the indicated time after induction. In effect, the same total amount of NaOH was added in each case. As seen from the OD_{600}

readings (Fig. 6.6A), early addition of NaOH (continuous or single dose at 0.5 h) has an inhibitory effect on growth. Corresponding to the lower OD_{600} readings, the amount of secreted SCI-57 detected for culture with early NaOH addition was also lower on a per-volume basis. However, when the biological activity of these samples was tested using our p-Akt signaling assay on 3T3-L1 adipocytes (Fig. 6.6B, C), the observed trend was the opposite of that seen in Fig. 6.6A. Cultures grown with early NaOH addition secrete similar, if not higher, levels of functional SCI-57 than cultures grown with late NaOH addition. This further confirms that the biological activity of secreted SCI-57 is more heavily affected by buffering conditions than total secretion level and suggests that earlier counterbalancing of the natural medium acidification increases the fraction of functional SCI-57 molecules.
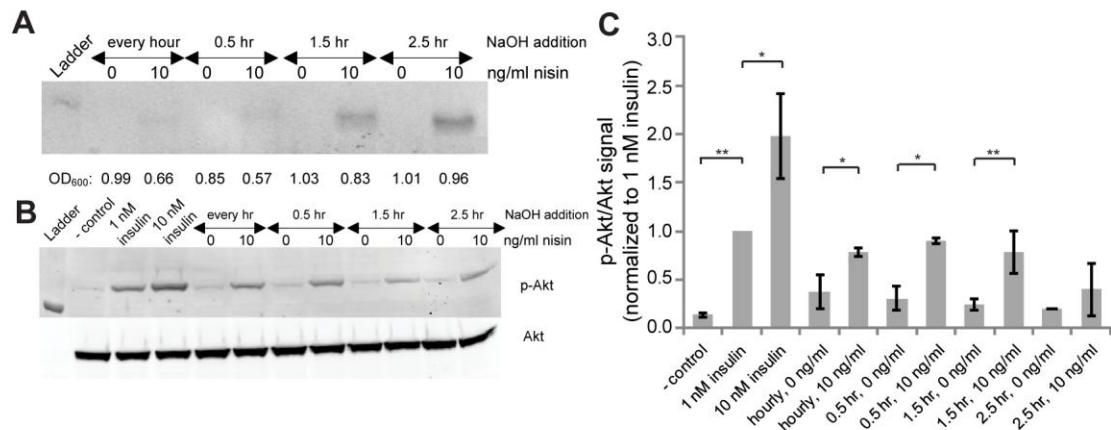
Figure 6.6. Effect of temporal pH modulation on SCI-57 secretion and bioactivity. (A) Western blot showing secreted SCI-57 with different NaOH treatments, with final OD_{600} readings given below. Cultures of NZ9000(pNZPnisA:usp-SCI-His) were induced at 2.5 h with 0 or 10 ng/ml nisin for an additional 4 h. For continuous pH regulation with NaOH, 2% culture volume of 5N NaOH was added at 1.5, 2.5, 3.5, 4.5, and 5.5 h after the 1:25 dilution. For single-dose modulation with NaOH, 10% culture volume of 5 N NaOH was added at 0.5, 1.5, or 2.5 h after nisin induction. RGS-His antibody was used to detect the RGS-

His-tagged SCI-57 secreted into the supernatant. (B) Representative Western blot showing relative p-Akt signals on differentiated 3T3-L1 adipocytes stimulated by concentrated supernatant in different buffering conditions. (C) Quantification of p-Akt signals, as shown in (B), from two independent experiments. Notation for the last eight samples is: NaOH addition time after nisin induction, nisin concentration. $^*P < 0.05$ and $^{**}P < 0.01$.

6.3.6 Oral delivery of SCI-57-secreting *L. lactis* to diabetic mice

Finally, a pilot study was conducted to test the performance of SCI-57-secreting *L. lactis* upon oral delivery to diabetic mice. Since a constitutive expression system might offer advantages *in vivo*, we evaluated *L. lactis* strains with both inducible and constitutive expression of SCI-57; specifically, the nisin-inducible promoter P*nisA* was compared to the constitutive promoter P*thyA*. P*thyA* was chosen because it had been successfully used to secrete interleukin-10 *in vivo*[8]. In addition, P*thyA* showed the highest level of SCI-57 secretion when compared to other strong lactococcal constitutive promoters P1[46], P23[47], and P59[47] by Western blotting (Fig. 6.7A). P*thyA* was able to secrete full-length untagged SCI-57, though at a slightly lower level compared to P*nisA*, (Fig. 6.7B).

Five streptozotocin (STZ)-induced diabetic mice were used to study the effect of using *L. lactis* as an oral delivery platform for the delivery of SCI-57. Both inducible (P*nisA*) and constitutive (P*thyA*) strains of *L. lactis* were used; in the inducible case, 10 ng/ml nisin was used for induction 1 h prior to *in vivo* administration. The cultures were pelleted by centrifugation and resuspended in buffered minimal media (BGM9), and oral gavage was used for oral delivery of *L. lactis*. There were no noticeable differences in activity level before and after oral gavage, suggesting that administration of *L. lactis* was

safe to the animals. Diabetic mice that received NZ9000(pNZPnisA:usp-SCI) had statistically significantly reductions in blood glucose levels after 3 h compared to both negative controls (no treatment or oral gavage with 1 U insulin solution) (Fig. 6.7C). Diabetic mice that received NZ9000(pNZPthyA:usp-SCI) showed a slight reduction in blood glucose level compared to the no-treatment control, but no significant difference compared to the control with 1 U insulin solution. After 24 h, blood glucose levels in both groups that received *L. lactis* treatment returned to near-normal levels, which is not unexpected since most of the bacteria should be excreted by this time[5].
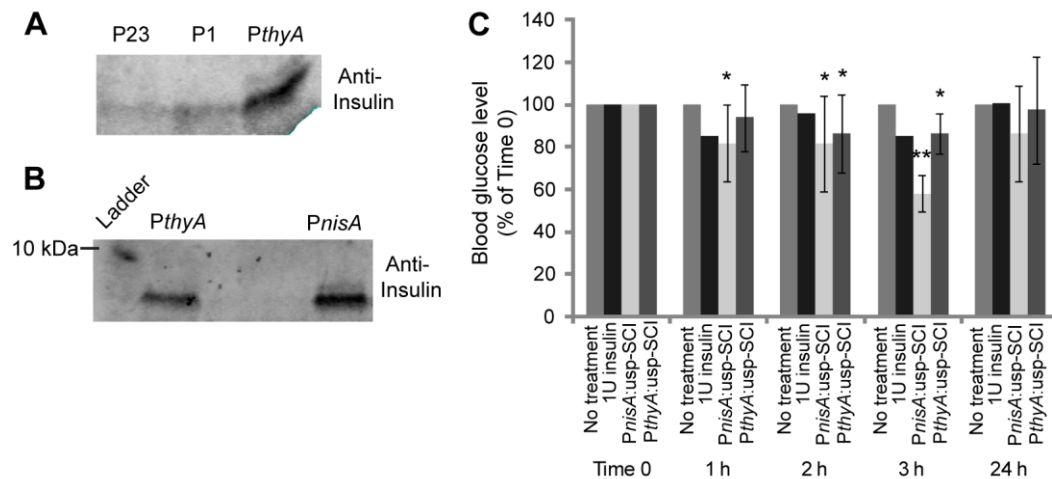
Figure 6.7. *In vitro* expression and *in vivo* delivery of SCI-57-secreting *L. lactis* to diabetic mice. (A) Western blot detection of SCI-57 in supernatant. Three constitutively expression strains, NZ9000(pNZP23:usp-SCI), NZ9000(pNZP1:usp-SCI), and NZ9000(pNZPthyA:usp-SCI) were grown for 3.5 h in GM17Cm after overnight dilutions. The supernatants were blotted with anti-insulin antibody to detect the level of secreted SCI-57. (B) Comparison of level of secreted SCI-57 between NZ9000(pNZPthyA:usp-SCI) and an inducible expression strain, NZ9000(pNZPnisA:usp-SCI). NZ9000(pNZPnisA:usp-SCI) was induced at 1.5 h with 10 ng/ml nisin. (C) Blood glucose levels of STZ-induced diabetic mice at different times after no treatment, oral gavage of 1 U insulin, or oral gavage of 10^{10} to 10^{11} CFU of NZ9000(pNZPnisA:usp-SCI), NZ9000(pNZPthyA:usp-SCI). Results represent averages from 1 to 4 independent experiments with 1 or 2 mice in each group. $^{*}P < 0.05$ and $^{**}P < 0.01$.

6.4 Discussion

As shown in Fig. 6.2, both nisin concentration and time of induction affect the growth rate. For a given induction time, a higher nisin concentration increases the doubling time (*i.e.*, slows the growth rate). This is possibly due to a combination of increased burden from producing heterologous protein and nisin toxicity (nisin can inhibit bacterial cell wall biosynthesis and can kill cells by forming pores in the cytoplasmic membrane[48]). For a given nisin concentration, a later induction time reduces the doubling time (*i.e.*, increases the growth rate). There is perhaps a critical OD_{600} threshold in early- to mid-log phase above which the burdens imposed by heterologous protein synthesis and nisin toxicity are offset by the faster growth rate in this phase.

The metabolic burden from heterologous protein synthesis can be observed more directly in the slower growth rate (Fig. 6.3). Furthermore, Western blot signals of SCI-57 in unconcentrated supernatant (Fig. 6.3A) are comparable to those of leptin in 20-fold concentrated supernatant (Fig. 6.3B), while signals of SCI-57-leptin fusion proteins in concentrated supernatants fell below the detection limit of Western blotting (though the bioassays clearly show that these proteins are present (Fig. 6.5B–D)). These observations reveal a further decrease in protein secretion by *L. lactis*, in addition to reduced growth rate, as the size of the heterologous produced protein is increased, although this may be specific to the proteins studies here. Quantification of secreted untagged SCI-57, leptin, and SCI-57-leptin fusions in *L. lactis* cultures showed around 10 nM, 2 nM, and 1 nM of

bioactive insulin, leptin, and insulin/leptin equivalents, respectively (Fig. 6.4 and 6.5). This effect is most likely due to reduced growth rate and expression, as described above, because there is no evidence that increasing protein size adversely affects the folding of protein since the usp45 signal peptide exports proteins unfolded via the secretion (Sec) pathway[6].

While leptin bioactivity was not sensitive to culture pH (Fig. 6.5B), the correlation between pH regulation and SCI-57 bioactivity is summarized in Table 6.3. *L. lactis* naturally acidifies the GM17 medium to pH < 5.7 when the culture is saturated. Regulating the pH with a base or buffering the media allows *L. lactis* to grow to a much higher density[49] and it has also been shown to increase the stability and biological activity of secreted interleukin-10[10,50]. In our study, buffering the medium may increase the solubility, stability, and/or folding of secreted SCI-57. SCI-57 has a pI similar to native insulin and is most soluble at neutral pH[13]. Also, wild-type insulin is most stable near neutral pH, due to deamidation at low pH and aggregation at high pH[51]. However, since the pH during our induction experiments stays relatively close to neutral pH, deamidation and aggregation are unlikely to have large effects. Furthermore, SCI-57 was designed with amino acid substitutions in the A and B chains to prevent dimerization and higher order assembly[13]. Thus, we hypothesize that the marked increase in biological activity upon addition of NaOH is primarily due to better folding. SCI-57, like native insulin, requires three disulfide bonds (A6–A11, A7–B7, A20–B19) for correct folding and biological activity. Previous work indicated that basic pH facilitates folding of insulin by deprotonating thiolate moieties and thus limiting aggregation of reduced B chains, giving

216

rise to more free B chains that can form productive disulfide bonds with A chains[52,53]. Also, thiol-disulfide exchange, the principal mechanism by which disulfide bonds are formed and rearranged in proteins, proceeds via a nucleophilic attack of the thiolate anion and alkaline conditions facilitate this reaction[54]. In our case, the addition of NaOH could provide a transient alkaline environment to more efficiently facilitate formation of the disulfide bonds of SCI-57, thus giving rise to higher biological activity. During the folding of insulin, there exists a critical folding intermediate containing the single disulfide A20–B19, which has been suggested to form first and guide subsequent folding[55]. The posited initial formation of this disulfide bond may also explain why SCI-57 without the RGS-His tag exhibits higher biological activity even without pH modulation. The RGS-His tag in our experiments is fused directly to the C-terminus of SCI- 57 (i.e., after A21), which might impede the formation of this initial disulfide bond.

Table 6.3. pH values during different induction conditions.

	Starting pH	Ending pH	pH spike	Insulin activity (with RGS-His tag)[a]	Insulin activity (no tag) [a]
M17 only	6.9	5.7	No	–	+
M17 with 50 mM sodium phosphates	7.2	6.5	No	–	+
M17 with NaOH addition every hour	6.9	7–7.7	No	++	N.D.
M17 with NaOH addition at 0.5 h	6.9	7–9.17	> 9	++	N.D.
M17 with NaOH addition at 1.5 h	6.9	6.9–7.1	> 7.7	++	+++
M17 with NaOH addition at 2.5 h	6.9	7.2–7.7	> 7.7	+	N.D.

[a] –: no activity; +: equivalent to < 1 nM commercial insulin; ++: equivalent to ≈ 1 nM commercial insulin; +++: equivalent to > 1 nM commercial insulin; N.D.: not determined.

One notable difference between the *in vitro* experiments and the actual *in vivo* environment is the pH-sensitive bioactivity. While pH decreases in *in vitro* experiments significantly hindered the bioactivity (e.g., without intervention with NaOH), the gut is highly pH-regulated and will not be affected by the acidification that occurs in unbuffered M17 medium *in vitro*. Furthermore, *L. lactis* resides in the jejunum and ileum of human intestine[7], which have pH values of ~6.2 and ~6.8–8.4, respectively[56]. These pH values were closely approximated in the experiments in which we observed functional SCI-57. In addition, SCI-57 administered *in vivo* will not have an RGS-His tag, elimination of which should substantially increase the biological activity of the protein.

For STZ-induced diabetic mice, subcutaneous injection of 0.5 to 1 U/kg insulin is usually sufficient to lower blood glucose to an acceptable level[57]. The same level of insulin delivered orally did not lower blood glucose (Fig. 6.7C), and the most likely reason is that the acidic pH and intestinal enzymes inactivated or degraded the drug[58]. The amount of untagged SCI-57 delivered by 10^{10} to 10^{11} CFU *L. lactis* was estimated to be the same order of magnitude, yet preliminary *in vivo* experiments with these doses of pre-induced *L. lactis* elicited a ~20−40% reduction in blood glucose within 2−3 hours of oral administration (Fig. 6.7C), suggesting an advantage in using *L. lactis* to deliver protein therapeutics rather than administrating the proteins directly. However, the trial-to-trial variations in these experiments are great, as well as the lack of proper negative and positive controls. Ideally, the mice should be fasted before and during experiments. An insulin-tolerance test should be carried out before the actual testing to make sure that the

mice respond to injected insulin. An appropriate negative control would be the oral gavage of uninduced bacteria. In addition, the insulin level in blood should be measured to ensure the reduction in blood glucose is due to insulin in the systemic circulation. Therefore, further optimization and *in vivo* studies are needed to draw more rigorous conclusions.

The modest reduction in blood glucose levels also points to the need for improving the amount of insulin delivered by *L. lactis*. Two complementary strategies are: 1) to increase the number of protein molecules delivered by a given number of cells and 2) to increase the fraction of secreted bioactive proteins that reach the systemic circulation (*i.e.*, improving the bioavailability). Since protein synthesis in our application is maximized by strong constitutive promoters (P*thyA*) and inducible promoters (P*nisA*), another way to increase the amount of protein delivered would be to enhance the secretion efficiency from *L. lactis*. We have investigated this possibility and our findings are presented in the next chapter (Chapter 7). Additionally, bioavailability can be improved by increasing absorption of delivered proteins, as proteins are often too large to readily cross the intestinal barrier[59]. To enhance absorption, a fusion partner that is able to cross the intestinal epithelial layer through receptor-mediated transcytosis[24] can be used to conjugate the therapeutic protein of interest. In this work, leptin was chosen as one such potential partner, and SCI-57-leptin fusions were shown to retain bioactivity of both components *in vitro*. Another, perhaps more promising, fusion partner with better understood mechanisms for transcytosis is the Fc region of immunoglobulin (IgG), which binds to FcRn receptors present on the intestinal epithelium to cross the epithelial barrier

intact[60]. IgG Fc domains that were engineered to be monomeric[61], and thus amenable to facile synthesis in *L. lactis*, were studied in Chapter 4. This remains an exciting area of research for furthering *L. lactis* as a platform for oral delivery of therapeutic proteins.

In summary, our study demonstrates that food-grade microorganisms can be engineered to secrete bioactive therapeutic proteins. We have constructed a *L. lactis* strain that can efficiently secrete SCI-57, leptin, and SCI-57-leptin fusion proteins that are biologically active at the physiological pH in the gut. From a biomedical perspective, the advantages of this system for oral delivery are threefold. First, *L. lactis* has long been shown to have a safe association with humans, and thus possible adverse effects (e.g., using synthetic particles) can be minimized. Second, the protein can by secreted locally in the small intestine, minimizing loss during passage through the upper digestive tract and potentially enabling 'on demand' secretion. Third, as a live delivery vehicle, the pharmacokinetics can be genetically tuned to match delivery requirements. From a biotechnological perspective, using SCI-57- or leptin-secreting *L. lactis* eliminates the needs for laborious synthesis, expensive protein purification, and temperature-sensitive storage of these protein therapeutics, offering cheaper and more convenient alternatives to traditional insulin and leptin replacement therapies.

6.5 References

1.   Morishita, M. & Peppas, N. A. Is the oral route possible for peptide and protein drug delivery? *Drug Discov. Today* **11**, 905–910 (2006).

2.   Muller, G. Oral Delivery of Protein Drugs: Driver for Personalized Medicine? *Curr. Issues Mol. Biol.* **13**, 13–24 (2011).

3.   Khafagy, E.-S., Morishita, M., Onuki, Y. & Takayama, K. Current challenges in non-invasive insulin delivery systems: A comparative review. *Adv. Drug Deliv. Rev.* **59**, 1521–1546 (2007).

4.   Balcázar, J. L. *et al.* Changes in intestinal microbiota and humoral immune response following probiotic administration in brown trout (*Salmo trutta*). *Br. J. Nutr.* **97**, 522–527 (2007).

5.   Klijn, N., Weerkamp, A. H. & de Vos, W. M. Genetic marking of *Lactococcus lactis* shows its survival in the human gastrointestinal tract. *Appl. Environ. Microbiol.* **61** , 2771–2774 (1995).

6.   Nouaille, S. *et al.* Heterologous protein production and delivery systems for *Lactococcus lactis*. *Genet. Mol. Res.* **2**, 102–111 (2003).

7.   Wells, J. M. & Mercenier, A. Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nat. Rev. Microbiol.* **6**, 349–362 (2008).

8.   Steidler, L. *et al.* Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nat. Biotechnol.* **21**, 785–789 (2003).

9.   Braat, H. *et al.* A Phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. *Clin. Gastroenterol. Hepatol.* **4**, 754–759 (2006).

10.   Steidler, L. *et al.* Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. *Science* **289** , 1352–1355 (2000).

11.   Gordon Still, J. Development of oral insulin: progress and current status. *Diabetes Metab. Res. Rev.* **18**, S29–S37 (2002).

12.   Hua, Q. Insulin: a small protein with a long journey. *Protein Cell* **1**, 537–551 (2010).

13.   Hua, Q. *et al.* Design of an active ultrastable single-chain insulin analog: synthesis, structure, and therapeutic implications. *J. Biol. Chem.* **283**, 14703–14716 (2008).

14.   Rajpal, G., Liu, M., Zhang, Y. & Arvan, P. Single-chain insulins as receptor agonists. *Mol. Endocrinol.* **23** , 679–688 (2009).

15.   Auwerx, J. & Staels, B. Leptin. *The Lancet* **351**, 737–742 (1998).

16.	Kim, Y. B., Uotani, S., Pierroz, D. D., Flier, J. S. & Kahn, B. B. *In vivo* administration of leptin activates signal transduction directly in insulin-sensitive tissues: overlapping but distinct pathways from insulin. *Endocrinology* **141**, 2328–2339 (2000).

17.	Dardeno, T. A. *et al.* Leptin in human physiology and therapeutics. *Front. Neuroendocrinol.* **31**, 377–393 (2010).

18.	Oral, E. A. *et al.* Leptin-replacement therapy for lipodystrophy. *N. Engl. J. Med.* **346**, 570–578 (2002).

19.	Welt, C. K. *et al*. Recombinant human leptin in women with hypothalamic amenorrhea. *N. Engl. J. Med.* **351**, 987-997 (2004).

20.	Naito, M. *et al.* Therapeutic impact of leptin on diabetes, diabetic complications, and longevity in insulin-deficient diabetic mice. *Diabetes* **60**, 2265-2273 (2011).

21.	Wang, M. *et al.* Leptin therapy in insulin-deficient type I diabetes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4813–4819 (2010).

22.	Cammisotto, P. G. *et al.* Receptor-mediated transcytosis of leptin through human intestinal cells *in vitro*. *Int. J. Cell Biol.* **2010**, 928169 (2010).

23.	Cammisotto, P. G., Gingras, D. & Bendayan, M. Transcytosis of gastric leptin through the rat duodenal mucosa. *Am. J. Physiol. Gastrointest. Liver Physiol.* **293**, G773–G779

24.	Swaan, P. Recent advances in intestinal macromolecular drug delivery via receptor-mediated transport pathways. *Pharm. Res.* **15**, 826–834 (1998).

25.	Mierau, I. & Kleerebezem, M. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. *Appl. Microbiol. Biotechnol.* **68**, 705–717 (2005).

26.	de Ruyter, P. G., Kuipers, O. P. & de Vos, W. M. Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl. Environ. Microbiol.* **62**, 3662–3667 (1996).

27.	Kuipers, O. P., de Ruyter, P. G. G. A., Kleerebezem, M. & de Vos, W. M. Quorum sensing-controlled gene expression in lactic acid bacteria. *J. Biotechnol.* **64**, 15–21 (1998).

28. de Ruyter, P. G., Kuipers, O. P., Beerthuyzen, M. M., van Alen-Boerrigter, I. & de Vos, W. M. Functional analysis of promoters in the nisin gene cluster of *Lactococcus lactis*. *J. Bacteriol.* **178** , 3434–3439 (1996).

29. Bermúdez-Humarán, L. G. *et al.* Controlled intra- or extracellular production of staphylococcal nuclease and ovine omega interferon in *Lactococcus lactis*. *FEMS Microbiol. Let.* **224**, 307–313 (2003).

30. van Asseldonk, M. *et al.* Cloning of usp45, a gene encoding a secreted protein from *Lactococcus lactis* subsp. *lactis* MG1363. *Gene* **95**, 155–160 (1990).

31. Leenhouts, K. *et al.* A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol. Gen. Genet.* **253**, 217–224 (1996).

32. Bahey-El-Din, M., Griffin, B. & Gahan, C. Nisin inducible production of listeriolysin O in *Lactococcus lactis* NZ9000. *Microb. Cell Fact.* **7**, 24 (2008).

33. Pope, B. & Kent, H. M. High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Res.* **24** , 536–537 (1996).

34. Nakashima, N., Sharma, P. M., Imamura, T., Bookstein, R. & Olefsky, J. M. The tumor suppressor PTEN negatively regulates insulin signaling in 3T3-L1 adipocytes. *J. Biol. Chem.* **275** , 12889–12895 (2000).

35. Bermúdez-Humarán, L. G. *et al.* Effects of intranasal administration of a leptin-secreting *Lactococcus lactis* recombinant on food intake, body weight, and immune response of mice. *Appl. Environ. Microbiol.* **73**, 5300–5307 (2007).

36. Le Loir, Y. *et al.* Signal peptide and propeptide optimization for heterologous protein secretion in *Lactococcus lactis*. *Appl. Environ. Microbiol.* **67**, 4119–4127 (2001).

37. van Asseldonk, M., de Vos, W. M. & Simons, G. Functional analysis of the *Lactococcus lactis* usp45 secretion signal in the secretion of a homologous proteinase and a heterologous α-amylase. *Mol. Gen. Genet.* **240**, 428–434 (1993).

38. Jiang, Z. Y. *et al.* Insulin signaling through Akt/protein kinase B analyzed by small interfering RNA-mediated gene silencing. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7569–7574 (2003).

39. Summers, S. A., Garza, L. A., Zhou, H. & Birnbaum, M. J. Regulation of insulin-stimulated glucose transporter GLUT4 translocation and Akt kinase activity by ceramide. *Mol. Cell. Biol.* **18**, 5457–5464 (1998).

40. Bates, S. H. *et al.* STAT3 signalling is required for leptin regulation of energy balance but not reproduction. *Nature* **421**, 856–859 (2003).

41. Zhang, F. *et al.* Crystal structure of the obese protein leptin-E100. *Nature* **387**, 206–209 (1997).

42. Lo, K.-M. *et al.* Engineering a pharmacologically superior form of leptin for the treatment of obesity. *Protein Eng. Des. Sel.* **18**, 1–10 (2005).

43. Arai, R., Ueda, H., Kitayama, a, Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* **14**, 529–32 (2001).

44. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook* **SE - 52***, 571–607 (2005).

45. Jenny, R. J., Mann, K. G. & Lundblad, R. L. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif.* **31**, 1–11 (2003).

46. Waterfield, N. R., Le Page, R. W., Wilson, P. W. & Wells, J. M. The isolation of lactococcal promoters and their use in investigating bacterial luciferase synthesis in *Lactococcus lactis*. *Gene* **165**, 9–15 (1995).

47. Morello, E. *et al. Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* **14**, 48–58 (2008).

48. McAuliffe, O., Ross, R. P. & Hill, C. Lantibiotics: structure, biosynthesis and mode of action. *FEMS Microbiol. Rev.* **25**, 285–308 (2001).

49. Tremillon, N. *et al.* Production and purification of staphylococcal nuclease in *Lactococcus lactis* using a new expression-secretion system and a pH-regulated mini-reactor. *Microb. Cell Fact.* **9**, 37 (2010).

50. Schotte, L., Steidler, L., Vandekerckhove, J. & Remaut, E. Secretion of biologically active murine interleukin-10 by *Lactococcus lactis*. *Enzyme and Microb. Technol.* **27**, 761–765 (2000).

51. Brange, J. & Langkjaer, L. Chemical stability of insulin. 3. Influence of excipients, formulation, and pH. *Acta Pharm. Nord.* **4**, 149–158 (1992).

52. Hua, Q. *et al.* Mechanism of insulin chain combination: asymmetric roles of a-chain α-helices in disulfide pairing. *J. Biol. Chem.* **277**, 43443–43453 (2002).

53. Weiss, M. A. Proinsulin and the genetics of diabetes mellitus. *J. Biol. Chem.* **284**, 19159–19163 (2009).

54. Rudolph, R. & Lilie, H. *In vitro* folding of inclusion body proteins. *FASEB J.* **10**, 49–56 (1996).

55. Yan, H., Guo, Z.-Y., Gong, X.-W., Xi, D. & Feng, Y.-M. A peptide model of insulin folding intermediate with one disulfide. *Protein Sci.* **12**, 768–775 (2003).

56. Charman, W. N., Porter, C. J. H., Mithani, S. & Dressman, J. B. Physicochemical and physiological mechanisms for the effects of food on drug absorption: The role of lipids and pH. *J. Pharm. Sci.* **86**, 269–282 (1997).

57. Cho, H. *et al.* Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB β). *Science* **292**, 1728–1731 (2001).

58. Carino, G. & Mathiowitz, E. Oral insulin delivery. *Adv. Drug Deliv. Rev.* **35**, 249–257 (1999).

59. Lee, V. H. L. & Yamamoto, A. Penetration and enzymatic barriers to peptide and protein absorption. *Adv. Drug Deliv. Rev.* **4**, 171–207 (1990).

60. Dickinson, B. L. *et al.* Bidirectional FcRn-dependent IgG transport in a polarized human intestinal epithelial cell line. *J. Clin. Invest.* **104**, 903–911 (1999).

61. Ying, T., Chen, W., Gong, R., Feng, Y. & Dimitrov, D. S. Soluble monomeric IgG1 Fc. *J. Biol. Chem.* **287**, 19399–19408 (2012).

CHAPTER 7

Engineering signal peptides for enhanced protein secretion from *Lactococcus lactis*

7.1 Introduction

Lactococcus lactis, traditionally used in the food fermentation industry for the production of cheese and buttermilk[1], has great potential as an oral delivery vehicle for the clinical delivery of therapeutic molecules. This bacterium has generally regarded as safe (GRAS) status in humans and has been used as a mucosal delivery vehicle because it can survive passage through the stomach acid and contact with bile[2,3]. *L. lactis* has been engineered to express and secrete a variety of therapeutic proteins, including interleukin-10 for treatment of inflammatory bowel disease[4], and bovine β-lactoglobulin for immunity against this cow milk allergen[5]. In the previous chapter, we demonstrated successful secretion of a single-chain insulin analog, leptin, and their fusions from this bacterium. However, although preliminary *in vivo* studies on oral delivery of insulin-secreting *L. lactis* to diabetic mice showed encouraging results, the level of glucose reduction was relatively modest and showed variability among the limited number of trials performed. The amount of insulin secreted by a feasible number of ingested bacteria was estimated to be approximately the same as which would be needed in a subcutaneous injection to lower blood glucose to an acceptable level, which would thus be insufficient

226

for oral delivery given the additional bottlenecks in delivery. Therefore, it is extremely important to enhance the secretion capability of *L. lactis*. Even in other biomedical applications in which a reasonable therapeutic dose can be achieved, maximizing secretion of the recombinant protein of interest would be desirable since the dose and/or administration frequency of *L. lactis* can be reduced.

In addition to clinical delivery of therapeutic proteins, there also has been increasing interest over the past two decades in the use of this bacterium for the biotechnological production of heterologous proteins. A major advantage of using *L. lactis* or other Gram-positive bacteria rather than Gram-negative bacteria (e.g., *Escherichia coli*) for protein production is that proteins can be easily secreted into the medium, which streamlines the often expensive and inefficient downstream purification process[6]. In contrast to many commonly used Gram-positive bacteria (e.g., *Bacillus* species), laboratory strains of *L. lactis* possess only one exported housekeeping protease, HtrA[7], and secrete only one major extracellular protein, Usp45[8], thus minimizing protein degradation by the extracellular proteolytic system and further simplifying downstream purification. Given its simple and well-known metabolism, its completely sequenced genome[9,10], and the availability of various genetic tools[11], *L. lactis* is an attractive host for many biotechnological applications. Enhancing the secretion capabilities of this host can reduce cost and maximize the production of heterologous proteins.

In *L. lactis*, most proteins are secreted unfolded via the secretion (Sec) pathway[12]. Proteins are synthesized as precursors containing the mature moiety of the protein with an N-terminal signal peptide. The signal peptide plays an important role in targeting the

protein to the cytoplasmic membrane, where the protein precursor is subsequently translocated by the Sec machinery[13]. Following cleavage of the signal peptide, the mature protein is released extracellularly[14]. The lactococcal signal peptides follow the common tripartite structure, including a positively charged N terminus (n-region), a central hydrophobic core (h-region), and a more polar C terminus (c-region) containing the signal peptide cleavage site[15]. The most widely used signal peptide for *L. lactis* secretion is that from Usp45 (Usp45sp), which has the best secretion efficiency[16]. Other reported lactococcal signal peptides, both natural (e.g., SP310[17], SPExp4[18], and AL9[19]) and engineered (e.g., SP310mut2[20]), have secretion efficiencies only as good as, and often worse than, that of Usp45sp[6,20,21]. Although improvements in secretion have been achieved by either introducing a synthetic propeptide[22] or complementing the lactococcal translocation machinery with SecDF[23], the inherent limitations of Usp45sp itself have not been addressed. Thus, optimization of this widely used signal peptide should not only directly lead to enhancements in the secretion efficiencies of a wide range of proteins but may also be combined with other improvements, such as those mentioned above, to fully exploit the secretion potential of *L. lactis*.

As the signal peptide of the only major secretory protein in *L. lactis*, it is unknown whether Usp45sp has been subjected to stringent evolutionary pressures that have maximized its secretory efficiency. Here, we screened libraries of silent and targeted mutations in Usp45sp and demonstrated that the secretion efficiency of this signal peptide can indeed be improved. We first showed that silent mutations of Usp45sp alone can give an increase of up to 16% in the activity of secreted α-amylase. We then used a simple

mathematical model to provide a quantitative explanation for this observed effect based on the free energy of the mRNA secondary structure around the ribosome binding site (RBS) and its consequent effect on protein translation. We further introduced nonsilent mutations in Usp45sp using a targeted mutagenesis approach and achieved an increase of up to ~50% in the activity of secreted α-amylase at the maximum induction level. Using the best clone from the targeted mutation library, we introduced silent mutations into this sequence and identified clones that also increase secretion levels at lower concentrations of inducer. Analyzing these silent mutation clones with our mathematical model, we show that the effect of optimizing the amino acid sequence is more dominant than that of relaxing mRNA folding in enhancing secretion levels. In summary, we have used two complementary directed evolution approaches to achieve progressive improvements in *L. lactis* secretion efficiencies solely by engineering the signal peptide; this work represents the first successful effort in achieving secretion yields higher than that attainable with Usp45sp and also provides general strategies for enhancing protein secretion in bacterial hosts.

7.2 Materials and methods

7.2.1 Bacterial strains, plasmids, and growth conditions

The strains and. *E. coli* and *L. lactis* used, and their respective growth conditions, were described in Section 6.2.1. Solid media were prepared by adding agar (15 g/liter) to the corresponding broths. For the nuclease plate assay, brain heart infusion (BHI) agar

(Acumedia, Lansing, MI) was used to grow *L. lactis*. When needed, chloramphenicol (Cm) was used at a final concentration of 10 μg/ml. Plasmids used in this work are listed in Table 7.1

Table 7.1. Plasmids used in this study.

	Characteristics	Sources
pNZPnisA:usp-SCI-His	Modified pNZ8048 containing P*nisA* promoter with Usp45sp fused with downstream RGS-His-tagged *SCI-57* gene; Cm^r	Chapter 6
pNZ8048m:usp45sp-MazF	Modified pNZ8048 containing P*nisA* promoter with Usp45sp fused with downsream RGS-His-tagged *E. coli MazF* gene; Cm^r	This work
pNZ8048m:usp45sp-Nuc	Modified pNZ8048 containing P*nisA* promoter with Usp45sp fused with downstream RGS-His-tagged *S. aureus* nuclease gene; Cm^r	This work
pNZ8048m:usp45sp-AmyE	Modified pNZ8048 containing P*nisA* promoter with Usp45sp fused with downstream RGS-His-tagged *B. subtilis* α-amylase gene (with internal EcoRI silent mutation); Cm^r	This work
pNZ8048m:SP310-Nuc	Modified pNZ8048 containing P*nisA* promoter with SP310 signal peptide fused with downstream RGS-His-tagged *S. aureus* nuclease gene; Cm^r	This work
pNZ8048m:SP310-AmyE	Modified pNZ8048 containing P*nisA* promoter with SP310 signal peptide fused with downstream RGS-His-tagged *B. subtilis* α-amylase gene (with internal EcoRI silent mutation); Cm^r	This work

7.2.2 DNA manipulations and plasmid construction

Isolation of plasmid DNA from *E. coli* and *L. lactis* was performed as described in Section 6.2.2. Phusion high-fidelity DNA polymerase, restriction enzymes, and T4 DNA ligase were purchased from NEB (Ipswich, MA) and were used as recommended by the manufacturer. Primers used for DNA amplification (Integrated DNA Technologies, Coralville, IA) are listed in Table 7.2. pNZ8048m:usp45sp-MazF was constructed by isolating the *mazF* gene from E. coli K-12 (GenBank accession no. NC000913) by

colony PCR with primers 1 and 2, digesting with BspEI and BamHI, and ligating into similarly cut pNZPnisA:usp-SCI-His (Usp45sp from GenBank accession no. M60178 but with two silent mutations to introduce a BspEI site, as described in 6.2.2). To incorporate an NcoI site at the junction between the coding sequences for Usp45sp and *mazF*, pNZ8048m:usp45sp-MazF was PCR amplified with primers 3 and 2, digested with BspEI and BamHI, and ligated into similarly cut pNZ8048m:usp45sp-SCI57his, giving rise to pNZ8048m:usp45sp-MazF. pNZ8048m:usp45sp-Nuc (Fig. 7.1A) was constructed by isolating the nuclease gene from *Staphylococcus aureus* (GenBank accession no. AM990992; courtesy of Dr. Mark Goulian, University of Pennsylvania) by colony PCR with primers 4 and 5, digesting with NcoI and BamHI, and ligating into similarly cut pNZ8048m:usp45sp-MazF. pNZ8048m:usp45sp-AmyE was constructed using the same method as described above after isolating *Bacillus subtilis* α-amylase (GenBank accession no. FJ643607; courtesy of Dr. Mark Goulian) with primers 6 and 7, except that an internal EcoRI site was eliminated by a silent mutation to facilitate subsequent library construction. This was done by PCR amplification of the α-amylase with primers 6 and 9 and primers 8 and 7, fusion of the two fragments by assembly PCR, digestion of the full-length product with NcoI and BamHI, and ligation into NcoI- and BamHI-digested pNZ8048m:usp45sp-MazF. To replace Usp45sp with the SP310 signal peptide, SP310 was constructed by extension PCR with primers 10 and 11 followed by digestion with EcoRI and NcoI and ligation into similarly cut pNZ8048m:usp45sp-Nuc and pNZ8048m:usp45sp-AmyE to yield pNZ8048m:SP310-Nuc and pNZ8048m:SP310-AmyE, respectively. All ligation mixtures were transformed into chemically competent

231

E. coli EC1000, sequenced, and electroporated into electrocompetent *L. lactis* NZ9000 as described previously in Section 6.2.2.
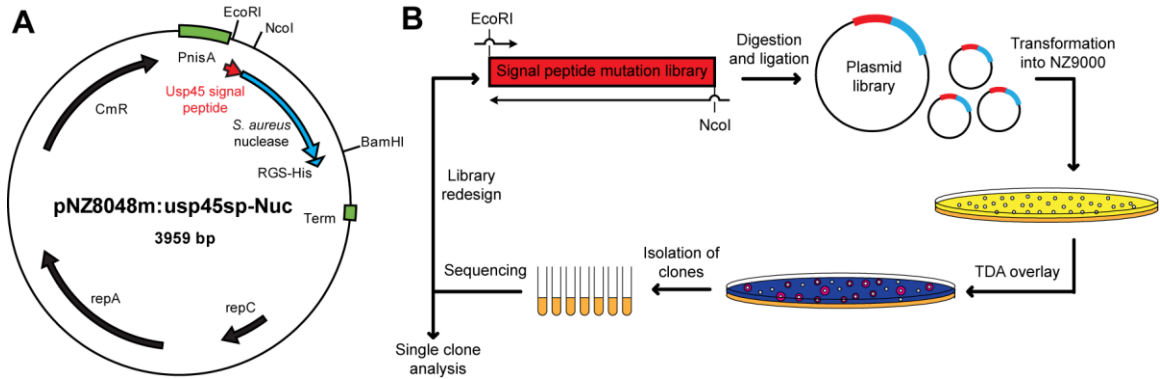
Figure 7.1. Schematics of the expression constructs and screening process. (A) Plasmid map of pNZ8048m:usp45sp-Nuc, showing promoter P*nisA*, usp45 signal peptide sequence, *S. aureus* nuclease gene, and RGS-His tag in the modified pNZ8048 backbone. (B) Flow diagram of the mutagenesis and screening process. The signal peptide mutations were encoded by degenerate PCR primers. Nucleotide libraries were constructed, digested, and ligated into the pNZ8048m vector backbone via the EcoRI and NcoI sites. After plasmid transformation into *L. lactis* NZ9000, the resulting library was plated and screened by carefully overlaying a toluidine blue-DNA agar (TDA) mixture with inducer (10 ng/ml nisin) onto the agar plates. Since the size of the pink halo around each colony corresponds to the amount of nuclease secreted, clones could be quickly screened by eye and subjected to further analyses (sequencing, amylase secretion assays, Western blotting, etc.). Subsequent libraries were designed based on data collected in the initial screen.

Table 7.2. Primers used in this study.

Primer Number	Primer Name	Nucleotide Sequence (5' → 3') [a]
1	MazF_BspEI_f	CTTACGTCCGGAGTTTACGCTATGGTAAGCCGATACGTAC
2	MazF_BamHI_r	TACGTTGGATCCTCTCCCAATCAGTACGTTAATTTTG
3	Usp45NcoIMazF-BspEI_f	TTCAGCTCCGGAGTTTACGCCATGGTAAGCC
4	SaurNuc_NcoI_f	TTTACGCCATGGTTTCACAAACAGATAATGGCGTAAATAG

5	SaurNuc_BamHI_r	ATGGTG<u>GGATCC</u>TCTTTGACCTGAATCAGCGTTGTCTTC
6	BsubAmy_NcoI_f	TTTACG<u>CCATGG</u>TTCTTACAGCACCGTCGATCAAAAG
7	BsubAmy_BamHI_r	ATGGTG<u>GGATCC</u>TCTATGGGGAAGAGAACCGCTTAAGC
8	BsubAmy_EcoRImut_f	CACGCAGAACTCATTGCTCG
9	BsubAmy_EcoRImut_r	CGAGCAATGAGTTCTGCGTG
10	SP310_EcoRI_f	CTCAAA<u>GAATTC</u>ATGAAATTTAATAAAAAAAGAGTTGCAATAG
11	SP310_NcoI_r	TTGTCG<u>CCATGG</u>CATTAGTTTGGTTATCTTGGATTGATG
12	pNZ_EcoRI_f	TTATAAGGAGGCACTCAAA<u>GAATTC</u>ATG
13	NucB_NcoI_r	CATTATCTGTTTGTGAAA<u>CCATGG</u>
14	Usp45SM_NcoI_r	CACTTT<u>CCATGG</u>CRTANACNCCNGAYAANGGNGCNGCNGCNGAN AGDATNACNGTNGACATYAADATNGCNGADATDATYTTYTTYTT CAT<u>GAATTC</u>TTTGAGTGCCTCCTTATAA
15	Usp45TM_NcoI_r	ATCGGCTTA<u>CCATGG</u>CMNNAACMNNGGAMNNCGGGGCTGCMNNM NNAAGMNNCACMNNMNNMNNTAAMNNAGCMNNMNNMNNCTTTTT TTTCAT<u>GAATTC</u>TTTGAGTGCCTCCTTATAA
16	TM8SM_NcoI_r [b]	GTGAAA<u>CCATGG</u>CYTTNACRTTNGADATNGGNGCNGCNGTNGTN AGCATNACNACNACNGC**N**AARTGNGCYTTNAGNACYTTYTTYTT CAT<u>GAATTC</u>TTTGAGTGCCTCCTTATAA

[a] Restriction sites are underlined.

[b] The bolded N should be a Y to encode for only silent mutations of leucine. Nevertheless, all clones selected for characterization in this work contained only silent mutations at this position.

7.2.3 Library construction and selection procedures

The selection procedure is outlined in Fig. 7.1B. To construct the mutation libraries for Usp45sp silent mutations, Usp45sp targeted mutations, and Usp45TM8 silent mutations, three separate extension PCRs were performed with forward primer 12 and degenerate reverse primers 14, 15, and 16, respectively. The constructs were digested with EcoRI and NcoI and ligated with a 1:1 molar ratio into > 1 µg similarly digested pNZ8048m:SP310-Nuc. The ligation products were electroporated directly into *L. lactis* NZ9000 and plated onto 10 150-mm-diameter petri dishes with BHI agar containing Cm. Colonies were allowed to grow overnight at 30°C.

Toluidine blue-DNA agar (TDA) was made with 36 mg toluidine blue O (Sigma-Aldrich, St. Louis, MO), 4 g NaCl, 4 g agar, 0.12 g salmon DNA (EMD Chemicals, Billerica, MA), and 0.4 ml of 0.01 M $CaCl_2$ dissolved in 400 ml of 0.05 M Tris (pH 9)[24]. Molten TDA (13 ml) containing 10 ng/ml nisin (Sigma-Aldrich) was carefully overlaid on each BHI agar plate and allowed to solidify. After incubation at 30°C for 1 h, colonies were evaluated by the size of their pink halos. Single colonies were picked, inoculated into 5 ml GM17Cm, and grown overnight at 30°C. Plasmids in selected clones were isolated and analyzed by DNA sequencing. To clone the selected signal peptide upstream of α-amylase, signal peptides were PCR amplified with primers 12 and 13, digested with EcoRI and NcoI, and ligated into similarly digested pNZ8048m:usp45sp-AmyE.

7.2.4 Nuclease plate assay

To evaluate the secretion efficiency and activity of nuclease secreted by individual clones, 10 ml molten TDA was spread into each 100-mm-diameter petri dish and allowed to solidify. Wells were created in the TDA with P1000 pipet tips (USA Scientific, Ocala, FL). Overnight cultures of *L. lactis* NZ9000 containing pNZ8048m:usp45sp-Nuc (or similar constructs with other signal peptides) were diluted 1:25 into fresh GM17Cm. Cultures were grown for 1.5 h and induced with the appropriate amount of nisin for another 2 h. The optical density at 600 nm (OD_{600}) was measured on a plate reader (Infinite M200; Tecan, Männedorf, Switzerland) to ensure that different cultures had similar growth rates. Cells were removed from the supernatant by a 5-min centrifugation

at 5,000*g*, 1 µl supernatant was added to each TDA well, and the plates were incubated at 30°C for 1 h. The size of the pink halo around each clone was measured.

7.2.5 α-Amylase activity assay and Western blotting

To evaluate the secretion efficiency and activity of α-amylase secreted by individual clones, starch azure assays were carried out as previously described[16] but with modifications. Cultures of *L. lactis* were grown, induced with nisin, and centrifuged as mentioned above. The supernatant (100 µl) was incubated with 20 mg starch azure (Sigma-Aldrich) and 900 µl α-amylase buffer (50 mM Tris-HCl [pH 7.5], 50 mM NaCl, 5 mM CaCl$_2$) at 37°C for 1 h with shaking. Centrifugation was carried out at 15,000 x *g* for 5 min, and supernatant absorbance at 595 nm was measured on a plate reader. α-Amylase activity was calculated from a standard curve (Fig. 7.2). The standard curve was obtained by incubating 0.0038 to 3.8 units of purified *B. subtilis* α-amylase (Sigma-Aldrich) with starch azure under the conditions described above.

$$y = 0.2041x^{0.3770}$$
$$R^2 = 0.9972$$

235
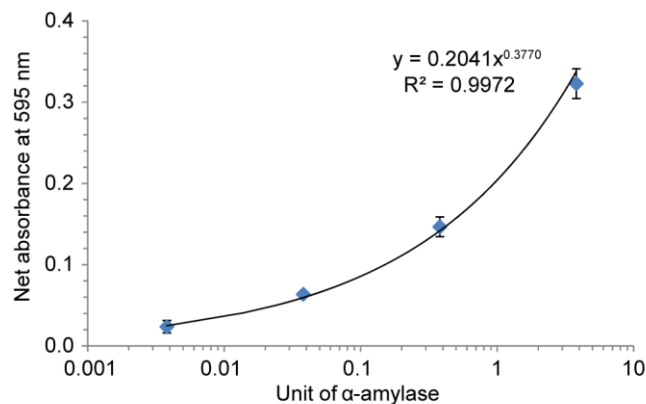
Figure 7.2. Standard curve of α-amylase starch azure test. *B. subtilis* α-amylase (100 µl of 0.0001 mg/ml to 0.1 mg/ml solution) was mixed with 20 mg starch azure, and then 900 µl α-amylase buffer was added. Samples were shaken horizontally at 37°C for 1 h, and absorbance of the supernatant was measured at 595 nm. The data represent the mean ± standard errors of the means (SEM) of two independent experiments.

For Western blotting, the supernatant was passed through a 0.22-µm-pore-size filter (Millipore, Billerica, MA) to remove any cells, and 13 µl supernatant (adjusted with phosphate-buffered saline as needed to achieve equivalent cell densities) was mixed with 2 µl 500 mM dithiothreitol and 5 µl lithium dodecyl sulfate sample buffer for analysis by SDS- PAGE in a 4–12% NuPAGE Bis-Tris gel (Invitrogen, Carlsbad, CA). Proteins were transferred and blotted as described previously in Chapter 6.

7.2.6 mRNA structure analysis and mathematical modeling

The mRNA sequence from the transcription start site to 6 nucleotides after the end of the signal peptide-coding region (135 bases total) was analyzed using the mfold web server[25]. The structure with the lowest (most negative) folding energy was used to calculate the free energy (ΔG) of the ribosome binding site (RBS). ΔG_{RBS} was calculated by adding up the ΔG contributions of the 6 bases in AGGAGG and an initiation ΔG of 3.4 kcal/mol[26].

We developed a mathematical model to quantitatively explain the differences in secretion due to silent mutations in the signal peptide. The model relates the stability of mRNA secondary structure around the ribosome binding site to translation and secretion

efficiencies based on a previous model with modifications[27] (Fig. 7.3). Nomenclature and parameter values are provided in Table 7.3.
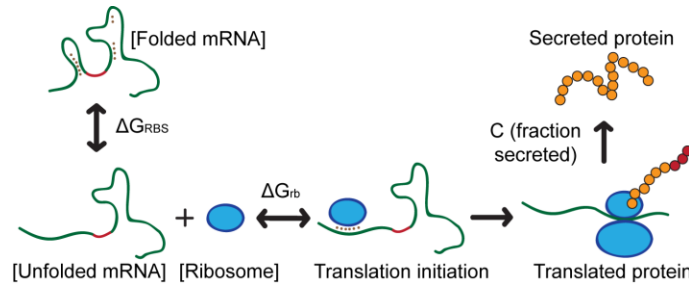
Figure 7.3. Schematic of the mathematical model. The model predicts that translational efficiency correlates with ribosome binding to accessible RBSs, which depends on the fraction of RBSs in the unfolded state, which, in turn, is determined by the ΔG of the RBS for a given mRNA sequence. Secretion efficiency is determined as a fraction of the total amount of translated protein and is specific to a particular amino acid sequence.

Table 7.3. Nomenclature and parameter values used in the mathematical model.

Symbol	Description	Value	Reference
F	mRNA with secondary structure around the RBS (folded)	Calculated from *fsolve*	N/A
U	mRNA with exposed RBS (unfolded)	Calculated from *fsolve*	N/A
Rb	Free ribosomes	Calculated from *fsolve*	N/A
Rb•U	Ribosomes bound to unfolded mRNA	Calculated from *fsolve*	N/A
K_f	Equilibrium association constant for mRNA folding around the RBS	Calculated from ΔG_{RBS}	Reference[25,26]
ΔG_{RBS}	Free energy of mRNA secondary structure around the RBS	Varies from -8.5 to -0.5 kcal/mol (calculated from mfold with initiation term)	Reference[25,26]
K_{rb}	Equilibrium association constant for binding of 30S ribosomal subunit to unfolded mRNA	Calculated from ΔG_{rb}	Reference[27]
ΔG_{rb}	Free energy of binding of 30S ribosomal	-14 kcal/mol	Reference[27]

	subunit to unfolded mRNA		
Rb_{tot}	Total concentration of ribosomes in the cell	40 μM (assumed to be equivalent to *E. coli* with the same doubling time)	Reference[28,29]
RNA_{tot}	Total concentration of mRNA transcripts (depending of induction level)	0.08 Rb_{tot}, 0.8 Rb_{tot}, and 8 Rb_{tot} for 0.1, 1, and 10 ng/ml nisin, respectively	This work
C	Peptide-specific proportionality constant relating protein expression to secretion efficiency	Usp45SM: 1.01; Usp45TM8_SM: 1.45	This work
R	Gas constant	1.986 cal/K mol	N/A
T	Temperature (in Kelvin)	303.15 K (*L. lactis* was grown at 30°C)	N/A

First, we modeled translation as a process initiated by a ribosome recognizing the ribosome binding site and unfolding of the initiation region in the mRNA[30,31]. The folding/unfolding equilibrium of mRNA is based specifically on the free energy of mRNA secondary structure around the ribosome binding site (RBS) (ΔG_{RBS}).

$$U \xleftarrow{K_f} F$$
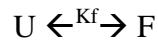
where $K_f = [F]/[U]$ and $\Delta G_{RBS} = - RT \ln(K_f)$ (1)

The 30S subunit of the ribosome binds to the unfolded mRNA with association constant K_{rb}.

$$Rb + U \xleftarrow{K_{rb}} Rb \cdot U$$
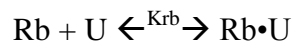
where $K_{rb} = [Rb \cdot U]/[Rb][U]$ and $\Delta G_{rb} = - RT \ln(K_{rb})$ (2)

Since initiation is the rate-limiting step in translation[32], translation was set to be proportional to the amount of ribosome-bound mRNA ($[Rb \cdot U]$). Finally, since *L. lactis*

only secretes a fraction of the expressed protein[14], secretion rate was calculated as a fraction of the total amount of protein translated:

secretion $\propto$ protein expression $\propto$ [Rb•U]

so,   secretion = C [Rb•U]/Rb$_{tot}$ (3)

where C is a constant that is unique to the amino acid sequence of a specific signal peptide.

Together with the two mass balance equations:

[U] + [F] + [Rb•U] = RNA$_{tot}$ (4)

and   [Rb] + [Rb•U] = Rb$_{tot}$,   (5)

the system of equations, (1), (2), (4) and (5), was solved in Matlab using *fsolve* to obtain the unknowns [U], [F], [Rb], and [Rb•U]. The C values for the Usp45sp and Usp45TM8 silent mutation libraries were determined experimentally from the secretion levels of wild-type Usp45sp and Usp45TM8, respectively, and Equation (3) was used to calculate the final normalized protein secretion.

7.3 Results

7.3.1 Signal peptide screening with nuclease and α-amylase reporters

To allow efficient screening of a large library of signal peptides, the *Staphylococcus aureus* nuclease was used as the reporter protein[33]. Secretion of *S. aureus* nuclease from *L. lactis* colonies can be easily detected by overlaying the library transformants with toluidine blue-DNA agar (TDA)[24]; the size of the pink halo indicates nuclease activity, which corresponds directly to secretion efficiency. For better quantification of secretion

efficiency, selected signal peptides were fused upstream of *Bacillus subtilis* α-amylase. Activity of secreted α-amylase in liquid culture can be easily quantified by incubation with starch azure[16] (Fig. 7.2). Furthermore, the use of two different heterologous proteins to evaluate secretion efficiencies can elucidate any protein-specific dependence on secretion and increase the likelihood that the best signal peptide selected would improve secretion of other useful proteins. These two proteins, *S. aureus* nuclease and *B. subtilis* α-amylase, were evaluated for their abilities to report secretion efficiencies of two known signal peptides, Usp45sp[16] and SP310[17]. A TDA plate test, carried out on *L. lactis* cultures secreting nuclease fused to either signal peptide, showed a larger pink halo with Usp45sp, indicating secretion levels qualitatively higher than those of SP310 (Fig. 7.4A). A more quantitative starch azure activity test on cultures secreting α-amylase also demonstrated that secretion with SP310 was only ~62% of that achieved with Usp45sp (Fig. 7.4B), in agreement with previous results[20]. We therefore used Usp45sp as the starting sequence for directed evolution to probe any inherent sequence limitations in using this signal peptide and to identify and alleviate bottlenecks to improve protein secretion in *L. lactis*. The signal peptide mutagenesis and screening process was performed as shown in Fig. 7.1B. To construct silent mutation libraries, we used primers with appropriate degeneracies in the third nucleotide of each codon, giving a maximum theoretical diversity of ~10^{12}. For the targeted mutation library, degenerate codons (NNK) were used to fully randomize specific amino acid residues, giving a maximum diversity of ~10^{19}. Our actual library sizes were limited by the transformation efficiency in *L. lactis* (~10^6 CFU/μg DNA). A TDA-nisin overlay allowed us to rapidly screen a large number of nuclease-secreting colonies by visualizing the sizes of the pink halos and

240

selecting individual clones for more quantitative testing with α-amylase. The maximum

subinhibitory concentration of nisin (10 ng/ml) was used for all screens[34].
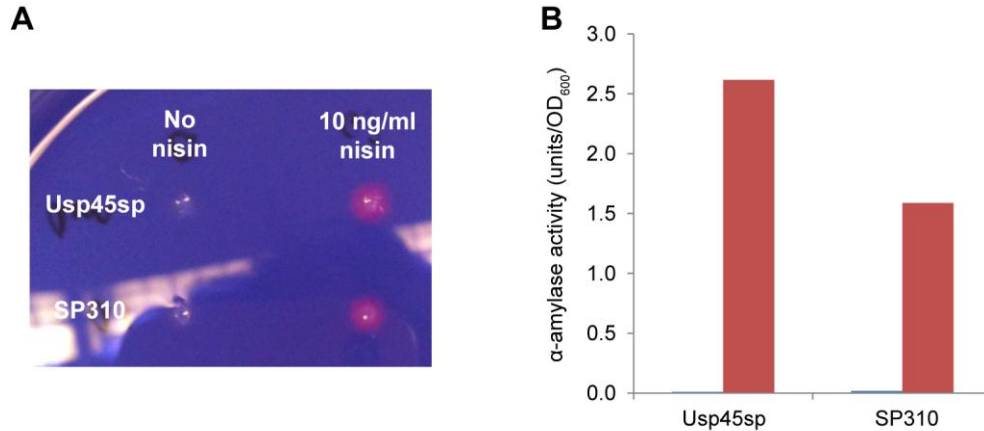
Figure 7.4. Nuclease plate test and α-amylase starch azure test of the two control signal peptides, Usp45sp and SP310. (A) The size of the pink halo on Toluidine blue-DNA-agar corresponds to the activity of secreted nuclease. (B) The starch azure test gives a more quantitative measure of α-amylase secretion (Fig. 7.2). Blue bar: no nisin induction; red bar: 10 ng/ml nisin induction. In both cases, the secretion efficiency of Usp45sp is higher than that of SP310. In the starch azure test, SP310 secretion was ~62% of Usp45sp secretion, in agreement with previous results[20].

7.3.2 Silent mutations in Usp45sp modestly increase secretion

First, a Usp45sp silent mutation library was constructed to examine if mRNA secondary structure, independent of amino acid sequence, has an effect on secretion and if secretion can be improved by relaxing mRNA structure, assuming that translation and secretion are directly correlated. It is known that silent mutations in the gene itself can alter mRNA folding and translation initiation, thus influencing gene expression[35]. Wild- type Usp45sp is predicted to have a considerable amount of mRNA folding and secondary structure, especially around the RBS (Fig. 7.5), and preliminary studies on signal peptides with

very highly folded mRNAs revealed that secretion was completely abolished. We therefore created a library to allow all possible silent mutations of Usp45sp and screened for (i) clones with representative low, medium, and high secretion levels to further probe the correlation between secondary structure and secretion rate and (ii) clones with secretion levels significantly higher than that of wild-type Usp45sp. Individual silent mutation clones were assayed for secretion of nuclease or α-amylase and were ordered by their ΔG values around the RBS (Table 7.4), which we hypothesized to be a major determinant of translation initiation and, thus, protein expression and secretion[27]. The majority of both nuclease and α-amylase secretion levels follows the ΔG trend: signal peptides with ΔG_{RBS} values that are more negative generally exhibited worse secretion than Usp45sp, and those with ΔG_{RBS} values that were less negative generally yielded better secretion than Usp45sp. To visualize putative mRNA folding, secondary structure predictions for Usp45sp, Usp45SM2, and Usp45SM3 — representing the wild-type sequence, a worse secreting sequence, and a better secreting sequence, respectively — were computed using mfold[25] (Fig. 7.5). As shown, Usp45SM2 has greater secondary structure around the RBS, which would be expected to hinder ribosome binding and lead to lower expression and secretion. On the other hand, Usp45SM3 shows a more relaxed mRNA conformation around the RBS, which would more readily allow translation initiation and, thus, result in higher secretion. Three of the best clones, Usp45SM10, Usp45SM5, and Usp45SM3, showed a small (~15%) but significant increase in secretion of α-amylase.
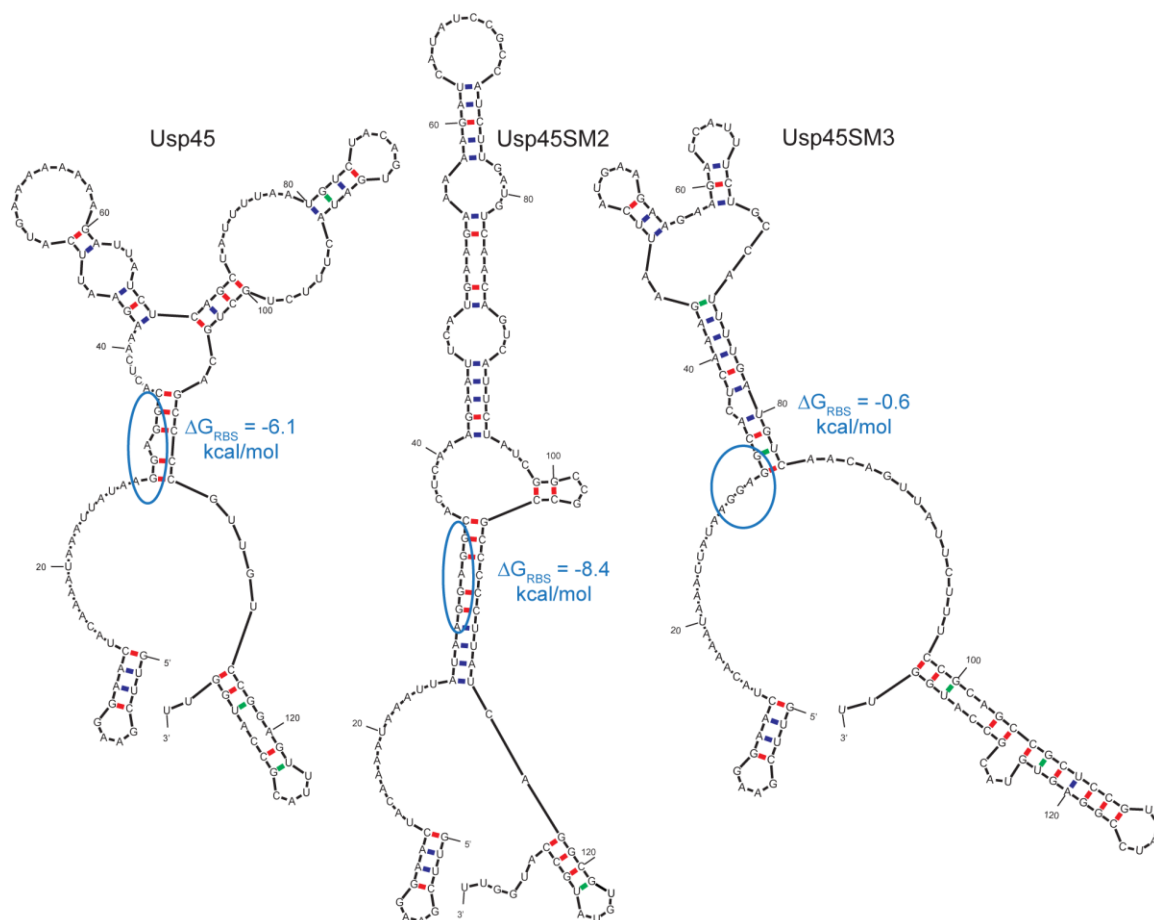
Figure 7.5. Representative mRNA structures, showing the differences in secondary structure and ΔG around the ribosome binding site (RBS) for wild-type Usp45 signal peptide, Usp45SM2, and Usp45SM3. mRNAs from the transcript start site to 6 nucleotides after the end of the signal peptide sequence were entered into mfold to obtain the secondary structures[25]. Structures with the lowest ΔG values are shown, with the RBS (AGGAGG) circled.

Table 7.4. mRNA folding energies of the AGGAGG ribosome binding site and activities of secreted nuclease and α-amylase for selected clones from the Usp45sp silent mutation library [a].

Usp45sp silent mutation clone	ΔG of ribosome binding site (AGGAGG)	Secreted nuclease activity (compared to Usp45sp) [b]	Secreted α-amylase activity (% of Usp45sp) [c]
Usp45SM2	-8.4	–	79 ± 5†

Usp45SM14	-8.4	– –	83 ± 13
Usp45SM11	-6.1	– –	85 ± 3††
Usp45sp	-6.1	*	100
Usp45SM16	-5.9	–	91 ± 12
Usp45SM19	-5.9	+	106 ± 16
Usp45SM4	-5.7	+	107 ± 6
Usp45SM10	-3.8	+	116 ± 5†
Usp45SM20	-3.7	+ +	103 ± 15
Usp45SM5	-3.7	+	113 ± 5†
Usp45SM8	-3.3	+	92 ± 5
Usp45SM15	-3.3	+	100 ± 6
Usp45SM9	-3.3	+	114 ± 7
Usp45SM7	-3.2	+	105 ± 14
Usp45SM6	-2.5	+ +	99 ± 12
Usp45SM12	-1.6	+ +	96 ± 12
Usp45SM3	-0.6	*	114 ± 3††

[a] Nuclease and α-amylase secretion levels are reported for cultures induced with 10 ng/ml nisin for2 h at mid-log phase. ΔG, mRNA folding energy. Nucleotide sequences are given in Table 7.5.

[b] Data represent levels of nuclease secretion. *, secretion similar to Usp45sp level; –, secretion slightly ($<$ 10%) lower than Usp45sp level; – –, secretion much ($>$ 10%) lower than Usp45sp level; +, secretion slightly ($<$ 10%) higher than Usp45sp level; + +, secretion much ($>$ 10%) higher than Usp45sp level.

[c] α-Amylase secretion data represent the means ± SEM of the results of three independent experiments. †, $P < 0.05$; ††, $P < 0.01$ (for statistical comparison of α-amylase activities of the respective clone and Usp45sp using one-tailed Student's t test).

Table 7.5. Nucleotide sequences of selected clones from the Usp45sp silent mutation library.

Usp45sp silent	Nucleotide sequence (5' → 3')

mutation clone

Usp45sp
```
ATGAAAAAAAAGATTATCTCAGCTATTTTAATGTCTACAGTGATACTTTCTGCTGCAG
CCCCGTTGTCCGGAGTTTACGCC
```

Usp45SM2
```
ATGAAGAAAAAGATCATATCCGCCATCTTGATGTCAACAGTCATTCTATCGGCCGCCG
CCCCCTTATCAGGCGTGTATGCC
```

Usp45SM3
```
ATGAAGAAGAAGATCATTTCTGCCATTTTGATGTCAACAGTTATTCTTTCCGCAGCCG
CTCCGTTATCCGGAGTGTACGCC
```

Usp45SM4
```
ATGAAGAAGAAAATCATTTCAGCGATATTAATGTCCACGGTCATCCTTTCTGCGGCCG
CTCCGTTGTCTGGTGTCTACGCC
```

Usp45SM5
```
ATGAAGAAGAAAATCATCTCAGCCATATTAATGTCCACAGTGATCCTCTCCGCGGCAG
CGCCATTGTCCGGCGTGTACGCC
```

Usp45SM6
```
ATGAAGAAGAAGATCATCTCCGCCATATTAATGTCCACTGTGATACTGTCAGCAGCTG
CGCCGTTGTCAGGTGTATATGCC
```

Usp45SM7
```
ATGAAAAAAAAAATCATTTCAGCCATCTTGATGTCCACAGTTATCCTCTCAGCAGCAG
CTCCCTTATCGGGAGTTTACGCC
```

Usp45SM8
```
ATGAAGAAGAAGATCATATCCGCGATCTTGATGTCGACAGTAATACTTTCGGCCGCGG
CGCCGTTATCGGGGGTATATGCC
```

Usp45SM9
```
ATGAAAAAGAAGATCATCTCTGCCATATTAATGTCCACCGTAATCCTCTCCGCCGCCG
CGCCCTTATCCGGAGTCTACGCC
```

Usp45SM10
```
ATGAAAAAGAAGATCATCTCAGCCATTTTGATGTCTACGGTTATTCTTTCTGCAGCAG
CTCCATTGTCCGGCGTATACGCC
```

Usp45SM11
```
ATGAAGAAGAAAATCATCTCCGCAATATTAATGTCCACCGTGATACTATCTGCAGCTG
CCCCGTTATCTGGAGTGTACGCC
```

Usp45SM12
```
ATGAAGAAGAAAATAATTTCCGCGATCTTAATGTCAACAGTGATCCTATCGGCAGCGG
CGCCATTATCAGGAGTGTACGCC
```

Usp45SM14
```
ATGAAAAAGAAGATCATATCCGCCATCTTGATGTCTACAGTAATCCTTTCCGCGGCCG
CCCCCTTATCAGGTGTGTATGCC
```

Usp45SM15
```
ATGAAGAAAAAGATCATCTCCGCAATTTTGATGTCGACAGTCATCCTCTCGGCTGCAG
CACCTTTGTCGGGAGTCTACGCC
```

Usp45SM16
```
ATGAAGAAGAAGATAATATCCGCCATCTTAATGTCGACCGTGATACTTTCGGCAGCTG
CTCCTTTATCGGGAGTATATGCC
```

Usp45SM19
```
ATGAAAAAAAAAATCATCTCTGCCATCTTAATGTCAACGGTTATTCTCTCCGCCGCTG
```

245

```
          CGCCTTTATCAGGGGTTTATGCC

Usp45SM20  ATGAAAAAGAAGATCATTTCGGCGATATTGATGTCCACCGTCATTCTATCCGCGGCTG
          CACCTTTGTCCGGGGTATACGCC
```

7.3.3 Mathematical model captures secretion levels based on ΔG around the RBS

To more quantitatively link predicted ΔG_{RBS} values and experimentally measured secretion levels, we developed a mathematical model based on modifications of previous work[27]. The model considers translation to be initiated once the ribosome binds to the RBS, which is assumed to occur only when the initiation region of the mRNA is unfolded[30,31] and is therefore determined by ΔG_{RBS}. In the model, translation is proportional to the amount of ribosome-bound mRNA and secretion is a fraction of the total translated protein. We plotted α-amylase secretion versus ΔG_{RBS} for three different nisin induction levels (0.1, 1, and 10 ng/ml) (Fig. 7.6A to C) and compared these experimental findings to model simulations with three analogously different mRNA transcript levels (all 10-fold apart) (Fig. 7.6D to F). Model predictions of protein secretion for all three transcript amounts show good agreement with experimental results. At all induction levels, tightly folded mRNA around the RBS yields low secretion. Secretion gradually increases as the mRNA structure around RBS is relaxed, but it eventually plateaus at high ΔG_{RBS} when either the mRNA transcripts or ribosomes become limiting. The agreement between experiment and model further supports the notion that the mRNA conformation around the RBS can explain differences in secretion among the Usp45sp silent mutation clones. The relaxation of the mRNA structure alone in this context can augment secretion by a ~15% increase at the maximum induction

level. This modest improvement is consistent with the model, since the value determined

for Usp45sp (-6.1 kcal/mol) is close to the plateau region of the curve (Fig. 7.6F).
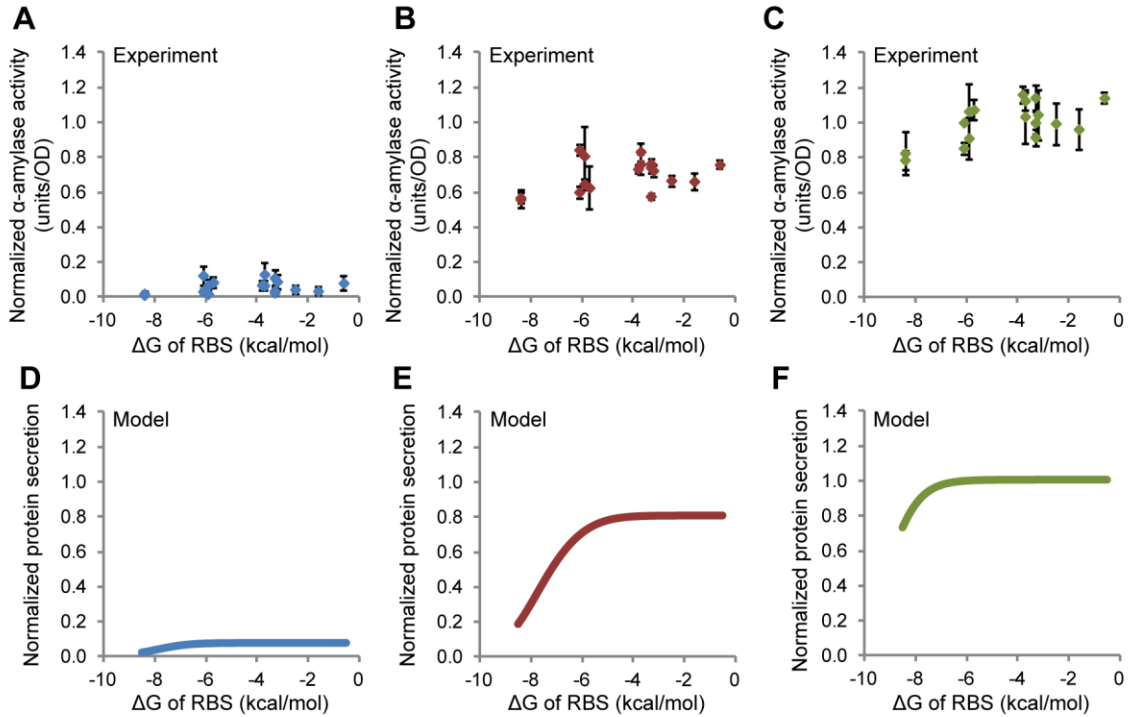
Figure 7.6. Secretion efficiencies of clones selected from the Usp45sp silent mutation library and comparison to results from a mathematical model. α-Amylase activity was plotted against the folding energy (ΔG) of the ribosome binding site (RBS) for induction at (A) 0.1 ng/ml, (B) 1 ng/ml, and (C) 10 ng/ml nisin. α-Amylase activity was normalized to the amount secreted by Usp45sp culture induced with 10 ng/ml nisin. ΔG_{RBS} was calculated for the AGGAGG RBS sequence from the mRNA structure with the lowest overall ΔG. The data represent the means ± SEM of the results of three independent experiments. Simulation results show protein secretion levels plotted against ΔG for (D) low (3.2 μM), (E) medium (32 μM), and (F) high (320 μM) mRNA transcript levels. Parameter values used in the model are given in Table 7.3.

7.3.4 Amino acid mutations in Usp45sp more significantly augment secretion of both nuclease and α-amylase

Since our experimental and modeling results suggest that the mRNA structure of Usp45sp is not a major bottleneck in achieving enhanced protein secretion, we constructed another signal peptide library consisting of targeted amino acid substitutions in Usp45sp. The length and tripartite structure of wild-type Usp45sp were kept: 27 amino acids in total, with 7, 13, and 7 residues in the n-, h-, and c-regions, respectively. Residues that were considered significant in maintaining the identities of the three regions were kept as follows: (i) charged lysines in the n-region, (ii) hydrophobic residues in the h-region, and (iii) residues important for structure and cleavage of the signal peptide in the c-region. The resulting library had 13 possible positions at which amino acid mutations could occur (Fig. 7.7A). After screening for library members with nuclease secretion, little to no conservation was observed in the 13 randomized positions (Fig. 7.7A). Five selected clones (Usp45TM8, -TM11, -TM12, -TM13, and -TM14) had increased positive charge in the n-region (from + 3 to + 4 or + 5), but only two of these clones (Usp45TM8 and -TM11) exhibited significantly higher secretion for both nuclease and α-amylase (Fig. 7.7A and B). The poorer performances of Usp45TM12 and -TM13 may be due to the fact that these sequences newly introduced a positive charge in the stretch of hydrophobic residues in the h-region. Comparing nuclease and α-amylase secretion across all clones, there was clearly a protein-specific effect on secretion for signal peptides Usp45TM13, -TM14, and-TM29 (Fig. 7.7A and B), but the two best clones from the nuclease screen, Usp45TM8 and -TM11, also exhibited enhanced secretion of α-amylase at both 1 and 10 ng/ml nisin induction (Fig. 7.7B). Western

blotting showed no differences among wild-type Usp45sp, Usp45TM8, and Usp45TM11 signal peptides in the size of the secreted α-amylase, indicating no change in the cleavage pattern when the engineered signal peptides were used (Fig. 7.7C). The best clone, Usp45TM8, gave a ~45% increase in α-amylase secretion, 30% higher than what was achieved by silent mutations alone.

A

Usp45 targeted mutation clone	Signal peptide amino acid sequence			Secreted nuclease activity
	N-region	H-region	C-region	
Usp45TM library	MKKKXXX	AXLXXXVXLXXAA	PXSXVXA	N/A
Usp45	MKKKIIS	AILMSTVILSAAA	PLSGVYA	*
Usp45TM6	MKKKLVT	AQLYGSVILLSAA	PYSCVNA	- -
Usp45TM7	MKKKSCT	AGLQLMVTLLTAA	PFSYVNA	*
Usp45TM8	MKKKVLK	AHLAVVVMLTTAA	PISNVKA	+
Usp45TM11	MKKKPRR	ALLCMLVGLALAA	PMSNVTA	++
Usp45TM12	MKKKRRM	AWLRMMVALGVAA	PPSDVLA	- -
Usp45TM13	MKKKRAS	ALLKLTVLLTTAA	PWSVVGA	*
Usp45TM14	MKKKSAR	AALMLSVNLFLAA	PASNVNA	-
Usp45TM28	MKKKPSS	AVLLCCVMLGIAA	PCSFVMA	-
Usp45TM29	MKKKPCT	ALLLWGVALAIAA	PKSFVHA	- -

Figure 7.7. Secretion efficiencies of clones selected from the targeted mutagenesis library of the Usp45 signal peptide. (A) Amino acid sequences and secreted nuclease activity for cultures induced at 10 ng/ml nisin. Amino acid mutation positions, indicated by red Xs in the Usp45TM library sequence, were encoded by degenerate NNK nucleotides. *, −, − −, +, and + + indicate secretion levels similar to and < 10% lower, > 10% lower, < 10% higher, and > 10% higher than Usp45sp levels, respectively. Blue and green residues indicate positively and negatively charged residues, respectively. Nucleotide sequences are given in Table 7.6. (B) α-Amylase activity of different clones normalized to the amount secreted by Usp45sp culture induced with 10 ng/ml nisin. Red bars represent induction at 1 ng/ml nisin, and green bars represent induction at 10 ng/ml nisin. The data represent the means ± SEM of the results of two independent experiments. (C) Western blot showing secreted α-amylase in the supernatant detected by anti-RGS-His antibody. The different targeted mutations did not alter the size of the secreted protein, indicating that the signal peptide cleavage site had been maintained.

Table 7.6. Nucleotide sequences of selected clones from the Usp45sp targeted mutation library.

Usp45sp targeted mutation clone	Nucleotide sequence (5' → 3')
Usp45sp	ATGAAAAAAAAGATTATCTCAGCTATTTTAATGTCTACAGTGATACTTTCTGCT GCAGCCCCGTTGTCCGGAGTTTACGCC
Usp45TM6	ATGAAAAAAAAGCTTGTTACTGCTCAGTTATATGGGAGTGTGATTCTTCTGTCT GCAGCCCCGTATTCCTGTGTTAATGCC
Usp45TM7	ATGAAAAAAAAGAGTTGTACGGCTGGTTTACAGTTGATGGTGACGCTTTTGACT GCAGCCCCGTTTTCCTATGTTAATGCC
Usp45TM8	ATGAAAAAAAAGGTGCTGAAGGCTCATTTAGCTGTGGTTGTGATGCTTACGACG GCAGCCCCGATTTCCAATGTTAAGGCC
Usp45TM11	ATGAAAAAAAAGCCGAGGCGGGCTTTGTTATGTATGCTTGTGGGTCTTGCTCTG GCAGCCCCGATGTCCAATGTTACGGCC
Usp45TM12	ATGAAAAAAAAGCGGCGGATGGCTTGGTTACGTATGATGGTGGCTCTTGGGGTT GCAGCCCCGCCTTCCGATGTTCTTGCC
Usp45TM13	ATGAAAAAAAAGCGTGCGTCTGCTTTGTTAAAGCTGACTGTGCTGCTTACTACG GCAGCCCCGTGGTCCGTGGTTGGGGCC
Usp45TM14	ATGAAAAAAAAGTCGGCGCGTGCTGCTTTAATGCTTTCTGTGAATCTTTTTTTG GCAGCCCCGGCGTCCAATGTTAATGCC
Usp45TM28	ATGAAAAAAAAGCCGTCGAGTGCTGTGTTATTGTGTTGTGTGATGCTTGGTATT GCAGCCCCGTGTTCCTTTGTTATGGCC
Usp45TM29	ATGAAAAAAAAGCCGTGTACTGCTCTTTTACTTTGGGGTGTGGCTCTTGCGATT GCAGCCCCGAAGTCCTTTGTTCATGCC

7.3.5 Silent mutations in Usp45TM8 further increase secretion

Finally, another round of silent mutation was performed on the best targeted mutation clone to combine the benefits of the two mutation strategies. We constructed a library with silent mutations in Usp45TM8 in order to determine if relaxing the mRNA secondary structure encoding this new signal peptide would have any consequent effect on secretion efficiency. In addition, comparison of Usp45TM8 and Usp45sp silent

mutation clones with highly folded mRNA structures could reveal if any secretion enhancement achieved by amino acid substitution could be eliminated by tighter mRNA folding. We again screened for representative low-, medium-, and high-secretion clones and analyzed them individually for α-amylase secretion. Table 7.7 shows Usp45TM8 silent mutation clones ordered by ΔG_{RBS}. All analyzed clones had an increase in secretion of at least 30% compared to wild-type Usp45sp, even though the lowest ΔG_{RBS} (Usp45TM8_SM15) was as low as those for the most poorly secreting Usp45sp silent mutation clones (Usp45SM2 and -SM14). This shows that the secretion enhancement brought about by the amino acid sequence of Usp45TM8 is more dominant than any inhibition of translation initiation due to mRNA folding (at least in this range of ΔG_{RBS}). As in the Usp45sp silent mutation library, α-amylase secretion generally increased (up to ~50%) when ΔG_{RBS} increased, which is consistent with our hypothesis that tight mRNA folding around the RBS inhibits translation initiation and subsequent secretion. We then plotted α-amylase secretion against ΔG_{RBS} for both the Usp45TM8 and Usp45sp silent mutation clones for comparison (Fig. 7.8A to C). At all induction levels, clones from the two silent mutation libraries exhibited the same trend, with low secretion at more-negative ΔG_{RBS} and higher saturated secretion at less-negative ΔG_{RBS}. Usp45TM8 silent mutation clones generally exhibited higher secretion than Usp45sp silent mutation clones at comparable ΔG_{RBS} values, with more pronounced differences observed at higher levels of induction. We also simulated secretion using the same mathematical model as described above, except we used a higher secretion constant for the Usp45TM8 amino acid sequence (see Discussion and Table 7.3), and there is good agreement between our

modeling and experimental results (Fig. 7.8). Our best clones, Usp45TM8_SM14 and TM8_SM25, yielded a ~50% increase in α-amylase secretion compared to Usp45sp.

Table 7.7. mRNA folding energies of the AGGAGG ribosome binding site and activities of secreted α-amylase of selected clones from the Usp45TM8 silent mutation library [a].

| Usp45TM8 silent mutation clone	ΔG of ribosome binding site (AGGAGG)	Secreted α-amylase activity (% of Usp45sp) [b]
Usp45TM8_SM15	-8.2	135 ± 2
Usp45TM8_SM7	-6.1	137 ± 6
Usp45TM8_SM22	-6.1	140 ± 2
Usp45TM8_SM25	-6.1	151 ± 3
Usp45TM8_SM9	-3.3	145 ± 5
Usp45TM8_SM10	-3.0	143 ± 9
Usp45TM8_SM14	-3.0	147 ± 5
Usp45TM8	-2.1	145 ± 10

[a] α-Amylase secretion levels are reported for cultures induced with 10 ng/ml nisin for 2 h at mid-log phase. ΔG, mRNA folding energy. Nucleotide sequences are given in Table 7.8.

[b] α-Amylase secretion data represent the means ± SEM of the results of three independent experiments. For each value in this column, $P < 0.01$ (for statistical comparison of α-amylase activities of the respective clone and Usp45sp using one-tailed Student's t test).

Table 7.8. Nucleotide sequences of selected clones from the Usp45TM8 silent mutation library.

Usp45TM8 silent mutation clone	Nucleotide sequence (5' → 3')
Usp45TM8	ATGAAAAAAAAAGGTGCTGAAGGCTCATTTAGCTGTGGTTGTGATGCTTACG ACGGCAGCCCCGATTTCCAATGTTAAGGCC

| Usp45TM8_SM7	ATGAAAAAAAAGGTCCTTAAAGCACACTTGGCCGTGGTAGTGATGCTAACG ACTGCTGCCCCGATTTCTAATGTAAAAGCC
Usp45TM8_SM9	ATGAAAAAGAAGGTGCTCAAAGCACATTTAGCTGTGGTGGTGATGCTTACA ACTGCTGCTCCCATTTCTAACGTTAAAGCC
Usp45TM8_SM10	ATGAAAAAAAAGGTCCTCAAGGCTCACTTGGCAGTGGTGGTAATGCTTACC ACAGCGGCCCCTATCTCTAATGTCAAGGCC
Usp45TM8_SM14	ATGAAAAAGAAGGTCCTCAAGGCCCACTTGGCCGTAGTCGTTATGCTAACC ACGGCTGCTCCAATATCTAACGTAAAAGCC
Usp45TM8_SM15	ATGAAAAAGAAGGTACTCAAAGCACATTTAGCTGTTGTCGTAATGCTCACA ACGGCCGCTCCTATATCGAATGTCAAAGCC
Usp45TM8_SM22	ATGAAAAAGAAGGTCCTTAAGGCACATTTAGCTGTAGTCGTTATGCTAACT ACGGCAGCCCCAATCTCAAATGTGAAAGCC
Usp45TM8_SM25	ATGAAAAAGAAGGTCCTCAAGGCCCACTTAGCAGTTGTTGTGATGCTCACC ACAGCTGCCCCTATCTCGAATGTTAAAGCC
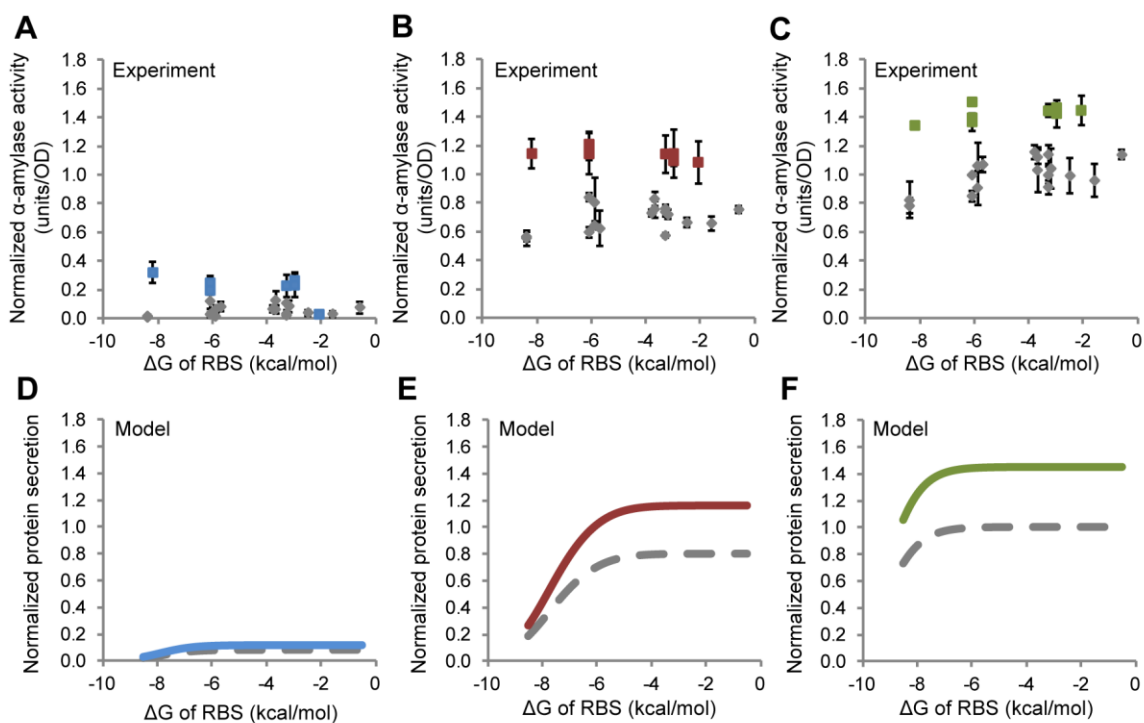
Figure 7.8. Secretion efficiencies of clones selected from Usp45TM8 silent mutation library compared to clones from Usp45sp silent mutation library. α-Amylase activity (normalized to the amount secreted by

wild-type Usp45sp at 10 ng/ml nisin) was plotted against ΔG of the RBS for induction at (A) 0.1 ng/ml, (B) 1 ng/ml, and (C) 10 ng/ml nisin for Usp45TM8 silent mutation clones (color squares; Usp45sp silent mutation clones as described for Fig. 7.6 are shown as gray diamonds for comparison). The data represent the means ± SEM of the results of three independent experiments. Simulation results show protein expression levels against folding energy for (D) low (3.2 μM), (E) medium (32 μM), and (F) high (32 μM) mRNA transcript levels. Color lines represent Usp45TM8 silent mutation clones, and dashed gray lines represent Usp45sp silent mutation clones (data from Fig. 7.6).

7.4 Discussion

Using directed evolution, we have systematically explored the mRNA and amino acid sequence space of Usp45sp to achieve enhanced protein secretion in *L. lactis* (Fig. 7.9). A minor limitation of wild-type Usp45sp lies in its significant mRNA secondary structure, which may lead to less-efficient translation initiation. By relaxing mRNA folding around the RBS (-6.1 kcal/mol for Usp45sp and -3.8 kcal/mol for Usp45SM10), we could slightly improve secretion (up to 16%). The modest nature of this effect could be quantitatively captured by a mathematical model that relates the free energy of folding around the RBS to protein secretion. A more significant limitation lies in the actual amino acid sequence of Usp45sp, which likely impacts the efficacy of the physical process of protein secretion. By mutating nonconserved positions in the tripartite signal peptide, we significantly improved secretion yields (by 41% and 45% for Usp45TM11 and Usp45TM8, respectively). As quantitatively predicted by our mathematical model and confirmed by additional directed evolution experiments, introducing silent mutations into the best amino acid sequence (Usp45TM8) minimally improved protein secretion at

maximum induction (6% additional enhancement for Usp45TM8_SM14 compared to Usp45TM8).
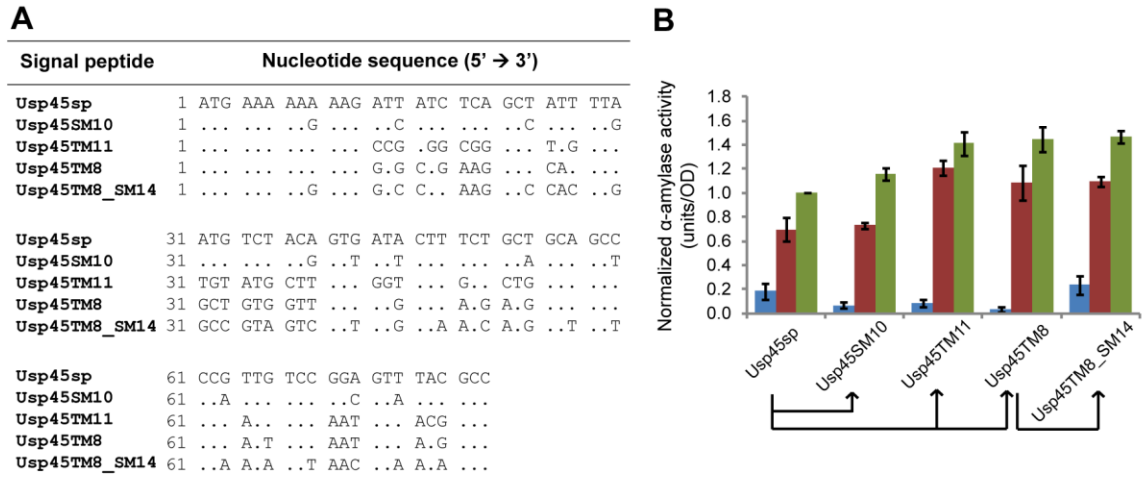
Figure 7.9. Progressive improvements in *L. lactis* secretion achieved by directed evolution. (A) Nucleotide sequence alignment of selected clones (only differences from Usp45sp are shown). (B) α-Amylase activity of selected clones at 0.1 ng/ml (blue), 1 ng/ml (red), and 10 ng/ml (green) nisin induction. The data represent the means ± SEM of the results of three independent experiments. Arrows below the graph show the evolutionary relationships between the different selected clones.

Intriguingly, the best silent mutation clones for Usp45TM8 were able to enhance secretion at lower induction levels. The Western blot in Fig. 7.7C and the starch azure assay in Fig. 7.9B show that secretion of α-amylase by Usp45TM8 and -TM11 at 0.1 ng/ml nisin is lower than that by wild-type Usp45sp despite enhancing secretion at higher nisin concentrations (1 and 10 ng/ml). Silent mutations in Usp45TM8 were able to restore wild-type secretion levels at 0.1 ng/ml nisin, and Usp45TM8_SM14 is shown as an illustrative example (Fig. 7.9B). Although the mechanism by which silent mutations

enhance submaximal secretion is not entirely clear, combining amino acid and silent mutation strategies has enabled us to augment the maximum secretion enhancement of 51% without jeopardizing secretion at lower induction level.

Our mathematical model is most successful in explaining the relationship between mRNA folding and protein secretion at a high nisin induction level (e.g., 10 ng/ml nisin) (Fig. 7.6 and 7.8). In this regime, the availability of ribosomes is limiting and translation rate is at its maximum. At lower nisin concentrations (1 or 0.1 ng/ml nisin), the number of mRNA transcripts becomes limiting and other factors not considered by the model (e.g., degradation of mRNA[36]) might become more important. Thus, it is possible that Usp45TM8, while as efficient as Usp45TM8_SM14 in translation and secretion, is subjected to a higher rate of mRNA degradation. This effect may be negligible at the high induction levels but becomes more pronounced when the mRNA transcript level becomes limiting. The importance of such secondary effects at lower induction levels (0.1 and 1 ng/ml nisin) may account for the disparity between the experiments and model simulations at more-negative ΔG_{RBS}. Clones with stronger folding around the RBS also tend to have a higher level of secondary structure overall, which might reduce mRNA degradation. By neglecting the counterbalancing effect of reduced mRNA degradation in the model, the simulations should underestimate secretion at low ΔG_{RBS} and this is indeed what is seen (Fig. 7.6 and 7.8). Another assumption in the model is that the mRNA transcript level varies linearly with nisin concentration for the range of inducer used in our experiments.

In analyzing our experimental results, we initially attempted to use ΔG of the entire translation initiation region to predict expression, as has been done in other work[35,37], but we did not see any correlations even when different mRNA windows were used. In our study, small windows could not capture the contributions of the entire signal peptide region (81 nucleotides from the start codon), while large windows included significant, but distal mRNA secondary structures that did not affect translation initiation. It was also not possible to use ΔG of the untranslated region around the RBS[27] because our silent mutations are contained in the coding region. Therefore, we simulated folding of the mRNA sequence from the transcription start site to six nucleotides after the signal peptide to capture all silent mutation contributions, but the relevant ΔG was calculated only from base pairing involving the RBS. These values were computed using ΔG data from mfold[25], but since ΔG contributions of individual base pairs were parsed from the overall mRNA structure, an initiation term of 3.4 kcal/mol was added[26]. The ΔG for unfolded mRNA binding to the 30S ribosomal subunit was -14 kcal/mol[27]. This value is consistent with experimentally measured values[38] with the additional consideration of translational standby sites[39].

With a limited number of targeted mutation clones, it is difficult to present a comprehensive picture of the molecular determinants in Usp45sp that govern protein secretion. Nevertheless, two of the best clones in our study, Usp45TM8 and -TM11, appeared to increase the overall positive charge in the n-region without introducing destabilizing charges in the hydrophobic h-region (Fig. 7.7). It has been suggested that positive charges in the n-region are responsible for interacting with the Sec translocation

machinery and the negatively charged phospholipids in the membrane during translocation[40]. Increasing the positive charge in this region has been shown to improve secretion efficiency in *E. coli* and *B. brevis*[20,41,42]. In the hydrophobic h-region, Usp45TM8 has an increased number of polar residues (four) and Usp45TM11 has a decreased number (two) compared to wild-type Usp45sp (three). Previous studies investigating the effect of introducing hydrophobic residues such as leucines in this region have given contradictory results; secretion has been shown to increase in *E. coli* and *B. brevis* but has been shown to decrease in *L. lactis*[20,41,42]. In the light of these findings and our own results, it is possible that the specific amino acid identities and their actual positions within the h-region are important factors in regulating secretion. For example, in Usp45TM11, a glycine is found in position 8 of the h-region, and this helix-breaking residue is commonly present in the middle of the hydrophobic core to allow the signal peptide to form a hairpin-like structure to facilitate membrane insertion[40]. Given our library design, mutations in the c-region should have relatively minor influences on the overall secretion efficiency since most of the important residues (proline and serine favoring β-turns; -3 and -1 residues for signal peptidase cleavage)[43] were held constant (Fig. 7.7A). Overall, the best signal peptides from our screens preserve the tripartite structure of the signal peptide but increase the positive charge of the n-region.

Many other factors, such as the overall charge balance, hydrophobicity profile, and interactions among the n-, h-, and c-regions, can influence secretion efficiency, so further design of signal peptides may still necessitate extensive trial-and-error. The lack of a blueprint for signal peptide engineering motivated us to use a directed evolution

approach to discover sequences that enhance secretion. In designing our targeted mutation library, we attempted to retain features that appear to be generally conserved in signal peptides while allowing for full randomization in the remaining positions. This library was far from exhaustive, but nevertheless, our simple screen, requiring only a transformation step and a plate assay, was successfully used to identify signal peptide mutants with an increase of up to 51% in secretion relative to the most efficient natural signal peptide in *L. lactis*.

Future work should include examining the potential context dependence of selected signal peptides, since it is possible, even likely, that the optimal signal peptide will change with different expression systems or coding sequences. In our screens, we found that signal peptides can exhibit different secretion efficiencies at different expression levels (e.g., low vs. high nisin concentration in Fig. 7.9B). Our screen was designed to select the best signal peptide at the highest expression level; however, if a promoter with an attenuated expression level were preferred, a different signal peptide might be optimal. We have also observed protein-specific effects on secretion by comparing nuclease and α-amylase secretion for the selected clones (Table 7.4 and Fig. 7.7). To address these concerns and to use this selection system for finding the optimal signal peptide for secretion of therapeutic or recombinant proteins of interest (e.g., insulin from Chapter 6), a signal peptide library can be built downstream of the desired promoter and upstream of a therapeutic-nuclease fusion reporter. The same nuclease screen can be employed, but any effects on secretion arising from construct-specific interactions within the upstream portion of either the encoding mRNA or the corresponding amino acid

sequence can now be accounted for. The protein of interest is fused directly to the signal peptide, which is an important consideration, as our work and previous work[44] have shown dependence of the 5' end of the encoding mRNA on expression. In fact, a toxic protein(MazF)-nuclease fusion reporter assay has been successfully used to select for *L. lactis* signal peptides with increased secretion, which reduced intracellular levels of the toxic protein and resulted in cell survival (Fig. 7.10).
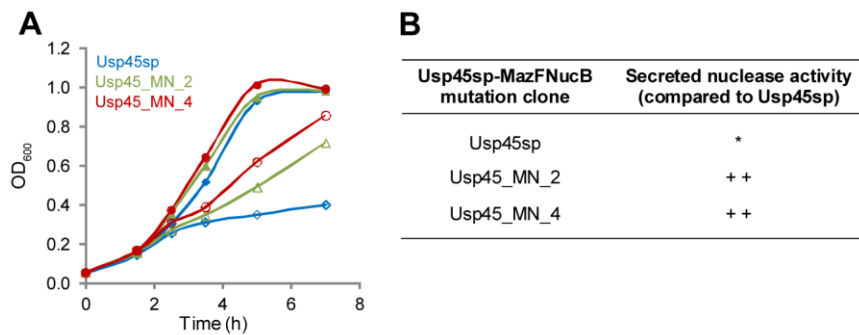
Figure 7.10. Selection results for Usp45sp silent mutation library fused to MazF-NucB reporter gene. (A) Growth curves of MazF-NucB cultures with Usp45sp (blue ♦: no nisin, red ◊: 1 ng/ml nisin), Usp45_MN_2 (green ▲: no nisin, green Δ: 1 ng/ml nisin), and Usp45_MN_4 (red ●: no nisin; red ○: 1 ng/ml nisin) showed improvements in cell survival in Usp45_MN_2 and Usp45_MN_4 variants, potentially due to greater toxin secretion. (B) Secreted nuclease activity for MazF-NucB cultures induced with 1 ng/ml nisin. * and ++ indicate secretion levels similar to and > 10% higher than Usp45sp levels respectively.

Additional studies are also needed to determine whether even greater improvements in secretion can be achieved solely by engineering the signal peptide, since other steps in the secretion process such as translocation efficiency[23], modification by chaperones[14], and transportation across the cell wall[45] might become limiting. Indeed, protein secretion in *L. lactis* has been improved by introduction of a synthetic

propeptide[22] or by complementation with the SecDF machinery[23]. Thus, other natural or engineered bacterial species with higher levels of protein secretion may serve as better fundamental models for studying inherent limitations of signal peptides. Nonetheless, this study produced signal peptides that can replace wild-type Usp45sp in many biotechnological and clinical applications of *L. lactis* and it also highlights the potential of directed evolution as a general approach for signal peptide engineering in bacteria.

7.5 References

1. Leroy, F. & De Vuyst, L. Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci. Technol.* **15**, 67–78 (2004).

2. Klijn, N., Weerkamp, A. H. & De Vos, W. M. Genetic marking of *Lactococcus lactis* shows its survival in the human gastrointestinal tract. *Appl. Environ. Microbiol.* **61** , 2771–2774 (1995).

3. Wells, J. M. & Mercenier, A. Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nat. Rev. Microbiol.* **6**, 349–362 (2008).

4. Steidler, L. *et al.* Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. *Science* **289** , 1352–1355 (2000).

5. Chatel, J. *et al*. Induction of mucosal immune response after intranasal or oral inoculation of mice with *Lactococcus lactis* producing bovine beta-lactoglobulin. *Clin. Vaccine Immunol.* **8**, 545–551 (2001).

6. Morello, E. *et al. Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* **14**, 48–58 (2008).

7. Poquet, I. *et al.* HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. *Mol. Microbiol.* **35**, 1042–1051 (2000).

8. Van Asseldonk, M. *et al.* Cloning of usp45, a gene encoding a secreted protein from *Lactococcus lactis* subsp. lactis MG1363. *Gene* **95**, 155–160 (1990).

9. Bolotin, A. *et al.* The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* **11**, 731–753 (2001).

10. Wegmann, U. *et al.* Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. cremoris MG1363. *J. Bacteriol.* **189**, 3256–3270 (2007).

11. Mierau, I. & Kleerebezem, M. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. *Appl. Microbiol. Biotechnol.* **68**, 705–717 (2005).

12. Nouaille, S. *et al.* Heterologous protein production and delivery systems for *Lactococcus lactis*. *Genet. Mol. Res.* **2**, 102–111 (2003).

13. Desvaux, M., Hébraud, M., Talon, R. & Henderson, I. R. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol.* **17**, 139–145 (2009).

14. Le Loir, Y. *et al.* Protein secretion in *Lactococcus lactis* : an efficient way to increase the overall heterologous protein production. *Microb. Cell Fact.* **4**:2 (2005).

15. Von Heijne, G. The signal peptide. *J. Membr. Biol.* **115**, 195–201 (1990).

16. Van Asseldonk, M., De Vos, W. M. & Simons, G. Functional analysis of the *Lactococcus lactis* usp45 secretion signal in the secretion of a homologous proteinase and a heterologous α-amylase. *Mol. Gen. Genet.* **240**, 428–434 (1993).

17. Ravn, P., Arnau, J., Madsen, S. M., Vrang, a & Israelsen, H. The development of TnNuc and its use for the isolation of novel secretion signals in *Lactococcus lactis*. *Gene* **242**, 347–356 (2000).

18. Poquet, I., Ehrlich, S. D. & Gruss, A. An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis*. *J. Bacteriol.* **180**, 1904–1912 (1998).

19. Perez-Martinez, G. *et al.* Protein export elements from *Lactococcus lactis*. *Mol. Gen. Genet.* **234**, 401–411 (1992).

20. Ravn, P., Arnau, J., Madsen, S. M., Vrang, A. & Israelsen, H. Optimization of signal peptide SP310 for heterologous protein production in *Lactococcus lactis*. *Microbiology* **149**, 2193–2201 (2003).

21. Mercenier, a. *et al.* Screening and construction of probiotic strains with enhanced protective properties against intestinal disorders. *Microb. Ecol. Health Dis.* **16**, 86–95 (2004).

22. Le Loir, Y., Gruss, A., Ehrlich, S. D. & Langella, P. A nine-residue synthetic propeptide enhances secretion efficiency of heterologous proteins in *Lactococcus lactis*. *J. Bacteriol.* **180**, 1895–1903 (1998).

23. Nouaille, S. *et al.* Complementation of the *Lactococcus lactis* secretion machinery with *Bacillus subtilis* SecDF improves secretion of staphylococcal nuclease. *Appl. Environ. Microbiol.* **72**, 2272–2279 (2006).

24. Lachica, R. V, Genigeorgis, C. & Hoeprich, P. D. Metachromatic agar-diffusion methods for detecting staphylococcal nuclease activity. *Appl. Microbiol.* **21**, 585–587 (1971).

25. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

26. Freier, S. M. *et al.* Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9373–9377 (1986).

27. de Smit, M. H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7668–7672 (1990).

28. Fegatella, F., Lim, J., Kjelleberg, S. & Cavicchioli, R. Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium *Sphingomonas* sp. strain RB2256. *Appl. Environ. Microbiol.* **64**, 4433–4438 (1998).

29. Ng, D. T. W. & Sarkar, C. A. Nisin-inducible secretion of a biologically active single-chain insulin analog by *Lactococcus lactis* NZ9000. *Biotechnol. Bioeng.* **108**, 1987–1996 (2011).

30. De Smit, M. & Van Duin, J. Control of translation by mRNA secondary structure in *Escherichia coli*: a quantitative analysis of literature data. *J. Mol. Biol.* **244**, 144–150 (1994).

31.   Kaminishi, T. *et al.* A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. *Structure* **15**, 289–297 (2007).

32.   Gualerzi, C. O. & Pon, C. L. Initiation of mRNA translation in prokaryotes. *Biochemistry* **29**, 5881–5889 (1990).

33.   Le Loir, Y., Gruss, A., Ehrlich, S. D. & Langella, P. Direct screening of recombinants in gram-positive bacteria using the secreted staphylococcal nuclease as a reporter. *J. Bacteriol.* **176**, 5153–5139 (1994).

34.   Pontes, D. S. *et al. Lactococcus lactis* as a live vector: Heterologous protein production and DNA delivery systems. *Protein Expr. Purif.* **79**, 165–175 (2011).

35.   Kudla, G., Murray, A., Tollervey, D. & Plotkin, J. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).

36.   Schlax, P. J. & Worhunsky, D. J. Translational repression mechanisms in prokaryotes. *Mol. Microbiol.* **48**, 1157–1169 (2003).

37.   Na, D., Lee, S. & Lee, D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Syst. Biol.* **4**:71 (2010).

38.   Calogero, R., Pon, C., Canonaco, M. & Gualerzi, C. Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6427–6431 (1988).

39.   De Smit, M. H. & Van Duin, J. Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. *J. Mol. Biol.* **331**, 737–743 (2003).

40.   Tjalsma, H., Bolhuis, a, Jongbloed, J. D., Bron, S. & Van Dijl, J. M. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* **64**, 515–547 (2000).

41.   Izard, J. W., Doughty, M. B. & Kendall, D. a Physical and conformational properties of synthetic idealized signal sequences parallel their biological function. *Biochemistry* **34**, 9904–9912 (1995).

42.   Takimura, Y., Kato, M., Ohta, T., Yamagata, H. & Udaka, S. Secretion of human interleukin-2 in biologically active form by *Bacillus brevis* directly into cultute medium. *Biosci. Biotechnol. Biochem.* **61**, 1858–1861 (1997).

43.   Von Heijne, G. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133**, 17–21 (1983).

44. Welch, M. *et al.* Design parameters to control synthetic gene expression in *Escherichia coli*. *PloS ONE* **4**, e7002 (2009).

45. Forster, B. M. & Marquis, H. Protein transport across the cell wall of monoderm Gram-positive bacteria. *Mol. Microbiol.* **84**, 405–413 (2012).

CHAPTER 8

Conclusions and future directions

8.1 Summary of work

In this work, we have sought to develop the next generation of therapeutics in two major ways: through molecular engineering approaches guided, when possible, by mathematical models, and through cellular engineering approaches to create a microorganism-based oral delivery platform. Our molecular engineering efforts have led to the development of two robust *in vitro* directed evolution protocols for mRNA display (Chapter 4) and ribosome display (Chapter 5), which were then employed to display and study a variety of interesting scaffold proteins and agonists. Through *in vivo* directed evolution, namely an *Escherichia coli* display system (Chapter 2) and a *Lactococcus lactis* platform (Chapter 7), proteins and peptides with new and useful properties have been evolved and these properties have been linked to desired responses through mathematical models that provide new mechanistic insights. In the area of antibody engineering, we successfully employed a CDR grafting strategy for scFv stabilization and humanization. Our work on Fc domain engineering identified specific binding properties of aglycosylated Fc mutants that enhance tumor cell phagocytosis; in a separate but related study, we altered a monomeric Fc domain by introducing mutations that stabilize and improve serum half-life, and this variant retained FcRn binding. For protein therapeutics from

nonimmunoglobulin domains, we elucidated stability determinants for a DARPin, as well as two important agonists, insulin and leptin. Throughout this work, mathematical modeling was critical for guiding directed evolution efforts and understanding outcomes. The molecular insights generated from the mechanistic models on aglycosylated Fc and signal peptide engineering can inform future engineering efforts. The thermodynamic ligation model and related software (Chapter 3) provide an easy way to rationally optimize ligation reactions for DNA library construction.

Complementary to our molecular engineering work, we have also proposed a novel cell-based platform for oral delivery of protein therapeutics based on *Lactococcus lactis* (Chapter 6). Single-chain insulin and leptin have been used as model agonists in this platform. We further improved the therapeutic potential of *L. lactis* through signal peptide engineering (Chapter 7). Our results open up the possibility of systemic delivery of protein therapeutics using live microorganisms.

8.2 Future directions in molecular engineering studies

A direct extension of this work would be the molecular evolution of some of the proteins already functionally displayed in either the ribosome display or mRNA display platform. With ribosome display, randomized internal repeats and/or capping repeats could be added to the truncated Off7 fragments and selected for MBP binding. The resulting binders could then be compared to wild-type Off7 to examine differences in sequence and affinity arising from different evolutionary paths. Single-chain insulin, leptin, and the

disulfide-free leptin variant could be evolved for higher stability or reduced proteolysis, particularly in the gut.

A new generation of hR6.5 with alternative stable frameworks could be tested for human ICAM-1 binding. If successful, the next step would be to evolve hR6.5 or hYN1 for cross-species binding. In particular, there is interest in an scFv capable of binding to rat ICAM-1, which is important for translational research since it will allow the study of *in vivo* biodistribution and protective effects in small animal models (C. Greineder, personal communication). In order to meet this evolutionary objective, DNA shuffling of hR6.5 and hYN1, coupled with additional point mutations introduced through random mutagenesis, may make a good starting library for ribosome display.

The bacterial display platform used to screen aglycosylated full length antibodies can be used to evolve Fc domains to enrich for FcγRIIa-H131/FcγRIIb selectivity since our mathematical model clearly suggested a dominant role for FcγRIIa-H131 in effector activation. Evolving more avid and selective binders to this new engineering target should lead to an even greater enhancement in antibody-dependent cellular phagocytosis.

The availability of a new streamlined mRNA display protocol, which utilizes highly stable, covalently assembled selection particles, should facilitate the development of novel selection methodologies that transcend simple binding and enrich for more complex behaviors directly on cells. One such function is transcytosis, where a new selection method with mRNA display utilizing cultured polarized epithelial cells can be investigated. Monolayers of Caco-2 or HT-29 cells grown on a permeable support (e.g., a

Transwell plate) differentiate and form tight junctions, resembling the intestinal epithelium[1,2]. In each round of mRNA display, the mRNA-protein complexes could be added to the apical side of the monolayer and selection particles that display transcytosis properties would be recovered on the basolateral side through RT-PCR of the mRNAs in the medium. Although there may be concerns about mRNA stability during this process, preliminary studies using ribosome display for selection of human leptin across Caco-2 monolayers showed that at least some of the mRNA from the ternary complexes added to the apical side can be recovered from the basolateral side (Fig. 8.1), suggesting that selection via transcytosis is in principle possible using *in vitro* display methods. For future studies, mRNA display might represent a better strategy because the covalently linked mRNA-protein complex is much more stable than the mRNA-ribosome-protein ternary molecule. A protein with greater potential than leptin for transcytosis in the gut might also be considered in developing this cell-based mRNA display protocol. A convenient choice would be mFc.67, since Fc/FcRn has more established mechanisms for transcytotis in the gut[3] and since mFc.67 has already been shown to specifically bind to FcRn in mRNA display format (Chapter 4). Overall, the idea that assays can be designed to directly select for protein functions in the context of a cell is extremely exciting as it can eliminate disparities that commonly arise between affinity engineering against a purified target and resulting cellular activity of such engineered variants.
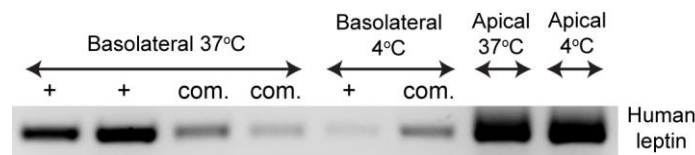
Figure 8.1. Transcytosis of human leptin across Caco-2 monolayer assessed by ribosome display. Caco-2 cells were grown on a 6-well Transwell plate and differentiated. Ternary complexes of human leptin were added to the apical side of the monolayer. "+" represents wells with only cells and "com." represents wells with cells in the presence of excess leptin competitor (5×10^{14}/well). Panning was performed for 1 h at 37°C or 4°C with shaking; then, 200 μl solution from the basolateral or apical side was taken to isolate mRNA according to protocols described in Chapters 4 and 5. Products after RT and 25 cycles of PCR were shown.

Extending this idea even further, another interesting molecule that displays high transcytotic potential is botulinum neurotoxin B (BoNT/B). This food toxin is generally regarded as one of the most potent poisons[4]. After ingestion of the toxin, BoNT/B binds to receptors on the mucosal side of the intestinal epithelium and undergoes transcytosis to reach systemic circulation, where it acts on peripheral cholinergic nerve endings to block acetylcholine release[5]. Fortunately, its binding and translocation domains are distinct from the catalytic domain that is responsible for its toxicity[6]. Upon removal of the toxic subunit, the BoNT/B binding subunit can potentially provide unique advantages over other carrier molecules for certain transcytosis applications (see next section).

8.3 Future directions in cell-based oral delivery studies

A more comprehensive study is needed to evaluate the *in vivo* efficacy of delivering SCI-57 orally to diabetic mice via *L. lactis*. However, our completed preliminary mice experiments suggest several areas for improvement of the platform. It is therefore prudent to first maximize the delivery potential of *L. lactis* before proceeding to further *in vivo*

testing. One significant issue is the dosage of therapeutic proteins in systemic circulation following oral delivery of *L. lactis*. Two challenges need to be addressed in this regard: 1) maximizing the amount of protein secreted by *L. lactis* in the intestinal lumen and 2) maximizing the fraction of this secreted protein that reaches systemic circulation.

To partially address the first challenge, we have evolved signal peptides capable of increasing the secretion level of *L. lactis*. However, since the library used was far from exhaustive and the selection was carried out with a generic reporter protein, further optimization of the signal peptide may be possible. The newly evolved signal peptides can then be fused to the therapeutic protein of interest to enhance secretion. Another strategy to improve secretion would be to overexpress the translocation machinery or complement missing members (SecDF)[7] to maximize the intrinsic secretion capacity of *L. lactis*. A third strategy would be to engineer stronger promoters so that more protein molecules can be synthesized. A library of P*nisA* (the strongest promoter tested so far) can be synthesized by keeping the transcription initiation site, -35 and -10 sites before the transcription initiation site, and ribosome binding site sequences constant while introducing stretches of randomized sequences around them; the fixed motif sequences ensure promoter function, while the random bases modulate efficiency[8]. The promoter library can then be introduced upstream of a reporter gene (e.g., green fluorescence protein) and stronger promoters can be enriched by fluorescence-activated cell sorting. In fact, the reagents needed for such promoter engineering have already been acquired. All of these strategies could eventually be combined to significantly increase the amount of

protein expressed and secreted by *L. lactis*, which would correspond to enhanced therapeutic delivery during treatment.

To address the second issue of bioavailability, mechanisms for efficient absorption of the delivered proteins are needed. Carrier-mediated transcytosis remains a promising approach for increasing absorption of bulky protein molecules. We briefly investigated this strategy using leptin as a transcytosis fusion partner, and we successfully secreted functional SCI-57-leptin fusion proteins from *L. lactis*. However, preliminary ELISA experiments assessing for transcytosis of leptin across Caco-2 cells gave a low yield on the basolateral side. Moreover, although there is some evidence of leptin transcytosis[9,10], there are other candidates with more well-characterized properties for intact transcytosis. Specifically, transcytosis partners such as mFc.67, mFc.67 variants, or BoNT/B might elicit greater success in enhancing intestinal absorption.

While molecular engineering can increase the proteolytic stability of the delivered protein, cellular engineering of *L. lactis* itself might improve the survival of the bacteria in harsh environments. For example, trehalose can be co-expressed by *L. lactis* to improve its viability after freeze-drying (for formulation) and survival in the gastric environment[11]. An evolutionary approach similar to the one described in Dekel and Alon[12] could also be used to evolve *L. lactis* strains that better adapt to the gastric or intestinal environment by serial dilution of a culture through hundreds of generations in progressively more unfavorable environments.

Further extensions to this work could include construction of a glucose sensor in *L. lactis* to express proteins when glucose is present. The *L. lactis* strain MG5267 can be used to allow growth with lactose as the carbon source[13]. In the presence of glucose, the constitutively expressed carbon catabolite control protein A (ccpA) activates the cre-containing promoter of the las operon and upregulates its downstream genes[14]. It might be possible to mimic these existing interactions to create a synthetic "glucose-sensing" promoter to trigger therapeutic protein production autonomously.

Although this work has mainly focused on systemic delivery of protein therapeutics, *L. lactis* could also be used for systemic delivery of vaccines[15]. It would be interesting to use the transcytosis fusion partner BoNT/B for this application since the fusion partner would be highly immunogenic and could serve as a potent adjuvant to maximize the immune response against the fused antigen. In addition, mucosal delivery of therapeutics using *L. lactis*, which does not depend on transcytosis, is a very exciting area of research[15]. For example, mucosal delivery of insulin was used to induce antigen-specific tolerance for treatment of Type I diabetes[16]. This mode of antigen delivery to the gut for induction of immune tolerance could be further exploited with other antigens. Moreover, other deliverables which act on cells in the gastrointestinal tract can also be easily delivered by *L. lactis*. In short, this simple and inexpensive cell-based platform has the potential to be used for oral delivery of a wide range of important protein therapeutics.

8.4 References

1. Hughson, E. J. & Hopkins, C. R. Endocytic pathways in polarized Caco-2 cells: identification of an endosomal compartment accessible from both apical and basolateral surfaces. *J. Cell Biol.* **110**, 337–348 (1990).

2. Louvard, D., Kedinger, M. & Hauri, H. P. The differentiating intestinal epithelial cell: establishment and maintenance of functions through interactions between cellular structures. *Annu. Rev. Cell Biol.* **8**, 157–195 (1992).

3. Dickinson, B. L. *et al.* Bidirectional FcRn-dependent IgG transport in a polarized human intestinal epithelial cell line. *J. Clin. Invest.* **104**, 903–911 (1999).

4. Davis, L. Botulinum toxin: from poison to medicine. *West. J. Med.* **158**, 25–29 (1993).

5. Ravichandran, E. *et al.* An initial assessment of the systemic pharmacokinetics of botulinum toxin. *J. Pharmacol. Exp. Ther.* **318**, 1343–1351 (2006).

6. Simpson, L. L., Maksymowych, A. B. & Kiyatkin, N. Botulinum toxin as a carrier for oral vaccines. *Cell. Mol. Life Sci.* **56**, 47–61 (1999).

7. Nouaille, S. *et al.* Complementation of the *Lactococcus lactis* secretion machinery with *Bacillus subtilis* SecDF improves secretion of staphylococcal nuclease. *Appl. Environ. Microbiol.* **72**, 2272–2279 (2006).

8. Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465–471 (2009).

9. Cammisotto, P. G., Gingras, D. & Bendayan, M. Transcytosis of gastric leptin through the rat duodenal mucosa. *Am. J. Physiol. Gastrointest. Liver Physiol.* **293**, G773–G779 (2007).

10. Cammisotto, P. G. *et al.* Receptor-mediated transcytosis of leptin through human intestinal cells *in vitro*. *Int. J. Cell Biol.* **2010**, 928169 (2010).

11. Termont, S. *et al.* Intracellular accumulation of trehalose protects *Lactococcus lactis* from freeze-drying damage and bile toxicity and increases gastric acid resistance. *Appl. Environ. Microbiol.* **72**, 7694–7700 (2006).

12. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592 (2005).

13. Mierau, I. & Kleerebezem, M. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. *Appl. Microbiol. Biotechnol.* **68**, 705–717 (2005).

14. Luesink, E. J., van Herpen, R. E. M. A, Grossiord, B. P., Kuipers, O. P. & de Vos, W. M. Transcriptional activation of the glycolytic *las* operon and catabolite repression of the *gal* operon in *Lactococcus lactis* are mediated by the catabolite control protein CcpA. *Mol. Microbiol.* **30**, 789–798 (1998).

15. Bermúdez-Humarán, L. G., Kharrat, P., Chatel, J.-M. & Langella, P. Lactococci and lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. *Microb. Cell Fact.* **10**, S4 (2011).

16. Takiishi, T. *et al.* Reversal of autoimmune diabetes by restoration of antigen-specific tolerance using genetically modified *Lactococcus lactis* in mice. *J. Clin. Invest.* **122**, 1717–1725 (2012).