



Universidade do Minho
Escola de Ciências da Saúde

Fábio André Eleutério Amaral

**Control of *Streptococcus pneumoniae*
Virulence by Manipulation of the Genetic
Sequence of Virulence Factor Pneumolysin**

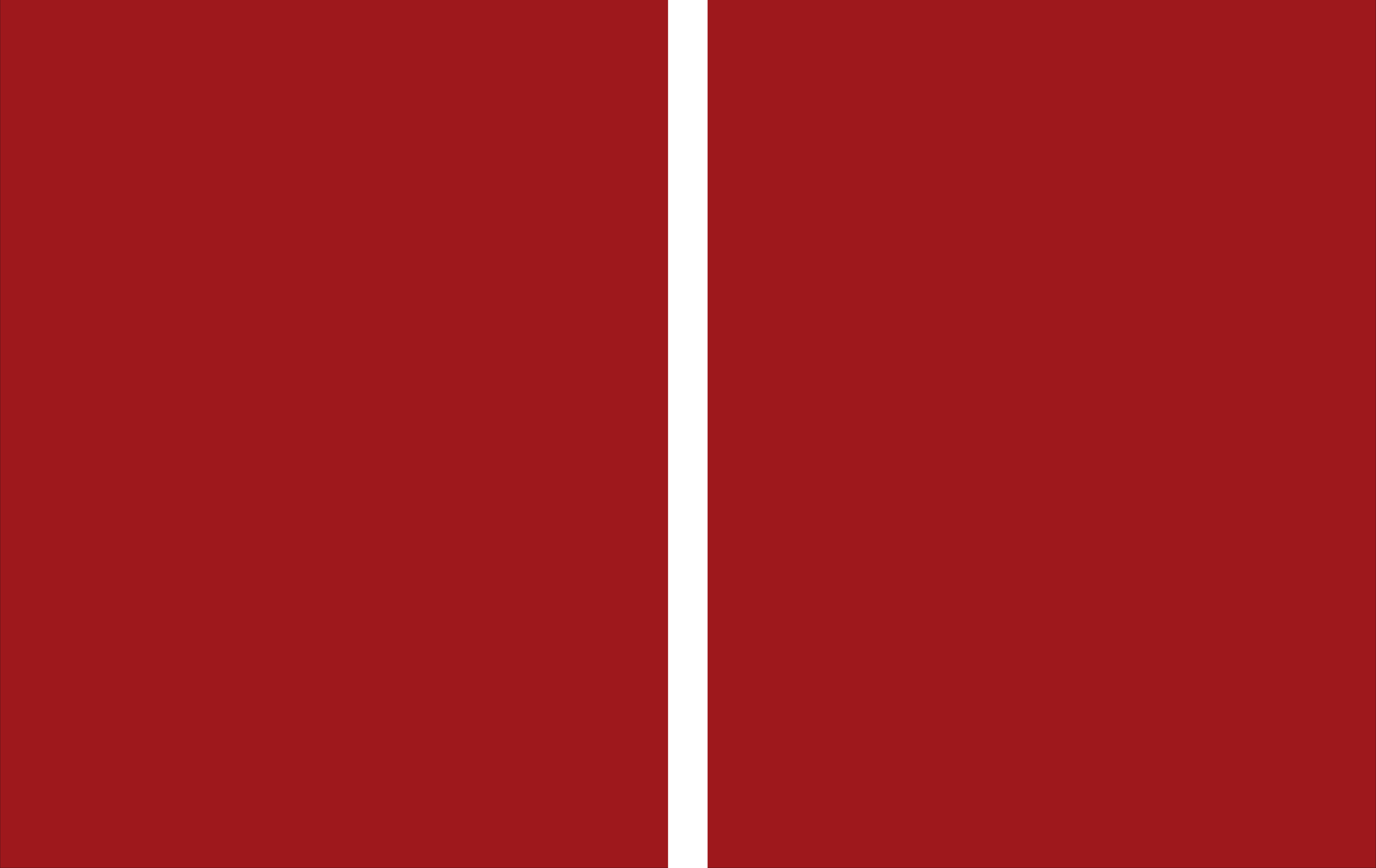
Fábio André Eleutério Amaral
Control of *Streptococcus pneumoniae* Virulence by Manipulation
of the Genetic Sequence of Virulence Factor Pneumolysin

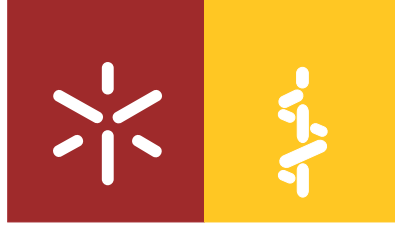
UMinho | 2016

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA



agosto de 2016





Universidade do Minho
Escola de Ciências da Saúde

Fábio André Eleutério Amaral

**Control of *Streptococcus pneumoniae*
Virulence by Manipulation of the Genetic
Sequence of Virulence Factor Pneumolysin**

**Controlo da Virulência de *Streptococcus*
pneumoniae por Manipulação da
Sequência Genética do Fator de Virulência
Pneumolisina**

Tese de Doutoramento em Medicina

Trabalho realizado sob a orientação de
Adam J. Ratner, M.D., M.P.H.
e
Prof. Doutor Fernando Rodrigues

agosto de 2016

DECLARAÇÃO

Nome: Fábio E. Amaral **E-mail:** fabio3amaral@gmail.com

Título da tese

Em inglês: Control of *Streptococcus pneumoniae* Virulence by Manipulation of the Genetic Sequence of Virulence Factor Pneumolysin

Em português: Controlo da Virulência de *Streptococcus pneumoniae* por Manipulação da Sequência Genética do Fator de Virulência Pneumolisina

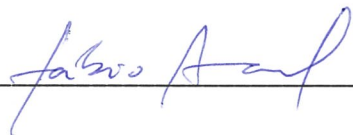
Orientadores: Adam J. Ratner, M.D., M.P.H. e Prof. Doutor Fernando Rodrigues

Ano de conclusão: 2016

Designação do Ramo de Conhecimento do Doutoramento: Medicina

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE (MÁXIMO DE 20 PÁGINAS), APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 4 de agosto de 2016.

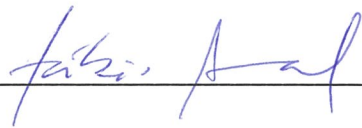


STATEMENT OF INTEGRITY

I, Fábio E. Amaral, hereby declare having conducted my PhD thesis with integrity. I affirm that I have not used plagiarism or any form of falsification of results throughout all of my work.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, August 4, 2016.



ACKNOWLEDGMENTS // AGRADECIMENTOS

I owe many acknowledgements to diverse individuals for helping me along the way.

In no particular order except, maybe, geographical for better organization. And so that I don't forget anyone. If I do, I am honestly sorry and I apologize in advance.

First, I want to thank Adam Ratner for his mentorship and ideas, for his openness to allow me freedom to pursue my own thinking, and for providing help whenever requested.

I would like to thank the Ratner lab for taking me in and providing a great environment for me to conduct my experiments, and for teaching me how to do it. I'd like to personally thank Ritwij Kulkarni for his invaluable friendship, valued advice, technical expertise, and basically every moment in the lab. I'd like to thank Timmy LaRocca for his friendship, interesting scientific discussions (and not), and help. I want to also thank the help and friendship of Ferdinand Los, Tara Randis, Kate Hensel Sapra (also for the great help with statistics), Ryan Rampersaud, Alice Wang, Swati Antala, Jorge Aguilar, Joanne Zaklama, Emma Lewis, Leor Akabas, Tom Hooven, Stephanie Eng, Leah Byland, and all Ratner lab former members.

I want to thank Mimi Shirasu-Hiza for her friendship, advice, mentoring and being there when necessary. I also want to acknowledge all the materials and know-how provided while working with flies. I would like to thank Mimi's lab for the great work environment and all the teaching provided, namely my friend Ben Fulton for everything, Elizabeth Stone Jacobs, Victoria Allen, Vanessa Hill, Albert and Paul Kim, and all other Shirasu-Hiza lab members.

I would like to thank Dane Parker for his friendship, happiness, and for the help with mice experiments. I also would like to acknowledge Grace Soong, Danielle Ahn, and the other Prince lab members for being great neighbors.

Likewise, I'd like to thank the Planet lab and its members for being great neighbors, in special Melanie Harasym for all help and kindness.

I want to thank Max Gottesman for his time to discuss my project and ideas, for his great input, enthusiasm and help.

I want to thank Jonathan Dworkin for providing help as part of my improvised Thesis Committee.

Special thank you to Dr. Steve Mackey, for his friendship, brilliant mentoring in clinical matters, for his enthusiasm and the time spent with me and us.

I want to thank Fred Loweff and Elizabeth Scott for all the logistics and bureaucratic help provided, always in a timely and efficient way.

Thank you to the Microbiology & Immunology Department for having integrated me and for allowing me to take part in the Scientific Retreats and presentations throughout the years, in a great and friendly community at Columbia University.

I want to thank Mike Shelanski for all his help from the beginning, as well as Ron Liem, and the Columbia MD/PhD Program for welcoming me as part of the program and integrating me in all daily aspects of MD/PhD student life.

To all my English-speaking friends I want to truly thank all the moments lived in NYC and everywhere else. Without them, all of this would have been even harder and not half the fun it has been. From Bard Hall Players, to Free Fridays, to Opera and Broadway, shows, circus, food, festivals, trips, bars, museums, parks, sports, parties, karaoke, and so on, and on... You are with me; I'll say names and nothing else =) Vikram, Tulsi, John Seeley, Cindy, Shez, Brie, Tiffany, Kally, Ian, Chris Tan, Dan Hu, Alex Baffet, Christine, Mi, Miki, Olya, Matt Borok, Nilushi, Sarah Deng, Dan K, Lucja, Sedef, Becca, Scott, among others, and Regina Lutz with a special acknowledgement for her precious editorial review of the general introduction.

//

Em português, quero primeiro agradecer ao Prof. Doutor Fernando Rodrigues pela sabedoria, ajuda sempre pronta a pedido, tutoria e apoio.

Depois, agradecer ao Prof. Doutor Nuno Sousa, à Prof.^a Doutora Joana Palha e à Prof.^a Doutora Margarida Correia Neves pela orientação do programa e apoio transatlântico que prestaram, sempre disponíveis.

À Dr.^a Maria José Tarroso agradeço pela diligência em todas as questões burocráticas.

Aos mentores científicos do meu passado, que me prepararam para estas provas, Prof.^a Doutora Paula Sampaio, Prof. Doutor Gil Castro, Prof. Doutor Jorge Pedrosa, Prof. Doutor Pedro Gomes, e Prof.^a Doutora Célia Pais, muito obrigado.

Pela amizade e recepção nos EUA, obrigado Tiago Oliveira, Nuno Lamas, e Ana Franky; abriram caminho e sem vocês era tudo muito mais difícil.

Agradecimento à Célia Soares, Tiago Dantas, Carla Abreu, pelos transportes, ementa tuga, e tudo o mais que mal pode ser expresso. Obrigado à Neide, Raquel, César, Sandro, e outros portugueses que a distâncias oceânicas ajudaram em tudo o que era divertido e não só, em especial à Ana Martins pela mobília, duas vezes, e por ser quem é.

Aos amigos que me visitaram, colegas que apoiaram, ao curso de Medicina ECS 2010, do qual sou delegado honorário, por não esquecerem, pelas festas surpresa e pelo encorajamento, em especial à Sónia Duarte e à Sofia Patrão, e depois ao gangue Romero, Berta, Cake, Suzy, Catarina, Christoph, Migas, e comunidade MTG cá de Braga.

Por fim, à minha família, pai, mãe, irmãs, tias, tios, avós, avôs, primos e primas, que me apoiaram à distância, que sempre se preocuparam sem necessidade, e que estão sempre comigo. O elevado reconhecimento escrito nunca será suficiente, mas aqui fica. O apoio incondicional permite concretizar os maiores projetos e ambições... só assim mantendo a minha sanidade física e mental!

Financial support

Fábio E. Amaral was supported by the Portuguese Foundation for Science and Technology (FCT) SFRH/BD/33901/2009, funded by QREN Portugal 2007-2013 POPH and by MCTES, and by the Luso-American Development Foundation.

This work was developed under the scope of the project NORTE-01-0145-FEDER-000013, supported by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through FEDER, through COMPETE, and by National funds, through FCT, under the scope of the project POCI-01-0145-FEDER-007038.

ABSTRACTS

The following include English and Portuguese abstracts for this work, as well as a lay summary in both languages.

Abstract

Control of *Streptococcus pneumoniae* Virulence by Manipulation of the Genetic Sequence of Virulence Factor Pneumolysin

Streptococcus pneumoniae is one of the most important pathogens of humans. It produces a toxin, pneumolysin (PLY), able to form extraordinarily large pores in the membrane of attacked cells, for which they are dependent on the presence of membrane cholesterol. Murine models of acute pneumonia have shown that PLY is essential for the survival of pneumococcus in both the upper and lower respiratory tracts, as well as it is required for bacterial spread from the lungs to the bloodstream. Besides colonization and invasive disease, PLY has a role in immunogenic protection response. It has been proposed as a strong serotype-independent vaccine candidate.

In this thesis, we demonstrate that manipulation of 5' mRNA folding free energy (ΔG) by introduction of synonymous (silent) mutations allows for projected changes in PLY translation by pneumococcus without the need for chemical inducers or heterologous promoters. Hence, we created a panel of isogenic strains differing only in a few silent mutations at the 5' end of the *ply* mRNA that are predicted to alter ΔG by a wide range.

In vitro, such manipulation allows rheostat-like control of PLY production and alters the cytotoxicity of whole *S. pneumoniae* on primary and immortalized human cells. We demonstrate that production of PLY increases due to faster translation initiation with unchanged mRNA stability and decay. *In vivo* virulence was assessed using a mouse model in which we show differential outcome of pneumococcal pneumonia according to the 5' mRNA ΔG . Inflammation features on lung histology and proinflammatory cytokine profile both positively and significantly correlate to PLY expression.

This strategy of rheostat control of virulence factors is useful and provides proof-of-principle as a new tool for understanding the role of specific factors upon host-bacterial interactions. Further research on mRNA ΔG manipulation is a resourceful tool in the study of bacterial pathogenesis.

Resumo

Controlo da Virulência de *Streptococcus pneumoniae* por Manipulação da Sequência Genética do Fator de Virulência Pneumolisina

Streptococcus pneumoniae é um dos agentes patogénicos mais importantes de humanos. Produz uma toxina, pneumolisina (PLY), capaz de formar poros extraordinariamente largos nas membranas de células atacadas, efeito para o qual é dependente da presença de colesterol membranar. Modelos de pneumonia aguda em ratinho mostraram que a PLY é essencial para a sobrevivência do pneumococo tanto no trato respiratório superior como inferior, assim como é requerida para a passagem bacteriana dos pulmões para a corrente sanguínea. Além de colonização e doença invasiva, a PLY tem um papel na resposta imunitária protetora. Tem sido proposta como forte candidata a vacina independente de serotipos.

Nesta tese, demonstramos que a manipulação da energia livre de *folding* (ΔG) de mRNA 5' por introdução de mutações sinónimas (silenciosas) permite alterações projetadas da tradução de PLY pelo pneumococo sem necessidade de indutores químicos ou promotores heterólogos. Então, criámos um painel de estirpes isogénicas diferindo apenas em poucas mutações silenciosas no terminal 5' do mRNA de *ply* que se prevê alterem a ΔG por uma grande margem.

In vitro, tal manipulação permite controlo como reóstato da produção de PLY e altera a citotoxicidade de *S. pneumoniae* em células humanas primárias ou imortalizadas. Demonstramos que a produção de PLY aumenta devido a início de tradução acelerado com estabilidade e decaimento de mRNA inalterados. Virulência *in vivo* foi avaliada usando o modelo em ratinho no qual mostramos resultados diferenciados de pneumonia pneumocócica de acordo com ΔG do mRNA 5'. Marcadores de inflamação em histologia pulmonar e perfil de citocinas proinflamatórias ambos correlacionam positiva e significativamente com a expressão de PLY.

Esta estratégia de controlo reóstato de fatores de virulência é útil e providencia prova de princípio de uma nova ferramenta para a compreensão do papel de fatores específicos durante interações hospedeiro-bactéria. A contínua investigação em manipulação de ΔG do mRNA é uma engenhosa arma no estudo da patogénese.

Lay summary // Resumo para Leigos

English Version

In this thesis, I have studied a toxin (pneumolysin) produced by a bacteria called pneumococcus which causes pneumonia in humans. We mutated its genetic coding to make the bacteria produce more or less toxin as we intended using a new technique. We then tested these mutants in human cells in the laboratory and in mouse models of pneumonia. We saw that increased amounts of toxin were worse for mice and human cells. We can now use this technique to engineer new vaccines against this microbe.

Versão Portuguesa

Nesta tese, estudei uma toxina (pneumolisina) produzida por uma bactéria chamada pneumococo que causa pneumonia em humanos. Nós mutámos o seu código genético para fazer a bactéria produzir mais ou menos toxina conforme quiséssemos usando uma técnica nova. Depois testámos estes mutantes em células humanas no laboratório e em modelos de pneumonia em ratinho. Vimos que doses aumentadas de toxina eram piores para os ratinhos e células humanas. Podemos agora usar esta técnica para desenhar novas vacinas contra este micróbio.

TABLE OF CONTENTS

DECLARAÇÃO	ii
STATEMENT OF INTEGRITY	iii
ACKNOWLEDGMENTS // AGRADECIMENTOS	v
Financial support	vii
ABSTRACTS	ix
Abstract	xi
Resumo	xiii
Lay summary // Resumo para Leigos	xv
TABLE OF CONTENTS	xvii
LIST OF FIGURES	xxi
ABBREVIATIONS AND GLOSSARY	xxiii
CHAPTER I. GENERAL INTRODUCTION	1
Background and State of the Art	1
I.I. <i>Streptococcus pneumoniae</i>	1
Pneumococcal virulence factors	4
Pneumolysin	5
Host immunity against pneumococcus and PLY	11
Animal models of pneumococcal infection	12
Vaccines against pneumococcus: old and new	14
I.II. Translation determinants	17
Codon usage bias	21
CAI of PLY in <i>S. pneumoniae</i>	22
Codon pair bias	23
mRNA folding free energy	24
Computational tools for nucleic acids	26
Figures	27
Research Problems	29

CHAPTER II. Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by <i>Streptococcus pneumoniae</i>	31
Author Contributions.....	31
Acknowledgments	31
CHAPTER II. Abstract.....	32
CHAPTER II. Introduction	33
CHAPTER II. Results.....	35
Production of PLY increases with increasing ΔG	35
Translation kinetics drive the relationship between ΔG and protein abundance .	36
<i>ply</i> mRNA ΔG affects host-pathogen interactions <i>in vitro</i>	37
CHAPTER II. Discussion	38
CHAPTER II. Materials and Methods	40
Ethics statement.....	40
Calculation of ΔG and codon adaptation index.....	40
Bacterial strains, primers, plasmids, transformation and growth conditions.....	40
Western blotting and ELISA.....	41
qRT-PCR and mRNA half-life.....	42
Hemolysis assay.....	42
Cell culture and cytotoxicity assay	43
Statistical analysis.....	43
CHAPTER II. Figures.....	45
CHAPTER III. Control of <i>S. pneumoniae</i> Virulence <i>in Vivo</i> by Rational Manipulation of mRNA Folding Free Energy	49
Author Contributions.....	49
Acknowledgments	49
CHAPTER III. Abstract.....	50
CHAPTER III. Introduction	51
CHAPTER III. Results.....	53
Distinct PLY expression <i>in vivo</i> according to ΔG	53
Bacterial burden is not correlated with PLY expression modulated by ΔG	53
Immune cell population shift correlates to differential PLY expression	54
Proinflammatory cytokine production correlates to ΔG	55

CHAPTER III. Discussion	56
CHAPTER III. Materials and Methods	58
Ethics statement	58
Calculation of ΔG	58
Bacterial strains, primers, transformation and growth conditions.....	58
Western blotting and ELISA	59
Murine infection model.....	59
Cell profiles by flow cytometry.....	60
Cytokine ELISA	60
Statistical analysis.....	60
CHAPTER III. Figures.....	61
CHAPTER IV. Translation Initiation Drives the Differential Rates of Protein Translation due to 5' <i>ply</i> mRNA Folding Free Energy Differences	65
Author Contributions.....	65
Acknowledgments	65
CHAPTER IV. Abstract	66
CHAPTER IV. Introduction.....	67
CHAPTER IV. Results	69
Increased production of PLY is due to accelerated translation initiation	69
CHAPTER IV. Discussion	71
CHAPTER IV. Materials and Methods	74
Calculation of ΔG	74
Bacteria, primers, and plasmids	74
<i>In vitro</i> translation	74
Statistical analysis.....	75
CHAPTER IV. Figures	77

CHAPTER V. GENERAL DISCUSSION	79
5' mRNA ΔG correlates with PLY expression <i>in vitro</i>	79
Control of pneumococcal virulence <i>in vivo</i> by 5' <i>ply</i> mRNA ΔG	82
Translation initiation drives the differential rates of protein translation	83
PLY and PLY toxoids as vaccine candidates	85
CONCLUSIONS	87
Future directions and contributions	89
GENERAL REFERENCES	91
CHAPTER VI. APPENDICES	109
A. <i>Curriculum Vitae</i>	111
Academic background	111
List of publications (6)	111
B. Lists of bacterial strains created	113
<i>Streptococcus pneumoniae</i> strains	115
<i>Escherichia coli</i> strains	116
C. Predicted mRNA folding secondary structures	117
D. Growth curves of <i>S. pneumoniae</i>	123
E. The limits of PLY production with increasing ΔG	125
F. Codon optimization of PLY	127
G. <i>Drosophila melanogaster</i> model for rheostat pneumolysin expression	131
H. <i>Caenorhabditis elegans</i> model for rheostat pneumolysin expression	133
I. Published article: "Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by <i>Streptococcus pneumoniae</i> "	135

LIST OF FIGURES

Figure I.1. Depiction of PLY domains undergoing conformational changes.	27
Figure I.2. Pneumolysin displaying the domains and significant residues.	28
Figure II.1. Production of PLY increases with increasing <i>ply</i> 5' mRNA folding free energy.	45
Figure II.2. Engineered differences in 5' mRNA ΔG do not alter <i>ply</i> mRNA quantity.	46
Figure II.3. Differences in PLY production do not correlate with mRNA decay kinetics.	46
Figure II.4. Differential PLY production by <i>S. pneumoniae</i> strains is directly related to PLY production.	47
Figure II.5. Lysis of human A549 respiratory epithelial cells as measured by LDH release correlates with predicted <i>ply</i> 5' mRNA ΔG	47
Figure III.1. (A) PLY levels of <i>S. pneumoniae</i> D39 strains created for this study, used for <i>in vivo</i> experiments, determined by Western blot. (B) Detection of PLY by ELISA of BAL fluid.	61
Figure III.2. Proinflammatory capacity of <i>S. pneumoniae</i> D39 increases with PLY production in a mouse model of upper airway infection.	62
Figure III.3. Proinflammatory capacity of <i>S. pneumoniae</i> D39 increases with PLY production in a mouse model of upper airway infection.	63
Figure IV.1. Production of PLY increases due to accelerated translation initiation.	77
Figure IV.2. Model for translation initiation delay when modulating 5' <i>ply</i> mRNA ΔG . .	78
Figure VI.1. Predicted 5' mRNA folding secondary structures.	117
Figure VI.2. Growth curves and doubling times of <i>S. pneumoniae</i> R6 strains created for this study.	123
Figure VI.3. Representative complete growth curve for WT <i>S. pneumoniae</i> R6.	124
Figure VI.4. Highest values of mRNA folding free energies do not correlate with maximal protein expression.	126
Figure VI.5. Codon optimization and deoptimization with regards to mRNA folding free energies.	129
Figure VI.6. Representative <i>D. melanogaster</i> infection results.	131
Figure VI.7. Representative <i>C. elegans</i> infection results.	133

ABBREVIATIONS AND GLOSSARY

ΔG	(mRNA) folding free energy
ABC	ATP-binding cassette
ATCC	American Type Culture Collection
BAL	Bronchoalveolar lavage
bp	Base pair(s)
CAI	Codon adaptation index
CDC	Cholesterol-dependent cytolysin(s)
CdO	Codon deoptimization(ed)
CFU	Colony forming unit(s)
CO	Codon optimization(ed)
CP	Codon pair(s)
<i>cps</i>	Capsule gene or locus
CS	Cigarette smoke
CSP	Competence stimulating peptide
CUB	Codon (usage) bias
CXC	One of the main chemokine subfamilies
DC	Dendritic cells
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
FACS	Fluorescence-activated cell sorting, applied in flow cytometry
GFP	Green fluorescent protein
HEG	Highly expressed genes
IFN	Interferon
IL	Interleukin
IUPAC	International Union of Pure and Applied Chemistry
KO	knockout
LB	Lysogeny broth
LDH	Lactate dehydrogenase
M Φ	Macrophages

MAPK	Mitogen-activated protein kinase(s)
MEM	Minimum essential medium
MHC	Major histocompatibility complex
NALT	Nasopharynx-associated lymphoid tissue
NK	Natural killer cells
<i>NLRP3</i>	NALP3 (NACHT, LRR and PYD domains-containing protein 3) gene
nt	Nucleotide(s)
ORF	Open reading frame
PCR	Polymerase chain reaction
PCV13	Pneumococcal conjugate vaccine (Prevnar 13®)
PFT	Pore-forming toxin(s)
PHEG	Predicted highly expressed genes (HEG)
PLY	Pneumolysin
PMN	Polymorphonuclear leukocytes
PPSV23	Pneumococcal polysaccharide vaccine (Pneumovax®)
PsaA	Pneumococcal surface antigen A
RBS	Ribosomal binding site
RFU	Relative fluorescence units
(*)RNA	(* m, messenger; s, small non-coding; t, transfer) Ribonucleic acid
RSCU	Relative synonymous codon usage
SEM	Standard error of the mean
tAI	tRNA adaptation index
Tm	Melting temperature(s)
TNF	Tumor necrosis factor
TSB	Tryptic soy broth
UTR	Untranslated region
WT	Wild-type

CHAPTER I. GENERAL INTRODUCTION

This chapter provides the motivation for this work, states the questions driving the research, and also reviews background information and literature on pneumococcus, one of its major virulence factors, pneumolysin, and translation determinants such as mRNA minimum folding free energy and codon usage bias.

Chapters II, III, and IV report the obtained results in manipulation repercussions of 5' mRNA folding free energy and mechanism, along with the respective discussions. The general discussion and conclusion of this work are in chapter V, which includes all the references for this thesis. The appendices contain helpful information for understanding the overall work and comprise unpublished results that delve into technical limitations and other models of infection.

Background and State of the Art

I.1. *Streptococcus pneumoniae*

S. pneumoniae is commonly known as the pneumococcus. It is a gram-positive alpha-hemolytic microaerophilic bacterium (Dockrell et al., 2012; Hajaj et al., 2012; Paton, 1996).

The pneumococcus has the ability to become naturally competent for transformation (Hakenbeck, 2000; Piotrowski et al., 2009). This means it can uptake foreign DNA material and incorporate it in its own genome. It secretes a competence stimulating peptide (CSP), a 17 amino acid peptide pheromone that, upon binding, triggers a cascade of signaling events, ultimately leading to upregulation of genes encoding proteins that are involved in DNA uptake and transformation (Zhu and Lau, 2011).

Because of this, it played a central role in demonstrating that genetic material consists of DNA. In 1928, Griffith turned a harmless pneumococcus into a lethal form by co-inoculating heat-killed virulent pneumococci along with the live pneumococci into a mouse (Lederberg, 1994). In 1944, Avery, MacLeod, and McCarty demonstrated that the transforming factor in Griffith's experiment was DNA, not protein as was widely believed at the time (Avery et al., 1944). Avery's work marked the birth of the molecular era of genetics.

As a laboratory tool, pneumococcus allowed numerous discoveries and breakthroughs. Its value is paralleled only by its clinical importance. It is one of the most important pathogens of humans and the most common cause of bacterial pneumonia.

S. pneumoniae induces the release of proinflammatory and procoagulative agents causing fatal tissue injury during the course of infection (Lüttge et al., 2012).

It leads the list of bacterial causative agents of potentially fatal diseases such as sepsis and meningitis (Marriott et al., 2008; Paton, 1996). Pneumococcus is also associated with other diseases from disorders related to the upper airway such as acute sinusitis and acute otitis media to systemic disorders such as bacteremia, osteomyelitis, septic arthritis, and also peritonitis, endocarditis, and pericarditis. The majority of disease is non-bacteremic with a 3:1 bacteremic pneumonia ratio having been documented (Feldman and Anderson, 2014). Pneumococcus is also a major cause of human ocular infections commonly isolated in cases of keratitis, and its major toxin plays a central role (Taylor et al., 2013).

Nonetheless it is present in the normal bacterial flora of the upper respiratory tract of a lot of people, more frequently in children. Because of the well-known colonizing capabilities it is said to be present in virtually every human. Carriage numbers are hard to quantify globally, whereas the numbers of disease are more often reported.

Current literature defines the transmission of pneumococcus from person to person via respiratory droplets but some authors have shown *S. pneumoniae* to be desiccation tolerant and infectious upon rehydration (Walsh and Camilli, 2011), which make it suitable for transmission by environmental surfaces (fomites). This route of infection has been linked to the spread of other respiratory pathogens.

Coinfection is a real issue. Pneumococcus can synergize with other bacteria and viruses, affecting its growth and virulence. In a very common association, *S. pneumoniae* synergizes with nontypeable *Haemophilus influenzae* to induce inflammation thus causing disease (Lim et al., 2008; Lysenko et al., 2005; Ratner et al., 2005). Coinfection has also been shown with the influenza virus (Vernatter and Pirofski, 2013) and occurs frequently during pandemics.

Throughout this work, we have used two main strains of pneumococcus: *S. pneumoniae* D39, capsular serotype 2, and a laboratory attenuated derivative *S. pneumoniae* R6. The D39 strain is a clinical isolate from 1916, used in the previously mentioned experiments by Avery and coworkers (Lanie et al., 2007), and high in prevalence (Kalin, 1998). The R6 strain was created in a laboratory around 1964, and it was derived from D39 therefore sharing a very similar genetic background (Lanie et al., 2007; Tomasz et al., 1964). It is avirulent and lacks the characteristic capsule due to a deletion in the *cps* locus, the major difference between strains. In addition, the R6 strain contains some single-base-pair mutations (71 changes) and a total of 6 deletions and 4 insertions (Lanie et al., 2007).

Pneumococcal virulence factors

The pathogenicity of pneumococcus is influenced by a number of important virulence factors (Jedrzejewski, 2001). One classic way to study their effects in pathogenesis is by means of deletion/knockout (Berry and Paton, 2000; Mitchell, 2000). The complete genome sequence of a virulent strain has shed some light on the complexity of pneumococcal physiopathogenesis and metabolism (Tettelin et al., 2001).

Virulence factors are a heterogeneous group of molecules. The major pneumococcal virulence factor, pneumolysin (PLY), is a toxin and will be described in detail in the following sections.

Biofilm formation and agglutination are virulence factors that prevent detection of pneumococcus by the immune system. Furthermore, pneumococcus can evade host defenses by two additional virulence factors, a thick polysaccharide capsule, which usually defines its serotype, and cell wall peptidoglycan, which is covalently bound to the capsule (Eberhardt et al., 2012). The cell wall protects the bacterium from mechanical damage and from osmotic rupture. The capsule may avoid cell entrapment in the nasal mucus and prevent effective opsonophagocytosis. Capsular polysaccharides play critical roles in bacterial survival strategies, and have important medical properties (Yother, 2011) such as immunization. We are currently aware of about 93 pneumococcal capsular serotypes (Kalin, 1998).

A characteristic property of pneumococcus is that it can spontaneously undergo autolysis when reaching the stationary phase of growth. This is mediated by virulence factors known as autolysins, such as the ubiquitous autolysin A (LytA), an amidase that cleaves pneumococcal peptidoglycan in the cell wall leading to cell lysis. LytA allows for cell wall turnover and it is unclear if it causes the release of PLY (Balachandran et al., 2001; Martner et al., 2008; Price et al., 2012). This capability may allow for escape from phagocytosis (Martner et al., 2009). Pneumococcus has the ability to change the opacity of its colonies, a poorly understood spontaneous phase variation, which is independent

of capsule expression or bacterial chain length, but rather may be connected to premature autolysis in transparent colonies (Weiser et al., 1994).

Many pneumococcal virulence factors are surface proteins that influence the interaction of the bacterium with its host. Pneumococcal surface antigen A (PsaA) is a component of the ABC transport system, which is involved in resistance to oxidative stress. Neuraminidase NanA, β -galactosidase (BgaA), and β -N-acetylglucosaminidase (StrH) act sequentially to cleave terminal sugars from human glycoconjugates, which may unmask receptors for adherence (Kadioglu et al., 2008). Pneumococcal surface protein A and C (PspA and PspC) play a role in virulence alone or in combination with other factors (Ogunniyi et al., 2007). PspA protects against binding of human complement and PspC, also known as choline-binding protein A, helps with translocation across the epithelium (Kadioglu et al., 2008). Hyaluronate lyase breaks down extracellular matrix components. Two-component systems are mechanisms by which organisms both sense the environment and respond to the stimuli. The majority of these have been linked to virulence in pneumococcus (Paterson et al., 2006).

The list of pneumococcal virulence factors keeps increasing as more insights are attained from new tools, methods, and studies becoming available.

Pneumolysin

Pneumolysin (PLY) is a major virulence factor with small genetic variability that is produced by virtually all pneumococcal serotypes. It was first described in the 1930s, and many biochemical, structural, and mutagenesis studies have been published since (Rosjohn et al., 1998).

Cholesterol-dependent cytolysins (CDCs) are a family of pore-forming toxins (PFTs) that exists in more than 20 species of bacteria, mainly gram-positives. Pneumococcus produces a toxin, PLY, which belongs to this family (Kadioglu et al., 2008; Marriott et al., 2008; Tweten, 2005). PLY is activated by thiols and is sensitive to inhibition by cholesterol like the others in the group of CDCs.

The CDCs are able to form extraordinarily large pores (200-400 angstrom) in the membrane of attacked cells, and for this, they depend on the presence of membrane cholesterol (Tilley et al., 2005; Tweten, 2005). See Figure I.1.

To form a pore, soluble monomers first interact with the membrane surface via their C-terminal Immunoglobulin-like domain 4 as suggested by X-ray crystallography, biophysical studies, and single particle cryo-electron microscopy experiments (Reboul et al., 2014). Membrane bound oligomers can then assemble into prepores, rearrange and finally form the pore (Gilbert et al., 1999b). While several studies have granted some insights into the protein behavior during this process, the exact conformational changes that occur are still poorly understood. See Figure I.1.

PLY is produced as a 52-53 kDa soluble monomeric protein that oligomerizes (~38-mer) in the membrane of target cells (Kadioglu et al., 2008). PLY is a four-domain protein (471 amino acids) that has a highly conserved undecapeptide in its domain 4 (D4) mediating host cell binding (Tweten, 2005).

PLY was initially thought to be localized not on the membrane but in the cytoplasm (Johnson, 1977). However, it has been shown more recently that it localizes to the cell wall (Price and Camilli, 2009), potentially on the external portion as it is accessible to extracellular proteases. How PLY is released from pneumococcal cells is still unclear. Known secretory pathways have been excluded as mechanisms of release, and PLY, unlike all other members of the CDCs, lacks a signal peptide for export outside the cell (Johnson, 1977; Price et al., 2012; Walker et al., 1987). Pneumococcus is capable of autolysis in stationary phase through activation of LytA, as previously mentioned; it has been hypothesized that this releases PLY into the environment (Martner et al., 2008; Paton et al., 1993). However, it is unclear if autolysis is necessary for the release of PLY as another study showed that autolysis is not required for PLY release (Balachandran et al., 2001). There is some evidence that there may be a PLY export mechanism coupled to the cell wall localization of the protein (Balachandran et al., 2001; Price and Camilli, 2009). The precise export mechanism remains uncharacterized but is known to depend on domain 2 of PLY (Price et al., 2012).

In addition to pore-formation to cause lysis of erythrocytes, immune cells, and epithelial cells (Martner et al., 2008), PLY is also able to activate complement (Alexander et al., 1998; Berry et al., 1995; Kadioglu et al., 2008; Paton et al., 1993; Tweten, 2005). Pneumococcal strains that expressed mutated PLY proteins which lacked hemolytic and complement-activating activity, were shown to be more virulent than a strain lacking PLY altogether (Kadioglu et al., 2008), thus revealing the existence of previously unknown virulence characteristics (Berry et al., 1999) and independent immunomodulatory effects (Khan et al., 2014).

Currently, there are studies that show that PLY, directly or not, is capable of:

1. being highly immunogenic (Lock et al., 1988) and therefore an important serotype-independent vaccine candidate (Alexander et al., 1994; Cockeran et al., 2005; Denoel et al., 2011; Kadioglu and Andrew, 2004; Malley et al., 2005; Moffitt and Malley, 2011; Tai, 2006);
2. upregulating prostaglandin E₂ and leukotriene B₄ release from polymorphonuclear leukocytes (PMNs) (Cockeran et al., 2001);
3. recruiting PMNs via transendothelial migration *in vitro* (Moreland and Bailey, 2006) and upregulating IL-6, IL-1 β , TNF- α , and KC expression combined with a M Φ and PMN influx *in vivo* (Dessing et al., 2009), and MIP-2 requiring the toxin's cytotoxic properties for PMN influx (Rijneveld et al., 2002);
4. activating toll-like receptors – TLR2 and TLR4 (Branger et al., 2004; Dessing et al., 2008; Dessing et al., 2009; Malley et al., 2003; McNeela et al., 2010; Srivastava et al., 2005; van Rossum et al., 2005) – enhancing the secretion of proinflammatory cytokines IL-1 α , IL-1 β , IL-6, IL-12, IL-23, and TNF- α by DC and IL-17A, and IFN- γ by splenocytes (McNeela et al., 2010); PLY-KO strains can infect TLR2-KO mice but not WT mice (Dessing et al., 2008) and it is argued that TLR4 can compensate even though *in vivo* response may involve both TLR2 and TLR4 (Dessing et al., 2009);
5. inducing IFN- γ and IL-17A in the lungs *in vivo*, independently of TLR4 (McNeela et al., 2010) – loss of TLR4 caused a diminished response to PLY, and TLR4-KO mice were more susceptible to lethal infection after intranasal colonization (Malley et al., 2003) with others finding that TLR4 plays a small role for this immune response to PLY (Branger et al., 2004; McNeela et al., 2010);

6. inducing Th1 proinflammatory response, with high levels of IFN- γ , IL-12, and IL-17 and low levels of IL-10 and IL-13, in a clinical study of human CD4⁺ T cells (Mureithi et al., 2009);
7. inducing IL-1 β *in vitro* (potassium influx-dependent), phagosome rupture, and inducing IL-6 *in vivo* (Los et al., 2013);
8. activating mitogen-activated protein kinases (MAPK) p38 (Ratner et al., 2006) and activation of JNK and ERK MAPKs *in vitro* (Aguilar et al., 2009) – interestingly, without specifying PLY, pneumococcus was known to induce p38 MAPK and JNK-mediated caspase-dependent apoptosis in human endothelial cells (N'Guessan et al., 2005);
9. inducing expression of MAPK phosphatase 1 (MKP1) via a TLR4-dependent MyD88-TRAF6-ERK signaling pathway, leading to upregulation of mucin production *in vivo* (Ha et al., 2008);
10. activating NLRP3 inflammasome (McNeela et al., 2010);
11. activating caspase-1 and inducing depending cytokines *in vitro* (Shoma et al., 2008);
12. activating type I IFN signaling mediated by pneumococcal DNA (Parker et al., 2011);
13. inducing nitric oxide production from M Φ (Braun et al., 1999);
14. facilitating peptidoglycan introduction into host cells, as well as inducing host signaling from osmotic changes both *in vitro* and *in vivo* (Lysenko et al., 2007; Ratner et al., 2007; Ratner et al., 2006);
15. breaching the endothelial layer, allowing *S. pneumoniae* to pass the blood-brain barrier (Hirst et al., 2004; Orihuela et al., 2004);
16. exerting direct toxic effects on the alveolus-capillary barrier leading to severe pulmonary hypertension in mice (Maus et al., 2004; Witzentrath et al., 2006);
17. inducing pulmonary edema by gap formation between epithelial cells in mouse lungs *in vitro* and *ex vivo* (Witzentrath et al., 2006);
18. promoting the initial steps of sepsis-causing infection establishment (Benton et al., 1995).

These effects occur mainly at high concentrations. However, even at lower concentrations that are sublytic, PLY is capable of triggering the lethal effects of pneumonia (del García-Suárez et al., 2007).

Studies show that low-dose PLY is known to:

19. induce apoptosis (del García-Suárez et al., 2007; Iliev et al., 2007; Wippel et al., 2011) *in vivo* in mouse airway epithelial cells through direct interaction with TLR4 but not TLR2 (Srivastava et al., 2005);
20. activate NLRP3 inflammasome and promote TLR4-mediated release of proinflammatory cytokines (Iliev et al., 2007; Khan et al., 2014);
21. induce the activation of the NADPH oxidase in human PMN (Martner et al., 2008);
22. potentiate the proinflammatory activities of PMN and MΦ and recruit them to the infection site (Cockeran et al., 2002; Jounblat et al., 2003);
23. produce rapid activation of Rho and Rac GTPases and formation of actin stress fibers, filopodia, and lamellipodia (Iliev et al., 2007).

A study by Rogers et al. further illustrates the extensive effects PLY has on host cells. They evaluated the expression of thousands of genes *in vitro* in a THP-1 human monocytic cell line after exposure to PLY, and found that the expression of 142 genes changed in a PLY-dependent manner, whereas the gene expression of 40 was responsive in a PLY-independent manner (Rogers et al., 2003).

PLY has 4 distinct domains that exert diverse functions. Some undergo conformational changes particularly during pore-formation. These specific parts have been studied separately. Case in point, Domain 4 (D4PLY) has been shown to activate Th17 cells in NALT (Gray et al., 2014). See Figure I.2.

Being a double-edged sword, PLY may play a role in colonization as well as in pathology, and even co-colonization. A study has shown that colonization with the PLY-sufficient pneumococcal strains correlates with attenuated recruitment and function of pulmonary inflammatory cells, which protects mice from influenza virus-induced disease (Wolf et al., 2014).

Little is known about the effects of PLY exposure to environmental elements (such as CS). CS condensate has been shown to exert effects on PLY, specifically attenuating the pore-formation activity which may favor colonization (Mutepe et al., 2013).

Host immunity against pneumococcus and PLY

The host inflammatory response also contributes to the pathogenesis of pneumococcal pneumonia. In this way, therapies that modulate it could help contain the damage sustained from inflammation (Vernatter and Pirofski, 2013) to improve outcomes.

During infection, NLRP3 is required for protective immunity against pneumococcus and PLY (McNeela et al., 2010; Witzzenrath et al., 2011). IL-1 β has been shown to be required for resistance to *S. pneumoniae* infection in mice (Kafka et al., 2008). In another pneumococcal pneumonia model, IL-1 β was shown to be a major part of a response plan by M Φ and epithelial cells that facilitated CXC chemokine expression and recruited PMNs (Marriott et al., 2012), a feature of pneumococcal disease. Furthermore in mice, initial resistance to infection with PLY-deficient strains appears to be independent of IL-1 β or IL-6 (Benton et al., 1998), which supports the effects of PLY on cytokine stimulation described above. The inflammasome is critical for activating caspase-1 and host innate resistance to pneumococcal infection (Fang et al., 2011).

TLR receptors play an important role in efficient bacterial clearance, although some authors argue the higher importance of TLR2 in a PLY-independent manner (Knapp et al., 2004), over the less significant TLR4 (van Rossum et al., 2005).

The inflammatory process upon the resolution of pneumococcal infection is orchestrated by $\gamma\delta$ T cells that regulate pulmonary DCs and alveolar M Φ (Kirby et al., 2007). *S. pneumoniae* creates an influx of PMN and DC whereas the resident M Φ diminish, which may resolve the infection, but subsequent re-establishment of alveolar and lung homeostasis post-infection is complex. One group has observed and suggested an important role of exudate in this situation but without apparent contribution from resident M Φ population (Taut et al., 2008).

One exemplar state of impairment of local host defenses is attributable to the effects of cigarette smoke (CS) on the host of a severe pneumococcal disease (Butler et al., 2000;

Mutepe et al., 2013). Even though CS effects on humans have been extensively studied (Sopori, 2002), less is known about the direct effects of smoke exposure on airway pathogens, or their virulence factors. Our previous work (Kulkarni et al., 2012) and that of others have shown effects of CS on biofilm formation by *Streptococcus pneumoniae* (unpublished data) and *Staphylococcus aureus* that resulted in significant dose-related augmentation of biofilm formation (Kulkarni et al., 2010; Mutepe et al., 2013). On the host side, CS exposure impairs pulmonary pneumococcal clearance and complement-mediated phagocytosis in alveolar M Φ (Phipps et al., 2010). Pneumococcal adherence was significantly upregulated in a rat model of chronic CS exposure (Ozlu et al., 2008) while acute exposure did not show similar results.

Animal models of pneumococcal infection

To study *in vivo* effects of *S. pneumoniae* or its toxin, pneumolysin, a large number of animal models have been developed and used, each with its own advantages and limitations. Traditionally, mouse, rat, and rabbit models have been used to recreate the diseases that pneumococcus induces in humans (Chiavolini et al., 2008). With enhanced genetic toolkits and ease of use (cost and time), fruit flies and worms have also been used.

In certain host models, PLY-KO strains, which are less virulent, grow at lower rates *in vivo* than strains with PLY. This indicates that effects of PLY on the host may stimulate growth directly and/or tissue damage provides conditions favorable for *S. pneumoniae* growth (Los et al., 2013). Mouse models of acute pneumonia have shown that PLY is essential for the survival of *S. pneumoniae* in both the upper and lower respiratory tracts, as well as for the spreading of bacteria from the lungs to the bloodstream (Kadioglu et al., 2008).

As a model for disease, mice have nasopharynx-associated lymphoid tissue (NALT) that is analogous to our Waldeyer's ring. PMN and pneumococcal interplay, aided by PLY, leads to the availability of microbial products for antigen delivery that, in part, promotes mucosal clearance (Matthias et al., 2008). Our colleagues have seen increased

pneumococcal nasal colonization in an IFN-deficient mouse model that indicated a role for IFN signaling in the mucosal response to infection (Parker et al., 2011).

In a mouse model of acute otitis media caused by pneumococcus, the two complement pathways play a critical role in host defense (Li et al., 2012). The classical pathway is generally more important against septicemia however, in this model, these authors have concluded that the transparent opacity phenotype of pneumococcus was killed by an alternative pathway-dependent process while the opaque variant was able to evade complement-mediated opsonophagocytosis and was associated with increased virulence (Li et al., 2012).

An interesting *ex vivo* model of murine blood-free perfused lungs, when ventilated and exposed to intravascular PLY, revealed a dose-dependent increase in vascular resistance by vasoconstriction effects (Witzenrath et al., 2011).

In pneumococcal meningitis models PLY causes part of the permanent neurological damage as PLY-KO strains have been shown to be less virulent in mouse model for meningitis (Wellmer et al., 2002). It has been found that a reduction of extracellular calcium strongly enhanced PLY pore-formation due to increased membrane binding (Wippel et al., 2011) in brain tissue. PLY leads to extensive stabilization of microtubules leading to axonal transport inhibition and neuropathy (Iliev et al., 2009). Similarly to what was seen after establishment of sepsis (Benton et al., 1995), PLY did not affect bacterial growth in the cerebrospinal fluid after intracisternal injection in a rat meningitis model (Reiss et al., 2011) but, likewise, it appears to be necessary for pneumococcus to cross the blood-brain barrier (Hirst et al., 2004; Orihuela et al., 2004). Afterwards, PLY seems not to be necessary for systemic spreading (Wellmer et al., 2002). Beyond providing insights on both pathophysiology and immunology upon infection, animal models have been excellent tools for vaccine development.

Vaccines against pneumococcus: old and new

While the availability of antibiotics has revolutionized treatment, they are becoming more and more ineffective due to the more frequent resistance genes. Consequently, pneumococcal infections still have high mortality rates, especially in risk groups (Chiavolini et al., 2008).

Prevention is key, but current vaccines have limitations. Vaccines that confer protection against pneumococcus are somewhat ineffective due to incomplete serotype coverage, absence of long lasting immunity, and low mucosal immunity (Assaad et al., 2012; Kadioglu et al., 2008).

Currently, at our disposal, we have two vaccines at our disposal: the pneumococcal conjugate vaccine (PCV13 or Prevnar 13[®]) and the pneumococcal polysaccharide vaccine (PPSV23 or Pneumovax[®]). The numbers indicate how many serotypes they protect against.

The Centers for Disease Control and Prevention institution recommends:

- PCV13 for all children at 2, 4, 6, and 12 through 15 months and adults 65 years or older;
 - risk groups from 19 years or older;
- PPSV23 for all adults 65 years or older;
 - risk groups from 2 years or older;
 - adults from 19 through 64 years old who smoke cigarettes or who have asthma.

Revaccination policies vary. We know that antibody levels decline after 5-10 years following vaccination with PPSV23 (Mufson et al., 1983; Mufson et al., 1987) indicating a need for new dose after 5 years.

Acquired immunity to *S. pneumoniae* is mediated mainly by major histocompatibility complex (MHC) class II-positive, CD4⁺ T cells, putatively through the effects of PLY (Kadioglu and Andrew, 2004; Malley et al., 2005; van Rossum et al., 2005) but it was thought to be mediated by B-cell production of antibodies against its capsular polysaccharides. Curiously, neither of the main functions of PLY has effect on B cells

(Jounblat et al., 2003) even though the absence of PLY caused reduced recruitment of T and B lymphocytes *in vivo* (Kadioglu et al., 2000). In a clinical study of human CD4⁺ T cells, exposure to PLY caused a Th1 proinflammatory response (Mureithi et al., 2009).

PLY is highly immunogenic (Lock et al., 1988) and a strong candidate for a serotype-independent vaccine as well as a drug target for the prevention and treatment of pneumococcal disease (Alexander et al., 1994; Cockeran et al., 2005; Denoel et al., 2011; Kadioglu and Andrew, 2004; Malley et al., 2005; Moffitt and Malley, 2011; Tai, 2006).

PLY has been demonstrated to be a protective monovalent detoxified recombinant derivative vaccine by itself as well as in conjugation with other protein candidates (Briles et al., 2003; Denoel et al., 2011; Goulart et al., 2013; Hu et al., 2013; Kamtchoua et al., 2013; Oloo et al., 2011; Pauksens et al., 2014; Salha et al., 2012) or when associated to existing vaccines (Sanders et al., 2013). In addition, domains of PLY, such as D4PLY, have been suggested as candidates for intranasal immunization (Gray et al., 2014). Several of these combinations have been shown to be safe for use in phase I (Berglund et al., 2014) and phase II randomized clinical trials (Leroux-Roels et al., 2014; Prymula et al., 2014).

Authors have observed high levels of mucosal antibodies to PLY, which are associated with lower acute otitis media occurrence in young children (Xu et al., 2015). A group studied monovalent recombinant protein vaccine candidates in an infant mouse model and showed protection from bacteremic pneumonia and sepsis mortality after vaccination with a detoxified PLY derivative and reduced lung bacterial burdens after immunization with pneumococcal histidine triad D, and pneumococcal choline-binding protein A (Verhoeven et al., 2014).

Another study, however, argued that a PLY-based vaccine could be beneficial as an adjuvant to stimulate cytokine production but would not be effective as primary vaccine component (Lu et al., 2014).

Vaccination and protection against *S. pneumoniae* is a field of prominent research.

I.II. Translation determinants

Life, as we know it, seems to follow a very consistent formula. From the elementary units being encoded in the genetic material of a being it is then transcribed from DNA to RNA. RNA, as a vector of information is called messenger RNA, or mRNA, and it will be the guiding instruction mold for protein production. This step known as translation utilizes transfer RNA (tRNA) bound to amino acids that will be the building blocks of a new molecule, the protein, and ribosomes to actually produce the protein. Far from being over, protein folding and post-translation modifications happen subsequently. Such processes involve many players from a multitude of enzymes that operate on nucleic acids, to transcription factors, and to energy, that in part will also regulate them. Regulation of translation efficiency and accuracy is varied and of very important as a topic in this thesis.

A number of factors determine the efficiency and rate of protein translation (Gingold and Pilpel, 2011). It appears that, depending on the specific situation, the significance of each of these factors can vary. Gene function and ontological classifications strongly influence translation revealing evolutionary optimization (Ciandrini et al., 2013) that may be explained by the interplay of the factors that modulate protein expression. Among them are:

- mRNA folding free energy (ΔG);
- codon adaptation index (CAI) measuring codon usage bias (CUB);
- codon pair (CP) bias;
- tRNA adaptation index (tAI) concerning codon-tRNA interaction and tRNA availability;
- GC%, which represents the percentage of G and C nt with regards to the 4 possible nt (A, T, C, G), and GC3%, which is that percentage counted only in the 3rd position of the codon.

The genetic code is redundant because there are more codons than amino acids (Crick, 1968). Most amino acids are encoded by 2 to 6 different codons, called synonymous codons. Synonymous or silent changes to a DNA sequence of a protein are often assumed to have no effect on the final protein because synonymous codons encode the same amino acid. Despite this fact, the frequency of synonymous codon usage varies among organisms and even among genes in a single organism (Hershberg and Petrov, 2008; Martín-Galiano et al., 2004). This phenomenon is known as CUB, and it can have significant effects on both rate and fidelity of protein synthesis and even on the overall growth kinetics of organisms (Hershberg and Petrov, 2008; Martín-Galiano et al., 2004). Generally, codon bias involves the selection of the base at the 3rd codon position to adapt the coding sequence to the most abundant tRNAs in the cell or those with more efficient codon-anticodon interaction kinetics (Hershberg and Petrov, 2009). This adaptation may be species-specific and can prove useful when we are studying the horizontal transfer of genes: the most recently acquired genes are likely not completely optimized for the CUB of the recipient genome (Martín-Galiano et al., 2004). Another way to look at CUB and its statistical significance is the codon deviation coefficient (Zhang et al., 2012). More about CUB on page 21.

Synonymous mutations may encode hidden stop codons. One theory suggests evolutionary selective pressure favoring out-of-frame stop codons (Seligmann and Pollock, 2004). This is supported by the idea that ribosomal frameshift, a translational error that alters the amino acid chain by reading an alternate ORF, wastes energy and resources, and may give rise to toxic byproducts (Bertrand et al., 2015).

One hypothesis postulates that a region of slow translation in the beginning of coding sequences is beneficial, and some authors saw some evidence linked to the gene copy number of cognate-tRNAs (Hockenberry et al., 2014). The tAI is a measure of efficiency of the adaptation of a coding sequence to that cell's tRNA pool, even though it rather seems that the tRNA pool adapts to the codon profile of the PHEG than the opposite way (Trotta, 2013). Efficiencies of the different codon-tRNA interactions are expected to vary among different organisms (Sabi and Tuller, 2014). Nonetheless, one group analyzed abundances of tRNA across several bacterial genera and found a consistent

positive correlation challenging the notion that accuracy of translation may play a role in driving CUB (Shah and Gilchrist, 2010).

The overall percentage of the nucleotides guanine and cytosine is often highlighted because, unlike adenine-thymine, GC pairs form three hydrogen bonds, which makes this pairing more stable with a higher bond energy. This way, the folding energy associated with an mRNA molecule varies according to its GC%. GC3% is the percentage of GC in the 3rd position of the codon. This is relevant because synonymous codons generally vary in the wobble position. GC3-related intrinsic DNA flexibility strongly associated with codon bias across 24 different prokaryotic multiple whole-genome alignments (Babbitt et al., 2014).

The mRNA folding free energy or Gibbs free energy is a property that can be predicted and calculated. In biochemical terms, a negative value is associated with the spontaneity of a molecular reaction. More specifically in the biology of mRNA, a lower ΔG means a more tightly folded molecule that is less accessible for translation and that is potentially more susceptible to degradation and less stable *in vivo*. On the other side, a higher ΔG correlates with increased protein expression. In other words, weaker structures correspond to higher protein levels. Many reports have shown that weaker mRNA structures attribute more importance to the CAI (Gingold and Pilpel, 2011; Plotkin and Kudla, 2011; Supek and Smuc, 2010; Tuller et al., 2010a; Tuller et al., 2010b). There appears to be a stronger correlation between mRNA structure and protein levels at the 5' end of the mRNA molecule, which was determined by calculating the ΔG in a sliding window of 42 nucleotides length (Kudla et al., 2009). More about ΔG on page 24.

The current literature suggests that fast-growing bacteria are adapted for fast protein synthesis as these species have a stronger selection pressure on codon usage than slow growing species and possess higher numbers of rRNA and tRNA genes (Ran and Higgs, 2012).

Highly translated genes need to utilize the translation machinery efficiently (Pop et al., 2014). Although these authors have shown a correlation between translation efficiency and the RNA structure, determined both computationally and from structure probing data, they saw no significant correlation with CUB or tAI (Pop et al., 2014).

Initiation efficiency and elongation dynamics have been shown to be responsible for translation outcomes (Ciandrini et al., 2013; Plotkin and Kudla, 2011). In terms of speed of translation, it was also shown that frequently used codons increase the rate of elongation of a protein, but, more importantly, the initiation of translation was shown to be the rate-limiting step (Jacques and Dreyfus, 1990; Kozak, 1999; Kudla et al., 2009; Plotkin and Kudla, 2011; Shah et al., 2013). The translation initiation rate is the rate at which ribosomes start translating the ORF. The initiation of translation involves the recruitment of many factors and strongly depends on many different elements, such as nutrients, stress conditions, cytoplasmic ribosome availability, initiation factors, and secondary structures on the long 5' UTRs (Ciandrini et al., 2013) generally found in bacterial mRNAs. Still, the ribosome binding is not fully understood (Espah Borujeni et al., 2014). Here, an increase in translation speed makes a larger number of ribosomes to be available for the synthesis of other proteins, thus enhancing the growth and fitness of the cell.

Selection for translational speed is a dominant effect in driving CUB in fast-growing bacteria, with selection for accuracy playing a small supplementary role (Ran and Higgs, 2012). However, other authors have argued that accuracy is more important than speed (Ran and Higgs, 2012). Studies have shown that selection on translational accuracy plays a role in CUB of both *D. melanogaster* and *E. coli* (Hershberg and Petrov, 2008).

Transcripts whose translation is regulated by tRNA modifications are key components of the oxidative stress response in budding yeast (Gu et al., 2014). Stress induces increases in tRNAs with specific anticodon wobble bases that are required for the optimal translation of significantly biased stress response transcripts that use degenerate codons set to these modified tRNA bases (Gu et al., 2014). Analogous to the tRNA adaptation, via CUB, proteins can adapt at the amino acid level to further increase their abundance (Tuller et al., 2007) creating another biasing level.

Codon usage bias

It is important to define some common measures pertaining to CUB that are important to establish the CAI of a gene (Sharp and Li, 1986, 1987). Firstly, the relative synonymous codon usage (RSCU) value of a codon is the observed frequency of that codon divided by the expected frequency when all synonymous codons for that amino acid are used equally meaning that values close to 1 indicate a lack of bias for that codon (Martín-Galiano et al., 2004). The normalized version of RSCU is called relative adaptiveness of a codon (w) and it is calculated as the quotient of the RSCU value of a specific codon and the highest RSCU value for codons encoding the same amino acid within the organism being studied (Martín-Galiano et al., 2004). The CAI value of a gene is the geometric mean of the w values from all its codons (Martín-Galiano et al., 2004). CAI values are useful ways to compare genes with regards to how well adapted they are to a particular genome's CUB. It is sometimes useful to calculate the CAI in relation to 1) all genes in an organism, 2) the predicted highly expressed genes (PHEG), which include some of the genes encoding: ribosomal proteins, energetic metabolism, amino acid metabolism, protein synthesis, transport proteins, and others, or 3) the ribosomal proteins, which are known to be highly expressed. Studying CUB using ribosomal genes has limitations and is controversial but also has a lot of value (Hershberg and Petrov, 2012). A single DNA sequence may have a very different CAI value depending on the organism it is in, and in relation to what it is calculated. See "Computational tools for nucleic acids" on page 26.

Codon bias can influence gene expression by optimization of the translation rate or by influencing the probability of amino acid misincorporation. This may confer metabolic advantages by reducing the number of misfolded proteins and energy spent or may simply allow more rapid growth by shortening the time needed to synthesize the required proteins (Kudla et al., 2009). Adherence to CUB for some organisms is inversely correlated with a coding region's average amino acid biosynthetic cost in a fashion that is independent of chemoheterotrophic, photoautotrophic, or thermophilic lifestyle

(Heizer et al., 2006) meaning that a higher use of major codons would be correlated with a lower average amino acid cost.

Overall, it is argued that an increase in the CAI of a gene does not increase its expression; however, it can improve cellular fitness to an extent that depends on its expression level (Kudla et al., 2009). Opposing this concept are some studies in relation to the *Adh* gene in *D. melanogaster* that showed that *in vivo* introduction of unpreferred codons results in reduced levels of alcohol dehydrogenase protein and that reduction of codon bias in this gene results in decreased ethanol tolerance of adult flies (Carlini, 2004; Carlini and Stephan, 2003). On the other hand optimized codon bias in the *Adh* gene leads to an increase in larval alcohol dehydrogenase activity (Hense et al., 2010).

CUB has been successfully used to help over-expression of proteins, for example, porcine pancreatic lipase could be expressed in *P. pastoris* as an extracellular, functional enzyme after codon optimization of the gene's 5'-terminus (Zhao et al., 2014).

On the other hand, CUB has been used for poliovirus attenuation by large-scale deoptimization of codons that led to reduction of protein synthesis rates (Mueller et al., 2006).

CAI of PLY in *S. pneumoniae*

It has been reported that the virulence factors of *S. pneumoniae*, such as pneumolysin, capsule, neuraminidase, autolysin, and others have CAI values of 0.250 to 0.350, and are thus considered medium-expressed genes (Martín-Galiano et al., 2004). There are no apparent explanations as to why this happens, but there are a number of possible reasons: 1) these genes might have been acquired through horizontal gene transfer very recently in evolutionary time; 2) they could be potentially toxic at higher concentrations assuming that increase in CAI would lead to more protein production; 3) synthesizing these genes may be energetically costly, which could impair cells that overexpress them and are in need of quick growth and division. Despite this, some authors have seen that rather than the costs of protein production, fitness disadvantage for silent mutations arose from lacking of beneficial protein (Agashe et al., 2012).

Codon pair bias

Rather than looking at CUB alone, certain authors have found pairs of codons that appear more frequently than others. This bias is independent of codon usage and different for different species (Coleman et al., 2011). Pairing codons also opens up the range of possible combinations up to 61×61 (3721).

Some authors have used deoptimization of CP bias to achieve protein expression reduction *in vivo* of PLY in pneumococcus and of the whole genome of poliovirus (Coleman et al., 2008; Coleman et al., 2011; Mueller et al., 2006). The group has shown in another study that immunization of mice by intranasal exposure to CP deoptimized influenza virus conferred protection against subsequent challenge with WT strain (Mueller et al., 2010).

However, others have seen that choosing underrepresented CPs may lead to unintended increases in CpG and UpA dinucleotide frequencies that attenuate viral replication while finding no effects of either CP bias and dinucleotide frequencies on translation efficiency (Atkinson et al., 2014; Tulloch et al., 2014). The question of CP deoptimization has mainly been addressed in viral systems. Authors have independently manipulated CP bias and CpG and UpA dinucleotide frequencies in picornavirus and echovirus 7. They concluded that dinucleotide frequencies primarily influenced the viral replication ability while they saw no fitness differences for CP bias variations (Atkinson et al., 2014; Tulloch et al., 2014). In fact, a recent study has found CP to be a consequence of dinucleotide bias with CpG playing a major role in deoptimization of protein expression (Kunec and Osterrieder, 2016).

mRNA folding free energy

RNA secondary structure computational prediction is a significant challenge and calculations of associated free energies are even more difficult because of contributions from complex tertiary RNA structures and ligand interactions (Wan et al., 2012). This is of particular importance with less rigidly defined structures, such as mRNAs and non-coding RNAs (Corley et al., 2015). However, they can also be determined experimentally by measuring RNA melting temperatures (T_m) via several methods (Wan et al., 2012). RNA conformation can also be assessed via polyacrylamide gel electrophoresis under native conditions and it allows the determination of the kinetics under various conditions (Woodson and Koculi, 2009).

The T_m is defined as the temperature at which half of the molecules of a double stranded molecule become single-stranded. Because of this it exhibits a relationship with folding energies. RNA structures of low T_m are more dynamic, exhibit lower energetic cost to unfold, are easily accessible to other factors or ribosomes, and are associated with a higher ΔG . Likewise, a high T_m represents a structure of lower ΔG that is chemically more stable but has higher energetic cost to unfold.

The role of T_m in protein translation is a controversial subject. Some authors found that, in endogenous *E. coli* genes, ΔG affects translation efficiency globally rather than in individual genes (Tuller et al., 2010b). That is to say that mRNA folding impacts the overall translation machinery and not so much the translation of a particular message. This illustrates the fact that the ribosome's interactions with mRNA are important and govern its translation rate (Espah Borujeni et al., 2014). In fact, ΔG in the mRNA region from -4 to +38 bases (1 is A in the first AUG) has been shown to most significantly affect gene expression (Kudla et al., 2009). Other groups observed that the influence of the secondary structure ΔG of the 5' leader sequence significantly correlates with translation rates (Ciandrini et al., 2013; Seo et al., 2009). One other group measured genome-wide RNA ΔG s in yeast using experimental and theoretical data, and found that the T_m s from coding versus noncoding RNAs were distinct and mapped to distinctive

cellular functions (Wan et al., 2012). This could represent an adaptation to stress conditions, such as heat shock. In addition, in yeast, the mRNA folding structure seems to be linked to codon bias through a mechanism that partially operates before pre-mRNA splicing (Trotta, 2013).

Computational tools for nucleic acids

The prediction of the secondary structure of single stranded nucleic acids can be done using the mfold web server, with access to RNA and DNA folding and hybridization software (Markham and Zuker, 2005; Zuker, 2003). The UNAFold software package is a collection of programs that simulate folding or secondary structure prediction for single-stranded RNA or DNA combining free energy minimization, hybridization, and melting pathways (Markham and Zuker, 2008). (See Appendix C, “Predicted mRNA folding secondary structures” on page 117.)

The Vienna RNA Website is another collection of tools for folding, design and analysis of RNA sequences (Gruber et al., 2008).

Computational algorithms to calculate the CAI of genes (Puigbo et al., 2008a), to determine the codon usage table of HEG (Puigbo et al., 2008b), and to optimize the codon usage of sequences (Puigbo et al., 2007) have been described. Tables and CUB measures are also centralized in a repository called the Codon Bias Database and they are available for a wide variety of genera, species, and strains (Hilterbrand et al., 2012).

One group designed a tool to alter mRNA secondary structures, increasing or decreasing the minimum free energy of the sequence, using synonymous codons and accounting for CAI and GC% (Gaspar et al., 2013)¹.

¹ At the time of this publication, we had already designed and tested our sequences.

Figures

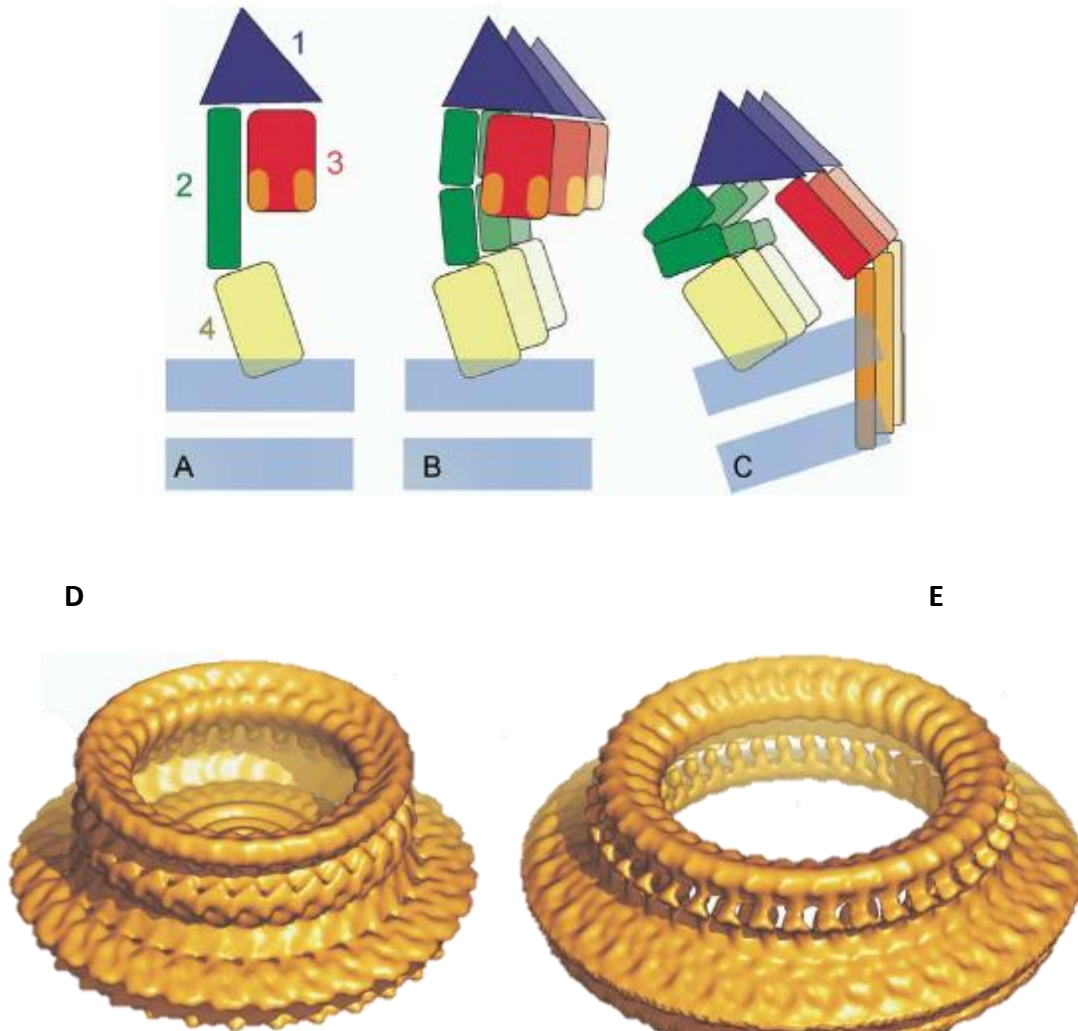


Figure I.1. Depiction of PLY domains undergoing conformational changes.

Cartoons of (A) binding, (B) prepore, and (C) pore-formation. 3D reconstructions of (D) prepore oligomer, 290 Å in diameter, and (E) 38-mer pore with a surrounding ring of membrane, 400 Å in diameter.

Adapted from (Tilley et al., 2005).

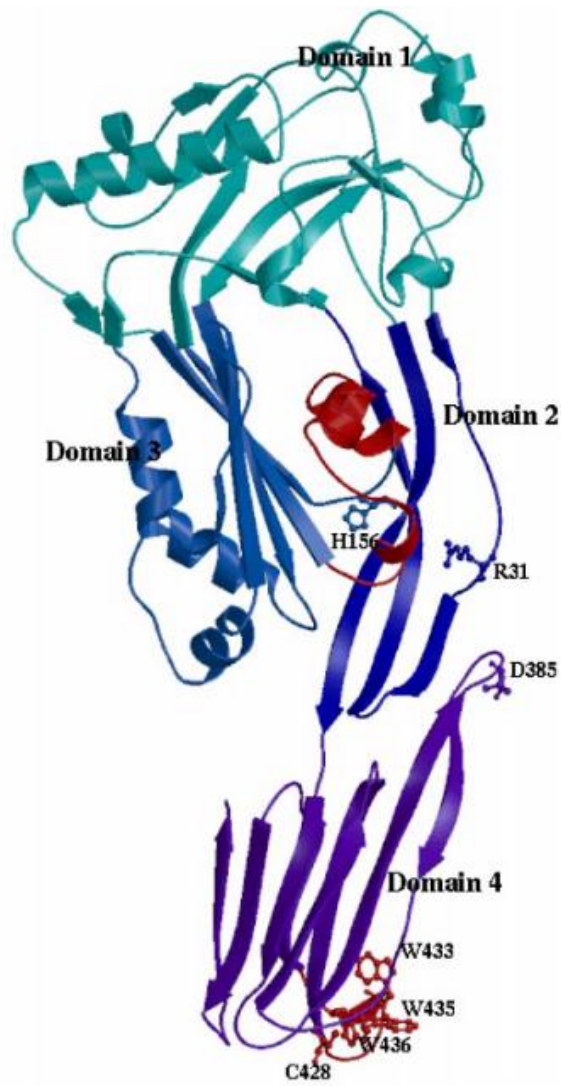


Figure I.2. Pneumolysin displaying the domains and significant residues.

The red colored region at the base of domain 4 is the Trp-rich loop involved in membrane binding.

Taken from (Gilbert et al., 1999a).

Research Problems

This thesis aims to develop a method to achieve rheostat control of a toxin (PLY) produced by pneumococci. This to be achieved without heterologous inducers or inhibitors in a minimalistic fashion that does not alter the produced protein. The main translation determinant implied here is the mRNA folding free energy accounted for the 5' region of the target gene. We want to understand if increasing or decreasing this value with synonymous mutations contained solely in the 5' mRNA can produce a gradient of PLY production across many strains. This system could then be used to study the role of PLY quantity in the pathogenesis pneumococcal pneumonia in a way not possible before. Likewise, this particular method can be useful in the development of new vaccination strategies when we consider the possibility of virulence factor expression fine-tuning. The level of expression of bacterial genes may play a role in pathogenesis, as has been noted in strains of *Vibrio cholerae*, group B *Streptococcus*, *Staphylococcus aureus*, and *Clostridium difficile* that have altered production of toxins or regulatory proteins, leading to hypervirulent phenotypes (Rasigade et al., 2013; Son et al., 2011; Stabler et al., 2008; Whidbey et al., 2013).

CHAPTER II. Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by *Streptococcus pneumoniae*

Fábio E. Amaral, Dane Parker, Tara M. Randis, Ritwij Kulkarni,

Alice S. Prince, Mimi M. Shirasu-Hiza, Adam J. Ratner.

This chapter contains the original research published in March 2015 by PLoS ONE. All of the experiments carried out and described in this chapter were executed by me (published article in Appendix I, on page 135).

Author Contributions

Conceived and designed the experiments: AJR FEA DP TMR ASP MMS. Performed the experiments: FEA DP TMR RK. Analyzed the data: FEA MMS AJR. Contributed reagents/materials/analysis tools: FEA ASP MMS AJR. Wrote the paper: FEA MMS AJR.

Acknowledgments

We thank Max Gottesman and Timothy LaRocca for helpful discussions.

CHAPTER II. Abstract

The contribution of specific factors to bacterial virulence is generally investigated through creation of genetic “knockouts” that are then compared to wild-type strains or complemented mutants. This paradigm is useful to understand the effect of presence vs. absence of a specific gene product but cannot account for concentration-dependent effects, such as may occur with some bacterial toxins. In order to assess threshold and dose-response effects of virulence factors, robust systems for tunable expression are required. Recent evidence suggests that the folding free energy (ΔG) of the 5' end of mRNA transcripts can have a significant effect on translation efficiency and overall protein abundance. Here we demonstrate that rational alteration of 5' mRNA folding free energy by introduction of synonymous mutations allows for predictable changes in pneumolysin (PLY) expression by *Streptococcus pneumoniae* without the need for chemical inducers or heterologous promoters. We created a panel of isogenic *S. pneumoniae* strains, differing only in synonymous (silent) mutations at the 5' end of the *ply* mRNA that are predicted to alter ΔG . Such manipulation allows rheostat-like control of PLY production and alters the cytotoxicity of whole *S. pneumoniae* on primary and immortalized human cells. These studies provide proof-of-principle for further investigation of mRNA ΔG manipulation as a tool in studies of bacterial pathogenesis.

CHAPTER II. Introduction

Delineating specific bacterial factors involved in interaction with the host is crucial to understanding mechanisms of pathogenesis and developing targeted therapies. Classically, such investigations involve construction of bacteria with disruptions in the genes encoding candidate factors (“knockouts”; KO) and comparison with wild-type (WT) or complemented strains. However, the level of expression of bacterial genes may also play a role in pathogenesis. The KO vs. WT paradigm cannot account for such differences, and although systems exist for inducible expression of specific bacterial genes, these generally rely on exogenous activators such as tetracycline that may have variable delivery to relevant sites during infection (Shockett and Schatz, 1996). Thus, it is desirable to develop strategies for genetically encoded, tunable, rheostat-like control of bacterial gene expression.

A number of factors determine the efficiency of protein production. As the genetic code is degenerate, choice among synonymous codons may play a role in translational efficiency (Hershberg and Petrov, 2008). Early models postulated a direct relationship between tRNA availability and protein production, implying that use of non-optimal codons (i.e. those that are rarely used among other genes in the species being studied) would lead to decreased protein production on the basis of tRNA scarcity (Hershberg and Petrov, 2008; Ikemura, 1985; Stoletzki and Eyre-Walker, 2007). Codon-optimization has been used as a means to increase production of heterologous genes in *E. coli* and other host species (Sharp and Li, 1987). Conversely, deoptimization of codons or codon pairs has been employed to rationally decrease protein production from natively transcribed genes (Coleman et al., 2008; Coleman et al., 2011). Kudla et al. examined a library of synonymous codon substitutions in heterologously expressed green fluorescent protein (GFP) in *E. coli* and demonstrated that the mRNA folding free energy (ΔG), particularly at the 5' terminus, correlates strongly with translational efficiency and with overall protein production (Kudla et al., 2009). Goodman et al. investigated the role

of both codon bias and mRNA ΔG in synthetic reporters and confirmed a prominent role for ΔG in shaping expression levels of individual genes (Goodman et al., 2013).

We used *Streptococcus pneumoniae* (pneumococcus), a major human pathogen with robust tools for genetic manipulation (Sung et al., 2001), as a model organism. We constructed a panel of isogenic pneumococcal strains differing only in synonymous codons at the 5' end of the gene encoding pneumolysin (PLY), an established virulence factor (Berry et al., 1989; Los et al., 2013), without antibiotic selection cassettes or exogenous promoters. These modifications altered the predicted mRNA ΔG and resulted in graded PLY production, thus affecting host-bacterial interactions and providing proof-of-principle for the use of rational modification of mRNA ΔG as a means to control protein production.

CHAPTER II. Results

Production of PLY increases with increasing ΔG

Using a degenerate nucleotide sequence derived from the *S. pneumoniae* R6 PLY protein sequence, we designed several candidate 5' mRNA regions for the *ply* gene (1416 nucleotides total) that differed in predicted minimum mRNA folding ΔG (calculated using the standard mfold algorithm over the region from -4 to +38 nucleotides (Markham and Zuker, 2005)). Each of these genes only contained synonymous mutations, leading to the production of identical amino acid sequences, but covering a wide range of predicted ΔG values. Because only a small number of codons was altered in each strain (4–7 codons/strain), the CAI for the full length *ply* gene based on highly expressed genes of *S. pneumoniae* R6 was unchanged (~ 0.38 in all strains). The two strains with the lowest 5' mRNA ΔG (-17.5 and -12.8) also had lower CAI over the first 39 bp (0.227 and 0.236, respectively) than the other three strains (0.404, 0.365, and 0.375 for the $\Delta G = -9.5, -7.3,$ and -3.4 strains, respectively). The predicted ΔG of the native *ply* 5' mRNA is -7.3 kcal/mol, and this strain was included as a control in all experiments (Figure II.1A). These alterations resulted in differential expression of PLY protein. With increasing ΔG values, we observed higher PLY levels as measured by both ELISA (Figure II.1B) and western blot (Figure II.1C). As an important control for changes in bacterial growth that would obviously impact total protein levels, we confirmed that all strains exhibited similar growth rates and doubling times (Figure II.1D). Thus alterations of the *ply* DNA sequence that alter the predicted ΔG of its resulting mRNA led to predictable and quantitative alterations in translated protein levels.

Translation kinetics drive the relationship between ΔG and protein abundance

In order to understand the mechanisms underlying the relationship between mRNA ΔG and protein production in *S. pneumoniae*, we tested whether differences in *ply* mRNA levels correlated with predicted ΔG values for the 5' mRNA and could account for differences in PLY protein levels. We used real-time reverse transcriptase PCR to quantify *ply* mRNA normalized to 16S rRNA for each strain. There were no statistically significant differences between wild-type and engineered strains in mRNA abundance (Figure II.2). Thus the absolute levels of mRNA abundance do not correlate with predicted ΔG values or with differences in PLY protein levels.

Decay kinetics of mRNA can also alter overall protein production. To test whether altered decay kinetics correlated with differences in PLY protein levels, we next assessed mRNA decay at several time points following arrest of transcription with rifampicin (Figure II.3A). Decay curves were similar and did not differ statistically. We also calculated mRNA half-lives based on mRNA abundance following transcription arrest. mRNA half-lives for each *ply* sequence were all approximately 1.5 min, and there were no statistically significant differences among groups (Figure II.3B). Taken together, these findings suggest that mRNA synthesis and decay are not altered and do not account for differences in PLY protein levels. Thus our results are consistent with the hypothesis that changes in mRNA ΔG lead to changes in post-transcriptional processing and changes in protein abundance.

ply* mRNA Δ G affects host-pathogen interactions *in vitro

The previous experiments were performed using bacteria in culture. To determine whether altered 5' mRNA Δ G also resulted in altered PLY production and cytotoxicity during infection of human cells, we examined PLY protein levels in two model systems: primary human erythrocytes and immortalized epithelial cells. PLY protein causes lysis of infected cells. We found that primary human erythrocytes exposed to *S. pneumoniae* strains exhibited lysis rates that correlated with predicted *ply* mRNA Δ G (Figure II.4A), suggesting that PLY protein levels expressed during infection correlate with predicted 5' mRNA Δ G. We confirmed these differences in PLY protein levels by direct measurement using ELISA (Figure II.4B). In parallel, we found that measurement of A549 epithelial cell cytotoxicity demonstrated a similar positive correlation between *ply* mRNA Δ G and cellular damage (Figure II.5). Thus predicted differences in *ply* mRNA Δ G positively correlate with differences in lysis rates and differences in PLY protein expression levels during active infection of human tissue culture cells.

CHAPTER II. Discussion

There is an emerging literature on the importance of 5' mRNA folding ΔG to efficiency of translation and to overall levels of protein production (Goodman et al., 2013; Kudla et al., 2009; Shah et al., 2013; Tuller et al., 2011; Zur and Tuller, 2012). Manipulation of codon bias, which may also impact protein levels, has been widely used as a method to enhance expression of genes in heterologous systems (Lee et al., 2009; Sharp and Li, 1987; Supek and Smuc, 2010). Rational alteration of 5' mRNA ΔG has been suggested as a technique that might further aid such production (Goodman et al., 2013; Kudla et al., 2009). Most studies of the ΔG /protein abundance relationship have focused on either synthetic constructs driving marker genes such as GFP (Goodman et al., 2013; Kudla et al., 2009) or on examination of naturally occurring coding sequences in model organisms (Roller et al., 2013; Shah et al., 2013; Tuller et al., 2010b; Zur and Tuller, 2012). In these studies, we rationally manipulated the predicted 5' mRNA ΔG of a known pneumococcal virulence factor in order to determine concentration-dependent effects. Such investigations are not possible using traditional KO vs. WT approaches, and determining the relationship between protein abundance and virulence may drive more detailed understanding of pathogenic mechanisms.

Using unmarked, isogenic pneumococcal strains, we showed that rheostat-like control of PLY production is feasible and that changes in ΔG are sufficient to alter host-pathogen interactions. Alteration in expression did not depend on mRNA abundance or stability but occurred at the level of translation. As predicted, there was a direct correlation between PLY production and target cell lysis using simplified *in vitro* systems, and we provided proof-of-principle for the use of such systems to assist in the study of host-pathogen interactions. Expanded testing of modified strains in animal models will allow assessment of the durability of altered PLY regulation during *S. pneumoniae* infection. Such models will facilitate understanding of the role of PLY expression levels at different stages and sites of infection.

Our findings are in agreement with those of Coleman et al., who used rational alteration of codon pair bias to alter PLY expression and found attenuation of inflammatory responses *in vivo* (Coleman et al., 2011). In contrast to that work, we alter PLY expression not by altering codon pair bias but by altering mRNA ΔG . Moreover, rather than altering the entire sequence, we were able to alter a short and specific part of the 5' region to alter PLY protein levels and lysis activity. The ability to alter a short portion of genetic sequence provides a potentially simple and straightforward tool for precisely regulating any bacterial protein involved in pathogenesis without having to alter large sections of the coding sequence and also without a priori knowledge of the host cell's codon preferences.

The correlation between 5' mRNA ΔG and overall protein production is not perfectly linear, and other factors, including codon bias, GC content, and alteration of folding by other segments of the mRNA may have substantial effects on efficiency. Because of the limited number of strains investigated, we were unable to formally evaluate the role of 5' mRNA ΔG while rigorously controlling each of these other variables, and, as noted above, our lowest expressing strains ($\Delta G = -17.5$ and -12.8) had low 5' CAI in addition to low 5' mRNA ΔG . However, the other three strains ($\Delta G = -9.5$, -7.3 , and -3.4) did not differ substantially in 5' CAI and had altered PLY production consistent with a role for 5' mRNA ΔG . Whether this technique is widely applicable to alter protein production across a range of target genes and hosts must be determined experimentally. Despite these limitations, rational alteration of 5' mRNA ΔG has potential as an important new tool in bacterial pathogenesis research. Given the increasing availability of whole genome sequences, much emphasis has been placed on the presence or absence of specific factors as markers of potential virulence in individual strains. Much less is known about the impact of relative expression levels on pathogenesis. The ability to investigate threshold and concentration-dependent effects of specific virulence factors without a requirement for non-native promoters, chemical inducers, or selective markers is extremely valuable and merits further exploration in a variety of systems.

CHAPTER II. Materials and Methods

Ethics statement

The use of primary human erythrocytes following written informed consent was approved by the Institutional Review Board of Columbia University Medical Center.

The human cell line A549 was obtained from ATCC (catalog number CCL-185).

Calculation of ΔG and codon adaptation index

Sequence folding free energies were calculated using the DINAMelt/Quikfold web server (<http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold>) (Markham and Zuker, 2005). The software uses predicted free energies at 37 °C and enthalpies from the Turner laboratory at the University of Rochester in Rochester, NY. For ΔG calculations we used version 3.0 free energies. Codon adaptation index (CAI) based on the set of highly expressed genes from *Streptococcus pneumoniae* strain R6 was calculated using the CAIcal server (<http://genomes.urv.es/CAIcal/>) (Puigbo et al., 2008a).

Bacterial strains, primers, plasmids, transformation and growth conditions

We used the Janus cassette (Sung et al., 2001), to create a panel of isogenic *S. pneumoniae* R6 derivatives with a wide range of ΔG values (-17.5 to -3.4 kcal/mol) for the 5' end of *ply* gene (-4 to +38 bp, with 1 being the A in the first ATG). To control for unexpected effects of transformations we recreated, and sequenced, a strain with the WT *ply* allele for comparison. Each of the new *ply* constructs was inserted into the native chromosomal locus, without resistance markers or alteration in predicted primary amino acid sequence. All transformed loci were confirmed by sequencing. All constructs were first assembled in *E. coli* TOP10 using plasmid pCR2.1-TOPO (Invitrogen), amplified

by PCR, and transformed into pneumococci. Forward primers differed in nucleotides encoding the initial PLY region according to Figure II.1, consistent with the following degenerate sequence: 5'-GAAR **ATG** GCN AAY AAR GCN GTN AAY GAY TTY ATH YTN GCN ATG AAT TAC G-3' (start codon in bold), and the reverse primer was 5'-CTA GTC ATT TTC TAC CTT ATC TTC T-3'. Constructs for pneumococcal transformation were assembled by overlap-extension PCR and included 570 bp upstream of *ply* (forward primer 5'-GGT TAT TGG CGA CAA GCA TT-3') and 769 bp downstream (reverse primer 5'-CCT GCT AAG ATG GTC TTG CC-3').

Briefly, pneumococcal transformation was achieved with use of semisynthetic casein plus yeast extract (C+Y) medium (Lacks and Hotchkiss, 1960). Cultures were grown at 37 °C in C+Y pH 6.8 to 0.15 OD₆₀₀, then diluted 1:20 fold into C+Y pH 8.0 (total volume 1 mL) at 30 °C, supplemented with 1 µg of CSP-1 and 11 µL of 1% CaCl₂ and incubated for 12 minutes. After adding 1 µg DNA, cultures were incubated for 40 minutes at 30 °C, followed by 90 minutes at 37 °C, followed by appropriate selection in agar plates incubated overnight at 37 °C, 5% CO₂.

E. coli were grown in lysogeny broth (LB) or agar supplemented with appropriate antibiotics for plasmid selection (ampicillin 200 µg/mL or kanamycin 50 µg/mL). *S. pneumoniae* were grown in tryptic soy broth (TSB) or agar supplemented with 200 U/mL of catalase (Worthington Biochemical), and appropriate antibiotics for strain selection (streptomycin 150 µg/mL or kanamycin 200 µg/mL).

Western blotting and ELISA

Protein separation, transfer, and immunoblotting were performed as previously described (Rampersaud et al., 2011). In blots, PLY was detected using 1:1000 dilution of mouse monoclonal anti-pneumolysin (1F11), PsaA was detected using 1:25000 dilution mouse monoclonal PsaA antibody (8G12) (Crook et al., 1998). Densitometry was performed using the gel analysis plugin implemented in ImageJ (version 1.45s; National Institutes of Health). PLY was detected by indirect ELISA using monoclonal anti-pneumolysin (1F11; 1:1000 dilution).

qRT-PCR and mRNA half-life

S. pneumoniae were grown to early exponential stage ($OD_{600} = 0.3$) and stored in RNAlater (Ambion). RNA was isolated using the RiboPure-Bacteria Kit (Ambion) with DNase treatment. cDNA was made using the SuperScript III RT Kit (Invitrogen), qRT-PCR was performed using Quanta SYBR Green PCR Master Mix in a StepOne Plus thermal cycler (Applied Biosystems).

The primers for the *ply* gene of *S. pneumoniae* (GenBank accession no. M17717) for realtime reverse-transcription PCR were previously described (Johansson et al., 2008).² Samples were normalized to *S. pneumoniae* 16S rRNA detected with forward (5'-GCC TAC ATG AAG TCG GAA TCG-3') and reverse (5'-TAC AAG GCC CGG GAA CGT-3') primers.³ In separate experiments, transcription was arrested by addition of rifampicin to 10 $\mu\text{g}/\text{mL}$ and mRNA abundance and decay were assessed by qRT-PCR of serially collected samples normalized to the 30 sec time point. mRNA half-life was calculated from best-fit exponential decay curves.

Hemolysis assay

Primary human erythrocytes (0.5% final concentration) were incubated for 30 min at 37 $^{\circ}\text{C}$ with an equal volume of heat-killed (52 $^{\circ}\text{C}$, 20 min) log-phase ($OD_{600} = 0.25$) *S. pneumoniae* R6 strains suspended in PBS. Intact erythrocytes were pelleted by centrifugation, and hemolysis was measured by assaying hemoglobin concentration in the supernatant (OD_{415} of a 100 μL sample). Hemolysis values were normalized to 0% (vehicle control) and 100% lysis (1% Triton X-100) controls as previously described (Gelber et al., 2008; Rampersaud et al., 2011).

² The nucleotide sequence of the forward primer was 5-AGC GAT AGC TTT CTC CAA GTG G-3 (positions 531 to 552), the sequence of the reverse primer was 5-CTT AGC CAA CAA ATC GTT TAC CG-3 (positions 605 to 583). (Greiner, O., Day, P.J., Bosshard, P.P., Imeri, F., Altwegg, M., and Nadal, D. (2001). Quantitative detection of *Streptococcus pneumoniae* in nasopharyngeal secretions by real-time PCR. *J Clin Microbiol* 39, 3129-3134, Johansson, N., Kalin, M., Giske, C.G., and Hedlund, J. (2008). Quantitative detection of *Streptococcus pneumoniae* from sputum samples with real-time quantitative polymerase chain reaction for etiologic diagnosis of community-acquired pneumonia. *Diagn Microbiol Infect Dis* 60, 255-261.)

³ This set of 16S primers works on R6, D39, and TIGR4 with a product size of 72 bp (amplified region 1314:1385).

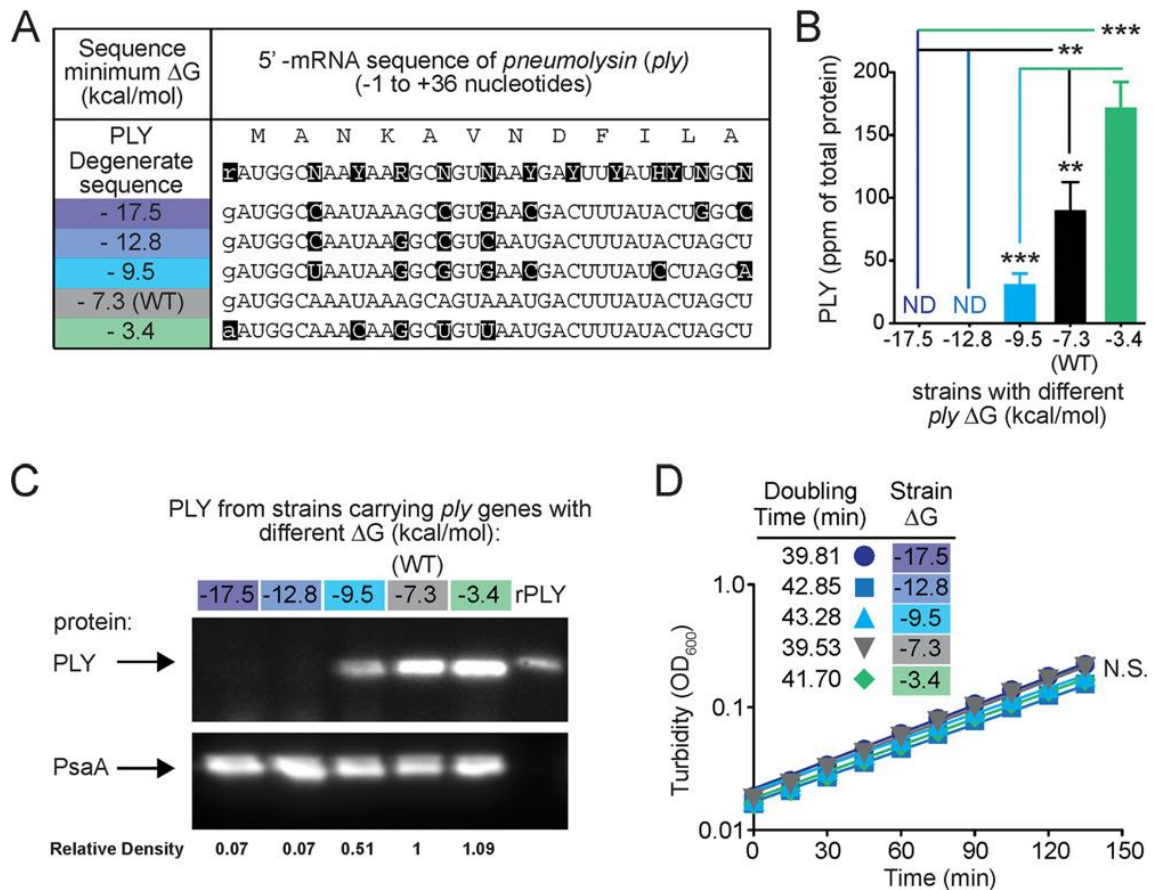
Cell culture and cytotoxicity assay

A549 respiratory epithelial cells (obtained from ATCC, catalog number CCL-185) were grown in minimum essential medium (MEM) with 10% fetal bovine serum as described (Aguilar et al., 2009). Cell monolayers (80–90% confluent in sterile 24-well plates) were incubated in MEM without serum overnight prior to stimulation with $\sim 10^8$ CFU of the indicated *S. pneumoniae* strains for 12 hrs. LDH cytotoxicity assay (Roche) was performed as per the manufacturer's instructions, and values were normalized to 0% (vehicle control) and 100% lysis (1% Triton X-100) controls as previously described (Randis et al., 2009).

Statistical analysis

Statistics were performed with Prism 5 software (GraphPad, Inc.). Data were analyzed by analysis of variance followed by post-tests to account for multiple comparisons, as appropriate. *In vitro* experiments were performed at least three times with at least three technical replicates. Bars represent SEM. Values of $P \leq 0.05$ were considered significant.

CHAPTER II. Figures

Figure II.1. Production of PLY increases with increasing *ply* 5' mRNA folding free energy.

- (A) Summary of the silent mutations engineered in the 5' mRNA region of the *ply* gene. IUPAC Nucleic Acid Codes: Y (pYrimidine); R (puRine); N (aNy); H (not G). See also Figure VI.4A for additional strains.
- (B) Increasing levels of predicted *ply* 5' mRNA ΔG correlate with increased PLY production as determined by ELISA normalized to total pneumococcal protein. Data are shown as mean of 4 experiments \pm SEM. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (ANOVA). See also Figure VI.4B for additional strains.
- (C) Western blot (top, anti-PLY; bottom, anti-PsaA loading control) demonstrates different levels of PLY production in engineered strains (lanes 1–5) compared to recombinant PLY (lane 6) as a control. Numbers beneath each lane indicate the ratio of PLY to PsaA band intensity, normalized to the value for the wild-type ($\Delta G = -7.3$) strain in lane 4.
- (D) Measurement of turbidity (OD_{600}) and calculated doubling times shows identical growth kinetics among isogenic *S. pneumoniae* strains with altered *ply* 5' sequences. (N.S. = not significant; ANOVA).

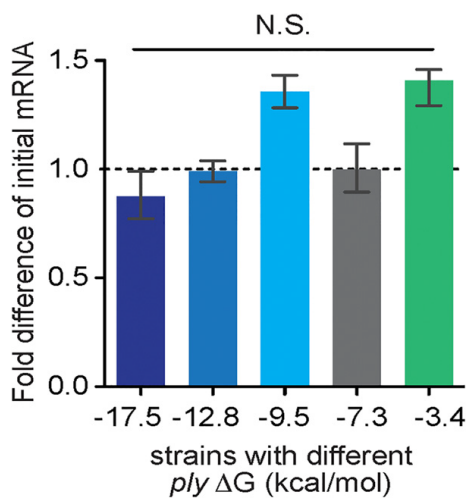


Figure II.2. Engineered differences in 5' mRNA ΔG do not alter *ply* mRNA quantity.

Fold difference in *ply* mRNA levels (y axis) was calculated relative to wild-type ($\Delta G = -7.3$) at early exponential phase. Relative mRNA amounts were calculated by $\Delta\Delta CT$ normalized to pneumococcal 16S rRNA. (N.S. = not significant; ANOVA).

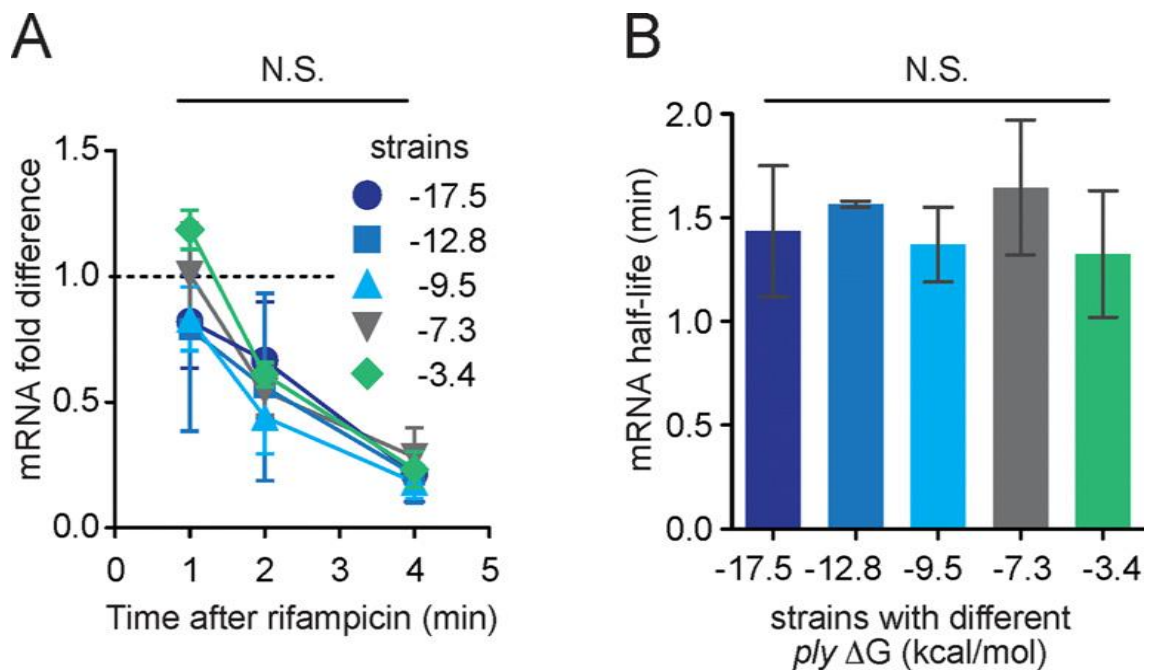


Figure II.3. Differences in PLY production do not correlate with mRNA decay kinetics.

- (A) Abundance of *ply* mRNA following transcriptional arrest with rifampicin. mRNA quantities were measured by qRT-PCR and normalized to 16S rRNA. (N.S. = not significant; ANOVA).
- (B) Differences in *ply* 5' mRNA ΔG do not alter mRNA half-life values, as calculated from best-fit exponential decay curves from data in (A). (N.S. = not significant; ANOVA).

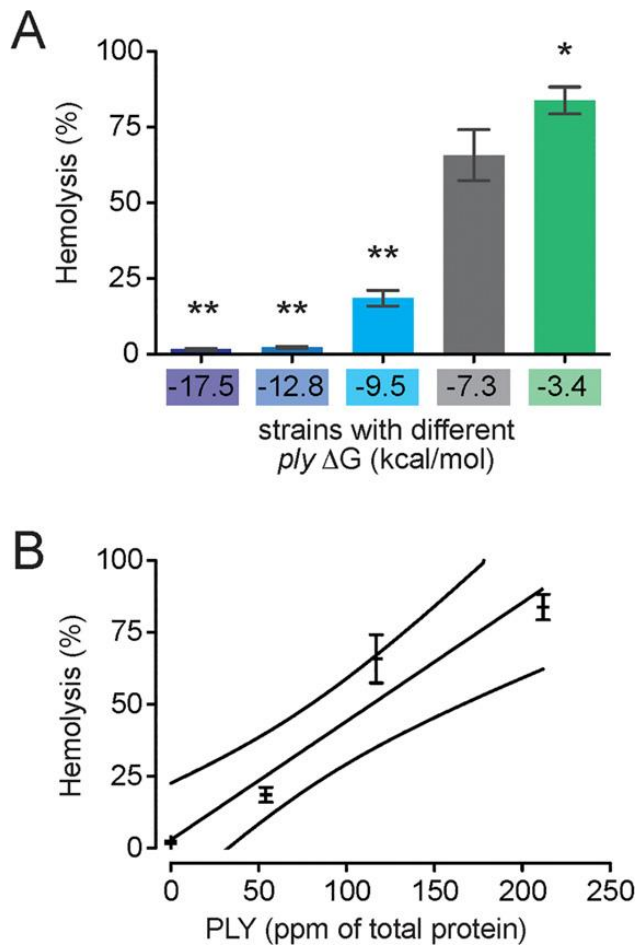


Figure II.4. Differential PLY production by *S. pneumoniae* strains is directly related to PLY production. Lysis of primary human erythrocytes as assessed by hemoglobin release assay correlates with (A) *ply* 5' mRNA ΔG and (B) PLY as measured by ELISA ($R^2 = 0.95$; outside lines represent 95% CI). *, $P < 0.05$; **, $P < 0.01$ (ANOVA).

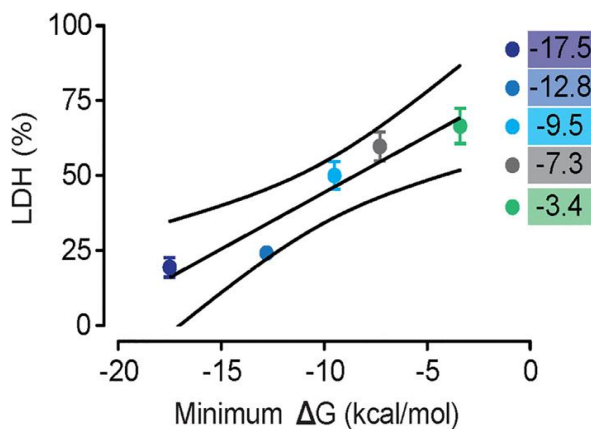


Figure II.5. Lysis of human A549 respiratory epithelial cells as measured by LDH release correlates with predicted *ply* 5' mRNA ΔG .

($R^2 = 0.91$; outside lines represent 95% CI; $P < 0.0001$, ANOVA for linear trend).

CHAPTER III. Control of *S. pneumoniae* Virulence in Vivo by Rational Manipulation of mRNA Folding Free Energy

Fábio E. Amaral, Dane Parker, Ritwij Kulkarni,

Alice S. Prince, Tara M. Randis, Adam J. Ratner.

This chapter contains original unpublished research material. All of the experiments carried out and described in this chapter were executed by me, or occasionally, with the help of people mentioned where appropriate.

Author Contributions

F.E.A., A.J.R., and D.P. conceived and designed the experiments, and analyzed the data. F.E.A. performed the experiments, except BAL performed by D.P. assisting with mice experiments. R.K. provided technical support. F.E.A., A.J.R., A.S.P., and T.M.R. contributed reagents, materials or analysis tools.

Acknowledgments

We thank Timothy J. LaRocca for the most interesting discussions furthering the understanding of this work, and Kate Hensel Sapro for the help with appropriate statistics, and Edwin Aedes for the generous gift of the PsaA antibody.

CHAPTER III. Abstract

Murine models demonstrate a role for PLY in colonization, invasive disease, and immunogenic protection response. However, prior evaluations of the role of PLY in virulence have generally relied on comparisons between PLY-sufficient and PLY-negative (knockout, KO) strains. We require rheostat-like control of PLY expression that happens natively to better understand dose-response effects. Thus, we explored rational manipulation of mRNA folding free energy (ΔG) for such control. In this work we show a range of PLY levels in strains derived from the same parental background. We have shown rheostat-like control of PLY production *in vitro* (Amaral et al., 2015) and now, we demonstrate this *in vivo* using a murine model, with differential outcome of pneumococcal pneumonia. We observed macrophage depletion and neutrophil recruitment in a manner that correlates to the production of PLY controlled by our engineered ΔG . Consequently, we show that proinflammatory cytokines are upregulated in a PLY-dependent way, which corresponds to the predicted ΔG . This model provides proof-of-principle for the rheostat control of virulence factors *in vivo* without the need for exogenous factors, using minimal manipulation of mRNA folding free energy.

CHAPTER III. Introduction

In our previous work, we created pneumococcal strains that differed in the 5' mRNA region of PLY, otherwise isogenic. This was done by rational introduction of silent mutations in that small region. These modifications altered the predicted mRNA ΔG and resulted in graded PLY production, thus affecting host-bacterial interactions *in vitro* (Amaral et al., 2015).

In this work, we continued the line of work by further testing these strains *in vivo* and providing proof-of-principle for the use of rational modification of mRNA ΔG in models of infection as a means to control protein production *in vivo*.

Pneumococcal encapsulated strains are capable of murine infection which make them suitable for this line of work. PLY is one of the major virulence factors of the pneumococcus and is directly responsible for lung cytotoxicity (del García-Suárez et al., 2007; Marriott et al., 2008) worthy of study. PLY is a pore-forming toxin that belongs to the cholesterol-dependent cytolysins produced by a lot of different bacteria. It depends on the presence of cholesterol in the membranes to oligomerize into a pore-forming structure (Martner et al., 2008). Another major action is the ability to activate complement (Alexander et al., 1998; Berry et al., 1995; Kadioglu et al., 2008; Paton et al., 1993; Tweten, 2005). Activating the membrane attack complex that clears pathogens away is antagonistic, but then again, pneumococcus has a capsule that confers resistance to this and to opsonization. The depletion of complement components, however, is useful in creating proinflammatory environment aiding invasion and survivability (Mitchell and Dalziel, 2014).

PLY is highly immunogenic (Lock et al., 1988) capable of activating toll-like receptors – TLR2 and TLR4 – and of inducing the expression of a number of proinflammatory cytokines like IL-1 α , IL-1 β , IL-6, IL-12, IL-23, TNF- α , IL-17A, and IFN- γ , and recruiting neutrophils (PMN) and dendritic cells (DC) to the infection site (Branger et al., 2004; Dessing et al., 2008; Dessing et al., 2009; Malley et al., 2003; McNeela et al., 2010;

Srivastava et al., 2005; van Rossum et al., 2005). This has been shown extensively *in vitro* and *in vivo*, upon the infection of human or murine hosts which makes this toxin a valid target for immunization (Cockeran et al., 2005; Mitchell and Dalziel, 2014).

Previous authors have shown strain attenuation by knocking out PLY (Berry and Paton, 2000). We hypothesize that different amounts of PLY can impact the host in distinct ways that may not be linear. This is particularly important when the mechanisms for protein regulation are unknown to us. Understanding the outcomes of PLY under or overexpression is a significant goal that can provide alternative paths to the study of bacterial pathogenesis beyond the KO vs. WT paradigm.

CHAPTER III. Results

Distinct PLY expression *in vivo* according to ΔG

In order to test the relevance of tunable PLY expression *in vivo*, it was necessary to use encapsulated *S. pneumoniae*. Thus, we constructed derivatives of strain D39 (serotype 2) with a subset of the leader sequences, using the genomes of the R6 background strains we previously tested. R6 is a lab strain derived from D39 therefore sharing a mostly similar genetic background (Lanie et al., 2007). These new D39 strains exhibited similar results to R6 strains with regards to PLY production (Figure III.1A), and hemolysis (data not shown). We then used these in a murine pneumonia model (Parker et al., 2011).

The day subsequent to infection, we noted general mouse behavior by comparison. Mice infected with strain -17.5 behaved like uninfected control mice; mice infected with strain -9.5 behaved sicklier though were still moving around; mice in groups -7.3 (WT) and -3.4 behaved very sickly and barely moved (data not shown). With the exception of -7.3 and -3.4, all three other groups shredded the nest-building material (data not shown).

Bacterial burden is not correlated with PLY expression modulated by ΔG

Twenty-four hours following infection, histopathological examination of lung sections revealed marked inflammation in the setting of infection, with increased consolidation and inflammatory infiltrate with increasing ΔG (Figure III.2A).

These findings are in contrast with the recovered CFU counts. Quantitative bacterial culture of lung homogenates revealed similar bacterial densities for all strains (Figure III.3B). However, in the bronchoalveolar lavage (BAL) CFU fraction we detected a significant increase in CFU load of the lowest ΔG strain compared to other groups (Figure III.2B). This suggests that sheer bacterial numbers are not causing the observed differential inflammatory state. Systemic invasion, as measured by splenic cultures, was infrequent and did not differ among strains (Figure III.3B).

We detected PLY in BAL fluid in amounts that correlated positively with predicted mRNA ΔG (Figure III.1B). Notably, even the -17.5 kcal/mol strain produced detectable PLY *in vivo*, in contrast to the *in vitro* results we previously obtained in R6 background (see CHAPTER II).

Immune cell population shift correlates to differential PLY expression

Next, we proceeded to characterize the inflammatory cell population by FACS. We analyzed the lung separately from the BAL cell population. During pneumococcal pneumonia, local depletion of macrophages (M Φ) via induction of apoptosis has been described and represents an important component of host defense (del García-Suárez et al., 2007; Dockrell et al., 2003). We found that such depletion was correlated with the level of PLY expression, with the highest PLY-expressing strains inducing maximal depletion in the lung (Figure III.2C). Polymorphonuclear leukocytes (PMN) and dendritic cells were recruited to the lungs in the setting of infection, but there were no significant differences among animals receiving the various pneumococcal strains (Figure III.2C and Figure III.3C). There was no significant accumulation of natural killer (NK) cells in the lungs in any of the infected groups (Figure III.3C).

Proinflammatory cytokine production correlates to ΔG

Lastly, we analyzed the cytokine profile that is generally induced by PLY. Because of the importance of local cytokine production for tissue damage and clinical outcome in pneumococcal pneumonia (Marriott et al., 2012), we determined the concentrations of CXCL1, TNF α , and IL-1 β in BAL fluid (Figure III.2D). In all cases, we found a relationship between PLY production and inflammation, with IL-1 β exhibiting the strongest direct correlation. We saw significant trends that positively correlated the increase in ΔG with increases in each cytokine. There were no significant differences in BAL total protein levels or cell populations among any of the infected groups (Figure III.3A and C).

We postulate that the differential CFU burden we observed in the BAL is consistent with the local antibacterial effects of cytokine release. The observed behavior patterns were in contrast with the recovered CFU in the BAL although matching the respectively characterized inflammatory state.

These differences meet statistical significance lending additional evidence that different amounts of PLY, from undetectable to higher than WT, impact the immune system uniquely.

CHAPTER III. Discussion

We generated *S. pneumoniae* D39 strains that were mutated in the 5' mRNA of *ply* gene in the same way as previously tested strains (see CHAPTER II). We observed that with increasing ΔG there is enhanced protein expression, manifested in both increased hemolysis and cytotoxicity *in vitro*. We wanted to evaluate these new strains *in vivo*. Mouse behavior upon one day post intranasal infection clearly demonstrates the differences in virulence of the different strains.

Histopathological findings revealed a correlation of inflammation markers in the lung sections of mice infected with higher ΔG strains. The more PLY was expressed more consolidation, hemorrhage and inflammatory population infiltrate was present in the samples. We argue these findings were not due to absolute bacterial burden as the recovered BAL CFUs were similar and only statistically increased in the strain with the lowest ΔG . Furthermore, the similar load of splenic CFUs shows that systemic invasion was not different among strains and, in the literature, it has been argued that PLY is less important for this stage of infection (Benton et al., 1995). Nonetheless, PLY has also been shown to be important for better fitness and growth rate, as well as the spread into the bloodstream (Kadioglu et al., 2008; Los et al., 2013). This further demonstrates the importance of the host factors in the inflammatory interplay. Even though these strains grow similarly *in vitro*, they present different characteristics *in vivo*. Perhaps due to low levels of PLY, the lowest ΔG strain did not elicit vigorous immune responses hence allowing for extended growth and lower bactericidal activity. Or, possibly, the overall translation fitness allowed better growth *in vivo*. This strain recapitulates what has been previously studied with PLY-KO (Kadioglu et al., 2008; Los et al., 2013; Rayner et al., 1995; Wellmer et al., 2002) in contrast to the WT and the strain that produces more PLY. PLY is very capable of inducing immune responses as mentioned previously. It synergizes with TLR agonists to enhance the secretion of some proinflammatory cytokines including IL-1 β , and TNF- α by DC or PMN, among others (McNeela et al., 2010), and is able to recruit PMN to the site of infection (Dessing et al., 2009; Moreland and Bailey, 2006).

Lack of PLY in a model of infection exhibited an immune response independent of IL-1 β or IL-6 (Benton et al., 1998) which may explain an evasion of the immune system and better proliferation for the strain with lowest PLY expression, much like it happens during asymptomatic colonization. IL-1 β , in turn, is required for resistance to pneumococcus (Kafka et al., 2008). Additionally, PLY is a known inflammasome activator (Fang et al., 2011; McNeela et al., 2010) and on its own is capable of triggering the pathology of pneumonia such as the destruction of lung parenchyma mediated by activation of caspase-1 and induction of apoptosis with PMN recruitment (del García-Suárez et al., 2007; Fang et al., 2011; Marriott et al., 2008). These observations are in accordance with our data. In the cytokine profile we evaluated, all proinflammatory cytokines (IL-1 β , TNF- α , and CXCL1) were upregulated in response to increasing ΔG strains that exhibited increasing PLY quantities.

While PLY abundance did not alter in PMN or DC recruitment in the lung in a very significant way, both local cytokine production and M Φ depletion were strongly correlated with *ply* mRNA ΔG . M Φ depletion exhibited a threshold phenotype, in that the WT and the $\Delta G = -3.4$ kcal/mol strains had similar amounts of M Φ , but local IL-1 β was significantly higher during infection with the latter strain. These findings are consistent with data linking PLY to activation of the NLRP3 inflammasome, with subsequent production of IL-1 β (Fang et al., 2011; McNeela et al., 2010; Witzentrath et al., 2011). Inflammasome signaling is a critical pathway in pulmonary detection of both infectious and environmental insults, and our data suggest that rheostat control of PLY production is useful for understanding quantitative aspects of such responses (Gasse et al., 2007; Kebaier et al., 2012; McNeela et al., 2010; Provoost et al., 2011; Wu et al., 2013).

We characterized cytokine levels, PMN and DC cell counts, and histopathological signs of inflammation, which were all significantly increased in accord with increase in strain ΔG . Conversely, the M Φ population in the lung, movement and shredding of nesting material were considerably diminishing with each increase in ΔG . This work appreciates a tangible application of protein rheostat control principle *in vivo* using minimalistic silent mutations in the 5' mRNA of said protein.

CHAPTER III. Materials and Methods

Ethics statement

All mouse infections were performed under the guidelines of the Columbia University Institutional Animal Care and Use Committee.

Calculation of ΔG

Sequence folding free energies were calculated using the DINAMelt/Quikfold web server (<http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold>) (Markham and Zuker, 2005). The UNAFold software uses predicted free energies at 37 °C and enthalpies from the Turner laboratory at the University of Rochester in Rochester, NY. For ΔG calculations we used the more up-to-date free energies for RNA denoted as version 3.0 energies.

Bacterial strains, primers, transformation and growth conditions

We used the isogenic *S. pneumoniae* R6 derivatives with a range of ΔG values (-17.5 to -3.4 kcal/mol) in the 5' end of *ply* mRNA created in our previous work (Amaral et al., 2015). We used those genomes to transform *S. pneumoniae* D39 and sequenced both *ply* and *cpsA* loci to ensure correct transformation.

Briefly, pneumococcal transformation was achieved with use of semisynthetic casein plus yeast extract (C+Y) medium (Lacks and Hotchkiss, 1960). Cultures were grown at 37 °C in C+Y pH 6.8 to 0.15 OD₆₀₀, then diluted 1:20 fold into C+Y pH 8.0 (total volume 1 mL) at 30 °C, supplemented with 1 µg of CSP-1 (Genscript) and 11.1 µL of 1% CaCl₂ and incubated for 12 minutes. After adding 1 µg DNA, cultures were incubated for 40 minutes at 30 °C, followed by 90 minutes at 37 °C, followed by appropriate selection in agar plates incubated overnight at 37 °C, 5% CO₂.

E. coli were grown in LB or agar supplemented with appropriate antibiotics for plasmid selection (ampicillin 200 µg/mL or kanamycin 50 µg/mL). *S. pneumoniae* were grown in TSB or agar supplemented with 200 U/mL of catalase (Worthington Biochemical Corporation), and appropriate antibiotics for strain selection (streptomycin 150 µg/mL or kanamycin 200 µg/mL).

Western blotting and ELISA

Protein separation, transfer, and immunoblotting were performed as previously described (Rampersaud et al., 2011). In blots, PLY was detected using 1:1000 dilution of mouse monoclonal anti-pneumolysin (1F11, sc-80500, Santa Cruz Biotechnology), PsaA was detected using 1:25000 dilution mouse monoclonal PsaA antibody (8G12) (Crook et al., 1998). Immunoblot assays were followed by secondary antibodies conjugated to horseradish peroxidase (Santa Cruz Biotechnology Inc.). Densitometry was performed using the gel analysis plugin implemented in ImageJ (version 1.45s; National Institutes of Health). PLY was detected by indirect ELISA using monoclonal anti-pneumolysin (1F11; 1:1000 dilution).

In vitro PLY was detected by indirect ELISA using monoclonal anti-pneumolysin as above. *In vivo* PLY in BAL was detected by sandwich ELISA using rabbit polyclonal anti-pneumolysin (ab71811, Abcam) as coating antibody and monoclonal anti-pneumolysin as detection antibody. Goat anti-mouse IgG-HRP conjugated antibody used was sc-2005 (Santa Cruz).

Imaging was done using FluorChem M (ProteinSimple) system.

Murine infection model

The mouse pneumonia model including FACS determination of cell populations and ELISA analysis of cytokines was performed as described (Parker et al., 2011; Parker and Prince, 2012) except that anesthetized mice were inoculated intranasally with 10^7 CFU *S. pneumoniae* in 50 µL of sterile PBS. Mice were sacrificed 24h after infection.

Cell profiles by flow cytometry

Staining cells for flow cytometry has been described elsewhere (Parker et al., 2011). Cells were labeled with an antibody cocktail consisting of PerCP-Cy5.5-labelled anti-CD11c (N418; Biolegend), allophycocyanin-labelled anti-MHC II (I-A/I-E; Biolegend), phycoerythrin-labelled anti-NK 1.1 (NKR-P1C, Ly-55; Biolegend) and fluorescein isothiocyanate-labelled anti-Ly-6G (Gr-1; RB6-8C5; Biolegend). PMN were identified as Ly6G⁺/MHCII⁻, MΦ as CD11c⁺/MHCII^{low-mid}, DC as CD11c⁺/MHCII^{high}, and NK as NK 1.1⁺. Data were analyzed using WinMDI (version 2.8; Joseph Trotter).

Cytokine ELISA

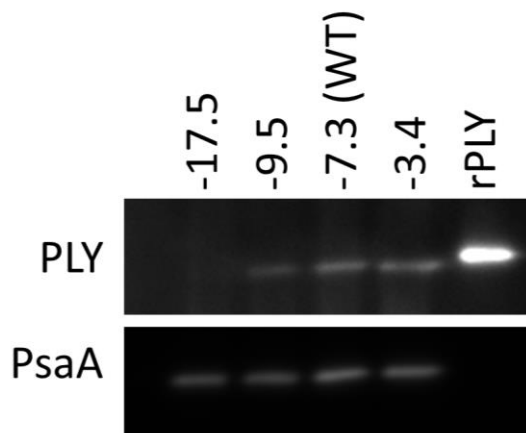
ELISA analysis of cytokines has been described elsewhere (Parker and Prince, 2012). Kits used for mouse functional IL-8 homologue CXCL1/KC (DuoSet ELISA, Catalog Number DY453, R&D Systems), for mouse IL-1β (ELISA Ready-SET-Go![®] Catalog Number: 88-7013, eBioscience), mouse TNF alpha (ELISA Ready-SET-Go![®] Catalog Number: 88-7324, eBioscience).

Statistical analysis

Statistics were performed with Prism 5 software (GraphPad, Inc.). Data were analyzed by analysis of variance or Kruskal-Wallis tests followed by post-tests to account for multiple comparisons, as appropriate. *In vitro* experiments were performed at least three times with at least three technical replicates (for experiments other than western blots). Sample sizes for *in vivo* experiments are indicated in the figure legends. Bars represent SEM. Values of $P \leq 0.05$ were considered significant.

CHAPTER III. Figures

A



B

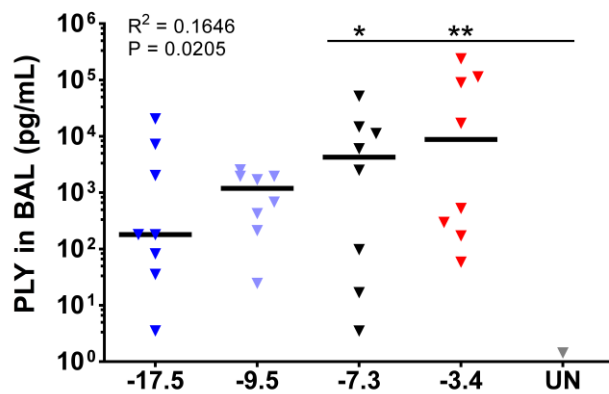


Figure III.1. (A) PLY levels of *S. pneumoniae* D39 strains created for this study, used for *in vivo* experiments, determined by Western blot. (B) Detection of PLY by ELISA of BAL fluid.

Data were analyzed by Kruskal-Wallis test followed by Dunn's Multiple Comparison Test against UN. Each symbol represents one mouse, and horizontal bars represent medians. * $p < 0.05$; ** $p < 0.01$.

A

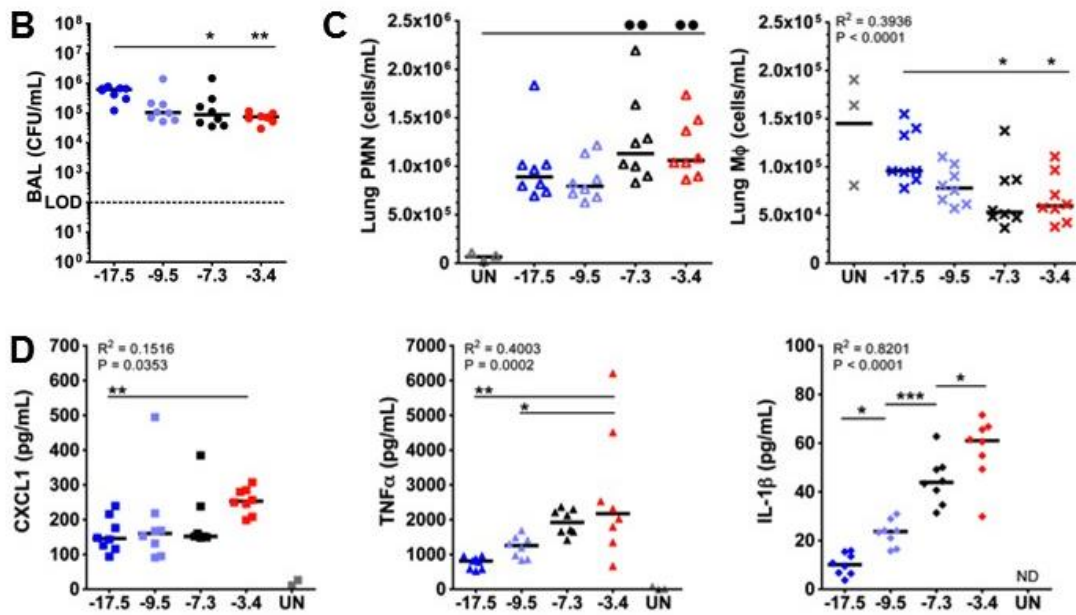
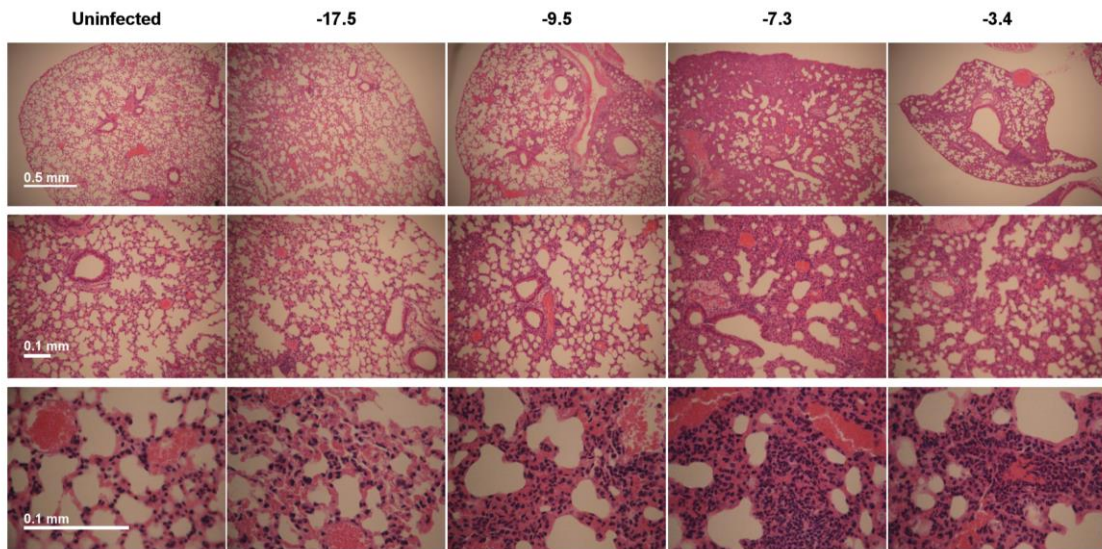


Figure III.2. Proinflammatory capacity of *S. pneumoniae* D39 increases with PLY production in a mouse model of upper airway infection.

- (A) H&E staining of lung sections from C57Bl/6 mice 24h post-infection demonstrate increasing consolidation and infiltrate in strains with higher predicted ΔG .
- (B) Quantitative culture of lung homogenates reveals similar colony counts in all groups. Colony forming units (CFU) of mouse bronchoalveolar lavage (BAL) 1 day post infection. Limit of detection (LOD) = 100 CFU/mL. Data were analyzed by Kruskal-Wallis test followed by Dunn's Multiple Comparison Test against -17.5. See also Figure III.3B.

- (C) Flow cytometric analysis of immune cell populations (MΦ, macrophages; PMN, polymorphonuclear leukocytes; DC, dendritic cells) in lung tissue 1 day post infection. Data were analyzed by Kruskal-Wallis test followed by Dunn's Multiple Comparison Test against UN (•) and excluding UN against -17.5 (*), and post test for linear trend, where noted. See also Figure III.3C.
- (D) Cytokine levels measured by ELISA in mouse BAL 1 day post infection. CXCL1 data were analyzed by Kruskal-Wallis test followed by Dunn's Multiple Comparison Test excluding UN against -17.5. TNFα data were analyzed by One-way ANOVA followed by Tukey's Multiple Comparison Test. IL-1β data were analyzed by One-way ANOVA followed by Bonferroni's Multiple Comparison Test. All cytokines were also analyzed by post test for linear trend, excluding uninfected, with the coefficient of determination for linear regression (R²) and P value displayed.

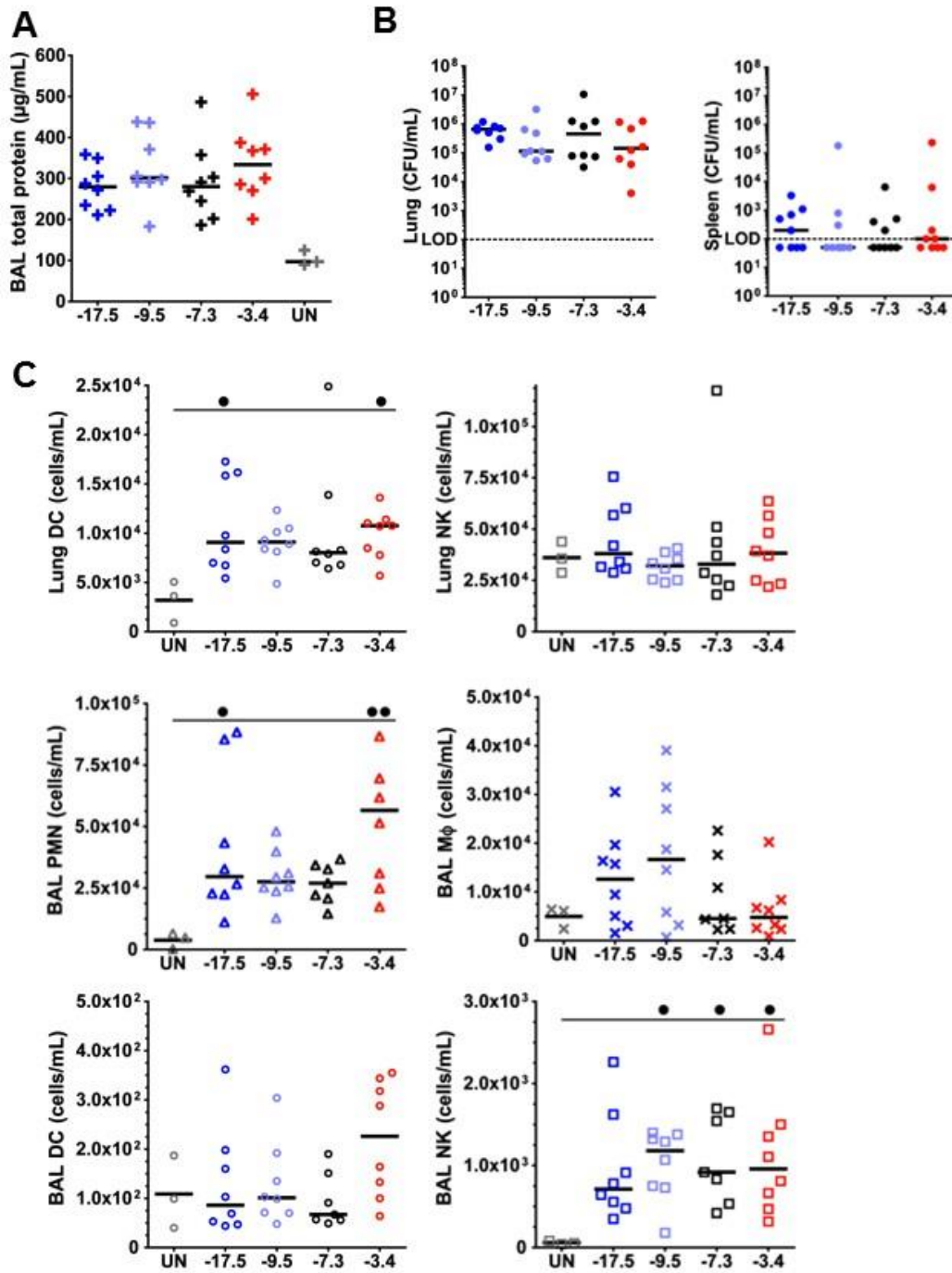
Each symbol represents one mouse, and horizontal bars represent medians. •/* p < 0.05; ••/** p < 0.01; *** p < 0.001. Animals: n = 8 per group; n = 3 in uninfected (UN) group.

(Figure III.3 on next page.)

Figure III.3. Proinflammatory capacity of *S. pneumoniae* D39 increases with PLY production in a mouse model of upper airway infection.

- (A) Total protein measurement of mouse bronchoalveolar lavage (BAL) 1 day post infection.
- (B) Colony forming units (CFU) of mouse lung and spleen 1 day post infection. Limit of detection (LOD) = 100 CFU/mL.
- (C) Flow cytometric analysis of immune cell population in mouse lung tissue and BAL 1 day post infection. Polymorphonuclear leukocytes (PMN), macrophages (MΦ), dendritic cells (DC), and natural killer cells (NK). Data were analyzed by Kruskal-Wallis test followed by Dunn's Multiple Comparison Test against UN (•) and excluding UN against -17.5.

Each symbol represents one mouse, and horizontal bars represent medians. • p < 0.05; •• p < 0.01. Animals: n = 8 to 9 per group; n = 3 in uninfected (UN) group.



(Figure III.3 legend on previous page.)

CHAPTER IV. Translation Initiation Drives the Differential Rates of Protein Translation due to 5' *ply* mRNA Folding Free Energy Differences

Fábio E. Amaral, Adam J. Ratner.

This chapter contains original unpublished research material. All of the experiments carried out and described in this chapter were executed by me.

Author Contributions

F.E.A. and A.J.R. conceived and designed the experiments, and analyzed the data. F.E.A. performed the experiments. F.E.A. and A.J.R. contributed reagents, materials or analysis tools.

Acknowledgments

We thank Ruben L. Gonzalez, Arthur G. Palmer III, Max E. Gottesman, Ritwij Kulkarni, and Timothy J. LaRocca for the most interesting discussions furthering the understanding of this work, and Kate Hensel Sagra for the help with appropriate statistics.

CHAPTER IV. Abstract

Amino acids can be encoded by more than one codon and there is a bias towards some preferred ones as a way to maybe regulate translation. This creates all sorts of bias in the evolution of genes that differ from specie to specie and create important features in the genetic sequence. Examples are codon usage, tRNA adaptation, GC%, and mRNA folding free energy (ΔG) that contribute differently to protein expression. In our work, we modulate *S. pneumoniae* toxin pneumolysin to include synonymous mutations that translate into the same protein (Amaral et al., 2015) but with different levels of translation that correlate to the 5' mRNA ΔG . We have seen no differences in mRNA abundance nor decay rates. In this work, we further characterize the mechanism by which protein translation is affected. Translation initiation is generally considered to be rate-limiting in terms of protein production. Thus, we have generated a fusion protein between the altered 5' region of PLY in frame with whole GFP and proceeded to evaluate *in vitro* protein production. Interestingly, unlike our previous *in vitro* and *in vivo* studies, we observed highest expression with the wild-type fusion and not the highest ΔG . Yet, we prove that production of PLY increases due to faster translation initiation rates.

CHAPTER IV. Introduction

Scientists have resorted to a number of ways to alter protein expression in mutant variants or for heterologous gene expression. Whether to increase expression or to abolish it by construction of knockouts, there have been widely used external inducers and inhibitors engineered in the genetic sequence. Other strategies have included the manipulation of the sequence itself using silent mutations. This results because amino acids can be encoded by more than one codon and there is a bias towards some preferred ones. This is called codon usage bias (CUB) and can be measured using the codon adaptation index (CAI) (Sharp and Li, 1986, 1987). The CUB has been successfully used to increase or decrease protein expression in natural host (Carlini, 2004; Carlini and Stephan, 2003; Hense et al., 2010) or heterologously. However, a number of translation determinants interfere in protein production (Gingold and Pilpel, 2011). The availability and production of tRNAs also play a role in protein translation and it too can create a bias that can be calculated by an adaptation index (tAI) (Trotta, 2013). Codon pairs (CP) also exhibit preferred and underused pairs, creating a bias that has been used to decrease the virulence of virus (by whole genome deoptimization) and bacteria (by means of altering toxin production) (Coleman et al., 2008; Coleman et al., 2011; Mueller et al., 2010; Mueller et al., 2006). Recently, CP bias has been implicated as a consequence of dinucleotide bias. CpG plays a role in deoptimization of gene expression (Kunec and Osterrieder, 2016) potentially because it is a frequent methylation target or due to the triple hydrogen bonds these nucleotides form. This last point is also implicated in the percentage of GC (GC%) nucleotides, particularly at the 3rd base of the codon (GC3%) (Babbitt et al., 2014). Lastly, the mRNA folding free (ΔG) energy has also been linked to protein expression and this correlation was found more positive in the 5' end of the mRNA molecule, by calculating the ΔG in a sliding window of 42 nucleotides length (Kudla et al., 2009). When the ΔG is higher, the structure is weaker. In this scenario, several authors agree that the CAI has more impact in gene expression (Gingold and Pilpel, 2011; Plotkin and Kudla, 2011; Supek and Smuc, 2010; Tuller et al.,

2010a; Tuller et al., 2010b). Overall, these factors contribute together to the total amount of protein expressed nonetheless, there is no consensus on which is more important.

During protein translation, translation initiation is considered rate-limiting (Jacques and Dreyfus, 1990; Kozak, 1986; Kozak, 1999; Kudla et al., 2009; Plotkin and Kudla, 2011; Shah et al., 2013) but whether elongation rate is equally important or not is disputed. An author (Xia, 2015) has evaluated the work of others (Kudla et al., 2009) to conclude that translation elongation accounts for about 17% of total variation in protein production. This reinforces the idea that no single factor detains unique control of protein translation. Naturally, the availability and the speed of ribosome binding, as well as mRNA molecule unfolding provide a basis for differential rates of translation initiation. In turn, these effects affect global translation efficiency that may reflect directly in enhancement of cell growth and fitness (Kudla et al., 2009).

In our previous work (see CHAPTER II), we modified the 5' region of *S. pneumoniae* toxin pneumolysin (PLY) to include synonymous mutations that translate into the same protein (Amaral et al., 2015) but with different levels of expression that correlate to the 5' mRNA ΔG . With regards to the mechanism inherent to rheostat control of protein, we quantified and studied the stability of the mRNA in the different strains. We concluded that there were no significant differences in abundance nor decay rates of mRNA. To further investigate this mechanism we looked at translation of the mRNA in a controlled *in vitro* setting. We were able to target the study of translation initiation since we analyzed specifically the translation differences of the 5' mRNA sequences. *In vitro* synthesis of proteins is accomplished in this work with the use of *E. coli* cell-free extracts and a T7 promoter translation system which allowed us to quantify the production of protein in real time and compare the diverse sequences.

CHAPTER IV. Results

Increased production of PLY is due to accelerated translation initiation

We observed that translation kinetics drive the relationship between ΔG and protein abundance because neither mRNA synthesis nor decay are altered among different ΔG strains (CHAPTER II, cf. Figure II.2 and Figure II.3). Because changes in mRNA ΔG may impact ribosomal access to translation initiation sites, we hypothesized that translation efficiency rather than mRNA abundance or decay would likely be responsible for altered protein expression.

We cloned *ply* 5' mRNA leader sequences upstream and in frame with *gfp* in order to evaluate the impact of 5' ΔG on protein production using an *in vitro* translation system based on cell-free *E. coli* extracts and T7 RNA polymerase. The fused gene is under a T7 promoter and includes a ribosomal binding site (RBS) similar to the Shine-Dalgarno sequence in the same position of WT *S. pneumoniae*. In this system, we observed an exponential increase in protein fluorescence over time with a peak at ~3 hours (Figure IV.1B). Fluorescence increased with increasing ΔG , with the exception of the WT *ply* leader sequence, which generated significantly higher GFP levels than the $\Delta G = -3.4$ kcal/mol sequence, in contrast to the protein levels observed in the *S. pneumoniae* strains (Amaral et al., 2015). In order to understand this result, we examined protein production at earlier time points, focusing on the period between 15-30 min following commencement of the reaction. This period captures the scenario before differences are established and detectable. By 30 min, the strain with the WT leader upstream of GFP had already generated significantly more GFP fluorescence than all of the other strains (Figure IV.1A), indicating that a very early step, likely initiation, was favored in the setting of the WT leader and differed between the cell-free *E. coli* extracts and the live *S. pneumoniae* systems.

Production rates (k) were calculated from best-fit logistic curves for protein synthesis according to the following equation:

Equation IV.1

$$f(x) = \frac{a}{1 + be^{-kx}}$$

where a is the curve's maximum value/upper asymptote, b is related to the value of $f(0)$, e is the natural logarithm base (Euler's number), and k is the slope of the curve also interpreted as growth/production rate.

In the first 30 minutes we found k to be negative in the lower ΔG strains (-0.016 and -0.006 for -17.5 and -9.5 kcal/mol, respectively), 0.004 for -3.4 kcal/mol, and highest for the WT -7.3 kcal/mol with 0.084. After this, and up to 3 hours, we found them to be similar and around 0.030 with the exception of the lowest ΔG -17.5 kcal/mol, which was lower (0.016, 0.031, 0.027, 0.034 for -17.5, -9.5, -7.3, -3.4 kcal/mol, respectively). Together, these results denote the variation occurring and the importance of early translation initiation in establishing differential protein production sustained in later time points.

CHAPTER IV. Discussion

This work further enlightens the mechanistic workings of mRNA folding free energy manipulation. We showed that the variation in gene expression is not due to mRNA abundance nor to altered half-life, as these were comparable across strains (see CHAPTER II). These facts support the idea that a few silent mutations in the 5' end of the gene do not significantly affect mRNA stability keeping the overall mRNA levels similar across time and strains.

To document the underlying mechanism behind the altered rate of protein production we performed *in vitro* translation in an *E. coli* cell-free lysate system. GFP fluorescence was assessed from fusion proteins with *ply* 5' leader sequences upstream of and in frame with the *gfp* coding sequence that vary in accordance to previously created *S. pneumoniae* strains (Amaral et al., 2015). Early (15-30 min) time points from *in vitro* translation experiments demonstrate increased GFP fluorescence from the strain with the WT leader sequence. Whereas in all other experiments we saw a gradient response varying according to ΔG , in this case the WT construct (-7.3 kcal/mol) exhibited the highest protein expression, beyond that of the highest ΔG construct (-3.4 kcal/mol). This implies that we do not yet fully comprehend the interplay between the translation determinants, however hinting at the importance of ΔG *in vivo*. Concerning the conditional differences, on one hand, *in vitro* constructs did not feature full length PLY but rather only the initial 12 codons in frame with full length GFP. This alone could be a cause for unforeseen alterations to the mRNA folding that influence the 5' region. On the other hand, the *in vitro* characterization is limited since reactions occur in a cell-free environment not restricted by whole cell volume neither bound to timing or to the cell cycle (about 40 min in pneumococcus, Figure VI.2). The reactions *in vitro* occur over a span of 6 hours and variable volumes. Also, being a technically engineered medium for protein expression, conditions such as ionic molarity may differ from the assumed by the algorithm upon the calculation of the sequences' ΔG ($[Na^+] = 1$ M and $[Mg^{2+}] = 0$ M at 37 °C; see also "CHAPTER IV. Materials and Methods" on page 74). Furthermore, *in*

in vitro reactions are free of any other DNA or mRNA molecules that would contaminate. It is known that translation of a single gene is affected differentially by the translome (Agashe et al., 2012; Babbitt et al., 2014; Ciandrini et al., 2013; Tuller et al., 2010b). Plus, the stability of mRNA controls rates of translation leading to both global and local synonymous codon biases in many unicellular organisms (Babbitt et al., 2014) highlighting the importance of the remaining translome. Therefore, the lack of other messages sequestering ribosomes and resources may well rework the conditions for the studied synthesis reaction. Other translation determinants unaccounted in our technique have also been shown to impact translation to a certain extent (see “I.II. Translation determinants” on page 17). With the cell extract originating in *E. coli*, we should assume an important distinction of parameters such as CAI, tAI, and GC%. In fact, the WT sequence exhibits the lowest GC and GC3% of all sequences; this characteristic may have been evolutionarily selected. If GC content plays a critical *in vitro* role, then the role for predicted ΔG would become secondary. More importantly, the artificial addition of tRNA to the reaction mixture certainly modifies the tAI which disrupts the expected CAI. This further reveals the importance of unbiased values such as ΔG and GC% accounting for protein expression. Lastly, the folding stability of intron sequences has been shown to influence CUB and mRNA levels (Trotta, 2013) and this *in vitro* system has altered this natural setting. Similarly, 5' UTRs secondary structures (Ciandrini et al., 2013) and candidate regulatory RNA motifs in 3' UTRs (Wan et al., 2012) are changed in the absence of true cells, more so of pneumococcus.

All things considered, we can't neglect the importance of evolution in producing the WT sequence when compared to the engineered ones. Nonetheless, we can appreciate the overall dissimilarity between sequences upon the reaction start and endpoint suggesting that protein production occurs much faster and earlier in the reaction according to favorable translation characteristics (such as ΔG and GC content but not CAI for *E. coli* HEG, Figure VI.4C). This demonstrates a role for translation initiation, as the sequence differences are exclusive to the beginning of the gene. We also know that the rate-limiting step in protein production is usually considered to be translation initiation (Jacques and Dreyfus, 1990; Kozak, 1999; Kudla et al., 2009; Plotkin and Kudla, 2011; Shah et al., 2013) and initiation is also controlled by the amount of single-stranded

surface area able to make contact with the ribosome (Espah Borujeni et al., 2014; Kozak, 1986).

Site accessibility, partial mRNA unfolding, and ribosomal sliding are entwined (Espah Borujeni et al., 2014) when adding up to determine a translation rate. The later time points of *in vitro* translation show clear differences in protein levels, although higher ΔG sequences exhibit comparable rates (seen in log scale; not shown). One possible explanation suggests that the reaction, attaining equilibrium between maximum mRNA molecules and maximum ribosome occupancy, is then limited by the release of ribosomes terminating translation. Previous studies suggest that the unwound 5'-secondary structures refold slowly relative to rates of translation initiation (Sagliocco et al., 1993) and that translation is typically limited by the availability of free ribosomes (Shah et al., 2013). Therefore, re-initiation with recycled ribosomes may not be affected by the thermodynamic stability of the 5' mRNA. This could explain the comparable rates of exponential synthesis in late but not early time points. Together, these data support the idea that translation initiation is a crucial limiting or enabling step in protein expression (Plotkin and Kudla, 2011) and suggest that it is causing our observed protein abundances.

Taking into account all of our experimental data and the available scientific knowledge we shaped a simplified model (Figure IV.2) of the consequences of mRNA folding free energy variations. Stability and overall abundance of mRNA molecules with silent mutations in the 5' region are comparable. Lower energies represent tightly folded secondary structures. These mRNAs are less accessible for ribosome binding and have slower translation initiation. Higher ΔG represent structures readily available for translation. The WT sequences tend to lie in between. Gene expression differences become apparent over time, taking into account cell division and bacterial growth thus rendering a gradient of PLY protein amounts (Figure IV.2).

CHAPTER IV. Materials and Methods

Calculation of ΔG

Sequence folding free energies were calculated using the DINAMelt/Quikfold web server (<http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold>) (Markham and Zuker, 2005). The UNAFold software uses predicted free energies at 37 °C and enthalpies from the Turner laboratory at the University of Rochester in Rochester, NY. For ΔG calculations we used the more up-to-date free energies for RNA denoted as version 3.0 energies.

Bacteria, primers, and plasmids

Plasmids for *in vitro* translation were assembled in *E. coli* TOP10 using plasmid pEXP5-NT/CALML3 (Invitrogen) via XbaI-SacII restriction sites. We included the ribosome binding site and fused 5'-PLY in frame with full length GFP using 4 forward primers (ultramers, IDT) that differed in PLY region according to Figure II.1A with the following degenerate 5'-GCC GCC TCT AGA AGG AGA TAGAAA **ATG** GCN AAY AAR GCN GTN AAY GAY TTY ATH YTN GCN **ATG** GCT AGC AAA GGA GAA GAA C-3' (start codons in bold), and the reverse primer was 5'-GGC GGC CCG CGG **TTA** TTT GTA GAG C-3' (stop ochre codon in bold).

E. coli were grown in LB or agar supplemented with appropriate antibiotics for plasmid selection (ampicillin 200 µg/mL or kanamycin 50 µg/mL).

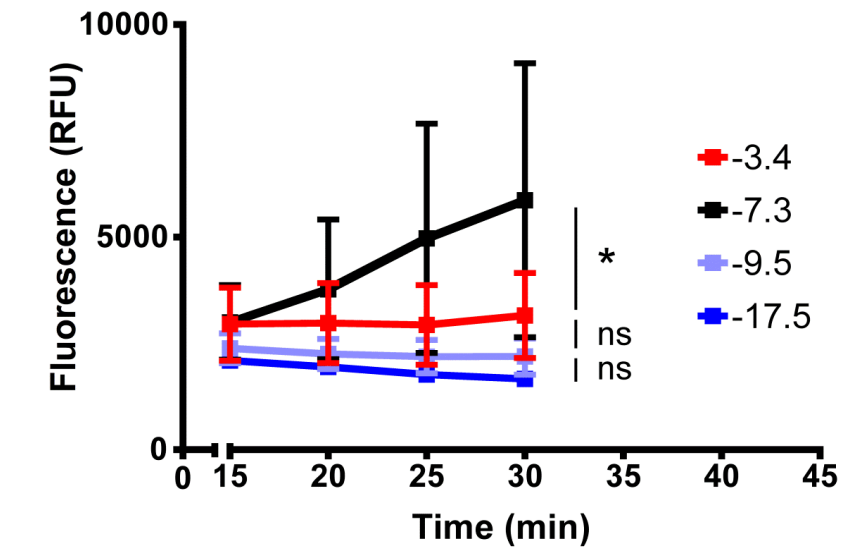
In vitro translation

We performed *in vitro* translation assays using Expressway™ Cell-Free *E. coli* Expression System (Invitrogen) following manufacturer's instructions. We monitored expression of fused PLY-GFP protein live with an Infinite® M200 (Tecan) plate reader.

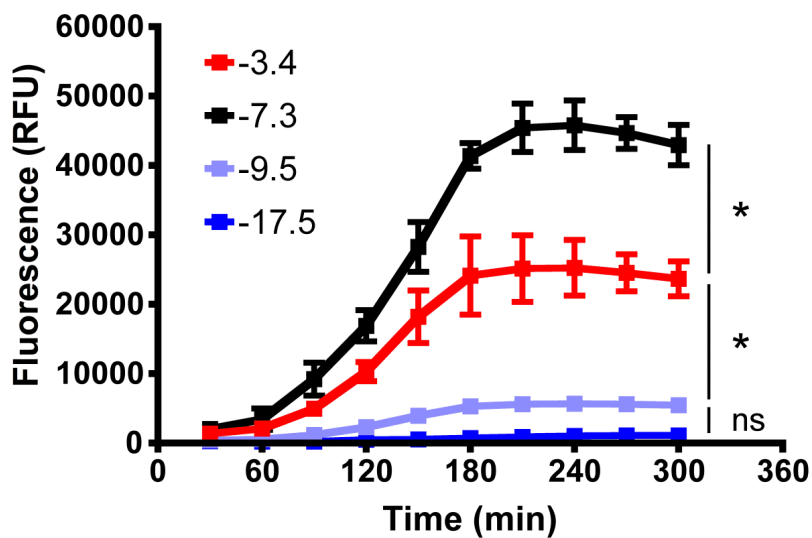
Statistical analysis

Statistics were performed with Prism 5 software (GraphPad, Inc.). Data were analyzed by analysis of variance or Kruskal-Wallis tests followed by post-tests to account for multiple comparisons, as appropriate. *In vitro* experiments were performed at least three times with at least three technical replicates (for experiments other than western blots). Sample sizes for *in vivo* experiments are indicated in the figure legends. Bars represent SEM. Values of $P \leq 0.05$ were considered significant.

CHAPTER IV. Figures



A



B

Figure IV.1. Production of PLY increases due to accelerated translation initiation.

GFP fluorescence, measured in relative fluorescence units (RFU), with added in-frame PLY 5' mRNA leading sequences with different mRNA folding energies as measure of protein production over time by *in vitro* translation assay in *E. coli* cell-free extract. (A) First 30 min and (B) subsequent 5h after feeding buffer addition. Data are shown as mean \pm SEM and were analyzed by One-way ANOVA followed by Bonferroni's Multiple Comparison Test. * $p < 0.05$; ns (not significant).

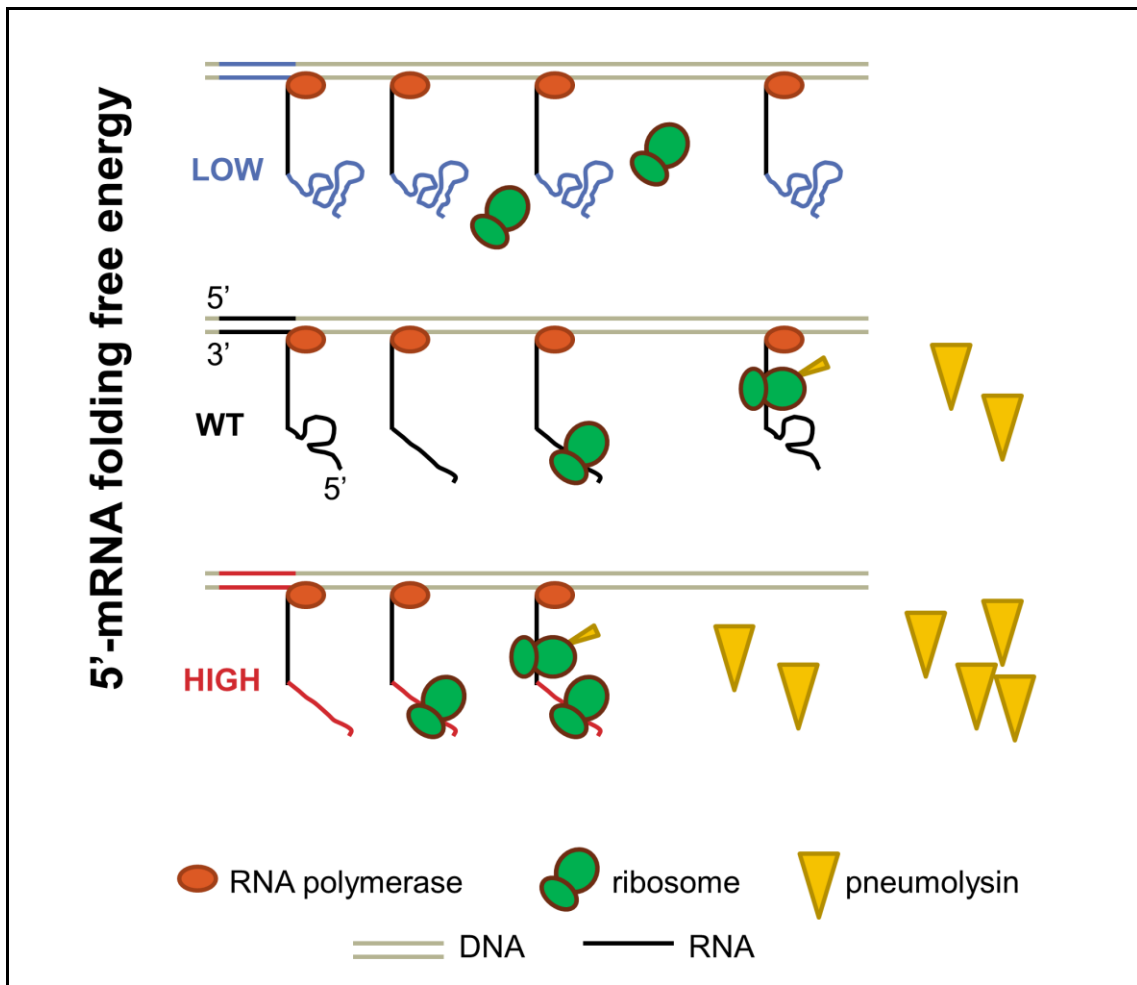


Figure IV.2. Model for translation initiation delay when modulating 5' *ply* mRNA ΔG .

Lower energies represent tightly folded secondary structures. These less accessible mRNAs take more time to bind ribosomes. Consequently, these sequences have slower translation initiation and will delay the recycling of ribosomes if the 5' region is allowed to refold. Higher ΔG represent the opposite; structures are readily available for translation that happens more efficiently. The WT sequence illustrated tends to present in between scenarios.

CHAPTER V. GENERAL DISCUSSION

This chapter will provide a critical analysis of results in light of the state of the art, conclude the research with a summary of key findings, and suggest continuing work.

The genetic code is redundant since the majority of amino acids may be coded by more than one codon (except for methionine and tryptophan). These coding possibilities and all the reasons and conditions stated in the introduction of this thesis, result in diverse biases in terms of codon usage, tRNA pool constitution, or percentage of GC and CpG dinucleotide. Horizontal gene transfer, evolution and selective pressure for efficiency and/or accuracy, genetic variations, as well as gene ontology and function (Ciandrini et al., 2013) all contribute to a diversity of CUB and of translation efficiencies. These are, in fact, very important with regards to the outcome of protein expression. Though, the crucial translation determinant that is the basis for this thesis was named after Willard Gibbs, who first described free enthalpy or Gibbs energy (ΔG) in 1873.

Pertaining to pneumolysin, the toxin of *S. pneumoniae*, the mechanism that naturally regulates it is barely understood. The manipulation of the folding free energy inherent to the 5' mRNA region of *ply* gene, by means of silent mutations, proved to be a successful way to engineer rheostat control.

5' mRNA ΔG correlates with PLY expression *in vitro*

We have generated a panel of bacteria (see Appendix B, "Lists of bacterial strains created" on page 113) that are isogenic within their groups with the exception of 5' *ply* gene. They represent a wide array of ΔG s that range from -17.5 up to -0.2 kcal/mol.

When calculating the ΔG there are a number of possible predicted structures associated to each value. Being that systems attempt to minimize the energy, each sequence is recognized by its minimum free energy (see Appendix C, “Predicted mRNA folding secondary structures” on page 117).

In vitro, we saw that *S. pneumoniae* R6 strains all grew alike (D39 as well; data not shown), at the same rate, with a doubling time of about 40 minutes. Even though they express different amounts of PLY there’s no apparent deficit in growth fitness (see Appendix D, “Growth curves of *S. pneumoniae*” on page 123).

In appendix E, “The limits of PLY production with increasing ΔG ” (page 125), we discuss the inability to increase protein expression beyond a certain cap. We hypothesize that the existence of multiple conformations with different than minimum energies may present altered translation initiation rates. The consequence of this is the noted inconsistency of protein production and potentially increase in the relative importance of other elements of translation efficiency. For this matter, we continued our studies with the established panel of strains that provide consistent results and have low number of possible foldings.

A couple of other designed strains had distinct features. It is agreeable that translation efficiency is determined by both codon bias and folding energy (Tuller et al., 2010b) and CAI seems to be more important when mRNA structure is weaker and folding free energy higher (Gingold and Pilpel, 2011; Jia and Li, 2005; Plotkin and Kudla, 2011; Supek and Smuc, 2010; Tuller et al., 2010a). To understand the effects of CAI in PLY expression we engineered a fully codon optimized (CO) *ply* gene and a strain whose 5’ sequence and ΔG (-6.2 kcal/mol) was CO. We also engineered a strain with codon deoptimization in that same 5’ region, which had the interesting high ΔG value of -2.0 kcal/mol. In appendix F, “Codon optimization of PLY” (page 127), we analyze these strains and theorize on what their results mean for protein translation determinants. In line with the limits of PLY production, these strains present multiple predicted foldings with different free energies associated. This alone could be a reason for inconsistency and lack of meaningful differences in results. The selection of preferred versus unpreferred codons in this case leads to an inversely correlated ΔG despite similar GC%. These dynamics acting centrifugally, in a theoretical way, are not the best showcases for each

determinant. Furthermore, full optimization of PLY seems to bring little difference to expression when compared to our high-PLY-yield strain or to the WT. When compared to the 5' CO strain we are more convinced of the importance of translation initiation in the entire process of efficiency as these strains behave similarly. Again, with the caveat of inconsistency, we pursued our studies with the established panel of strains that present similar CAI as these are considerably more aligned with the ΔG .

To evaluate virulence parameters *in vitro* as a function of PLY abundance we used human red-blood cell lysis assays and immortalized human respiratory epithelial cells cytotoxicity assays measuring LDH release (Figure II.4 and Figure II.5). The ability of PLY to form very relatively large pores in the membranes of these cell types leads to lysis in one assay and release of LDH in the other. More PLY present leads to enhanced lysis. We were able to significantly and positively correlate 5' mRNA ΔG with lysis and LDH release which would be due to the amounts of PLY present in the samples. We also measured PLY production across the strains to see observe this correlation both by western blot analysis and by ELISA (Figure II.1). This is in accordance to previous studies that link 5' mRNA ΔG and translation efficiency (Goodman et al., 2013; Kudla et al., 2009). Unlike other studies, we were able to solely manipulate the 5' region of the mRNA to produce changes in protein expression whether abolishing (Coleman et al., 2011) or enhancing it.

Recently a group of researchers demonstrated that synonymous codon usage may in fact alter translation kinetics but also produce changes in protein folding (Buhr et al., 2016). Potentially, the rates of translation have secondary repercussions on the chance of exposure to chaperons and protein modifiers. In our strains, ratios between measured protein and hemolysis percentage as a measure of function were all similar to WT (data not shown) discarding the possibility of misfolding. Nonetheless, the point is valid and important to mind when conceiving protein synthesis modifiers.

Control of pneumococcal virulence *in vivo* by 5' *ply* mRNA Δ G

Several animal models are available to put these strains to test. However, with two backgrounds of pneumococcus and their manipulation particularities, we tried more than a model.

We used *Drosophila melanogaster* because they are susceptible to R6 strains, provide a high number of animals in short periods of time, with a simplistic model of infection that could provide clearness to results (see Appendix G, “*Drosophila melanogaster* model for rheostat pneumolysin expression” on page 131). Nevertheless, fly results were very inconsistent and several attempts of optimization didn't provide enough evidence to answer our hypothesis.

PLY activates p38 (Ratner et al., 2006) and to study this we had access to a *Caenorhabditis elegans* model of infection, that was simple to setup and quick to provide results (see Appendix H, “*Caenorhabditis elegans* model for rheostat pneumolysin expression” on page 133). Results in this model weren't very evident and would have required further work.

The better model for our hypothesis was the mouse. We used the intranasal mode of infection in WT mice. We know that PLY does not affect bacterial growth *in vitro* but conditions *in vivo* are not the same, and PLY may play a role in growth (Los et al., 2013). This effect may contribute to the observed *in vivo* results. We hypothesize that low PLY production facilitates immune system evasion and may explain the higher CFUs of the strain with the lowest 5' mRNA Δ G (Figure III.2). However, some authors found that PLY-KO was associated with lower numbers of pneumococcus in the respiratory tract (Kadioglu et al., 2002). Mouse lung sections showed an increase in inflammatory signs that was related to the strain Δ G (Figure III.2) by way of diverse PLY amounts acting in the lung parenchyma (Maus et al., 2004; Witzentrath et al., 2006). As previously studied, PLY is able to activate, induce, enhance, and potentiate host immune responses (Cockeran et al., 2002; Los et al., 2013). Whether in response to PLY or to pneumococci, we saw immune cell recruitment not found in uninfected animals. PMN and DC were

significantly recruited to the site of infection, more so in the higher ΔG strains, as previously shown (Dessing et al., 2009; Moreland and Bailey, 2006) but this was not statistically significant. M Φ depletion was seen and it correlated with increase in ΔG (Figure III.2 and Figure III.3). One outcome of the cell influx is the production and release of proinflammatory cytokines such as IL-1 β and TNF- α . We quantified an upregulation in response to PLY and ΔG further supporting our hypothesis. Lastly, we were able to retrieve and measure PLY in the BAL fluid (Figure III.1). We show the expected gradient of PLY concentrations in accordance with ΔG but, surprisingly, we detected PLY in the lower ΔG strains. This was unexpected because we could not do it *in vitro*. We conjecture that *in vivo* conditions present other stimuli to pneumococci that upregulate PLY protein expression. This is acceptable since the *ply* gene is present with only few synonymous mutations. More research is necessary to better our understanding of PLY-host interactions and this technique is a great tool for that purpose.

The strategy to rationally change gene expression has allowed us to study the importance of PLY concentration on host-bacterial interactions as they strongly correlate with 5' *ply* mRNA ΔG .

Translation initiation drives the differential rates of protein translation

In our strains, we found that *ply* mRNA was present in the same amounts across them (Figure II.2). Since, the mRNA molecules would be the important theoretical difference across the strains – aside from the DNA – we needed to evaluate the stability and decay of these molecules (Figure II.3). Whether by decay or use of small non-coding RNAs, called sRNAs, the mRNA has been known to regulate protein expression (Richards et al., 2008). In bacteria, however, there seems to exist unspecific machinery for degradation such as the RNases that don't distinguish mRNAs and classes of RNA (Deutscher, 2006). Across all tested strains we saw no differences in decay rates and half-life times. These results pointed to the importance of translation rate and efficiency. To address this issue we conceived an *in vitro* setting to test this hypothesis. Our results indicate the relevance of translation initiation as the engine driving the observed distinct PLY

quantities and effects on hosts (Figure IV.1) as other authors have seen (Jacques and Dreyfus, 1990; Kozak, 1999; Kudla et al., 2009; Plotkin and Kudla, 2011; Shah et al., 2013). Nonetheless, this experiment presented several limitations already addressed in “CHAPTER IV. Discussion” (page 71), namely that the WT sequence exhibited the highest yields of protein expression.

The debate about which determinant of translation is more important is situational at best. Weak 5' mRNA structures exhibit more correlation with CAI (Angov, 2011; Gingold and Pilpel, 2011; Plotkin and Kudla, 2011; Supek and Smuc, 2010; Tuller et al., 2010a; Tuller et al., 2010b). It is consensual that the choice of unpreferred codons leads to decrease in gene expression perhaps due to tRNA scarcity or poor codon-tRNA interactions (Hockenberry et al., 2014). This may, however, serve as an important functional aspect that can be encoded thus. Some authors critique the generic measure of CUB (i.e. CAI) disregarding the contribution of individual codons at specific sites. The choice of a codon may not be so random when secondary structure is considered in the equation (Trotta, 2013). A new mathematical model to show that individual CUBs follow a position-dependency for each codon in *E. coli* was developed and could be of interest (Hockenberry et al., 2014). This postulates that some codons may be preferred in certain positions but also be suboptimal in other regions of the gene – or other genes. Some authors used *Saccharomyces cerevisiae* genome-wide experimental data sets of ribosomal density on mRNAs integrated with a stochastic simulation model describing ribosome traffic dynamics during translation elongation to reveal that in a position-dependent manner, rather than simply CUB, codons have key roles in determining translational efficiency (Ciandrini et al., 2013). They have also estimated mRNA-specific translation initiation rates across the entire transcriptome and identified different classes of mRNAs characterized by their initiation rates, their ribosome traffic dynamics, and by their response to ribosome availability (Ciandrini et al., 2013). Ribosomes interact with structured sites upstream of the Shine-Dalgarno sequences and can modulate translation initiation rates (Espah Borujeni et al., 2014) however, in this work, regions upstream, downstream and, for that matter, the great majority (99.4 to 99.7%) of *ply* gene are unchanged. Translation in yeast, can be limited by the availability of free ribosomes (Shah et al., 2013).

We designed a model, which can be seen in Figure IV.2, to attempt a visual mechanistic clarification. Lower energies represent tightly folded secondary structures. These less accessible mRNAs take more time and energy to bind ribosomes. Consequently, these sequences have slower translation initiation and will delay the recycling of ribosomes if the 5' region is allowed to refold. Higher ΔG represent the opposite; structures are readily available for translation that happens more efficiently. The WT sequence illustrated tends to present in between scenarios. Taking into account cell cycle, bacterial growth, outside stimuli, and unknown determinants therefore resulting in PLY translation differences associated to the 5' *ply* mRNA ΔG .

PLY and PLY toxoids as vaccine candidates

We showed that bacterial virulence modulation *in vivo* can be engineered using the 5' mRNA ΔG associated to virulence factors without the need for chemical inducers or heterologous promoters. This is a good setting for toxoid manipulation in live attenuated vaccination strategies. Even though, in the wild, differences in virulence among certain pneumococcal serotypes have been shown not attributable to differences in PLY production (Benton et al., 1997) the manipulation of this toxin's abundance may prove useful in the design of new vaccines.

PLY and PLY toxoids are vaccine candidates that have been studied multiple times and can provide capsular serotype-independent immunity (Alexander et al., 1994; Cockeran et al., 2005; Denoel et al., 2011; Kadioglu and Andrew, 2004; Malley et al., 2005; Moffitt and Malley, 2011; Tai, 2006; van Rossum et al., 2005). Phase I/II studies have proven the safety of PLY toxoids (Berglund et al., 2014; Leroux-Roels et al., 2014; Prymula et al., 2014). Authors are divided supporting the efficacy of immunity acquired with these novel vaccines (Verhoeven et al., 2014; Xu et al., 2015) or considering PLY toxoids as adjuvants to other molecules (Lu et al., 2014). Our findings are an important step towards better engineering of new vaccines, potentially live attenuated.

As the relative age of our population increases whereas also increasing continuously in numbers, the settings of pneumonia occurrence appear as more problematic. Together

with antibiotic failure, lack of effective vaccination that is long-lasting and of universal coverage provides difficult circumstances that need to be thought about. New therapies are being researched but further work is still necessary (Lucas et al., 2013).

Evolution prepared pathogens for mutation, adaptation, and change. We, as hosts, are also under pressure to better resist and endure microbes. This continued turnover and balance is an effect called Red Queen hypothesis. There is a risk of Red Queen dynamics in the field of pneumococcal vaccination, particularly because of serotype exclusion (Jefferies et al., 2011). Urgent research in new methodologies and strategies is required.

CONCLUSIONS

Using unmarked, isogenic pneumococcal strains, we showed that rheostat-like control of PLY production is feasible and that changes in 5' mRNA ΔG are sufficient to alter gene expression both *in vitro* and during *in vivo* infection using a mouse model of pneumonia. Alteration of expression did not depend on mRNA abundance or stability but rather it seemed to occur at the level of translation initiation. As predicted, there was a direct correlation between PLY production and target cell lysis using simplified *in vitro* systems, but the role of PLY *in vivo* appears to be more complex. Even so, we were able to show inflammatory states that escalate in a PLY-dependent manner.

With minimal manipulation of 5' mRNA ΔG we can achieve various protein expression patterns encoded in the native locus, under the same controlling mechanisms, subjected to normal feedback loops.

In conclusion, the effects of major virulence factors, or any protein, cannot be thoroughly explained using reductionist KO vs WT models. We confirm that rheostat control of virulence factors is useful and provides proof-of-principle as a new tool for understanding the role of specific factors upon infection.

Future directions and contributions

Further work to understand the mechanism could encompass imaging of polysomes by transmission electron microscopy or by ribosome footprinting. These techniques may illustrate ribosome density in varyingly designed sequences. Other techniques can include crystallography or nuclear magnetic resonance spectroscopy (Felden, 2007). Alternatively, the overall technique may be enhanced by considering other translation determinants in the design, such as GC%, CAI, CPs, codon position, and tAI.

Synonymous codon usage also plays a role in higher organisms. Maybe surprisingly, but there are a number of human disorders that arise from silent mutations that affect mRNA in its stability or splicing, thus affecting protein function or regulation, among other unknown effects (Bali and Bebok, 2015). Studying synonymous mutations has profound impacts in human health on multiple levels that merit support and continuous research.

PLY and PLY toxoids are vaccine candidates that can provide capsular serotype-independent immunity (Alexander et al., 1994; Cockeran et al., 2005; Denoel et al., 2011; Kadioglu and Andrew, 2004; Malley et al., 2005; Moffitt and Malley, 2011; Tai, 2006; van Rossum et al., 2005). Our findings are an important step towards better engineering of new vaccines, potentially live attenuated.

Alternate pathways to synthetically design the reduction of translation efficiency have been studied as safe methods to generate live attenuated vaccines. Conversely, the ability to overexpress an engineered safe (Berglund et al., 2014) non-hemolytic PLY toxoid under native pneumococcal regulatory elements can potentially generate more effective serotype-independent vaccines, in the case of pneumococcus.

Given the increasing availability of whole genome sequences, a lot can be learned from the impact of relative expression levels on pathogenesis. Ultimately, this technique can be used to achieve rheostat control of virtually any protein with minimal genomic manipulation for applications in medicine, science and industry.

GENERAL REFERENCES

(A to Z)

- Agashe, D., Martinez-Gomez, C.N., Drummond, A.D., and Marx, C.J. (2012). Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Molecular Biology and Evolution*.
- Aguilar, J.L., Kulkarni, R., Randis, T.M., Soman, S., Kikuchi, A., Yin, Y., and Ratner, A.J. (2009). Phosphatase-Dependent Regulation of Epithelial Mitogen-Activated Protein Kinase Responses to Toxin-Induced Membrane Pores. *PLoS ONE* 4.
- Alexander, J., Berry, A., Paton, J., Rubins, J., Andrew, P., and Mitchell, T. (1998). Amino acid changes affecting the activity of pneumolysin alter the behaviour of pneumococci in pneumonia. *Microbial pathogenesis* 24, 167-174.
- Alexander, J., Lock, R., Peeters, C., Poolman, J., Andrew, P., Mitchell, T., Hansman, D., and Paton, J. (1994). Immunization of mice with pneumolysin toxoid confers a significant degree of protection against at least nine serotypes of *Streptococcus pneumoniae*. *Infection and immunity* 62, 5683-5688.
- Amaral, F.E., Parker, D., Randis, T.M., Kulkarni, R., Prince, A.S., Shirasu-Hiza, M.M., and Ratner, A.J. (2015). Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by *Streptococcus pneumoniae*. *PloS one* 10.
- Angov, E. (2011). Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnology Journal* 6, 650-659.
- Assaad, U., El-Masri, I., Porhomayon, J., and El-Solh, A.A. (2012). Pneumonia immunization in older adults: review of vaccine effectiveness and strategies. *Clin Interv Aging* 7, 453-461.
- Atkinson, N.J., Witteveldt, J., Evans, D.J., and Simmonds, P. (2014). The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic acids research* 42, 4527-4545.
- Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type III. *J Exp Med* 79, 137-158.
- Babbitt, G.A., Alawad, M.A., Schulze, K.V., and Hudson, A.O. (2014). Synonymous codon bias and functional constraint on GC3-related DNA backbone dynamics in the prokaryotic nucleoid. *Nucleic Acids Res* 42, 10915-10926.

- Balachandran, P., Hollingshead, S.K., Paton, J.C., and Briles, D.E. (2001). The Autolytic Enzyme LytA of *Streptococcus pneumoniae* Is Not Responsible for Releasing Pneumolysin. *Journal of Bacteriology* 183, 31083116.
- Bali, V., and Bebok, Z. (2015). Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *Int J Biochem Cell Biol* 64, 58-74.
- Benton, K.A., Everson, M.P., and Briles, D.E. (1995). A pneumolysin-negative mutant of *Streptococcus pneumoniae* causes chronic bacteremia rather than acute sepsis in mice. *Infection and immunity* 63, 448-455.
- Benton, K.A., Paton, J.C., and Briles, D.E. (1997). Differences in virulence for mice among *Streptococcus pneumoniae* strains of capsular types 2, 3, 4, 5, and 6 are not attributable to differences in pneumolysin production. *Infection and immunity* 65, 1237-1244.
- Benton, K.A., VanCott, J.L., and Briles, D.E. (1998). Role of tumor necrosis factor alpha in the host response of mice to bacteremia caused by pneumolysin-deficient *Streptococcus pneumoniae*. *Infection and immunity* 66, 839-842.
- Berglund, J., Vink, P., Tavares Da Silva, F., Lestrade, P., and Boutriau, D. (2014). Safety, immunogenicity, and antibody persistence following an investigational *Streptococcus pneumoniae* and *Haemophilus influenzae* triple-protein vaccine in a phase 1 randomized controlled study in healthy adults. *Clin Vaccine Immunol* 21, 56-65.
- Berry, A., Alexander, J., Mitchell, T., Andrew, P., Hansman, D., and Paton, J. (1995). Effect of defined point mutations in the pneumolysin gene on the virulence of *Streptococcus pneumoniae*. *Infection and immunity* 63, 1969-1974.
- Berry, A., and Paton, J. (2000). Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infection and immunity* 68, 133-140.
- Berry, A.M., Ogunniyi, A.D., Miller, D.C., and Paton, J.C. (1999). Comparative virulence of *Streptococcus pneumoniae* strains with insertion-duplication, point, and deletion mutations in the pneumolysin gene. *Infection and immunity* 67, 981-985.
- Berry, A.M., Yother, J., Briles, D.E., Hansman, D., and Paton, J.C. (1989). Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infect Immun* 57, 2037-2042.
- Bertrand, R.L., Abdel-Hameed, M., and Sorensen, J.L. (2015). Limitations of the 'ambush hypothesis' at the single-gene scale: what codon biases are to blame? *Molecular genetics and genomics* : MGG 290, 493-504.
- Branger, J., Knapp, S., Weijer, S., Leemans, J.C., Pater, J.M., Speelman, P., Florquin, S., and van der Poll, T. (2004). Role of Toll-Like Receptor 4 in Gram-Positive and Gram-Negative Pneumonia in Mice. *Infection and Immunity* 72, 788794.
- Braun, J.S., Novak, R., Gao, G., Murray, P.J., and Shenep, J.L. (1999). Pneumolysin, a protein toxin of *Streptococcus pneumoniae*, induces nitric oxide production from macrophages. *Infect Immun* 67, 3750-3756.

- Briles, D.E., Hollingshead, S.K., Paton, J.C., Ades, E.W., Novak, L., van Ginkel, F.W., and Benjamin, W.H., Jr. (2003). Immunizations with pneumococcal surface protein A and pneumolysin are protective against pneumonia in a murine model of pulmonary infection with *Streptococcus pneumoniae*. *J Infect Dis* 188, 339-348.
- Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M.V., and Komar, A.A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol Cell* 61, 341-351.
- Butler, J.C., Farley, M.M., Harrison, L.H., McGeer, A., Kolczak, M.S., Breiman, R.F., and the Team, A. (2000). Cigarette Smoking and Invasive Pneumococcal Disease. *The New England Journal of Medicine*.
- Carlini, D.B. (2004). Experimental reduction of codon bias in the *Drosophila* alcohol dehydrogenase gene results in decreased ethanol tolerance of adult flies. *J Evol Biol* 17, 779-785.
- Carlini, D.B., and Stephan, W. (2003). In vivo introduction of unpreferred synonymous codons into the *Drosophila* Adh gene results in reduced levels of ADH protein. *Genetics* 163, 239-243.
- Chiavolini, D., Pozzi, G., and Ricci, S. (2008). Animal models of *Streptococcus pneumoniae* disease. *Clin Microbiol Rev* 21, 666-685.
- Ciandrini, L., Stansfield, I., and Romano, C.M. (2013). Ribosome Traffic on mRNAs Maps to Gene Ontology: Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation. *PLoS Computational Biology* 9.
- Cockeran, R., Anderson, R., and Feldman, C. (2002). The role of pneumolysin in the pathogenesis of *Streptococcus pneumoniae* infection. *Current Opinion in Infectious Diseases* 15, 235.
- Cockeran, R., Anderson, R., and Feldman, C. (2005). Pneumolysin as a vaccine and drug target in the prevention and treatment of invasive pneumococcal disease. *Arch Immunol Ther Exp (Warsz)* 53, 189-198.
- Cockeran, R., Steel, H.C., Mitchell, T.J., Feldman, C., and Anderson, R. (2001). Pneumolysin Potentiates Production of Prostaglandin E2 and Leukotriene B4 by Human Neutrophils. *Infection and Immunity* 69, 3494-3496.
- Coleman, R.J., Papamichail, D., Skiena, S., Fitcher, B., Wimmer, E., and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science (New York, NY)* 320, 1784-1787.
- Coleman, R.J., Papamichail, D., Yano, M., García-Suárez, M., and Pirofski, L.-A. (2011). Designed reduction of *Streptococcus pneumoniae* pathogenicity via synthetic changes in virulence factor codon-pair bias. *The Journal of infectious diseases* 203, 1264-1273.
- Corley, M., Solem, A., Qu, K., Chang, H.Y., and Laederach, A. (2015). Detecting riboSNitches with RNA folding algorithms: a genome-wide benchmark. *Nucleic acids research* 43, 1859-1868.
- Crick, F.H. (1968). The origin of the genetic code. *J Mol Biol* 38, 367-379.

- Crook, J., Tharpe, J.A., Johnson, S.E., Williams, D.B., Stinson, A.R., Facklam, R.R., Ades, E.W., Carlone, G.M., and Sampson, J.S. (1998). Immunoreactivity of five monoclonal antibodies against the 37-kilodalton common cell wall protein (PsaA) of *Streptococcus pneumoniae*. *Clin Diagn Lab Immunol* 5, 205-210.
- del García-Suárez, M., Flórez, N., Astudillo, A., Vázquez, F., Villaverde, R., Fabrizio, K., Pirofski, L.-A., and Méndez, F.J. (2007). The role of pneumolysin in mediating lung damage in a lethal pneumococcal pneumonia murine model. *Respiratory Research* 8, 3.
- Denoel, P., Philipp, M.T., Doyle, L., Martin, D., Carletti, G., and Poolman, J.T. (2011). A protein-based pneumococcal vaccine protects rhesus macaques from pneumonia after experimental infection with *Streptococcus pneumoniae*. *Vaccine* 29, 5495-5501.
- Dessing, M.C., Florquin, S., Paton, J.C., and Poll, T. (2008). Toll - like receptor 2 contributes to antibacterial defence against pneumolysin - deficient pneumococci. *Cellular Microbiology* 10, 237-246.
- Dessing, M.C., Hirst, R.A., de Vos, A.F., and van der Poll, T. (2009). Role of Toll-Like Receptors 2 and 4 in Pulmonary Inflammation and Injury Induced by Pneumolysin in Mice. *PLoS ONE* 4.
- Deutscher, M.P. (2006). Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Research* 34, 659-666.
- Dockrell, D.H., Marriott, H.M., Prince, L.R., Ridger, V.C., Ince, P.G., Hellewell, P.G., and Whyte, M.K. (2003). Alveolar macrophage apoptosis contributes to pneumococcal clearance in a resolving model of pulmonary infection. *J Immunol* 171, 5380-5388.
- Dockrell, D.H., Whyte, M.K., and Mitchell, T.J. (2012). Pneumococcal pneumonia: mechanisms of infection and resolution. *Chest* 142, 482-491.
- Eberhardt, A., Hoyland, C.N., Vollmer, D., Bisle, S., Cleverley, R.M., Johnsborg, O., Håvarstein, L.S., Lewis, R.J., and Vollmer, W. (2012). Attachment of capsular polysaccharide to the cell wall in *Streptococcus pneumoniae*. *Microbial drug resistance (Larchmont, NY)* 18, 240-255.
- Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* 42, 2646-2659.
- Fang, R., Tsuchiya, K., Kawamura, I., Shen, Y., Hara, H., Sakai, S., Yamamoto, T., Fernandes-Alnemri, T., Yang, R., Hernandez-Cuellar, E., et al. (2011). Critical roles of ASC inflammasomes in caspase-1 activation and host innate resistance to *Streptococcus pneumoniae* infection. *J Immunol* 187, 4890-4899.
- Felden, B. (2007). RNA structure: experimental analysis. *Current Opinion in Microbiology* 10, 286-291.
- Feldman, C., and Anderson, R. (2014). Recent advances in our understanding of *Streptococcus pneumoniae* infection. *F1000Prime Rep* 6, 82.
- Gaspar, P., Moura, G., Santos, M.A.S., and Oliveira, J. (2013). mRNA secondary structure optimization using a correlated stem-loop prediction. *Nucleic Acids Research* 41.

- Gasse, P., Mary, C., Guenon, I., Noulin, N., Charron, S., Schnyder-Candrian, S., Schnyder, B., Akira, S., Quesniaux, V.F., Lagente, V., et al. (2007). IL-1R1/MyD88 signaling and the inflammasome are essential in pulmonary inflammation and fibrosis in mice. *J Clin Invest* 117, 3786-3799.
- Gelber, S.E., Aguilar, J.L., Lewis, K.L.T., and Ratner, A.J. (2008). Functional and phylogenetic characterization of Vaginolysin, the human-specific cytolysin from *Gardnerella vaginalis*. *Journal of bacteriology* 190, 3896-3903.
- Gilbert, R., Heenan, R., Timmins, P., Gingles, N., Mitchell, T., Rowe, A., Rossjohn, J., Parker, M., Andrew, P., and Byron, O. (1999a). Studies on the structure and mechanism of a bacterial protein toxin by analytical ultracentrifugation and small-angle neutron scattering. *Journal of molecular biology* 293, 1145-1160.
- Gilbert, R., Jiménez, J.L., Chen, S., Tickle, I.J., Rossjohn, J., Parker, M., Andrew, P.W., and Saibil, H.R. (1999b). Two Structural Transitions in Membrane Pore Formation by Pneumolysin, the Pore-Forming Toxin of *Streptococcus pneumoniae*. *Cell* 97.
- Gingold, H., and Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. *Molecular Systems Biology* 7.
- Goodman, D.B., Church, G.M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science (New York, NY)* 342, 475-479.
- Goulart, C., da Silva, T.R., Rodriguez, D., Politano, W.R., Leite, L.C.C., and Darrieux, M. (2013). Characterization of protective immune responses induced by pneumococcal surface protein A in fusion with pneumolysin derivatives. *PloS one* 8.
- Gray, C., Ahmed, M.S., Mubarak, A., Kasbekar, A.V., Derbyshire, S., McCormick, M.S., Mughal, M.K., McNamara, P.S., Mitchell, T., and Zhang, Q. (2014). Activation of memory Th17 cells by domain 4 pneumolysin in human nasopharynx-associated lymphoid tissue and its association with pneumococcal carriage. *Mucosal immunology* 7, 705-717.
- Greiner, O., Day, P.J., Bosshard, P.P., Imeri, F., Altwegg, M., and Nadal, D. (2001). Quantitative detection of *Streptococcus pneumoniae* in nasopharyngeal secretions by real-time PCR. *J Clin Microbiol* 39, 3129-3134.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. *Nucleic Acids Res* 36, W70-74.
- Gu, C., Begley, T.J., and Dedon, P.C. (2014). tRNA modifications regulate translation during cellular stress. *FEBS letters* 588, 4287-4296.
- Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology* 22, 346-353.
- Ha, U.-H., Lim, J., Kim, H.-J., Wu, W., Jin, S., Xu, H., and Li, J.-D. (2008). MKP1 Regulates the Induction of MUC5AC Mucin by *Streptococcus pneumoniae* Pneumolysin by Inhibiting the PAK4-JNK Signaling Pathway. *Journal of Biological Chemistry* 283, 30624-30631.

- Hajaj, B., Yesilkaya, H., Benisty, R., David, M., Andrew, P.W., and Porat, N. (2012). Thiol peroxidase is an important component of *Streptococcus pneumoniae* in oxygenated environments. *Infection and immunity* 80, 4333-4343.
- Hakenbeck, R. (2000). Transformation in *Streptococcus pneumoniae*: mosaic genes and the regulation of competence. *Research in Microbiology* 151, 453-456.
- Heizer, E.M., Raiford, D.W., Raymer, M.L., Doom, T.E., Miller, R.V., and Krane, D.E. (2006). Amino Acid Cost and Codon-Usage Biases in 6 Prokaryotic Genomes: A Whole-Genome Analysis. *Molecular Biology and Evolution* 23, 1670-1680.
- Hense, W., Anderson, N., Hutter, S., Stephan, W., Parsch, J., and Carlini, D.B. (2010). Experimentally increased codon bias in the *Drosophila Adh* gene leads to an increase in larval, but not adult, alcohol dehydrogenase activity. *Genetics* 184, 547-555.
- Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. *Annu Rev Genet* 42, 287-299.
- Hershberg, R., and Petrov, D.A. (2009). General rules for optimal codon choice. *PLoS genetics* 5.
- Hershberg, R., and Petrov, D.A. (2012). On the limitations of using ribosomal genes as references for the study of codon usage: a rebuttal. *PloS one* 7.
- Hilterbrand, A., Saelens, J., and Putonti, C. (2012). CBDB: The codon bias database. *BMC Bioinformatics* 13, 62.
- Hirst, R.A., Kadioglu, A., O'Callaghan, C., and Andrew, P.W. (2004). The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin Exp Immunol* 138, 195-201.
- Hockenberry, A.J., Sirer, M.I., Amaral, L.A., and Jewett, M.C. (2014). Quantifying position-dependent codon usage bias. *Molecular biology and evolution* 31, 1880-1893.
- Hu, L., Joshi, S.B., Liyanage, M.R., Pansalawatta, M., Alderson, M.R., Tate, A., Robertson, G., Maisonneuve, J., Volkin, D.B., and Middaugh, C.R. (2013). Physical characterization and formulation development of a recombinant pneumolysin protein-based pneumococcal vaccine. *J Pharm Sci* 102, 387-400.
- Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2, 13-34.
- Iliev, A.I., Djannatian, J., Nau, R., Mitchell, T.J., and Wouters, F.S. (2007). Cholesterol-dependent actin remodeling via RhoA and Rac1 activation by the *Streptococcus pneumoniae* toxin pneumolysin. *Proceedings of the National Academy of Sciences* 104, 2897-2902.
- Iliev, A.I., Djannatian, J., Opazo, F., Gerber, J., Nau, R., Mitchell, T.J., and Wouters, F.S. (2009). Rapid microtubule bundling and stabilization by the *Streptococcus pneumoniae* neurotoxin pneumolysin in a cholesterol - dependent, non - lytic and Src - kinase dependent manner inhibits intracellular trafficking. *Molecular Microbiology* 71, 461-477.

- Jacques, N., and Dreyfus, M. (1990). Translation initiation in *Escherichia coli*: old and new questions. *Molecular Microbiology* 4, 1063-1067.
- Jedrzejewski, M.J. (2001). Pneumococcal Virulence Factors: Structure and Function. *Microbiology and Molecular Biology Reviews* 65, 187-207.
- Jefferies, J.M., Clarke, S.C., Webb, J.S., and Kraaijeveld, A.R. (2011). Risk of red queen dynamics in pneumococcal vaccine strategy. *Trends Microbiol* 19, 377-381.
- Jia, M., and Li, Y. (2005). The relationship among gene expression, folding free energy and codon usage bias in *Escherichia coli*. *FEBS letters* 579, 5333-5337.
- Johansson, N., Kalin, M., Giske, C.G., and Hedlund, J. (2008). Quantitative detection of *Streptococcus pneumoniae* from sputum samples with real-time quantitative polymerase chain reaction for etiologic diagnosis of community-acquired pneumonia. *Diagn Microbiol Infect Dis* 60, 255-261.
- Johnson, M.K. (1977). Cellular location of pneumolysin. *FEMS Microbiology Letters* 2, 243-245.
- Jounblat, R., Kadioglu, A., Mitchell, T.J., and Andrew, P.W. (2003). Pneumococcal Behavior and Host Responses during Bronchopneumonia Are Affected Differently by the Cytolytic and Complement-Activating Activities of Pneumolysin. *Infection and Immunity* 71, 7239-7239.
- Kadioglu, A., and Andrew, P.W. (2004). The innate immune response to pneumococcal lung infection: the untold story. *Trends in Immunology* 25, 143-149.
- Kadioglu, A., Gingles, N.A., Grattan, K., Kerr, A., Mitchell, T.J., and Andrew, P.W. (2000). Host Cellular Immune Response to Pneumococcal Lung Infection in Mice. *Infection and Immunity* 68, 4925-4931.
- Kadioglu, A., Taylor, S., Iannelli, F., Pozzi, G., Mitchell, T.J., and Andrew, P.W. (2002). Upper and Lower Respiratory Tract Infection by *Streptococcus pneumoniae* Is Affected by Pneumolysin Deficiency and Differences in Capsule Type. *Infection and Immunity* 70, 2886-2890.
- Kadioglu, A., Weiser, J.N., Paton, J.C., and Andrew, P.W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nature Reviews Microbiology* 6, 288-301.
- Kafka, D., Ling, E., Feldman, G., Benharroch, D., Voronov, E., Givon-Lavi, N., Iwakura, Y., Dagan, R., Apte, R.N., and Mizrahi-Nebenzahl, Y. (2008). Contribution of IL-1 to resistance to *Streptococcus pneumoniae* infection. *International Immunology* 20, 1139-1146.
- Kalin, M. (1998). Pneumococcal serotypes and their clinical relevance. *Thorax* 53, 159-162.
- Kamtchoua, T., Bologa, M., Hopfer, R., Neveu, D., Hu, B., Sheng, X., Corde, N., Pouzet, C., Zimmermann, G., and Gurusathan, S. (2013). Safety and immunogenicity of the pneumococcal pneumolysin derivative PlyD1 in a single-antigen protein vaccine candidate in adults. *Vaccine* 31, 327-333.
- Kebaier, C., Chamberland, R.R., Allen, I.C., Gao, X., Broglie, P.M., Hall, J.D., Jania, C., Doerschuk, C.M., Tilley, S.L., and Duncan, J.A. (2012). *Staphylococcus aureus*

alpha-hemolysin mediates virulence in a murine model of severe pneumonia through activation of the NLRP3 inflammasome. *J Infect Dis* 205, 807-817.

- Khan, M.N., Coleman, J.R., Vernatter, J., Varshney, A.K., Dufaud, C., and Pirofski, L.-A.A. (2014). An ahemolytic pneumolysin of *Streptococcus pneumoniae* manipulates human innate and CD4⁺ T-cell responses and reduces resistance to colonization in mice in a serotype-independent manner. *The Journal of infectious diseases* 210, 1658-1669.
- Kirby, A.C., Newton, D.J., Carding, S.R., and Kaye, P.M. (2007). Pulmonary dendritic cells and alveolar macrophages are regulated by $\gamma \delta$ T cells during the resolution of *S. pneumoniae* - induced inflammation. *The Journal of Pathology* 212, 29-37.
- Klumpp, S., Dong, J., and Hwa, T. (2012). On ribosome load, codon bias and protein abundance. *PloS one* 7.
- Knapp, S., Wieland, C.W., van Veer, C., Takeuchi, O., Akira, S., Florquin, S., and van der Poll, T. (2004). Toll-like receptor 2 plays a role in the early inflammatory response to murine pneumococcal pneumonia but does not contribute to antibacterial defense. *Journal of immunology (Baltimore, Md : 1950)* 172, 3132-3138.
- Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences* 83, 2850-2854.
- Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187208.
- Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science (New York, NY)* 324, 255-258.
- Kulkarni, R., Antala, S., Wang, A., Amaral, F.E., Rampersaud, R., Larussa, S.J., Planet, P.J., and Ratner, A.J. (2012). Cigarette smoke increases *Staphylococcus aureus* biofilm formation via oxidative stress. *Infection and immunity* 80, 3804-3811.
- Kulkarni, R., Rampersaud, R., Aguilar, J.L., Randis, T.M., Kreindler, J.L., and Ratner, A.J. (2010). Cigarette smoke inhibits airway epithelial cell innate immune responses to bacteria. *Infection and immunity* 78, 2146-2152.
- Kunec, D., and Osterrieder, N. (2016). Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias. *Cell reports* 14, 55-67.
- Lacks, S., and Hotchkiss, R.D. (1960). A study of the genetic material determining an enzyme in *Pneumococcus*. *Biochim Biophys Acta* 39, 508-518.
- Lanie, J.A., Ng, W.-L., Kazmierczak, K.M., Andrzejewski, T.M., Davidsen, T.M., Wayne, K.J., Tettelin, H., Glass, J.I., and Winkler, M.E. (2007). Genome Sequence of Avery's Virulent Serotype 2 Strain D39 of *Streptococcus pneumoniae* and Comparison with That of Unencapsulated Laboratory Strain R6. *Journal of Bacteriology* 189, 38-51.
- Lederberg, J. (1994). Anecdotal, historical and critical commentaries on genetics. *The Transformation of Genetics by DNA: An Anniversary Celebration of Avery, MacLeod and McCarty (1944)*. *Genetics* 136, 423-426.

- Lee, S.F., Li, Y.J., and Halperin, S.A. (2009). Overcoming codon-usage bias in heterologous protein expression in *Streptococcus gordonii*. *Microbiology* 155, 3581-3588.
- Leroux-Roels, G., Maes, C., De Boever, F., Traskine, M., Rüggeberg, J.U., and Borys, D. (2014). Safety, reactogenicity and immunogenicity of a novel pneumococcal protein-based vaccine in adults: a phase I/II randomized clinical study. *Vaccine* 32, 6838-6846.
- Li, Q., Li, Y.X., Douthitt, K., Stahl, G.L., Thurman, J.M., and Tong, H.H. (2012). Role of the alternative and classical complement activation pathway in complement mediated killing against *Streptococcus pneumoniae* colony opacity variants during acute pneumococcal otitis media in mice. *Microbes Infect* 14, 1308-1318.
- Lim, J., Ha, U., Sakai, A., Woo, C.-H., Kweon, S.-M., Xu, H., and Li, J.-D. (2008). *Streptococcus pneumoniae* synergizes with nontypeable *Haemophilus influenzae* to induce inflammation via upregulating TLR2. *BMC Immunology* 9, 40.
- Lock, R.A., Paton, J.C., and Hansman, D. (1988). Comparative efficacy of pneumococcal neuraminidase and pneumolysin as immunogens protective against *Streptococcus pneumoniae*. *Microbial Pathogenesis* 5, 461-467.
- Los, F.C.O., Randis, T.M., Aroian, R.V., and Ratner, A.J. (2013). Role of Pore-Forming Toxins in Bacterial Infectious Diseases. *Microbiology and Molecular Biology Reviews* 77, 173-207.
- Lu, J., Sun, T., Hou, H., Xu, M., Gu, T., Dong, Y., Wang, D., Chen, P., Wu, C., Liang, C., et al. (2014). Detoxified pneumolysin derivative Plym2 directly protects against pneumococcal infection via induction of inflammatory cytokines. *Immunological investigations* 43, 717-726.
- Lucas, R., Czikora, I., Sridhar, S., Zemskov, E., Gorshkov, B., Siddaramappa, U., Oseghale, A., Lawson, J., Verin, A., Rick, F.G., et al. (2013). Mini-review: novel therapeutic strategies to blunt actions of pneumolysin in the lungs. *Toxins* 5, 1244-1260.
- Lüttge, M., Fulde, M., Talay, S.R., Nerlich, A., Rohde, M., Preissner, K.T., Hammerschmidt, S., Steinert, M., Mitchell, T.J., Chhatwal, G.S., et al. (2012). *Streptococcus pneumoniae* induces exocytosis of Weibel - Palade bodies in pulmonary endothelial cells. *Cellular Microbiology* 14, 210-225.
- Lysenko, E.S., Clarke, T.B., Shchepetov, M., Ratner, A.J., Roper, D.I., Dowson, C.G., and Weiser, J.N. (2007). Nod1 Signaling Overcomes Resistance of *S. pneumoniae* to Opsonophagocytic Killing. *PLoS Pathogens* 3.
- Lysenko, E.S., Ratner, A.J., Nelson, A.L., and Weiser, J.N. (2005). The Role of Innate Immune Responses in the Outcome of Interspecies Competition for Colonization of Mucosal Surfaces. *PLoS Pathogens* 1.
- Malley, R., Henneke, P., Morse, S.C., Cieslewicz, M.J., Lipsitch, M., Thompson, C.M., Kurt-Jones, E., Paton, J.C., Wessels, M.R., and Golenbock, D.T. (2003). Recognition of pneumolysin by Toll-like receptor 4 confers resistance to pneumococcal infection. *Proceedings of the National Academy of Sciences* 100, 1966-1971.
- Malley, R., Trzcinski, K., Srivastava, A., Thompson, C.M., Anderson, P.W., and Lipsitch, M. (2005). CD4+ T cells mediate antibody-independent acquired immunity to

- pneumococcal colonization. *Proceedings of the National Academy of Sciences of the United States of America* 102, 4848-4853.
- Markham, N.R., and Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research* 33.
- Markham, N.R., and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods in molecular biology (Clifton, NJ)* 453, 3-31.
- Marriott, H., Mitchell, T., and Dockrell, D. (2008). Pneumolysin: A Double-Edged Sword During the Host-Pathogen Interaction. *Current Molecular Medicine* 8, 497509.
- Marriott, H.M., Gascoyne, K.A., Gowda, R., Geary, I., Nicklin, M.J., Iannelli, F., Pozzi, G., Mitchell, T.J., Whyte, M.K., Sabroe, I., et al. (2012). Interleukin-1beta regulates CXCL8 release and influences disease outcome in response to *Streptococcus pneumoniae*, defining intercellular cooperation between pulmonary epithelial cells and macrophages. *Infect Immun* 80, 1140-1149.
- Martín-Galiano, A.J., Wells, J.M., and de la Campa, A.G. (2004). Relationship between codon biased genes, microarray expression values and physiological characteristics of *Streptococcus pneumoniae*. *Microbiology* 150, 2313-2325.
- Martner, A., Dahlgren, C., Paton, J.C., and Wold, A.E. (2008). Pneumolysin released during *Streptococcus pneumoniae* autolysis is a potent activator of intracellular oxygen radical production in neutrophils. *Infection and immunity* 76, 4079-4087.
- Martner, A., Skovbjerg, S., Paton, J.C., and Wold, A.E. (2009). *Streptococcus pneumoniae* autolysis prevents phagocytosis and production of phagocyte-activating cytokines. *Infection and immunity* 77, 3826-3837.
- Matthias, K.A., Roche, A.M., Standish, A.J., Shchepetov, M., and Weiser, J.N. (2008). Neutrophil-toxin interactions promote antigen delivery and mucosal clearance of *Streptococcus pneumoniae*. *Journal of immunology (Baltimore, Md : 1950)* 180, 6246-6254.
- Maus, U.A., Srivastava, M., Paton, J.C., Mack, M., Everhart, M.B., Blackwell, T.S., Christman, J.W., Schlondorff, D., Seeger, W., and Lohmeyer, J. (2004). Pneumolysin-induced lung injury is independent of leukocyte trafficking into the alveolar space. *Journal of Immunology* 173, 1307-1312.
- McNeela, E.A., Burke, Á., Neill, D.R., Baxter, C., Fernandes, V.E., Ferreira, D., Smeaton, S., El-Rachkidy, R., McLoughlin, R.M., Mori, A., et al. (2010). Pneumolysin Activates the NLRP3 Inflammasome and Promotes Proinflammatory Cytokines Independently of TLR4. *PLoS Pathogens* 6.
- Mitchell, T.J. (2000). Virulence factors and the pathogenesis of disease caused by *Streptococcus pneumoniae*. *Research in Microbiology* 151, 413419.
- Mitchell, T.J., and Dalziel, C.E. (2014). The biology of pneumolysin. *Sub-cellular biochemistry* 80, 145-160.
- Moffitt, K.L., and Malley, R. (2011). Next generation pneumococcal vaccines. *Current opinion in immunology* 23, 407-413.

- Moreland, J.G., and Bailey, G. (2006). Neutrophil transendothelial migration in vitro to *Streptococcus pneumoniae* is pneumolysin dependent. *American Journal of Physiology - Lung Cellular and Molecular Physiology* 290.
- Mueller, S., Coleman, J.R., Papamichail, D., Ward, C.B., Nimnual, A., Futcher, B., Skiena, S., and Wimmer, E. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol* 28, 723-726.
- Mueller, S., Papamichail, D., Coleman, R.J., Skiena, S., and Wimmer, E. (2006). Reduction of the Rate of Poliovirus Protein Synthesis through Large-Scale Codon Deoptimization Causes Attenuation of Viral Virulence by Lowering Specific Infectivity. *Journal of Virology* 80, 9687-9696.
- Mufson, M.A., Krause, H.E., and Schiffman, G. (1983). Long-term persistence of antibody following immunization with pneumococcal polysaccharide vaccine. *Proc Soc Exp Biol Med* 173, 270-275.
- Mufson, M.A., Krause, H.E., Schiffman, G., and Hughey, D.F. (1987). Pneumococcal antibody levels one decade after immunization of healthy adults. *Am J Med Sci* 293, 279-284.
- Mureithi, M.W., Finn, A., Ota, M.O., Zhang, Q., Davenport, V., Mitchell, T.J., Williams, N.A., Adegbola, R.A., and Heyderman, R.S. (2009). T cell memory response to pneumococcal protein antigens in an area of high pneumococcal carriage and disease. *The Journal of infectious diseases* 200, 783-793.
- Mutepe, N.D., Cockeran, R., Steel, H.C., Theron, A.J., Mitchell, T.J., Feldman, C., and Anderson, R. (2013). Effects of cigarette smoke condensate on pneumococcal biofilm formation and pneumolysin. *Eur Respir J* 41, 392-395.
- N'Guessan, P., Schmeck, B., Ayim, A., Hocke, A.C., Brell, B., Hammerschmidt, S., Rosseau, S., Suttorp, N., and Hippenstiel, S. (2005). *Streptococcus pneumoniae* R6x induced p38 MAPK and JNK-mediated caspase-dependent apoptosis in human endothelial cells. *Thrombosis and haemostasis* 94, 295-303.
- Ogunniyi, A.D., LeMessurier, K.S., Graham, R.M., Watt, J.M., Briles, D.E., Stroehel, U.H., and Paton, J.C. (2007). Contributions of pneumolysin, pneumococcal surface protein A (PspA), and PspC to pathogenicity of *Streptococcus pneumoniae* D39 in a mouse model. *Infect Immun* 75, 1843-1851.
- Oloo, E.O., Yethon, J.A., Ochs, M.M., Carpick, B., and Oomen, R. (2011). Structure-guided Antigen Engineering Yields Pneumolysin Mutants Suitable for Vaccination against Pneumococcal Disease. *Journal of Biological Chemistry* 286, 12133-12140.
- Orihuela, C.J., Gao, G., Francis, K.P., Yu, J., and Tuomanen, E.I. (2004). Tissue - Specific Contributions of Pneumococcal Virulence Factors to Pathogenesis. *The Journal of Infectious Diseases* 190, 1661-1669.
- Ozlu, T., Celik, I., Oztuna, F., Bulbul, Y., and Ozsu, S. (2008). *Streptococcus pneumoniae* adherence in rats under different degrees and durations of cigarette smoke. *Respiration* 75, 339-344.

- Parker, D., Martin, F.J., Soong, G., Harfenist, B.S., Aguilar, J.L., Ratner, A.J., Fitzgerald, K.A., Schindler, C., and Prince, A. (2011). Streptococcus pneumoniae DNA Initiates Type I Interferon Signaling in the Respiratory Tract. *mBio* 2, 11.
- Parker, D., and Prince, A. (2012). Staphylococcus aureus Induces Type I IFN Signaling in Dendritic Cells Via TLR9. *The Journal of Immunology* 189, 4040-4046.
- Paterson, G.K., Blue, C.E., and Mitchell, T.J. (2006). Role of two-component systems in the virulence of Streptococcus pneumoniae. *Journal of medical microbiology* 55, 355-363.
- Paton, J.C. (1996). The contribution of pneumolysin to the pathogenicity of Streptococcus pneumoniae. *Trends Microbiol* 4, 103-106.
- Paton, J.C., Andrew, P.W., Boulnois, G.J., and Mitchell, T.J. (1993). Molecular analysis of the pathogenicity of Streptococcus pneumoniae: the role of pneumococcal proteins. *Annu Rev Microbiol* 47, 89-115.
- Pauksens, K., Nilsson, A.C., Caubet, M., Pascal, T.G., Van Belle, P., Poolman, J.T., Vandepapeliere, P.G., Verlant, V., and Vink, P.E. (2014). Randomized controlled study of the safety and immunogenicity of pneumococcal vaccine formulations containing PhtD and detoxified pneumolysin with alum or adjuvant system AS02V in elderly adults. *Clin Vaccine Immunol* 21, 651-660.
- Phipps, J.C., Aronoff, D.M., Curtis, J.L., Goel, D., O'Brien, E., and Mancuso, P. (2010). Cigarette smoke exposure impairs pulmonary bacterial clearance and alveolar macrophage complement-mediated phagocytosis of Streptococcus pneumoniae. *Infection and immunity* 78, 1214-1220.
- Piotrowski, A., Luo, P., and Morrison, D.A. (2009). Competence for genetic transformation in Streptococcus pneumoniae: termination of activity of the alternative sigma factor ComX is independent of proteolysis of ComX and ComW. *Journal of bacteriology* 191, 3359-3366.
- Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 12, 32-42.
- Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Molecular systems biology* 10, 770.
- Price, K.E., and Camilli, A. (2009). Pneumolysin localizes to the cell wall of Streptococcus pneumoniae. *J Bacteriol* 191, 2163-2168.
- Price, K.E., Greene, N.G., and Camilli, A. (2012). Export requirements of pneumolysin in Streptococcus pneumoniae. *J Bacteriol* 194, 3651-3660.
- Provoost, S., Maes, T., Pauwels, N.S., Vanden Berghe, T., Vandenabeele, P., Lambrecht, B.N., Joos, G.F., and Tournoy, K.G. (2011). NLRP3/caspase-1-independent IL-1beta production mediates diesel exhaust particle-induced pulmonary inflammation. *J Immunol* 187, 3331-3337.
- Prymula, R., Pazdiora, P., Traskine, M., Ruggeberg, J.U., and Borys, D. (2014). Safety and immunogenicity of an investigational vaccine containing two common

pneumococcal proteins in toddlers: a phase II randomized clinical trial. *Vaccine* 32, 3025-3034.

- Puigbo, P., Bravo, I.G., and Garcia-Vallve, S. (2008a). CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* 3, 38.
- Puigbo, P., Guzman, E., Romeu, A., and Garcia-Vallve, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res* 35, W126-131.
- Puigbo, P., Romeu, A., and Garcia-Vallve, S. (2008b). HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. *Nucleic Acids Res* 36, D524-527.
- Rampersaud, R., Planet, P.J., Randis, T.M., Kulkarni, R., Aguilar, J.L., Lehrer, R.I., and Ratner, A.J. (2011). Inerolysin, a cholesterol-dependent cytolysin produced by *Lactobacillus iners*. *Journal of bacteriology* 193, 1034-1041.
- Ran, W., and Higgs, P.G. (2012). Contributions of speed and accuracy to translational selection in bacteria. *PLoS one* 7.
- Randis, T.M., Kulkarni, R., Aguilar, J.L., and Ratner, A.J. (2009). Antibody-Based Detection and Inhibition of Vaginolysin, the *Gardnerella vaginalis* Cytolysin. *PLoS ONE* 4.
- Rasigade, J.P., Trouillet-Assant, S., Ferry, T., Diep, B.A., Sapin, A., Lhoste, Y., Ranfaing, J., Badiou, C., Benito, Y., Bes, M., et al. (2013). PSMs of hypervirulent *Staphylococcus aureus* act as intracellular toxins that kill infected osteoblasts. *PLoS One* 8, e63176.
- Ratner, A.J., Aguilar, J.L., Shchepetov, M., Lysenko, E.S., and Weiser, J.N. (2007). Nod1 mediates cytoplasmic sensing of combinations of extracellular bacteria. *Cellular Microbiology* 9, 1343-1351.
- Ratner, A.J., Hippe, K.R., Aguilar, J.L., Bender, M.H., Nelson, A.L., and Weiser, J.N. (2006). Epithelial Cells Are Sensitive Detectors of Bacterial Pore-forming Toxins. *Journal of Biological Chemistry* 281, 12994-12998.
- Ratner, A.J., Lysenko, E.S., Paul, M.N., and Weiser, J.N. (2005). Synergistic proinflammatory responses induced by polymicrobial colonization of epithelial surfaces. *Proceedings of the National Academy of Sciences of the United States of America* 102, 3429-3434.
- Rayner, C., Jackson, A., Rutman, A., Dewar, A., Mitchell, T., Andrew, P., Cole, P., and Wilson, R. (1995). Interaction of pneumolysin-sufficient and -deficient isogenic variants of *Streptococcus pneumoniae* with human respiratory mucosa. *Infection and immunity* 63, 442-447.
- Reboul, C.F., Whisstock, J.C., and Dunstone, M.A. (2014). A new model for pore formation by cholesterol-dependent cytolysins. *PLoS computational biology* 10.
- Reiss, A., Braun, J.S., Jäger, K., Freyer, D., Laube, G., Bühner, C., Felderhoff-Müser, U., Stadelmann, C., Nizet, V., and Weber, J.R. (2011). Bacterial pore-forming cytolysins induce neuronal damage in a rat model of neonatal meningitis. *The Journal of infectious diseases* 203, 393-400.
- Richards, J., Sundermeier, T., Svetlanov, A., and Karzai, W.A. (2008). Quality control of bacterial mRNA decoding and decay. *Biochimica et biophysica acta* 1779, 574-582.

- Rijneveld, A.W., van den Dobbelsteen, G.P., Florquin, S., Standiford, T.J., Speelman, P., van Alphen, L., and van der Poll, T. (2002). Roles of interleukin-6 and macrophage inflammatory protein-2 in pneumolysin-induced lung inflammation in mice. *J Infect Dis* 185, 123-126.
- Rogers, P.D., Thornton, J., Barker, K.S., McDaniel, D.O., Sacks, G.S., Swiatlo, E., and McDaniel, L.S. (2003). Pneumolysin-Dependent and -Independent Gene Expression Identified by cDNA Microarray Analysis of THP-1 Human Mononuclear Cells Stimulated by *Streptococcus pneumoniae*. *Infection and Immunity* 71, 20872094.
- Roller, M., Lucic, V., Nagy, I., Perica, T., and Vlahovicek, K. (2013). Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic Acids Res* 41, 8842-8852.
- Rossjohn, J., Gilbert, R., Crane, D., Morgan, P., Mitchell, T., Rowe, A., Andrew, P., Paton, J., Tweten, R., and Parker, M. (1998). The molecular mechanism of pneumolysin, a virulence factor from *Streptococcus pneumoniae*. *Journal of molecular biology* 284, 449-461.
- Sabi, R., and Tuller, T. (2014). Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res* 21, 511-526.
- Sagliocco, F.A., Laso, V.M.R., Zhu, D., Tuite, M.F., McCarthy, J.E., and Brown, A.J. (1993). The influence of 5'-secondary structures upon ribosome binding to mRNA during translation in yeast. *The Journal of biological chemistry* 268, 26522-26530.
- Salha, D., Szeto, J., Myers, L., Claus, C., Sheung, A., Tang, M., Ljutic, B., Hanwell, D., Ogilvie, K., Ming, M., et al. (2012). Neutralizing antibodies elicited by a novel detoxified pneumolysin derivative, PlyD1, provide protection against both pneumococcal infection and lung injury. *Infect Immun* 80, 2212-2220.
- Sanders, M.E., Taylor, S., Tullos, N., Norcross, E.W., Moore, Q.C., 3rd, Thompson, H., King, L.B., and Marquart, M.E. (2013). Passive immunization with Pneumovax(R) 23 and pneumolysin in combination with vancomycin for pneumococcal endophthalmitis. *BMC Ophthalmol* 13, 8.
- Seligmann, H., and Pollock, D.D. (2004). The ambush hypothesis: hidden stop codons prevent off-frame gene reading. *DNA and cell biology* 23, 701-705.
- Seo, S., Yang, J., and Jung, G. (2009). Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in *Escherichia coli*. *Biotechnology and Bioengineering* 104, 611-616.
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-Limiting Steps in Yeast Protein Translation. *Cell* 153.
- Shah, P., and Gilchrist, M.A. (2010). Effect of Correlated tRNA Abundances on Translation Errors and Evolution of Codon Usage Bias. *PLoS Genetics* 6.
- Sharp, P.M., and Li, W.H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol* 24, 28-38.

- Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15, 1281-1295.
- Shockett, P.E., and Schatz, D.G. (1996). Diverse strategies for tetracycline-regulated inducible gene expression. *Proc Natl Acad Sci U S A* 93, 5173-5176.
- Shoma, S., Tsuchiya, K., Kawamura, I., Nomura, T., Hara, H., Uchiyama, R., Daim, S., and Mitsuyama, M. (2008). Critical Involvement of Pneumolysin in Production of Interleukin-1 α and Caspase-1-Dependent Cytokines in Infection with *Streptococcus pneumoniae* In Vitro: a Novel Function of Pneumolysin in Caspase-1 Activation. *Infection and Immunity* 76, 1547-1557.
- Son, M.S., Megli, C.J., Kovacicova, G., Qadri, F., and Taylor, R.K. (2011). Characterization of *Vibrio cholerae* O1 El Tor biotype variant clinical isolates from Bangladesh and Haiti, including a molecular genetic analysis of virulence genes. *J Clin Microbiol* 49, 3739-3749.
- Sopori, M. (2002). Effects of cigarette smoke on the immune system. *Nature Reviews Immunology* 2, 372-377.
- Srivastava, A., Henneke, P., Visintin, A., Morse, S., Martin, V., Watkins, C., Paton, J., Wessels, M., Golenbock, D., and Malley, R. (2005). The apoptotic response to pneumolysin is Toll-like receptor 4 dependent and protects against pneumococcal disease. *Infection and immunity* 73, 6479-6487.
- Stabler, R.A., Dawson, L.F., Phua, L.T., and Wren, B.W. (2008). Comparative analysis of BI/NAP1/027 hypervirulent strains reveals novel toxin B-encoding gene (tcdB) sequences. *J Med Microbiol* 57, 771-775.
- Stoletzki, N., and Eyre-Walker, A. (2007). Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol* 24, 374-381.
- Sung, C.K., Li, H., Claverys, J.P., and Morrison, D.A. (2001). An rpsL Cassette, Janus, for Gene Replacement through Negative Selection in *Streptococcus pneumoniae*. *Applied and Environmental Microbiology* 67, 5190-5196.
- Supek, F., and Smuc, T. (2010). On relevance of codon usage to expression of synthetic and natural genes in *Escherichia coli*. *Genetics* 185, 1129-1134.
- Tai, S.S. (2006). *Streptococcus pneumoniae* protein vaccine candidates: properties, activities and animal studies. *Crit Rev Microbiol* 32, 139-153.
- Taut, K., Winter, C., Briles, D.E., Paton, J.C., Christman, J.W., Maus, R., Baumann, R., Welte, T., and Maus, U.A. (2008). Macrophage Turnover Kinetics in the Lungs of Mice Infected with *Streptococcus pneumoniae*. *American Journal of Respiratory Cell and Molecular Biology* 38, 105113.
- Taylor, S.D., Sanders, M.E., Tullos, N.A., Stray, S.J., Norcross, E.W., McDaniel, L.S., and Marquart, M.E. (2013). The cholesterol-dependent cytolysin pneumolysin from *Streptococcus pneumoniae* binds to lipid raft microdomains in human corneal epithelial cells. *PLoS one* 8.

- Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. (2001). Complete Genome Sequence of a Virulent Isolate of *Streptococcus pneumoniae*. *Science* 293, 498506.
- Tilley, S.J., Orlova, E.V., Gilbert, R., Andrew, P.W., and Saibil, H.R. (2005). Structural Basis of Pore Formation by the Bacterial Toxin Pneumolysin. *Cell* 121.
- Tomasz, A., Jamieson, J.D., and Ottolenghi, E. (1964). The Fine Structure of *Diplococcus pneumoniae*. *J Cell Biol* 22, 453-467.
- Trotta, E. (2013). Selection on codon bias in yeast: a transcriptional hypothesis. *Nucleic Acids Res* 41, 9382-9395.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010a). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* 141.
- Tuller, T., Girshovich, Y., Sella, Y., Kreimer, A., Freilich, S., Kupiec, M., Gophna, U., and Ruppin, E. (2011). Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res* 39, 4743-4755.
- Tuller, T., Kupiec, M., and Ruppin, E. (2007). Determinants of Protein Abundance and Translation Efficiency in *S. cerevisiae*. *PLoS Computational Biology* 3.
- Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010b). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences* 107, 3645-3650.
- Tulloch, F., Atkinson, N.J., Evans, D.J., Ryan, M.D., and Simmonds, P. (2014). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *eLife* 3, e04531.
- Tweten, R.K. (2005). Cholesterol-Dependent Cytolysins, a Family of Versatile Pore-Forming Toxins. *Infection and Immunity* 73, 61996209.
- van Rossum, A.M.C., Lysenko, E.S., and Weiser, J.N. (2005). Host and Bacterial Factors Contributing to the Clearance of Colonization by *Streptococcus pneumoniae* in a Murine Model. *Infection and Immunity* 73, 77187726.
- Verhoeven, D., Perry, S., and Pichichero, M.E. (2014). Contributions to protection from *Streptococcus pneumoniae* infection using the monovalent recombinant protein vaccine candidates PcpA, PhtD, and PlyD1 in an infant murine model during challenge. *Clin Vaccine Immunol* 21, 1037-1045.
- Vernatter, J., and Pirofski, L.A. (2013). Current concepts in host-microbe interaction leading to pneumococcal pneumonia. *Curr Opin Infect Dis* 26, 277-283.
- Walker, J., Allen, R., Falmagne, P., Johnson, M., and Boulnois, G. (1987). Molecular cloning, characterization, and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of *Streptococcus pneumoniae*. *Infection and immunity* 55, 1184-1189.
- Walsh, R.L., and Camilli, A. (2011). *Streptococcus pneumoniae* Is Desiccation Tolerant and Infectious upon Rehydration. *mBio* 2.

- Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, D.L., Nutter, R.C., Segal, E., and Chang, H.Y. (2012). Genome-wide measurement of RNA folding energies. *Molecular cell* 48, 169-181.
- Weiser, J.N., Austrian, R., Sreenivasan, P.K., and Masure, H.R. (1994). Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect Immun* 62, 2582-2589.
- Wellmer, A., Zysk, G., Gerber, J., Kunst, T., Von Mering, M., Bunkowski, S., Eiffert, H., and Nau, R. (2002). Decreased virulence of a pneumolysin-deficient strain of *Streptococcus pneumoniae* in murine meningitis. *Infect Immun* 70, 6504-6508.
- Whidbey, C., Harrell, M., Burnside, K., Ngo, L., Becraft, A.K., Iyer, L.M., Aravind, L., Hitti, J., Waldorf, K.M., and Rajagopal, L. (2013). A hemolytic pigment of Group B *Streptococcus* allows bacterial penetration of human placenta. *The Journal of Experimental Medicine* 210, 1265-1281.
- Wippel, C., Förtsch, C., Hupp, S., Maier, E., Benz, R., Ma, J., Mitchell, T.J., and Iliev, A.I. (2011). Extracellular calcium reduction strongly increases the lytic capacity of pneumolysin from *streptococcus pneumoniae* in brain tissue. *The Journal of infectious diseases* 204, 930-936.
- Witzenrath, M., Gutbier, B., Hocke, A.C., Schmeck, B., Hippenstiel, S., Berger, K., Mitchell, T.J., de los Toyos, J.R., Rosseau, S., Suttorp, N., et al. (2006). Role of pneumolysin for the development of acute lung injury in pneumococcal pneumonia. *Critical Care Medicine* 34, 1947.
- Witzenrath, M., Pache, F., Lorenz, D., Koppe, U., Gutbier, B., Tabeling, C., Reppe, K., Meixenberger, K., Dorhoi, A., Ma, J., et al. (2011). The NLRP3 Inflammasome Is Differentially Activated by Pneumolysin Variants and Contributes to Host Defense in Pneumococcal Pneumonia. *The Journal of Immunology* 187, 434-440.
- Wolf, A.I., Strauman, M.C., Mozdzanowska, K., Williams, K.L., Osborne, L.C., Shen, H., Liu, Q., Garlick, D., Artis, D., Hensley, S.E., et al. (2014). Pneumolysin expression by *streptococcus pneumoniae* protects colonized mice from influenza virus-induced disease. *Virology* 462-463, 254-265.
- Woodson, S.A., and Koculi, E. (2009). Analysis of RNA folding by native polyacrylamide gel electrophoresis. *Methods in enzymology* 469, 189-208.
- Wu, J., Yan, Z., Schwartz, D.E., Yu, J., Malik, A.B., and Hu, G. (2013). Activation of NLRP3 inflammasome in alveolar macrophages contributes to mechanical stretch-induced lung inflammation and injury. *J Immunol* 190, 3590-3599.
- Xia, X. (2015). A major controversy in codon-anticodon adaptation resolved by a new codon usage index. *Genetics* 199, 573-579.
- Xu, Q., Casey, J.R., and Pichichero, M.E. (2015). Higher levels of mucosal antibody to pneumococcal vaccine candidate proteins are associated with reduced acute otitis media caused by *Streptococcus pneumoniae* in young children. *Mucosal immunology* 8, 1110-1117.
- Yother, J. (2011). Capsules of *Streptococcus pneumoniae* and Other Bacteria: Paradigms for Polysaccharide Biosynthesis and Regulation. *Microbiology* 65, 563-581.

- Zhang, Z., Li, J., Cui, P., Ding, F., Li, A., Townsend, J.P., and Yu, J. (2012). Codon Deviation Coefficient: a novel measure for estimating codon usage bias and its statistical significance. *BMC bioinformatics* 13, 43.
- Zhao, H., Chen, D., Tang, J., Jia, G., Long, D., Liu, G., Chen, X., and Shang, H. (2014). Partial optimization of the 5-terminal codon increased a recombinant porcine pancreatic lipase (opPPL) expression in *Pichia pastoris*. *PloS one* 9.
- Zhu, L., and Lau, G.W. (2011). Inhibition of Competence Development, Horizontal Gene Transfer and Virulence in *Streptococcus pneumoniae* by a Modified Competence Stimulating Peptide. *PLoS Pathogens* 7.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.
- Zur, H., and Tuller, T. (2012). Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*. *EMBO reports* 13, 272-277.

CHAPTER VI. APPENDICES

A.	<i>Curriculum Vitae</i>	111
	Academic background	111
	List of publications (6)	111
B.	Lists of bacterial strains created	113
	<i>Streptococcus pneumoniae</i> strains.....	115
	<i>Escherichia coli</i> strains.....	116
C.	Predicted mRNA folding secondary structures	117
D.	Growth curves of <i>S. pneumoniae</i>	123
E.	The limits of PLY production with increasing ΔG	125
F.	Codon optimization of PLY.....	127
G.	<i>Drosophila melanogaster</i> model for rheostat pneumolysin expression.....	131
H.	<i>Caenorhabditis elegans</i> model for rheostat pneumolysin expression.....	133
I.	Published article: “Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by <i>Streptococcus pneumoniae</i> ”	135

A. Curriculum Vitae

Academic background

- 2004-2014: Integrated Master of Medicine, University of Minho, Portugal.

Master's Thesis in Medicine title:

“Cigarette smoke induces virulence and antibiotic resistance of bacteria.”

- 2009-2013: PhD research period at Columbia University in the City of New York, USA.
- 2004-2007: Bachelor of Basic Sciences of Medicine, University of Minho, Portugal.

List of publications (6)

- **Amaral, F.E.**, Parker, D., Randis, T.M., Kulkarni, R., Prince, A.S., Shirasu-Hiza, M.M., and Ratner, A.J. (2015). Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by *Streptococcus pneumoniae*. *PLoS one* 10. (See Appendix I on page 135.)
- Paragas, N., Kulkarni, R., Werth, M., Schmidt-Ott, K.M., Forster, C., Deng, R., Zhang, Q., Singer, E., Klose, A.D., Shen, T.H., Francis, K.P., Ray, S., Vijayakumar, S., Seward, S., Bovino, M.E., Xu, K., Takabe, Y., **Amaral, F.E.**, Mohan, S., Wax, R., Corbin, K., Sanna-Cherchi, S., Mori, K., Johnson, L., Nickolas, T., D'Agati, V., Lin, C.S., Qiu, A., Al-Awqati, Q., Ratner, A.J., and Barasch, J. (2014). α -Intercalated cells defend the urinary system from bacterial infection. *The Journal of clinical investigation* 124, 2963-2976.
- Randis, T.M., Zaklama, J., LaRocca, T.J., Los, F.C., Lewis, E.L., Desai, P., Rampersaud, R., **Amaral, F.E.**, and Ratner, A.J. (2013). Vaginolysin drives epithelial ultrastructural responses to *Gardnerella vaginalis*. *Infection and immunity* 81, 4544-4550.

- Kulkarni, R., Randis, T.M., Antala, S., Wang, A., **Amaral, F.E.**, and Ratner, A.J. (2013). β -Hemolysin/cytolysin of Group B *Streptococcus* enhances host inflammation but is dispensable for establishment of urinary tract infection. *PloS one* 8.
- Kulkarni, R., Antala, S., Wang, A., **Amaral, F.E.**, Rampersaud, R., Larussa, S.J., Planet, P.J., and Ratner, A.J. (2012). Cigarette smoke increases *Staphylococcus aureus* biofilm formation via oxidative stress. *Infection and immunity* 80, 3804-3811.
- Sampaio, P., Santos, M., Correia, A., **Amaral, F.E.**, Chavéz-Galarza, J., Costa-de-Oliveira, S., Castro, A.G., Pedrosa, J., and Pais, C. (2010). Virulence attenuation of *Candida albicans* genetic variants isolated from a patient with a recurrent bloodstream infection. *PloS one* 5.

B. Lists of bacterial strains created

The following table describes the selected *ply* designed sequences from the total of 294912 sequences (PLY seq) in increasing order (#) of calculated 5' mRNA folding free energy. The corresponding *E. coli* and *S. pneumoniae* strains that use those sequences are indicated by codes and can be found in the following pages. The number of silent mutations differing from the WT strain regarding the *ply* gene is noted (Point mutations total).

#	ΔG (kcal/mol)	<i>E. coli</i>	R6	D39	PLY seq	Point mutations total	Notes on <i>ply</i>
1	-17.5	ARE354	ARP34	ARP50	112940	6	lowest calculated energy
2	-12.8	ARE416	ARP42	-	223932	4	mutated until nt18
3	-9.5	ARE415	ARP43	ARP49	6496	7	equal to -0.2 after nt16
4	-7.3	-	ARP16	ARP48	223754	0	wild-type (WT)
5	-6.2	ARE356	ARP35	-	287208	8	first 38nt are equal to fCO (CO)
6	-6.2	ARE350	ARP44	-	coPLY	398	fully codon optimized (fCO)
7	-3.4	ARE320	ARP41	ARP47	224241	5	mutated until nt18
8	-2.5	ARE349	ARP32	-	227715	6	differs from -0.2 at nt36
9	-2.0	ARE488	ARP45	-	188254	7	codon deoptimized 39nt (CdO)
10	-0.2	ARE357	ARP36	-	6531	7	highest calculated energy

In addition to the previous table, the following table details *ply* mRNA sequence from nucleotide -4 to 39, highlighting the differences from the WT strain (in red) with the degenerate sequence presented atop. Nucleotides 1 to 3 are codon AUG, the start codon for *ply* gene.

Deg seq:	g	a	a	r	A	U	G	G	C	N	A	A	Y	A	A	R	G	C	N	G	U	N	
#	ΔG	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	-17.5	g	a	a	g	A	U	G	G	C	C	A	A	U	A	A	A	G	C	C	G	U	G
2	-12.8	g	a	a	g	A	U	G	G	C	C	A	A	U	A	A	G	G	C	C	G	U	C
3	-9.5	g	a	a	g	A	U	G	G	C	U	A	A	U	A	A	G	G	C	G	G	U	G
4	-7.3	g	a	a	g	A	U	G	G	C	A	A	A	U	A	A	A	G	C	A	A	U	A
5	-6.2	g	a	a	g	A	U	G	G	C	U	A	A	C	A	A	A	G	C	U	G	U	U
6	-6.2	g	a	a	g	A	U	G	G	C	U	A	A	C	A	A	A	G	C	U	G	U	U
7	-3.4	g	a	a	a	A	U	G	G	C	A	A	A	C	A	A	G	G	C	U	G	U	U
8	-2.5	g	a	a	a	A	U	G	G	C	C	A	A	C	A	A	A	G	C	A	A	U	U
9	-2.0	g	a	a	g	A	U	G	G	C	G	A	A	U	A	A	G	G	C	G	G	U	G
10	-0.2	g	a	a	a	A	U	G	G	C	C	A	A	C	A	A	A	G	C	A	A	U	U
Deg seq:	...	A	A	Y	G	A	Y	U	U	Y	A	U	H	Y	U	N	G	C	N	A	U	g	
#	ΔG	...	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1	-17.5	...	A	A	C	G	A	C	U	U	U	A	U	A	C	U	G	G	C	C	A	U	g
2	-12.8	...	A	A	U	G	A	C	U	U	U	A	U	A	C	U	A	G	C	U	A	U	g
3	-9.5	...	A	A	C	G	A	C	U	U	U	A	U	C	C	U	A	G	C	A	A	U	g
4	-7.3	...	A	A	U	G	A	C	U	U	U	A	U	A	C	U	A	G	C	U	A	U	g
5	-6.2	...	A	A	C	G	A	C	U	U	C	A	U	C	C	U	U	G	C	U	A	U	g
6	-6.2	...	A	A	C	G	A	C	U	U	C	A	U	C	C	U	U	G	C	U	A	U	g
7	-3.4	...	A	A	U	G	A	C	U	U	U	A	U	A	C	U	A	G	C	U	A	U	g
8	-2.5	...	A	A	C	G	A	C	U	U	U	A	U	C	C	U	A	G	C	U	A	U	g
9	-2.0	...	A	A	U	G	A	U	U	U	U	A	U	A	C	U	G	G	C	G	A	U	g
10	-0.2	...	A	A	C	G	A	C	U	U	U	A	U	C	C	U	A	G	C	A	A	U	g

IUPAC Nucleic Acid Codes: Y (pYrimidine); R (puRine); N (aNy); H (not G).

NOTE: nucleotide 31 could be a C or U but that codon has to translate a leucine, so, codons UUU and UUC (phenylalanine) were exclude from the generated sequences, totalizing 294912.

Streptococcus pneumoniae strains

The following table describes the pneumococcal strains created (ARP...), background, selection markers, and corresponding *ply* mRNA folding free energy where appropriate. Order numbers (#) are linked to the other tables in this appendix and thesis, and correspond to ΔG . Selection antibiotics ($\mu\text{g}/\text{mL}$): Km (kanamycin); Sm (streptomycin); S, sensitive, or R, resistant.

#	ΔG	Strain #	Description	Selection	Notes
		ARP12	R6 SmR	Sm150	From Weiser P917 R6 SmR
KO		ARP15	R6 <i>ply</i> ::Janus SmS KmR	Km200	ARP12 transformed with Janus cassette <i>ply</i> ::Janus (from Malley)
1	-17.5	ARP34	R6 SmR PLY112940	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE354
2	-12.8	ARP42	R6 SmR PLY223932	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE416
3	-9.5	ARP43	R6 SmR PLY6496	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE415
4	-7.3	ARP16	R6 SmR transformed WT	Sm150	ARP15 transformed with upstream+wtPLY PCR from ARP12
5	-6.2	ARP35	R6 SmR PLY287208 (CO)	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE356
6	-6.2	ARP44	R6 SmR coPLY	Sm150	ARP15 transformed with 1k+coPLY PCR on pARE350
7	-3.4	ARP41	R6 SmR PLY224241	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE320
8	-2.5	ARP32	R6 SmR PLY227715	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE349
9	-2.0	ARP45	R6 SmR PLY188254(CdO)	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE488
10	-0.2	ARP36	R6 SmR PLY6531	Sm150	ARP15 transformed with upstream+wtPLY PCR on pARE357
		ARP21	D39	none	From Weiser P210 D39
		ARP39	D39 SmR	Sm150	ARP21 spontaneous SmR mutant
cKO		ARP46	D39 <i>ply</i> ::Janus SmS KmR	Km200	ARP39 transformed with genome from ARP15
c1	-17.5	ARP50	D39 SmR PLY112940	Sm150	ARP46 transformed with genome ARP34
c3	-9.5	ARP49	D39 SmR PLY6496	Sm150	ARP46 transformed with genome ARP43
c4	-7.3	ARP48	D39 SmR transformed WT	Sm150	ARP46 transformed with genome ARP39
c7	-3.4	ARP47	D39 SmR PLY224241	Sm150	ARP46 transformed with genome ARP41

Escherichia coli strains

The following table describes the supporting *E. coli* strains created (ARE...), selection markers, and corresponding *ply* mRNA folding free energy where appropriate. Order numbers (#) are linked to the other tables in this appendix and thesis, and correspond to ΔG . Selection antibiotics ($\mu\text{g}/\text{mL}$): Km (kanamycin); Amp (ampicillin).

#	ΔG	Strain #	Description	Selection	Notes
		ARE213	NEB10B+pUC57/coPLYregion	Amp200	pUC57 + coPLY (GenScript) + flanked by ARP12 genome
		ARE244	NEB10B+TOPO/outer+coPLY	Km50	NEB10B + TOPO + coPLY with outer regions from ARP12 genome
1	-17.5	ARE354	TOP10+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 112940
2	-12.8	ARE416	NEB10B+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 223932
3	-9.5	ARE415	NEB10B+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 6496
5	-6.2	ARE356	TOP10+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 287208 (CO)
6	-6.2	ARE350	TOP10+TOPO/1k-coPLY	Km50	coPLY with 1kbp outer regions from ARP12 genome; inserted in reverse
7	-3.4	ARE320	TOP10+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 224241
8	-2.5	ARE349	TOP10+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 227715
9	-2.0	ARE488	TOP10+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 188254 (CdO)
10	-0.2	ARE357	TOP10+pCR2.1-TOPO/outup+PLY	Km50	1986bp insert with upstream wt region + PLY seq 6531
		ARE531	TOP10+TOPOpCR2.1/PLYGFP	Km50	TOPO cloned GFP + WT PLY 5' end + RBS for T7 <i>in vitro</i> translation
		ARE532	TOP10+pEXP5-NT/CALML3	Amp100	Control vector from Expressway™ Expression System (Invitrogen)
g1	-17.5	ARE533	TOP10+pEXP5-NT/PLY1GFP	Amp100	PLY1 5' end + GFP construct from PCR, in vector from ARE532
g3	-9.5	ARE534	TOP10+pEXP5-NT/PLY3GFP	Amp100	PLY3 5' end + GFP construct from PCR, in vector from ARE532
g4	-7.3	ARE535	TOP10+pEXP5-NT/PLY4GFP	Amp100	PLY4 5' end + GFP construct (from pARE531), in vector from ARE532
g7	-3.4	ARE536	TOP10+pEXP5-NT/PLY7GFP	Amp100	PLY7 5' end + GFP construct from PCR, in vector from ARE532

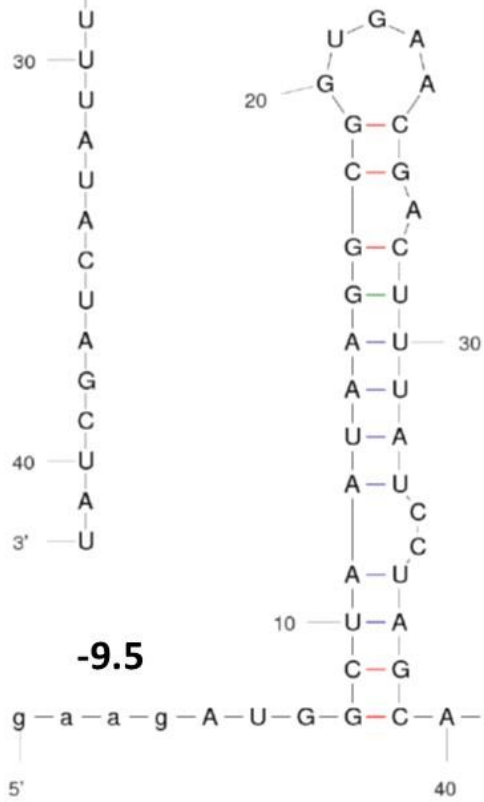
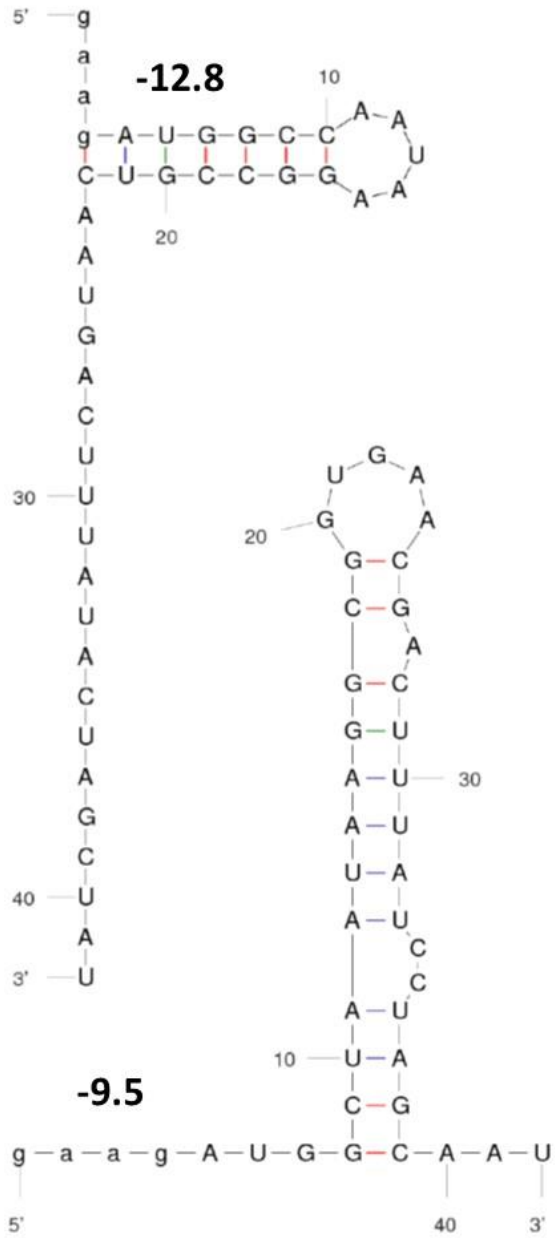
C. Predicted mRNA folding secondary structures

In the following 5 pages, Figure VI.1 in 5 slides.

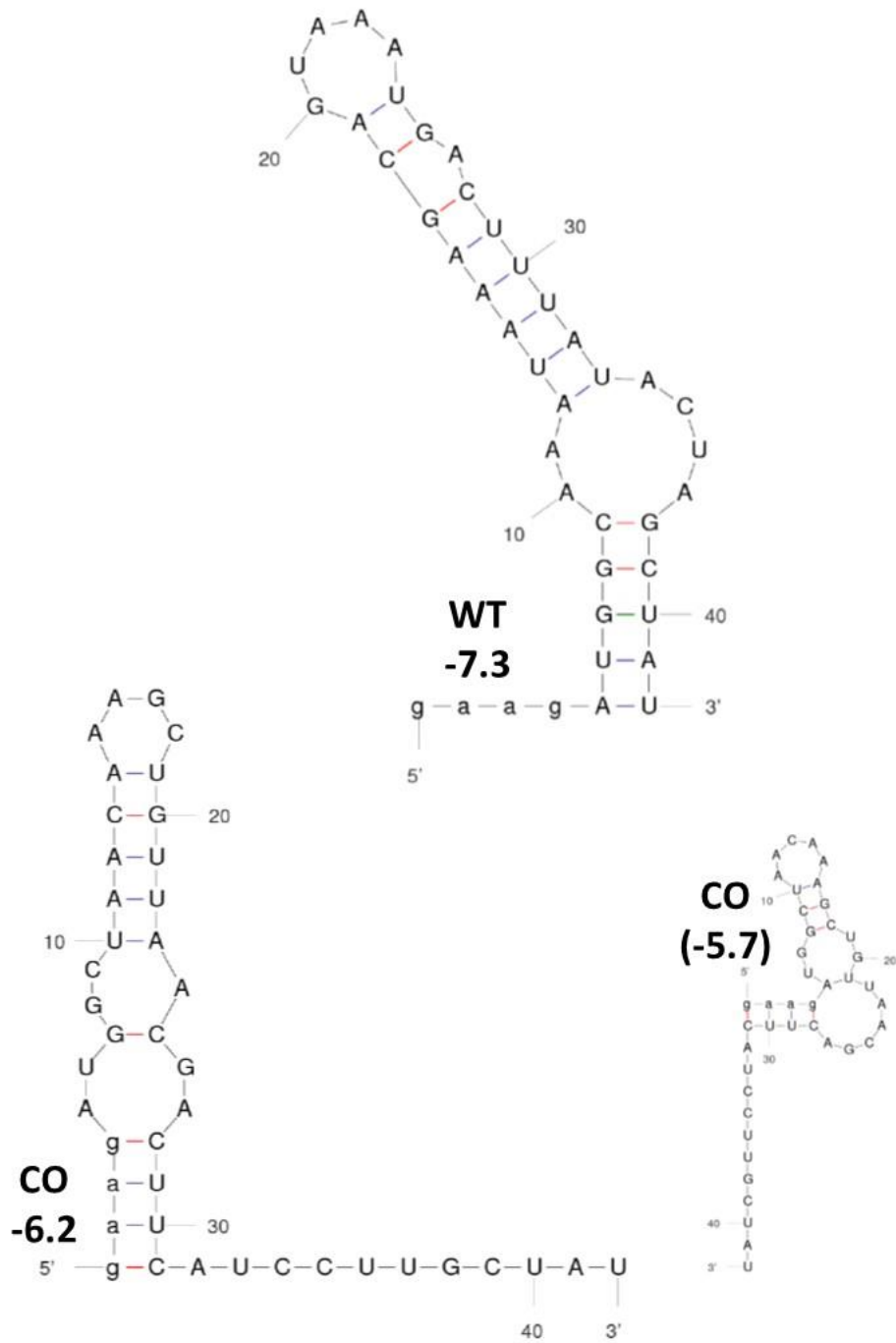
Figure VI.1. Predicted 5' mRNA folding secondary structures.

Using mfold web server (Markham and Zuker, 2005, 2008; Zuker, 2003) we computed the designed 5' mRNA sequences. There are 42 nucleotides to which corresponds the defining ΔG and folding structures, with the starting AUG codon corresponding to nucleotides 5, 6, and 7.

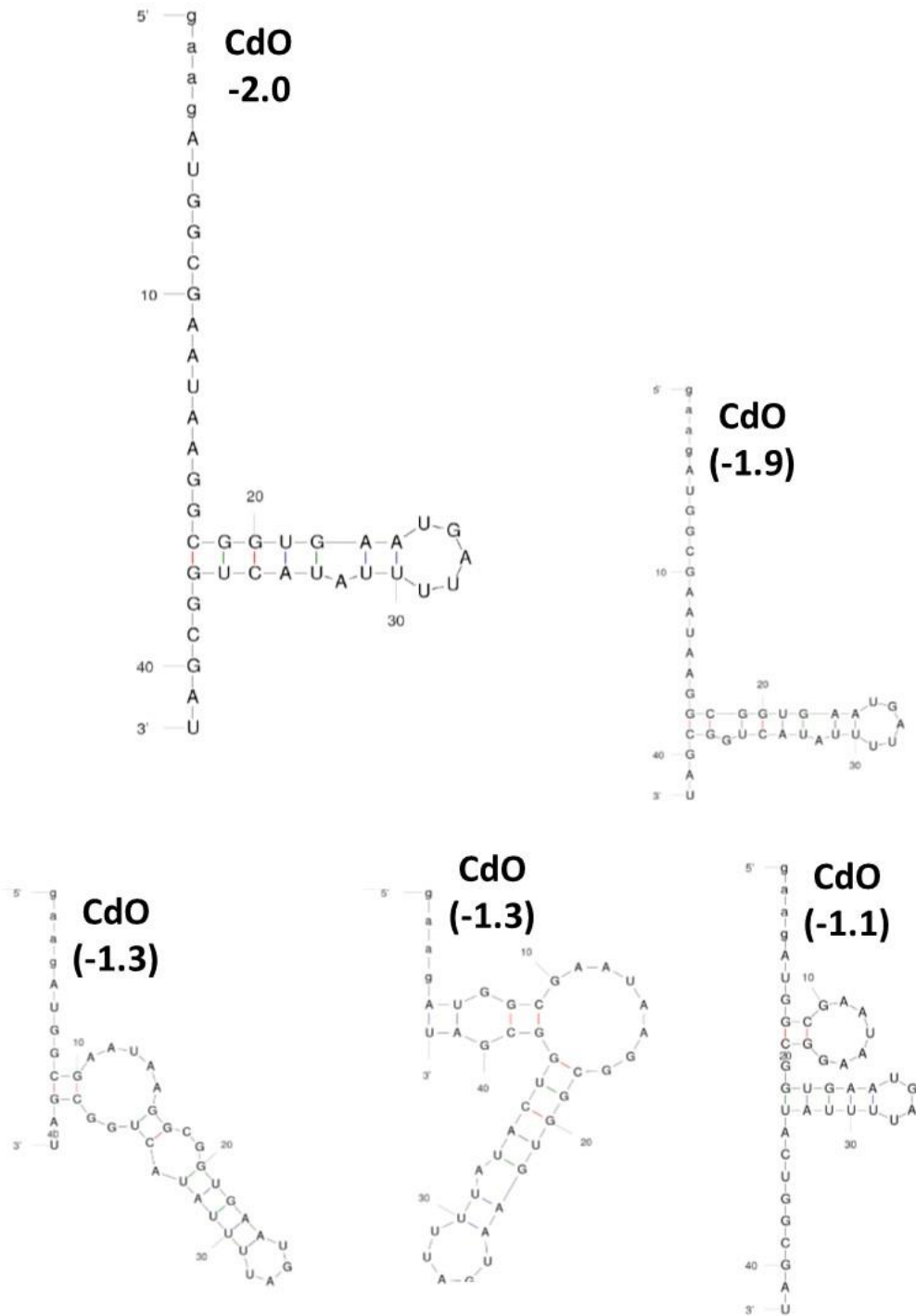
- (A) Strains with minimum ΔG -17.5, -12.8, and -9.5.
- (B) Strains with minimum ΔG -7.3 (WT) and -6.2 (CO with 2 structures).
- (C) Strains with minimum ΔG -3.4 (with 2 minimum structures) and -2.5 (with 2 structures).
- (D) Strain with minimum ΔG -2.0 (CdO with 5 structures).
- (E) Strain with minimum ΔG -0.2 (with 2 minimum structures and 5 higher structures).



A



B



D

D. Growth curves of *S. pneumoniae*

PLY lacks the ability to affect bacterial growth *in vitro* (Los et al., 2013). Therefore, since our strains are otherwise isogenic, the silent mutations that lead to distinct PLY expression, and PLY itself, don't seem to affect exponential phase growth curves depicted in the next figure.

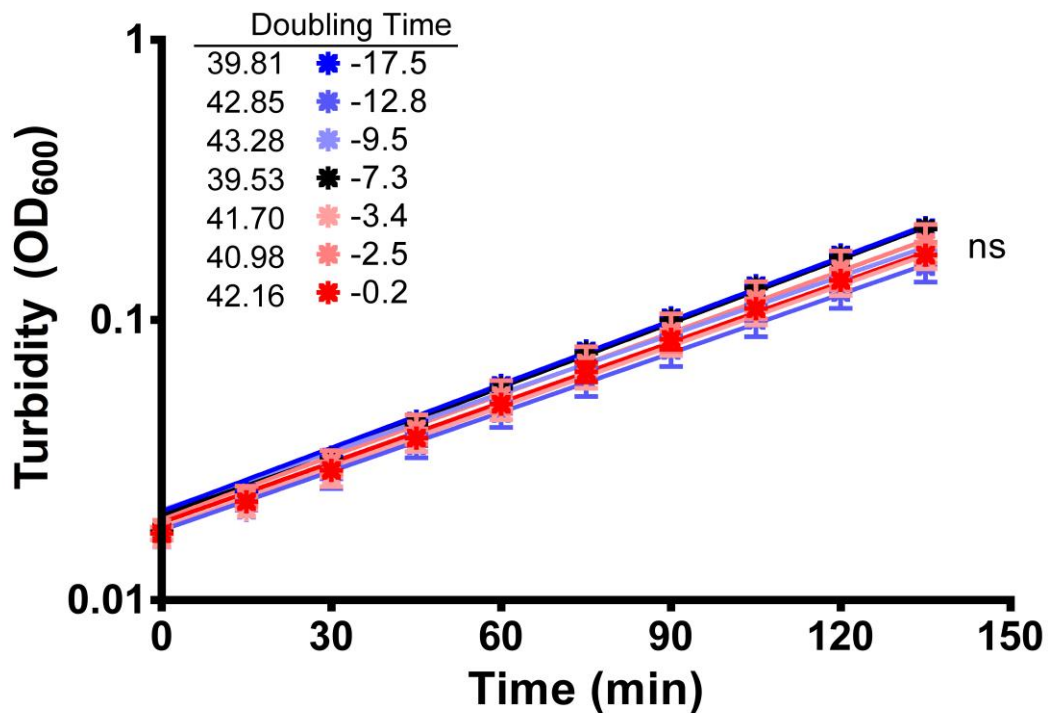


Figure VI.2. Growth curves and doubling times of *S. pneumoniae* R6 strains created for this study.

Data are shown as mean of 5 experiments \pm SEM and were analyzed by One-way ANOVA followed by Tukey's Multiple Comparison Test. ns (not significant). Doubling times were calculated from best-fit exponential growth curves.

Doubling times calculated for all strains in another experiment were found to be similar and around 40 min: 40.5; 41.3; 43.6; 36.7; 37.7; 39.8; 39.4; 37.3; 38.7; 43.3 min for, respectively, -17.5; -12.8; -9.5; -7.3 (WT); -6.2 (CO); -6.2 (fCO); -3.4; -2.5; -2.0 (CdO); -0.2.

A complete representative growth curve without logarithmic scale axis is shown for illustration in the next figure.

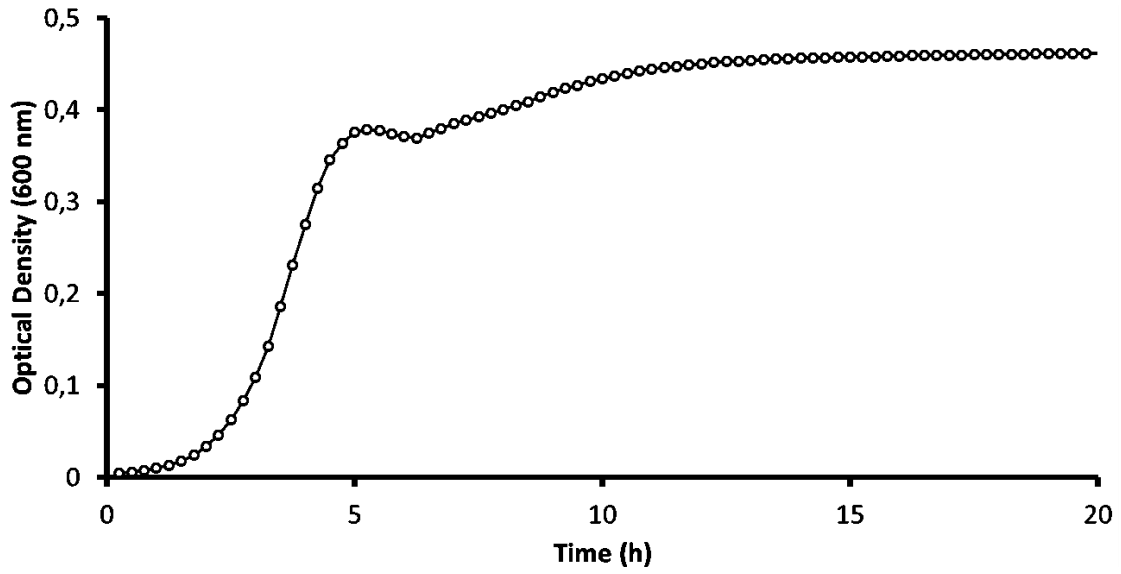


Figure VI.3. Representative complete growth curve for WT *S. pneumoniae* R6.

E. The limits of PLY production with increasing ΔG

We have shown that production of PLY increases with increasing mRNA ΔG at the 5' region. In this part of the study, we present the limits to the maximal levels of PLY that can be produced by this technique. These limits are lower than the theoretical capacity for the system that would otherwise be achieved close to zero kcal/mol of minimum folding free energy. Previous studies have presented similar limits for protein production in an *E. coli* model (ΔG maximum at approximately -4) (Kudla et al., 2009).

The strains we designed have 5' mRNA regions for the *ply* gene that differ in minimum folding free energy ranging from -17.5 to -3.4 kcal/mol and were transformed into *S. pneumoniae* R6 strains (CHAPTER II, cf. Figure II.1). Additionally, we designed strains in which the ΔG was higher, up to -0.2 kcal/mol (Figure VI.4A). All strains exhibited similar growth rates and doubling times (Figure VI.2).

For strains with $\Delta G > -3.4$, we noted less consistency in protein expression and overall values that did not differ significantly from WT (Figure VI.4B), so we focused the remainder of the experiments on strains in the range of $\Delta G = -17.5$ to -3.4 kcal/mol. For each of these 7 strains, we report all the 5' mRNA ΔG values for the predicted secondary structures, their CAI, and their GC% contents (Figure VI.4C). The predicted mRNA folding secondary structures can be found in Appendix C on page 117.

The sequences selected for strain creation have mRNA ΔG values which represent the minimum free energy of the predicted possible structures. The algorithm predicts that sequences with higher minimum values also have more possible secondary folding structures (Figure VI.4C and Figure VI.5B). Some of these may represent unnatural or improbable conformations due to unfavorable reaction conditions as are the examples of the positive ΔG values. Systems try to achieve minimum free energies. Hence, we hypothesize this to be one of the reasons for obtaining less consistent results with higher ΔG strains and also for attaining a cap in PLY production. For these reasons, we opted to continue our work with a selected panel of strains limited up to -3.4 kcal/mol.

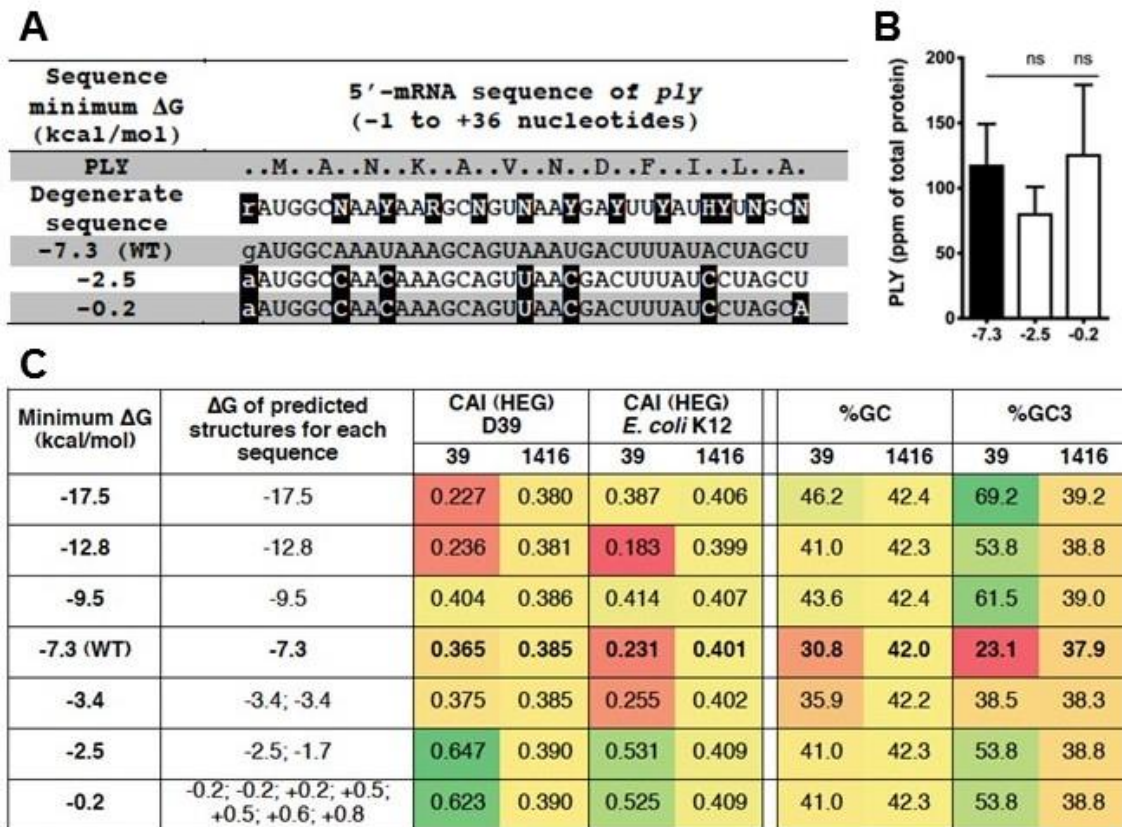


Figure VI.4. Highest values of mRNA folding free energies do not correlate with maximal protein expression.

- (A) Summary of the silent mutations engineered in the 5' mRNA region of the pneumolysin gene in 2 additional otherwise isogenic *Streptococcus pneumoniae* strains constructed for this study. IUPAC Nucleic Acid Codes: Y (pYrimidine); R (puRine); N (aNy); H (not G).
- (B) PLY quantification by ELISA for additional R6 strains. Data are shown as mean of 5 experiments \pm SEM and were analyzed by One-way ANOVA followed by post test for linear trend and Dunnett's Multiple Comparison Test against -7.3 (WT). ns (not significant).
- (C) CAI and GC% heat maps created with Microsoft® Excel 2010 with values for each sequences' first 13 codons (39 nucleotides) or complete *ply* gene (1416 nucleotides). Summary of folding free energies for the predicted structures of each sequence produced using the Quikfold server. Codon Adaptation Indexes (CAI) were calculated based on the predicted highly expressed genes (HEG) database for *S. pneumoniae* D39 and *E. coli* K12 using the CAIcal server (<http://genomes.urv.es/CAIcal/>).

F. Codon optimization of PLY

It has been reported that PLY, like other virulence factors have CAI values of 0.250 to 0.350, and is considered a medium-expressed gene (Martín-Galiano et al., 2004). Some authors have denoted the importance of local context in an individual gene selection with regards to some codons that may be favored across the genome but may act against it locally (Agashe et al., 2012). Other authors have seen codon position-dependency generating CUB (Hockenberry et al., 2014) in which the importance of a codon has different weights with regards to the position it occupies in the gene.

It has been shown that ΔG is more important for the expression of a single gene and it correlates with protein production better than CAI or other determinants of translation (Kudla et al., 2009), although other authors found that codon usage is actually relevant if the 5'-mRNA structures are weak (Supek and Smuc, 2010). Arguing in favor of CUB correlation with translation, some authors provide a ribosome load model denoting an evolutionary link explaining the difference between natural and synthetic genes encoding the same protein (Klumpp et al., 2012).

Codon optimization (CO) has been successfully used in bioengineering for protein expression outside of its natural proteome and genome (Gustafsson et al., 2004). On the other end of the spectrum, deoptimization (CdO) of PLY CP bias has also served good purpose in reducing virulence of pneumococcus (Coleman et al., 2011).

We designed CO and CdO strains in which only the 5' mRNA was mutated like our previous work (see CHAPTER II) and the ΔG s were -6.2 and -2.0 kcal/mol, respectively (Figure VI.5B). Additionally we created the fully CO strain (fCO). Again, strains exhibited similar growth rates and doubling times (Figure VI.2).

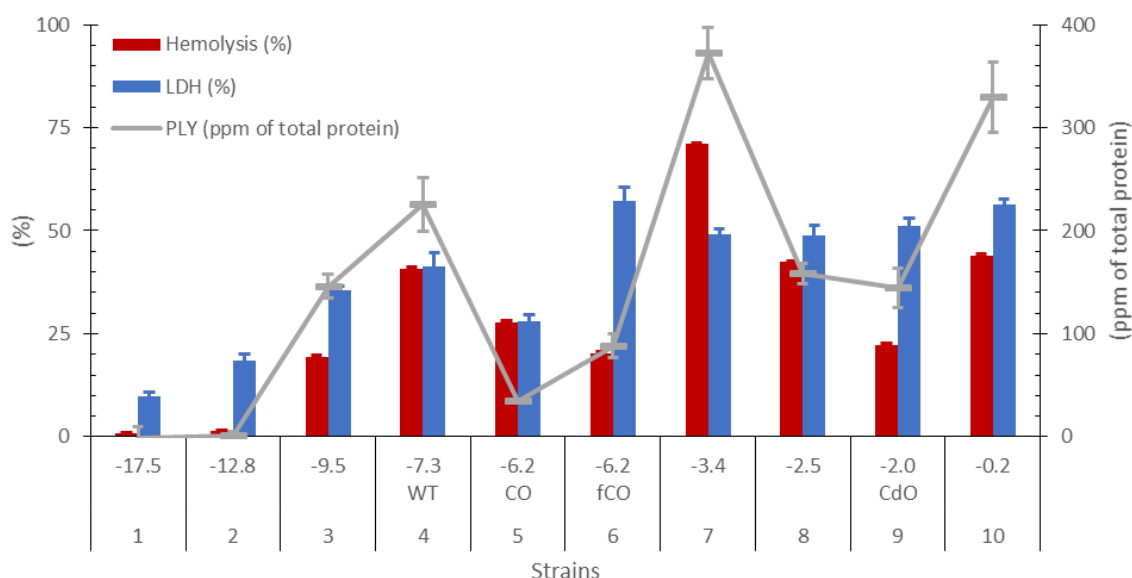
On Figure VI.5A, a plot with a representative experiment using all *S. pneumoniae* R6 strains for comparison in assays of hemolysis, LDH cytotoxicity, and PLY quantification by ELISA.

For this point, we will be comparing strains #4 WT, #5 5'-CO, #6 full CO, and #9 5'-CdO. Interestingly, the highest energy strain #9 uses the least preferred codons during translation initiation. Perhaps CAI and ΔG working against each other revealed no important differences in hemolysis between CO and CdO at the 5' *ply*. However, attributing more relevance to 5' ΔG , CdO strain showed higher cytotoxicity and detectable PLY production than CO strain. When comparing CO and fCO we see comparable levels of PLY, slightly higher for fCO but consequent hemolysis slightly lower. However cytotoxicity for the fCO was the highest.

Overall, other strains (#1, #2, #3, #4, #7, #8, and #10), when compared to these, exhibit a normal range of CAI based on D39 HEG (Figure VI.5B). They also present less predicted folding structures – with the exceptions of #8 and #10 discussed in Appendix E, “The limits of PLY production with increasing ΔG ” on page 125. Potentially because of these points being controlled in those strains, and aligned with the ΔG , they present consistent results discussed in the rest of this thesis.

Some authors argue that the selection of preferred versus unpreferred codons may be too simplistic when looking solely at PHEG. Indeed, increased usage of AT rich codons shortly after the gene start can be a structural requirement of mRNAs (Hockenberry et al., 2014). This can potentially explain position-dependent CUB that postulates the degree of codon usage to be different depending on the region it stands. The implication particular to this work is that the choice of optimal codons to the region start may be wrong because their position was unaccounted for.

This experiment would merit more research because of a lack of consistency in results that are difficult to explain in light of the controversial knowledge on this subject.

A**B**

mRNA ΔG (kcal/mol)	predicted structures	D39 CAI (HEG)		GC%		GC3%	
		39	1416	39	1416	39	1416
-17.5	-17.5	0,227	0,380	46,2	42,4	69,2	39,2
-12.8	-12.8	0,236	0,381	41,0	42,3	53,8	38,8
-9.5	-9.5	0,404	0,386	43,6	42,4	61,5	39,0
-7.3 (WT)	-7.3	0,365	0,385	30,8	42,0	23,1	37,9
-6.2 (CO)	-6.2; -5.7	1,000	0,394	41,0	42,3	53,8	38,8
-6.2 (fCO)	-6.2; -5.7	1,000	1,000	41,0	41,2	53,8	29,7
-3.4	-3.4; -3.4	0,375	0,385	35,9	42,2	38,5	38,3
-2.5	-2.5; -1.7	0,647	0,390	41,0	42,3	53,8	38,8
-2.0 (CdO)	-2.0; -1.9; -1.3; -1.3; -1.1	0,148	0,376	43,6	42,4	61,5	39,0
-0.2	-0.2; -0,2; +0,2; +0,5; +0,5; +0,6; +0,8	0,623	0,390	41,0	42,3	53,8	38,8

Figure VI.5. Codon optimization and deoptimization with regards to mRNA folding free energies.

- (A) Representative experiment of Hemolysis, LDH cytotoxicity, and PLY quantification by ELISA for all *S. pneumoniae* R6 strains. Data are shown as average values \pm STD.
- (B) CAI and GC% heat maps created with Microsoft® Excel 2013 with values for each sequences' first 13 codons (39 nucleotides) or complete *ply* gene (1416 nucleotides). Summary of folding free energies for the predicted structures of each sequence produced using the Quikfold server. Codon Adaptation Indexes (CAI) were calculated based on the predicted highly expressed genes (HEG) database for *S. pneumoniae* D39 using the CAIcal server (<http://genomes.urv.es/CAIcal/>).

G. *Drosophila melanogaster* model for rheostat pneumolysin expression

This appendix contains original unpublished research material that yielded unclear results that are worth noting and merit ponderation. Experiments carried out by me, in the laboratory of Mimi Shirasu-Hiza, to whom I am fondly grateful for the help, materials, and animals.

We used WT *Drosophila* as another *in vivo* model because they are susceptible to R6 strains, and since they present a number of technical advantages despite their limitations. However, as we can see in the following representative figure, despite promising CFU results (at 20h, increasing CFU counts with increments in ΔG), fly survival curves provided very inconsistent results and no conclusions can be truly attained.

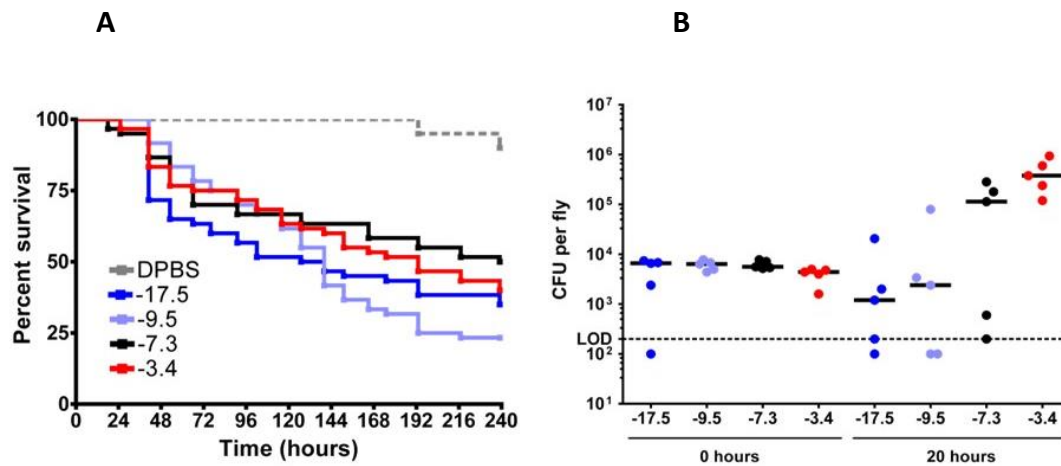


Figure VI.6. Representative *D. melanogaster* infection results.

Adult male WT flies with 20 days of age were anesthetized with CO_2 (for no longer than 15 min at a time). Flies were injected intra-abdominally with approximately 50 nL of sterile DPBS with or without *S. pneumoniae* R6, using pulled glass capillary needles. Bacterial strains grown to similar exponential phase OD were resuspended in DPBS and OD_{600} was corrected to 0.300. 60 Flies per experimental group and 20 flies in the control group were injected and kept at 29 °C. CFUs for whole homogenized fly were counted at 0 and 20 hours post infection. 5' mRNA folding free energies are color coded in the legend. (A) Survival curves and (B) CFUs per fly.

H. *Caenorhabditis elegans* model for rheostat pneumolysin expression

This appendix contains original unpublished research material that yielded unclear results that are worth noting and merit ponderation. Experiments carried out by me, with the help of Ferdinand Los, to whom I acknowledge all the help provided.

We tested *C. elegans* as another *in vivo* model of infection as we wanted to assess p38 responses to our strains that exhibit a wide range of PLY amounts. They also possess a number of technical advantages despite their limitations. We used WT and a p38 deficient mutant. We show in the following representative figure that infection after 72h by oral ingestion of pneumococcus was in part lethal for the WT and very detrimental for the Δ p38 worms. Results hint at a trending effect that inversely correlates with the amount of PLY with higher alive fractions with increments in Δ G.

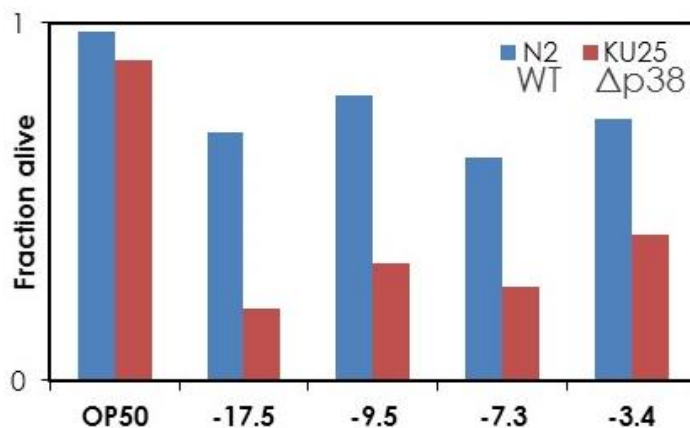


Figure VI.7. Representative *C. elegans* infection results.

WT and p38-deficient worms were incubated for 72h with bacteria grown to similar exponential phase, resuspended, and OD₆₀₀ corrected to 2.000 – *E. coli* OP50 (inoculum around 2.5x10⁹ CFU/mL) and mRNA folding free energy of *S. pneumoniae* D39 (inoculum around 4.0x10⁸ CFU/mL) strains (-7.3 is WT D39).

I. Published article: “Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by *Streptococcus pneumoniae*”

Amaral, F.E., Parker, D., Randis, T.M., Kulkarni, R., Prince, A.S., Shirasu-Hiza, M.M., and Ratner, A.J. (2015). PLoS ONE 10(3): e0119823. doi:10.1371/journal.pone.0119823.


RESEARCH ARTICLE

Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by *Streptococcus pneumoniae*

Fábio E. Amaral^{1,4,5}, Dane Parker¹, Tara M. Randis¹, Ritwij Kulkarni¹, Alice S. Prince^{1,2}, Mimi M. Shirasu-Hiza³, Adam J. Ratner^{1*}

1 Department of Pediatrics, Columbia University, New York, NY United States of America, **2** Department of Pharmacology, Columbia University, New York, NY, United States of America, **3** Department of Genetics & Development, Columbia University, New York, NY, United States of America, **4** Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal, **5** ICVS/3B's, PT Government Associate Laboratory, Braga/Guimarães, Portugal

* ar127@columbia.edu


 OPEN ACCESS

Citation: Amaral FE, Parker D, Randis TM, Kulkarni R, Prince AS, Shirasu-Hiza MM, et al. (2015) Rational Manipulation of mRNA Folding Free Energy Allows Rheostat Control of Pneumolysin Production by *Streptococcus pneumoniae*. PLoS ONE 10(3): e0119823. doi:10.1371/journal.pone.0119823

Academic Editor: Indranil Biswas, University of Kansas Medical Center, UNITED STATES

Received: November 7, 2014

Accepted: January 16, 2015

Published: March 23, 2015

Copyright: © 2015 Amaral et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by the National Institutes of Health (www.nih.gov) (R01 AI092743 and R21 AI111020 to A.J.R.). F.E.A. was supported by the Portuguese Foundation for Science and Technology (www.fct.pt) SFRHBD/33901/2009 and the Luso-American Development Foundation (www.fad.pt). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The contribution of specific factors to bacterial virulence is generally investigated through creation of genetic “knockouts” that are then compared to wild-type strains or complemented mutants. This paradigm is useful to understand the effect of presence vs. absence of a specific gene product but cannot account for concentration-dependent effects, such as may occur with some bacterial toxins. In order to assess threshold and dose-response effects of virulence factors, robust systems for tunable expression are required. Recent evidence suggests that the folding free energy (ΔG) of the 5' end of mRNA transcripts can have a significant effect on translation efficiency and overall protein abundance. Here we demonstrate that rational alteration of 5' mRNA folding free energy by introduction of synonymous mutations allows for predictable changes in pneumolysin (PLY) expression by *Streptococcus pneumoniae* without the need for chemical inducers or heterologous promoters. We created a panel of isogenic *S. pneumoniae* strains, differing only in synonymous (silent) mutations at the 5' end of the PLY mRNA that are predicted to alter ΔG . Such manipulation allows rheostat-like control of PLY production and alters the cytotoxicity of whole *S. pneumoniae* on primary and immortalized human cells. These studies provide proof-of-principle for further investigation of mRNA ΔG manipulation as a tool in studies of bacterial pathogenesis.

Introduction

Delineating specific bacterial factors involved in interaction with the host is crucial to understanding mechanisms of pathogenesis and developing targeted therapies. Classically, such

Competing Interests: A.J.R. is a Section Editor at PLOS ONE. A.S.P. is an Editorial Board Member at PLOS Pathogens. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

investigations involve construction of bacteria with disruptions in the genes encoding candidate factors ("knockouts"; KO) and comparison with wild-type (WT) or complemented strains. However, the level of expression of bacterial genes may also play a role in pathogenesis. The KO vs. WT paradigm cannot account for such differences, and although systems exist for inducible expression of specific bacterial genes, these generally rely on exogenous activators such as tetracycline that may have variable delivery to relevant sites during infection [1]. Thus, it is desirable to develop strategies for genetically encoded, tunable, rheostat-like control of bacterial gene expression.

A number of factors determine the efficiency of protein production. As the genetic code is degenerate, choice among synonymous codons may play a role in translational efficiency [2]. Early models postulated a direct relationship between tRNA availability and protein production, implying that use of non-optimal codons (i.e. those that are rarely used among other genes in the species being studied) would lead to decreased protein production on the basis of tRNA scarcity [2–4]. Codon-optimization has been used as a means to increase production of heterologous genes in *E. coli* and other host species [5]. Conversely, deoptimization of codons or codon pairs has been employed to rationally decrease protein production from natively transcribed genes [6,7]. Kudla et al. examined a library of synonymous codon substitutions in heterologously expressed green fluorescent protein (GFP) in *E. coli* and demonstrated that the mRNA folding free energy (ΔG), particularly at the 5' terminus, correlates strongly with translational efficiency and with overall protein production [8]. Goodman et al. investigated the role of both codon bias and mRNA ΔG in synthetic reporters and confirmed a prominent role for ΔG in shaping expression levels of individual genes [9].

We used *Streptococcus pneumoniae* (pneumococcus), a major human pathogen with robust tools for genetic manipulation [10], as a model organism. We constructed a panel of isogenic pneumococcal strains differing only in synonymous codons at the 5' end of the gene encoding pneumolysin (PLY), an established virulence factor [11,12], without antibiotic selection cassettes or exogenous promoters. These modifications altered the predicted mRNA ΔG and resulted in graded PLY production, thus affecting host-bacterial interactions and providing proof-of-principle for the use of rational modification of mRNA ΔG as a means to control protein production.

Materials and Methods

Ethics statement

The use of primary human erythrocytes following written informed consent was approved by the Institutional Review Board of Columbia University Medical Center. The human cell line A549 was obtained from ATCC (catalog number CCL-185).

Calculation of ΔG and codon adaptation index

Sequence folding free energies were calculated using the DINAMelt/Quickfold web server (<http://mfold.ma.albany.edu/?q=DINAMelt/Quickfold>) [13]. The software uses predicted free energies at 37°C and enthalpies from the Turner laboratory at the University of Rochester in Rochester, NY. For ΔG calculations we used version 3.0 free energies. Codon adaptation index (CAI) based on the set of highly expressed genes from *Streptococcus pneumoniae* strain D39 was calculated using the CAIcal server (<http://genomes.urv.es/CAIcal/>) [14].

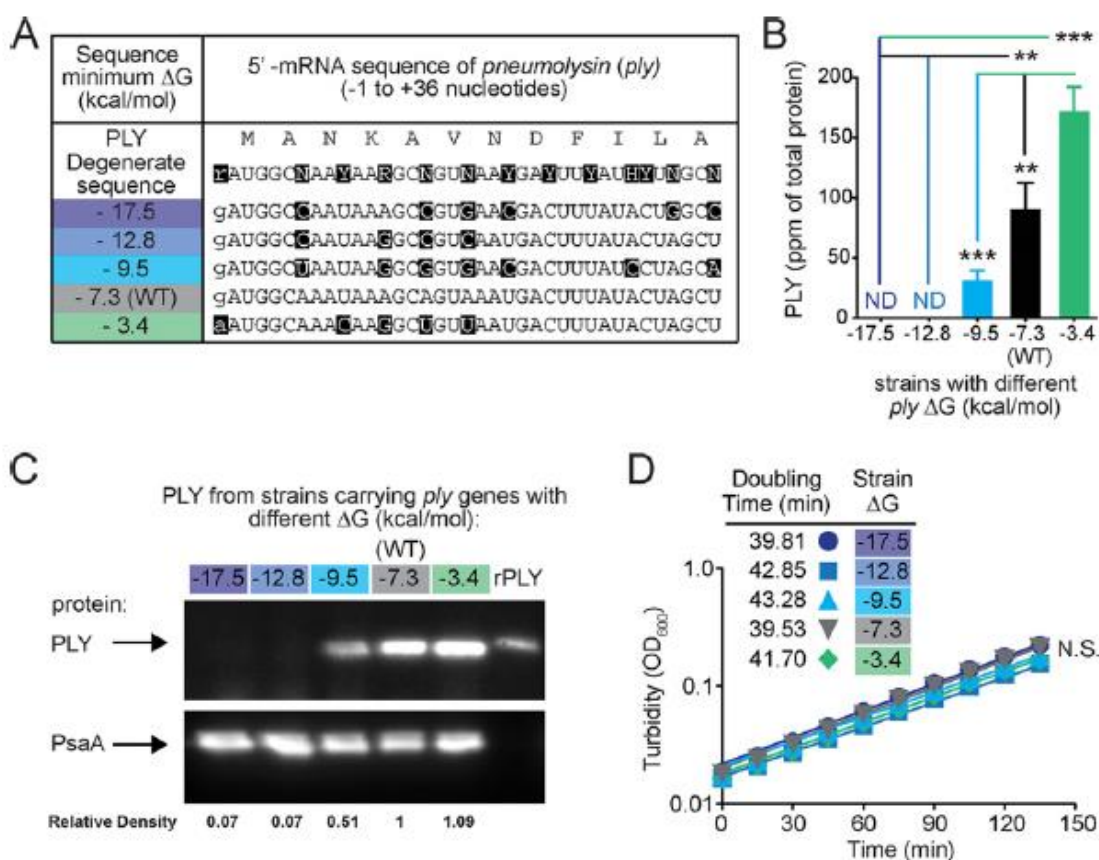


Fig 1. Production of PLY increases with increasing *ply* 5' mRNA folding free energy. (A) Summary of the silent mutations engineered in the 5'-mRNA region of the *ply* gene. IUPAC Nucleic Acid Codes: Y (pYrimidine); R (puRine); N (aNy); H (not G). (B) Increasing levels of predicted *ply* 5' mRNA ΔG correlate with increased PLY production as determined by ELISA normalized to total pneumococcal protein. Data are shown as mean of 4 experiments \pm SEM. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (ANOVA). (C) Western blot (top, anti-PLY; bottom, anti-PsaA loading control) demonstrates different levels of PLY production in engineered strains (lanes 1–5) compared to recombinant PLY (lane 6) as a control. Numbers beneath each lane indicate the ratio of PLY to PsaA band intensity, normalized to the value for the wild-type ($\Delta G = -7.3$) strain in lane 4. (D) Measurement of turbidity (OD_{600}) and calculated doubling times shows identical growth kinetics among isogenic *S. pneumoniae* strains with altered *ply* 5' sequences. (N.S. = not significant; ANOVA)

doi:10.1371/journal.pone.0119823.g001

Bacterial strains, primers, plasmids, transformation and growth conditions

We used the Janus cassette [10], to create a panel of isogenic *S. pneumoniae* R6 derivatives with a wide range of ΔG values (-17.5 to -3.4 kcal/mol) for the 5' end of *ply* gene (-4 to +38 bp, with 1 being the A in the first ATG). To control for unexpected effects of transformations we recreated, and sequenced, a strain with the WT *ply* allele for comparison. Each of the new *ply* constructs was inserted into the native chromosomal locus, without resistance markers or alteration in predicted primary amino acid sequence. All transformed loci were confirmed by sequencing. All constructs were first assembled in *E. coli* TOP10 using plasmid pCR2.1-TOPO (Invitrogen), amplified by PCR, and transformed into pneumococcus. Forward primers differed in nucleotides encoding the initial PLY region according to Fig 1, consistent with the

following degenerate sequence: 5'-GAAR **ATG** GCN AAY AAR GCN GTN AAY GAY TTY ATH YTN GCN ATG AAT TAC G-3' (start codon in bold), and the reverse primer was 5'-CTA GTC ATT TTC TAC CTT ATC TTC T-3'. Constructs for pneumococcal transformation were assembled by overlap-extension PCR and included 570 bp upstream of *ply* (forward primer 5'-GGT TAT TGG CGA CAA GCA TT-3') and 769 bp downstream (reverse primer 5'-CCT GCT AAG ATG GTC TTG CC-3').

Briefly, pneumococcal transformation was achieved with use of semisynthetic casein plus yeast extract (C+Y) medium [15]. Cultures were grown at 37°C in C+Y pH 6.8 to 0.15 OD₆₀₀, then diluted 1:20 fold into C+Y pH 8.0 (total volume 1 mL) at 30°C, supplemented with 1 µg of CSP-1 and 11 µL of 1% CaCl₂ and incubated for 12 minutes. After adding 1 µg DNA, cultures were incubated for 40 minutes at 30°C, followed by 90 minutes at 37°C, followed by appropriate selection in agar plates incubated overnight at 37°C, 5% CO₂.

E. coli were grown in lysogeny broth (LB) or agar supplemented with appropriate antibiotics for plasmid selection (ampicillin 200 µg/mL or kanamycin 50 µg/mL). *S. pneumoniae* were grown in tryptic soy broth (TSB) or agar supplemented with 200 U/mL of catalase (Worthington Biochemical), and appropriate antibiotics for strain selection (streptomycin 150 µg/mL or kanamycin 200 µg/mL).

Western blotting and ELISA

Protein separation, transfer, and immunoblotting were performed as previously described [16]. In blots, PLY was detected using 1:1000 dilution of mouse monoclonal anti-pneumolysin (1F11), PsaA was detected using 1:25000 dilution mouse monoclonal PsaA antibody (8G12) [17]. Densitometry was performed using the gel analysis plugin implemented in ImageJ (version 1.45s; National Institutes of Health). PLY was detected by indirect ELISA using monoclonal anti-pneumolysin (1F11; 1:1000 dilution).

qRT-PCR and mRNA half-life

S. pneumoniae were grown to early exponential stage (OD₆₀₀ = 0.3) and stored in RNAlater (Ambion). RNA was isolated using the RiboPure-Bacteria Kit (Ambion) with DNase treatment. cDNA was made using the SuperScript III RT Kit (Invitrogen), qRT-PCR was performed using Quanta SYBR Green PCR Master Mix in a StepOne Plus thermal cycler (Applied Biosystems).

The primers for the *ply* gene of *S. pneumoniae* (GenBank accession no. M17717) for real-time reverse-transcription PCR were previously described [18]. Samples were normalized to *S. pneumoniae* 16S rRNA detected with forward (5'-GCC TAC ATG AAG TCG GAA TCG-3') and reverse (5'-TAC AAG GCC CGG GAA CGT-3') primers. In separate experiments, transcription was arrested by addition of rifampicin to 10 µg/mL and mRNA abundance and decay were assessed by qRT-PCR of serially collected samples normalized to the 30 sec time point. mRNA half-life was calculated from best-fit exponential decay curves.

Hemolysis assay

Primary human erythrocytes (0.5% final concentration) were incubated for 30 min at 37°C with an equal volume of heat-killed (52°C, 20 min) log-phase (OD₆₀₀ = 0.25) *S. pneumoniae* R6 strains suspended in PBS. Intact erythrocytes were pelleted by centrifugation, and hemolysis was measured by assaying hemoglobin concentration in the supernatant (OD₄₁₅ of a 100µl sample). Hemolysis values were normalized to 0% (vehicle control) and 100% lysis (1% Triton X-100) controls as previously described [16,19].

Cell culture and cytotoxicity assay

A549 respiratory epithelial cells (obtained from ATCC, catalog number CCL-185) were grown in minimum essential medium (MEM) with 10% fetal bovine serum as described [20]. Cell monolayers (80–90% confluent in sterile 24-well plates) were incubated in MEM without serum overnight prior to stimulation with $\sim 10^8$ cfu of the indicated *S. pneumoniae* strains for 12 hrs. LDH cytotoxicity assay (Roche) was performed as per the manufacturer's instructions, and values were normalized to 0% (vehicle control) and 100% lysis (1% Triton X-100) controls as previously described [21].

Statistical analysis

Statistics were performed with Prism 5 software (GraphPad, Inc.). Data were analyzed by analysis of variance followed by post-tests to account for multiple comparisons, as appropriate. In vitro experiments were performed at least three times with at least three technical replicates. Bars represent SEM. Values of $P \leq 0.05$ were considered significant.

Results

Production of PLY increases with increasing ΔG

Using a degenerate nucleotide sequence derived from the *S. pneumoniae* R6 PLY protein sequence, we designed several candidate 5' mRNA regions for the *ply* gene (1416 nucleotides total) that differed in predicted minimum mRNA folding ΔG (calculated using the standard mfold algorithm over the region from -4 to +38 nucleotides [13]). Each of these genes only contained synonymous mutations, leading to the production of identical amino acid sequences, but covering a wide range of predicted ΔG values. Because only a small number of codons was altered in each strain (4–7 codons/strain), the CAI for the full length *ply* gene based on highly expressed genes of *S. pneumoniae* D39 was unchanged (~ 0.38 in all strains). The two strains with the lowest 5' mRNA ΔG (-17.5 and -12.8) also had lower CAI over the first 39 bp (0.227 and 0.236, respectively) than the other three strains (0.404, 0.365, and 0.375 for the $\Delta G = -9.5$, -7.3, and -3.4 strains, respectively). The predicted ΔG of the native *ply* 5' mRNA is -7.3 kcal/mol, and this strain was included as a control in all experiments (Fig. 1A). These alterations resulted in differential expression of PLY protein. With increasing ΔG values, we observed higher PLY levels as measured by both ELISA (Fig. 1B) and western blot (Fig. 1C). As an important control for changes in bacterial growth that would obviously impact total protein levels, we confirmed that all strains exhibited similar growth rates and doubling times (Fig. 1D). Thus alterations of the *ply* DNA sequence that alter the predicted ΔG of its resulting mRNA led to predictable and quantitative alterations in translated protein levels.

Translation kinetics drive the relationship between ΔG and protein abundance

In order to understand the mechanisms underlying the relationship between mRNA ΔG and protein production in *S. pneumoniae*, we tested whether differences in *ply* mRNA levels correlated with predicted ΔG values for the 5' mRNA and could account for differences in PLY protein levels. We used real-time reverse transcriptase PCR to quantify *ply* mRNA normalized to 16S rRNA for each strain. There were no statistically significant differences between wild-type and engineered strains in mRNA abundance (Fig. 2). Thus the absolute levels of mRNA abundance do not correlate with predicted ΔG values or with differences in PLY protein levels.

Decay kinetics of mRNA can also alter overall protein production. To test whether altered decay kinetics correlated with differences in PLY protein levels, we next assessed mRNA decay

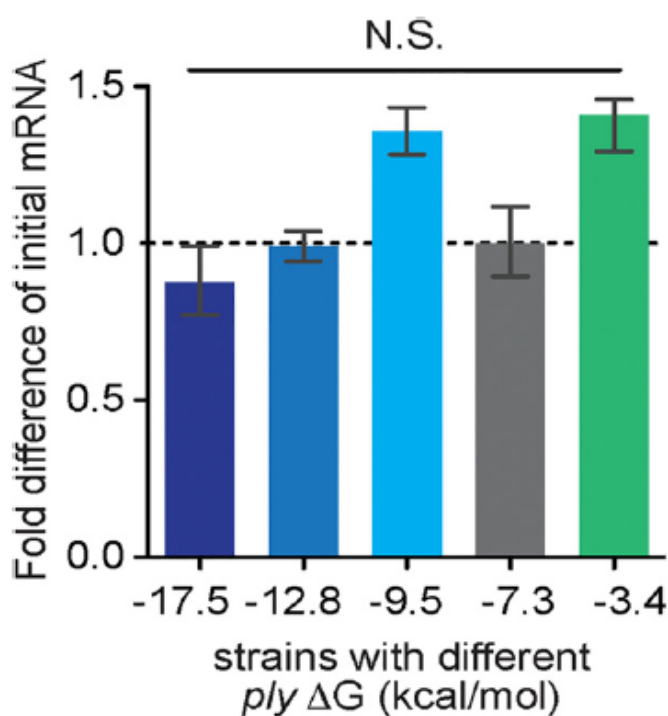


Fig 2. Engineered differences in 5' mRNA ΔG do not alter *ply* mRNA quantity. Fold difference in *ply* mRNA levels (y axis) was calculated relative to wild-type ($\Delta G = -7.3$) at early exponential phase. Relative mRNA amounts were calculated by $\Delta\Delta C_T$ normalized to pneumococcal 16S rRNA. (N.S. = not significant; ANOVA)

doi:10.1371/journal.pone.0119823.g002

at several time points following arrest of transcription with rifampicin (Fig. 3A). Decay curves were similar and did not differ statistically. We also calculated mRNA half-lives based on mRNA abundance following transcription arrest. mRNA half-lives for each *ply* sequence were all approximately 1.5 min, and there were no statistically significant differences among groups (Fig. 3B). Taken together, these findings suggest that mRNA synthesis and decay are not altered and do not account for differences in PLY protein levels. Thus our results are consistent with the hypothesis that changes in mRNA ΔG lead to changes in post-transcriptional processing and changes in protein abundance.

ply mRNA ΔG affects host-pathogen interactions in vitro

The previous experiments were performed using bacteria in culture. To determine whether altered 5' mRNA ΔG also resulted in altered PLY production and cytotoxicity during infection of human cells, we examined PLY protein levels in two model systems: primary human erythrocytes and immortalized epithelial cells. PLY protein causes lysis of infected cells. We found that primary human erythrocytes exposed to *S. pneumoniae* strains exhibited lysis rates that correlated with predicted *ply* mRNA ΔG (Fig. 4A), suggesting that PLY protein levels expressed during infection correlate with predicted 5' mRNA ΔG . We confirmed these differences in PLY protein levels by direct measurement using ELISA (Fig. 4B). In parallel, we found that measurement of A549 epithelial cell cytotoxicity demonstrated a similar positive correlation between

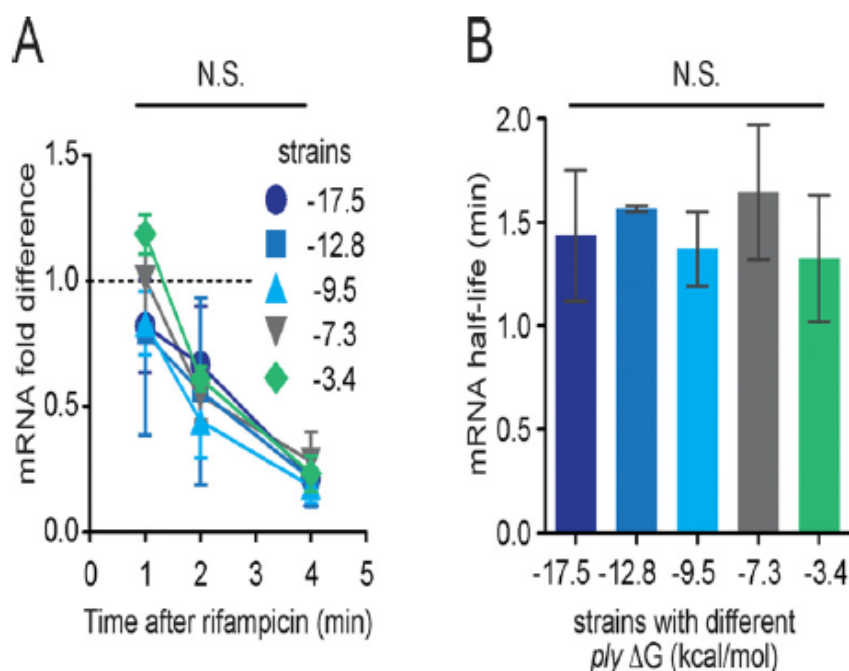


Fig 3. Differences in PLY production do not correlate with mRNA decay kinetics. (A) Abundance of *ply* mRNA following transcriptional arrest with rifampicin. mRNA quantities were measured by qRT-PCR and normalized to 16S rRNA. (N.S. = not significant; ANOVA). (B) Differences in *ply* 5' mRNA ΔG do not alter mRNA half-life values, as calculated from best-fit exponential decay curves from data in (A). (N.S. = not significant; ANOVA)

doi:10.1371/journal.pone.0119823.g003

ply mRNA ΔG and cellular damage (Fig. 5). Thus predicted differences in *ply* mRNA ΔG positively correlate with differences in lysis rates and differences in PLY protein expression levels during active infection of human tissue culture cells.

Discussion

There is an emerging literature on the importance of 5' mRNA folding ΔG to efficiency of translation and to overall levels of protein production [8,9,22–24]. Manipulation of codon bias, which may also impact protein levels, has been widely used as a method to enhance expression of genes in heterologous systems [5,25,26]. Rational alteration of 5' mRNA ΔG has been suggested as a technique that might further aid such production [8,9]. Most studies of the ΔG /protein abundance relationship have focused on either synthetic constructs driving marker genes such as GFP [8,9] or on examination of naturally occurring coding sequences in model organisms [22,23,27,28]. In these studies, we rationally manipulated the predicted 5' mRNA ΔG of a known pneumococcal virulence factor in order to determine concentration-dependent effects. Such investigations are not possible using traditional KO vs. WT approaches, and determining the relationship between protein abundance and virulence may drive more detailed understanding of pathogenic mechanisms.

Using unmarked, isogenic pneumococcal strains, we showed that rheostat-like control of PLY production is feasible and that changes in ΔG are sufficient to alter host-pathogen interactions. Alteration in expression did not depend on mRNA abundance or stability but occurred

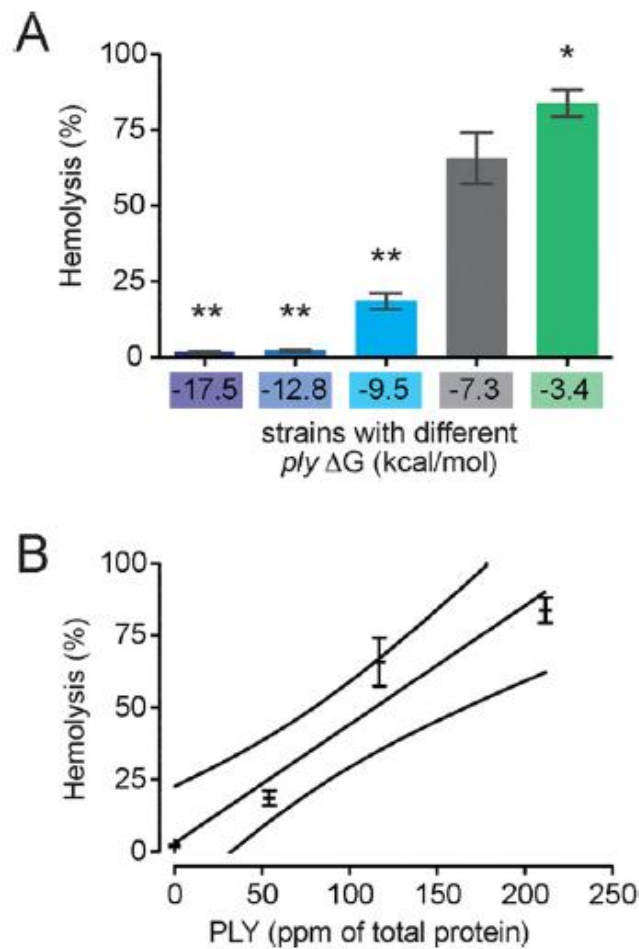


Fig 4. Differential PLY production by *S. pneumoniae* strains is directly related to PLY production. Lysis of primary human erythrocytes as assessed by hemoglobin release assay correlates with (A) *ply* 5' mRNA ΔG and (B) PLY as measured by ELISA ($R^2 = 0.95$; outside lines represent 95% CI). *, $P < 0.05$; **, $P < 0.01$ (ANOVA).

doi:10.1371/journal.pone.0119823.g004

at the level of translation. As predicted, there was a direct correlation between PLY production and target cell lysis using simplified *in vitro* systems, and we provided proof-of-principle for the use of such systems to assist in the study of host-pathogen interactions. Expanded testing of modified strains in animal models of will allow assessment of the durability of altered PLY regulation during *S. pneumoniae* infection. Such models will facilitate understanding of the role of PLY expression levels at different stages and sites of infection.

Our findings are in agreement with those of Coleman et al., who used rational alteration of codon pair bias to alter PLY expression and found attenuation of inflammatory responses *in vivo* [7]. In contrast to that work, we alter PLY expression not by altering codon pair bias but by altering mRNA ΔG. Moreover, rather than altering the entire sequence, we were able to alter a short and specific part of the 5' region to alter PLY protein levels and lysis activity. The

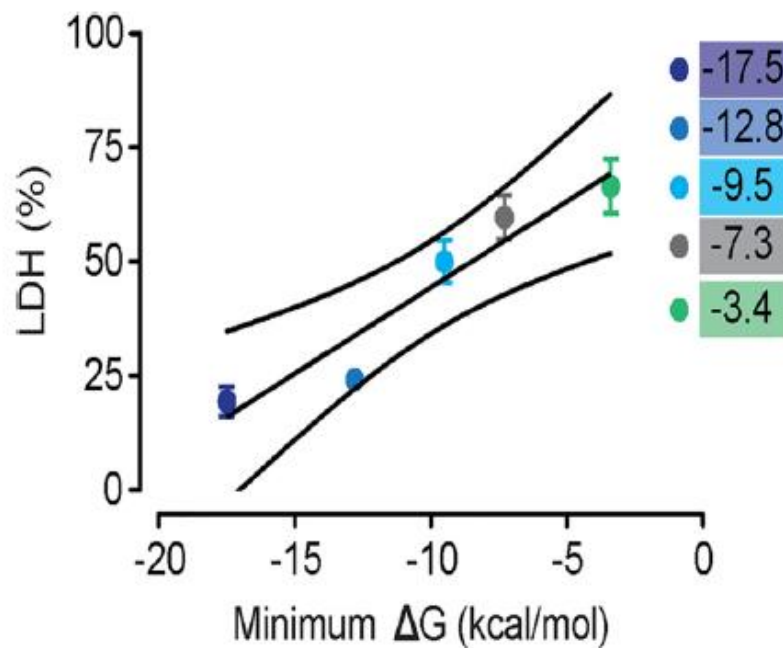


Fig 5. Lysis of human A549 respiratory epithelial cells as measured by LDH release correlates with predicted *ply* 5' mRNA ΔG ($R^2 = 0.91$; outside lines represent 95% CI; $P < 0.0001$, ANOVA for linear trend).

doi:10.1371/journal.pone.0119823.g005

ability to alter a short portion of genetic sequence provides a potentially simple and straightforward tool for precisely regulating any bacterial protein involved in pathogenesis without having to alter large sections of the coding sequence and also without *a priori* knowledge of the host cell's codon preferences.

The correlation between 5' mRNA ΔG and overall protein production is not perfectly linear, and other factors, including codon bias, GC content, and alteration of folding by other segments of the mRNA may have substantial effects on efficiency. Because of the limited number of strains investigated, we were unable to formally evaluate the role of 5' mRNA ΔG while rigorously controlling each of these other variables, and, as noted above, our lowest expressing strains ($\Delta G = -17.5$ and -12.8) had low 5' CAI in addition to low 5' mRNA ΔG . However, the other three strains ($\Delta G = -9.5$, -7.3 , and -3.4) did not differ substantially in 5' CAI and had altered PLY production consistent with a role for 5' mRNA ΔG . Whether this technique is widely applicable to alter protein production across a range of target genes and hosts must be determined experimentally. Despite these limitations, rational alteration of 5' mRNA ΔG has potential as an important new tool in bacterial pathogenesis research. Given the increasing availability of whole genome sequences, much emphasis has been placed on the presence or absence of specific factors as markers of potential virulence in individual strains. Much less is known about the impact of relative expression levels on pathogenesis. The ability to investigate threshold and concentration-dependent effects of specific virulence factors without a requirement for non-native promoters, chemical inducers, or selective markers is extremely valuable and merits further exploration in a variety of systems.

Acknowledgments

We thank Max Gottesman and Timothy LaRocca for helpful discussions.

Author Contributions

Conceived and designed the experiments: AJR FEA DP TMR ASP MMS. Performed the experiments: FEA DP TMR RK. Analyzed the data: FEA MMS AJR. Contributed reagents/materials/analysis tools: FEA ASP MMS AJR. Wrote the paper: FEA MMS AJR.

References

- Shockett PE, Schatz DG. Diverse strategies for tetracycline-regulated inducible gene expression. *Proc Natl Acad Sci USA*. 1996; 93:5173–5176. PMID: [8643548](#)
- Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet*. 2008; 42:287–299. doi: [10.1146/annurev.genet.42.110807.091442](#) PMID: [18983258](#)
- Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985; 2:13–34. PMID: [3916708](#)
- Stoletzki N, Eyre-Walker A. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol*. 2007; 24:374–381. PMID: [17101719](#)
- Sharp PM, Li WH. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987; 15:1281–1295. PMID: [3547335](#)
- Coleman JR, Papamichail D, Skiena S, Fletcher B, Wimmer E, Mueller S. Virus attenuation by genome-scale changes in codon pair bias. *Science*. 2008; 320:1784–1787. doi: [10.1126/science.1155761](#) PMID: [18583614](#)
- Coleman JR, Papamichail D, Yano M, Garcia-Suarez MDM, Pirofski L-A. Designed reduction of *Streptococcus pneumoniae* pathogenicity via synthetic changes in virulence factor codon-pair bias. *J Infect Dis*. 2011; 203:1264–1273. doi: [10.1093/infdis/jir010](#) PMID: [21343143](#)
- Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*. 2009; 324:255–258. doi: [10.1126/science.1170160](#) PMID: [19359587](#)
- Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. *Science*. 2013; 342:475–479. doi: [10.1126/science.1241934](#) PMID: [24072823](#)
- Sung CK, Li H, Claverly JP, Morrison DA. An rpsL cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl Environ Microbiol*. 2001; 67:5190–5196. PMID: [11679344](#)
- Berry AM, Yother J, Briles DE, Hansman D, Paton JC. Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infect Immun*. 1989; 57:2037–2042. PMID: [2731982](#)
- Los FCO, Randis TM, Aroian RV, Ratner AJ. Role of Pore-Forming Toxins in Bacterial Infectious Diseases. *Microbiol Mol Biol Rev*. 2013; 77:173–207. doi: [10.1128/MMBR.00052-12](#) PMID: [23699254](#)
- Markham NR, Zuker M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res*. 2005; 33:W577–581. PMID: [15980540](#)
- Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct*. 2008; 3:38. doi: [10.1186/1745-6150-3-38](#) PMID: [18796141](#)
- Lacks S, Hotchkiss RD. A study of the genetic material determining an enzyme in *Pneumococcus*. *Biochim Biophys Acta*. 1960; 39:508–518. PMID: [14413322](#)
- Rampersaud R, Planet PJ, Randis TM, Kulkarni R, Aguilar JL, Lehrer RI, et al. Inerolysin, a Cholesterol-Dependent Cytolysin Produced by *Lactobacillus iners*. *J Bacteriol*. 2011; 193:1034–1041. doi: [10.1128/JB.00694-10](#) PMID: [21169489](#)
- Crook J, Tharpe JA, Johnson SE, Williams DB, Stinson AR, Facklam RR, et al. Immunoreactivity of five monoclonal antibodies against the 37-kilodalton common cell wall protein (PsaA) of *Streptococcus pneumoniae*. *Clin Diagn Lab Immunol*. 1998; 5:205–210. PMID: [9521144](#)
- Johansson N, Kalin M, Giske CG, Hedlund J. Quantitative detection of *Streptococcus pneumoniae* from sputum samples with real-time quantitative polymerase chain reaction for etiologic diagnosis of community-acquired pneumonia. *Diagn Microbiol Infect Dis*. 2008; 60:255–261. PMID: [18036762](#)
- Gelber SE, Aguilar JL, Lewis KLT, Ratner AJ. Functional and phylogenetic characterization of Vaginolysin, the human-specific cytolysin from *Gardnerella vaginalis*. *J Bacteriol*. 2008; 190:3896–3903. doi: [10.1128/JB.01965-07](#) PMID: [18390664](#)

20. Aguilar JL, Kulkarni R, Randis TM, Soman S, Kikuchi A, Yin Y, et al. Phosphatase-dependent regulation of epithelial mitogen-activated protein kinase responses to toxin-induced membrane pores. *PLOS ONE*. 2009; 4:e8076. doi: [10.1371/journal.pone.0008076](https://doi.org/10.1371/journal.pone.0008076) PMID: [19956644](https://pubmed.ncbi.nlm.nih.gov/19956644/)
21. Randis TM, Kulkarni R, Aguilar JL, Ratner AJ. Antibody-based detection and inhibition of vaginolysin, the *Gardnerella vaginalis* cytotoxin. *PLOS ONE*. 2009; 4:e5207. doi: [10.1371/journal.pone.0005207](https://doi.org/10.1371/journal.pone.0005207) PMID: [19370149](https://pubmed.ncbi.nlm.nih.gov/19370149/)
22. Zur H, Tuller T. Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*. *EMBO Rep.*; 2012; 13:272–277. doi: [10.1038/embor.2011.262](https://doi.org/10.1038/embor.2011.262) PMID: [22249164](https://pubmed.ncbi.nlm.nih.gov/22249164/)
23. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-limiting steps in yeast protein translation. *Cell*. 2013; 153:1589–1601. doi: [10.1016/j.cell.2013.05.049](https://doi.org/10.1016/j.cell.2013.05.049) PMID: [23791185](https://pubmed.ncbi.nlm.nih.gov/23791185/)
24. Tuller T, Girshovich Y, Sella Y, Kreimer A, Freilich S, Kupiec M, et al. Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res.* 2011; 39:4743–4755. doi: [10.1093/nar/gkr054](https://doi.org/10.1093/nar/gkr054) PMID: [21343180](https://pubmed.ncbi.nlm.nih.gov/21343180/)
25. Supek F, Šmuc T. On relevance of codon usage to expression of synthetic and natural genes in *Escherichia coli*. *Genetics*. 2010; 185:1129–1134. doi: [10.1534/genetics.110.115477](https://doi.org/10.1534/genetics.110.115477) PMID: [20421604](https://pubmed.ncbi.nlm.nih.gov/20421604/)
26. Lee SF, Li Y-J, Halperin SA. Overcoming codon-usage bias in heterologous protein expression in *Streptococcus gordonii*. *Microbiology*. 2009; 155:3581–3588. doi: [10.1099/mic.0.030064-0](https://doi.org/10.1099/mic.0.030064-0) PMID: [19696103](https://pubmed.ncbi.nlm.nih.gov/19696103/)
27. Tuller T, Waldman YY, Kupiec M, Ruppin E. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA*. 2010; 107:3645–3650. doi: [10.1073/pnas.0909910107](https://doi.org/10.1073/pnas.0909910107) PMID: [20133581](https://pubmed.ncbi.nlm.nih.gov/20133581/)
28. Roller M, Lucic V, Nagy I, Perica T, Vlahovicek K. Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic Acids Res.* 2013; 41:8842–8852. doi: [10.1093/nar/gkt673](https://doi.org/10.1093/nar/gkt673) PMID: [23921637](https://pubmed.ncbi.nlm.nih.gov/23921637/)