**NusA solubility in *escherichia coli*:**

**discordant homologues**

by

Leslie A. Strathern

B.A. (St. Mary’s College of Maryland) 2002

THESIS

Submitted in partial satisfaction of the requirements

for the degree of

MASTER OF SCIENCE

in

BIOMEDICAL SCIENCE

in the

GRADUATE SCHOOL

of

HOOD COLLEGE

March 2018

Accepted:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dominic Esposito, Ph.D. Ann Boyd, Ph.D.

Committee Member Director, Biomedical Science Program

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ricky R. Hirschhorn, Ph.D.

Committee Member

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Craig Laufer, Ph.D. April M. Boulton, Ph.D.

Thesis Adviser Dean of the Graduate School

**DEDICATION**

I dedicate this manuscript to my family. Without all of your loving support and encouragement this thesis work would not have been possible. First, to my daughter, Kaylee Anne Garvey, who has been cheering me on these last two years as I write. You may only be four years old, but your strength, ingenuity and positivity has the power to move mountains. Second, to my parents Jeff and Ann Strathern. You have been there for me my entire life without fail, cheering for me when I succeed and picking me up when I fall. Your unwavering faith in my abilities helped keep my dream alive when I was ready to give up. To my brother Russ Strathern and my sunny sister-in-law Alli Houseworth. You showed me that a good smile and sheer force of will can get you to a better place. This has been a very long and winding road with more detours than I would have liked, but I have learned a great deal along the way and all of you have helped me get here. Thank you!

Lastly, I would like to thank Joe Simoldoni for his tireless support and encouragement this last year. The many catered dinners and words of encouragement made the late nights locked in front of the computer much more tolerable.

**ACKNOWLEDGEMENTS**

I would like to thank Dr. Dominic Esposito of the Protein Expression Lab at the Frederick National Laboratories for Cancer Research for his support as an advisor on this project and for providing materials and laboratory space for the work reported in this manuscript.

Special thanks go to Jennifer Mehalko and Miranda Matthews for technical assistance with cloning and sample preparation. This project had an extensive number of clones to generate and test along the way. Their help is greatly appreciated.

I would like to thank Dr. David Waugh of the Frederick National Laboratories for Cancer Research for the donation of genomic DNA from various bacterial sources that was used for the initial homologue comparison study and included the *Yersinia enterocolitica* DNA.

The structural models used in figure 11 were generated by Dr. Simon Messing. I would like to thank him for his assistance in making this figure and for his insights on the possible interactions of the amino acids in what we refer to as the hydrophobic pocket in the globular N-terminal domain.

Finally, I would like to thank several people for backing me as I petitioned to return to Hood to complete my degree after a significant break. They are Dr. Dominic Esposito, Dr. Craig Laufer, Dr. Ricky Hirschhorn, Dr. Rachel Beyer, and Dr. Ann Boyd. Your faith in me and unwavering support allowed me to finally reach a goal I set out for myself 12 years ago. I could not ask for a more amazing group of scientific mentors.

**STATEMENT OF USE AND COPYRIGHT WAIVER**

I do authorize Hood College to lend this thesis, or reproductions of it, in total or part, at the request of other institutions or individuals for the purpose of scholarly research.

TABLE OF CONTENTS

Page

[ABSTRACT vii](#_Toc511338584)

[LIST OF TABLES viii](#_Toc511338585)

[LIST OF FIGURES ix](#_Toc511338586)

[LIST OF ABBREVIATIONS x](#_Toc511338587)

[INTRODUCTION 1](#_Toc511338588)

[MATERIALS AND METHODS 10](#_Toc511338589)

[General Molecular Biology Techniques 10](#_Toc511338590)

[Bacterial Cell Lines 10](#_Toc511338591)

[Media 11](#_Toc511338592)

[Templates 11](#_Toc511338593)

[Oligonucleotides 12](#_Toc511338594)

[Polymerase Chain Reactions 14](#_Toc511338595)

[Destination Vector Construction 17](#_Toc511338596)

[Gateway Cloning 18](#_Toc511338597)

[Site-Directed Mutagenesis 20](#_Toc511338598)

[Site-Specific Random Mutants 21](#_Toc511338599)

[Protein Expression 23](#_Toc511338600)

[Protein Extraction 23](#_Toc511338601)

[RESULTS 25](#_Toc511338602)

[DISCUSSION 33](#_Toc511338603)

[REFERENCES 39](#_Toc511338604)

ABSTRACT

Solubility is a recurrent challenge when overexpressing non-native proteins in *Escherichia coli* (*E. coli*)for purification. N-utilizing substance A (NusA) is often used as a fusion tag to enhance the solubility of its partner protein. However, NusA is not always successful in rescuing solubility. We sought a bacterial homologue with stronger solubility enhancing properties and discovered that the NusA from a closely related species, *Yersinia enterocolitica (Y. ent)*, was completely insoluble as a fusion partner when expressed in *E. coli*. To investigate the cause of the marked difference in solubility, we made chimeras between the sequences of the two homologues and tested the effects both with and without the burden of a fusion partner. We narrowed the region of interest down to the portion of the N-terminal domain between amino acids 62 and 121. We then used site directed mutagenesis to investigate the seven amino acid substitutions that were present in that region of the *Y. ent* NusA. We identified amino acids 63 and 64 as the pair that were responsible for the difference in solubility. Replacing alanine in position 64 of the *Y. ent* NusA with valine results in a protein that is much more soluble than the wild type *Y. ent*, but is not as soluble as the V63L/A34V double mutant previously mentioned. We then made random mutants at those locations to see if we could improve the solubility. We were unable to significantly improve the solubility enhancement capabilities of NusA beyond that of the native *E. coli* protein, but we discovered that an isoleucine at position 64 in the *E. coli* sequence was equally as soluble as the wildtype *E. coli* protein. We will discuss how the characteristics of the amino acids in both sequences and their location within those sequences may affect solubility.

LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| Table |  | Page |
| 1 | | Summary of oligonucleotides used for PCR and mutagenesis. | 13 |
| 2 | Summary of PCRs including oligos used to amplify wildtype and chimeric NusA from *E. coli* and *Y. ent* for the construction of gateway destination vectors. | 16 |
| 3 | Summary of PCRs used to amplify wildtype and chimeric NusA from *E. coli* and *Y. ent* for the construction of gateway entry clones. | 20 |
| 4 | Summary of site-directed mutagenesis reactions used to create point mutations in both *E. coli* and *Y. ent* wild type NusA entry clones for Gateway Multisite recombination reactions. | 21 |
| 5 | Summary of expression clones generated by Gateway Multisite recombination. | 22 |

LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| Figure |  | Page |
| 1 | Overlap PCR schematic diagram. | 15 |
| 2 | Gateway Destination vector cloning schematic. | 18 |
| 3 | Amino acid sequence alignment of *E. coli* and *Y. ent* NusA showing the 49 discordant residues. | 25 |
| 4 | Partial structural diagram of *T. maritima* NusA showing approximate divisions for the chimeric segments | 26 |
| 5 | Diagram of chimeras made between *E. coli* and *Y. ent*. | 26 |
| 6 | SDS page gel of NusA wild type and chimeric proteins fused to an insoluble protein partner (human GRIM19). | 28 |
| 7 | SDS page gel of wild type, chimeric and point mutations of NusA without a fusion partner. | 30 |
| 8 | SDS page gel of NusA wild type and mutant proteins without a fusion partner. | 31 |
| 9 | SDS page gels of NusA wild type and mutant proteins fused to human GFER. | 32 |
| 10 | Sequence relationship of NusA Homologues. | 34 |
| 11 | Structure model showing proposed fit of valine at position 64 in the *E. coli* NusA sequence. | 36 |

LIST OF ABBREVIATIONS

aa-tRNAs aminoacyl-tRNAs

*C. psychrerythraea Colwellia psychrerythraea*

*E. coli* *Escherichia coli*

ELS GroEL-GroES chaperone complex

IPTG isopropyl-beta-D-thiogalactopyranoside

KJE DnaK-DnaJ-GrpE chaperone team

LB Lysogeny broth

MBP maltose-binding protein

mRNA messenger RNA

NusA N-utilizing substance A

Oligos oligonucleotides

PCR polymerase chain reaction

PTMs post-translational modifications

RNAP RNA polymerase

SDS sodium dodecyl sulfate

TAE tris-acetate ethylenediamine tetraacetic acid

TF trigger factor

tRNA transfer RNA

*T. maritima Termatoga maritima*

*Y. ent Yersinia enterocolitica*

INTRODUCTION

Recombinant protein expression in *Escherichia coli* (*E. coli*) is an established method for low cost and high yield production of proteins for structural and biochemical assays. Despite decades of use, it is often hampered by solubility complications. Each protein of interest comes with its own set of challenges and there appears to be no solution to the solubility problem that works across the board. It is thought that the major contributing factors are the lack of post-translational modifications that the foreign proteins would normally receive (Sahdev et al. 2008) and the difference in the translation rate between *E. coli* and the eukaryotic organisms from which most of the transgenes come ([Widmann and Christen 2000](#_ENREF_52)). Attempts to increase the likelihood that a foreign protein is soluble when expressed in *E. coli* have resulted in many tools being developed over the years, ranging from specialized expression strains and the use of inducible promoters to co-expression of chaperone proteins or the addition of solubility-enhancing fusion tags ([Rosano and Ceccarelli 2014](#_ENREF_37)).

Differences in host biology from that of the donor organism can play a significant role in the yield of soluble recombinant proteins*.*  One issue with *E. coli* as a host organism is the lack of post-translational modifications (PTMs) that would normally occur to a protein produced in a eukaryotic cell. For example, many eukaryotic proteins carry recognition sites for myristoylation, and glycosylation. These modifications can change the tertiary structure of the final product and consequently, might alter its solubility. Eukaryotic N-terminal myristoylation is catalyzed by N-myristoyltransferase which covalently links myristatic acid (a 14-carbon saturated fatty acid) to the N-terminal glycine after removal of the initiator methionine ([Xu et al. 2015](#_ENREF_53)). This addition of the fatty acid allows the protein to interact with lipid membranes and facilitates protein-protein interactions, both of which could contribute to stability and solubility. Eukaryotic glycosylation, in contrast, can begin as the addition of a single sugar moiety onto the hydroxyl oxygen of serine or threonine (*O-*glycosylation) or the amide nitrogen of asparagine residues (*N-*glycosylation) which can then be modified by glycosidases and/or glycosyltransferases into more complex chains ([Chang et al. 2007](#_ENREF_6)). It is estimated that two-thirds of eukaryotic proteins have glycosylation sties ([Nothaft and Szymanski 2010](#_ENREF_33)). Glycosylation has been shown to affect ligand binding and protein-protein interactions which could contribute to stability and solubility ([Chang et al. 2007](#_ENREF_6)). It stands to reason that expressing recombinant proteins in a host that cannot provide the appropriate sugar or lipid modifications could impede solubility. If the goal of the recombinant protein expression is to represent the protein in its most native form with all appropriate PTMs, *E. coli* will be a poor choice as a host. However, if the goal of expression does not require these (and other) PTMs, then simply finding ways to compensate for any possible loss of solubility due to missing PTMs may be sufficient.

Translation dynamics is another challenge that can play a role in protein solubility. The rate of translation in *E. coli* is four to ten times faster than in eukaryotes ([Widmann and Christen 2000](#_ENREF_52)). This increased protein production rate necessitates that recombinant proteins fold faster than would be required in their native environment which increases the chance that misfolding can occur. Differential codon bias is an additional obstacle. The degeneracy of the genetic code creates options for alternative codon usage at many places in a nucleotide sequence without effecting the amino acid sequence. The preference of these synonymous codons has drifted during evolution due to changes in tRNA gene redundancy and increasing genome size ([dos Reis et al. 2004](#_ENREF_13)). Expressing eukaryotic genes in *E. coli* can be hampered by a lack of availability of the aminoacyl-tRNAs (aa-tRNAs) that are common in eukaryotes, but underrepresented in the *E. coli* genome. Eukaryotic genes with excessive rare codon occurrences will have decreased translation rates and more paused ribosomes which can lead to mRNA decay ([Hanson and Coller 2018](#_ENREF_20); [Presnyak et al. 2015](#_ENREF_35)) and/or translational abandonment via the process of trans-translation which also tags the incomplete protein for degradation ([Buchan and Stansfield 2007](#_ENREF_4); [Sundermeier et al. 2008](#_ENREF_49)). One solution to this issue has been to optimize the DNA sequence of the recombinant gene to only contain codons for abundant *E. coli* tRNAs ([Sorensen and Mortensen 2005](#_ENREF_46)). Alternatively, *E. coli* expression strains, such as Rosetta (Novagen), have been introduced that contain a plasmid that provides copies of the underrepresented tRNAs to compensate for the codon bias ([Novy et al. 2001](#_ENREF_34)). Recent research indicates that *E. coli* utilize strategically placed “rare” codons to modulate the translation rate and assist protein folding ([Brule and Grayhack 2017](#_ENREF_3); [Buhr et al. 2016](#_ENREF_5)). There are concerns that altering the tRNA dynamics either with the introduction of “rare” tRNAs or via depletion of the supply due to overexpression of the transgene during recombinant protein production can negatively affect the growth and endogenous gene expression of *E. coli* ([Sogaard and Norholm 2016](#_ENREF_45)).

In addition to solubility challenges, toxicity effects on the host cell due to over-expression of recombinant proteins are another factor affecting soluble protein yield. This can be due to gene specific responses (Mujacic et al. 1999; Nannenga and Baneyx 2012), or the accumulation of inclusion bodies ([Hartley and Kane 1988](#_ENREF_22)). Tightly controlled inducible promoters can circumvent toxicity effects and increase yield through regulation of transcription rates. Using these promoters, expression of the recombinant gene is kept silent until a high cell density can be achieved, at which point the gene of interest is activated and the protein is made. There are many promoter options available for heterologous protein expression ([Rosano and Ceccarelli 2014](#_ENREF_37)). The *lac* promoter along with its derivatives the *lac*UV5 and *tac* promoters have been used extensively for expression of recombinant proteins or cofactors for protein expression ([Kaur et al. 2018](#_ENREF_25); [Rosano and Ceccarelli 2014](#_ENREF_37); [Sahdev et al. 2008](#_ENREF_40); [Sorensen and Mortensen 2005](#_ENREF_46)). The *lac* promoter is induced by the presence of lactose or the non-hydrolysable sugar molecule isopropyl-beta-D-thiogalactopyranoside (IPTG) ([Hannig and Makrides 1998](#_ENREF_19)). The *lac*UV5 promoter is a mutant of *lac* that is active in the presence of glucose ([Silverstone et al. 1970](#_ENREF_44)). These promoters have some basal level of expression which can be a problem when expressing highly toxic proteins. Promoters that have stronger regulation give greater control over expression. Two such examples are the T7 RNA polymerase based pET system ([Moffatt and Studier 1987](#_ENREF_29)) and the arabinose inducible pBAD system ([Guzman et al. 1995](#_ENREF_18)). Both systems require a transactivating factor to be supplied either in the host cell or on the expression vector. The pET system requires that T7 RNA polymerase (RNAP), under the control of a *lac*UV5 promoter, be supplied in the host cell line. Expression is then initiated by IPTG induction of T7 RNAP expression which recognizes the T7 promoter upstream of the gene of interest and initiates transcription. Basal level transcription is kept in check by co-expression of a repressor, the T7 lysozyme, which binds to T7 RNA polymerase and inhibits transcription ([Moffatt and Studier 1987](#_ENREF_29)). Once T7 RNAP expression is induced it overwhelms the amount of available lysozyme allowing it to bind the promoter and initiate translation. The transcriptional regulator AraC, on the other hand, silences gene expression in the absence of arabinose by direct binding to the *araP*BAD promoter sequence and inhibiting transcription of the gene. When arabinose is present, it binds to AraC which then switches function to become a transcriptional activator ([Schleif 2010](#_ENREF_41)). Choosing the right promoter for the job is essential and there are many options available depending on the challenges presented by the protein of interest. For a more extensive discussion of available promoters see Kaur et. al ([2018](#_ENREF_25)).

While there are limitations to what *E. coli* can do as far as replicating eukaryotic PTMs and translation dynamics, there are modifications that can be made to either the host strain or the expression construct that can positively influence solubility and yield. The BL21(DE3) strain of *E. coli* (New England Biolabs, Beverley, MA) are *lon* and *ompT* protease deficient ([Daegelen et al. 2009](#_ENREF_8)). The absence of these proteases decreases (but does not eliminate) the risk of proteolysis of the recombinant protein and thus increases potential yield ([Ryan and Henehan 2013](#_ENREF_39)). This strain also contains the λDE3 prophage which expresses the T7 RNA polymerase that is necessary for the pET expression system discussed above ([Studier and Moffatt 1986](#_ENREF_48)). The BL21 star (DE3) line (Invitrogen) has a mutation in the RNaseE gene (*rne*131) that increases mRNA stability by reducing the level of endogenous RNases which are responsible for mRNA degradation. Co-expression of chaperone proteins is another method for modifying strains to improve the yield of soluble recombinant proteins. Chaperones reversibly bind to hydrophobic residues of improperly folded proteins and provide a protected microenvironment in which those proteins can refold ([Mamipour et al. 2017](#_ENREF_28)). The same level of complexity is true with chaperones as with all other tools in the solubility enhancement box, there are many choices of chaperones and there is no universal answer ([de Marco and De Marco 2004](#_ENREF_11); [de Marco et al. 2007](#_ENREF_12)). One major chaperone found in *E. coli* is the ribosome-associated trigger factor (TF). Overexpression of TF decreases the amount of recombinant protein aggregation and in combination with other overexpressed chaperones can improve solubility ([Nishihara et al. 2000](#_ENREF_32)). In addition to TF, there are two teams of chaperones that are used in *E. coli* recombinant protein expression; DnaK-DnaJ-GrpE (KJE) and GroEL-GroES (ELS) ([Nishihara et al. 1998](#_ENREF_31)). DnaK and GroEL function as foldases in that they hydrolyze ATP to change conformation and assist in the refolding of misfolding proteins. Other chaperone proteins like DnaJ function as holdases which bind to misfolded proteins and “hold” them until they can be refolded, preventing them from forming aggregates ([Baneyx and Mujacic 2004](#_ENREF_1); [Kaur et al. 2018](#_ENREF_25)). The mechanics of both chaperone teams have been reviewed nicely several times ([Baneyx and Mujacic 2004](#_ENREF_1); [de Marco 2011](#_ENREF_10); [Hartl and Hayer-Hartl 2002](#_ENREF_21)). DnaJ recruits DnaK to misfolded protein sites and triggers ATP hydrolysis which binds DnaK to the target protein. The reaction is reversed by GrpE mediated ADP/ATP exchange which releases DnaK from the target protein. This process can trigger refolding into a more stable confirmation or the protein can be passed to the ELS team for further work. The ELS complex creates a folding chamber inside which a misfolded protein can attempt to find more stable conformations while shielded from cytoplasmic interactions. The chamber is made up of two multimeric GroEL rings capped on either side by the cochaperone GroES. ATP hydrolysis cycles one end of the chamber open to test new conformations and then closed to reinitiate refolding if necessary. All three chaperone sets (TF, KJE and ELS) can be expressed simultaneously from 5 different plasmids that are available from Takara Bio (Mountain View, CA).

When modifications to the host and expression construct are insufficient to produce soluble product, the next step is to modify the protein itself. Fusing a highly soluble partner protein to the N-terminal or C-terminal end of the protein of interest can sometimes cause that protein to remain soluble long enough to be extracted from the cell and purified ([Gopal and Kumar 2013](#_ENREF_17)). The hypotheses on how solubility tags function vary widely and it is possible that a given tag can have more than one mode of action and function differently in different fusion situations. The proposed explanations fall into two broad categories; blocking formation of insoluble aggregates or altering the folding of the passenger protein itself. The former model suggests that the simple proximity of the large, hydrophilic protein tag is sufficient to keep the passenger protein from forming insoluble aggregates via “electrostatic shielding” ([Zhang et al. 2004](#_ENREF_55)). In this case, the polar surface of the protein tag, and possibly key charged residues, are sufficient to hide any nonpolar amino acids that are left exposed by improper folding of the passenger protein and this in turn prevents the hydrophobic protein-protein interactions that result in aggregate formation. Another suggestion is that a group of passenger proteins might be encapsulated inside a sphere of soluble tags in mini-aggregates much like a micelle (Nomine et al. 2001). The alternative model proposes that a large, N-terminal, properly folded tag seeds the folding process for the passenger protein and acts as an anchor that limits the number of possible confirmations the protein can take and thus increases the folding rate ([Kapust and Waugh 1999](#_ENREF_24)). Another suggestion is that the large N-terminal tag slows the folding of the passenger protein, possibly functioning as a cis-acting molecular chaperone, so that it is more likely to fold into a more stable conformation ([Fox et al. 2001](#_ENREF_15); [Li et al. 2013](#_ENREF_27)).

Several solubility-enhancing fusion partners have been identified and used successfully in protein production (Esposito and Chatterjee 2006; Rosano and Ceccarelli 2014). Among the most successful are N-utilizing substance A (NusA) and maltose-binding protein (MBP) from *E. coli* ([Davis et al. 1999](#_ENREF_9); [Kapust and Waugh 1999](#_ENREF_24)). N-utilizing substance A is an essential transcription elongation factor with an elongated structure that is arranged into discrete domains that are connected by a flexible linker ([Qayyum and Dey 2016](#_ENREF_36); [Yang and Lewis 2010](#_ENREF_54)). The crystal structure has been solved for the homologues from *Thermatoga maritima* and *Mycobacterium tuberculosis,* but a complete structure for the *E. coli* protein does not exist ([Qayyum and Dey 2016](#_ENREF_36); [Shin et al. 2003](#_ENREF_43)). Much is known about the many functions of NusA in transcriptional regulation ([Mondal et al. 2017](#_ENREF_30); [Yang and Lewis 2010](#_ENREF_54)), but less is known about its ability to function as a solubility enhancer. The working hypothesis has been that electrostatic shielding responsible for the effect that NusA has on its passenger proteins, however more recently is has been suggested that NusA can form oligomers that may have chaperone abilities ([Li et al. 2013](#_ENREF_27)). In our hands, we have seen that NusA can solubilize some proteins that other high performing solubility tags like MBP cannot. Sometimes the solubility enhancement is only partial leaving a good deal of the target protein in insoluble aggregates.

Maltose-binding protein has been widely used as a solubility enhancer and as such has been studied in greater detail than NusA ([Fox et al. 2001](#_ENREF_15); [Walker et al. 2010](#_ENREF_50)). MBP is 44 kDa in size making it slightly smaller than the 57 kDa NusA protein. It has the inherent ability to bind amylose which makes it useful as an affinity tag as well as a solubility enhancer. Like NusA, the primary method of solubility enhancement appears to be electrostatic shielding, but it has been suggested that the protein acts as a molecular chaperone as well (Fox et al. 2001). A comparison of homologues of MBP from other bacteria revealed solubility differences when expressed as fusions in *E. coli* ([Fox et al. 2003](#_ENREF_16)). The study also showed that the homologue from *Pyrococus furiosus* may be a better fusion partner for solubility enhancement.

We set out to test homologues of NusA with the goal of finding a better solubility enhancer. Unlike the previously cited MBP work, we were unsuccessful in finding a NusA homologue that was a better solubility partner than *E. coli* NusA (unpublished data). However, we discovered that even though the amino acid sequence is greater than 90% identical to that of *E. coli,* the NusA from *Yersinia enterocolitica* (*Y. ent*)was nearly insoluble when expressed in *E. coli*. Interrogating the differences between homologues of NusA with high sequence identity and extremely discordant solubility may shed some light on how NusA functions as a solubility enhancer and possibly lead to opportunities for improvement.

MATERIALS AND METHODS

**General Molecular Biology Techniques**

All DNA extractions from *E. coli* cultures were performed using the FastPlasmid DNA Kit (Eppendorf, Hamburg, Germany) following the manufacturer specified protocol. Polymerase Chain Reaction (PCR) purifications were performed using the QIAquick® PCR Purification Kit (QIAGEN Inc., Valencia, CA) and gel extractions were purified using the MinElute Gel Extraction Kit (QIAGEN Inc., Valencia, CA).

Plasmid DNA and PCR products were size-verified by agarose gel electrophoresis on 0.8 % agarose gels with TAE buffer with ethidium bromide (Embi Tech, San Diego, CA). Linear DNA was size-verified according to ReadyLoad 1 kb Plus DNA Ladder (Invitrogen, Carlsbad, CA). Plasmid DNA was size-verified according to Supercoiled DNA Ladder (Invitrogen, Carlsbad, CA).

Protein samples were analyzed on 4-15% gradient Criterion Tris-HCl SDS-Page polyacrylamide gels (Bio-Rad, Hercules, CA) with Tris-Glycine SDS running buffer (Bio-Rad, Hercules, CA). Unstained BenchMark Protein Ladder (Invitrogen Corporation, Carlsbad, CA) was used as a size standard on all gels. Gels were stained with an in-house protein stain ([Esposito et al. 2009](#_ENREF_14)).

**Bacterial Cell Lines**

*E. coli* *ccdB* Survival competent cells (Invitrogen, Carlsbad, CA) are required for cloning of destination vectors which contain the causes cell death B gene (*ccdB* gene) ([Bernard and Couturier 1992](#_ENREF_2)). *E. coli* DH10B chemically competent cells (Invitrogen Corporation, Carlsbad, CA) were used for all other cloning purposes as well as site directed mutagenesis.

Our expression cell line is *E. coli* DE09s. These are a modified version of the commercially available BL21 Star (DE3) chemically competent *E. coli* strain (Invitrogen, Carlsbad, CA). These cells also contain the pRare plasmid which confers the ability to translate non-native DNA sequences by providing tRNA sequences for codons that are underrepresented in the *E. coli* genome and therefore rare in the cell ([Novy et al. 2001](#_ENREF_34)).

**Media**

*E. coli* were grown on Lysogeny Broth (LB) agar plates and colonies were picked into Superior Broth liquid medium (Athena ES, Baltimore, MD) and grown overnight at 37 ˚C. Appropriate antibiotics were used in plates and liquid media; spectinomycin (50 μg/mL, Sigma-Aldrich, St. Louis, MO) for Entry clones, carbenicillin (50 μg/mL, Sigma-Aldrich, St. Louis, MO) for Expression clones, and a combination of ampicillin (100μg/mL, Sigma-Aldrich, St. Louis, MO) and chloramphenicol (15 μg/mL, Sigma-Aldrich, St. Louis, MO) for expression strains.

Prior to expression testing, all expression strains were grown in a non-inducing minimal media, MDAG ([Studier 2005](#_ENREF_47)) in the presence of ampicillin and chloramphenicol at 37 ˚C overnight.

**Templates**

Genomic DNA from *Y. ent* was generously provided by Dr. David Waugh (National Cancer Institute). The Gateway destination vector pDest-544 (Addgene, Cambridge, MA. plasmid # 11519) containing *E. coli* NusA was used as a template for PCRs as well as a control vector for comparison during expression testing. Once the full-length wild type *Y. ent* NusA destination vector was generated (pDest-849), it was used as the template to make the chimeric destination vectors for the first round of screening. These destination vectors were then used as templates to make a second round of chimeras as entry clones to be used in Multisite Gateway Reactions.

**Oligonucleotides**

Oligonucleotides (oligos) were ordered from Eurofins MWG Operon (Huntsville, AL). The 5’ and 3’ most oligos for each NusA construct intended to make the destination vectors had BamHI restriction sites added to the 5’ end of the oligo for restriction enzyme cloning into pDest-590 digested with BglII. Oligos for entry clone construction had an attB4 site added to the 5’ end oligo and an attB1r site added to the 3’end oligo for Gateway multisite recombination. Table 1 shows the oligo sequences for all oligos.

Table 1: Summary of oligonucleotides used for PCR and mutagenesis.

| **Oligo ID** | **Oligo Name** | **Oligo Sequence** |
| --- | --- | --- |
| 6616 | Eco-Nus-N-ter | GCATCTATGGGATCCCATCACCATCACCAT |
| 6617 | Eco-Nus-N-ter-rev | ACGCACTTTCTGCACGATAACCTGTTTTGC |
| 6618 | Yen-Nus-MidC-ter | TTATCGTGCAGAAAGTGCGTGAAGCTGAGCGTGCTATGGT |
| 6619 | Yen-M-C-rev | TGACTGGTGGGATCCGCGTTATCGCCAAACCAACAGATAT |
| 6620 | Eco-N-M-C-rev | ACGGAACAGTTCGATCAGCATTTCCGGCTT |
| 6621 | Yen-C-term | TGCTGATCGAGTTGTTCCGCATTGAAGTGCCAGAAATCGG |
| 6622 | Yen-N-term | TGCTGCAATGGGATCCCACCATCACCATCACCATAACAAA |
| 6623 | Yen-N-rev | ACGTACTTTTTGTACGATAACTTGCTTGGC |
| 6624 | Eco-m-c | TTATCGTACAAAAAGTACGTGAAGCCGAACGTGCGATGGT |
| 6625 | Eco-m-c-r | TGACTGGTGGGATCCCCAGTCGCTTCGTCA |
| 6626 | Yen-mid-rev | GCGGAACAACTCGATCAGCATTTCAGGACG |
| 6627 | Eco-C-term | TGCTGATCGAGTTGTTCCGCATTGAAGTGCCAGAAATCGG |
| 6984 | Yent-NusA-C | TGACTGGTGGGATCCCGCGTTATCGCCAAACCAACAGATA |
| 6985 | Ecol-NusA-C | TGACTGGTGGGATCCAGTCGCTTCGTCACC |
| 7552 | YeNusA1-60 | ACGACGGAAGGTGTCGAAATCACCGGTTTT |
| 7553 | EcNusA61-e | ATTTCGACACCTTCCGTCGTTGGTTAGTTGTTGATGAAGT |
| 7554 | EcNusA1-60 | GCGACGGAAAGTGTCAAAATCACCGCTTTT |
| 7555 | Ye61-120 | ATTTTGACACTTTCCGTCGCTGGGTTGCCGTTGACGAAGT |
| 7556 | Ye61-120r | AGCTTCACGTACTTTTTGTACGATAACTTG |
| 7557 | Ec.122-e | TACAAAAAGTACGTGAAGCTGAACGTGCGATGGTGGTTGA |
| 8405 | 5'Ec-NusA | AACAAAGAAATTTTGGCTGTAGTTGAAGCC |
| 8407 | B4-RBS | GTTTAACTTTAAGAAGGAGATATACAT |
| 8406 | RBS-H6-EcNus | GTTTAACTTTAAGAAGGAGATATACATATGAGATCCCACCATCACCATCACCATAACAAAGAAATTTTGGCTGT |
| 8408 | B1r-EcNusA | TGATCTAGTCGCTTCGTCACCGAACCAGCA |
| 8409 | B1r-YeNusA | GATATCAGATCCCGCGTTATCGCCAAACCA |
| 8679 | 7000-E05F | GGTGATTTCGACACCTTCCGTCGTTGGTTAGTTGTTGACGAAGTCACGATGCCAACCCGC |
| 8680 | 7000-E05R | GCGGGTTGGCATCGTGACTTCGTCAACAACTAACCAACGACGGAAGGTGTCGAAATCACC |
| 8681 | 7000-E06F | CGTTGGGTTGCCGTTGACGAAGTCACGCAGCCAACCAAGGAAATTACGTTAGAAGCGGCTCAATAC |
| 8682 | 7000-E06R | GTATTGAGCCGCTTCTAACGTAATTTCCTTGGTTGGCTGCGTGACTTCGTCAACGGCAACCCAACG |
| 8683 | 7000-E07F | CCCGCGAAATTACGTTAGAAGCGGCTCGTTACGAAGATCCTTCGCTCCAGTTGGGC |
| 8684 | 7000-E07R | GCCCAACTGGAGCGAAGGATCTTCGTAACGAGCCGCTTCTAACGTAATTTCGCGGG |
| 8685 | 7000-E08F | CGTTAGAAGCGGCTCAATACGAAGATGAATCGCTCAACTTGGGCGATTATGTAGAAGATCAGATTG |
| 8686 | 7000-E08R | CAATCTGATCTTCTACATAATCGCCCAAGTTGAGCGATTCATCTTCGTATTGAGCCGCTTCTAACG |
| 8996 | YeNusA-F | AACAAAGAGATTCTGGCTGTTGTAGAAGCA |
| 8997 | YeNusA-R | CGCGTTATCGCCAAACCAACAGATATTAC |
| 9283 | E09Fwd | TGATTTTGACACTTTCCGTCGCTGGGTTGCCGTTGATGAAGTCACCCAGCCGACCA |
| 9284 | E09Rev | TGGTCGGCTGGGTGACTTCATCAACGGCAACCCAGCGACGGAAAGTGTCAAAATCA |
| 9285 | E10Fwd | TGATTTCGACACCTTCCGTCGTTGGTTAGCCGTTGACGAAGTCACGATGCCAA |
| 9286 | E10Rev | TTGGCATCGTGACTTCGTCAACGGCTAACCAACGACGGAAGGTGTCGAAATCA |
| 9287 | E11Fwd | TTTCGACACCTTCCGTCGTTGGGTTGTTGTTGACGAAGTCACGATGCCAACCC |
| 9288 | E11Rev | GGGTTGGCATCGTGACTTCGTCAACAACAACCCAACGACGGAAGGTGTCGAAA |
| 9289 | Y. ent NusA\_BF | GACGAAGTCACGATGCCAACCCGCGAAATTACGTTAGAAGCGGCTCA |
| 9290 | Y. ent NusA\_63 | TTCGCGGGTTGGCATCGTGACTTCGTCAACGGCVNNCCAACGACGGAAGGTGTCGAAATC |
| 9291 | Y. ent NusA\_64 | TTCGCGGGTTGGCATCGTGACTTCGTCAACVNNAACCCAACGACGGAAGGTGTCGAAATC |
| 9636 | E. coli NusA\_BF | GATGAAGTCACCCAGCCGACCAAGGAAATCACCCTTGAAG |
| 9637 | E. coli NusA\_64 | TGGTCGGCTGGGTGACTTCATCAACvnnTAACCAGCGACGGAAAGTGTCAAAA |
| 10105 | NusA-Ec79Fwd | GCGGTGATTTTGACACTTTCCGTCGCTGGTTATTAGTTGATGAAGTCACCCAGCCGACCAAGGAAATCACCC |
| 10106 | NusA-Ec79Rev | GGGTGATTTCCTTGGTCGGCTGGGTGACTTCATCAACTAATAACCAGCGACGGAAAGTGTCAAAATCACCGC |
| 10107 | NusA-Ye80Fwd | CCGGTGATTTCGACACCTTCCGTCGTTGGGTTCTTGTTGACGAAGTCACGATGCCAACCCGCGAAATTACGT |
| 10108 | NusA-Ye80Rev | ACGTAATTTCGCGGGTTGGCATCGTGACTTCGTCAACAAGAACCCAACGACGGAAGGTGTCGAAATCACCGG |
| 10119 | EcNusA-V64AF | GCGGTGATTTTGACACTTTCCGTCGCTGGTTAGCCGTTGATGAAGTCACCCAGCCGACCAAGGAAATCACCC |
| 10120 | EcNusA-V64AR | GGGTGATTTCCTTGGTCGGCTGGGTGACTTCATCAACGGCTAACCAGCGACGGAAAGTGTCAAAATCACCGC |

**Polymerase Chain Reactions**

All PCRs were performed using 2× Phusion High Fidelity Master Mix (New England Biolabs, Beverley, MA) and following the manufacturer’s recommended thermal cycler conditions. Chimeras were made by overlap PCR where two separate PCRs were performed to amplify a portion of the different templates and then a third PCR was used to combine the two segments. For example (see Figure 1), the EYY vector was made with the *E. coli* forward oligo (6616) and the *E. coli*-overlap reverse oligo (6617) in the first reaction to amplify the N-terminal region of the gene, the overlap-*Y. ent* forward (6618) and the *Y. ent* 3’ reverse (6984) oligos in the second reaction for the middle and C-terminal region of the gene. These products were size verified on an agarose gel, PCR purified and then DPNI (New England Biolabs, Beverley, MA) digested to remove any residual template. Then the *E. coli* forward oligo (6616) and the *Y. ent* 3’ reverse oligo (6984) were used to combine the two segments in the third (overlap) PCR. Table 2 contains a list of the oligos used to construct the chimeras at each stage. PCR products were verified by size on agarose gels and purified prior to restriction digest.

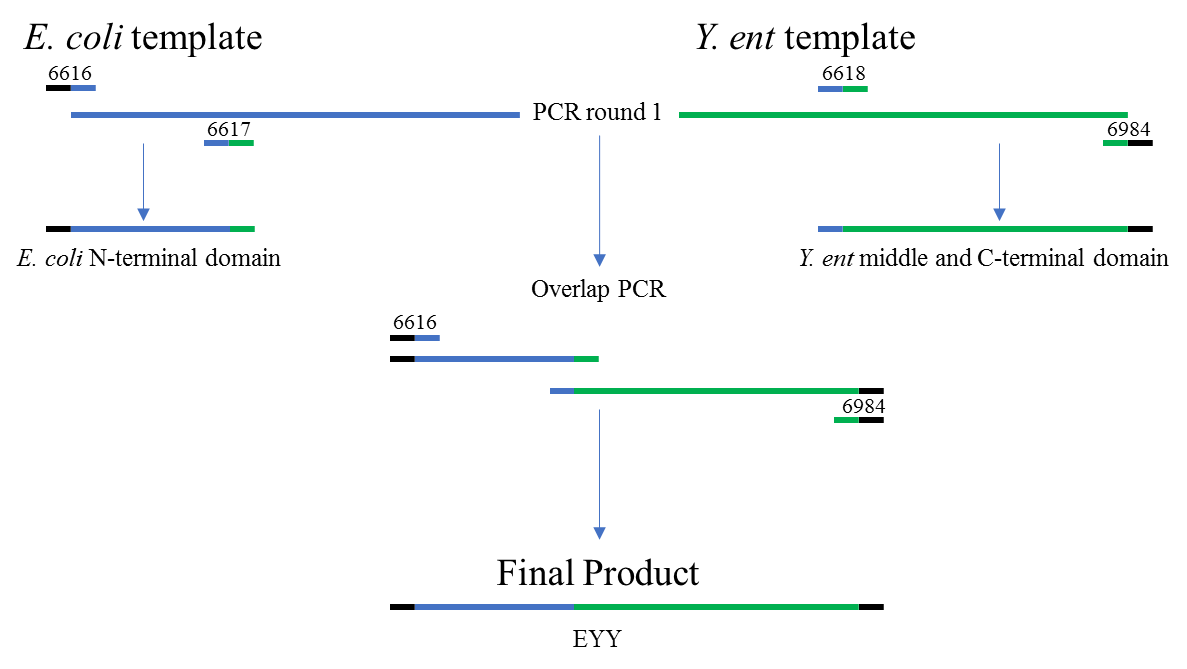


Figure 1: Overlap PCR schematic diagram. Chimeric NusA genes were generated using overlap PCR. This example is for the construction of the chimera with the *E. coli* N-terminal domain and the *Y. ent* middle and C-terminal domains (denoted EYY). *E. coli* sequence is represented by the blue line and the *Y. ent* sequence is represented in green. Oligo numbers are listed above the short lines for reference in Table 1. Oligos 6616 and 6984 have BamHI restriction sites added to the 5’ end (represented in black) for use in cloning into the expression vector. Oligos 6617 and 6618 were designed to have partial complementary sequences so that the products from round one would have regions that would allow for annealing and elongation during the second round of PCR to generate the chimeric construct.

Table 2. Summary of PCRs including oligos used to amplify wildtype and chimeric NusA from *E. coli* and *Y. ent* for the construction of gateway destination vectors.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PCR ID** | **Details** | **Template** | **Forward Oligo** | **Reverse Oligo** | **pDest** |
| 7000-P12 | Y. ent NusA | Y. ent | 6026 | 6035 | 849 |
| 7000-P13 | E. coli NusA N-term | pDest-544 | 6616 | 6617 |  |
| 7000-P14 | Y. ent NusA Mid-C | pDest-849 | 6618 | 6984 |  |
| 7000-P15 | E. coli N-term/ Y. ent Mid-C (EYY) | P13+P14 | 6616 | 6984 | 851 |
| 7000-P16 | E. coli N-Mid | pDest-544 | 6616 | 6620 |  |
| 7000-P17 | Y. ent C-term | pDest-849 | 6621 | 6984 |  |
| 7000-P18 | E. coli N-mid / Y. ent C-term (EEY) | P16+P17 | 6616 | 6984 | 852 |
| 7000-P19 | Y. ent N-term | pDest-849 | 6622 | 6623 |  |
| 7000-P20 | E. coli Mid-C | pDest-544 | 6624 | 6985 |  |
| 7000-P21 | Y. ent N-term / E. coli Mid-C (YEE) | P19+P20 | 6622 | 6985 | 853 |
| 7000-P22 | Y. ent N-Mid | pDest-849 | 6622 | 6626 |  |
| 7000-P23 | E. coli C-term | pDest-544 | 6627 | 6985 |  |
| 7000-P24 | Y. ent N-Mid / E. coli C-term (YYE) | P22+P23 | 6622 | 6985 | 854 |
| 7000-P25 | Y. ent Mid | pDest-849 | 6618 | 6626 |  |
| 7000-P26 | E. coli N-term / Y. ent Mid / E. coli C-term (EYE) | P13+P23+P25 | 6616 | 6985 | 855 |
| 7000-P27 | E. coli Mid | pDest-544 | 6624 | 6620 |  |
| 7000-P28 | Y. ent N-term / E. coli Mid / Y. ent C-term (YEY) | P19+P17+P27 | 6622 | 6984 | 856 |
| 7000-P29 | Y. ent 1-61 | pDest-849 | 6622 | 7552 |  |
| 7000-P30 | E. coli 62-e | pDest-544 | 7553 | 6985 |  |
| 7000-P31 | NusA(yeEE) | P29+P30 | 6622 | 6985 | 870 |
| 7000-P32 | E. coli 1-61 | pDest-544 | 6616 | 7554 |  |
| 7000-P33 | Y. ent 62-121 | pDest-849 | 7555 | 7556 |  |
| 7000-P34 | E. coli 122-end | pDest-544 | 7557 | 6985 |  |
| 7000-P35 | NusA (eyEE) | P32+P33+P34 | 6616 | 6985 | 871 |
| 7000-P36 | NusA (yeYY) | P36A+P36B | 8113 | 8114 | 872 |
| 7000-P36A | Y. ent 1-61/E. coli 62-121 | pDest-870 | 8111 | 8110 |  |
| 7000-P36B | Y. ent 122-end | pDest-851 | 8109 | 8112 |  |

**Destination Vector Construction**

Gateway destination vectors containing the chimeric NusA gene sequences were generated by restriction enzyme cloning into pDest-590 ([Chatterjee and Esposito 2006](#_ENREF_7)) to make Gateway destination vectors. The genes were under the control of a T7 promoter which is induced via IPTG induction of T7 RNA polymerase under the *lac*UV5 promoter in the BL21 star DE3 *E. coli* expression strain. The PCR products were digested with BamHI (New England Biolabs, Beverley, MA) post purification and ligated to pDest-590 digested with BglII (New England Biolabs, Beverley, MA) using T4 QuickLigase (New England Biolabs, Beverley, MA) and following manufacturer recommended procedure. Figure 2 shows the cloning strategy for the destination vectors. *E. coli* *ccdB* Survival cells were transformed and grown on LB agar plates with ampicillin and chloramphenicol for selection. The destination vectors were sequence verified for directionality and sequence accuracy.

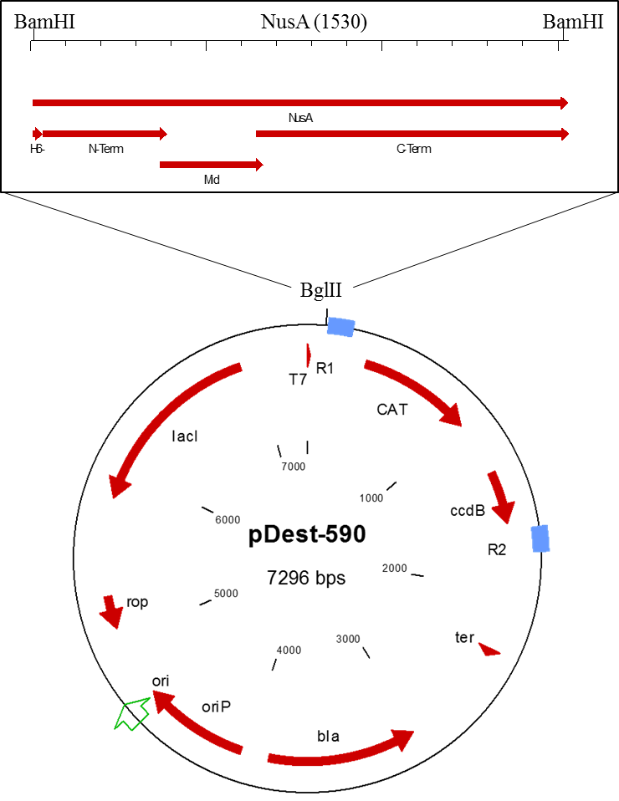


Figure 2: Gateway Destination vector cloning schematic. This diagram shows the insertion of the BamHI digested NusA containing PCR products into the BglII cleaved pDest-590 backbone vector. The light blue boxes represent the gateway attB recombination sites. The red arrows show direction of transcription of genes on the plasmid. The boxed diagram above depicts the sectioning of the gene for the chimeric clones. See Figure 1 for details on overlap PCR.

**Gateway Cloning**

All proteins expressed in this project were made using Gateway cloning methods ([Hartley et al. 2000](#_ENREF_23)). For a more detailed description of the methods used here please see Esposito, Garvey and Chakiath (2009) and Wall et al.(2014). The destination vectors containing the different NusA chimeras were used to make a panel of expression clones containing the NusA gene fused to proteins that have a history of solubility problems when expressed in *E. coli* (data not published). These expression clones were made by Gateway LR reaction utilizing LR Clonase II enzyme mix (Invitrogen, Carlsbad, CA) and following the standard protocol.

The second set of chimeras were generated as entry clones for use in Multisite Gateway cloning ([Wall et al. 2014](#_ENREF_51)). PCR reactions were performed using the previous destination vectors as a template with a forward oligo containing the attB4 sequence and a reverse oligo containing the attB1r sequence for gateway cloning. Table 1 shows the oligos used for these PCRs. Following size verification and PCR purification, these products were then Gateway cloned into pDonr-233 utilizing the BP Clonase II Kit (Invitrogen, Carlsbad, CA) and following the standard protocol. pDonr-233 is a Gateway donor vector made from pDonr-223 ([Rual et al. 2004](#_ENREF_38)) with the attP4-attP1r cassette. These entry clones were sequence verified for accuracy and then used to make fusion proteins via Gateway multisite LR reactions into pDest-880. Table 3 shows the entry clone combinations used to make the expression clones.

Table 3. Summary of PCRs used to amplify wildtype and chimeric NusA from *E. coli* and *Y. ent* for the construction of gateway entry clones. PCRs P42, P43 and P44 utilize degenerate oligos which create amino acid substitutions at positions 63 or 64. The resulting entry clones can be found in the multisite table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PCR ID** | **Details** | **Template** | **Forward Oligo** | **Reverse Oligo** | **Entry Clone** |
| 7000-P37 | B4-EcNusA-B1r | 7000-P37a | 8406 | 8408 | 7000-E03 |
| 7000-P37a | E.c.NusA-B1r | pDest-544 | 8405 | 8408 |  |
| 7000-P38 | B4-Y.eNusA-B1r | pDest-849 | 8407 | 8409 | 7000-E04 |
| 7000-P39 | B4-NusA (eyEE)-B1r | 7000-P39a | 8406 | 8408 | 7000-E39 |
| 7000-P39a | NusA (eyEE)-B1r | pDest-871 | 8405 | 8408 |  |
| 7000-P40 | B4-NusA (yeYY)-B1r | pDest-872 | 8407 | 8409 | 7000-E40 |
| 7000-P42 | YeNusA(V63X) | P42a+P42b | 8407 | 8409 | Y. e V63 Deg. |
| 7000-P42a | YeNusA(V63X) OLA | pDest-849 | 8407 | 9290 |  |
| 7000-P42b | YeNusA OLB | pDest-849 | 9289 | 8409 |  |
| 7000-P43 | YeNusA(A64X) | P43a + P42b | 8407 | 8409 | Y. e A64 Deg. |
| 7000-P43a | YeNusA(A64X) OLA | pDest-849 | 8407 | 9291 |  |
| 7000-P44 | EcNusA(V64X) | P44a+P44b | 8407 | 8408 |  |
| 7000-P44a | E.c.NusA | pDest-544 | 8405 | 9637 | E. c V64 Deg |
| 7000-P44b | E.c.NusA | pDest-544 | 9636 | 8408 |  |

**Site-Directed Mutagenesis**

Once the region of interest was determined to be between amino acids 62 and 121 of the NusA N-terminal domain, site-directed mutagenesis was used to determine which of the eight amino acid substitutions could be responsible for the phenotypic change. The wildtype entry clone for *Y. ent* was used as a template and oligos were designed to make point mutations to the *E. coli* sequence (Table 4). Reactions were performed using the QuickChange II Site-Directed Mutagenesis Kit (Agilent, Santa Clara, CA) and following the standard protocol.

Table 4: Summary of site-directed mutagenesis reactions used to create point mutations in both *E. coli* and *Y. ent* wild type NusA entry clones for Gateway Multisite recombination reactions. Template 7000-E04 is the wild-type *Y. ent* sequence and 7000-E03 is the *E. coli* wild-type sequence. To make the V63L/A64V/Q81R triple mutant (7000-E12) the V63L/A64V double mutant (7000-E05) was used as a template instead of one of the wild-type entry clones.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entry Clone** | **Gene Info** | **Template** | **Forward Primer** | **Reverse Primer** |
| 7000-E05 | Y. ent NusA (V63L/A64V) | 7000-E04 | 8679 | 8680 |
| 7000-E06 | Y. ent NusA (M70Q/R73K) | 7000-E04 | 8681 | 8682 |
| 7000-E07 | Y. ent NusA (Q81R) | 7000-E04 | 8683 | 8684 |
| 7000-E08 | Y. ent NusA (P85E/Q88N) | 7000-E04 | 8685 | 8686 |
| 7000-E09 | E. coli NusA (L63V/V64A) | 7000-E03 | 9283 | 9284 |
| 7000-E10 | Y. ent NusA (V63L) | 7000-E04 | 9285 | 9286 |
| 7000-E11 | Y. ent NusA (A64V) | 7000-E04 | 9287 | 9288 |
| 7000-E12 | Y. ent NusA (V63L/A64V/Q81R) | 7000-E05 | 8683 | 8684 |
| 7000-E80 | Y. ent NusA (A64L) | 7000-E04 | 10107 | 10108 |
| 7000-E79 | E. coli NusA (V64L) | 7000-E03 | 10105 | 10106 |
| 7000-E13 | E. coli NusA (V64A) | 7000-E03 | 10119 | 10120 |

**Site-Specific Random Mutants**

Once the amino acid 63 and 64 sites were identified as significant, degenerate oligos were ordered for overlap PCR so that constructs of varying sequences could be made and tested. The reverse oligo for the front half of the overlap PCR was used to introduce the mutations. As such the oligos were designed so that the codon for the amino acid of interest had the sequence VNN, where N can be any nucleotide and V can be A, C, or G but not T. The purpose of this is to avoid the introduction of stop codons (TAA, TGA) in the complement strand. This design allows for all possible amino acid substitutions with only one possible stop codon (TAG). Table 3 shows the PCR design for the overlap pieces. Table 5 shows the resulting entry clones and the multisite expression clones that they became.

Table 5: Summary of expression clones generated by Gateway Multisite recombination. The second entry clone contains a stop codon just after the att1 recombination site yielding a very small C-terminal fusion tag. The pDest-880 destination vector was used for all reactions. Entry clones numbered 7000-E50 to E75 are the result of the degenerate oligo PCR reactions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Expression ID** | **Exp Gene Name** | **Entry Clone1** | **Entry Clone2** |
| 7000-M010-880 | His6-Ec NusA- stop | 7000-E03 | pEL124 |
| 7000-M020-880 | His6-Ye NusA- stop | 7000-E04 | pEL124 |
| 7000-M030-880 | NusA (eyEE) stop | 7000-E39 | pEL124 |
| 7000-M040-880 | NusA (yeYY) stop | 7000-E40 | pEL124 |
| 7000-M050-880 | YeNusA(V63L/A64V) stop | 7000-E05 | pEL124 |
| 7000-M060-880 | YeNusA(M70Q/R73K) stop | 7000-E06 | pEL124 |
| 7000-M070-880 | YeNusA(Q81R) stop | 7000-E07 | pEL124 |
| 7000-M080-880 | YeNusA(P85E/Q88N) stop | 7000-E08 | pEL124 |
| 7000-M090-880 | EcNusA(L63V/V64A) stop | 7000-E09 | pEL124 |
| 7000-M100-880 | YeNusA(V63L) stop | 7000-E10 | pEL124 |
| 7000-M110-880 | YeNusA(A64V) stop | 7000-E11 | pEL124 |
| 7000-M130-880 | EcNusA(V64A) stop | 7000-E13 | pEL124 |
| 7000-M500-880 | YeNusA(V63N/D66N) stop | 7000-E50 | pEL124 |
| 7000-M510-880 | YeNusA(V63I) stop | 7000-E51 | pEL124 |
| 7000-M520-880 | YeNusA(V63R) stop | 7000-E52 | pEL124 |
| 7000-M530-880 | YeNusA(V63A) stop | 7000-E53 | pEL124 |
| 7000-M600-880 | YeNusA(A64T) stop | 7000-E60 | pEL124 |
| 7000-M610-880 | YeNusA(A64I) stop | 7000-E61 | pEL124 |
| 7000-M620-880 | YeNusA(A64N) stop | 7000-E62 | pEL124 |
| 7000-M630-880 | YeNusA(A64T) stop | 7000-E63 | pEL124 |
| 7000-M640-880 | YeNusA(A64F) stop | 7000-E64 | pEL124 |
| 7000-M650-880 | YeNusA(A64S) stop | 7000-E65 | pEL124 |
| 7000-M660-880 | YeNusA(A64Y) stop | 7000-E66 | pEL124 |
| 7000-M670-880 | YeNusA(A64P) stop | 7000-E67 | pEL124 |
| 7000-M680-880 | YeNusA(A64R) stop | 7000-E68 | pEL124 |
| 7000-M690-880 | YeNusA(A64K) stop | 7000-E69 | pEL124 |
| 7000-M700-880 | EcNusA(V64R) stop | 7000-E70 | pEL124 |
| 7000-M710-880 | EcNusA(V64P) stop | 7000-E71 | pEL124 |
| 7000-M720-880 | EcNusA(V64T) stop | 7000-E72 | pEL124 |
| 7000-M730-880 | EcNusA(V64G) stop | 7000-E73 | pEL124 |
| 7000-M740-880 | EcNusA(V64I) stop | 7000-E74 | pEL124 |
| 7000-M750-880 | EcNusA(V64S) stop | 7000-E75 | pEL124 |
| 7000-M790-880 | EcNusA(V64L) stop | 7000-E79 | pEL124 |
| 7000-M800-880 | YeNusA(A64L) stop | 7000-E80 | pEL124 |

**Protein Expression**

Small scale protein expression was used to evaluate solubility of the different NusA fusion tags that were generated at all stages of this project. Sixteen hours in advance of a scheduled test, 2 ml cultures of non-inducing MDAG medium was inoculated and incubated at 37 ˚C while shaking overnight. The next day, 2 wells of a 24-well plate were seeded with 2 μl of overnight culture into 700 μl of Superior Broth and loaded into a BMG Fluostar Omega plate reader. Growth curves were monitored by absorbance at 600 nm and cultures were induced with IPTG when culture density reached 200 milliOD units above background. Cultures were allowed to grow for 2 hours at 37 ˚C or 16 hr at 16 ˚C at which point samples were taken for whole cell lysis and soluble fractionation.

**Protein Extraction**

Whole cell lysates were generated from cell pellets made from 100 μl samples of culture taken at the end of the expression test. Cells were centrifuged at maximum speed in a microcentrifuge to create pellets and supernatant was gently removed. Pellets were resuspended in Benzonase Cracking Buffer (25 mM Tris—HCl pH 8.0, 5 mM MgCl2) and then frozen on dry ice for at least 5 min. Samples were then thawed and 1 μl of Benzonase (EMD Biosciences, Darmstadt, Germany) was added. Samples were incubated at 37 ˚C for 10 minutes after which 7.5 μl of NuPage 4x LDS Sample buffer (Invitrogen, Carlsbad, CA) and 2 μl of TCEP: Tris(2-carboxyethyl)phosphine hydrochloride (Pierce, Rockford, IL) was added.

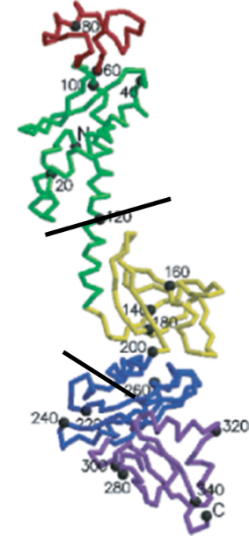
Soluble and insoluble fractions were separated via Ready Prep protocol detailed here. *E. coli* cell pellets from 2 ml growths were prepared as follows. To each wet cell pellet 250 μl buffer A (50 mM Tris–HCl pH 7.5, 0.9% glucose, 0.1% Ready-Lyse lysozyme (non-avian; Epicentre), 0.2% Omni-Cleave nuclease (Epicentre) were added, followed by mixing at 300 rpm on an Eppendorf Thermomixer at room temperature for 15 min. Then 250 μl of buffer B (0.1% sodium deoxycholate, 5 mM MgCl2) were added and mixing was repeated for another 15 min. The samples were centrifuged at 2787 xg for 15 min at 4 °C. Supernatants were carefully removed and insoluble pellets were resuspended in 8 M Urea. Samples were stored on ice or at −80 °C pending gel analysis. Amount of soluble protein was estimated by comparison of total protein to amount retained in the soluble fraction as visible on Coomasie stained polyacrylamide gel.

RESULTS

We used recombinant DNA techniques to compare and contrast the NusA from *Y. enterocolitica* and *E. coli* in order to better understand the marked differences in solubility that exist between the two homologues despite their 90% amino acid sequence identity. Figure 3 shows an alignment of the two sequences with the discordant amino acids highlighted. Utilizing the partial structure from *T. maritima* (Shin et al. 2003) as a reference, we partitioned the NusA protein into sections and created chimeras from the *E. coli* and *Y. ent* sequences to narrow down the region responsible for the solubility differences. Figure 4 shows the structural diagram from *T. maritima* and the sites of the partitions. The chimera scheme is shown in Figure 5.



Figure 3: Amino acid sequence alignment of *E. coli* and *Y. ent* NusA showing the 49 discordant residues. Positions 63 and 64 are circled in red. Alignment performed in Geneious software version 9.1 (<http://www.geneious.com>, Kearse et al., 2012).



N

M

C

Figure 4: Partial structural diagram of *T. maritima* NusA showing approximate divisions for the chimeric segments. N-terminal domain, middle region and C-terminal domain are denoted with N, M and C respectively. Structural diagram is figure 2 from Shin et al. 2003.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Vector Number** | **Construct Code** | **Na** | **Nb** | **Middle** | **C-term** |
| 544 | EEE |  |  |  |  |
| 849 | YYY |  |  |  |  |
| Round 1: | |  |  |  |  |
| 851 | EYY |  |  |  |  |
| 852 | EEY |  |  |  |  |
| 853 | YEE |  |  |  |  |
| 854 | YYE |  |  |  |  |
| 855 | EYE |  |  |  |  |
| 856 | YEY |  |  |  |  |
| Round 2: | |  |  |  |  |
| 870 | yeEE |  |  |  |  |
| 871 | eyEE |  |  |  |  |
| 872 | yeYY |  |  |  |  |
| 873 | eyYY |  |  |  |  |

Figure 5: Diagram of chimeras made between *E. coli* and *Y. ent.* The N-terminal region consists of amino acids 1 to 121. The middle region contains amino acids 122 to 251. The C-terminal region comprises the remainder of the protein (252 to the end). The construct code includes a set of three to 4 letters that represent the source of the sequence in each domain of the chimera: *E. coli* (E) or *Y. ent*. (Y). In round 2, the N-terminal region was divided into two parts (Na; amino acids 1-61 and Nb; amino acids 62-121). Lowercase letters represent the organism from which the “a” and “b” subsegments of the N-terminal domain were derived.

For the first round of experiments we tested the NusA chimeras as a fusion with a set of partner proteins that had previously been determined to be insoluble when expressed in *E. coli.* Note that the predicted isoelectric point of NusA is 4.3, meaning it will be negatively charged at neutral pH and bind more SDS, causing it to migrate more slowly than would be expected based on size alone. We find that NusA by itself migrates at a size of 65 kDa as compared to the predicted size of 57 kDa. When fused to the partner protein GRIM19 for solubility enhancement testing, the fused protein migrates close to 90 kDa as compared to its predicted size of 75 kDa (Figure 6). The *E. coli* wildtype fusion protein was nearly completely soluble in contrast to the wild type *Y. ent* fused to GRIM19, which was completely insoluble. The chimeras that had *E. coli* N-terminal domains enhanced the solubility of their GRIM19 partners. The EYY chimera was more soluble than the EEY chimera. In contrast, both chimeras with the *Y. ent* N-terminal domain were incapable of solubilizing the partner protein. The preliminary results indicated that the N-terminal domain was responsible for the observed solubility difference.

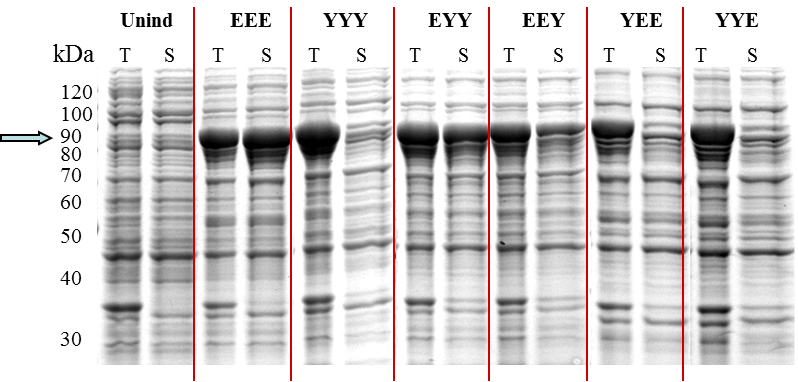


Figure 6: SDS page gel of NusA wild type and chimeric proteins fused to an insoluble protein partner (human GRIM19). This image shows total (T) and soluble (S) fractions from bacterial cell lysates. The first two lanes are uninduced control samples. All other samples are from induced cultures that were grown at 16 ˚C for 16 hr. Sample sets are labeled with three letter codes to denote the three domains of the chimera. The letters “E” for *E. coli* and “Y” for *Y. ent* correspond to the source of the amino acid sequence in that region of the NusA protein. The wild type *E. coli* (EEE) and wild type *Y. ent.* (YYY) serve as controls with which to compare the chimeric proteins. Sizes at the left are approximate based on the location of native *E. coli* proteins in the uninduced samples.

We made a second round of chimeras (see Figure 5) to further divide the N-terminal domain and narrow down the number of amino acid substitutions that could be causing this phenotype. We chose to express the proteins without a fusion partner for this round of testing to remove any possible complications caused by interactions with the partner protein. Figure 7 shows the total and soluble fractions for the most significant constructs. The *E. coli* wildtype NusA (EEE) is very soluble when expressed at 37 ˚C while the *Y. ent* NusA (YYY) is largely insoluble. The eyEE NusA was nearly insoluble while the complementary yeYY NusA was vary soluble. This suggests that the second half of the N-terminal domain carries amino acids that are capable of altering the solubility of these NusA homologues. The yeEE chimera was similar to the *E. coli* wildtype and the eyYY maintained the insoluble nature of the *Y. ent* wildtype (data not shown). There are seven amino acid substitutions in the *Y. ent* sequencethat exist in the second half of the N-terminal domain between amino acids 62 and 121. We made point mutants in the *Y. ent* background to test these amino acid substitutions. Figure 7 also includes the results of this expression test. The Q81R mutant shows a very small amount of soluble protein, which we chose not to pursue further as the level was not highly significant. The V63L/A64V pair of mutants is the only construct that shows appreciable soluble product, indicating that these amino acids are responsible for the majority of the rescue effect we have observed. We made additional mutations at positions 63 and 64 in an attempt to generate a protein with better solubility than the *E. coli* NusA. Table 5 shows all of the substitutions we isolated using a degenerate PCR approach. We were not able to exhaust all possibilities at the two sites. Of the mutations which were screened, the only amino acid substitution that showed a significant positive effect on solubility was isoleucine at position 64. Figure 8 shows the isoleucine results as well as a few representative mutants.

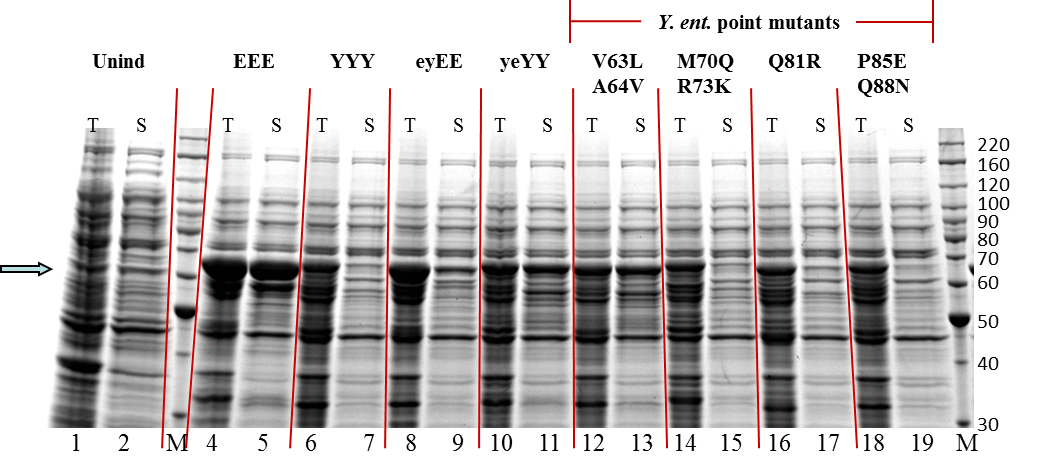


Figure 7: SDS Page gel of wild type, chimeric and point mutations of NusA without a fusion partner. This image shows total (T) and soluble (S) fractions from bacterial cell lysates. Lanes 1 and 2 are uninduced control cells. All other samples are from induced cultures that were grown at 37 ˚C for 2 hr. Sample sets are labeled with E for *E. coli* and Y for *Y. ent*. Three letter codes correspond to the source of the amino acid sequence in that region of the NusA protein. The wild type *E. coli* (EEE) and wild type *Y. ent.* (YYY) serve as controls with which to compare the chimeric proteins. Lower case letters over lanes 8 through 11 indicate the sub-sectioning of the N-terminal domain. The final four sets of samples are the point mutants made by site-directed mutagenesis to the wildtype *Y. ent* sequence.

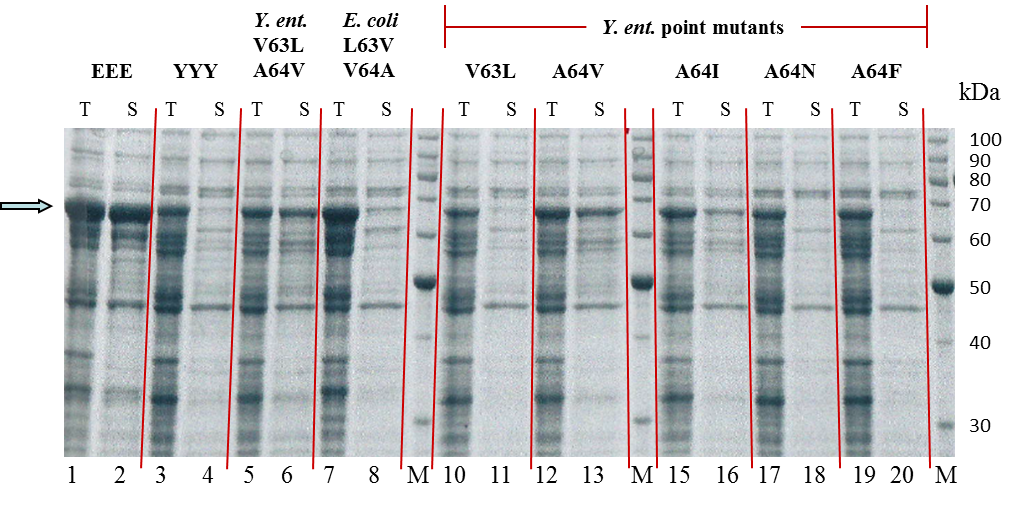
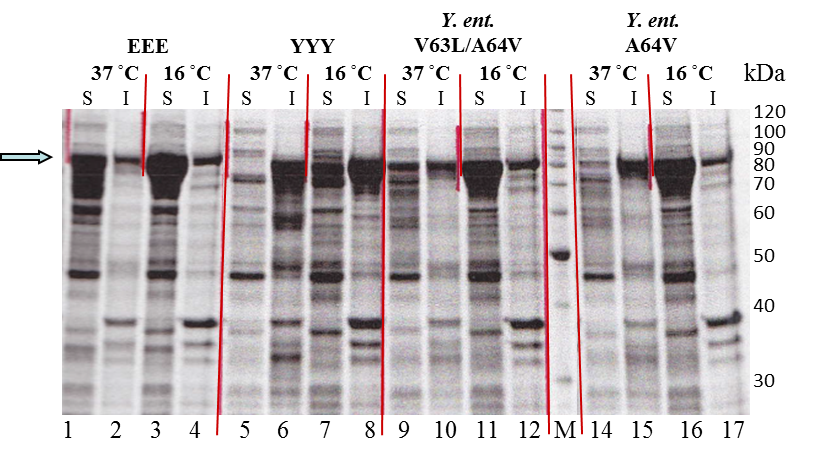


Figure 8: SDS page gel of NusA wild type and mutant proteins without a fusion partner. Samples are total (T) and soluble (S) fractions of bacterial cell lysates from induced cultures that were grown at 37C for 2hr. The wild type *E. coli* (EEE) and wild type *Y. ent*. (YYY) serve as controls, along with the double mutant clones in lanes 5 through 8, for comparison to the single mutant clones in the remaining lanes.

We also tested these mutants as fusion partners with the human GFER protein which is insoluble when expressed without a tag. This protein is more easily solubilized than GRIM19 when tested with other solubility enhancing tags (data not shown). We found subtle differences in expression between the single mutants and the double mutant that we could not see when they were expressed alone (see Figure 9). For this analysis we show soluble and insoluble fractions instead of total and soluble fractions on the gels to highlight the subtle differences. It is also necessary to examine expression at both 37 °C and 16 °C. The double mutant was as soluble as the native *E. coli* protein, where the single mutant A64V was more soluble than the wild type *Y. ent*, but not as soluble as the double mutant. The difference can be seen in the 37 °C sample lanes.

A.

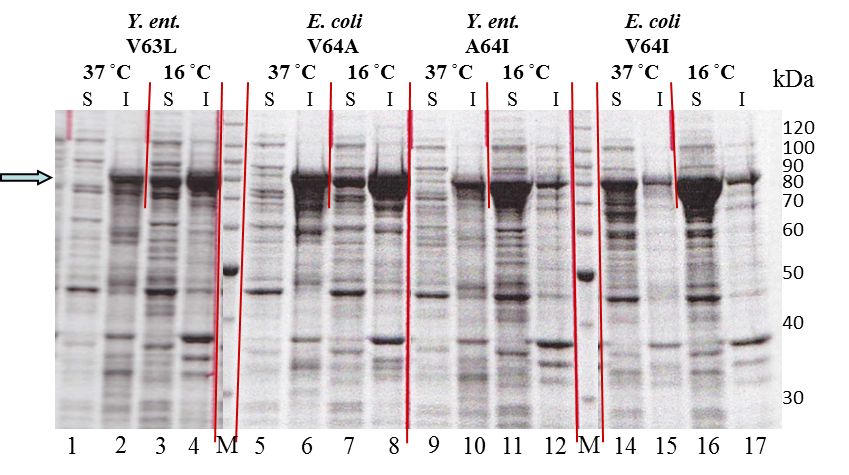
B.

Figure 9: SDS page gels of NusA wild type and mutant proteins fused to human GFER. These images show soluble (S) and insoluble (I) fractions from cultures expressed at 37̊C and 16̊C. The wild type *E. coli* (EEE) and wild type *Y. ent* (YYY) serve as controls. Lanes 5 through 8 of gel A are the double mutant in the *Y. ent* sequence. The single mutants in lanes 14 through 17 of gel A and 1 through 9 of gel B are the complementing amino acid substitutions from the homolog. The final two clones are the successful mutants from the random mutant set that was previously tested without a fusion partner. All point mutants were made by site-directed mutagenesis to the wildtype sequences as noted.

DISCUSSION

N-utilizing substance A has been successfully used as a solubility tag for *E. coli* transgenic protein expression; however, there are situations where it is not capable of solubilizing its passenger protein. We asked if homologues of NusA from other bacteria would be better at this job but found that the *E. coli* native protein was the best of the ones we tested. Figure 10 shows an evolutionary tree of the NusA proteins which we tested. Interestingly, we discovered that the NusA from *Y. ent* was completely insoluble when expressed in *E. coli* despite the high sequence identity with *E. coli* NusA of 90%. We thought that investigating this further could help us understand how NusA functions as a solubility enhancer. We made chimeras between the sequences of the two homologues and tested the effects both with and without the burden of a fusion protein, and found that the C-terminal end of the protein, while closest to the fusion partner, did not seem to change the solubility of the chimeras. Similarly, the middle region could be swapped out without affecting the solubility ratio. The N-terminal region however, transferred the solubility phenotype from the *E. coli* NusA to the *Y. ent* protein. With a second round of chimeras, we narrowed the region of interest down to the portion of the N-terminal domain between amino acids 62 and 121. This region contained just seven amino acid variances between the two sequences. We used site directed mutagenesis to investigate these differences and identified amino acids 63 and 64 as the pair that were responsible for the effects on solubility (Figure 7). We studied the mutations individually without fusion partners and found that the *Y. ent* A64V mutant was fully soluble at 37̊C while the *Y. ent* V63L mutant was insoluble (Figure 8). We then made random mutations at those locations to see if we could improve the solubility. We were unable to significantly improve the solubility over the *E. coli* wildtype, but we discovered that an isoleucine at position 64 in the *Y. ent* sequence recovered some solubility over the wild type *Y. ent* (Figures 8 and 9). Interestingly, we see isoleucine in position 64 of the *Colwellia psychrerythraea* indicating that there may be a naturally occurring biological correlation to the result we observed.

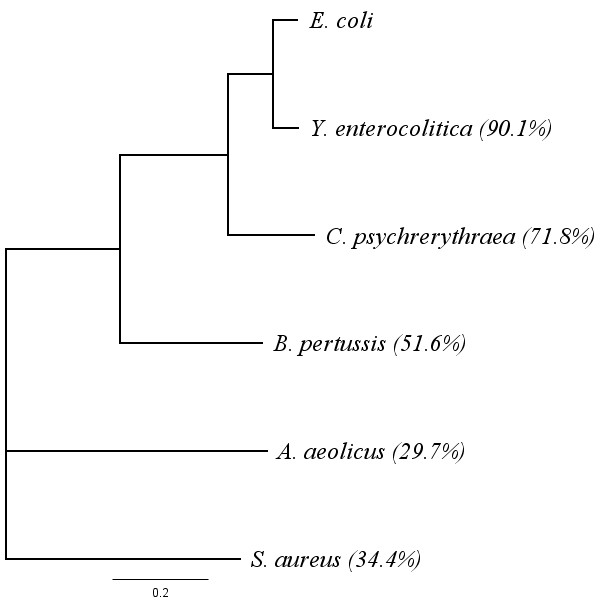
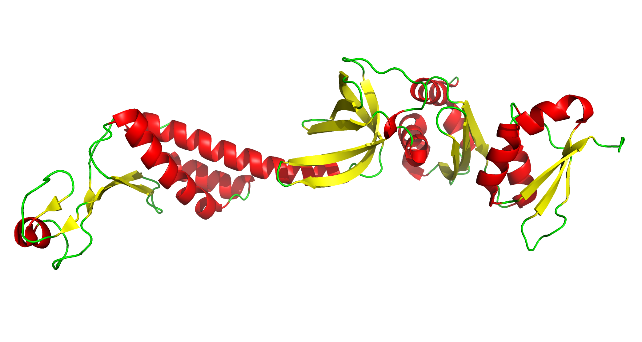
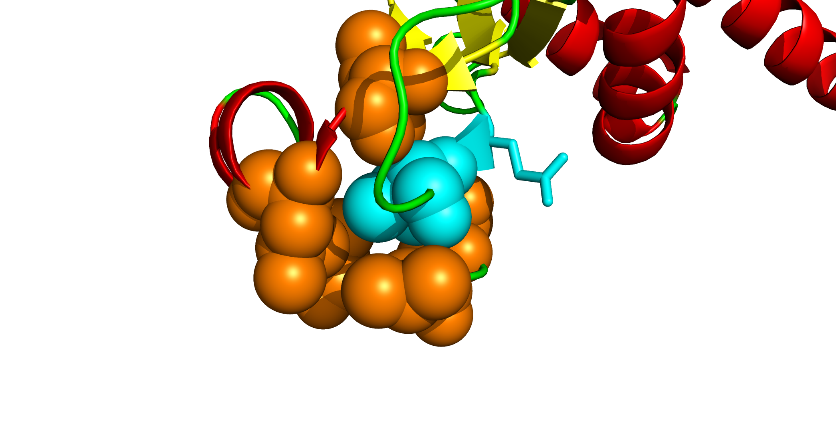


Figure 10: Sequence relationship of NusA homologues. The tree was generated in Geneious software using a ClustalW multiple sequence alignment with a Blosum62 cost matrix for the amino acid sequence alignment and a Jukes-Cantor, neighbor-joining tree building method. Numbers in parenthesis indicate the percent identity relative to the *E. coli* sequence. The inset image shows a focused alignment around the region of interest

Fusions with the GFER protein show that the *Y. ent* A64I has the same solubility as *Y. ent* A64V which is better than the wild type *Y. ent*, but not as good as the V63L/A64V double mutant. An isoleucine substitution in the *E. coli* sequence at position 64 retains the solubility profile of the wild type. This suggests that either valine or isoleucine can be used in that location successfully. A substitution of alanine for valine at position 64 in the *E. coli* sequence is sufficient to drop the solubility to that of the *Y. ent* wild type.

We can use the characteristics of the amino acids in both sequences and their location as clues to the cause of the observed solubility differences. Alanine, valine, leucine and isoleucine are all nonpolar amino acids, which differ in size and structure. These differences can greatly influence the tertiary structure of a folded protein. Nonpolar amino acids are most usually found on the inside of folded structures under aqueous conditions. Alanine has a much smaller side chain than valine and isoleucine. It is possible that, at position 64, the side chain of alanine is too short to properly interact with other amino acids in a hydrophobic pocket. Combine that with a valine at position 63 in *Y. ent* which is shorter than the leucine in *E. coli* and you could create an empty pocket and destabilize the tertiary structure. *T. maritima* has a valine at position 64 like *E. coli* and if we look at the structure model built from its amino acid sequence the prediction is that the valine at position 64 does in fact point to the interior of the folded N-terminal domain (Figure 11). It appears that the A64V substitution is more important for rescuing solubility based on the experimental data. However, while necessary, it is not sufficient to restore solubility of the *Y. ent* NusA to the level of the native *E. coli* protein*.* When combined with the V63L substitution the solubility enhancement capability of *Y. ent* NusA matches that of *E. coli* native protein when fused to a passenger protein.



64

63

75

91

89

87

77

Figure 11: Structure model showing proposed fit of valine at position 64 in the *T. maritima* NusA sequence. The model on the right is a predicted tertiary structure based on the amino acid sequence. The figure on the left is a zoomed in image with space filling used to show the proposed hydrophobic pocket. The light blue amino acid that is represented in stick form is glutamic acid in position 63 as opposed to the leucine that is in the *E. coli* sequence. The light blue amino acid represented in sphere form is valine at position 64. These models were built by Dr. Simon Messing in the MacPyMol software version 1.7.6.2 Enhanced for OS X ([Schrodinger 2015](#_ENREF_42)).

We suggest that the inability of *Y. ent* NusA to serve as a solubility-enhancing fusion tag is directly related to improper folding of the NusA protein itself due to the amino acid differences at position 63 and 64. We see these effects both when NusA is expressed by itself and as a fusion to an insoluble protein partner. We can reverse these effects by substituting the *E. coli* amino acids at these locations in the *Y. ent* sequence. It will be important to perform size exclusion chromatography on protein lysates from the *E. coli* wild type, *Y. ent* wild type, and the *Y. ent* A64V and *E. coli* V64A mutants to look for protein aggregates and profile the mutants against the wild type proteins. Additionally, it may be possible to crystallize the *Y. ent* A64V mutant as it is slightly less soluble than the native *E. coli* protein which has presumably been too soluble to crystallize to date.

Further work will be necessary with the isoleucine mutants as well. The *E. coli* V64I mutant showed some promise as a better solubility enhancer than the native *E. coli* protein, but Coomassie stained gels are not sensitive enough to tell the difference (Figure 9). It is possible that a fluorescent tag could help capture the subtle differences. A 63I/64I double mutant would be worth testing in both backgrounds as well. The isoleucine at position 64 showed solubility enhancement in the *Y. ent* NusA and did not decrease solubility in the *E. coli* protein (Figures 8 and 9). It is possible that an additional isoleucine could be substituted at position 63 and increase solubility. The isoleucine at 63 in *Y. ent* would be larger than valine and has a different sidechain conformation than leucine. Both of these characteristics might change the fold of the protein in this region, possibly leading to a more stable and soluble conformation.

Further characterization of the *Y. ent* V63L/A64V/Q81R triple mutant is also necessary as this mutant was fully soluble when expressed without a fusion partner (data not shown). The Q81R mutant was slightly soluble when expressed without the other two amino acid substitutions (Figure 7). In this case the glutamine is substituted for arginine. Both are polar amino acids and are on the outside of the folded protein structure, however arginine is charged while glutamine is not. This makes it capable of forming a salt bridge that can increase stability if an acidic amino acid is available in the right region of the protein.

One additional puzzle that remains is how the *Y. ent* NusA can be soluble in its host organism while totally insoluble in its close relative, *E. coli*. One possibility is that there is a binding partner that attaches to the N-terminal domain and encourages NusA to be soluble. If these proteins co-evolved then there could be a complimentary difference in the binding partner in *Y. ent* that allows the *Y. ent*. wildtype NusA to be soluble in its native host. When expressed in *E. coli*, this binding interaction could be incompatible resulting in solubility issues. A pull-down assay would be a good way to figure out if this is the case. It might also explain how the *E. coli* protein becomes insoluble with a single amino acid change. If this is the case, then over expression of NusA without over expression of the binding partner could also create a solubility limiting situation, as the supply of the binding partner could become exhausted. Should this prove to be true, co-expressing the binding partner could increase the yield of soluble protein.

In conclusion, we utilized the insoluble NusA from *Y. ent* and rescued it to the level of solubility of the native *E. coli* protein by point mutation at amino acids 63 and 64. Additionally, we may have discovered a mutant that is slightly more soluble than the native *E. coli* protein. These discoveries lead us to believe that the N-terminal domain of NusA is extremely important in proper folding of the protein despite the elongated and flexible structure of the protein as a whole.

REFERENCES

Baneyx F, Mujacic M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nature Biotechnology. 22(11):1399-1408.

Bernard P, Couturier M. 1992. Cell killing by the f plasmid ccdb protein involves poisoning of DNA-topoisomerase ii complexes. Journal of Molecular Biology. 226(3):735-745.

Brule CE, Grayhack EJ. 2017. Synonymous codons: Choose wisely for expression. Trends in Genetics: TIG. 33(4):283-297.

Buchan JR, Stansfield I. 2007. Halting a cellular production line: Responses to ribosomal pausing during translation. Biology of the Cell. 99(9):475-487.

Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, Rodnina MV, Komar AA. 2016. Synonymous codons direct cotranslational folding toward different protein conformations. Molecular Cell. 61(3):341-351.

Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, Yu C, Boles KS, Evans EJ, Stuart DI et al. 2007. Glycoprotein structural genomics: Solving the glycosylation problem. Structure. 15(3):267-273.

Chatterjee DK, Esposito D. 2006. Enhanced soluble protein expression using two new fusion tags. Protein Expression and Purification. 46(1):122-129.

Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF. 2009. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REl606 and BL21(DE3). Journal of Molecular Biology. 394(4):634-643.

Davis GD, Elisee C, Newham DM, Harrison RG. 1999. New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnology and Bioengineering. 65(4):382-388.

de Marco A. 2011. Molecular and chemical chaperones for improving the yields of soluble recombinant proteins. Methods in Molecular Biology (Clifton, NJ). Humana Press 705:31-51.

de Marco A, De Marco V. 2004. Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. Journal of Biotechnology. 109(1-2):45-52.

de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B. 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. BMC Biotechnology. 7(1):32.

dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: A test for translational selection. Nucleic Acids Research. 32(17):5036-5044.

Esposito D, Garvey LA, Chakiath CS. 2009. Gateway cloning for protein expression. Methods in Molecular Biology. (Totowa, NJ): Humana Press. 498:31-54.

Fox JD, Kapust RB, Waugh DS. 2001. Single amino acid substitutions on the surface of *escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. Protein Science. 10(3):622-630.

Fox JD, Routzahn KM, Bucher MH, Waugh DS. 2003. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. FEBS Letters. 537(1-3):53-57.

Gopal GJ, Kumar A. 2013. Strategies for the production of recombinant protein in *Escherichia coli*. Protein Journal. 32(6):419-425.

Guzman LM, Belin D, Carson MJ, Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose p(BAD) promoter. Journal of Bacteriology. 177(14):4121-4130.

Hannig G, Makrides SC. 1998. Strategies for optimizing heterologous protein expression in *Escherichia coli*. Trends in Biotechnology. 16(2):54-60.

Hanson G, Coller J. 2018. Codon optimality, bias and usage in translation and mRNA decay. Nature Reviews Molecular Cell Biology. 19(1):20-30.

Hartl FU, Hayer-Hartl M. 2002. Molecular chaperones in the cytosol: From nascent chain to folded protein. Science. 295(5561):1852-1858.

Hartley DL, Kane JF. 1988. Properties of inclusion bodies from recombinant *Escherichia coli*. Biochemical Society Transactions. 16(2):101-102.

Hartley JL, Temple GF, Brasch MA. 2000. DNA cloning using in vitro site-specific recombination. Genome Research. 10(11):1788-1795.

Kapust RB, Waugh DS. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused Protein Science. 8(8):1668-1674.

Kaur J, Kumar A, Kaur J. 2018. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. International Journal of Biological Macromolecules. 106:803-822.

Kearse M, Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647-1649.

Li K, Jiang T, Yu B, Wang L, Gao C, Ma C, Xu P, Ma Y. 2013. *Escherichia coli* transcription termination factor NusA: Heat-induced oligomerization and chaperone activity. Science Reports. 3:2347.

Mamipour M, Yousefi M, Hasanzadeh M. 2017. An overview on molecular chaperones enhancing solubility of expressed recombinant proteins with correct folding. International Journal of Biological Macromolecules. 102:367-375.

Moffatt BA, Studier FW. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell. 49(2):221-227.

Mondal S, Yakhnin AV, Babitzke P. 2017. Modular organization of the NusA- and NusG-stimulated RNA polymerase pause signal that participates in the *Bacillus subtilis* *trp* operon attenuation mechanism. Journal of Bacteriology. 199(14).

Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of japanese cedar pollen, Cryj2, in *Escherichia coli*. Applied and Environmental Microbiology. 64(5):1694-1699.

Nishihara K, Kanemori M, Yanagi H, Yura T. 2000. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. Applied and Environmental Microbiology. 66(3):884-889.

Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: Sweeter than ever. Nature Reviews Microbiology. 8(11):765-778.

Novy R, Drott D, Yaeger K, Mierendorf R. 2001. Overcoming the codon bias of *E. coli* for enhanced protein expression. inNovations. 12:1-3.

Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR et al. 2015. Codon optimality is a major determinant of mRNA stability. Cell. 160(6):1111-1124.

Qayyum MZ, Dey D. 2016. Transcription elongation factor NusA is a general antagonist. Journal of Biological Chemistry. 291(15):8090-8108.

Rosano GL, Ceccarelli EA. 2014. Recombinant protein expression in *Escherichia coli*: Advances and challenges. Frontiers in Microbiology. 5:172.

Rual JF, Hirozane-Kishikawa T, Hao T, Bertin N, Li S, Dricot A, Li N, Rosenberg J, Lamesch P, Vidalain PO et al. 2004. Human orfeome version 1.1: A platform for reverse proteomics. Genome Research. 14(10b):2128-2135.

Ryan BJ, Henehan GT. 2013. Overview of approaches to preventing and avoiding proteolysis during expression and purification of proteins. Current Protocols in Protein Science. Chapter 5:Unit5.25.

Sahdev S, Khattar SK, Saini KS. 2008. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. Molecular and Cellular Biochemistry. 307(1-2):249-264.

Schleif R. 2010. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of arac action. FEMS Microbiology Reviews. 34(5):779-796.

Schrodinger L. 2015. The PyMOL molecular graphics system, version 2.0.

Shin DH, Nguyen HH, Jancarik J, Yokota H, Kim R, Kim SH. 2003. Crystal structure of NusA from *Thermotoga maritima* and functional implication of the N-terminal domain. Biochemistry. 42(46):13429-13437.

Silverstone AE, Arditti RR, Magasanik B. 1970. Catabolite-insensitive revertants of *lac* promoter mutants. Proceedings of the National Academy of Sciences. 66(3):773-779.

Sogaard KM, Norholm MH. 2016. Side effects of extra tRNA supplied in a typical bacterial protein production scenario. Protein Science. 25(11):2102-2108.

Sorensen HP, Mortensen KK. 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. Journal of Biotechnology. 115(2):113-128.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. Protein Expression and Purification. 41(1):207-234.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology. 189(1):113-130.

Sundermeier T, Ge Z, Richards J, Dulebohn D, Karzai AW. 2008. Studying tmRNA-mediated surveillance and nonstop mRNA decay. Methods in Enzymology. 447:329-358.

Walker IH, Hsieh PC, Riggs PD. 2010. Mutations in maltose-binding protein that alter affinity and solubility properties. Applied Microbiology and Biotechnology. 88(1):187-197.

Wall VE, Garvey LA, Mehalko JL, Procter LV, Esposito D. 2014. Combinatorial assembly of clone libraries using site-specific recombination. Methods in Molecular Biology (Clifton, NJ). 1116:193-208.

Widmann M, Christen P. 2000. Comparison of folding rates of homologous prokaryotic and eukaryotic proteins. The Journal of Biological Chemistry. 275(25):18619-18622.

Xu M, Xie L, Yu Z, Xie J. 2015. Roles of protein N-myristoylation and translational medicine applications. Critical Reviews in Eukaryotic Gene Expression. 25(3):259-268.

Yang X, Lewis PJ. 2010. The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation. Transcription. 1(2):66-69.

Zhang YB, Howitt J, McCorkle S, Lawrence P, Springer K, Freimuth P. 2004. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. Protein Expression and Purification. 36(2):207-216.