




2018

A Genetically Encoded Fluorescent Amino Acid Reveals Protein Dynamics Regulating The Bacterial Dna Damage Response

Zachary Michael Hostetler

University of Pennsylvania, zhostetler@gmail.com

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Biochemistry Commons](#), [Biophysics Commons](#), and the [Microbiology Commons](#)

Recommended Citation

Hostetler, Zachary Michael, "A Genetically Encoded Fluorescent Amino Acid Reveals Protein Dynamics Regulating The Bacterial Dna Damage Response" (2018). *Publicly Accessible Penn Dissertations*. 3129.
<https://repository.upenn.edu/edissertations/3129>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/3129>
For more information, please contact repository@pobox.upenn.edu.

A Genetically Encoded Fluorescent Amino Acid Reveals Protein Dynamics Regulating The Bacterial Dna Damage Response

Abstract

Diversification of the genetic code in response to selective pressures can render organisms more fit to particular stresses. In many bacteria, the inducible prokaryotic DNA damage (SOS) response facilitates survival and adaptation to genotoxic stresses by upregulating genes involved in both high-fidelity and pro-mutagenic DNA damage repair. Within pathogenic bacteria, treatment with genotoxic antibiotics can induce the SOS response and lead to the acquisition of antibiotic resistance. Interest in disarming the SOS-dependent ability of bacteria to evade antibiotics has prompted investigation into the mechanisms underlying SOS pathway activation. Two proteins, the repressor LexA and the DNA recombinase RecA, operate together to regulate SOS gene expression. In response to DNA damage, RecA directly stimulates an autoproteolysis reaction within LexA that leads to SOS induction. Although this direct interaction is necessary, how complex formation results in autoproteolysis remains unknown. Here, we aimed to use the fluorescent unnatural amino acid acridonylalanine as a probe of the interaction between LexA and RecA. In this work, we first demonstrate how directing the evolution of a tRNA synthetase against the incorporation of unwanted contaminants can result in large increases in the selectivity of this enzyme for acridonylalanine. Recognizing that acridonylalanine incorporation may be poorly tolerated at certain positions in either LexA or RecA, we also describe a systematic evaluation of the effect of its incorporation at different positions on soluble protein expression. While acridonylalanine incorporation at different positions affects soluble protein expression, we could not determine any amino acid properties that reliably correlate with protein solubility. Finally, we show how a fluorescently-labeled LexA variant can be used to monitor the kinetics of association with RecA*. With this assay, we report the kinetic and thermodynamic parameters underlying the interaction of full-length LexA with RecA. Additionally, we provide direct evidence for a binding site on LexA for RecA. Altogether, the work presented here demonstrates how deliberately expanding the genetic code through scientific means enabled the design of new tools for studying protein function, and, in this case, allowed us to probe a protein-protein interaction that regulates a pathway naturally involved in diversifying the genetic code.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Rahul M. Kohli

Keywords

Antibiotic Resistance, Fluorescence, Genetic Code Expansion, LexA, SOS Response, Unnatural Amino Acids

Subject Categories

Biochemistry | Biophysics | Microbiology

A GENETICALLY ENCODED FLUORESCENT AMINO ACID REVEALS PROTEIN DYNAMICS
REGULATING THE BACTERIAL DNA DAMAGE RESPONSE

Zachary M. Hostetler

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Rahul M. Kohli, M.D. Ph.D., Assistant Professor of Medicine

Graduate Group Chairperson

Daniel S. Kessler, Ph.D., Associate Professor of Cell and Developmental Biology

Dissertation Committee

Frederic D. Bushman, Ph.D., William Maul Measey Professor in Microbiology

Yale E. Goldman, M.D. Ph.D., Professor of Physiology

Mark Goulian, Ph.D., Charles and William L. Day Distinguished Professor in the Natural Sciences

E. James Petersson, Ph.D., Associate Professor of Chemistry

*This work is dedicated to my grandfather, John D. Hostetler,
for his unwavering enthusiasm and steadfast support.*

ACKNOWLEDGMENTS

I am indebted to a number of people for making this body of work technically feasible, scientifically rigorous, and intellectually valuable. To my dissertation mentor, Dr. Rahul Kohli, I am grateful for both the opportunity and autonomy with which you afforded me to pursue my scientific interests. I cannot begin to enumerate all the ways that you have supported me these past several years, but your advice and guidance have proven invaluable. To other members of the lab, both past and present, I must recognize your ready willingness to provide technical assistance or scientific knowledge upon my request of either. I want to also particularly acknowledge Amanda Samuels and Jamie DeNizio, two coworkers and friends with whom I shared not only a common workspace but also my every success and failure while a member of the lab.

I also credit the vibrant and collaborative research environment at the University of Pennsylvania for my ability to execute this work. I want to highlight the intellectual contributions of my dissertation committee: Dr. Frederic Bushman, Dr. Yale Goldman, Dr. Mark Goulian, and Dr. E. James Petersson. In particular, an enthusiastic and productive collaboration between the Kohli and Petersson labs was an absolute requirement for much of my research. I especially want to extend gratitude to two individuals from the Petersson lab, Itthipol Sungwienwong and John Ferrie: your help and assistance throughout my dissertation work were essential. Finally, for friends and colleagues from a number of training programs, including the Medical Scientist Training Program and the Chemistry Biology Interface Training Grant, as well as several graduate groups, including Cell and Molecular Biology, Biochemistry and Biophysics, and Chemistry, you all deserve special thanks for constructive and insightful comments and feedback on my dissertation work.

Undoubtedly, I would not have accomplished any of the work described here without the irreplaceable support of friends and family. Strong and lasting friendships with a number of people during these years, including Robert Holtzinger, Jordan Derk, Brian Fulton, and Michael Vido, provided me with an avenue for celebrating triumphs and overcoming challenges. I am also incredibly fortunate to have enjoyed the benefits of having not one, but two, loving families that live

within driving distance of Philadelphia, PA. Naturally, I heavily relied on my parents and sisters and grandfather for their perpetual support and faith in me and my abilities. Without all your love, I would not have reached this milestone in my training. But I must also acknowledge the kindness and generosity of a family that I will soon join shortly after the deposition of this thesis; indeed, I will not ever forget the support of the Khetarpal family throughout graduate school. Your help and encouragement were indispensable resources. Finally, I am eternally grateful to Risha Khetarpal, my best friend and partner for life, for her tireless support and overwhelming love. Risha, your determination, intelligence, independence, and persistence have provided me with the inspiration and motivation I needed to accomplish my goals. I cannot begin to thank you enough.

ABSTRACT

A GENETICALLY ENCODED FLUORESCENT AMINO ACID REVEALS PROTEIN DYNAMICS REGULATING THE BACTERIAL DNA DAMAGE RESPONSE

Zachary M. Hostetler

Rahul M. Kohli

Diversification of the genetic code in response to selective pressures can render organisms more fit to particular stresses. In many bacteria, the inducible prokaryotic DNA damage (SOS) response facilitates survival and adaptation to genotoxic stresses by upregulating genes involved in both high-fidelity and pro-mutagenic DNA damage repair. Within pathogenic bacteria, treatment with genotoxic antibiotics can induce the SOS response and lead to the acquisition of antibiotic resistance. Interest in disarming the SOS-dependent ability of bacteria to evade antibiotics has prompted investigation into the mechanisms underlying SOS pathway activation. Two proteins, the repressor LexA and the DNA recombinase RecA, operate together to regulate SOS gene expression. In response to DNA damage, RecA directly stimulates an autoproteolysis reaction within LexA that leads to SOS induction. Although this direct interaction is necessary, how complex formation results in autoproteolysis remains unknown. Here, we aimed to use the fluorescent unnatural amino acid acridonylalanine as a probe of the interaction between LexA and RecA. In this work, we first demonstrate how directing the evolution of a tRNA synthetase against the incorporation of unwanted contaminants can result in large increases in the selectivity of this enzyme for acridonylalanine. Recognizing that acridonylalanine incorporation may be poorly tolerated at certain positions in either LexA or RecA, we also describe a systematic evaluation of the effect of its incorporation at different positions on soluble protein expression. While acridonylalanine incorporation at different positions affects soluble protein expression, we could not determine any amino acid properties that reliably correlate with protein solubility. Finally, we show how a fluorescently-labeled LexA variant can be used to monitor the kinetics of association

with RecA*. With this assay, we report the kinetic and thermodynamic parameters underlying the interaction of full-length LexA with RecA. Additionally, we provide direct evidence for a binding site on LexA for RecA. Altogether, the work presented here demonstrates how deliberately expanding the genetic code through scientific means enabled the design of new tools for studying protein function, and, in this case, allowed us to probe a protein-protein interaction that regulates a pathway naturally involved in diversifying the genetic code.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
ABSTRACT	v
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF ILLUSTRATIONS	x
Chapter 1: Introduction	1
1.1 Bacterial Adaptation and Antibiotic Resistance.....	1
1.2 The Inducible SOS Pathway and Antimicrobial Evasion	2
1.3 Signal transduction of the SOS response at a molecular level	4
1.4 Deciphering RecA*-Stimulated LexA Autoproteolysis at an Atomic Level.....	6
1.5 Approaches to studying protein structural dynamics	10
1.6 Site-specific incorporation of unnatural amino acids in proteins	12
1.7 Fluorescent unnatural amino acids as tools of protein structure and function.....	15
1.8 An improved fluorescent unnatural amino acid	18
1.9 Thesis Objectives	20
Chapter 2: Improving Target Amino Acid Selectivity in a Permissive Aminoacyl tRNA Synthetase Through Counter-Selection	22
2.1 Abstract.....	22
2.2 Introduction.....	22
2.3 Results and Discussion	26
2.3.1 AcdRS Optimization	27
2.3.2 Analysis of AcdRS Selectivity	29
2.4 Conclusions	35
2.5 Contributions	36
2.6 Acknowledgements.....	36
Chapter 3: Systematic evaluation of soluble protein expression using a fluorescent unnatural amino acid reveals no reliable predictors of tolerability	37
3.1 Abstract.....	37
3.2 Introduction.....	37
3.3 Results and Discussion	39
3.4 Conclusion	46
3.5 Contributions	47
3.6 Acknowledgments.....	47
Chapter 4: A Fluorescent Amino Acid Probe Reveals the Kinetic and Molecular Basis of the Interaction Between LexA and RecA	48
4.1 Abstract.....	48
4.2 Introduction.....	48

4.3 Results	51
4.3.1 Design of a Fluorescent Full-Length LexA Binding Reporter	51
4.3.2 Rates of LexA:RecA* Complex Formation	53
4.3.3 Competition Between Labeled and Unlabeled LexA for RecA* Binding Sites	56
4.3.4 Comparison of Competitive Binding Strengths Among Modified LexA Variants	58
4.4 Discussion	61
4.4.1 Association Rate and Affinity of the LexA:RecA* Binding Step	62
4.4.2 A Structural Region in LexA Strongly Impacts Binding to RecA*	64
4.4.3 A Refined Catalytic Role for RecA*	65
4.5 Methods	66
4.5.1 Protein cloning, expression, and purification	66
4.5.2 RecA Activation Reactions	67
4.5.3 Fluorescence Anisotropy Measurements	68
4.5.4 Measuring LexA:RecA* Association Rates and Fitting a Binding Model	68
4.5.5 Measuring a LexA:RecA* Dissociation Rate	69
4.5.6 Competition Binding Assays with Unlabeled LexA Variants	69
4.6 Acknowledgments	71
Chapter 5: Future Directions and Concluding Remarks	72
5.1 Investigating the Binding of Other LexA-Family Proteins to RecA*	74
5.2 Probing the Conformational Dynamics of the LexA Peptide Loop	76
5.3 Identifying a Minimal LexA:RecA* Complex	80
5.4 Concluding Remarks	82
APPENDIX	85
Supporting Information for Chapter 2	85
General Information	85
Chemical Synthesis	85
Synthetase Selection	87
<i>In Vivo</i> Selectivity Analysis	89
Cloning and Purification of Synthetases	94
<i>In Vitro</i> Aminoacylation Analysis	96
Computational Modeling	98
Supporting Information for Chapter 3	100
Experimental Methods	100
Supplemental Figures	103
Supplemental Tables	110
BIBLIOGRAPHY	117

LIST OF TABLES

Table 1.1: Properties of Fluorescent Unnatural Amino Acids.....	17
Table 2.1. AcdRS <i>In Vitro</i> Enzymology and <i>In Vivo</i> MS Selectivity Parameters	33
Table 3.1: List of properties examined for association with Uaa tolerability.....	40
Supplemental Table 1: Sequence of top performing AcdRSs.	88
Supplemental Table 2: CaM 113 Trypsin Digest Intensity Scaling.....	92
Supplemental Table 3: Measured Total and Soluble Amounts of Fluorescent LexA	110
Supplemental Table 4: Summary of Linear Regression Fits for Categorical Properties With LexA.....	111
Supplemental Table 5: Summary of Linear Regression Fits for Numerical Properties with LexA.....	112
Supplemental Table 6: Two-Factor Linear Regression Fits for Categorical Properties with LexA.....	113
Supplemental Table 7: Two-Factor Linear Regression Fits for Numerical Properties with LexA.....	114
Supplemental Table 8: Measured Total and Soluble Amounts of Fluorescent RecA.....	115
Supplemental Table 9: Summary Statistics of Linear Regression Models with RecA.....	116

LIST OF ILLUSTRATIONS

Figure 1.1. The Bacterial DNA Damage (SOS) Response.....	4
Figure 1.2. Possible Catalytic Roles of RecA* in LexA Autoproteolysis.....	7
Figure 1.3. The Structural Model of LexA.....	8
Figure 1.4. Possible Binding Orientations for LexA and RecA*.....	9
Figure 1.5. Biological Components for Unnatural Amino Acid Mutagenesis.....	14
Figure 1.6. Fluorescent Unnatural Amino Acids that can be Incorporated In Vivo.....	16
Figure 2.1. Synthesis of Acd by H ₂ SO ₄ or Polyphosphoric Acid (PPA) Routes.....	25
Figure 2.2. AcdRS Selection.....	27
Figure 2.3. In Vivo AcdRS Selectivity.....	29
Figure 2.4. CaM ₁₁₃ Typysin Digest Data for AcdRS Selectivity Analysis.....	31
Figure 2.5. tRNA Aminoacylation Kinetics.....	32
Figure 2.6. AcdRS Homology Models.....	34
Figure 3.1: Scanning a variety of positions in LexA for Acd tolerability.....	41
Figure 3.2: Features associated with soluble Acd-labeled LexA proteins.....	43
Figure 3.3: Features associated with soluble Acd-labeled RecA proteins.....	45
Figure 4.1. Macromolecular Interactions with LexA Regulate the SOS Response.....	49
Figure 4.2. Screening Positions that Report on Specific RecA* Binding.....	51
Figure 4.3. A Fluorescent LexA Binding Reporter.....	52
Figure 4.4. Association of LexA with RecA*.....	54
Figure 4.5. Dissociation of LexA from RecA*.....	55
Figure 4.6. Competitive Binding of LexA Proteins for RecA*.....	57
Figure 4.7. Affinities of LexA Truncations for RecA*.....	58
Figure 4.8. Affinities of LexA Proteolytic Fragments for RecA*.....	59
Figure 4.9. Affinities of LexA Loop Mutants for RecA*.....	61
Figure 5.1. The Overall LexA:RecA* Reaction Pathway Model.....	74
Figure 5.2. Examples of LexA-Family Proteins.....	75
Figure 5.3. Revealing LexA's Conformational Dynamics.....	77
Figure 5.4. Design of Oligomeric RecA Proteins.....	81
Supplemental Figure 1. Fluorescence measurements of RSs with GFP reporter.....	89
Supplemental Figure 2. In Vivo AcdRS Selectivity.....	90
Supplemental Figure 3. In Vivo AcdRS Selectivity.....	91
Supplemental Figure 4. In Vivo AcdRS Selectivity.....	93
Supplemental Figure 5. AcdRS Malachite Green Activity Assay.....	95
Supplemental Figure 6. AcdRS Purification.....	95
Supplemental Figure 7. Representative primary data for aminoacylation.....	97
Supplemental Figure 8. Monitoring a Conformational Change in αS.....	98
Supplemental Figure 9. Measuring the Kinetics of LexA-DNA Binding.....	99

Supplemental Figure 10. Sampling of Numerical Properties by Chosen Positions in LexA.....	103
Supplemental Figure 11. Sampling of Categorical Properties by Chosen Positions in LexA.....	104
Supplemental Figure 12. Dynamic Range Determination from Purified LexA Standards.	105
Supplemental Figure 13. Reproducibility of Experimental Approach.	105
Supplemental Figure 14. Effect of Neighboring Nucleotides on Amber Suppression Efficiency.....	106
Supplemental Figure 15. Effect of Individual Numerical Properties on LexA Solubility.....	107
Supplemental Figure 16. Effect of Individual Categorical Properties on LexA Solubility.....	108
Supplemental Figure 17. Predicting Protein Solubility through Simulation of Acd Incorporation in LexA.	109
Supplemental Figure 18. Predicting Protein Solubility through Simulation of Acd Incorporation in RecA....	109

Chapter 1: Introduction

1.1 Bacterial Adaptation and Antibiotic Resistance

Bacterial evolution largely depends on the natural selection of individual cells from a diverse population that are more able to adapt to some selective pressure. Over the last several decades, circumstances have transformed this intellectual abstraction into a deadly reality. Antibiotic overuse or misuse combined with declines in antibiotic development programs have combined to produce fertile ground for the rapid rise of organisms that are well-adapted to the current arsenal of antibiotics. Globally, antimicrobial resistance currently leads to 700,000 deaths each year (Review on Antimicrobial Resistance, 2016). Infections caused by drug-resistant bacteria are nearly twice as likely to result in higher rates of morbidity and mortality and inarguably lead to additional utilization of healthcare resources (N. D. Friedman, 2016). Further, losing antibiotic efficacy could cripple the foundation of modern medicine. Many procedures rely on effective antibiotic prophylaxis or treatment, such as cancer chemotherapy, organ transplantation, invasive surgery, and intensive care intubation. Without deliberate action, the annual death toll attributable to antimicrobial resistance could rise to 10,000,000 by 2050, accompanied by economic losses of up to \$100 trillion (Review on Antimicrobial Resistance, 2016).

Justifiably, the sense of urgency surrounding antimicrobial resistance has grown. Efforts to mitigate the global overuse and misuse of antibiotics have resulted in new awareness campaigns and public health interventions that could lessen the demand for existing antibiotics. However, the inevitability of resistance to current antibiotics necessitates a continued need to expand our supply of new antimicrobials. While wholesale abandonment of new antibiotic development programs in the pharmaceutical sector contributed to this crisis over the past several decades (Brown, 2016; The Pew Charitable Trusts, 2016), the more recent direction of resources and financial support to several key intellectual efforts, notably distinct from traditional small molecule approaches,

promises to renew the antibiotic pipeline again (World Economic Forum, 2018). Among these include pursuing conditionally-essential genes as targets (Brown, 2016), chemically modifying existing drugs to increase permeability or overcome intrinsic resistance (Richter, 2017; The Pew Charitable Trusts, 2016), and discovering new natural antimicrobials (Hover, 2018; Ling, 2015; Walsh, 2003). As awareness around the problem of antimicrobial resistance grows, increased investment in these strategies will hopefully follow.

Unfortunately, progress achieved under any of these strategies could be quickly compromised due to our insufficient knowledge of how bacteria acquire mechanisms that permit tolerance to and escape from new antibiotics. The introduction of new antimicrobials in the clinic exposes pathogenic bacteria to novel evolutionary pressures, which can result in the selection of genetic changes in those bacteria. These changes include, but are not limited to, modifications in the drug target or increases in processes that inactivate the drug itself. This phenomenon of acquired resistance can result from mutagenesis leading to genomic point mutations or from horizontal gene transfer leading to expression of novel genes. While many of these acquired resistance mechanisms are well-studied, the underlying mechanisms that lead to acquired resistance remain underexplored. Mechanistic insights into how acquired resistance arises could help researchers and clinicians actively deter the rise of acquired resistance or reveal potentially unexploited therapeutic targets for expanding our antibiotic arsenal (Culyba, 2015; P. A. Smith, 2007).

1.2 The Inducible SOS Pathway and Antimicrobial Evasion

Bacteria possess a number of pathways that respond to a diverse array of environmental stresses by inducing adaptive mechanisms that promote genetic diversity, such as the upregulation of error-prone DNA polymerases and the repression of error-correcting enzymes (Foster, 2007). Among these pathways, the bacterial DNA damage (SOS) response—an inducible network of genes involved in the repair of and adaptation to DNA damage—serves a crucial role in the response to genotoxic stresses (Radman, 1975; Schlacher, 2007). Inducing stimuli for the SOS

response include nearly any environmental insult that damages DNA or leads to the arrest of DNA replication, including UV-induced cross-links, double-stranded DNA breaks caused by ionizing radiation, thymine starvation, or even temperature shifts in strains with temperature-sensitive DNA replication (Witkin, 1976). Certainly, treatment with a number of antibiotics can directly lead to these sorts of DNA damage, such as fluoroquinolones that incite double-stranded DNA breaks, mitomycin C or other DNA-alkylating agents, and trimethoprim or other agents that target DNA metabolism. Further, bactericidal antibiotics that do not directly target DNA may indirectly cause DNA damage, through the stress-induced accumulation of reactive oxygen species that subsequently cause direct oxidative damage on DNA nucleobases (Dixon, 2014; Kohanski, 2007).

Genes regulated by the SOS response, and the functions those genes provide, vary widely across different bacterial species likely as a result of the particular types of genotoxic stresses to which those bacteria have adapted. Within *Escherichia coli*, the SOS response regulates the expression of over forty genes (Courcelle, 2001), whose gene products range from proteins involved in nucleotide excision repair and homologous recombination (Kowalczykowski, 2015; Truglio, 2006), error-prone polymerases capable of translesion synthesis (Tippin, 2004), to toxin-antitoxin systems that arrest cell growth (Dörr, 2009). In other bacterial species, SOS effector genes provide even more diverse functions, from the formation of biofilms (Bernier, 2013), induction of virulence factors and toxins (Quinones, 2005), activation of integrons that can lead to horizontal gene transfer (Guerin, 2009; Hocquet, 2012), to even regulation of direct antibiotic resistance genes (Da Re, 2009). Even among bacteria with relatively few SOS-regulated genes, the observation that a core group of SOS regulated genes provides inducible translesion synthesis function suggests an essential role in overcoming damage to the genome (Erill, 2007). Undoubtedly, the aptly-named SOS response can perform a number of “life-saving” antibiotic evasion roles for bacterial pathogens awash in a sea of possible antibiotic therapies.

1.3 Signal transduction of the SOS response at a molecular level

Interest in preventing the bacterial evasion of antimicrobials has driven decades of mechanistic studies into how the stimulus of DNA damage is transduced at a molecular level into the expression of SOS response genes. In the absence of genotoxic stress, SOS genes are maintained under transcriptional repression by the repressor LexA (Butala, 2009; Little, 1984, 1994). Broadly conserved across bacteria, LexA is comprised of a DNA-binding N-terminal domain (NTD) and a serine protease C-terminal domain (CTD) that also contains a dimerization interface. In the basal state, dimers of LexA (Mohana-Borges, 2000) repress the 40 genes comprising the SOS genetic regulon in *E. coli* by binding to specific operator DNA sequences within the promoter DNA and consequently inhibiting transcription (Brent, 1981; Erill, 2007) (Figure 1.1, left). SOS operator sequences are not strictly conserved; sequence differences alter the affinity and kinetics of LexA

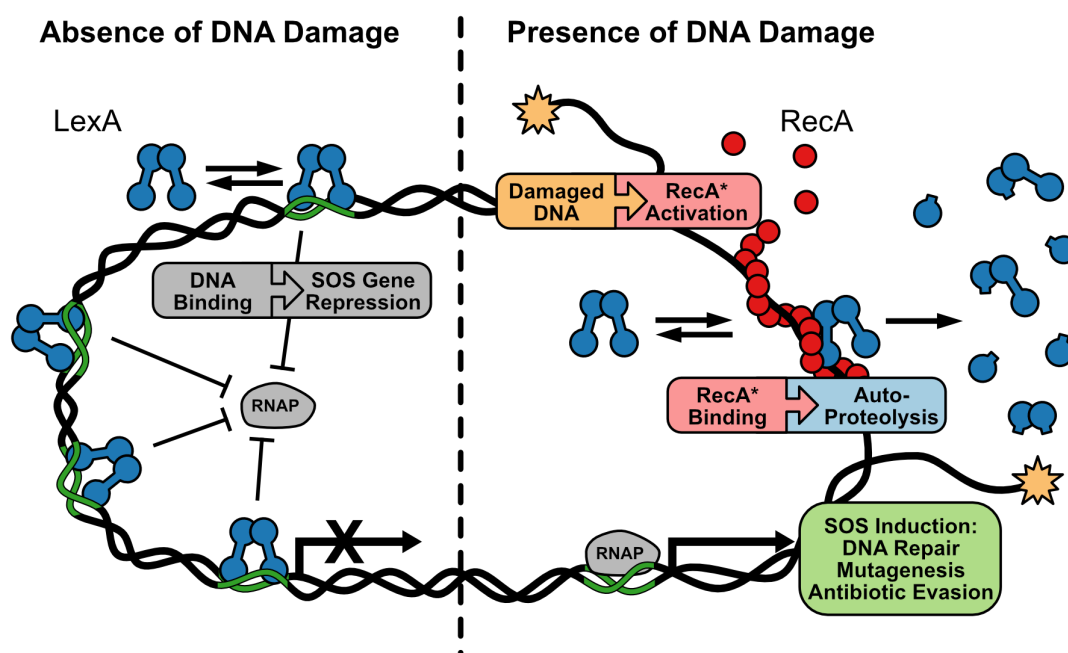


Figure 1.1. The Bacterial DNA Damage (SOS) Response. Left: State of a bacterial cell in the absence of DNA damage. LexA repressor dimers (blue) bind to specific operator sequences in the promoters of SOS genes (green), repressing transcription. Right: State of a bacterial cell in the presence of DNA damage. Damaged DNA (yellow) results in the production of single-stranded DNA. Sensing this, the protein RecA (red) polymerizes along the single-stranded DNA, forming activated RecA* nucleoprotein filaments. The direct interaction of free LexA with RecA* results in the autoproteolysis of LexA repressors. As LexA is depleted in the cell due to autoproteolysis, the equilibrium of DNA-bound LexA is disrupted and LexA repressors dissociate from SOS gene promoters. As a result, SOS genes are expressed, resulting in a number of effects that promote survival of DNA damage.

binding to DNA such that the strength and timing of expression can differ between SOS genes (Culyba, 2018; Ronen, 2002).

Both DNA damage and the other genotoxic stresses described above (See “The Inducible SOS Pathway and Antimicrobial Evasion”) ultimately lead to elevated levels of single-stranded DNA (ssDNA), whether as a result of replication fork collapse or resection of double-stranded DNA breaks. Increases in ssDNA initiate the first stages of SOS activation (Lovett, 1994; Sassanfar, 1990) by providing a substrate on which the bacterial ATPase RecA polymerizes, forming extended nucleoprotein filaments (Cox, 2007). Although this activated form of RecA (RecA*) plays a well-studied role in homologous recombination pathways, it also interacts with LexA (Figure 1.1, right). This interaction stimulates an autoproteolytic reaction within the LexA repressor, leading to intramolecular self-cleavage of LexA and subsequent destabilization of the dimer (Slilaty, 1987). Additionally, cleaved LexA fragments are rapidly degraded by cellular proteases (Neher, 2003). Both the destabilization of LexA dimers and depletion of overall LexA levels disrupt the equilibrium of DNA-bound LexA, resulting in the derepression of SOS genes.

Notably, both the *recA* and *lexA* genes are under SOS response genetic regulation, which allows a more complicated genetic circuit to form (Brent, 1981). At a single cell level, the combination of positive feedback regulation from RecA and negative feedback regulation from LexA results in pulses of SOS pathway induction for as long as DNA damage persists (N. Friedman, 2005). Upon successful repair of DNA damage, re-expression of LexA repressors and dissolution of RecA* filaments restores the initial state of repression to the cell.

Directed genetic studies against the RecA* and LexA regulatory axis have highlighted the importance of SOS response inducibility in decreasing antibiotic tolerance and preventing or reversing resistance. For example, deletion of *recA* or inactivation of LexA self-cleavage decreases the MIC of DNA damaging antibiotics by as much as 32-fold (Mo, 2016). Additionally, overexpression of inactivated LexA via phage transduction can resensitize fluoroquinolone-resistant *E. coli* (Lu, 2009). In an *in vivo* infection model, genetic inactivation of LexA inhibited the

acquisition of antibiotic resistance to fluoroquinolones (Cirz, 2005). Based on these observations, several groups have speculated that targeted small molecule inhibition of the SOS response could potentiate the current antibiotic arsenal by lowering MICs for DNA damaging antibiotics and suppressing acquired resistance (Cirz, 2007; Rodríguez-Rojas, 2013). The identification of natural protein factors, such as DinI, that modulate or disrupt the LexA/RecA* signaling axis provides indirect evidence of the feasibility of targeting the SOS response (Galkin, 2011). However, therapeutic efforts remain stymied by the absence of information describing the interaction between RecA* and LexA either at an atomic level or with sufficient mechanistic detail.

1.4 Deciphering RecA*-Stimulated LexA Autoproteolysis at an Atomic Level

Induction of the SOS response depends on an activated RecA* nucleoprotein filament stimulating the latent autoproteolysis activity in the LexA repressor. However, the question of how a protein-protein interaction catalyzes intra-molecular peptide bond hydrolysis has bedeviled the field for decades. Early explanations for this phenomenon offered a variety of possible mechanisms (Lin, 1989). First, activated RecA* could bind both the target scissile bond and active site simultaneously, either by forcing them into apposition or stabilizing a pre-existing conformation (See 1 in Figure 1.2). Alternatively, close association of RecA* with the active site of LexA could change the local chemical environment of the catalytic dyad, better activating it for nucleophilic attack (See 2 in Figure 1.2). Finally, RecA* might bind in such a way to stabilize the energetically unstable tetrahedral intermediate of the forward peptide bond hydrolysis reaction, akin to the stabilization provided by a key glycine residue in chymotrypsin (See 3 in Figure 1.2).

Using a series of genetic and biochemical studies, Little and colleagues began to test the above hypothetical mechanisms for the catalytic role of RecA* in LexA autoproteolysis. An *in vivo* screen for LexA mutants deficient in RecA-mediated cleavage revealed the essential role of Lys156 as part of the catalytic dyad (Lin, 1988). The observation that a K156R substitution in LexA retained basal autoproteolysis activity at alkaline pH strongly suggested that a deprotonated Lys or Arg residue at position 156 is required to abstract a proton from Ser119 (Lin, 1989). Additionally, several

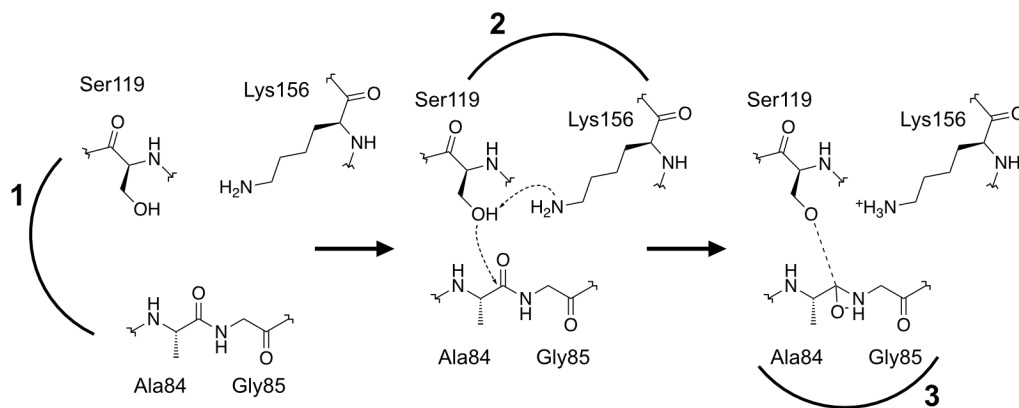


Figure 1.2. Possible Catalytic Roles of RecA* in LexA Autoproteolysis. (1) RecA* bringing the catalytic dyad residues (Ser119 and Lys156) and the target scissile bond (Ala84-Gly85) into close proximity. (2) RecA* activating the catalytic dyad residues for nucleophilic attack. (3) RecA* stabilizing the tetrahedral intermediate of the proteolysis reaction.

lines of evidence identified LexA mutants with increased *in vitro* and *in vivo* rates of autoproteolysis—independent of RecA*—suggesting that these substitutions were overcoming a rate-limiting step in autoproteolysis, perhaps in a manner similar to the role of RecA* (Mo, 2014; Roland, 1992; M. H. Smith, 1991). This alteration of the rate-limiting step could involve either reducing the free energy of a peptide bond tetrahedral intermediate or destabilizing the LexA ground state conformation.

Some clarification of the autoproteolysis mechanism arrived in the form of protein crystal structures of LexA, revealing for the first time how the target scissile bond might interact with the C-terminal protease domain (Y. Luo, 2001; A. P. P. Zhang, 2010) (Figure 1.3, left). The structures of different *E. coli* LexA mutants, relying on amino acid substitutions identified in the aforementioned studies, revealed a mobile peptide loop within the CTD that can adopt two distinct conformations (Figure 1.3, middle), even as the structure of the catalytic core remains static (RMSD = 0.5 Å). This peptide loop, which contains the Ala84-Gly85 scissile bond, undergoes a striking ~20 Å shift between different structures, placing it either away from, or in close proximity to, the active site Ser119/Lys156 catalytic dyad (Figure 1.3, right). Moreover, in the cleavage competent conformation, the active site catalytic dyad residues were effectively buried from the surface of the

protein at a distance of about 6 Å, making a model in which RecA* directly increases the nucleophilicity of the catalytic dyad less likely. Instead, a proposed model in which the conformation of LexA switches to a cleavage-proficient state in the presence of RecA* was proposed (Y. Luo, 2001).

In light of this “conformational switch” model, a number of previously identified rate-enhancing or rate-diminishing substitutions were evaluated or re-evaluated for putative effects on the conformation of the mobile peptide loop. For instance, substitutions in the Gly80 position of LexA are thought to destabilize the cleavage-proficient state by reducing the propensity of that flexible glycine to rotate and form a β bulge motif (Y. Luo, 2001; Mo, 2014). In contrast, substitution of the Gln92 position to an aromatic residue might stabilize the cleavage-proficient state (Roland, 1992), potentially through a favorable cation- π interaction between that residue and Arg148 (Y. Luo, 2001). Additionally, the energetically-unfavorable burial of the charged catalytic Lys156 residue appears to offer a thermodynamic explanation for why the cleavage-proficient conformation is disfavored. Consistent with these expectations, most substitutions thought to either stabilize or destabilize the cleavage-proficient conformation were shown to either strongly or weakly inhibit RecA*-mediated autoproteolysis of wildtype LexA, respectively (Giese, 2008; Lin, 1989).

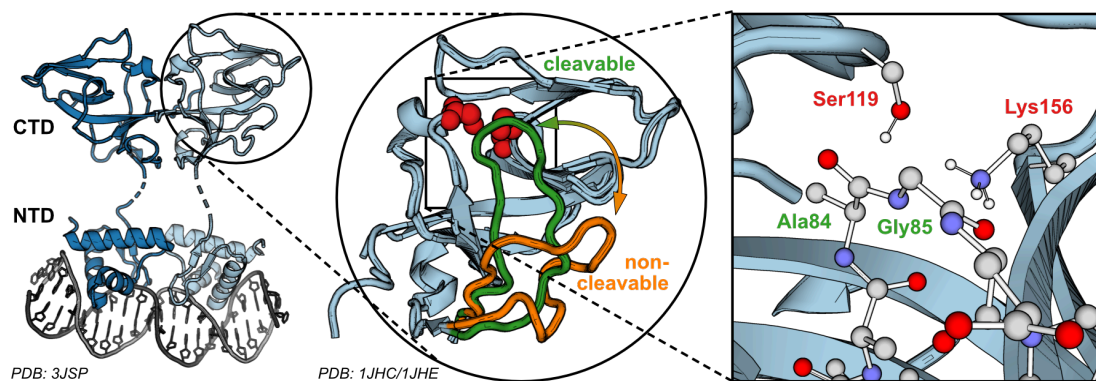


Figure 1.3. The Structural Model of LexA. Left: The LexA dimer bound to DNA. Middle: Two aligned crystal structures of the mobile peptide loop in the C-terminal domain of LexA. The A84-G85 scissile bond at the apex of the loop shifts about 20 Å between the two pictured conformations. Right: Close-up of the LexA active site architecture.

Although the above genetic and biochemical evidence provides a sound foundation for a model in which RecA* induces or stabilizes an energetically unfavorable, cleavage-proficient state of LexA, there is incomplete and often conflicting data regarding this binding interaction. Attempts to visualize the LexA:RecA* complex using cryo-EM investigations of RecA* filaments incubated with the higher-affinity LexA-K156A mutant revealed low-resolution density attributable to LexA in the deep helical groove of RecA* (Yu, 1993). Even though atomic-level interactions between LexA and RecA* were not resolved in this study, mass analysis of LexA occupancy provided evidence for a model in which LexA randomly binds to the RecA* filament with a binding site size of two RecA subunits (McGhee, 1974). To date, higher resolution structural studies of the LexA:RecA* complex are still lacking. Elucidation of the X-ray crystal structure of the activated RecA* nucleoprotein filament (Z. Chen, 2008) encouraged renewed efforts to model the LexA:RecA* complex. Using methods spanning from genetic analysis to chemical cross-linking, several groups proposed models for the binding of LexA to RecA* (Adikesavan, 2011; Kovačič, 2013). However, significant disagreement between these models exists, specifically regarding the proposed orientations in

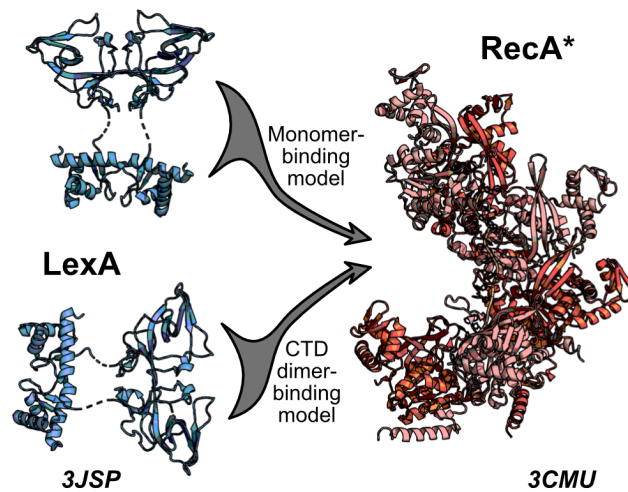


Figure 1.4. Possible Binding Orientations for LexA and RecA*. Solution of the activated RecA* nucleoprotein structure resulted in the proposal of a number of conflicting models of the LexA:RecA* complex. In the monomer-binding model, one LexA monomer binds the groove of RecA*. In the C-terminal dimer binding model, the two C-terminal domains of a LexA dimer wedge into the RecA* groove.

which LexA engages RecA* (Figure 1.4). Further, close inspection of either model does not readily reveal a biophysical rationale for how interaction with RecA* induces or stabilizes the cleavage-proficient conformation of LexA. The limited results delivered by the above approaches highlight the need to probe the LexA:RecA* interface with different methodologies.

1.5 Approaches to studying protein structural dynamics

Attempts to describe how proteins function very often rely on defining protein behavior in terms of conformational changes, molecular binding events, or even folding pathways and domain rearrangements. These protein structural dynamic concepts are almost always borne out of the high-resolution, three-dimensional structural models of the proteins themselves. Experts using these conventional structural biology approaches—nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, and cryo-electron microscopy (cryo-EM)—have continued to push against the resolution limits for larger proteins and complexes and have even sought to describe protein motions as well (Bai, 2015; Van Den Bedem, 2015). Nevertheless, the constraints of data collection for these methods, which include requirements for highly-concentrated proteins, non-physiologic conditions or long timescales, are ill-suited for definitively capturing fast protein movements or rearrangements in conditions that are physiologically relevant both *in vitro* and *in vivo*.

Fluorescence spectroscopic approaches, by contrast, are well-suited for the study of complex biologic processes in a physiologic context (Lakowicz, 2006). Fluorescent labeling of a macromolecule affords researchers with the ability to directly measure changes in fluorescence with unparalleled sensitivity in real time using conditions that more closely mimic the native context of that system. Widespread development of numerous methods for labeling peptides and proteins has intensified interest in studying a variety of dynamic processes that are central to biologic systems, such as protein folding pathways, conformational rearrangements, macromolecular interactions, or catalytic events. Improvements in fluorescent labeling approaches for proteins have

even made the study of protein function, from the level of single molecules all the way to whole organisms, entirely routine.

Despite the broad adoption of fluorescent labeling methodologies, many conventional fluorescent techniques are burdened with disadvantages that are frequently overlooked. Naturally-fluorescent proteins such as green fluorescent protein (GFP) are popular largely due to the ease with which they can be genetically encoded as translational fusions to target proteins. Enhanced GFP variants with improved spectral properties or different colors have resulted in a number of standard applications, from localizing fluorescent proteins fusions *in vivo* to monitoring molecular interactions, using fluorescent protein fusions as Förster resonance energy transfer (FRET) reporters. However, fluorescent proteins are very often restricted to only labeling either the N- or C- terminus of a target protein and, given their size, they can often perturb the structure or function of the target protein (Charbon, 2011). Other fusion protein labeling approaches have similar drawbacks attributable to their steric bulk, such as the HALO-tag, Ligand-Link, and SNAP-tag or CLIP-tag systems (Speight, 2014).

Several alternative strategies not only avoid the steric bulk associated with GFP or other fusion proteins but also maintain the ease of genetically encoding a label in a position-specific manner. Certain genetically-encoded small peptide tags are capable of binding small fluorescent molecules, such as the binding of tetracysteine amino acid motifs (CysCysProGlyCysCys) to biarsenical fluorescein dyes (FIAsH) (Griffin, 1998). Alternatively, fluorescent dyes can be enzymatically attached, such as the Probe Incorporation Mediated by Enzymes (PRIME) method, in which lipoic acid ligase attaches a 7-hydroxycoumarin derivative to a Lys residue within an encoded 13 amino acid recognition site (Uttamapinant, 2013). Expanding on these methods, reactive residues within a protein can be selectively labeled, such as the conjugation of maleimide-containing fluorophores to Cys residues, or the bio-orthogonal conjugation of azide- or alkyne-containing fluorophores to a complementarily-reactive handle present in an unnatural amino acid. However, all of the preceding technologies share common limitations. First, these methods require, at minimum, the addition of

an exogenous reagent such that the protein can be post-translationally labeled with the desired fluorophore. Specific labeling requires a uniquely reactive site, and methods that rely on the labeling of native Cys or Lys residues may suffer from nonspecific labeling. The relative accessibility of the encoded reactive site or tag can also limit the efficiency with which labeling occurs. Finally, many of these exogenous chromophores are attached to the labeling site via flexible linkers capable of moving significant distances, an effect that can obscure the biophysically-relevant behavior being examined.

Given the limitations of the above fluorescent labeling approaches, the optimal fluorescent probe for studying protein structural dynamics will both minimize perturbations to protein structure or function and maximize proximity to the position under investigation. Additionally, installation of this probe should be sufficiently straightforward and efficient such that quantitatively-labeled protein can be obtained in good yield with a minimal amount of effort. The site-specific genetic incorporation of fluorescent unnatural amino acids best satisfies these requirements by avoiding the limitations and artifacts introduced by long linkers and bulky fluorophores. The advent of methods for highly-efficient incorporation of unnatural amino acids *in vivo* permits the application of fluorescent amino acids to a number of applications aimed at studying protein structural dynamics. Below, we describe the technology of unnatural amino acid mutagenesis and describe several examples of fluorescent unnatural amino acids and their technical applications.

1.6 Site-specific incorporation of unnatural amino acids in proteins

As early as the 1960s, the discovery and elaboration of the transfer RNAs (tRNAs) and aminoacyl tRNA synthetases (RSs) that are instrumental in the translation of the genetic code raised interest in altering these molecules for the deliberate introduction of non-canonical or unnatural amino acids into target proteins (Hendrickson, 2004). Toward this goal, Hecht and colleagues successfully modified tRNA^{Phe} with non-phenylalanine residues by using T4 RNA ligase to ligate any of several chemically misacylated dinucleotides to the 3' terminal tRNA nucleotide (Heckler, 1984). This development effectively established a route for chemically misacylating any

given tRNA to an unnatural amino acid of choice. Shortly after, Schultz and coworkers adapted these chemical acylation approaches to include an amber suppressor tRNA (Noren, 1989), which overcame the issue of a misacylated tRNA competing with a wildtype tRNA. The use of a bio-orthogonal, misacylated yeast tRNA_{CUA} in an *E. coli in vitro* expression system enabled site-specific incorporation of unnatural phenylalanine analogs into β -lactamase only at a position where the amber stop codon was encoded. Chamberlin and coworkers independently replicated this approach, using a misacylated amber suppressor tRNA to specifically incorporate ¹²⁵I-tyrosine into a sixteen-residue peptide (Bain, 1989). Consequently, chemical biologists by the 1990s had the means and motivation to begin incorporating unnatural amino acids at specific sites in peptides or proteins in earnest.

Despite these advances, there remained two major hurdles for the incorporation of unnatural amino acids via chemical misacylation of amber suppressor tRNA molecules: Scientists often lacked the technical expertise required for *in vitro* protein translation and they experienced inherently low yields as a result of the stoichiometric nature of chemically-misacylated tRNA. Recognizing these issues, Schultz and coworkers aimed to generate the necessary biologic components for enhancing the site-specific *in vivo* incorporation of unnatural amino acids (Figure 1.5). Fortunately, nature provided the conceptual solution in the form of the non-canonical amino acid pyrrolysine (Pyl), which is incorporated in response to the amber stop codon (UAG) by an orthogonal tRNA/tRNA synthetase pair in certain organisms. Drawing from this example, Schultz and coworkers screened tRNA synthetases and cognate tRNAs from organisms other than *E. coli* for pairs that were orthogonal to the *E. coli* translational machinery (L. Wang, 2001). Ultimately, they selected a tyrosyl tRNA (TyrRS) and cognate tRNA from the archaeabacteria

Methanocaldococcus jannaschii for reprogramming. To evolve this pair for specific incorporation of an unnatural amino acid, the anticodon of the *M. jannaschii* tRNA was first mutated to CUA to recognize the amber stop codon. Next, they constructed a library of *E. coli* cells containing mutated *M. jannaschii* TyrRS enzymes and subjected the library to alternating rounds of positive and negative selection. During rounds of positive selection, TyrRS mutants are transformed with a chloramphenicol acetyl transferase (CAT) gene containing an internal amber stop codon and challenged with chloramphenicol in the presence of the target unnatural amino acid, in this case O-methyl-L-tyrosine. Successful amber suppression by a TyrRS mutant with O-methyl-L-tyrosine, or with an endogenous canonical amino acid, led to survival. During rounds of negative selection, the TyrRS mutants are transformed with a toxic barnase gene containing multiple amber stop codons. Survival of these mutants in the absence of O-methyl-L-tyrosine requires that the TyrRS mutants lack activity with any of the canonical 20 amino acids.

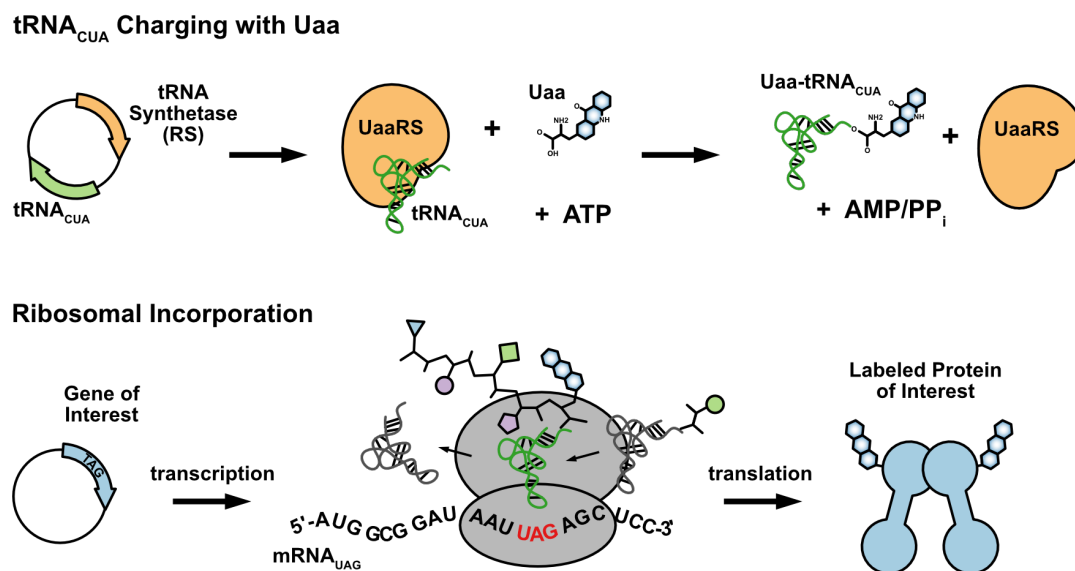


Figure 1.5. Biological Components for Unnatural Amino Acid Mutagenesis. Top: Genes for an orthogonal tRNA synthetase and tRNA pair are provided on a plasmid. Expression of this tRNA/tRNA synthetase pair allows for the specific synthetase-catalyzed reaction of the tRNA with a target Uaa substrate. Bottom: A protein of interest is encoded with an internal stop codon, here the amber stop codon. Ribosomal biosynthesis of this mRNA transcript along with the Uaa-tRNA results in the site-specific incorporation of the Uaa into the protein of interest.

The principles of positive and negative selection during the evolution of a tRNA synthetase ensures the orthogonality of the system: both that the suppressor tRNA is not charged with canonical amino acids by any endogenous tRNA synthetases and that the orthogonal tRNA synthetase charges its cognate suppressor tRNA with only the target unnatural amino acid. The success of this strategy for evolving orthogonal tRNA/tRNA synthetase pairs has resulted in the rapid expansion of available synthetases for a wide variety of unnatural amino acids across a number of host organisms. Some examples include evolved forms of *Methanosarcina* spp. pyrrolysine tRNA synthetases that incorporate lysine derivatives in *E. coli* (Brabham, 2017), as well as *E. coli* leucyl synthetase pairs that have been evolved for orthogonal incorporation of a number of unnatural amino acids in yeast and mammalian cells (Chatterjee, 2013). Other ongoing efforts aimed at recoding other elements of the *E. coli* genome and translational machinery promise to deliver continued improvements in efficient incorporation of a growing number of unnatural amino acids.

1.7 Fluorescent unnatural amino acids as tools of protein structure and function

The development of techniques for the *in vivo* incorporation of unnatural amino acids stimulated a number of efforts to identify and evolve orthogonal tRNA/tRNA synthetase pairs for a wide variety of unnatural amino acids, from those with reactive chemical handles for exogenous labeling to those that have photocaged groups that react on exposure to light. While these various unnatural amino acids have great value in particular applications, the remainder of this background will focus on the currently available fluorescent unnatural amino acids that can be genetically encoded *in vivo* to produce minimally perturbing and maximally sensitive fluorescently-labeled proteins with relative ease. Of the numerous fluorescent unnatural amino acids synthesized and developed to date, only a handful also possess the requisite orthogonal tRNA/tRNA synthetase pairs for *in vivo* incorporation (Figure 1.6). These fluorescent unnatural amino acids are mostly based on the α -amino acid versions of common fluorophores, such as naphthalenes or coumarins. Selected

spectral properties are given in Table 1.1, and individual uses and applications of each of these fluorescent unnatural amino acids are described below.

Given their small size and covalent linkage to the polypeptide backbone, many fluorescent amino acids are frequently used as probes of changes in local environment. Frequently, solvatochromic amino acids, which display changes in spectral properties upon changes in solvent polarity, are used to monitor the folded state of a protein. In one example, Schultz and coworkers tracked local differences in protein unfolding by genetically incorporating L-(7-hydroxycoumarin-4-yl)ethylglycine (Hco) into either helix A or helix C of holomyoglobin and then titrating in urea. Because the Hco label in helix A undergoes a fluorescence increase at lower urea concentrations than the Hco label in helix C, they concluded that helix A is more prone to unfolding and disorder than helix C (J. Wang, 2006). In a parallel effort, the guanidinium-induced denaturation of human superoxide dismutase labeled with dansyl-L-alanine (Dana) showed that the fluorescence of Dana decreases and blue-shifts as the protein unfolds (Summerer, 2006). Two related unnatural amino acids, L-4-cyanophenylalanine (Cnf) and L-4-ethynylphenylalanine (Enf), have also been used to probe protein unfolding because they can both act as FRET donors to intrinsic Trp residues (Miyake-Stoner, 2009; Ries, 2016).

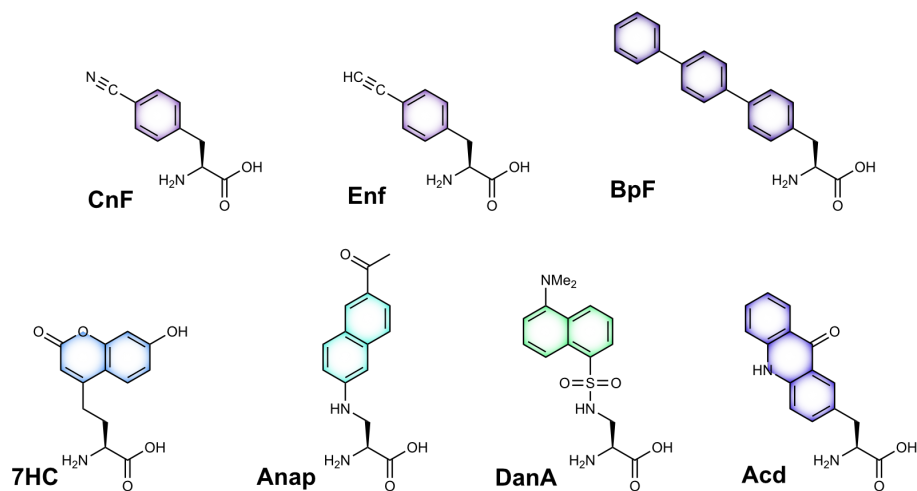


Figure 1.6. Fluorescent Unnatural Amino Acids that can be Incorporated In Vivo. Structures of reported unnatural amino acids with intrinsic fluorescence are colored to approximately reflect their emission maxima.

The sensitivity of fluorescent amino acids enables not only the monitoring of protein folding but also the measurement of protein interactions and conformational changes. Genetic incorporation of the solvatochromic α -amino acid derivative of acetylnaphthalene (Anap) into the *E. coli* glutamine-binding protein (QBP) results in a fluorescent version of QBP that, upon binding to glutamine, demonstrates an increase and blue-shift in its Anap fluorescence signal (Lee, 2009). The genetic incorporation of 4-biphenyl-L-phenylalanine (Bpf), which can form a FRET pair with Hco (S. Chen, 2013), has recently been demonstrated using a substrate-permissive tRNA synthetase originally evolved to incorporate the α -amino acid derivative of naphthalene (Lampkowski, 2015).

In addition to the *in vitro* study of protein function and structural dynamics, fluorescent unnatural amino acids also prove valuable in localizing and studying proteins *in vivo*, especially in cases where using a fluorescent fusion protein becomes problematic. In one notable example, fusing GFP to the bacterial homolog of tubulin, FtsZ, abrogates the contractile activity of this cytoskeletal protein during cytokinesis, but installing the fluorescent Hco into FtsZ *in vivo* permits the subcellular localization of fully-functional FtsZ during cell division in *E. coli* (Charbon, 2011). Further, the

Table 1.1: Properties of Fluorescent Unnatural Amino Acids

Name	λ_{ex} (nm)	λ_{em} (nm)	Extinction coefficient, ϵ_{max} ($\text{M}^{-1} \text{cm}^{-1}$)	Quantum yield, Φ	tRNA Synthetase (RS)
Cnf ^a	233	295	13000	0.11	<i>M. jannaschii</i> TyrRS
Enf ^b	240	299	NA	0.09	<i>M. jannaschii</i> TyrRS
Bpf ^c	280	342	NA	0.49	<i>E. coli</i> LeuRS
Hco ^d	329	464	9000	0.49	<i>M. jannaschii</i> TyrRS
Dana ^e	340	580	33600	0.37	<i>E. coli</i> LeuRS
Anap ^f	360	490	17500	0.48	<i>E. coli</i> LeuRS
Acd ^g	386	410-450	6000	0.95	<i>M. jannaschii</i> TyrRS
a. (Schultz, 2006; Serrano, 2010; Tang, 2009) b. (Miyake-Stoner, 2009) c. (S. Chen, 2013; Lampkowski, 2015) d. (Brun, 2004; J. Wang, 2006) e. (Summerer, 2006) f. (Chatterjee, 2013; Lee, 2009; Mitchell, 2017) g. (Hamada, 2005; Speight, 2013; Taki, 2010)					

genetic incorporation of Anap in a voltage-gated potassium channel allowed researchers to monitor, for the first time, depolarization-dependent rearrangements on the previously-inaccessible cytosolic side of the channel using voltage-clamp fluorometry in *Xenopus* cells (Kalstrup, 2013). Recently, these probes have even gained traction in the subcellular localization of proteins in mammalian cells (Chatterjee, 2013).

1.8 An improved fluorescent unnatural amino acid

The above unnatural amino acids have clear value as alternatives to conventional protein labeling with fluorescent proteins or exogenous dyes. However, several disadvantages inherent to these unnatural amino acids have likely constrained their use as routine fluorescent probes. While sensitivity to changes in solvent can be a desirable property (Lakowicz, 2006), interpreting fluorescent signal changes in these experiments can be tricky, especially when the fluorophore, such as Hco, is sensitive not only to polarity but pH changes. Additionally, tradeoffs typically exist between the size of these fluorescent unnatural amino acids and their spectral properties; using smaller labels usually results in decreased brightness and blueshifted excitation and emission spectra. As a result, the sensitivity and specificity of these can probes suffer, especially in contexts where intrinsic fluorescence from Trp and Tyr residues may be a contributing factor. Resistance to photobleaching can also limit the utility of some of these labels, including Hco, during extended and continuous fluorescence measurements.

Previously, synthesis of the fluorescent unnatural amino acid L-2-acridonylalanine (Acd) yielded a fluorescent probe with appealing properties (Szymańska, 2003). Similar in size to the aforementioned fluorescent amino acids, Acd boasts a much higher quantum yield in aqueous solution ($\Phi = 0.95$), possesses a red-shifted excitation maximum (λ_{ex}), is insensitive to pH, and demonstrates robust photostability (<5% decrease in 60 min) (Hamada, 2005). Notably, Acd also contains a substantially longer fluorescent lifetime ($\tau = 15$ ns) than many commercially available fluorophores (usually $\tau < 4$ ns). Given these improved fluorescent and photophysical properties, the Petersson laboratory first devised an efficient route to synthesize Acd (Speight, 2013). Next, in

collaboration with Mehl and coworkers, they demonstrated that Acd could be incorporated into proteins in *E. coli* using an evolved *M. jannaschii* tyrosyl tRNA/tRNA synthetase pair known for broad substrate specificity (Cooley, 2014; Speight, 2013; Stokes, 2009). While incorporation of Acd was robust and reproducible using this permissive synthetase, the relatively indiscriminate substrate specificity of this enzyme also resulted in the unwanted and favorable incorporation of N-phenyl-p-aminophenylalanine (Npf), a trace byproduct of the Acd synthetic route. To overcome this issue, Petersson and colleagues pursued an alternative synthetic strategy that eliminated the contaminating Npf byproduct, but at the cost of lower yields of racemic Acd.

Several studies have demonstrated the utility of Acd as a probe of protein dynamics. Sisido and coworkers first explored the environmental sensitivity of Acd fluorescence, which blueshifts slightly and decreases as solvent polarity increases (Hamada, 2005). Taking advantage of this property, they attempted to measure the binding of biotin to Acd-labeled streptavidin through changes in Acd fluorescence. Of the three positions they chose for Acd incorporation in streptavidin, only two—positions 83 and 120—yielded decreases in fluorescence on binding to biotin. The third mutant, with Acd incorporated at position 22, showed only a minor decrease in fluorescence, highlighting not only the exquisite sensitivity of Acd to local environment but also the challenge in identifying positions suitable for reporting on a biophysical event. The Petersson laboratory has also leveraged features of Acd fluorescence to study protein dynamics (Speight, 2013). Incorporation of Acd in the calcium-binding protein calmodulin provided a means to monitor binding to the peptide ligand pOCNC. By synthesizing this peptide with a 7-methoxycoumarin label (Mcm), they were able to monitor increases in energy transfer from the FRET donor Mcm to the FRET acceptor Acd as more labeled pOCNC was added. Continued efforts to refine Acd as a FRET label have resulted in the recently reported site-specific installation of two (Mcm and Acd) or three (Trp, Mcm, and Acd) FRET labels within recombinant protein complexes (Ferrie, 2017). Additionally, the Petersson group explored the possibility of monitoring the urea-dependent unfolding of triose phosphate isomerase by exploiting a short-distance quenching phenomenon

that occurs when Acd is very close to Tyr or Trp. By engineering Acd in close proximity to a native Trp residue in the folded protein structure, they observed Trp-dependent quenching of Acd as a reduction in the measured fluorescence lifetime of Acd ($\tau \sim 4\text{--}5$ ns). Addition of urea disrupted this quenching interaction, restoring the measured fluorescence lifetime of Acd to its unusually long value in water ($\tau \sim 15$ ns). Further, Acd fluorescence can also be quenched by small fluorescence-quenching probes, such as thioamides, expanding the possible applications for this fluorescent unnatural amino acid (Goldberg, 2013). Given its ability to sensitively report on conformational changes and macromolecular binding events, Acd appears well-suited for the task of probing protein dynamics.

1.9 Thesis Objectives

Genetically encoded fluorescent amino acids are valuable labeling tools with an established record of uncovering the details of protein structure and function in biologic systems recalcitrant to conventional methodologies. In this body of work, I present our attempts to refine the technology and strategy underlying efficient *in vivo* incorporation of one of these fluorescent amino acids, Acd, into target proteins. Given the potential advantages of Acd as a fluorescent probe, we also describe our successful use of Acd in studying the protein dynamics underlying the activation of the SOS response, a key pathway involved in antibiotic evasion. In Chapter 2, I describe an investigation into one consequence of relying on a permissive tRNA synthetase for unnatural amino acid mutagenesis—the unwanted incorporation of trace byproducts—and offer a general counter-selection strategy for the evolution of synthetases that can discriminate against these byproducts. In Chapter 3, I document our systematic efforts to identify amino acid features that correlate with increased tolerability to unnatural amino acid mutagenesis and provide justifications for why no clear predictors emerge from this study. In Chapter 4, I detail how we leveraged our methodologic findings from Chapters 2 and 3 to site-specifically incorporate Acd into tolerant positions in LexA, generating fully-labeled fluorescent LexA constructs. I then discuss how these fluorescent LexA proteins are used to investigate the first step of the RecA*-mediated LexA autoproteolysis reaction:

the binding of LexA to RecA^{*}. By collecting high resolution kinetic data of this interaction, we provide quantitative insights into the strength and mode of LexA:RecA^{*} binding. Together, these insights not only build a mechanistic foundation for future studies on the protein dynamics regulating the SOS response in bacteria but also establish a methodologic framework for using Acd as a fluorescent probe of protein-protein interactions more broadly.

Chapter 2: Improving Target Amino Acid Selectivity in a Permissive Aminoacyl tRNA Synthetase Through Counter-Selection

The contents of this chapter have been published:

Sungwienwong, I.;* Hostetler, Z. M.;* Blizzard, R. J.; Porter, J. J.; Driggers, C. M.; Mbengi, L. Z.; Villegas, J. A.; Speight, L. C.; Saven, J. G.; Perona, J. J.; Kohli, R. M.; Mehl, R. A.; and Petersson, E. J. Improving Target Amino Acid Selectivity in a Permissive Aminoacyl tRNA Synthetase through Counter-Selection. *Org. Biomol. Chem.* **2017**, 15 (17), 3603–3610.

* These authors contributed equally to this work.

2.1 Abstract

The amino acid acridon-2-ylalanine (Acd) can be a valuable probe of protein dynamics, either alone or as part of a Förster resonance energy transfer (FRET) or photo-induced electron transfer (eT) probe pair. We have previously reported the genetic incorporation of Acd by an aminoacyl tRNA synthetase (RS). However, this RS, developed from a library of permissive RSs, also incorporates *N*-phenyl-amino-phenylalanine (Npf), a trace byproduct of one Acd synthetic route. We have performed negative selections in the presence of Npf and analyzed the selectivity of the resulting AcdRSs by *in vivo* protein expression and detailed kinetic analyses of the purified RSs. We find that selection conferred a ~50-fold increase in selectivity for Acd over Npf, eliminating incorporation of Npf contaminants, and allowing one to use a high yielding Acd synthetic route for improved overall expression of Acd-containing proteins. More generally, our report also provides a cautionary tale on the use of permissive RSs, as well as a strategy for improving selectivity for the target amino acid.

2.2 Introduction

It is now well-established that protein folding and dynamics play essential roles in health and disease. For example, the small protein calmodulin (CaM) undergoes a dramatic conformational rearrangement to carry out its calcium sensor function in eukaryotic cells (Chin, 2000; Vetter, 2003). In a second example, the bacterial repressor-protease LexA uses a complex

sequence of RecA-induced structural change, self-proteolysis, and dissociation of subunits to sense DNA damage and activate genes that ultimately lead to antibiotic resistance (Butala, 2009; Culyba, 2015). Finally, the conformational flexibility of the neuronal protein α -synuclein (α S) is a liability, as it leads α S to misfold and form amyloid fibrils that contribute to the pathogenesis of Parkinson's disease (Lashuel, 2013; Spillantini, 1997). Fluorescence spectroscopy is a powerful tool for studying such processes, as it allows one to observe protein motions in real time under physiological conditions, including measurements in live cells (Giepmans, 2006; Haney, 2015). One can even obtain low resolution structural information using distance-dependent chromophore interactions such as Förster resonance energy transfer (FRET) and quenching by photo-induced electron transfer (eT) (Lakowicz, 2006). To appropriately model protein motions, one needs a set of probes that are capable of accurately reporting on distance changes without disrupting the fold and function of the protein of interest (Speight, 2014).

Recent developments in genetic code expansion and biorthogonal chemistry have made the site-specific incorporation of non-canonical amino acids (ncAAs) and installation of fluorophores through post-translational modification straightforward, even in cells or lysates (Lang, 2014). Unfortunately, the fluorophores used in these approaches are often relatively large and are attached by long flexible linkers, which have a non-trivial impact on the observed FRET measurements. Additionally, larger fluorophores cannot be introduced co-translationally to allow packing into the interior of a folded protein. Thus, they will be restricted to surface-accessible positions, limiting the regions of the protein for which conformational changes can be studied.

Smaller probes that are more closely tied to the backbone are better able to report on conformational changes of the protein. The Petersson laboratory has developed small fluorescent probes and quenchers that should be non-perturbing to proteins and which are closely tied to the protein backbone. These include thioamide substitutions of the backbone itself and intrinsically fluorescent ncAAs such as acridon-2-ylalanine (Acd or **1**) (Speight, 2013). These new approaches are significant because, of the >100 ncAAs that have been genetically encoded in *E. coli*, only four

are fluorescent (Chatterjee, 2013; Dumas, 2015; Harkiss, 2016; Lee, 2009; J. Luo, 2014; Speight, 2013; Summerer, 2006; J. Wang, 2006).

Acd is a blue-wavelength fluorescent amino acid that is a useful fluorophore because of its small size (222 Å³), near unity quantum yield in water ($\phi = 0.95$), unusually long lifetime ($\tau \sim 15$ ns) and high photostability (<5% degradation after 3 h irradiation) (Hamada, 2005; Szymańska, 2003; Taki, 2010). Previous work in the Petersson laboratory has shown that Acd can be efficiently quenched by a thioamide through an eT mechanism (Goldberg, 2012). We have also shown that it can be a valuable intermolecular FRET acceptor from Trp or methoxycoumarin, and a donor to more red-shifted dyes such as nitrobenzoxadiazole (NBD) or fluorescein (Speight, 2013). Further examples of the utility of Acd in intramolecular FRET with Mcm and in fluorescence polarization studies of protein binding are shown in Supplemental Figure 8 and Supplemental Figure 9.

Previously, the Mehl and Petersson laboratories developed an *in vivo* system for Acd incorporation using methods pioneered by Schultz (L. Wang, 2001). These methods require the generation of an aminoacyl tRNA synthetase (RS) that is selective for the ncAA and a tRNA that can be selectively charged by the ncAARS to deliver the tRNA to an unassigned codon, typically the amber stop codon UAG (tRNA_{CUA}) (Lei Wang, 2004). An RS was selected from a library of permissive *M. janaschii* (*Mj*) tyrosyl RS mutants that had previously been shown to incorporate bulky aromatic amino acids such as 4-(2'-bromoisobutyramido) phenylalanine (Brb) and *p*-benzoyl phenylalanine (Miyake-Stoner, 2010; Peeler, 2010; Stokes, 2009). The most active mutant from this library (G2) was used to express Acd-containing variants of CaM, triose phosphate isomerase, and α S (Speight, 2013). Here, we will refer to this mutant as AcdRS1.

At the same time, we also developed an efficient synthetic route, starting from the natural amino acid Tyr (**2**) and using a Buchwald-Hartwig coupling to O-methyl anthranilate (**4**) to form **5** (Speight, 2013) (See Figure 2.1). We obtained an 86% overall yield from Tyr using the route shown in Scheme 1 with H₂SO₄ used in the Friedel-Crafts cyclization step (no racemization observed by high performance liquid chromatography, HPLC). Unfortunately, this route also produced a

decarbonylated by-product, *N*-phenyl-aminophenyl alanine (Npf or **6**) in trace amounts. We found that Npf was in fact incorporated much more efficiently by AcdRS1 than Acd. We were able to initially address this problem by converting **5** to Acd using LiOH deprotection followed by PPA cyclization. This eliminated Npf formation but limited us to a 44% overall yield from Tyr with complete racemization (a 22% yield of the requisite L-Acd form). While we were able to express proteins containing exclusively Acd, this was not a very satisfactory solution, and we sought to obtain an AcdRS that was sufficiently selective against Npf that we could use the higher yielding H₂SO₄ synthetic route.

A typical selection experiment to evolve RSs for ncAAs consists of rounds of positive selection (performed in the presence of the ncAA and the 20 canonical amino acids) and negative selection (performed in the presence of only the 20 canonical amino acids). In this work, we sought to obtain an Acd-selective RS by performing these standard selections, as well as negative selections where Npf was included in the selection media to eliminate those AcdRS mutants charging tRNA with Npf. This resulted in two additional RSs, AcdRS2a (clone G11, from traditional negative selection) and AcdRS2b (clone A9, from Npf counter-selection), which both showed good

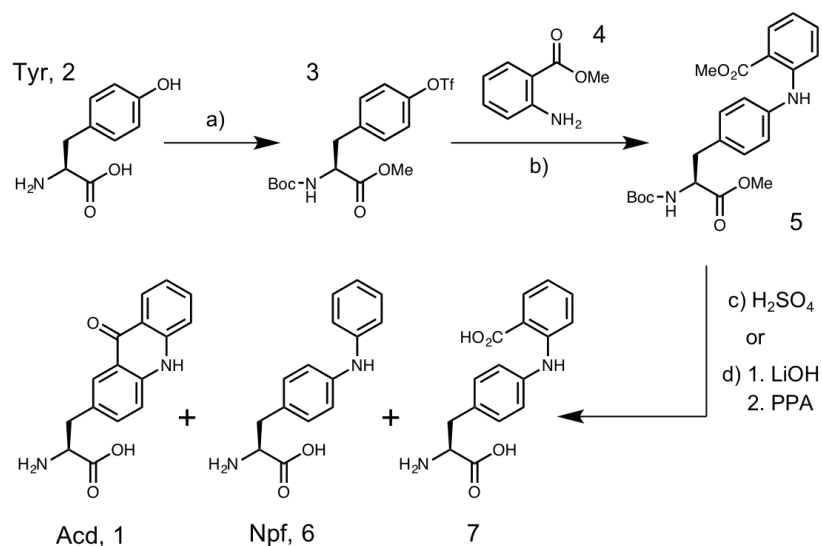


Figure 2.1. Synthesis of Acd by H₂SO₄ or Polyphosphoric Acid (PPA) Routes. (a) 1. SOCl₂, MeOH 2. Boc₂O, Na₂CO₃, THF/H₂O (3:1) 3. PhNTf₂, DMAP, NEt₃, CH₂Cl₂; 88% yield over 3 steps; b) Pd(OAc)₂, *rac*-BINAP, Cs₂CO₃; 92% yield; c) H₂SO₄, reflux; 98% yield; d) 1. LiOH, 2. PPA; 53% yield over two steps (racemic mixture)

selectivity in an initial screen in which the ncAAs were incorporated into green fluorescent protein (GFP).

We have performed detailed studies of their *in vivo* selectivities in expressions of CaM, α S, and LexA and found that AcdRS2b has superior selectivity for Acd when compared to AcdRS2a. We have also expressed and purified AcdRS1 and AcdRS2b to measure their Acd and Npf activation kinetics. We are able to rationalize their selectivities in terms of the X-ray crystal structure of AcdRS1 and a homology model of AcdRS2b. Our study provides an improved Acd incorporation method for fluorescent labeling of proteins, and it also validates a general strategy for how one may optimize a permissive RS to eliminate incorporation of an unwanted contaminant.

2.3 Results and Discussion

CaM, LexA and α S constructs with UAG mutations were expressed in *E. coli* along with plasmids encoding AcdRS1 and its cognate tRNA_{CUA} species. We analyzed the selectivity of AcdRS1 based on matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) data of both intact and trypsin-digested, purified proteins, including CaM, LexA, and α S. Data for incorporation at position 113 in CaM are shown in Figure 2.3 and Figure 2.4; additional data for LexA and α S are shown in the Appendix. In all cases, when the proteins are expressed using media containing pure Npf or Acd produced using the PPA route, a single peak for the intended product is obtained in the MALDI spectra. When the proteins are expressed in media containing Acd produced using the H₂SO₄ route, we observe a roughly 45:55 Npf/Acd ratio of the two CaM species, even though Npf is present only in trace quantities. We also considered the possibility that Npf could be generated *in vivo* by several possible routes. For example, a carboxy-lyase such as YigC could convert **7** (also a <1% contaminant in the H₂SO₄ Acd synthesis) into Npf (Díaz, 2001), or some fraction of Acd could be converted to Npf. Our PPA Acd expression data allow us to exclude metabolic processing of Acd, since we see no Npf incorporation in this case. When we use media in which we intentionally include 1% or 10% Npf with Acd, we detect products containing Npf/Acd ratios of 65:35 and >95:5, respectively (Supplemental Figure 2). These data and similar data for

other proteins (see Appendix) indicated to us that the selectivity of our nominal AcdRS—designed to incorporate Acd specifically—in fact favored Npf incorporation by approximately 100-fold. Despite the incongruity between its intended use and its actual fidelity, one should keep in mind that AcdRS1 was selected based on expression yields of GFP in the presence of media containing Acd from the H₂SO₄ route (and therefore containing ~1% Npf). Similar contamination, barely detectable by HPLC analysis, may be present in other stocks of home-synthesized or commercial batches of ncAAs used in RS selection.

2.3.1 AcdRS Optimization

RS Selection to Remove Npf Activity. To reduce the incorporation of Npf, we screened mutant *Mj* TyrRS libraries with positions in the amino acid binding pocket randomized using a GFP

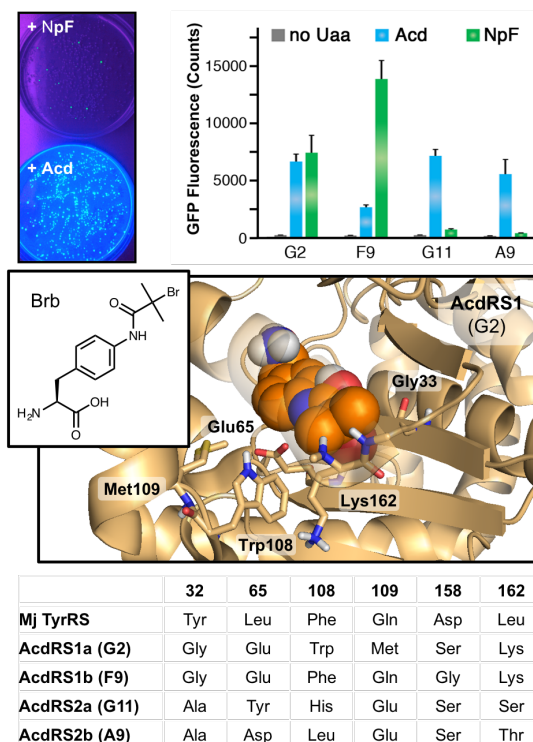


Figure 2.2. AcdRS Selection. Top Left: Images of *E. coli* agar plates used in rounds of positive (+ Acd) and negative (+ Npf) selection. Media for both plates also contain sources of the 20 canonical amino acids. Top Right: Fluorescence of suspensions of *E. coli* cells expressing GFP with a TAG codon at position 150 using the indicated RS clone and amino acid mixture. Emission was measured at 528 nm with excitation at 485 nm. Middle: Image of the AcdRS1 (G2) active site with radical polymerization initiator Brb bound. (PDB ID: 4PBR) A favourable hydrogen bond between the carboxylate of Glu₆₅ and the aniline N-H of Brb can be seen. Bottom Sequences of *Mj* RS clones used for incorporation of Acd. Additional sequences of clones from GFP-based screening are given in the Appendix.

expression screen common to the Mehl laboratory (Cooley, 2014; Peeler, 2010). Two rounds of positive and negative selection were performed according to standard protocols, with the desired ncAA (i.e., Acd) present in the media for rounds of positive selection and only the 20 canonical amino acids present in the media for rounds of negative selection. In parallel, we also performed a similar two round selection experiment with a novel negative selection step that included the undesired Npf in the media. Thus, in this second “counter-selection” protocol, we explicitly selected against Npf incorporation rather than relying on high activity for Acd to be mutually exclusive of activity for Npf. The *Mj* RS libraries were based around the G2 (AcdRS1) and F9 clones, both of which incorporate Acd, and have very similar sequences. The level of Npf misincorporation is noticeably higher for F9 than for G2, highlighting the idea that small sequence changes can have a large impact on selectivity.

From these two selection protocols, eleven RS clones were identified that showed high levels of Acd incorporation. Of these clones, only A9 derived from the Npf counter-selection. Among the ten clones from the standard negative selection, the G11 clone demonstrated the highest level of selectivity as measured by the GFP fluorescence of cell suspensions expressed in media containing either 1 mM Acd or 1 mM Npf (Figure 2.2 and Supplemental Figure 1). Therefore, the G11 and A9 RSs were cloned into the pDule2 vector for expression of other proteins, and are referred to as AcdRS2a and AcdRS2b, respectively.

Comparison of AcdRS Active Sites. The availability of an X-ray crystal structure of AcdRS1 (G2) allowed us to examine the sequences of various library members to understand how changes might confer increased selectivity for Acd relative to Npf. Although we were never able to obtain suitably diffracting crystals of an Acd complex with AcdRS1, the Mehl group has previously published a structure of this RS with Brb bound (Figure 2.2, Middle). Recall that “AcdRS1” is a permissive RS, as can be seen in an analysis of its activation of a variety of aromatic ncAAs in Cooley *et al.* Brb, Npf, and Acd share the feature of a nitrogen atom in the *para* position of the phenylalanine ring, which seems to be a key recognition determinant as it can hydrogen bond with

Glu₆₅. Both of the RS clones that give the highest selectivity for Acd, AcdRS2a (G11) and AcdRS2b (A9), have a mutation at this position, but one that maintains the potential hydrogen bond acceptor functionality. Other library members with Pro, Trp, Ile, or Val at position 65 had less than ten-fold selectivity for Acd over Npf in the GFP assay (See Appendix for details). Further analysis of AcdRS1 (G2), AcdRS2a (G11), and AcdRS2b (A9) selectivity based on homology modelling and docking studies is given below.

2.3.2 Analysis of AcdRS Selectivity

In Vivo Characterization of AcdRS Selectivity. To more rigorously investigate the selectivity of AcdRSs 2a and 2b, we expressed proteins in *E. coli*, purified them, and analyzed ncAA incorporation by MALDI MS of whole proteins and trypsin digests. As above, each experiment was carried out under five media conditions, varying in the amino acid provided and/or the synthetic route by which it was obtained: with H₂SO₄ Acd, PPA Acd, Npf, 1% Npf/Acd, and 10% Npf/Acd. Data for incorporation at position 113 in CaM are shown in Figure 2.3 and Figure 2.4; data for αS and LexA are shown in the Appendix. As anticipated from the GFP screening data, we found that both AcdRS2a and 2b have improved selectivity against Npf. However, AcdRS2a still gives a mixture of Npf- and Acd-containing protein when H₂SO₄ Acd is used in the growth media. In contrast, proteins expressed using AcdRS2b contain only Acd, even when challenged with 10% Npf in the media. This level of selectivity was observed in CaM, αS, and LexA (Figure 2.3,

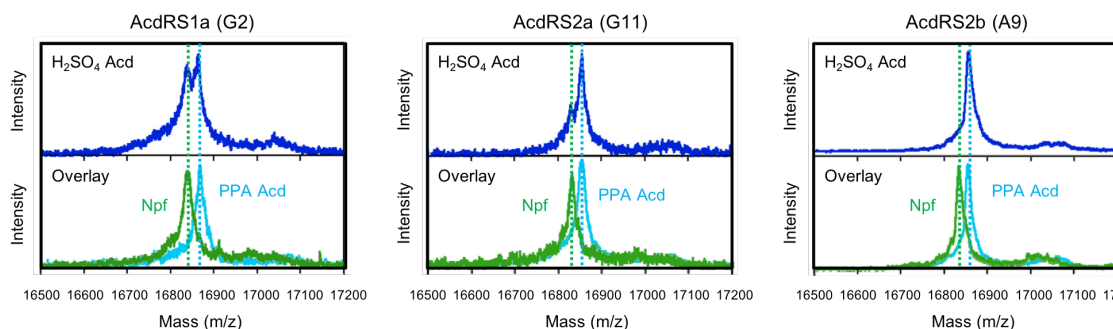


Figure 2.3. In Vivo AcdRS Selectivity. CaM (UAG codon at 113) was expressed in minimal media containing 1 mM ncAA: Acd, synthesized either using the H₂SO₄ route or PPA route, or Npf. Expression was performed with one of the three AcdRSs indicated. Significant incorporation of Npf is seen for AcdRS2a (G11) when using H₂SO₄ Acd, but only Acd-containing protein is seen with AcdRS2b (A9). Experimental details are given in the Appendix.

Supplemental Figure 2, Supplemental Figure 3, and Supplemental Figure 4). It should also be noted that AcdRS2b selectivity is observed in a variety of media. CaM and α S are expressed in minimal media or Luria broth (LB) with isopropyl β -D-1-thiogalactopyranoside (IPTG) induction, while LexA is expressed in richer, lactose auto-induction media. Based on these data, we selected the AcdRS2b (A9) for further characterization.

For a more rigorous, quantitative analysis of selectivity, we analyzed the CaM trypsin digest data by normalizing the intensity of the peak for the 108-116 fragment, containing Acd or Npf at position 113, to the intensity of the peak for the 117-127 fragment (Figure 2.4). CaM₁₁₇₋₁₂₇ should be produced in a 1:1 ratio with CaM₁₀₈₋₁₁₆ when the protein is completely digested by trypsin, and this can be confirmed by varying the digest time and observing that the intensity ratios do not change (data not shown). Normalization using an internal standard is essential to interpreting the intensity data correctly. The CaM₁₀₈₋₁₁₆Acd₁₁₃ fragment (**A** in Figure 2.4) ionizes 5.5-fold better than the CaM₁₁₇₋₁₂₇ fragment (***** in Figure 2.4), while the CaM₁₀₈₋₁₁₆Npf₁₁₃ fragment (**N** in Figure 2.4) ionizes 6.1-fold worse than the CaM₁₁₇₋₁₂₇ fragment. This can be seen by examining the peak ratios for PPA Acd (i.e., Acd only) and Npf expressions. After peak scaling, one obtains an MS-based Acd/Npf selectivity ratio (MS Sel) of 2.6×10^{-3} for AcdRS1 and 0.20 for AcdRS2b, calculated as follows:

$$MS\ Sel = (Scaled\ Acd / Scaled\ Npf) / (Acd/Npf\ ratio\ in\ media)$$

The AcdRS1 MS Sel value is in good agreement with our estimate of 100-fold selectivity for Npf based on the less quantitative whole protein MALDI MS data. While the AcdRS2b MS Sel value may seem surprisingly low given the absence of any obvious CaM₁₀₈₋₁₁₆Npf₁₁₃ peak in the MALDI spectra, it is important to keep in mind that the RS is only being challenged with at most 10% Npf in the media and that CaM₁₀₈₋₁₁₆Npf₁₁₃ ionizes 30-fold worse than CaM₁₀₈₋₁₁₆Acd₁₁₃. Using these Acd/Npf selectivity ratios, we determined that selection resulted in a 76-fold improvement in

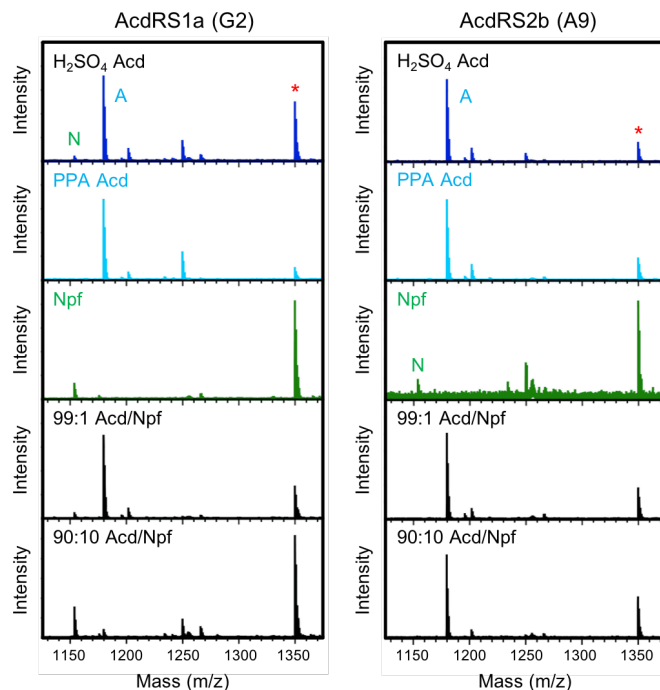
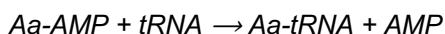
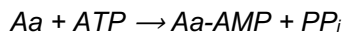


Figure 2.4. CaM₁₁₃ Typsin Digest Data for AcRS Selectivity Analysis. CaM (UAG codon at 113) was expressed in minimal media containing 1 mM ncAA: Acd, synthesized either using the H₂SO₄ route or PPA route, Npf, or either a 99:1 or 90:10 mixture of PPA Acd and Npf. Expression was performed with one of the three AcRSs indicated. The peaks for the (M+H)⁺ masses of the CaM₁₀₈₋₁₁₆Acd₁₁₃ fragment (A, 1179.8 Da), CaM₁₀₈₋₁₁₆Npf₁₁₃ fragment (N, 1153.8 Da), and CaM₁₁₇₋₁₂₇ fragment (*, 1349.9 Da) are indicated. The intensities of the CaM₁₀₈₋₁₁₆Acd₁₁₃ and CaM₁₀₈₋₁₁₆Npf₁₁₃ fragment peaks were normalized using the intensity of the CaM₁₁₇₋₁₂₇ fragment peak.

selectivity for AcRS2b relative to AcRS1. We note that the effective *in vivo* Acd selectivities for both AcRS1 and AcRS2b are less than one and are probably influenced by preferential uptake of Npf into cells.

***In Vitro* Characterization of RS Activity.** To better understand AcRS1 and 2b selectivity, we generated His-tagged variants of the enzymes, then expressed and purified them for *in vitro* activity assays. Charging of tRNA by RSs is a two-step process, where the first step is the RS-catalyzed reaction of amino acid (Aa) with ATP to form an aminoacyl-adenylate intermediate (Aa-AMP), releasing inorganic pyrophosphate (PP_i); and the second step is the reaction of this enzyme-bound adenylate with the 2' or 3' hydroxyl group on A76 at the 3' end of the tRNA (Ibba, 2000).



While some prior studies have used assays that measure only the first step of the aminoacylation reaction to demonstrate that *in vitro* activities are consistent with ncAA incorporation (Francklyn, 2008; Nehring, 2012; L. Wang, 2001), it has been shown by the Perona laboratory that full tRNA aminoacylation assays correlate well with *in vivo* observations of RS activity (Rauch, 2016). This is expected, since amino acid incorporation into protein *in vivo* can only occur upon aminoacylation. Thus, aminoacylation of an *in vitro* transcribed ^{32}P -labeled tRNA_{CUA} was measured under single-turnover conditions at a variety of Acd or Npf concentrations for AcdRS1 and AcdRS2b. Plots of the first-order rate constants as a function of amino acid concentration were used to determine k_{obs} and K_{d} for each amino acid and RS combination (Figure 2.5 and Table 2.1), as previously described (Rodríguez-Hernández, 2011). For these measurements, k_{obs} corresponds to the microscopic rate constant for the chemical steps of aminoacylation, or to a closely linked first-order rearrangement that follows tRNA binding and precedes aminoacylation.

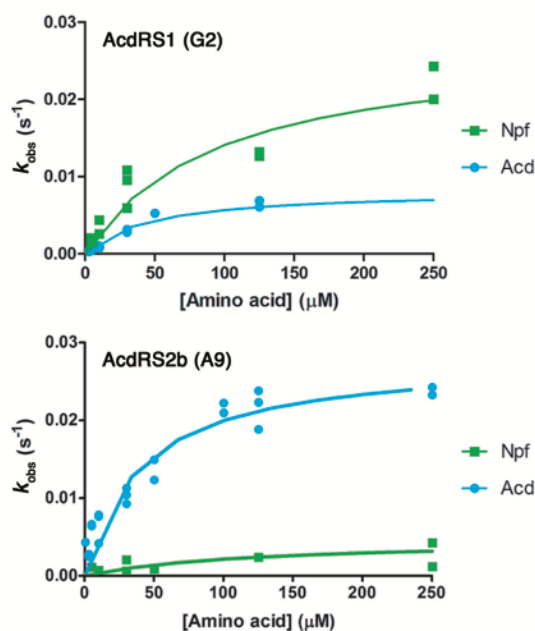


Figure 2.5. tRNA Aminoacylation Kinetics. Plot of k_{obs} for reactions of AcdRS1 and AcdRS2b as a function of Acd or Npf concentration. Reaction rates were determined using thin layer chromatography following nuclease digestion of aminoacylated tRNA as described in the Appendix. Sample primary data are shown in Fig. S7. Rate constants were determined from progress curves for each enzyme at varying amino acid concentrations.

Comparative kinetic analysis of AcdRS1 and AcdRS2b reveal catalytic preferences that correlate well with measurements of amino acid incorporation *in vivo* and mass spectrometry of proteins incorporating Acd and Npf. AcdRS1 aminoacylates tRNA_{CUA} with Acd three-fold slower than with Npf, while AcdRS2b aminoacylates tRNA_{CUA} with Acd five-fold faster than with Npf (Table 2.1). These changes in k_{obs} are primarily responsible for the change in selectivity between the RSs, as there are only relatively minor differences in K_d for the ncAAs. The enzymological selectivity (Enz Sel) calculated as

$$Enz\ Sel = [k_{obs}(Acd)/K_d(Acd)] / [k_{obs}(Npf)/K_d(Npf)]$$

changes from about a two-fold Npf preference by AcdRS1 to 16-fold Acd selectivity for AcdRS2b. Overall, AcdRS2b is improved in selectivity for Acd relative to Npf by 34-fold, in reasonable agreement with the 76-fold specificity shift derived from the *in vivo* measurements made based on trypsin digest peak intensities. Note that while the relative change in selectivity is consistent between the *in vitro* and *in vivo* data, the actual selectivities are quite different, presumably because the *in vivo* MS Sel is the product of Enz Sel and factors affecting the availabilities of the amino acids inside the cells. We estimate that the effective Npf concentration in cells is about 100-fold greater than the Acd concentration. This could be the result of preferential uptake of Npf, or sequestration of Acd by binding to other targets. Such targets could include DNA or RNA, both of which are known targets of acridone-based intercalator molecules (Belmont, 2007). Thus, to prevent Npf misincorporation, AcdRS2b must be significantly Acd-selective, even though Npf contaminants are only present at about 1% (10 μ M) in the media.

Table 2.1. AcdRS *In Vitro* Enzymology and *In Vivo* MS Selectivity Parameters

AcdRS	$k_{obs}(Acd)$ $\times 10^{-3} \text{ s}^{-1}$	$K_d(Acd)$ μM	$k_{obs}(Npf)$ $\times 10^{-3} \text{ s}^{-1}$	$K_d(Npf)$ μM	Enz Sel	MS Sel
1 (G2)	8.0 ± 2.0	45.1 ± 18.8	24.0 ± 1.0	64.1 ± 10.8	0.47	2.6×10^{-3}
2b (A9)	28.0 ± 2.0	39.9 ± 6.8	5.0 ± 1.0	114 ± 10.0	16	0.20

Modeling of AcdRSs complexed with Npf and Acd. To inform the differences in Acd and Npf charging activity, models were built based upon the crystallographic structure of AcdRS1 complexed with Brb (PDB ID: 4PBR) (Cooley, 2014). In generating a model of AcdRS1 (G2), the phenyl ring on the Brb ligand guided the placement of Acd and Npf within the binding pocket Figure 2.6. Hydrogen atoms were added to the protein, and the ligand structure was energy minimized within the fixed protein structure. For Acd, a void volume remains between the molecule and the G₃₂-G₃₄ β -strand. In each G2 complex, Glu₆₅ is poised to accept a hydrogen-bond from the N-H of the ligand's side chain. For Npf, the additional free volume allows the molecule to relax to a twisted conformation, alleviating repulsive internal steric contacts between the ligand's two phenyl rings. Using the same protein structure, computational protein design methods developed in the Saven laboratory were used to generate a model of A9 (Bender, 2007; Calhoun, 2003; Fry, 2013; Kono, 2001), followed by energy minimization of the atomic coordinates of the ligands and the mutated

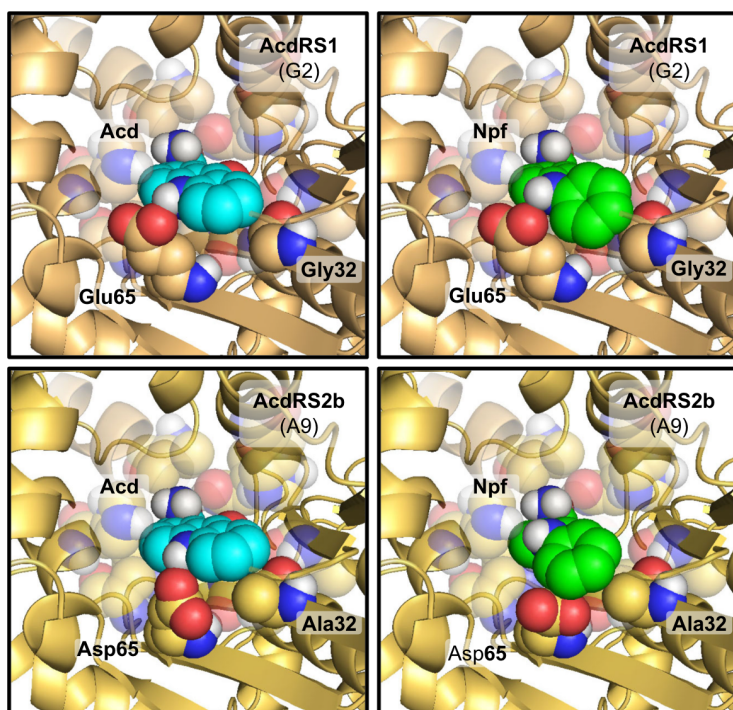


Figure 2.6. AcdRS Homology Models. Acd or Npf were energy minimized in the active site of AcdRS1 (G2), taken directly from PDB structure file 4PBR, or a model of AcdRS2b (A9). Npf is accommodated in the AcdRS1 active site where the phenyl rings can become non-planar with respect to each other. In the A9 active site, Npf twisting is restricted by the Ala₃₂ sidechain, and a hydrogen bond cannot be made with Asp₆₅.

side chains. In A9, the Gly₃₂Ala mutation forces Npf to adopt a more planar conformation of the ligand's two phenyl rings in order to form a hydrogen bond with Asp₆₅. Instead, Npf relaxes to avoid this high-energy distortion, which not only prevents hydrogen bonding of the Npf aniline N-H with Asp₆₅, but also forces the Asp₆₅ sidechain to move. While it is possible that the actual conformations adopted in the active site are different in the presence of tRNA, these models are consistent with AcdRS2b's increased activity toward Acd relative to Npf. Thus, it appears that the combination of the Gly₃₂Ala and Glu₆₅Asp mutations provides the increase in Acd selectivity for AcdRS2b (A9). Further experimental studies of single point mutations are required to validate this hypothesis and will be reported when complete.

2.4 Conclusions

Our elucidation of the process by which trace Npf in the Acd media was incorporated into expressed proteins and our subsequent optimization of AcdRS1 provide a cautionary tale to the field of genetic code expansion, as well as a potentially general solution to the problem. Explicit inclusion of the unwanted Npf in the counter-selection step allowed us to obtain AcdRS2b, which had much higher selectivity toward its cognate ncAA than AcdRS2a, obtained through standard rounds of positive and negative selection (against only the 20 canonical amino acids). Trace contaminants in batches of ncAAs, synthesized either in-house or bought from commercial vendors, may indeed be utilized by many RSs in the literature which have never been purposefully selected for or against charging of these ncAAs. Moreover, variations of *in vivo* availability (resulting from differential uptake, for example) may exacerbate these problems, as appears to be the case with Npf.

Permissive RSs in libraries, which provide a very efficient way of obtaining RSs for new ncAAs, are likely to be susceptible to this problem. Indeed, RSs with open binding sites that accept bulky amino acids such as Acd and Brb may be particularly prone to this difficulty. However, using such a permissive RS with an open pocket is sometimes necessary to finding an RS for an amino acid such as Acd. In some cases, RS selection for bulky ncAAs has required an initial round of

selection using a more moderate sized amino acid to expand the binding pocket before selection for charging of a bulky aromatic ncAA (Lee, 2009). Such RSs may be prone to misincorporation of amino acids similar to that intermediate amino acid. Here, we have seen that small sequence changes (i.e. Gly→Ala and Glu→Asp) can result in large increases in *in vitro* and *in vivo* selectivity for permissive RSs. Thus, we expect that the counter-selection strategy used here will be generally useful to the genetic code expansion community.

2.5 Contributions

I.S. and L.C.S. synthesized amino acids and performed *in vivo* protein expression and MALDI MS analysis. Z.M.H. cloned and expressed LexA and AcdRSs. R.J.B. and J.J.Po. performed AcdRS selections. C.M.D. and L.Z.M. performed AcdRS kinetic measurements. J.A.V. generated computational models. J.G.S., J.J.Pe., R.M.K., R.A.M., and E.J.P. supervised these studies.

2.6 Acknowledgements

This work was supported by funding from the National Science Foundation (NSF CHE-1150351 to E.J.P., NSF MCB-1518265 to R.A.M., NSF CHE-1508318 to J.G.S.), National Institutes of Health (NIH DP2-GM105444 to R.M.K.) and the Searle Scholars Program (10-SSP-214 to E.J.P.). Instruments supported by the National Science Foundation and National Institutes of Health include: LCMS/HRMS (NIH RR-023444), MALDI MS (NSF MRI-0820996), NMR (NIH RR-022442), and a stopped flow fluorometer (NSF CHE-1337449). J.G.S. acknowledges infrastructural support from the Penn Laboratory for Research on the Structure of Matter (NSF DMR 1120901) and use of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant ACI-1053575 under grant TG-CHE 110041. I.S. thanks the Royal Thai Foundation for fellowship support. Z.M.H. and J.A.V. thank the the NIH for funding through the Chemistry Biology Interface Training Program (T32 GM071399). J.A.V. and J.G.S. acknowledge assistance and preliminary computational studies from Lu Gao.

Chapter 3: Systematic evaluation of soluble protein expression using a fluorescent unnatural amino acid reveals no reliable predictors of tolerability

The contents of this chapter have been published:

Hostetler, Z. M.; Ferrie, J. J.; Bornstein, M. R.; Sungwienwong, I.; Petersson, E. J.; Kohli, R. M. Systematic evaluation of soluble protein expression using a fluorescent unnatural amino acid reveals no reliable predictors of tolerability. *ACS. Chem. Biol.* **2018**, 13 (10), 2855-2861.

3.1 Abstract

Improvements in genetic code expansion have made preparing proteins with diverse functional groups almost routine. Nonetheless, unnatural amino acids (Uaas) pose theoretical burdens on protein solubility, and determinants of position-specific tolerability to Uaas remain underexplored. To broadly examine associations, we systematically assessed the effect of substituting the fluorescent Uaa, acridonylalanine, at more than fifty chemically, evolutionarily, and structurally diverse residues in two bacterial proteins—LexA and RecA. Surprisingly, properties that ostensibly contribute to Uaa tolerability—like conservation, hydrophobicity, or accessibility—demonstrated no consistent correlations with resulting protein solubility. Instead, solubility closely depended on the location of the substitution within the overall tertiary structure, suggesting that intrinsic properties of protein domains, and not individual positions, are stronger determinants of Uaa tolerability. Consequently, those who seek to install Uaas in new target proteins should consider broadening, rather than narrowing, the types of residues screened for Uaa incorporation.

3.2 Introduction

Technological advances in genetic code expansion have encouraged the design of proteins with a wide range of reactive residues, post-translational modifications, photocaged groups, or intrinsic fluorophores (Neumann-Staubitz, 2016; Xiao, 2016; Young, 2010). Nonsense codon suppression using orthogonal tRNA/aminoacyl-tRNA synthetase pairs enables direct incorporation of chemically diverse unnatural amino acids (Uaas, also known as non-canonical amino acids) into proteins *in vivo*. Many efforts have sought to boost the efficiency of Uaa incorporation, including evolving more efficient aminoacyl-tRNA synthetases and recoding the *E.*

coli genome to remove competing translational release factors (Chatterjee, 2013; Lajoie, 2013). Although these developments can improve total yields of modified proteins, factors governing the position-dependent effects of Uaa substitution on protein solubility remain understudied.

Recent reports have demonstrated that the position of a Uaa can affect the level of total protein expressed, both in cell-free and cell-based systems (Albayrak, 2013; Goerke, 2009; Hamada, 2005; Hammill, 2007; Hino, 2006). Investigations of 20 positions in IFN- α and 33 positions in VSV glycoprotein revealed varying total protein yields, from 0 to 95% of wildtype (B. Zhang, 2015; Y. Zheng, 2015). Despite these observations, explanations for position-dependent differences in total amounts of Uaa-containing proteins have been limited, and no studies have explicitly addressed UAA incorporation versus the resulting protein solubility.

Unnatural amino acid mutagenesis could hypothetically operate under well-accepted principles that govern the effects of natural amino acid mutation. For example, substitution of a nonpolar for a polar residue within the hydrophobic core generally destabilizes proteins, whereas mutations on the solvent-exposed surface less frequently affect solubility (Lau, 1990; Markiewicz, 1994). Unsurprisingly, evolutionarily-conserved residues largely disfavor mutation (Campbell-Valois, 2005; Lim, 1989; Romero, 2015). Substituting bulkier and more chemically-diverse Uaas into a protein can restrict function (J. Luo, 2014) and therefore could pose similar burdens on folding and solubility. Nevertheless, the applicability of principles of natural amino acid mutagenesis to Uaa mutagenesis remains unknown.

Suggested guidelines or approaches for choosing Uaa-tolerant sites have been proposed. Some groups favor residues with structural similarity to the Uaa (Hammill, 2007). Others assert that candidate positions should be first assessed for mutational tolerability with natural amino acids (Hino, 2006) or that proteins should be thoroughly screened by random incorporation of Uaas into protein-GFP fusions to reveal positions that label with high efficiency (Arpino, 2015; Reddington, 2015). Nonetheless, the feasibility of using position-specific properties to increase soluble protein expression remains untested.

To address these open questions, we aimed to explore factors that impact Uaa incorporation and soluble protein production. By employing an intrinsically fluorescent Uaa, acridonylalanine (Hamada, 2005; Speight, 2013; Sungwienwong, 2017), we directly detect labeled protein in cell lysate samples, overcoming the inability of past studies to measure levels of both total and soluble expressed protein. Our systematic survey of more than fifty sites across two proteins reveals that while incorporation efficiency is relatively similar, protein solubility, and by extension Uaa tolerability, varies widely across different positions. However, most position-specific physicochemical, evolutionary, and structural properties, some of which have been previously suggested to improve yield, were minimally predictive; instead, solubility more strongly associated with the identity of the protein domain. After controlling for this domain effect, we found that only a few factors, such as a tolerance for aromatic residues, moderately trended with protein solubility. To our knowledge, this work currently represents the most systematic effort evaluating predictive factors for producing soluble Uaa-containing proteins.

3.3 Results and Discussion

The bacterial protein LexA, a multi-domain repressor of the DNA damage response, has characteristics that made it well-suited to this broader survey. Wild-type *E. coli* LexA is well-behaved in overexpression and has previously tolerated selective unnatural amino acid (Uaa) incorporation (Sungwienwong, 2017). Additionally, the availability of protein crystal structures and a multiple sequence alignment for LexA enabled retrieval of position-specific properties from databases or servers that require these data as inputs. For every position in LexA, we calculated established metrics across different classes of properties: physicochemical, such as hydrophobicity; evolutionary, such as conservation; and structural, such as solvent accessibility (Table 3.1). Using these metrics, we selected 32 positions spanning both domains of LexA, deliberately avoiding known deleterious mutants as well as the most conserved or hydrophobic positions (Figure 3.1a). Our selected positions sample the remaining metrics well (Figure 3.1b, Supplemental Figure 10, and Supplemental Figure 11), indicating that this series is well-positioned

to explore how aromatic, accessible, or poorly-conserved residues might differentially tolerate Uaa incorporation.

Historically, measuring Uaa incorporation efficiencies *in vivo* has overlooked protein solubility issues, while labeling Uaa-containing proteins *in vitro* has suffered from incomplete

Table 3.1: List of properties examined for association with Uaa tolerability

Property	Details
Physicochemical	
Hydrophobicity	Discrete number describing experimentally-determined hydrophobic indices (usually kcal/mol)
Similar to Phe, Trp, or Tyr	Discrete number calculated from a substitution matrix similarity score table
Volume	Size of residue (\AA^3)
Evolutionary	
Conservation	Calculated score describing the degree of conservation from a multiple sequence alignment
Tolerance to Phe, Trp, or Tyr	Presence or absence of a particular residue substitution within a multiple sequence alignment
Structural	
Solvent Accessible Area	Surface area of residue exposed to solvent (\AA^2)
Accessibility	Ratio of solvent accessible area relative to the theoretical maximum surface area of a residue
Fractional Loss of Accessible Area	Area lost when a residue is buried upon folding (\AA^2)
Surrounding Hydrophobicity	Numerical sum of local hydrophobic indices assigned to residues within 8 \AA
Average hydrophobic gain/ratio	Total increase or a ratio describing the difference in surrounding hydrophobicity between unfolded to folded state
Position	Residue number in primary sequence of protein
Secondary/tertiary structure	Categorical assignment to secondary structure type or classification into a protein domain
Nearby contacts	Discrete number of contacts within 8 or 14 \AA , either using C_α or C_β atoms
Noncovalent contacts	Presence or absence of interaction with another residue through a H-bond, cation- π , hydrophobic, or polar contact
Long Range Order	Presence or absence of contacts with residues close in space but far in sequence
Surrounding Residues	Number of residues within 8 \AA contextualized by sequence position

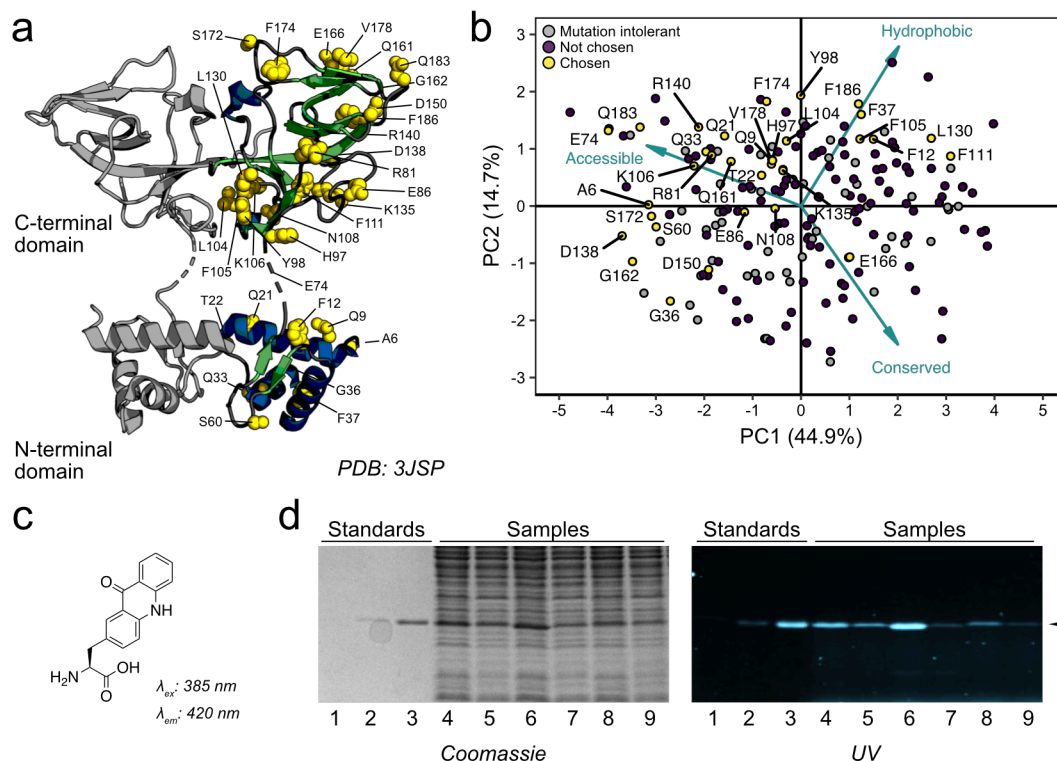


Figure 3.1: Scanning a variety of positions in LexA for Acrid tolerability. (a) Positions chosen for Uaa incorporation in the LexA dimer. Chosen positions are depicted in yellow, α -helices in blue, and β -sheets in green. (b) Principal component analysis (PCA) of LexA positions determined by multiple structural, evolutionary, and physicochemical properties (see methods). All residues in LexA were scored and plotted against the first two principal components, with positions chosen for Uaa incorporation highlighted in yellow. Arrow segments represent a few notable variables among those used in PCA loaded onto the plotted data. (c) Chemical structure of Acrid with indicated excitation and emission peaks. (d) Acrid-labeled LexA samples visualized in 15% SDS-PAGE gels by Coomassie staining (left) or UV excitation (right). Lanes 1–3 show purified LexA standards. Lanes 4–11 show paired total and soluble fractions from four individual mutants as representative examples.

sample recovery and detection. Crucially, we chose to measure both total and soluble protein levels by using the fluorescent Uaa acridonylalanine (Acrid, Figure 3.1c), which already possesses an optimized tRNA/tRNA synthetase pair for *in vivo* incorporation (Speight, 2013; Sungwienwong, 2017). This system offers several advantages. First, Acrid incorporation occurs during protein overexpression without post-translational labeling. Second, measurements of Acrid fluorescence at the expected size on an SDS-PAGE gel are directly proportional to levels of protein with successfully-incorporated Acrid. Finally, gel-based detection of Acrid demonstrates a broad dynamic range, enabling us to detect quantitative differences in the expression of Acrid-containing LexA mutants (Supplemental Figure 12).

Expression levels for a single protein can range widely due to experimental variability, making quantitative comparison between different proteins difficult. To overcome this challenge, we overexpressed the 32 LexA mutants in the presence of both Acd and the Acd-specific tRNA/tRNA synthetase using autoinduction media for consistency in the timing and duration of protein production. Following overexpression, we measured fluorescence intensity levels of Acd-containing LexA protein in both the whole cell lysate and soluble fraction (Figure 3.1d). The use of purified Acd-containing LexA as a standard enabled quantitative and reproducible comparisons of protein amounts across independent experiments (Supplemental Figure 13).

Parallel overexpression of all 32 LexA mutants allowed us to investigate how amounts of total expressed Acd-labeled LexA proteins differed (Supplemental Table 3). A plot of logarithmically-transformed total protein amounts shows uniformly high protein expression (mean = 3.1) with minor variability (SD = 0.16) (Figure 3.1a). While past studies have suggested that the identity of nucleotides surrounding the stop codon can impact nonsense codon suppression efficiencies (Miller, 1983; Pott, 2014; Xu, 2016), we did not observe this relationship (Supplemental Figure 14). Rather, the small 4.5-fold difference between measurements of the lowest and highest-expressing samples suggests that changing the position of Acd does not alter Acd incorporation rates *in vivo*, and that incorporation is not a bottleneck with regards to solubility.

Recognizing the consistency in total levels of expressed protein, we next evaluated whether levels of soluble protein differed. A distribution of logarithmically-transformed soluble protein amounts (Figure 3.2a) reveals more variability (mean = 2.2, SD = 0.86). Measured soluble protein amounts ranged nearly 40-fold from the lowest detectable measurements to the highest, a ten-fold increase over the range of total protein amounts. Because both measurements are paired, we can isolate the position-dependent effect of Acd incorporation on solubility by calculating the soluble fraction of total protein, which should exclude variability due to differences in total protein production. The soluble fractions of Acd-labeled LexA mutants still vary considerably, from 0% to nearly 70% of total protein expressed (Figure 3.2b, Supplemental Table 3). This result not only

corroborates previous observations of position-dependent effects on total protein expression (B. Zhang, 2015; Y. Zheng, 2015), but it also establishes the heightened sensitivity of protein solubility to Uaa incorporation.

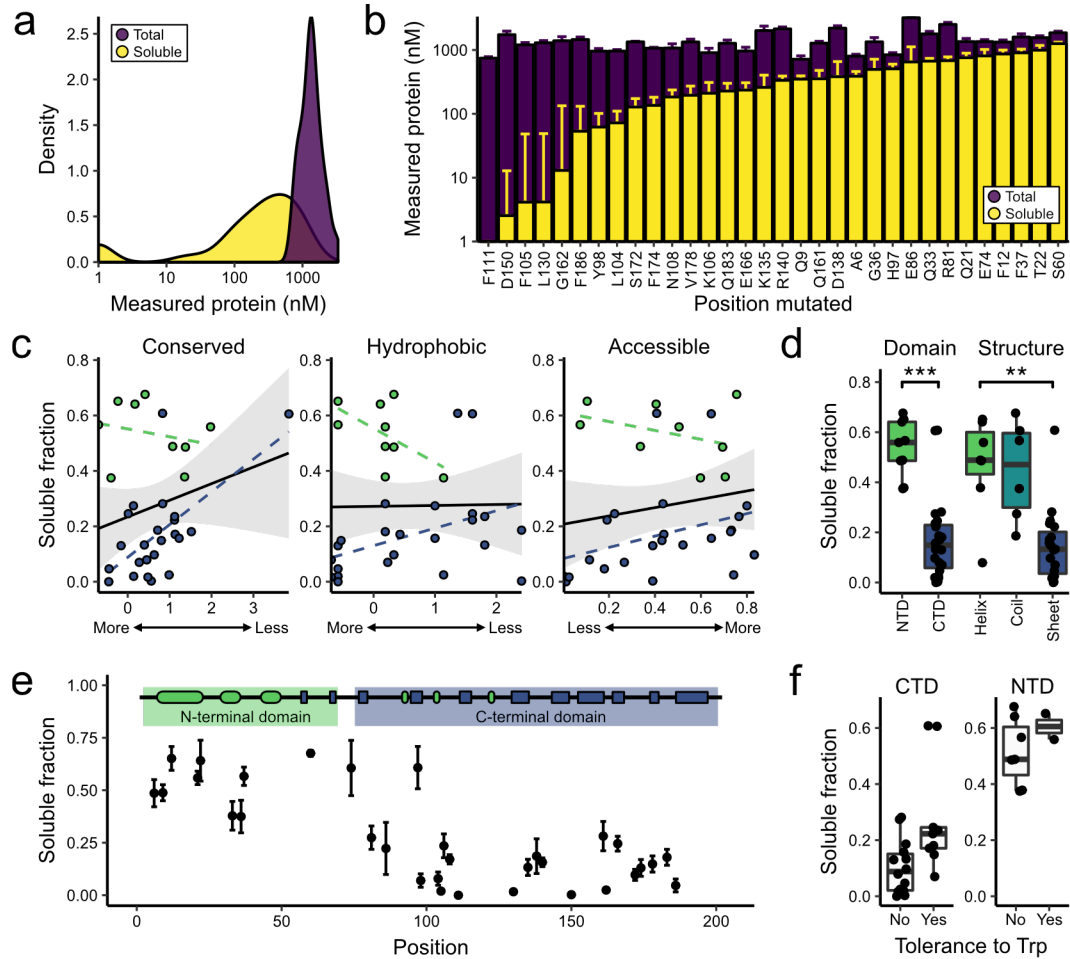


Figure 3.2: Features associated with soluble Acid-labeled LexA proteins. (a) Smoothed density plots of log10-transformed amounts of total protein or soluble protein. (b) Average log10-transformed soluble protein amounts overlaid on average log10-transformed total protein amounts for each mutant. Error bars indicate the standard deviation from three individual replicates each derived from separate clones. (c) Plots of the average fraction of soluble protein as a function of three selected parameters: conservation, hydrophobicity, and accessibility. Other parameters were also examined (Figures S6 and S7). Fits for the entire LexA dataset to individual linear regression models yield best fit lines (solid black) and 95% confidence intervals (shaded gray). Fits of data from each separate LexA domain yield best fit lines for the NTD (dashed green) or CTD (dashed blue). (d) Boxplots comparing the average fraction of soluble protein against either domain or secondary structure, with individual averages overlaid. Differences between groups were evaluated using Tukey's HSD test for multiple pairwise comparisons (** = p-value < 0.01; *** = p-value < 0.001). (e) Plot of the average fraction of soluble protein as a function of position in the LexA sequence, with error bars indicating the standard deviation from three replicates. Above, the secondary and tertiary structure of LexA is indicated; α -helices are depicted as green ovals and β -sheets as blue rectangles. (f) Separate boxplots for each LexA domain indicating the relationship between average fraction of soluble protein and evolutionary tolerance at each position to tryptophan, as one example of an aromatic residue.

Observing that the position of Acd can substantially impact protein solubility, we next asked which of the properties that ostensibly affect Uaa tolerability might correlate with solubility. We fitted the soluble fraction as a response variable to each property in individual linear regression models (Supplemental Table 4 and Supplemental Table 5). For almost all of the properties we evaluated, the explained variability (adj. R^2) was about 5% or less, indicating that if any property-specific effect exists, it is insubstantial and likely below our ability to detect with a sample size of 32 (Supplemental Figure 15 and Supplemental Figure 16). We note that particular properties—such as accessibility, conservation, and hydrophobicity—did not explain any substantial variation in our data, despite past suggestions that choosing accessible, less-conserved, and chemically-similar residues may yield more soluble Uaa-containing protein (Figure 3.2c).

Conspicuously, several highly-correlated properties each explained around 50% of the variability in our data, including individual residue position (adj. $R^2 = 0.53$), secondary structure (adj. $R^2 = 0.45$), and overall protein domain (adj. $R^2 = 0.53$) (Figure 3.2d and Figure 3.2e). Specifically, we obtained more soluble protein when Acd was incorporated within the first 74 residues of LexA, which includes all three of the α -helices that comprise the N-terminal domain. By contrast, Acd incorporation within the β -sheets of the C-terminal domain resulted in much lower proportions of soluble protein. The nearly uniform secondary structure composition of each domain limited our ability to interpret whether Acd tolerability is due to local secondary structure effects or global protein domain stabilities.

Excluding the effect that secondary or tertiary structure has on protein solubility could reveal minor trends obscured in the overall dataset. To address this possibility, in individual linear regression models, we fitted each property along with protein domain as explanatory factors for the soluble fraction (Supplemental Table 6 and Supplemental Table 7). By controlling for domain, we detected a minor correlation between the soluble fraction and the evolutionary tolerance of any given position to an aromatic residue (Figure 3.2f). However, remaining factors—including conservation, hydrophobicity, and accessibility—either did not explain any substantial variation in

the data or demonstrated inconsistent trends between domains (Figure 3.2c). Consequently, our extended LexA analysis reaffirmed that the tolerability of a protein domain to Acd—or possibly the tolerability of a secondary structure type—overwhelmingly determines protein solubility.

Studying Acd incorporation in a distinct protein scaffold with mixed α/β character could help dissect the similar effects we observed from the highly-correlated domain and secondary structure factors with LexA. Thus, we extended our survey to RecA, a bacterial ATPase that binds LexA to suppress its repressor function (Culyba, 2015). We selected positions in *E. coli* RecA that satisfied one or more criteria: high accessibility, low conservation, few inter-residue contacts, or prior functional tolerance to mutation (Figure 3.3a) (McGrew, 2003). After expressing these mutants with

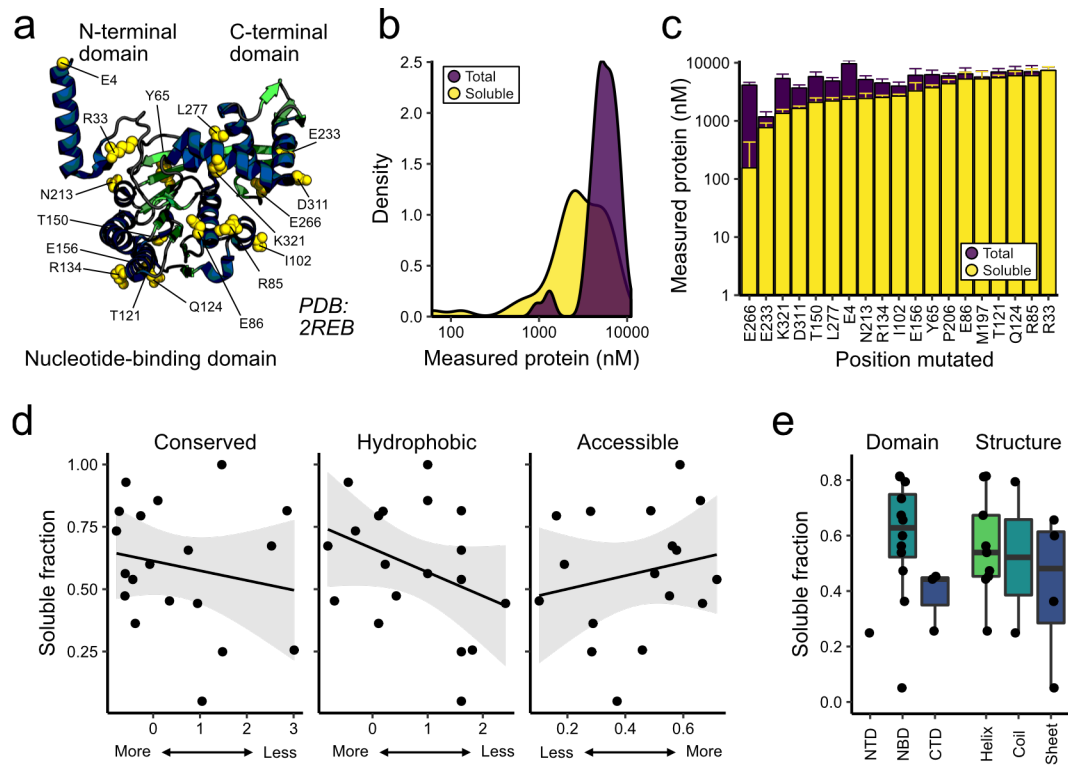


Figure 3.3: Features associated with soluble Acd-labeled RecA proteins. (a) Positions chosen for Acd incorporation in RecA. Chosen positions are depicted in yellow, α -helices in blue, and β -sheets in green. (b) Smoothed density plots of \log_{10} -transformed amounts of total protein or soluble protein. (c) Average \log_{10} -transformed soluble protein amounts overlaid on average \log_{10} -transformed total protein amounts for each mutant. Error bars indicate the standard deviation from three individual replicates each derived from separate clones. (d) Plots of the average fraction of soluble protein as a function of three selected parameters: conservation, hydrophobicity, and accessibility. Fits to individual linear regression models yield best fit lines (solid black) and 95% confidence intervals (shaded gray). (e) Boxplots comparing the average fraction of soluble protein against domain or secondary structure, with individual averages overlaid.

Acid and measuring protein amounts, we again observed greater variability in logarithmically-transformed soluble protein levels (mean = 3.42, SD = 0.40) compared to total protein levels (mean = 3.72, SD = 0.17) (Figure 3.3b, Figure 3.3c, Supplemental Table 8). Similar to LexA, most properties examined did not explain much variation in the fractions of soluble protein (Figure 3.3d), with the exception that solubility modestly trended with domain type and tolerance to aromatics (Supplemental Table 9). However, unlike in LexA, no clear relationship existed between protein solubility and type of secondary structure (Figure 3.3e), a result consistent with a more limited prior survey of GFP (Albayrak, 2013). This survey in RecA bolsters a model in which the intrinsic Uaa tolerability of a protein domain remains the key obstacle for the production of soluble protein.

Searching for easily-determined properties that correlate with Acid tolerability may have eliminated from consideration more complicated properties with higher predictive ability. Additionally, linear regression modeling may have over-simplified the inter-dependence of certain properties and protein solubility. Previously, Rosetta modeling has predicted the $\Delta\Delta G$ associated with a particular mutation and identified tolerated mutations within a protein (Alford, 2017; Kellogg, 2011; C. A. Smith, 2011). Speculating that Rosetta modeling could recapitulate our experimental results, we used the Rosetta Modeling Suite to simulate the resulting energy associated with Acid incorporation in LexA or RecA. However, we observed no significant correlations between simulated energies and soluble fractions of LexA or RecA (Supplemental Figure 17 and Supplemental Figure 18). Incidentally, we noted that nearly all high-energy positions in LexA experimentally yielded insoluble protein and may therefore have been useful in filtering out those positions; however, we did not observe a similar energy threshold effect for RecA. Accordingly, further refinement towards predicting Uaa incorporation using Rosetta is required in order to recapitulate experimental data and exclude higher-energy and lower-solubility mutants.

3.4 Conclusion

The expression of soluble protein is a major bottleneck for the study of protein function. Here, we leveraged the fluorescence of Acid to study how protein solubility is impacted by Uaa

mutagenesis. In two bacterial proteins, we demonstrated the dramatic impact that Uaa position has on protein solubility. Surprisingly, a number of amino acid properties that purportedly contribute to Uaa tolerability—including low evolutionary conservation, similar hydrophobic character, or high surface accessibility—were unreliable predictors of protein solubility. Instead, these inconsistent relationships suggest that consideration of specific amino acid features for successful Uaa mutagenesis is less critical than previously thought. Rather, we speculate that the Uaa tolerability of a protein domain may matter more. Our results also emphasize a continued need to explore, through theory and experiment, the steric and chemical burdens different Uaas pose to the expression of soluble protein. In the absence of reliable predictors or refined simulation algorithms for Uaa tolerability, a chemical biologist pursuing Uaa incorporation in a new protein, as of now, should broaden rather than narrow the types of residues screened for Uaa tolerability when possible.

3.5 Contributions

Z.M.H., J.J.F., M.R.B., I.S., E.J.P., and R.M.K. designed the experiments. Z.M.H. performed all the experiments with assistance from M.R.B. J.J.F. performed the Rosetta simulations. I.S. synthesized Acd. E.J.P. and R.M.K. supervised these studies.

3.6 Acknowledgments

We thank members of the Kohli and Petersson laboratories for general advice, and we are grateful to E. Schutsky for input in preparing the manuscript. This work was supported by the National Institutes of Health (R01-GM127593 to R.M.K. and E.J.P.) and the National Science Foundation (NSF, CHE-1708759 to E.J.P.). Z.M.H. was supported by the NIH Chemistry Biology Interface Training Program (T32-GM071399). J.J.F. was supported by the NSF Graduate Research Fellowship Program (DGE-1321851). I.S. was supported by the Royal Thai Foundation.

Chapter 4: A Fluorescent Amino Acid Probe Reveals the Kinetic and Molecular Basis of the Interaction Between LexA and RecA

The contents of this chapter are currently being prepared as a manuscript for publication.

4.1 Abstract

In many bacteria, activation of the DNA damage response pathway (the SOS response) is controlled by the interplay between two proteins: the repressor LexA and the activated DNA damage sensor RecA. Following DNA damage, RecA directly interacts with LexA to promote an intramolecular self-cleavage reaction that renders LexA unable to repress SOS genes. While RecA is crucial for this LexA autoproteolysis reaction *in vivo*, the strength and mode of this macromolecular interaction remain unknown. Here, we demonstrate the use of the minimally-perturbing and sensitive fluorescent amino acid acridonylalanine (Acd) as a kinetic probe of the LexA:RecA complex. Using an Acd-labeled LexA reporter, we define the kinetic and thermodynamic parameters of the reversible binding of full-length *E. coli* LexA to activated RecA (RecA*) in solution. We also show that residues 75–84 of LexA—located on a dynamic peptide loop N-terminal to the scissile bond—strongly contribute to RecA* binding. Together, these findings establish a recognition interface in LexA with sub-micromolar affinity for RecA as the key factor for initiating the LexA autoproteolysis reaction during the SOS response. Our results also establish Acd as a valuable chemical biology tool for investigating protein dynamics in other biological systems.

4.2 Introduction

In bacteria, striking the right balance between the high-fidelity repair and the mutagenic bypass of damaged DNA requires the precise regulation of a number of genes. The ‘off’ or ‘on’ state of the DNA damage response pathway (SOS response) is regulated by the conserved repressor-protease LexA and its interactions between two different macromolecular partners (Figure 4.1) (Butala, 2009). In the absence of DNA damage (the ‘off’ state), LexA dimers bind to genomic DNA at specific operator sequences in SOS gene promoters to impede their transcription

(Brent, 1981; Culyba, 2018). The presence of DNA damage leads to the ATP-dependent polymerization of the sensor protein, RecA, along ssDNA to generate “activated” nucleoprotein filaments, termed RecA* (Cox, 2007). The ‘on’ state is accessed after the direct interaction of RecA* with LexA stimulates an intramolecular peptide bond hydrolysis reaction in LexA (autoproteolysis) (Little, 1984). This self-cleavage reaction abrogates LexA’s ability to bind DNA and function as a repressor, and expression of SOS genes ensues. Critically, genetic inactivation of LexA autoproteolysis has been shown to decrease bacterial survival and prevent or reverse antimicrobial resistance upon treatment with DNA-damaging antibiotics (Cirz, 2005; Lu, 2009; Mo, 2016), making it an appealing target for addressing the challenges posed by increasingly resistant pathogens.

Mechanistic details of the LexA autoproteolysis reaction are best revealed by its structure. LexA is comprised of an N-terminal DNA binding domain (NTD, residues 1–69) connected by a flexible linker to a C-terminal dimerization and protease domain (CTD, residues 75–202) that harbors a Ser119/Lys156 catalytic dyad (Slilaty, 1987; A. P. P. Zhang, 2010). Within the CTD, a structurally-dynamic peptide loop (residues 79–95) bearing the target scissile bond (Ala84-Gly85) can sample multiple conformations (Y. Luo, 2001); however, structural snapshots suggest that only one “cleavable” conformation positions the scissile bond for nucleophilic attack by Ser119. The observation of a conformationally-dynamic loop in LexA has led to debate over an allosteric role for RecA* in promoting the rate of LexA autoproteolysis (Giese, 2008; Y. Luo, 2001).

Direct characterization of the strength and nature of the LexA:RecA* interaction would be an important step towards resolving the mechanistic role of RecA*. Yet, this protein complex has defied numerous efforts aimed at its elucidation. Structural studies have not revealed the interaction

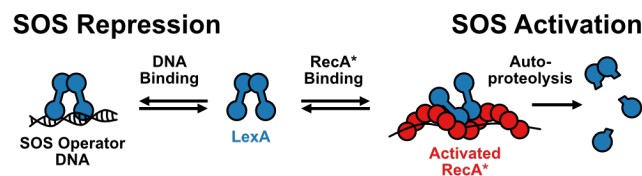


Figure 4.1. Macromolecular Interactions with LexA Regulate the SOS Response. A schematic of the multiple interactions in which LexA can participate during repression (left) or activation (right) of the SOS response.

at atomic resolution (VanLoock, 2003; Yu, 1993), and proposed models of the complex based on the high-resolution structure of the activated RecA* filament conflict with one another (Adikesavan, 2011; Kovačič, 2013). Researchers have not yet identified amino acid substitutions in LexA that specifically disrupt RecA*-mediated proteolysis without impacting the slower basal autoproteolysis rate (Lin, 1988; Mo, 2014; Shepley, 1996). Further, direct SPR studies of LexA:RecA* interaction kinetics have complicated our understanding by suggesting that either the NTD or CTD alone could bind RecA*, albeit at high concentrations and in non-physiologic conditions (Kovačič, 2013).

Given the central role of this reaction in regulating the SOS response and the implications for acquired antibiotic resistance, more precise approaches are needed to directly probe RecA*-mediated LexA autoproteolysis, critically starting with the binding of LexA to RecA*. Fluorescence spectroscopic methods offer an underutilized approach in the SOS field that offer potentially precise, sensitive, and direct measurements of the protein dynamics inherent to the LexA:RecA* interaction (Speight, 2014). Recognizing the limitations of conventional labeling strategies that employ bulky fluorescent proteins or exogenous fluorophores attached by long linkers, here, we instead employ a fluorescent unnatural amino acid as a minimally-perturbing and maximally-sensitive site-specific probe that can directly and specifically report on LexA binding to DNA or RecA* (Speight, 2013). Using this fluorescent LexA binding reporter, we address several longstanding unknowns in the SOS field, including the affinity of full-length LexA for RecA* as well as structural determinants in LexA for RecA* binding. Our findings establish a scientific and methodologic foundation for future investigation of the RecA*-mediated LexA autoproteolysis reaction and validate the utility of minimally-perturbing fluorescent probes in the study of protein dynamics.

4.3 Results

4.3.1 Design of a Fluorescent Full-Length LexA Binding Reporter

Prior approaches lacked the requisite sensitivity or specificity to examine the kinetics of individual steps in the RecA*-mediated LexA autoproteolysis reaction pathway (Giese, 2008; Kovačič, 2013; Takahashi, 1986). In contrast, fluorescence spectroscopy provides measurements that can be directly attributed to the biophysical phenomenon under investigation, such as the binding of proteins in solution (Lakowicz, 2006). However, not knowing the location of the interface between LexA and RecA* necessitates a labeling strategy that minimizes the risk of disrupting this interaction. Previously, we identified a number of positions in full-length LexA (Chapter 3:) that can tolerate the genetic incorporation of acridonylalanine (Acd or δ), a small, intrinsically-fluorescent unnatural amino acid (Speight, 2013). Drawing from this insight, we began by purifying multiple catalytically-inactive LexA proteins (LexA-S119A) containing Acd at several of these tolerant positions. We next investigated whether the fluorescence signals of these Acd-labeled LexA variants can be used to detect specific binding to RecA* by incubating each Acd-labeled LexA variant alone, with non-activated RecA, or with specifically-activated RecA*. Equilibrium

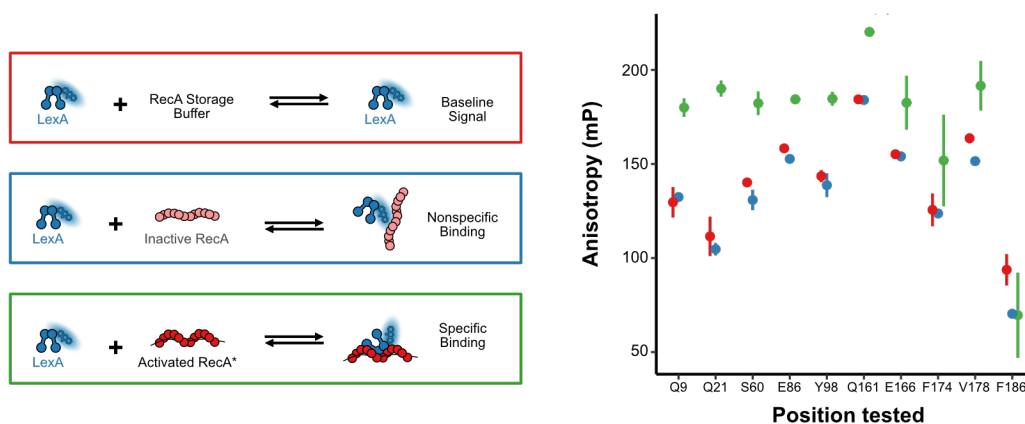


Figure 4.2. Screening Positions that Report on Specific RecA* Binding. Left: Reactions used to identify signals attributable to specific RecA* binding for each LexA construct. Right: Plot of anisotropy values by position for each reaction condition. Increases observed for addition of activated RecA* (green) compared to inactive RecA (blue) or no RecA (red) are consistent with signals due to specific binding.

measurements of fluorescence anisotropy revealed that most of these labeled LexA proteins demonstrate significant changes in anisotropy upon specific binding to RecA* (Figure 4.2).

Of the Acd-labeled LexA mutants we screened, we advanced LexA-Q161δ for further study. Modeling of the Q161δ substitution in LexA suggested that it is sufficiently removed from either the catalytic dyad or target scissile bond (Figure 4.3a), two regions in LexA that are

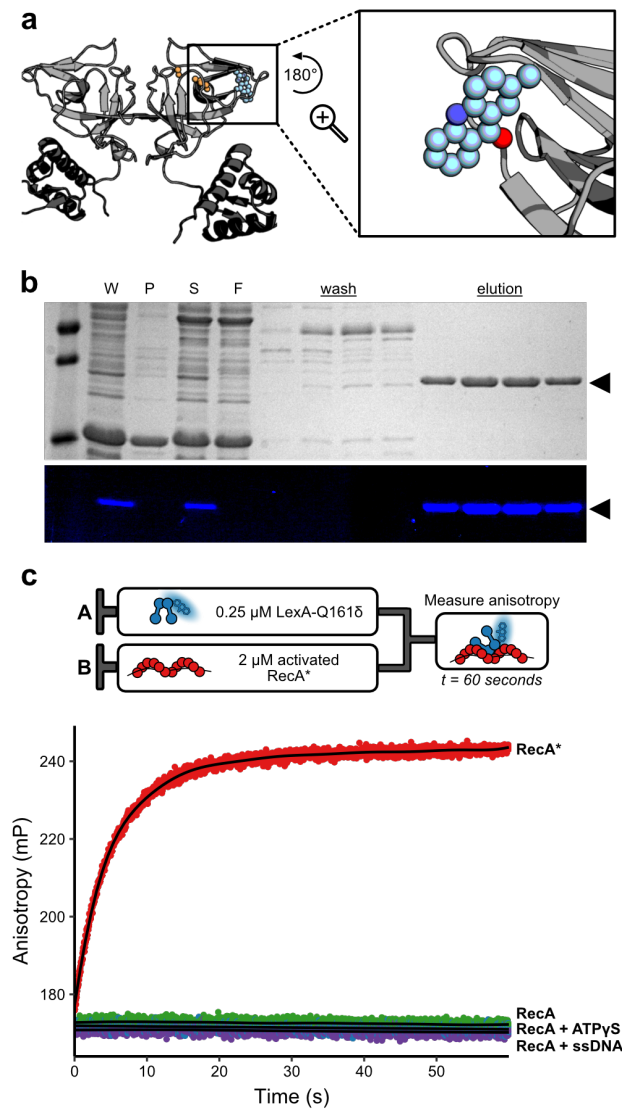


Figure 4.3. A Fluorescent LexA Binding Reporter. (a) Modeled incorporation of Acd at position 161 in LexA. The inset shows a 180° close-up of the incorporated Acd label. (b) SDS-PAGE gels of over-expressed and purified LexA-Q161δ, visualized by Coomassie staining (top) or UV illumination (bottom). (c) Plots of time-dependent anisotropy increases over time, showing specific binding of LexA-Q161δ to RecA*. 250 nM of LexA-Q161δ were rapidly mixed in a stopped-flow with 2 μM RecA reactions containing only RecA (green), RecA + ATPγS (blue), RecA + ssDNA (purple), or RecA + ATPγS + ssDNA (red). Individual data points are overlaid with smoothed fits to the data.

suggested to come into close proximity with RecA* (Lin 1989, Luo 2001). We overexpressed and purified LexA-Q161 δ and verified successful Acd incorporation by visualizing the fluorescent protein gel band (Figure 4.3b). Finally, we confirmed that the Q161 δ substitution does not disrupt native LexA activity by reverting the S119A mutation and determining that the proteolytic activity of this catalytically-active Acd-labeled LexA is similar to that of wildtype LexA (data not shown).

In equilibrium measurements, LexA-Q161 δ displayed a large increase in fluorescence anisotropy on specific binding to RecA* in our screen (Figure 4.2). We reasoned that this signal shift would permit kinetic resolution of LexA as either a free or bound species over time. Indeed, mixing 250 nM LexA-Q161 δ with 2 μ M activated RecA* in a stopped-flow apparatus for 60 seconds revealed a time-dependent change in fluorescence anisotropy of about 70 mP (Figure 4.3c). This increase in the anisotropy of LexA-Q161 δ occurs only in the presence of activated RecA*; the exclusion of ATPyS or ssDNA substrates, which are required for RecA activation, eliminates this anisotropy increase. We reasoned that this large shift in anisotropy could be due to the unusually long lifetime of Acd (15 ns) (Hamada, 2005), because fluorophores with longer lifetimes provide a better dynamic range for the study of macromolecular interactions (Pope, 1999).

4.3.2 Rates of LexA:RecA* Complex Formation

The temporal resolution provided by measuring LexA-Q161 δ on the stopped-flow enables quantitative kinetic analysis of the binding of full-length LexA with RecA* in solution. As before, we rapidly mixed 250 nM LexA-Q161 δ with activated RecA* reactions spanning a nearly 1000-fold concentration range (from 16 nM to 10.24 μ M) and monitored fluorescence anisotropy measurements for 60 seconds. Examination of the measured anisotropy as a function of time reveals that the apparent rate of LexA binding to RecA* increases as a function of RecA* concentration (Figure 4.4a). Further, at intermediate RecA* concentrations, the LexA binding curves do not approach the same maximum anisotropy values as those at the plateaus for the

highest RecA* concentration curves, consistent with a model of reversible binding between LexA and RecA*.

To identify a kinetic model of the LexA:RecA* interaction supported by our association rate data, we tested a variety of binding models via reaction simulation and global fitting. Despite testing

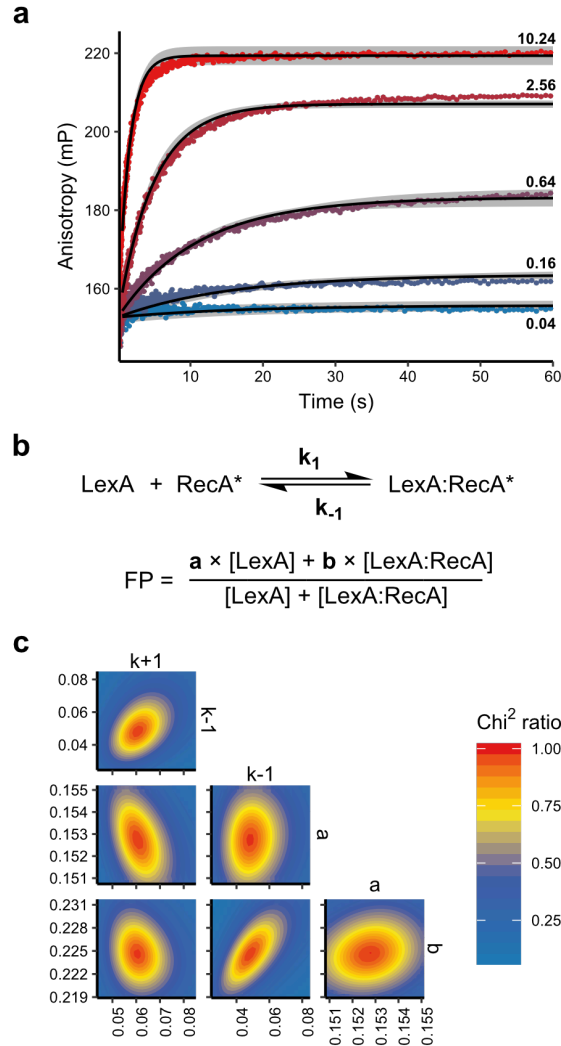


Figure 4.4. Association of LexA with RecA*. (a) Time-dependent changes in fluorescence anisotropy of 250 nM LexA-Q161δ upon addition of a concentration series of activated RecA*. The label for each curve indicates the concentration of activated RecA* in μM. Solid lines represent the best-fit curves from reaction simulations after global fitting with five independent concentration series experiments, and the gray intervals indicate the boundaries of the simulation at a Chi² ratio of 0.70. (b) The reaction model used for global fitting. Parameters fitted during the simulation are in bold, and include the two rate constants, k_1 and k_{-1} , as well as two factors, a and b , for calculating the fluorescence anisotropy signal (FP) of unbound and bound LexA-Q161δ. (c) Two-dimensional confidence intervals for the fitted parameters, with high Chi² ratio values indicating best-fit values. The presence of clear boundaries around the best fitted values is consistent with a model well-constrained by the data.

more complex models, we found that a one-step reversible binding model (Figure 4.4b) was a sufficient model in which all the fitted parameters were well-constrained by the data we collected across three independent binding experiments, each spanning the full concentration series (Figure 4.4c). Using the fitted values for the association rate ($k_1 = 0.061 \mu\text{M}^{-1} \text{sec}^{-1}$) and dissociation rate ($k_{-1} = 0.048 \text{sec}^{-1}$), we calculated an apparent dissociation constant, K_D , of $0.79 \mu\text{M}$ (95% CI = 0.53 to $1.18 \mu\text{M}$) for the interaction of full-length LexA with RecA*.

We next tested the strength of this one-step reversible binding model by directly assessing the dissociation rate of LexA from RecA* in an alternative experimental setup. Reported values for LexA:DNA binding rates indicate rapid binding (Culyba, 2018), and it has been reported that DNA-bound LexA can no longer bind RecA* (Butala, 2011). We predicted that we could pre-form the LexA:RecA* complex, rapidly mix with specific operator DNA, and then monitor dissociation of LexA from RecA* as it preferentially binds to DNA (Figure 4.5a). By using a short operator dsDNA

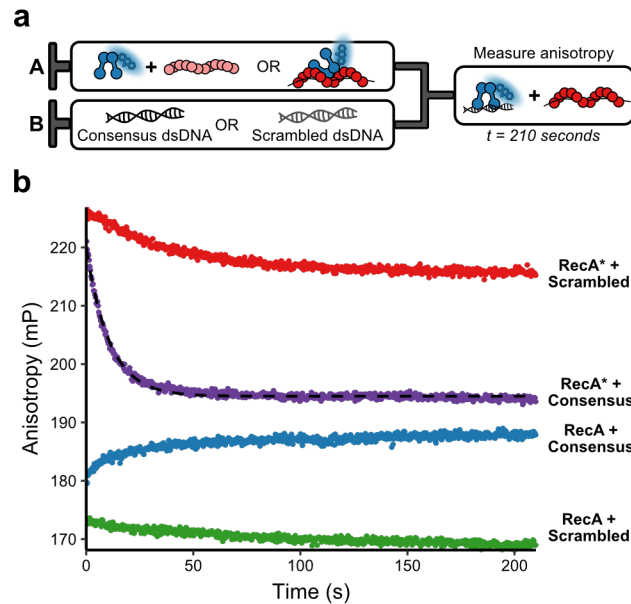


Figure 4.5. Dissociation of LexA from RecA*. (a) Experimental setup for measuring the dissociation of LexA from RecA*. Pre-incubated mixtures of LexA-Q161Δ with from non-activated or activated RecA* are rapidly mixed with 44mer DNA substrates containing the consensus operator sequence or a scrambled operator sequence. The high affinity of LexA for DNA acts as a quench for LexA as it dissociates for RecA*. (b) Plots of anisotropy changes for LexA-Q161Δ over time, with the combinations of reagents indicated for each curve. A dashed line overlaid on the LexA-Q161Δ/RecA* + consensus dsDNA curve indicates a best-fit single exponential decay curve.

substrate, we hypothesized that the anisotropy differences for LexA:DNA and LexA:RecA* complexes should be distinct. To test this idea, we pre-incubated 250 nM LexA-Q161δ with either 5 μM RecA* or 5 μM of inactive RecA as a control. Using the stopped-flow, we mixed our pre-incubated LexA:RecA* reactions with 5 μM of 44mer dsDNA substrates bearing either the consensus operator sequence or a scrambled operator sequence (Figure 4.5b). As expected, we observed negligible changes in anisotropy when dsDNA with the scrambled sequence was mixed with our LexA:RecA* reactions. In contrast, addition of dsDNA with the consensus sequence produced an observable shift in anisotropy, which we attribute to LexA dissociation from RecA* and subsequent association with dsDNA. we obtained an estimate of the LexA:RecA* dissociation rate ($k_{-1} = 0.093 \text{ sec}^{-1}$) that was only two-fold faster than the best fit value ($k_{-1} = 0.048 \text{ sec}^{-1}$) from our above binding model.

4.3.3 Competition Between Labeled and Unlabeled LexA for RecA* Binding Sites

In the above mixing experiments, we show that the binding of LexA with RecA* can be monitored with high kinetic precision. By mixing in different concentrations of RecA*, we are also able to detect a wide range of apparent association rates. Given the positive correlation between RecA* concentration and the observed association rates, we devised a competition experimental setup as a potential means of assessing the binding strength (K_D) of unlabeled LexA variants for RecA*. In this design, we would first allow unlabeled LexA to come to equilibrium with RecA*. The availability of free RecA* sites would then be determined by the dissociation constant of this interaction (Figure 4.6a). After rapid mixing with LexA-Q161δ, the observed rate of association should be proportional to free RecA* availability.

To test this concept, we purified unlabeled, catalytically-inactive LexA-S119A and pre-incubated varying amounts of this protein with RecA*. Next, we rapidly mixed this pre-incubated sample with LexA-Q161δ and monitored changes in anisotropy on the stopped-flow instrument for 15 seconds. We selected a final RecA* concentration of 2 μM because, at this condition, the association rate is 80-90% of its maximum, giving us the ability to sensitively detect decreases over

a wide range of rates. Pre-incubation with increasing amounts of unlabeled LexA correlated with a decrease in the observed rate of LexA-Q161 δ association with RecA*, confirming the expected competition (Figure 4.6b). Next, we calculated the apparent initial rate for each mixing experiment by smoothing the data and deriving the slope at time zero. A plot of these rates versus unlabeled LexA concentration revealed that these initial association rate decreased nonlinearly with increasing unlabeled LexA (Figure 4.6c).

In our experimental design, at time zero in these reactions, the apparent association rates should be proportional to the amount of available LexA binding sites in RecA*. This value can be described using a modified form of the quadratic binding equation in terms of total unlabeled LexA and the concentration of open LexA binding sites on the activated RecA* filament (see Section 4.5.6). Next, we fit the rate versus unlabeled LexA concentration data to this modified quadratic

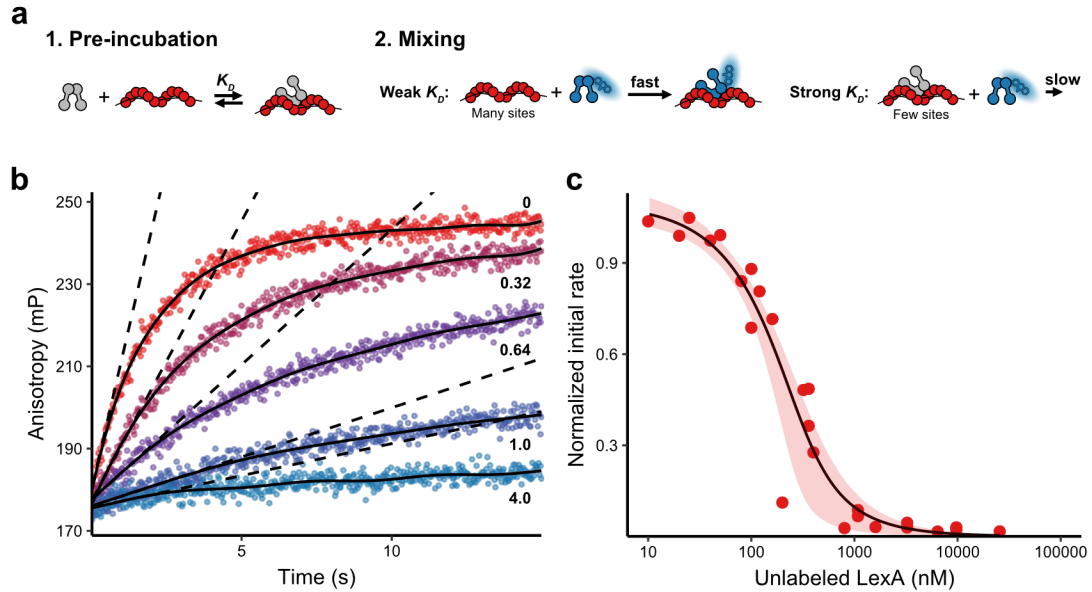


Figure 4.6. Competitive Binding of LexA Proteins for RecA*. (a) Experimental setup for measuring the strength of competitive binding for RecA* between labeled LexA-Q161 δ and unlabeled LexA. Left: Unlabeled LexA and RecA* are mixed and come to an equilibrium. Right: The association rate of LexA-Q161 δ depends on the availability of free RecA* in the unlabeled LexA:RecA* equilibrium. (b) Individual plots of anisotropy changes for LexA-Q161 δ with pre-incubated mixture of 2 μ M RecA* and the indicated amounts of unlabeled LexA in μ M. Solid lines represent a smoothed function of the collected data, and dashed lines represent initial association rates by calculating the slope of the curves at time zero. (c) A plot of initial association rates versus concentration of unlabeled LexA. Data points represent the median slope from at least four individual reaction curves. The solid line indicates the best fit to a modified quadratic function model of binding, and the shaded interval represents bootstrapped 95% confidence intervals.

equation, which revealed an acceptable fit of this model to our data (Figure 4.6c). This fit yielded a K_D^* estimate of 0.63 μM for the interaction of unlabeled LexA with bulk RecA*, a value similar to the K_D of 0.79 μM for LexA-Q161 δ bound to RecA* that we report above.

4.3.4 Comparison of Competitive Binding Strengths Among Modified LexA Variants

Next, we reasoned that we could determine the apparent K_D values for other unlabeled variants of LexA by employing the competition binding assay described above. By comparing the apparent affinities between unlabeled LexA variants, we could gain insights into the regions involved in binding RecA*. As an initial region for study, we targeted the N-terminal domain (Figure 4.7a), because conflicting models exist regarding its role in binding RecA* (Adikesavan, 2011; Kovačič, 2013). We decided to remove each α -helix of the N-terminal domain in turn, generating LexA(23-202), LexA(36-202), and LexA(61-202). We also sequentially shortened a stretch of

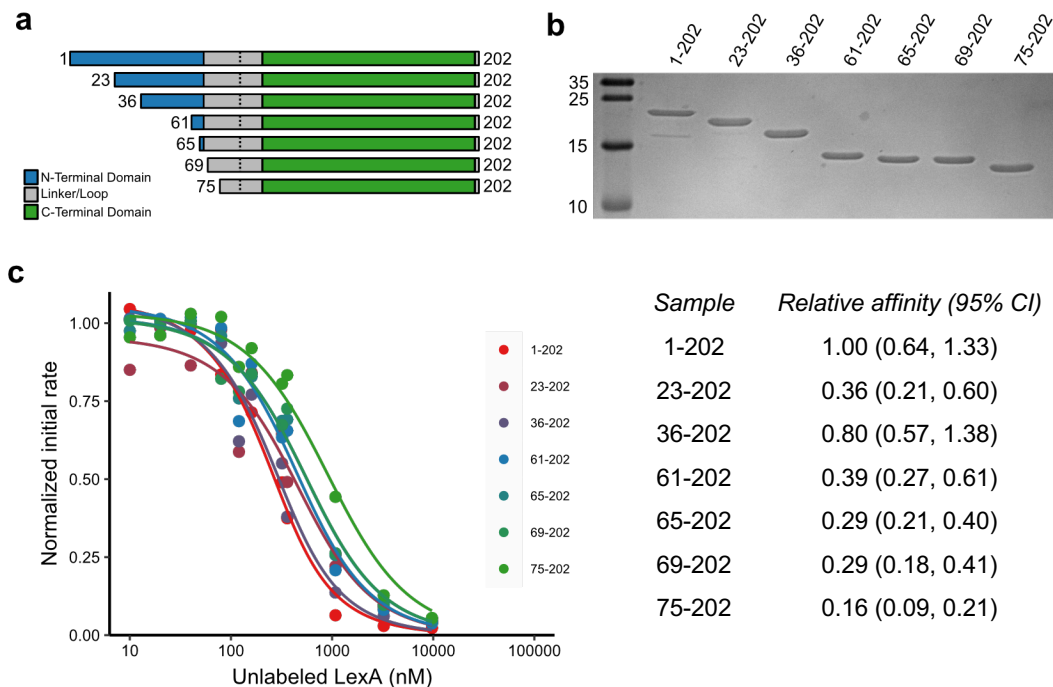


Figure 4.7. Affinities of LexA truncations for RecA*. (a) Design of LexA truncation series. (b) SDS-PAGE gel of overexpressed and purified LexA truncations. (c) Plots of initial association rates as a function of concentration of unlabeled LexA truncations. Data points represent the median slope from at least four individual reaction curves. The solid line indicates the best fit to a modified quadratic function model of binding, and the shaded interval represents bootstrapped 95% confidence intervals.

residues connecting the N-terminal domain to the cleavage loop, yielding LexA(65-202), LexA(69-202), and LexA(75-202). After purifying these truncations (Figure 4.7b), we mixed different concentrations of each one with activated RecA* and performed our competition binding assay as described above. Plotting the initial rates as a function of unlabeled truncation concentration revealed a similar dependence of the initial rates on unlabeled LexA concentration (Figure 4.7c). However, none of the fitted binding curves were substantially different from the curve for full-length LexA(1-202). Instead, we only observed minor rightward shifts in the fitted curves as the length of LexA's N-terminal domain decreased, inconsistent with a major role for the N-terminal domain in binding to RecA*.

We next investigated the role of the mobile peptide loop in interacting with RecA*. While many studies targeting this loop have implicated it in determining the rate of LexA autoproteolysis,

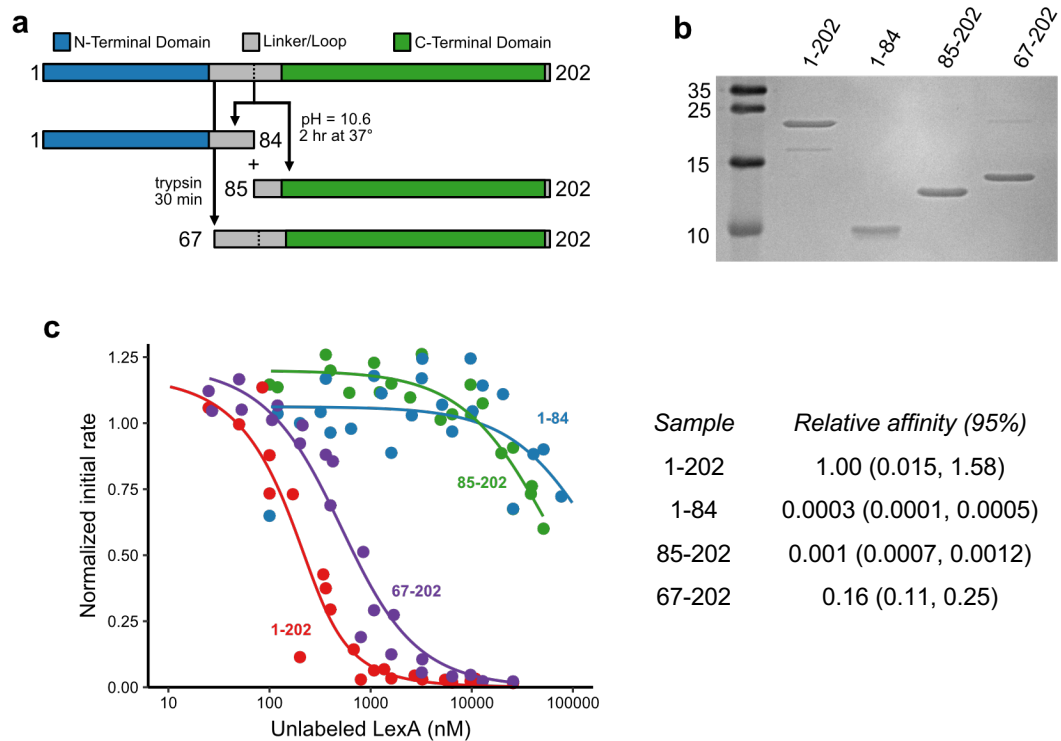


Figure 4.8. Affinities of LexA Proteolytic Fragments for RecA*. (a) Schematic for generating LexA proteolytic fragments. (b) SDS-PAGE gel of overexpressed and purified LexA proteolytic fragments. (c) Plots of initial association rates as a function of concentration of unlabeled LexA fragments. Data points represent the median slope from at least four individual reaction curves. The solid line indicates the best fit to a modified quadratic function model of binding, and the shaded interval represents bootstrapped 95% confidence intervals.

no direct evidence has been reported to support the binding of this region to RecA* (Lin, 1988; Mo, 2014). To assess the role of this loop, we first decided to split this loop in half at the natural site of autoproteolysis. Incubation of full-length, wildtype LexA at high pH can be used to generate the N-terminal and C-terminal fragments, each bearing one half of the β -hairpin loop (Figure 4.8a,b). In parallel, we also prepared full-length, catalytically-inactive LexA and trypsin-digested LexA; the loops in both of these unlabeled variants remain intact. We then tested each of these variants for their ability to inhibit LexA-Q161 δ association with RecA* in our competition binding assay described above. A plot of the observed initial rates versus unlabeled LexA fragment concentration reveals the substantial differences between these curves (Figure 4.8c). Whereas full-length and trypsin-digested LexA prevent the association of labeled LexA-Q161 δ at low micromolar concentrations, both proteolytic fragments demonstrated dramatically weaker affinities with the only evidence of binding competition occurring at very high, non-physiologic concentrations.

A stark difference in binding affinities between intact LexA and either proteolytic fragment implicates the intact mobile peptide loop as a critical structural determinant for RecA* binding. However, the molecular nature of how this loop contributes to binding is not apparent; for example, the interaction could depend on direct contact with residues in the loop or it could require loop flexibility for a recognition interface to form. To probe the mechanism underlying this interaction, we selected two positions, Gly80 and Val82 (Figure 4.9a), in which certain amino acid substitutions are known to decrease the rates of both basal and RecA*-mediated LexA autoproteolysis (Mo, 2014). Here, we mutated and purified unlabeled LexA proteins with a G80P or V82M substitution alongside our LexA-S119A protein as a control. By titrating each of these LexA proteins with 2 μ M RecA* and then mixing these pre-incubated reactions with LexA-Q161 δ , we could again monitor the effect of each substitution on the apparent initial rate of association. A plot of the calculated initial rates versus the concentration of each unlabeled LexA showed the expected decrease in association rates as unlabeled LexA increased (Figure 4.9b). Notably, the association rate decreases attributable to the S119A mutant occurred at lower unlabeled concentrations than the

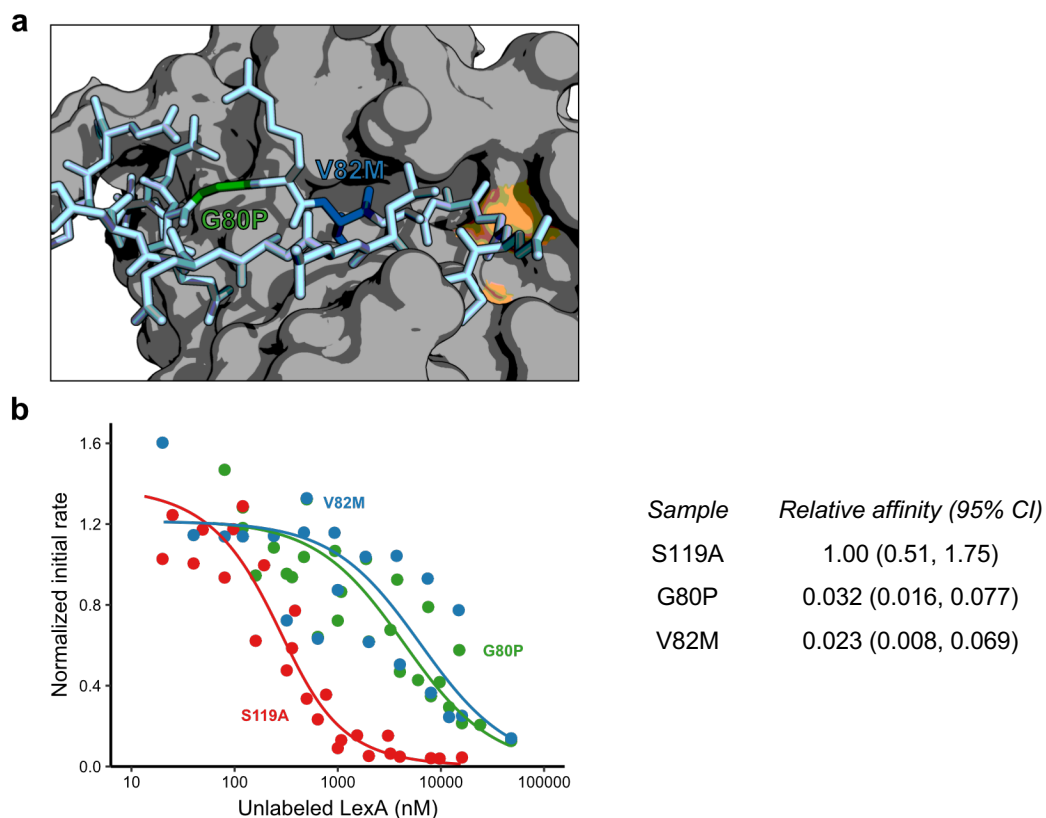


Figure 4.9. Affinities of LexA Loop Mutants for RecA*. (a) Structural model of the mobile peptide loop within the LexA C-terminal domain active site. The loop is pictured as light blue sticks with the two loop mutants highlighted (G80P = green, V82M = blue). The remainder of the C-terminal domain is shown as a gray surface-filling model, with the active site residues indicated in orange. (b) Plots of initial association rates as a function of concentration of unlabeled LexA mutants. Data points represent the median slope from at least four individual reaction curves. The solid line indicates the best fit to a modified quadratic function model of binding, and the shaded interval represents bootstrapped 95% confidence intervals.

rates for either the G80P or V82M mutants, and the fitted binding curves revealed a pronounced difference in affinities for either of the loop mutants compared to the S119A mutant. Because neither substituted position contains a bulky or charged residue with which to disrupt a potential interaction with RecA*, these amino acid substitutions are consistent instead with alterations that impact the flexibility or dynamics of the peptide loop.

4.4 Discussion

Interest in targeting the SOS response for its role in mutagenesis and antibiotic evasion has spurred numerous investigations into the mechanisms underlying its activation. Genetic

inactivation of the SOS response regulator, LexA, has been shown to decrease bacterial survival and even reverse resistance to DNA-damaging antibiotics (Cirz, 2005; Lu, 2009; Mo, 2016). However, developing small molecule inhibitors of this response is challenging due to our insufficient mechanistic insight into the activation of the pathway. One key knowledge gap in the SOS response is understanding the strength and mode of the molecular interface between RecA* and LexA. Despite past efforts, few approaches have come close to resolving this protein-protein interaction. To overcome deficiencies of past studies, we employed the fluorescent unnatural amino acid Acd as a small probe of protein dynamics. This strategy enables us to report for the first time the thermodynamic and kinetic constants governing the interaction between full-length LexA and activated RecA* in solution. This approach also allows us to identify LexA's mobile peptide loop as a major factor in modulating RecA* binding.

4.4.1 Association Rate and Affinity of the LexA:RecA* Binding Step

Generating an Acd-labeled LexA variant, LexA-Q161 δ , enabled us to precisely monitor rates of association and dissociation of full-length LexA with RecA* in solution and allowed us to develop a kinetic model of LexA:RecA* binding. By generating the labeled LexA-Q161 δ variant with a catalytically-inactivating S119A amino acid substitution, we excluded the cleavage step and isolated a single, reversible binding step. Because evidence for and speculation about the complexity of the LexA:RecA* interaction is well-documented (Adikesavan, 2011; Kovačič, 2013; Mustard, 2000; Yu, 1993), we recognize that the one-step receptor-ligand binding model that best fits our data is likely an oversimplification, especially given the assumptions this model required. For example, we fixed activated RecA* as a single species competent for LexA binding in our direct examination of LexA-Q161 δ :RecA* binding, even though its biology and activation kinetics are quite complex (Bell, 2016; Cox, 2007). Notably, the binding model for our competition binding experiment (Figure 4.6) required that we include a parameter to account for multiple RecA* subunits in one LexA binding site. Our fitted value of 8 RecA subunits for this binding site size parameter is consistent with past reports of extended LexA binding sites in the RecA* filament (Adikesavan,

2011; Yu, 1993). While some uncertainty of the stoichiometry between LexA and RecA* remains (given the wide confidence interval around this binding site parameter), our proposed binding model is well-supported by the data and provides a robust and quantitative framework against which more complicated models can be designed and tested.

Our calculated K_D value of 0.79 μM for the LexA:RecA* interaction compares similarly to reported K_M values of either 0.5 μM for the RecA*-catalyzed auto-proteolysis of full-length LexA (Lin, 1988) or 0.9 μM for the auto-proteolysis of a truncated LexA containing residues 65-202 (Giese, 2008). That our measured K_D value falls within a range of reported K_M values suggests that k_2 , the step representing the chemistry of LexA auto-proteolysis, may be rate-limiting because it must be slower (i.e. smaller) than the dissociation rate in order for the K_D and K_M values to be similar. Additionally, our finding that neither N-terminal nor C-terminal auto-proteolytic fragment of LexA competitively inhibits LexA-Q161 Δ binding at physiologic concentrations strongly suggests that product inhibition is not a factor in the auto-proteolysis reaction. The weak affinities we inferred for these fragments implies that, following the chemistry step of auto-proteolysis, product release is not rate-limiting. However, our study did not examine the formation of full-length/fragment heterodimers, a possible physiologic species that may interact with RecA* differently to modulate SOS activation (Giese, 2008).

Within the cellular context of a physiologic response to DNA damage, our results support a non-equilibrium view of the SOS response (Culyba, 2018). We find that LexA binding to RecA* is demonstrably slower and weaker than these reported binding rates and affinities of LexA to operator sites in the genome. In the absence of persistent DNA damage, stochastically-formed and short-lived RecA* filaments are likely not in sufficient quantity to disrupt LexA:DNA binding equilibria. Only upon DNA damage do sudden and large changes in the operative levels of RecA* overcome, through mass effect, the favored kinetics of LexA binding to operator DNA sites. The absence of tight binding by either LexA proteolytic fragment establishes a driving force for flux

through the kinetic pathway, which is reinforced by the rapid degradation of both fragments by cellular proteases (Neher, 2003).

4.4.2 A Structural Region in LexA Strongly Impacts Binding to RecA*

Efforts to localize residues or regions in LexA that bind RecA* have yielded a number of results that are difficult to reconcile. An early genetic screen of LexA mutants had implicated a region from 80-84 as important in RecA*-mediated autoproteolysis (Lin, 1988), but further examination of substitutions in this region revealed that these also impact alkaline-induced LexA auto-proteolysis in the absence of RecA* (Mo, 2014; Shepley, 1996). Certain truncated versions of LexA, either including residues 65-202 or 75-202, can undergo RecA*-mediated auto-proteolysis at rates comparable to wildtype LexA, but these studies do not further resolve a region involved in RecA* binding (Giese, 2008; Mo, 2018). The most direct examination of the LexA:RecA* interaction, using chemical cross-linking and SPR, suggested that either the N-terminal domain (1-69) or C-terminal domain (75-202) of LexA can bind RecA*, albeit at high, likely non-physiologic, concentrations (Kovačič, 2013).

Our results using a quantitative, competitive binding assay clarify these past findings. Testing the competitive binding of our series of LexA truncations revealed that the removal of N-terminal residues up to Glu74 only nominally altered affinity for RecA* (Figure 4.7). Should the N-terminal domain interact with the RecA* filament, our results suggest it does so only through weak and diffuse interactions across its surface, as no single truncation in that series produced a dramatic shift in affinity. Instead, we only observed a dramatic shift in affinity when we tested the C-terminal proteolytic fragment (Figure 4.8). Compared to the LexA(75-202) truncation, removal of the additional residues 75-84 in the C-terminal proteolytic fragment dramatically decreased affinity for RecA*. However, these residues, on their own, are not sufficient for binding to RecA*. The N-terminal proteolytic fragment, LexA(1-84), does possess this region but is similarly unable to bind to RecA* in the competition binding assay. Clearly, the combination of residues 75-84 with an intact

mobile peptide loop or with the C-terminal catalytic core is required to form the molecular region that permits recognition and binding to RecA*.

4.4.3 A Refined Catalytic Role for RecA*

RecA* has been previously described as a “co-protease” or catalyst of LexA self-cleavage, because it is capable of promoting the autoproteolysis of multiple LexA “substrates” (Giese, 2008). In the absence of RecA*, however, LexA autoproteolysis occurs very slowly. At physiologic pH, Lys156 is protonated and likely acts as an energetic barrier to the dynamic peptide loop adopting the cleavable conformation. The observation that exposing LexA to alkaline conditions increases the basal autoproteolysis rate is consistent with a requirement for deprotonation of Lys156. These structural and biochemical observations have led to the proposal that RecA* accelerates LexA autoproteolysis by inducing or stabilizing an energetically-unfavorable cleavable conformation in LexA.

Despite this model, the role of RecA* as a catalyst has remained poorly defined. Our study begins to shed light on how binding to RecA* might promote the rate of LexA autoproteolysis. While others have suggested that specific residues might interact directly with RecA*, our findings are inconsistent with this interpretation. Substitution of Gly80 or Val82 had a pronounced effect on the ability of those LexA proteins to bind RecA* (Figure 4.9). Each substitution we tested is thought to impact the flexibility or dynamics of the loop. Position G80 is thought to form a hinge for the amino-terminal end of the mobile peptide loop, and its modification to Pro would dramatically reduce its flexibility. Val82 is thought to insert into a binding pocket in the catalytic core of the C-terminal domain, and its substitution for a polar Met would likely perturb this interaction. Given our results, either substitution appears likely to drastically alter the shape or flexibility of the loop, disrupting a molecular interface for RecA* binding in turn. By extension, it is reasonable to assume that RecA* has low affinity for a misshaped mobile peptide loop in the LexA C-terminal domain. Binding of RecA* to the peptide loop might therefore stabilize its “cleavable” conformation through an allosteric mechanism, consequently trapping the scissile bond and preventing the loop from sampling other

conformations. Likewise, allosteric stabilization could promote deprotonation of Lys156 by making other kinetic steps for the loop less likely, resulting in an increased probability of proteolysis.

Our findings that neither proteolytic fragment can bind RecA* supports this allosteric model. Because RecA* binds to a molecular interface in full-length LexA that forms when amino acids 75-84 are present, after bond cleavage, residues 75-84 are no longer covalently tethered at their C-terminus and are free to diffuse away from the C-terminal domain. Whether by loss of the β -sheet structure or by increased loop disorder, it is unlikely that the RecA* molecular interface remains formed. As a result, fragment binding becomes unfavorable, both fragments release rapidly, and neither fragment engages in competitive inhibition. Altogether, our findings begin to illuminate how LexA:RecA* complex formation leads directly to autoproteolysis.

4.5 Methods

4.5.1 Protein cloning, expression, and purification

Overexpression of LexA variants containing amber stop codons in the presence of Acd was accomplished as previously described (Hostettler, 2018), with the following changes. To scale up expressions, cultures were grown in 400 mL of MDA-5052 auto-induction media (Studier, 2014) in 4 L baffled flasks with solubilized Acd added to 1 mM. After growth for 24 hours, cells were harvested and stored at -80 °C. Cells were resuspended in buffer containing 20 mM sodium phosphate pH 7.2, 500 mM NaCl, and 25 mM imidazole; cells were then lysed using a benchtop microfluidizer (Microfluidics). Lysates were centrifuged for 40 min at 20,000 rpm. Clarified lysates were affinity purified with HisPur resin (Thermo Fisher) using a 25 to 400 mM imidazole gradient. Fractions containing purified Acd-labeled LexA eluted in the second half of the gradient. Purified LexA was dialyzed into buffer containing 50 mM Tris-HCl pH 7.6 (at 4°), 200 mM NaCl, 0.5 mM EDTA, and 10% glycerol and then stored at -80 °C.

Plasmids bearing the genes for N- or C-terminally HIS-tagged LexA were used as the starting point for generating catalytically-inactive unlabeled LexA truncations, proteolytic fragments,

or point mutants. Oligonucleotides used in Phusion polymerase site-directed mutagenesis or in Gibson assembly were ordered from Integrated DNA Technologies (IDT) and are available upon request. Successful cloning was confirmed through restriction mapping of the plasmids and sequencing the insert gene (GeneWiz). For protein overexpression, clones were transformed into BL21(DE3) cells. Overnight cultures of transformed cells were used to inoculate 400 mL of LB media with 30 µg/mL kanamycin. Cultures were grown at 37 °C to an OD₅₉₅ of 0.4-0.7 and induced with IPTG to a concentration of 1 mM. For certain LexA variants that had a tendency to aggregate with these expression conditions, they were instead switched to 16 °C after reaching the critical OD₅₉₅ and induced with 0.1 mM IPTG and grown overnight. Cells were harvested and lysed similar to the Acd-labeled proteins described above, with the exception that these were purified in batches according to recommended protocol (Thermo Fisher). Following dialysis, HIS-tags were removed via TEV protease cleavage, and remaining contaminants were removed by flowing sample back over HisPur resin. The thrombin and pH proteolytic fragments were similarly purified following the proteolysis step. Proteins were stored as above.

A plasmid encoding an N-terminally HIS-tagged wildtype RecA gene (Hostetler, 2018) was transformed into BLR(DE3) cells (BL21(DE3) with *recA*⁻ genotype). Cultures were grown in 2 L of LB media plus 30 µg/mL kanamycin at 37 °C for 2 hours and then harvested and stored at -80 °C. Purification of HIS-RecA was carried out as previously described for wildtype RecA (Shibata, 1983). Notably, HIS-RecA eluted from hydroxyapatite resin at much higher sodium phosphate concentration than that reported for wildtype RecA. Following hydroxyapatite chromatography, RecA was deemed sufficiently pure for biochemical assays, and was dialyzed into storage buffer and stored as previously described (Shibata, 1983).

4.5.2 RecA Activation Reactions

RecA^{*} activation reactions were performed as follows, unless otherwise indicated in the text. A 2x concentrated reaction was prepared in reaction buffer containing 70 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 150 mM NaCl, 5 mM DTT, 50 µg/mL BSA with 0.25 mM ATPγS, 16 µM poly(dT)

ssDNA, and 4 μM HIS-RecA. The reactions were incubated overnight on ice, a condition previously described to result in optimal rates of LexA cleavage (Giese, 2008). For experiments involving concentration series of RecA*, the ssDNA substrate was maintained at a ratio of 4:1 nucleotides to RecA monomers. Non-activated RecA controls were prepared similarly, with the exclusion of ATP γ S (unless otherwise indicated in the text). Upon mixing with an equal volume of sample in the stopped-flow, the concentration of RecA* decreased to a final concentration of 2 μM .

4.5.3 Fluorescence Anisotropy Measurements

All stopped-flow measurements were made in a KinTek AutoSF-120 stopped-flow spectrofluorometer, with the following settings. A Xenon/Mercury lamp (Hamamatsu) with a monochromator (Oriel) was set to emit light through a fiber optic at 385 nm with a 3.16 mm diffraction grating. Incident light was plane polarized using a Glen-Taylor polarized optic and passed into the observation chamber. Emitted light was filtered with two 440 \pm 40 nm emission filters (Edmund Optics) with polarized films in the vertical or horizontal direction for either photomultiplier tube (PMT). PMT voltages were calibrated so that a fluorescence reading of 5.0 was achieved, and a grating factor (G-factor) was obtained at the start of each instrument run. Buffer-matched samples for each syringe were prepared at twice the final concentration, such that Acd-labeled LexA reached a final concentration of 250 nM in each experiment. The concentrations of other biological components in the text are indicated at their final concentrations. Samples measurements were collected for the indicated times, and anisotropy was calculated as follows:

$$r = \frac{I_{polarized}}{I_{total}} = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} + 2GI_{\perp}}$$

4.5.4 Measuring LexA:RecA* Association Rates and Fitting a Binding Model

To track association rates, mixtures of Acd-labeled LexA at 250 nM and RecA* at concentrations indicated in the text were monitored for changes in fluorescence anisotropy for 60 seconds. Following exclusion of low-quality traces, the remaining curves were averaged to generate a mean and standard deviation for each time point. The data were imported into KinTek

Explorer global fitting software, a program that simulates reactions from a user-specified model and provides robust tools for error analysis (Johnson, 2009b). Parameters in a one-step reversible binding model with LexA and RecA in a 1:1 stoichiometry were manually adjusted until they approximated the data, and then improved fits were obtained via nonlinear regression. Next, using the FitSpace tool in the software, the fitted parameters were systematically examined for ranges of values that still yielded an acceptable fit (Johnson, 2009a). More complex models were also tested, and these were excluded because certain parameters were no longer well-constrained by the data per FitSpace analysis. Finally, error bars were generated by adjusting the χ^2 ratio in the fitting process until most of the data points were contained in the intervals, thus providing upper and lower limits to each estimated parameter.

4.5.5 Measuring a LexA:RecA* Dissociation Rate

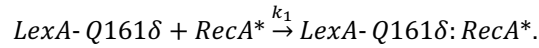
To monitor the dissociation of LexA from RecA*, labeled LexA was incubated with non-activated or activated RecA to give final concentrations of 250 nM and 2 μ M, respectively. An equal volume of sample containing a 44mer dsDNA substrate containing the consensus operator sequence or a scrambled version of that operator sequence (Culyba, 2018) were used as quench molecules for LexA that dissociates from RecA. Reactions were mixed and fluorescence anisotropy monitored for the indicated time. An estimate of the dissociation rate from RecA* was generated by fitting the appropriate curve to a single exponential decay equation.

4.5.6 Competition Binding Assays with Unlabeled LexA Variants

A concentration series of unlabeled LexA variants was prepared for each experiment, using the range of concentrations indicated in the text. Each unlabeled LexA variant was prepared at 4x the final concentration. Activated RecA reactions were also prepared at 4x the final concentration. Equal volumes of the 4x unlabeled LexA variants and 4x RecA* samples were mixed and incubated for 1 hour so that an equilibrium was reached. These mixtures, with both components now at 2x the final concentration, were mixed with an equal volume of 2x LexA-Q161 δ on the stopped flow.

Individual traces from competition binding assay were examined and low-quality traces were discarded. For each trace at a particular concentration of unlabeled LexA, a smoothed spline model representing the data is computed using a Gaussian Process smoothing model from the “mgcv” package for R. To approximate the initial rate for each smoothed curve, the instantaneous rate at time = 0 sec was estimated using a finite difference interpolation method. To account for differences between experiments, the calculated initial rates for the experimental samples are normalized against calculated slopes from positive (RecA* with no unlabeled LexA) and negative (non-activated RecA) control pre-incubation reactions.

At early time points in the association reaction, we assume that no LexA-Q161δ:RecA complex has yet formed, therefore, we can simplify this non-equilibrium state to an irreversible, bimolecular binding reaction,



The apparent initial rate of this irreversible reaction at early time points is given by

$$v_0 = k_1 [\text{LexA-Q161}\delta] [\text{RecA}^*].$$

Using the quadratic solution for the binding equilibrium between an unlabeled LexA variant and RecA, the $[\text{RecA}^*]$ term, representing available sites in RecA* for binding, can be described in terms of $[\text{LexA}]_0$, the initial concentration of unlabeled LexA variant, and K_D , the apparent equilibrium constant between unlabeled LexA and RecA*. Because more than one RecA subunit may be involved in binding to LexA, the total concentration of RecA is divided by a binding site correction factor, m , which was fixed at a value of 8 for all fits performed here. The solution of this relationship is given below, where LexA and RecA have been shortened to “L” and “R” for simplicity:

$$v_0 = \left(\frac{v_{max}}{[R^*]_0/m} \right) \left([R^*]_0/m - \frac{([L]_0 + [R^*]_0/m + K_D) \pm \sqrt{([L]_0 + [R^*]_0/m + K_D)^2 - 4[L]_0 [R^*]_0/m}}{2} \right)$$

Here, the $k_1 [\text{LexA-Q161}\delta]$ term from above has been replaced by the $\left(\frac{v_{max}}{[R^*]_0/m} \right)$ term, which can be measured for each experiment using a positive control reaction. This relationship was used to fit the initial rate data as a function of unlabeled LexA concentration using the nls.multstart non-linear

regression package in R. 95% confidence intervals for the best fit curves and model parameter estimates were obtained by bootstrapping the data in R and obtaining the 2.5 and 97.5 percentiles of each value. At least 1000 iterations were performed for each round of bootstrapping.

4.6 Acknowledgments

We thank members of the Kohli and Petersson laboratories for general advice as well as M. Culyba and R. Kubanoff for technical assistance with the AutoSF-120 stopped-flow spectrofluorometer. This work was supported by the National Institutes of Health (R01-GM127593 to R.M.K. and E.J.P.) and the National Science Foundation (NSF, CHE-1708759 to E.J.P.). Z.M.H. was supported by the NIH Chemistry Biology Interface Training Program (T32-GM071399). J.J.F. was supported by the NSF Graduate Research Fellowship Program (DGE-1321851).

Chapter 5: Future Directions and Concluding Remarks

In the studies presented here, we sought to uncover the protein dynamics governing the activation of the pro-mutagenic bacterial SOS response by refining and using a minimally-perturbing fluorescence labeling strategy. In addition to the mechanistic insights we hoped to gain into the LexA:RecA* interaction, our efforts were also purposefully designed to expand the utility and applicability of using unnatural amino acid (Uaa) mutagenesis techniques to fluorescently label proteins. Our primary focus was with acridonylalanine (Acđ). Our work in Chapter 2 provides important insight for researchers using genetic code expansion technologies, especially those looking to expand the range of unnatural amino acids that can be incorporated *in vivo*. By characterizing the substrate specificity of our evolved Acđ tRNA synthetase (AcđRS), we contributed to a growing body of evidence for how AcđRS and other Uaa synthetases discriminate between a target Uaa and closely-related, but undesired, α -amino acid contaminants. In this case, a number of active site substitutions, including the relatively modest Gly→Ala and Glu→Asp substitutions, transformed a permissive AcđRS into a specific one. These results join other examples of tRNA synthetase evolution that establish permissive synthetases as an excellent starting point for evolving Uaa-specific enzymes (Cooley, 2014; Lee, 2009). Determining the structural and mechanistic elements that govern tRNA synthetase specificity should be the focus of future theoretical and experimental studies so that we can better understand how to tailor these enzymes to fit the growing needs of the genetic code expansion community.

In the last two decades, rapid development has centered around synthesizing new Uaas as well as improving the necessary biological tools for their incorporation in a wide range of organisms. The introduction of these additional chemistries and functions to the proteinogenic genetic code has ushered in many fundamental and applied discoveries in the study of proteins. However, the additional steric and chemical burdens of unnatural and non-canonical amino acids on protein structure and function remain underexplored. In Chapter 3, we attempted to address this gap in knowledge by systematically surveying the effect of Uaa incorporation on soluble protein

expression across fifty positions in both LexA and RecA. While we established that introduction of a Uaa into these proteins has a demonstrable effect on soluble protein expression, we were unable to identify amino acid properties that correlated with protein solubility. Instead, for each protein domain in which we sampled an adequate number of positions (i.e. the N- and C-terminal domains of LexA, the nucleotide-binding domain of RecA), we observed differing levels of intrinsic tolerability to Uaa incorporation. There are several implications for our finding that Uaa incorporation is related to overall protein solubility. First, it suggests that some aspect of protein stability or protein folding is affected by the presence of a Uaa, and this effect varies between domains. Because our analysis was designed to identify position-level features, future investigations should shift attention to understanding how different types of protein folds or domains tolerate Uaa incorporation. Second, our results suggest that constraints specific to a certain type of amino acid feature, such as evolutionary conservation or surface exposure, are less important on the success of Uaa mutagenesis. Instead researchers screening multiple positions for Uaa incorporation should include counter-intuitive sites, as our results demonstrate the tolerability of a position to Uaa incorporation is difficult to predict at this point.

The methodologic findings for incorporation of Acd in Chapters 2 and 3 directly positioned us for our work probing the LexA:RecA* interaction in Chapter 4. Through the specific incorporation of Acd at tolerant sites in LexA, we generated a fluorescent LexA protein that can report on interactions with RecA* with excellent sensitivity and kinetic resolution. By focusing on and isolating this step of LexA:RecA* binding, we were able to provide the first quantitative estimates of the binding affinity and rates of association and dissociation for this elusive protein-protein interaction. We also leveraged the robust nature of our assay to identify residues and structural elements in LexA that are important for binding to RecA*. Taken together, these results provide direct experimental evidence for a molecular recognition interface in LexA with sub-micromolar affinity for RecA*. Our results also provide compelling evidence for a model in which RecA* promotes the autolysis of LexA through an allosteric binding event. However, additional investigations into

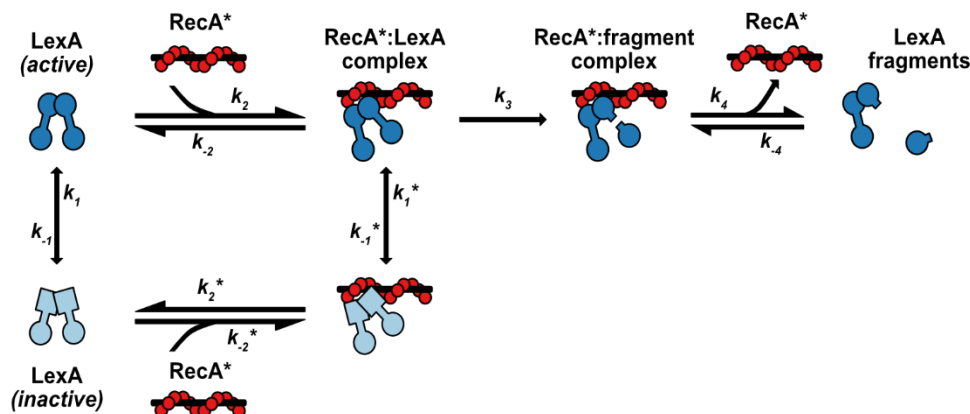


Figure 5.1. The Overall LexA:RecA* Reaction Pathway Model. Individual steps of RecA*-mediated LexA autoproteolysis are shown, representing proposed steps of the reaction pathway. This model does not exclude the possibility of more complicated kinetic steps.

the protein dynamics of this reaction pathway are required before settling the debate surrounding the catalytic role of RecA* as well as what residues and regions are involved in its binding to LexA. Nevertheless, this study represents the necessary initial work in a multi-pronged effort to dissect and elucidate the underlying mechanistic and structural details of the RecA*-mediated LexA reaction (Figure 5.1).

The work summarized in this thesis provides the technical foundation and intellectual framework for future studies aimed at unraveling the molecular mechanisms behind SOS activation. It also provides a starting point towards targeting LexA cleavage as a means of disarming bacterial survival strategies in their response to DNA-damaging antibiotics. In the sections that follow, I will describe areas of ongoing study and future research directions, all of which share the common goal of delivering new insights into the mechanisms underlying activation of the SOS response through the use of new and innovative biochemical and biophysical approaches.

5.1 Investigating the Binding of Other LexA-Family Proteins to RecA*

LexA is conserved across many bacterial species, and the unusual features of its C-terminal protease domain—the dynamic peptide loop and its intra-molecular protease activity—are found not only in the LexA proteins of other species, but also in homologs of LexA with more diverse

roles. Among these inter-related LexA-family proteins, the λ cl repressor and the UmuD protein are well-studied examples that both share remarkable structural similarity in their protease domains (Figure 5.2). Functionally, these two LexA-like proteins also undergo autoproteolysis faster in the presence of RecA* (Mustard, 2000). Similar to LexA autoproteolysis, cleavage of λ cl prohibits this repressor from dimerizing and binding to the promoters of lytic cycle genes preventing their activation. In contrast to LexA, however, UmuD becomes activated upon RecA*-mediated proteolysis, and the truncated protein comprises an essential part of the mutagenic Pol V polymerase. Notably, these homologs demonstrate slower proteolytic rates than that of LexA, raising questions regarding how well-conserved the interaction is between RecA* and each of these LexA-family proteins. Beyond these two examples, LexA proteins from other bacterial species also demonstrate a wide range of autoproteolysis rates (Amanda Samuels, personal communication), further emphasizing a need to understand how complex formation between RecA* and any of its protein substrates leads to autoproteolysis.

Given these observations, investigating the affinities and modes of interaction for any number of these LexA-family proteins with RecA* is a reasonable next step. Our results from Chapter 4 highlighted the value of using a competition binding assay to parse out differences in relative binding affinities between different variants of *E. coli* LexA. We would predict that other LexA-like proteins that are found within *E. coli* and are known to interact with *E. coli* RecA, including

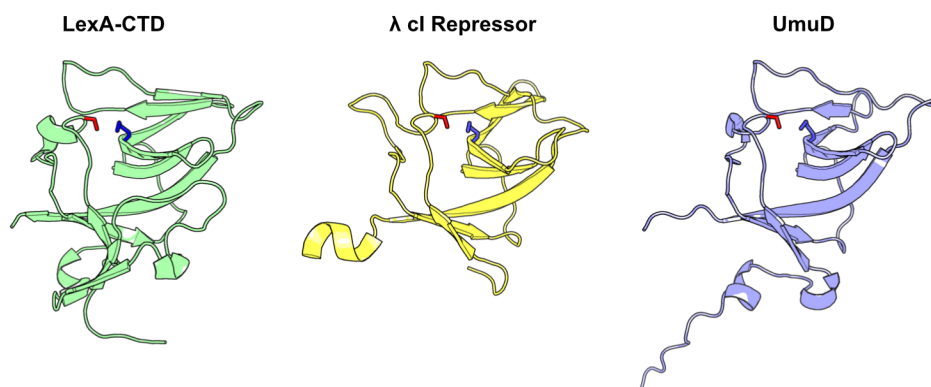


Figure 5.2. Examples of LexA-Family Proteins. Crystal structures of λ cl repressor (PDB: 3BDN) and UmuD (PDB: 1UMU) are represented in the same orientation as the C-terminal domain of LexA (PDB: 1JHC).

λ cl and UmuD, would be ideal initial experimental candidates in our competitive binding assay. If these proteins reveal differences in binding affinities for RecA* relative to LexA, follow-up studies could attempt to mutagenize or shuffle residues from different hypothesized recognition interfaces to localize determinants of weak or strong binding. Alternatively, similar binding affinities for these LexA-like proteins would suggest all three LexA family members share a similar molecular interface for RecA*.

Although our competition binding assay is designed for *E. coli* RecA* and a labeled form of *E. coli* LexA to provide the fluorescent signal, it should be readily adaptable to examine LexA or RecA proteins from other species. Prior work has demonstrated that *Pseudomonas aeruginosa* LexA can bind *E. coli* RecA and that *Bacillus subtilis* RecA can bind *E. coli* LexA (Mo, 2014; Namsaraev, 1998). This raises the intriguing possibility of testing a number of permutations of RecA and LexA proteins from different species for their ability to interact. The modularity of the Acd-labeled LexA competition binding assay we describe in Chapter 4 should allow for the straightforward exchange of unlabeled *E. coli* LexA for unlabeled *Pseudomonas aeruginosa* LexA or even the switch of *E. coli* RecA for *Bacillus subtilis* RecA. At this stage, the remaining obstacle is the overexpression and purification of the bacterial homologs of *E. coli* RecA and LexA, an ongoing and non-trivial endeavor that others in the laboratory are pursuing. Hopefully, this analysis will reveal which features of either protein, LexA or RecA, are conserved and amenable to perturbation by small molecule inhibitors in an effort to disrupt this interaction and disarm the SOS response in other bacteria.

5.2 Probing the Conformational Dynamics of the LexA Peptide Loop

As we described in Chapter 4, our investigation into the interaction of LexA with RecA* represented an initial foray into studying the individual steps of the RecA*-mediated LexA autoproteolysis reaction pathway (Figure 5.1). By isolating our experiments through the use of catalytically-inactive LexA constructs, we were able to examine complex formation with high kinetic resolution. However, this binding step represents only one aspect of the overall LexA:RecA*

reaction, and the simplified binding model we proposed relies on a number of assumptions. A major simplifying assumption is that LexA exists as a single conformational species to which RecA* can readily bind. Several biochemical studies—including the work we present in Chapter 4—suggest that this may not be correct, and instead point towards an allosteric relationship between RecA* binding and the position of the mobile peptide loop in LexA (Giese, 2008; Kovačič, 2013; Mo, 2014). However, these conventional biochemical assays cannot distinguish conformational change from

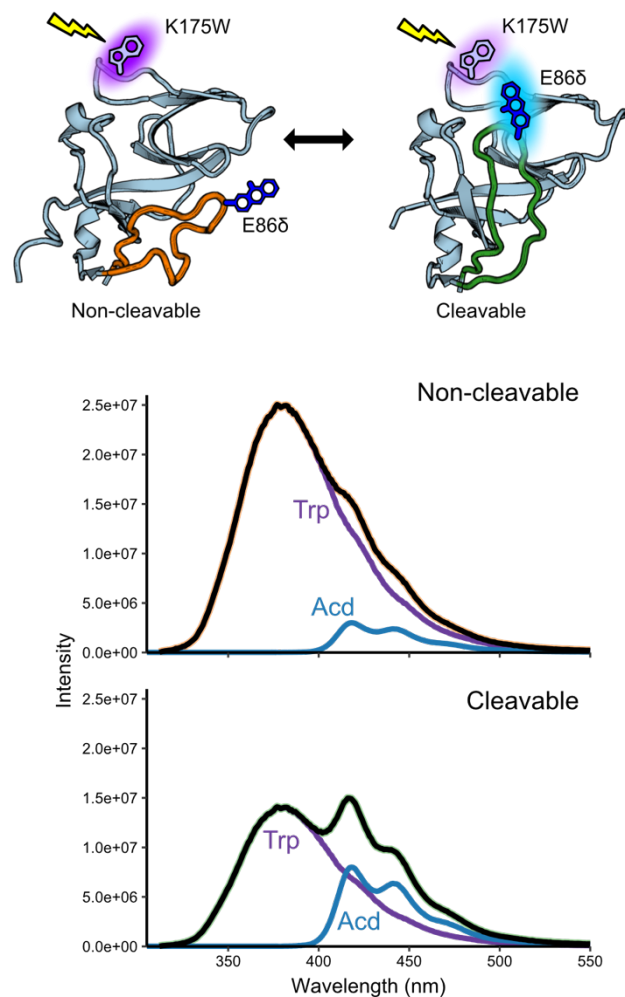


Figure 5.3. Revealing LexA's Conformational Dynamics. Top: Structures of the two snapshots of the mobile peptide loop, with positions indicated for the incorporation of a Trp/Acd FRET pair. At these positions, FRET efficiency is predicted to increase when the loop moves from its non-cleavable position to its cleavable one. Bottom: Fluorescence spectra of a non-cleavable and cleavable mutant excited at 285 nm. Deconvolution into the component emission spectra for each label reveals a relative increase in Acd fluorescence in the cleavable mutant, consistent with increased FRET for this conformation.

proteolysis, because the two are necessarily linked. As a result, no clear biophysical evidence of a dynamic conformational loop exists, other than the static snapshots from crystal structures (Y. Luo, 2001; A. P. P. Zhang, 2010). Therefore, questions remain regarding the conformational preference of the loop in the absence and presence of RecA*.

Normally, determining the real time conformational dynamics of a peptide loop, such as this one, in physiologic conditions poses a challenging experimental problem. However, the minimally-perturbing fluorescent amino acids we described in Chapter 1 appear particularly well-suited to solve this problem. Small, and bound directly to the backbone of the loop, a probe like Acd could sensitively report on the location of the loop. Rational installation of a Trp residue as a FRET donor to Acd elsewhere in LexA could provide a means of monitoring conformational dynamics via FRET efficiency. Previously, this laboratory identified positions in the *Pseudomonas aeruginosa* LexA loop (P4 and P2') that were tolerant to mutagenesis (Mo, 2014). In Chapter 3, we tested the incorporation of Acd into both of these positions in the LexA protein (Arg81 and Glu86) and found that, of the C-terminal positions tested, these positions showed high tolerability to Acd incorporation. In unpublished work, we also examined the effect of Acd incorporation at either of these positions on the proteolytic activity of LexA. Surprisingly, Acd incorporation at either Arg81 or Glu86 did not dramatically alter the rate of basal auto-proteolysis or impact the RecA*-mediated proteolysis reaction (data not shown), suggesting that this approach may be feasible for monitoring loop conformational dynamics.

To gather preliminary evidence as to whether we could monitor the position of the loop using the genetically-encoded Trp/Acd FRET pair, we first generated amino acid substitutions in LexA that, through a combination of biochemical and structural insights, are believed to putatively trap the loop either in its cleavable conformation ("quadruple mutant" = L89P, Q92F, E152A, K156A) or its non-cleavable conformation (G85D). By swapping the native Trp201 for a Phe, we could then genetically encode Acd into position 86 and a new Trp residue into position 175 (Figure 5.3). Using the crystal structure as a guide, the distance between Acd and Trp shifts from nearly

30 Å away in the non-cleavable conformation to about 10 Å away in the cleavable conformation, a range that encompasses the calculated Förster radius (at which FRET efficiency is 50%) of the Acd/Trp pair of 23 Å (Speight 2013). Initial fluorescence spectra of these two “conformationally-trapped” LexA proteins revealed an increase in fluorescence due to Acd emission and a concomitant decrease in Trp emission for the cleavable conformation compared to the non-cleavable conformation (Figure 5.3). Though preliminary, these results strongly align with a model in which residues around the active site of LexA can modulate the conformational preference of the mobile peptide loop. Future studies will investigate the rate of loop interconversion and how the conformation of native LexA, without any “trapping” amino acid substitutions, is affected by the addition of activated RecA*.

These preliminary results suggest a promising method for directly investigating the conformational dynamics of the C-terminal loop in LexA. However, certain drawbacks of the Trp/Acd FRET pair should be considered. Reliance on the canonical Trp as a FRET donor can complicate interpretations, especially if native Trp residues are not removed. While this was straightforward for exchanging a Trp201 residue in LexA for Phe, exchanging the native Trp residues in RecA is non-trivial and may alter its native function (Berger, 2001). Further, Acd itself can be excited at wavelengths near the Trp excitation maximum; due to direction excitation, additional single-label controls are typically required to deconvolute FRET data. Instead, deployment of an entirely non-canonical, genetically-encoded FRET pair with lower energy excitation requirements would be preferable. Recently, the Petersson group has launched investigations into Acd derivatives with improved fluorescence properties (Sungwienwong, 2018). One of these derivatives, aminoacridonylalanine (Aad), served as an improved FRET acceptor for another unnatural fluorescent amino acid, methoxycoumarin (Mcm). This improved Mcm/Aad FRET pair could avoid the aforementioned difficulties with Trp/Acd while retaining the benefits of using minimally-perturbing labels to probe the loop in the first place. Using this pair in LexA,

however, would first necessitate the evolution of a specific synthetase for each Uaa, an effort that is currently ongoing.

5.3 Identifying a Minimal LexA:RecA* Complex

Throughout this work, we have documented a number of studies aimed at structurally and biochemically characterizing the interface between LexA and RecA* (Kovačič, 2013; Mustard, 2000; Yu, 1993). Disrupting the LexA:RecA* interface has been of significant interest in the scientific community for decades because the SOS response is directly involved in contributing to the acquisition of antibiotic resistance. Despite a plethora of work, a consensus model of this interaction has yet to emerge. Parallel to these studies, efforts to understand RecA's structure and function with respect to its role in homologous recombination have yielded two instructive insights that should be applicable to the SOS field. First, the biology of RecA is tremendously complex, especially considering that its activated form requires the polymerization of many RecA subunits along exposed, single-stranded DNA (Bell, 2016). In all likelihood, the complicated dynamics of RecA* nucleoprotein filaments add to the inherent difficulty of studying the LexA:RecA* interaction. Second, RecA's complex and dynamic biology can be overcome through clever biochemical manipulations. Pavletich and colleagues biochemically simplified the nucleoprotein filaments by generating concatenated RecA subunits with N- and C-terminal modifications that restrict polymerization (Z. Chen, 2008). Purification of these oligomeric RecA constructs, consisting of five or six RecA monomers covalently tethered together, was an essential first step for solving the crystal structure of the activated RecA* nucleoprotein filament.

Drawing inspiration from the approach used for crystallizing RecA* in its active nucleoprotein conformation, we have explored the feasibility of generating covalently-linked and polymerization-restricted RecA constructs each consisting of a discrete number of RecA subunits—two, three, four, or five—translationally fused together (Figure 5.4). Each subunit is covalently tethered with an engineered flexible peptide linker. A truncation of the N-terminal RecA subunit and a trio of amino acid substitutions in the C-terminal RecA subunit should alter the ability of these

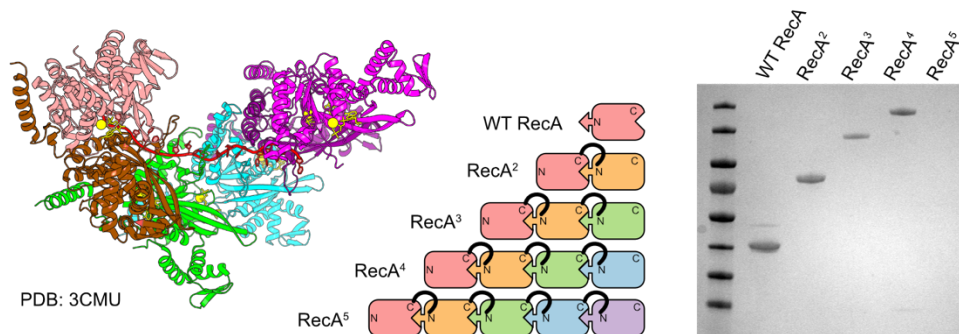


Figure 5.4. Design of Oligomeric RecA Proteins. Left: Published crystal structure of the activated RecA* filament using an oligomeric RecA construct with six RecA subunits. Middle: Schematic of designed oligomeric RecA constructs for biochemical characterization. Internal RecA subunits are capable of polymerizing, but the N- and C-terminal subunits are modified to prevent polymerization. Right: SDS-PAGE gel of purified oligomeric RecA proteins.

“capping” subunits to engage in polymerization. Despite these modifications, we show that we can successfully overexpress and purify these differently-sized RecA^N constructs (Figure 5.4). Optimization of protocols for large-scale purification of each construct is ongoing, and following recovery of each RecA^N, we expect to perform a full panel of biochemical characterization to assess how the function of RecA is altered as the length of its nucleoprotein filament is constrained.

A major aim of these studies is identifying the minimal number of RecA subunits in an activated RecA* filament that is required to bind to and promote the proteolysis of LexA. Characterization of this minimal RecA* unit will provide a valuable biochemical tool for further investigation into the pathway. Assuming that three RecA subunits comprise the minimal RecA* unit, a covalently-linked construct would enable us to introduce amino acid substitutions into the first, second, and third RecA subunits individually. The ability to precisely modify specific residues along an activated RecA* filament positions us to begin precisely mapping the LexA:RecA* interface. Using our findings from Chapter 3, we can begin engineering small fluorescent amino acids at the dozens of positions in both proteins that tolerated incorporation of Acd. With the right combination of FRET donor and acceptor, such as an Mcm/Aad pair (see Probing the Conformational Dynamics of the LexA Peptide Loop), measurements of FRET efficiency between

an Mcm-labeled RecA³ protein and an Aad-labeled LexA protein could indicate whether or not these positions are in close proximity in the bound complex. Even by preparing only a small number of labeled proteins with different probes in different positions, the permutations with which interposition distances between RecA* and LexA can be measured begin to grow exponentially, ensuring the reproducibility and redundancy in this approach. Given the high sensitivity of FRET efficiency measurements, a high resolution model of the LexA:RecA* complex using these distance constraints could be mapped. Additionally, our identification of RecA* binding determinants in LexA in Chapter 4: suggests minimal structural requirements in LexA for binding to RecA*. By combining this minimal LexA construct to a putative minimal RecA* construct, the opportunity to crystallize this protein complex for the first time becomes more promising. These complementary approaches highlight the value of reductionistic biochemistry to decrease the complexity of an inherently dynamic system. Thus for the first time, elucidating this elusive interaction may be possible.

5.4 Concluding Remarks

This thesis provides a framework for the development of Uaa labeling techniques and their application to the study of protein dynamics within biologic pathways. Specifically, we improve the methods underlying genetic incorporation of a fluorescent amino acid, Acd, and using this label as a probe to gain insight into the molecular mechanisms activating the bacterial DNA damage response. Given the growing interest in the use of unnatural amino acids to study fundamental or applied biology, any insight into expanding the scope of this technology is valuable. Our findings from Chapter 2 will inform efforts to evolve orthogonal tRNA synthetases with increased specificity for increasingly diverse arrays of unnatural amino acid substrates. Our conclusions from Chapter 3 offer a general strategy for adapting the technology of unnatural amino acid mutagenesis to new proteins and biological systems. As practitioners of genetic code expansion continue to test the boundaries of this technique, continued optimization and refinement of these methods will be required to expand their reach across scientific fields and disciplines.

A valid critique of our work questions the necessity of using a fluorescent amino acid label to produce the mechanistic findings into the SOS response shown in Chapter 4. Certainly, other non-fluorescent biochemical and biophysical techniques are well-established and could provide some insight into the affinity and mode of binding between LexA and RecA*. Already, other groups had made preliminary progress towards this question, either using gel-based assays or surface plasmon resonance techniques (Giese, 2008; Kovačič, 2013). However, the benefits of using fluorescence spectroscopy to study protein dynamics, particularly in this work, strongly outweigh the time and effort required to design and validate fluorescently-labeled proteins. Incorporation of Acd into LexA provided us with a minimally-perturbing and maximally-sensitive probe of the dynamics of LexA's structure and function in real time. Changes in the fluorescence anisotropy of our Acd-labeled LexA construct directly report on specific binding to RecA*, and, due to the insensitivity of fluorescence polarization to small changes in concentration or environmental conditions, this signal is less prone to artifacts and misinterpretation (Pope, 1999). By improving the efficiency for specific Acd incorporation in Chapter 2 and identifying tolerant positions both in LexA and RecA in Chapter 3, we directly addressed two of the major challenges commonly associated with these techniques. As a result, the field is well-positioned to begin investigating the other steps of the RecA*-mediated LexA autoproteolysis reaction using fluorescence spectroscopy. In contrast to the binding step we focused on in Chapter 4, the protein dynamics involved in other steps, including conformational changes or domain rearrangements, would be even more difficult to study with non-fluorescent techniques.

This thesis highlights the immense challenge of investigating the protein dynamics required to activate the bacterial SOS response. We can safely assume that RecA*-mediated LexA autoproteolysis is a complicated and dynamic process *in vivo*, and, in this body of work, we targeted and aimed to explain only one aspect of that reaction—the formation of the LexA:RecA* complex—and study it *in vitro*. Likely, elucidating other mechanistic aspects of this reaction will offer new challenges and require comparable degrees of effort. Further, additional work is required to

transform these insights into practical methods towards inhibiting RecA*-mediated LexA autoproteolysis. An ongoing effort in this laboratory is focused on characterizing small molecule inhibitors of RecA*-mediated autoproteolysis reaction (Mo, 2018). It is anticipated that we will combine insights from these lead compounds with our increasing mechanistic understanding of the LexA autoproteolysis reaction to deliver promising therapeutic strategies for disarming the SOS response—not only in a laboratory strain of *E. coli* but also in a wide range of pathogenic bacteria. While the efforts described in this work represent one incremental effort towards realizing this goal, the threats posed by antimicrobial resistant organisms—losing our ability to treat common infections, or to perform routine surgeries, or to keep immunocompromised populations alive—make such efforts arguably essential for the survival of our species.

APPENDIX

Supporting Information for Chapter 2

General Information

Materials. L-Tyrosine, thionyl chloride, di-*tert*-butyl dicarbonate (Boc anhydride), methyl 2-aminobenzoate, and phenylsepharose CL-4B resin were purchased from Sigma-Aldrich (St. Louis, MO, USA). *N*-Phenyl-bis(trifluoromethane sulfonylimide) was purchased from Oakwood Chemical (West Columbia, SC, USA). *E. coli* BL21(DE3) cells were purchased from Stratagene (La Jolla, CA, USA). *E. coli* ElectroMAX DH10B cells were purchased from Invitrogen (Grand Island, NY, USA). Milli-Q filtered (18 M Ω) water was used for all solutions (Millipore; Billerica, MA, USA). Bradford reagent assay kits were purchased from BioRAD (Hercules, CA, USA). Amicon Ultra centrifugal filter units (3 kDa MWCO) were purchased from EMD Millipore. All other reagents and solvents were purchased from Fisher Scientific or Sigma-Aldrich unless otherwise specified. DNA sequencing was performed at the University of Pennsylvania DNA sequencing facility.

Instruments. Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. UV-Vis absorption spectra were acquired on a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III instrument. Analytical HPLC was performed on an Agilent 1100 Series HPLC system. Preparative HPLC was performed on a Varian Prostar HPLC system (currently Agilent Technologies). HPLC columns were purchased from W. R. Grace & Company (Columbia, MD, USA).

Chemical Synthesis

Methyl-(S)-2-((*tert*-butoxycarbonyl)amino)-3-(4-(((trifluoromethyl)sulfonyl)oxy)phenyl)propanoate (Boc-L-Tyr(OTf)-OMe, 3). L-Tyrosine (**2**) (5.07 g, 27.99 mmol) was reacted with SOCl₂ 10.21 mL (139.94 mmol) in 50 mL of cooled MeOH. The reaction was allowed to stir overnight at room temperature under Ar. The reaction mixture was concentrated under vacuum and washed with MeOH. 6.44 g of crude L-tyrosine methyl ester (99% yield) was obtained after drying under vacuum overnight. The crude product was further dissolved in 120 mL tetrahydrofuran (THF)/H₂O (3:1). Na₂CO₃ 3.37 g (31.80 mmol) was then added, and the reaction was stirred for 10 min. Then 6.98 g Boc₂O was added, and the solution was stirred at room temperature overnight. 100 mL H₂O was added to the solution and then the reaction was acidified with 3 M HCl to pH 3. The organic layer was extracted with ethyl acetate (EtOAc), dried with MgSO₄, and concentrated under reduced pressure. The product can be recrystallized by dissolution in THF/hexane (3:1). *N*-Boc-L-tyrosine methyl ester (4.50 g, 15.24 mmol) was further dissolved in 30 mL CH₂Cl₂. Triethylamine (6.38 mL, 45.72 mmol), 4-(dimethylamino) pyridine (0.186 g, 1.524 mmol), and *N*-phenyl bis(trifluoromethanesulfonylimide) (7.62g, 21.34mmol) were then added, respectively. The reaction was stirred overnight. The reaction mixture was concentrated under reduced pressure and worked up with NH₄Cl and then extracted

by EtOAc. The crude product was purified by flash chromatography, eluting with 8:2 hexanes/EtOAc. (R_f = 0.26). The purified product Boc-L-Tyr-OTf-OMe (**3**) was obtained as colorless oil 5.77 g (88.62% yield). ESI-LRMS and NMR spectra matched previous reports (Speight, 2013).

(S)-Methyl-2-((4-(2-((*tert*-butoxycarbonyl)amino)-3-methoxy-3-oxopropyl)phenyl)amino)

benzoate (5). 75 mL degassed toluene was added to Boc-L-Tyr-OTf-OMe (**3**) (3.000 g, 7.02 mmol) in a dried round-bottom flask, followed by methyl 2-aminobenzoate (1200 μ L, 9.27 mmol). The solution was degassed with Ar for 5 min. Then palladium(II) acetate (0.082 g, 0.365 mmol), and racemic 2,2'-bis(diphenylphosphino)-1,1'-binaphthyl (0.054 g, 0.087 mmol) were added to the flask. Cesium carbonate (6.88 g, 21.1 mmol) was ground and added to the flask. The flask was then fitted with a reflux condenser and heated to 135 °C for 23 h. After the solution was allowed to cool to ambient temperature, the contents were filtered through a short plug of silica gel using CH₂Cl₂ to transfer the material to the silica (200 mL), and then ethyl acetate (300 mL) was used to elute the product. The clarified solution was then concentrated under reduced pressure and further purified by flash column chromatography (9:1 Hexane/EtOAc) afforded 2.759 g of compound **5** (6.44 mmol, 91.7%) as a yellow oil. (R_f = 0.2 in 85:15 Hexane/EtOAc) ESI-LRMS and NMR spectra matched previous reports (Speight, 2013).

Acridon-2-ylalanine (Acid, 1) (H₂SO₄ route). A solution of 13.5 M sulfuric acid (12 mL) was added to a flask containing **5** (1.02 g, 2.38 mmol). The flask was then fitted with a reflux condenser and heated to 115 °C for 16 h in an oil bath. 80 mL water was then added to the flask and allowed to stir for 15 min. The reaction was then removed from the hot oil bath and allowed to cool down. Upon reaching ambient temperature, the solution was cooled to 4 °C and allowed to stand for 2 h. 100 g of ion-exchange resin (Dowex[®] 50WX8 hydrogen form, strongly acidic cation exchange resin) was made into a slurry with 1.8 M aqueous H₂SO₄ and applied to a flash chromatography column. The resin was washed with 350 mL 1.8 M aqueous H₂SO₄, 2 L of water, 1 L of 1.5 M aqueous NH₄OH, and 4 L of water. Following these washes, the resin was dried by passing air through the column. The cooled Acid solution was then vacuum filtered on a Büchner funnel to remove precipitated material, and the clarified solution was applied to the washed and dried ion-exchange resin. The resulting resin slurry was shaken in the chromatography column for 5 min before the solution was drained. This solution was then reapplied to the dried resin and shaken for an additional 5 min. The twice-passed solution was then set aside. The loaded resin was washed with 4 L water before the compound of interest was eluted with 1.45 L of 1.5 M NH₄OH. The solution was concentrated to 50 mL by rotary evaporation, and then lyophilized to dryness, yielding a crop of Acid (**1**) as a yellow powder (0.6375 g 2.26 mmol 94.9%). The ion-exchange resin was recycled by washing with 4 L of water and dried until further use. To maximize yield, the twice-passed solution was reapplied to washed and dried ion-exchange resin, and the process was repeated to yield a second crop of Acid (0.0232 g of 0.082 mmol 3.4%).

Acridon-2-ylalanine (Acid, 1) (PPA route). Compound **5** 2.16 g (5.04 mmol) was added to a round-bottom flask. Then 50 mL THF was added and the solution cooled to 4 °C. LiOH (3.31 g, 138.15 mmol) was dissolved with 150 mL water. The LiOH solution was then slowly added to the reaction. The mixture was stirred under Ar at 4 °C overnight. The pH of the reaction mixture was adjusted to 3 with 6 M HCl. Crude product was extracted with CH₂Cl₂ (3 x 100 mL), dried with MgSO₄, and concentrated under reduced pressure. The

resulting crude oil was re-dissolved in 25 mL CH₂Cl₂ and cooled to 4 °C for 16 h. A first crop of recrystallized product was collected by vacuum filtration and washed with 100 mL of cold CH₂Cl₂ (1.63 g, 4.08 mmol, 80.8 %). The product was obtained as a white powder (82.6 % yield). The product was further reacted with polyphosphoric acid (PPA).

66.62 g of polyphosphoric acid (PPA) was added into a round-bottom flask with a stir bar. A round-bottom flask was heated to 135 °C in an oil bath. The deprotected compound **7** (1.43 g, 3.57 mmol) was added to the flask and stirred for 2 h. Then 50 mL water was slowly added (2 mL portions over 10 min) and the reaction allowed to cool to 60 °C. After stirring for 1 h at 60 °C, the reaction was cooled to ambient temperature. Insoluble impurities were removed by vacuum filtration and the clarified solution was adjusted the pH to 4 by addition of 8 M NaOH. Then the solution was cooled to 4 °C for precipitation about 16 h. Crude product was collected by vacuum filtration. After drying, the crude material was re-suspended in 50 mL water and brought into solution by adjusting the pH to 9.0. Insoluble impurities were removed by vacuum filtration. The pH of the clarified solution was then adjusted to 5.5 with 6 M HCl. The yellow precipitate was collected by vacuum filtration and dried overnight affording 0.66 g of Acd (**1**) (1.86 mmol, 65.3% yield).

Synthetase Selection

Positive and Negative Selection for Acd Specificity. A standard double sieve selection was performed in DH10B *Escherichia coli* (*E. coli*) with *Methanococcus jannaschii* (Mj) aminoacyl tRNA synthetase (RS) library 3D (3D-Lib) (Blizzard, 2015). The first positive and negative rounds of selection contained 1 mM Acd and no ncAA in the selection media, respectively. The second positive selection contained 1 mM Acd, while the second negative selection contained 1 mM Npf in the media. One microliter of the remaining library resulting from the second positive and the second negative rounds of selection was transformed with 60 µL of pALS plasmid containing DH10B cells (Blizzard, 2015). The pALS plasmid contains a sfGFP reporter with a TAG codon at residue 150 as well as tyrosyl-tRNA_{CUA}. The cells were rescued for 1 h in 1 mL of SOC media (37 °C, 250 rpm). A 250 µL or 50 µL volume of cells from each library were plated on autoinducing media agar with 25 µg/mL kanamycin (Kn), 50 µg/mL tetracycline (Tet), and 1 mM Acd. Plates were grown at 37 °C for 24 h, and then grown on the bench top, at room temperature, for an additional 24 h. Autoinducing agar plates were prepared by combining the reagents with an autoclaved solution of 40 g of agar in 400 mL water (Blizzard, 2015). Sterile water was added to a final volume of 500 mL. Antibiotics were added to a final concentration of 25 µg/mL Tet and 50 µg/mL Kn.

A total of 96 visually green colonies were selected from the 1 mM Acd plates and used to inoculate a 96-well plate containing 0.5 mL per well non-inducing media with 25 µg/mL Kn and 25 µg/mL Tet (Blizzard, 2015). After 24 h of growth (37 °C, 250 rpm), 5 µL of these non-inducing samples were used to inoculate three 96-well plates with 0.5 mL autoinduction media containing 25 µg/mL Kn and 25 µg/mL Tet. For the three 96-well plates, one plate contained 1 mM Acd, the second contained 1 mM Npf and the third contained no ncAA.

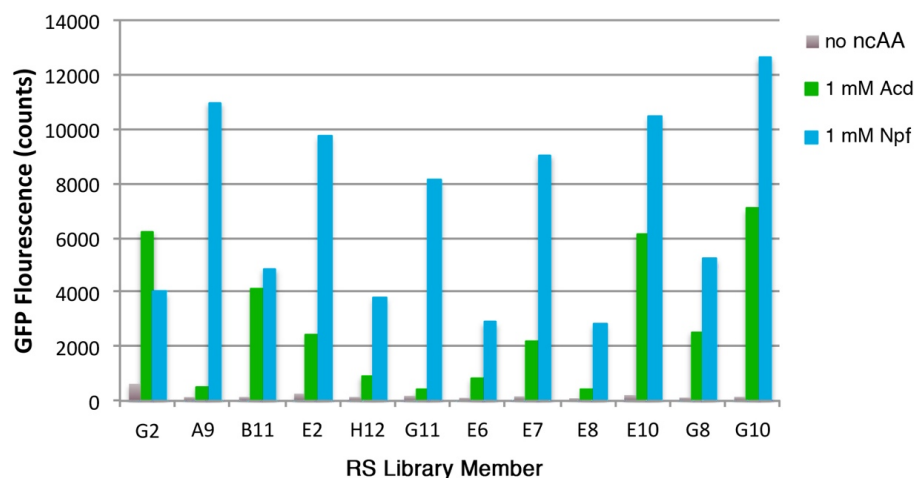
Fluorescence measurements of the cultures were collected 36 h after inoculation using a BIOTEK Synergy 2 Microplate Reader. The emission was measured at 528 nm (20 nm bandwidth) with excitation at 485 nm (20 nm bandwidth). Samples were prepared by diluting suspended cells directly from culture 2-fold with phosphate buffer saline (PBS, 10 mM PO₄³⁻, 137 mM NaCl, 2.7 mM KCl pH 7.4). Eleven unique RSs

showed high efficiency with Acd, good fidelity without ncAA, and reduced ability to incorporate Npf. Fluorescence data from the G2, F9, G11, and A9 clones are shown in Figure 2.2 in the main text, and the sequences of all 11 RS variants are shown in Supplemental Table 1.

Quantitative Analysis of Highest-Fluorescing Clones. Non-inducing cultures (3 mL) with 25 µg/mL Kn and 25 µg/mL Tet were grown to saturation (37 °C with shaking at 250 rpm) from the top eleven expressions in the 96 well plate analysis. Autoinduction cultures (5 mL) with 25 µg/mL Kn and 25 µg/mL Tet were inoculated with 30 µL of non-inducing cultures and grown with 1 mM Acd, 1 mM Npf, or without ncAA at 37 °C with shaking at 250 rpm. All cultures were grown in triplicate. After approximately 40 hours, fluorescence was assessed as described (Supplemental Figure 1).

Supplemental Table 1: Sequence of top performing AcdRSs.

Variant	Residue number								
	32	65	70	108	109	155	158	159	162
WT	Y	L	H	F	Q	Q	D	I	L
G2	G	E	H	W	M	Q	S	I	K
A9	A	D	H	L	E	Q	S	I	T
B11	G	V	H	L	L	Q	E	I	T
E2	G	V	H	A	S	Q	S	I	E
H12	A	W	H	T	E	Q	G	I	A
G11	A	Y	H	H	E	Q	S	I	S
E6	G	P	H	E	S	Q	G	I	H
E7	G	P	H	E	S	Q	G	I	H
E10	G	I	H	E	S	Q	G	I	T
G8	A	P	H	S	Q	Q	D	I	A
G10	A	W	H	R	E	Q	G	I	S



Supplemental Figure 1. Fluorescence measurements of RSs with GFP reporter. Grey, green, and blue represent fluorescence from colonies induced in media containing no ncAA, 1 mM Acd, or 1 mM Npf, respectively. Expressions of 500 μ L were grown for 48 hours before 2-fold dilution of suspended cells directly from culture with PBS. Fluorescence measurements were collected using a BIOTEK® Synergy 2 Microplate Reader.

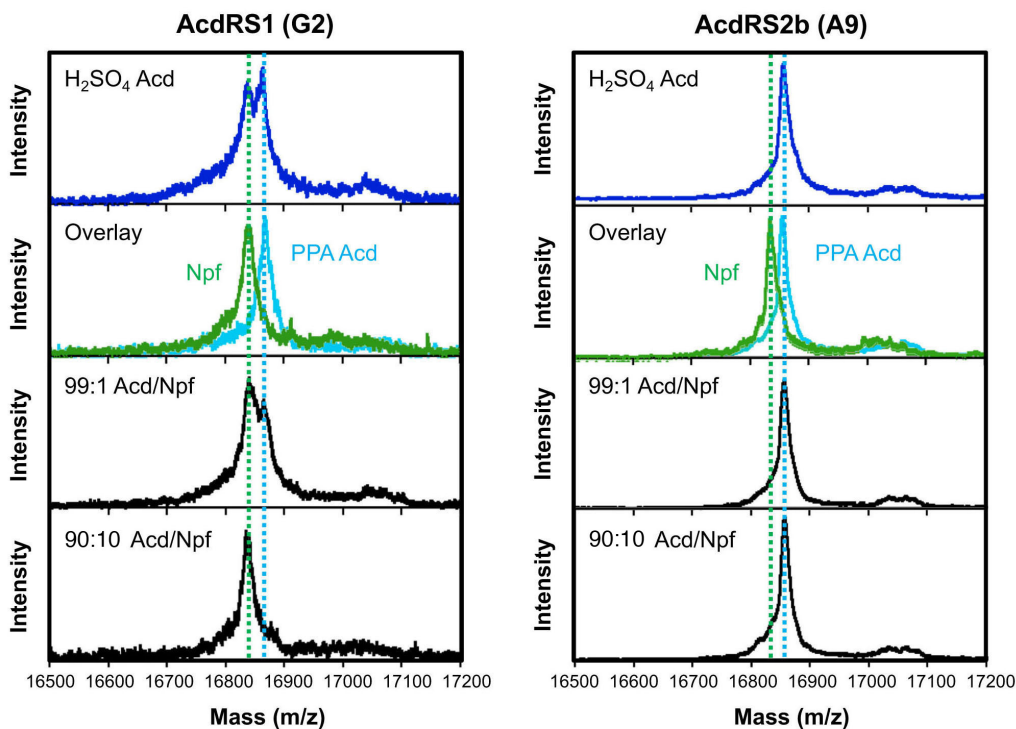
***In Vivo* Selectivity Analysis**

Acid Mutant Calmodulin Protein Expression. For calmodulin (CaM) mutant expression, pCaM-L₁₁₃TAG and pDule2-AcdRS plasmids were used to transform *E. coli* BL21(DE3) cells (Speight, 2013). The pDule2-AcdRS1, pDule2-AcdRS2a, and pDule2-AcdRS2b plasmids encode AcdRS1(G2), AcdRS2a(G11) and AcdRS2b(A9), respectively. The pCaM-L₁₁₃TAG plasmid confers ampicillin (Amp) resistance and the pDule2-AcdRS plasmids confer streptomycin (Strep) and spectinomycin (Spec) resistance. Transformed cells were selected on the basis of Amp and Strep or Amp and Spec resistance. Single colonies were used to inoculate 5 mL of LB media supplemented with Amp and Strep (100 μ g/mL each). To an autoclaved solution containing 42 mM Na₂HPO₄, 22 mM KH₂PO₄, 19 mM NH₄Cl, and 86 mM NaCl (M9 salts), the following autoclaved solutions were added per liter of M9 salts: 1 mL of 2 M MgSO₄, 1 mL of 15 mg/mL FeCl₂ (in 1.0 M HCl), 1 mL of 15 mg/mL ZnCl₂ (in acidified H₂O), 6.25 mL of 40% glucose, 100 μ L of 1 M CaCl₂ and 2 mL of 10% Bacto™ Yeast Extract. The primary 5 mL culture was incubated at 37 °C with shaking at 250 rpm for 4 h. Upon reaching saturation, the primary culture was added to 1 L of M9 minimal media supplemented with Amp and Strep (or Spec). The 100 mL culture was incubated at 37 °C with shaking at 250 rpm until the OD₆₀₀ reached 0.7. At this point, the different amounts of unnatural amino acid were added to the culture depending on the following conditions:

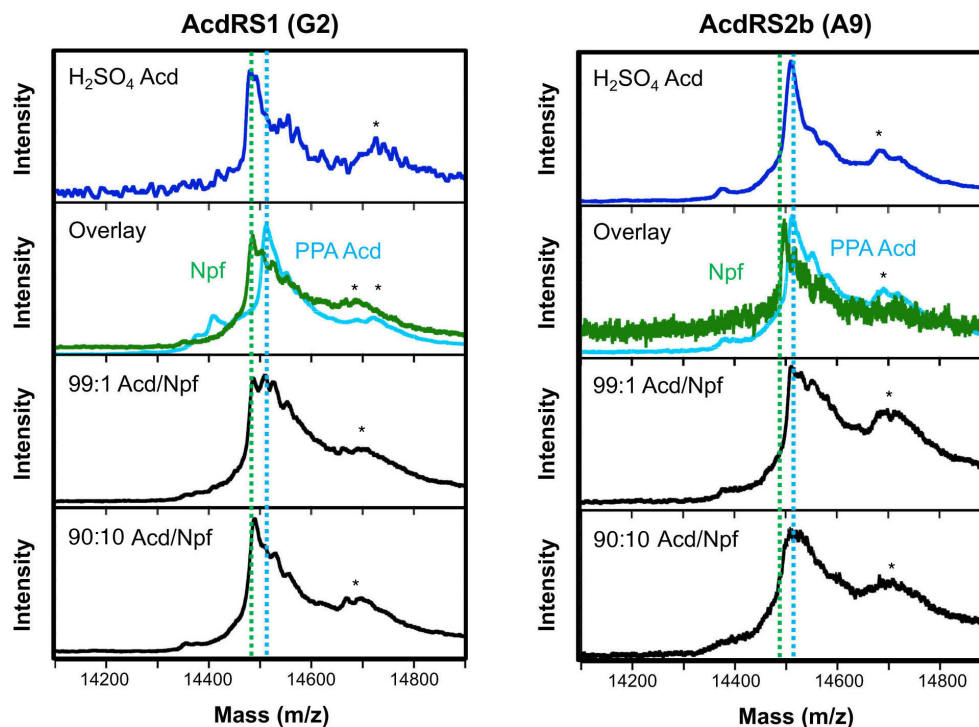
- A: a solution of 28.23 mg Acid (1) from H₂SO₄ route in 5 mL sterile water
- B: a solution of 28.23 mg Acid (1) from PPA route
- C: a solution of 25.63 mg Npf (6)
- D: a solution of 0.26 mg Npf (6) and 27.95 mg Acid (1)
- E: a solution of 2.56 mg Npf (6) and 25.41 mg Acid (1)

Acid was solubilized with 5 drops 10 M NaOH. The protein expression was then induced with IPTG. The culture was incubated at 37 °C for an additional 16 h. The cells were harvested at 5000 x g for 15 min, and the resulting pellet was suspended in 15 mL of 3-(N-morpholino) propanesulfonic acid (MOPS) resuspension buffer (50 mM MOPS, 100 mM KCl, 1 mM EDTA, pH 7.5). Following sonication, the cell lysate was allowed to cool on ice for 5 min. CaCl₂ was added to the sonicated lysate to a final concentration of 5 mM prior to centrifugation for 20 min at 30,000 x g, 4 °C.

CaM was purified from the cleared cell lysate using a phenyl-sepharose (PhS) CL-4B column with EDTA as eluent. Using a total resin bed volume of 10 mL, the column was first equilibrated with 4 column volumes of PhS Buffer A (50 mM Tris base, 1 mM CaCl₂, pH 7.5). After the cleared cell lysate was loaded and allowed to pass through the resin, the column was washed with 4 column volumes of PhS Buffer A, 4 column volumes of high-salt PhS Buffer B (50 mM Tris base, 0.5 M NaCl, 0.1 mM CaCl₂, pH 7.5), and an additional 2 column volume washes of PhS Buffer A to restore low-salt conditions. CaM was eluted with PhS Buffer C (10 mM Tris base, 10 mM EDTA, pH 7.5) and collected in 1 mL fractions. The presence of protein was detected by SDS-PAGE. Fractions containing protein were combined and dialyzed against water for 16 h at 4 °C.



Supplemental Figure 2. In Vivo AcridRS Selectivity. CaM (UAG codon at 113) was expressed in minimal media containing 1 mM ncAA: Acrid, synthesized either using the H₂SO₄ route or PPA route, Npf, or either a 99:1 or 90:1 mixture of PPA Acrid and Npf. Expression was performed with either AcdRS1 (G2) or AcdRS2b (A9). AcdRS2b (A9) can tolerate at least 10% Npf in the growth media with no apparent Npf incorporation.



Supplemental Figure 3. In Vivo AcdRS Selectivity. α S (UAG codon at 94) was expressed in LB media containing 1 mM ncAA: Acd, synthesized either using the H_2SO_4 route or PPA route, Npf, or either a 99:1 or 90:1 mixture of PPA Acd and Npf. Expression was performed with either AcdRS1 (G2) or AcdRS2b (A9). AcdRS2b (A9) can tolerate at least 10% Npf in the growth media with no apparent Npf incorporation. * indicates MALDI matrix adduct. αSNpf_{94} , Calc'd: 14482; αSAcd_{94} , Calc'd: 14508.

Acd Mutant α -Synuclein Protein Expression. For α -synuclein (α S) mutant expression, *E. coli* BL21-Gold (DE3) cells were transformed with the α S-F₉₄TAG plasmid and a pDule2 plasmid containing the AcdRS and tRNA_{CUA} pair. Cells were selected for resistance to both Amp (100 $\mu\text{g}/\text{mL}$) and Strep (100 $\mu\text{g}/\text{mL}$) or Spec (50 $\mu\text{g}/\text{mL}$). Single colonies were used to inoculate 4 mL of LB media. The primary culture was grown at 37 °C with shaking at 250 rpm for 4 h. The 100 mL culture was incubated at 37 °C with shaking at 250 rpm until the OD₆₀₀ reached 0.8 AU. At this point, the different amounts of unnatural amino acid were added to the culture grown under conditions A-E as in the CaM protocol above. The cells were harvested at 5000 x g for 15 min and the resulting pellet was resuspended in 20 mM Tris, pH 8.0 with 1 mM PMSF and sonicated. Following sonication, the cell lysate was boiled for 15 min prior to centrifugation for 20 min at 13,200 x g, 4 °C. The cleared lysate was dialyzed overnight against 20 mM Tris, pH 8.0 at 4 °C prior to purified using a HiTrap Q HP column (GE Healthcare). FPLC fractions were dialyzed against 20 mM Tris, pH 8.0 and stored at 4 °C.

Trypsin Digest Analysis of Acd Mutants. Protein Acd mutants were precipitated using 1:4 8.75 M trichloroacetic acid/protein sample and incubated at 4 °C for 15 minutes. The precipitate was centrifuged for 15 min at 13,200 rpm to pellet protein. The protein pellet was then washed three times with cold acetone to

remove trace trichloroacetic acid. Trace acetone was removed by incubating protein pellets in a 95 °C water bath for 5 min open to the atmosphere. Protein pellets were then re-suspended in 6 M guanidinium hydrochloride with 50 mM Tris pH 8.0, and denatured by boiling at 95 °C for 10 minutes. Protein samples were then diluted to 0.75 M guanidinium hydrochloride with 50 mM Tris pH 7.6 and 1 mM calcium chloride. Sequencing grade modified trypsin (0.6 µg, Promega) was used to digest samples for 24 hours at 37 °C. Trypsin digest aliquots (1 µL) were combined with α-cyano-4-hydroxycinnamic acid (1 µL of a saturated solution in 1:1 H₂O/CH₃CN with 1 % TFA) and analyzed by MALDI-MS.

Peak intensity scaling to determine Acd/Npf selectivity was performed as follows: For PPA Acd expressions and Npf expressions, which should have only Acd or Npf in the CaM₁₀₈₋₁₁₆ trypsin fragment, a ratio of the peak for this fragment (CaM₁₀₈₋₁₁₆Npf₁₁₃: 1153.8 m/z, CaM₁₀₈₋₁₁₆Acd₁₁₃: 1179.8 m/z) was calculated relative to the peak for CaM₁₁₇₋₁₂₇ (1349.9 m/z). This ratio represents the relative ease of ionization of these

Supplemental Table 2: CaM 113 Trypsin Digest Intensity Scaling.

ncAAs	G2			A9		
	m/z	Peak Height	Scaled Peak	m/z	Peak Height	Scaled Peak
Acd (H ₂ SO ₄)	1153.8	391	2277	1153.7	13	57.6
	1179.8	8166	1458	1179.7	5878	1083.1
	1349.9	5661		1349.8	1348	
Acd (PPA)	1153.7	24		1153.7	38	
	1179.8	7574		1179.7	15487	
	1349.9	1091		1349.8	4132	
Npf	1153.8	518		1153.7	90	
	1179.8	15		1179.5	10	
	1349.9	3500		1349.8	493	
1% Npf/Acd	1153.8	1224	7281	1153.4	17	73
	1179.8	20839	3801	1179.7	5386	984
	1349.9	7968		1349.8	1922	
10% Npf/Acd	1153.8	3034	18197	1153.5	22	121
	1179.8	754	15	1179.7	1621	294
	1350.0	10266		1349.8	790	
CaM ₁₀₈₋₁₁₆ Acd ₁₁₃ / CaM ₁₁₇₋₁₂₇ Scaling Factor: 5.47						
CaM ₁₀₈₋₁₁₆ Npf ₁₁₃ / CaM ₁₁₇₋₁₂₇ Scaling Factor: 0.165						
CaM ₁₀₈₋₁₁₆ Npf ₁₁₃ / CaM ₁₁₇₋₁₂₇ Background Factor: 0.0156						
CaM ₁₀₈₋₁₁₆ Acd ₁₁₃ / CaM ₁₁₇₋₁₂₇ Background Factor: 0.0122						

fragments. Similarly, the intensities of the peaks for the absent fragment were used to determine a level of background counts at this mass. These values were used to correct the data to give the scaled peak intensities in Table S2. The scaled intensities were then used to calculate an Acd/Npf MS incorporation selectivity factor using the following equation:

$$MS\ Sel = (Scaled\ Acd\ Peak / Scaled\ Npf\ Peak) / (Acd / Npf\ ratio\ in\ media)$$

For each AcdRS, MS Sel values were averaged between 1% Npf and 10% Npf data sets. LexA and α S trypsin digests did not consistently show complete digestion, preventing such analysis.

LexA Cloning/Expression/Purification. To create a C-terminal His-tagged LexA expression construct, a recombinant *E. coli* *lexA* gene containing the S₁₁₉A catalytically-inactive mutation was PCR amplified with forward

A: 5'-GGCAGCCATATGAAAGCGTTAACGG-3'

and reverse

B: 5'-AATCTCGAGCCAGTCGCCGTTGC-3'

Primers (Mo, 2016). This amplified *lexA* gene was sub-cloned into the NdeI and XhoI sites in pET-41a(+) plasmid (EMD Millipore) to produce pET41-LexA-S₁₁₉A-HIS. To generate an amber stop codon in the S₆₀ position, the pET41-LexA-S₁₁₉A-HIS plasmid was amplified with forward

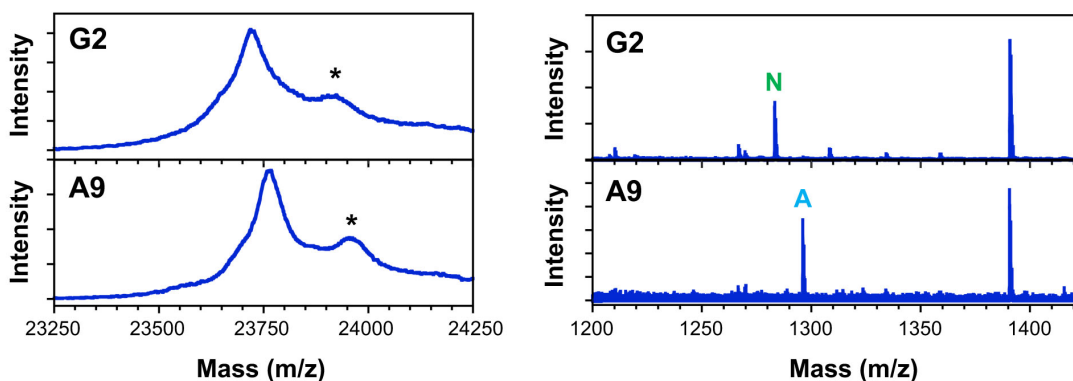
C: 5'-ATTGTTTAGGGCGCATCACGCGGAATTCGTC-3'

and reverse

D: 5'-GTGATGCGCCCTAAACAATTTCAATAACGCC-3'

primers designed for site-directed mutagenesis with Phusion polymerase, creating pET41-LexA-S₆₀TAG-S119A-HIS (L. Zheng, 2004).

Chemically competent BLR(DE3) cells (EMD Millipore) were transformed with either of the above pET41-LexA plasmids (Kn resistance) and either of the pDule2-AcdRS1a or pDule2-AcdRS2b plasmids



Supplemental Figure 4. In Vivo AcRS Selectivity. LexA (UAG codon at 60) was expressed in auto-induction media containing 1 mM Acd, synthesized using the H₂SO₄ route. Expression was performed with either AcdRS1 (G2) or AcdRS2b (A9). Although the whole protein peaks (left) are very broad, trypsin digest producing the LexA₅₃₋₆₄ fragment (right) indicates the presence of only Npf in the G2 sample and only Acd in the A9 sample. N denotes LexA₅₃₋₆₄Npf₆₀. A denotes LexA₅₃₋₆₄Acd₆₀. * indicates MALDI matrix adduct. LexA₅₃₋₆₄Npf₆₀, Calc'd: 1366.5, Obsv'd: 1366.6; LexA₅₃₋₆₄Acd₆₀, Calc'd: 1392.5, Obsv'd: 1392.4

(streptomycin resistance) and selected on LB + Kn (30 µg/mL) + Strep (100 µg/mL) plates. Isolated colonies were grown overnight to saturation in liquid LB + Kn + Strep. Overnight cultures were inoculated at a 1:100 ratio into 25 mL MD-5051 auto-inducing medium containing 150 µg/mL Kn, 100 µg/mL Strep, and 1 mM acridone (H₂SO₄ route) (Studier, 2014). Cultures were grown for 24 hours in sterile 250 mL polypropylene centrifuge tubes (Corning) in a shaking 37 °C incubator. Cells were harvested at 4000 x g for 30 min at 4 °C.

Cell pellets were resuspended in 4 mL of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.1) containing 1x BugBuster Protein Extraction Reagent (EMD Millipore), 0.5 mg/mL lysozyme, 25 U/µL benzonase, and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). After incubation at room temperature for 30 min, lysed cells were centrifuged at 20000 x g for 30 min at 4 °C. LexA proteins were purified from the clarified cell lysate using 0.5 mL HisPur cobalt spin columns per manufacturer's instructions (Thermo Scientific), using 300 mM imidazole as eluent. LexA proteins were detected by running purification samples on 15% SDS-PAGE gels. Elutions containing LexA were pooled and buffer-exchanged into low salt buffer (20 mM Tris-HCl, 25 mM NaCl, pH 7.6) using Vivaspin 500 concentration devices.

Cloning and Purification of Synthetases

AccdRS Cloning/Expression/Purification. Synthetic genes (IDT) for AccdRS1 and AccdRS2b were amplified with forward

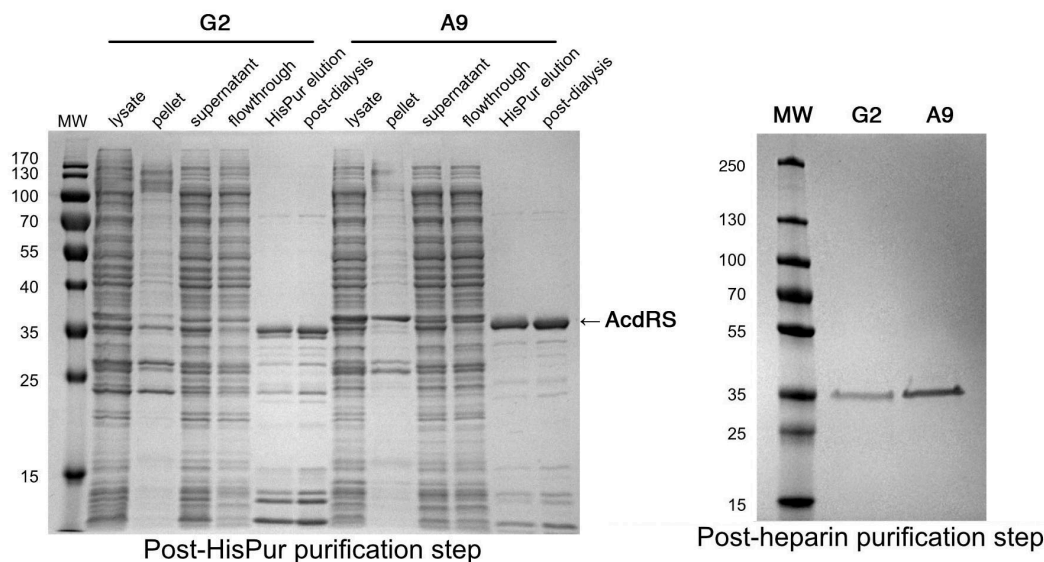
E: 5'-GAGCGGATAACAATTCCCCTCTAG-3'

and reverse

F: 5'-GTGGTGGTGCTCGAGTCTCTTTC-3'

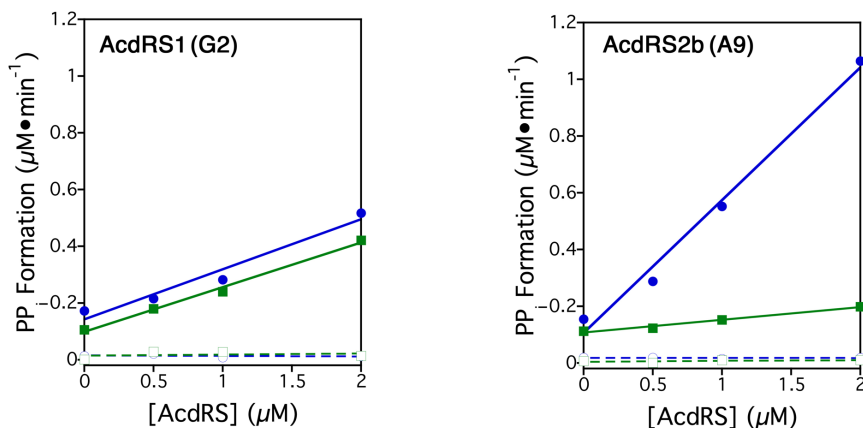
primers. The resulting PCR product was digested with NdeI and XhoI and sub-cloned into a pBAD24 plasmid modified with a C-terminal His tag, resulting in the three pBAD24-AccdRS-HIS constructs. Chemically competent BL21(DE3) cells were transformed with one of the three above pBAD24-AccdRS-HIS plasmids (Amp resistance) and selected on LB + Amp (100 µg/mL) plates. Isolated colonies were inoculated into 200 mL arabinose auto-inducing medium containing 100 µg/mL carbenicillin (Carb) (Hammill, 2007). Cultures were grown for 24 hours in 2 L flasks in a shaking 37° incubator. Cells were harvested at 4000g for 15 min at 4 °C.

Cell pellets were resuspended in 30 mL of wash buffer (50 mM HEPES-KOH, 500 mM NaCl, 10 mM imidazole, pH 7.9) containing 1 mM PMSF and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche). Cell suspensions were lysed using a French pressure cell (FA-032, Thermo Electron Corp) and then centrifuged at 20000g for 30 min at 4 °C (Sorvall RC 6+). Clarified lysates were loaded with a BioLogic LP System (Bio-Rad) onto a column containing 4 mL of packed HisPur resin (Thermo Scientific), and proteins were eluted with a linear imidazole gradient. AccdRS proteins were detected by running elution fractions on 12% SDS-PAGE gels. Elutions containing AccdRS were pooled and dialyzed three times against 2 L of 20 mM Tris-HCl pH 8.5, 50 mM NaCl, 10 mM 2-mercaptoethanol, and 20% glycerol. Dialyzed samples were loaded onto a 5 mL HiTrap Heparin HP column (GE Healthcare) and proteins were separated with a linear NaCl gradient. FPLC fractions were analyzed by 12% SDS-PAGE, and those containing AccdRS proteins were concentrated using Amicon Ultra centrifugal filter units (Sigma), buffer-exchanged (50 mM HEPES-KOH pH 7.5, 20 mM NaCl, 10 mM MgCl₂, 1 mM DTT, and 50% glycerol), and stored at -80 °C.



Supplemental Figure 6. AcdRS Purification. Left: SDS-PAGE gel showing purification samples of AcdRS1 (G2, lanes 2-7) and AcdRS2b (A9, lanes 8-13) taken throughout the cell lysis and HisPur affinity resin steps. Right: SDS-PAGE gel showing final, purified samples of both AcdRS enzymes following the purification scheme described in the supplemental text. For the lanes in each gel, an equal volume of sample and Laemmli buffer (Bio-Rad) were mixed, boiled at 95°, centrifuged, and loaded onto 12% SDS-PAGE gels, which were run for 50 min at 200 V.

AcdRS Activity Test. The activity of each AcdRS enzyme was verified by monitoring pyrophosphate release during the initial condensation reaction of Acd or Npf with ATP. To measure pyrophosphate, an enzyme-coupled, colorimetric assay recently described in the literature was used (Cestari, 2013). Briefly, pyrophosphate generated during each reaction is hydrolyzed by excess inorganic pyrophosphatase (PPiase), and free orthophosphate is incubated with malachite green under acidic conditions (Malachite Green



Supplemental Figure 5. AcdRS Malachite Green Activity Assay. Plots showing rate of inorganic pyrophosphate formation versus AcdRS concentration for AcdRS1 (G2, left) and AcdRS2b (A9, right). Data are from reactions at 37° containing either 1 mM Acd (blue) or 1 mM Npf (green) and in the presence (solid points) or absence (open points) of 0.2 mM ATP. The linear dependence of activity on AcdRS concentration was demonstrated by fitting a straight line through the data points.

Phosphate Assay Kit, Sigma) to generate a malachite green phosphomolybdate complex (600-660 nm absorbance). Several concentrations of each AcdRS enzyme (0, 0.5, 1.0, and 2.0 μ M) were incubated in reaction buffer (50 mM HEPES-KOH pH 7.5, 20 mM NaCl, 10 mM MgCl₂, 1 mM DTT) with 2 U/ml PPIase (NEB), 1 mM of either Acd or Npf, and with or without 0.2 mM ATP for 35 min at 37 °C. Reactions were quenched with chilled EDTA to a final concentration of 10 mM. Samples were developed per manufacturer's instructions in a clear 96-well plate (Corning Costar), and absorbances at 620 nm were read in an Infinite F200 plate reader (Tecan). Standard curves were used to interpolate observed phosphate concentrations, and the data describing calculated pyrophosphate release rate versus AcdRS concentration were fit to a straight line (Supplemental Figure 5).

***In Vitro* Aminoacylation Analysis**

Preparation and Labeling of tRNA Transcripts for Aminoacylation. The tRNA substrate for aminoacylation by the AcdRS1 (G2) and AcdRS2b (A9) isolates was prepared by *in vitro* transcription with T7 RNA polymerase (Rauch, 2016; Sherlin, 2001), using DNA templates encoding a self-cleaving hammerhead ribozyme immediately upstream of the T7 promoter (Price, 1995). DNA templates were assembled from four overlapping synthetic oligonucleotides, listed below (overlapping regions are underlined). The tRNA^{Tyr} species used is the orthogonalized tRNA (Orthog-tRNA) isolated by genetic selections when the *Mj* TyrRS-tRNA^{Tyr} system was first developed for ncAA incorporation (L. Wang, 2001).

G: 5'-AATTCCTGCAGTAATACGACTCACTATAGGGAGACCGGCTGATGAGTC-3'

H: 5'-CCGGGACGGTACCGGGTACCGTTTCGTCCTACGGACTCATCAGCCGGTCTCCC-3'

I: 5'-CCCGGTACCGTCCCGGCGGTAGTTCAGCCTGGTAGAACGGCGGACTGTAG-3'

J: 5'-TGGTCCGGCGGGCCGGATTGAACACGCGACATGCGGATCTACAGTCCGCCGTTCTACC-3'

Oligonucleotide assembly by PCR was performed as described (Rauch, 2016). A fifth oligonucleotide was used, along with oligonucleotide **G**, to amplify the assembled template:

K: 5'-[2'-OMe]U[2'-OMe]GGTCCGGCGGGCCGG-3'

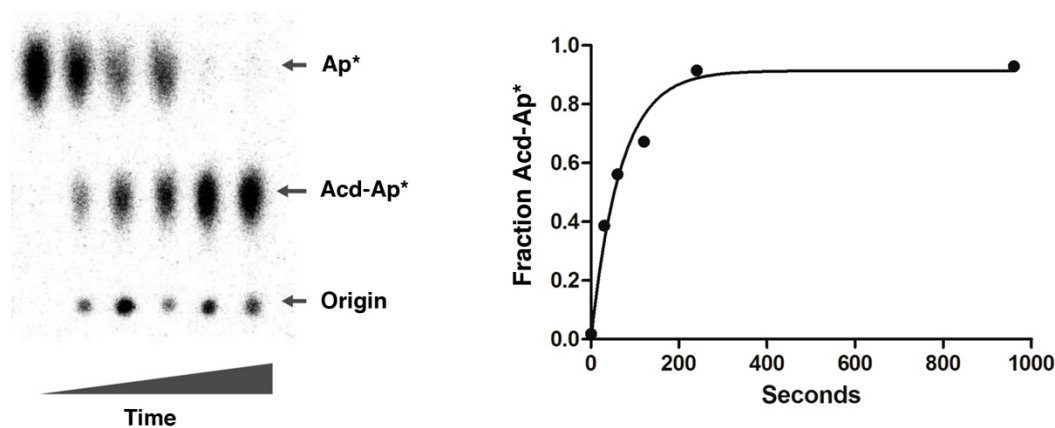
The 2'-O-methyl modifications inhibit run-over transcription by T7 RNA polymerase (Kao, 1999). Amplification of the full-length template, template purification, and transcription reactions were performed as described (Rauch, 2016). The enzyme for transcription was the T7 RNA polymerase Δ 172-173 variant, which was expressed and purified in the laboratory as described (Lyakhov, 1998). The transcription reaction mixture was incubated at 37 °C for 16-24 h.

Cleavage of the hammerhead ribozyme was performed as described (Rauch, 2016). After the reaction, the mature tRNA was purified from uncleaved transcripts and free hammerhead ribozyme by electrophoresis through a denaturing 15 cm gel containing 15% polyacrylamide:bisacrylamide (29:1), 8 M urea and 1X TBE. tRNA was visualized by UV shadowing and excised from the gel. tRNA was extracted from the gel band with 10 volumes of TE6 buffer (10 mM BisTris-HCl and 1 mM EDTA, pH 6.0), shaking for 16-24 h at room temperature. Insoluble gel debris were removed by centrifugation at 5000 rcf for 10 min. The supernatant was passed through a 0.45 μ m filter and washed with TE6 buffer using a centrifugal filtration device (Amicon; 10 kDa MWCO) such that urea was diluted to a concentration below 1 μ M.

The tRNA substrates for aminoacylation were radiolabeled at the 3' internucleotide linkage in accordance with established methods (Rauch, 2016; Wolfson, 2002), using *E. coli* nucleotidyltransferase, purified in the Perona laboratory as described (McGann, 1980). Prior to labeling, tRNA was denatured by incubating for 10 min at 65° C in ~50 μ L water, and refolded by slow-cooling to room temperature over the course of one hour.⁹ Labeled tRNA was purified by phenol:chloroform:isoamyl alcohol extraction (Sigma; 25:24:1; pH 7.8) followed by polyacrylamide gel electrophoresis. tRNA was then extracted from the gel band with 2 volumes of TE6 buffer without shaking, in a 16-24 h incubation at ambient temperature (Rauch, 2016).

Aminoacylation Reaction Assays. Aminoacylation assays were conducted at 37° C in siliconized conical tubes containing 20 μ L of 50 mM HEPES KOH (pH 7.5), 20 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 2.5 mM MgATP and various concentrations of amino acid. All reactions were conducted under single-turnover conditions, and control experiments were performed to ascertain saturating concentrations for tRNA. For every combination of enzyme (AcdRS1 or AcdRS2b) and amino acid (Acd or Npf), these experiments established that 1.0 μ M enzyme with 0.1 μ M tRNA allowed full tRNA binding. These concentrations were therefore employed in all measurements.

Methods used were similar to those employed in the Perona laboratory for the study of other RSs (Rauch, 2016; Rodríguez-Hernández, 2011). tRNAs were refolded in the presence of 33 mM MgCl₂ prior to aminoacylation, with refolding performed as described (Rauch, 2016). Time points in aminoacylation assays were quenched by diluting 1:3 into a solution containing 200 mM NaOAc (pH 5.2), 0.2% SDS, and subsequently digested for 15 minutes at room temperature with *Penicillium citrinum* P1 nuclease (Sigma) at a concentration of 0.0125 U/ μ L (Rodríguez-Hernandez, 2013). Free nucleotides were separated by thin layer chromatography across 10 cm PEI cellulose sheets (Sigma) with a solvent containing 1 M ammonium acetate and 5% acetic acid. Spots corresponding to AMP and aminoacyl-AMP were visualized by phosphorimaging and quantified with the software package ImageQuant 5.2. These data were used to determine single-turnover



Supplemental Figure 7. Representative primary data for aminoacylation. Left: TLC plate showing reactions of the AcdRS2b(A9) enzyme with saturating levels of orthog-tRNA and Acd at varying times. There is almost complete aminoacylation by the end of the timecourse (about 5 min). Right: Timecourse for the AcdRS2b(A9) reaction fit to derive k_{obs} (see the text).

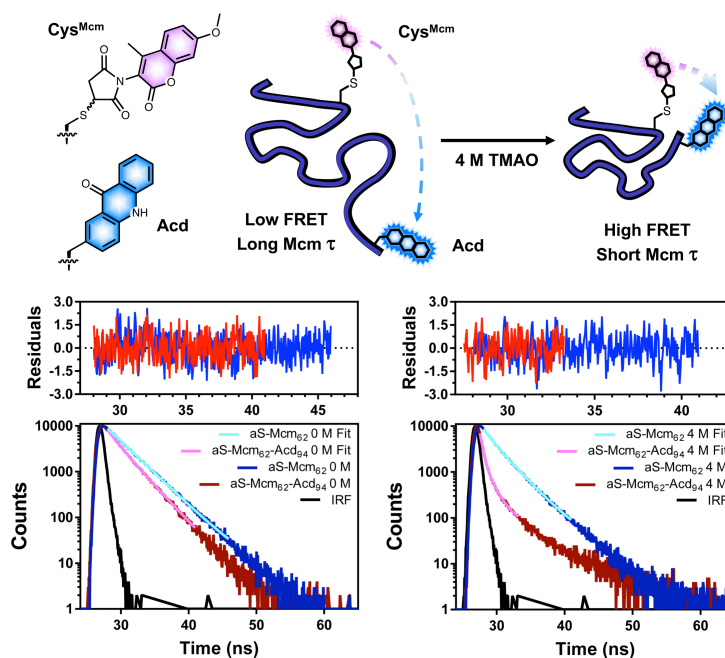
rate constants for a given AcdRS and amino concentration as shown in Supplemental Figure 7. To extract k_{obs} , the timecourses were fit to the single exponential equation $Y = A_1 \cdot e^{-k_{\text{obs}} \cdot t}$. To derive K_d , k_{obs} was determined at several amino acid concentrations and the plot of k_{obs} versus initial amino acid concentration (S_0) was fit to the hyperbolic binding curve $Y = S_0 \cdot k_{\text{chem}} / S_0 + K_d$. Prism 5 was used for all data fitting.

Computational Modeling

Geometry Optimization of Ligands. Structures of Acd and Npf were constructed in ChemDraw and saved in Molfile format. Acd and Npf Molfiles were submitted to the PRODRG webserver for energy minimization (Schüttelkopf, 2004). Coordinates of Npf were further optimized using density functional theory (DFT) with the B3LYP hybrid functional and the 6-311+G** basis set in Gaussian09. The UltraFine integration grid was used with tight convergence criteria.

Molecular Mechanics Parameterization of Ligands. Force field parameters and partial charges for Acd and Npf were obtained from the CGenFF webserver (Vanommeslaeghe, 2010, 2012, 2012). Bond orders were imposed based upon the connectivity. The partial charges of the main chain and beta carbon atoms were changed to those of the free tyrosine amino acid in CHARMM22 (MacKerell, 1998). An adjustment of ± 0.002 was made to the partial charge of the gamma carbon to make the molecule entirely neutral.

Ligand Modeling in AcdRS1 (G2). Atomic coordinates for G2 were obtained from the PDB structure 4PBR. Amino acid sidechain coordinates for sites with partial occupancy were taken from the A conformations.

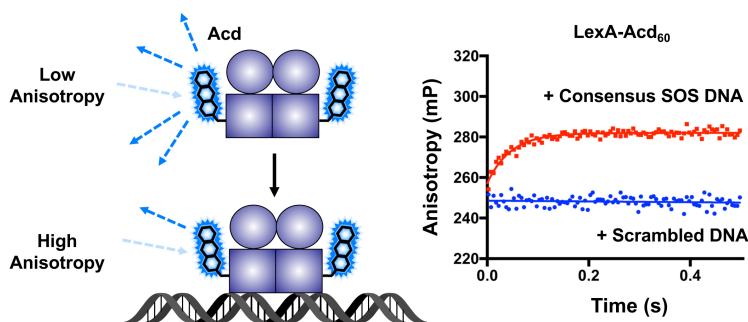


Supplemental Figure 8. Monitoring a Conformational Change in α S. A Cys₆₂Acd₉₄ α S mutant was expressed and labeled with Mcm-maleimide to give α S- Cys^{Mcm}₆₂Acd₉₄. Mcm/Acd FRET was measured in buffer (0 M) and in a solution of 4 M trimethylamine *N*-oxide (TMAO), which causes the protein to compact. The Mcm fluorescence lifetime (τ) was measured with excitation at 340 nm and emission at 380 nm using a PTI Quantamaster 40 fluorometer with time correlated single photon counting detection. A decrease in the the Mcm lifetime indicates increased FRET with Acd.

Alignments by superimposition of the first benzyl rings of Acd and Npf with the Brb benzyl ring were carried out in PyMol (Schrödinger), and atomic coordinates for the equivalent amino acid atoms of the ligands were adjusted to those of Brb. PSF files and PDB structure files with hydrogen atoms were generated using the AutomaticPSF builder tool in VMD (Humphrey, 1996). 10,000 steps of energy minimization of ligand atoms and all hydrogen atoms were carried out using the method of conjugated gradients in NAMD (Phillips, 2005). The CHARMM22 force field was used for all protein atoms, and the CGenFF force field for all interactions involving the ligands.

Generation of the AcdRS2b (A9) Structure. A statistical, computationally assisted design strategy (Kono, 2001) was used to determine atomic coordinates for the sidechains of mutated residues. Atomic coordinates of the protein backbone, wild-type residues and the Acd ligand, along with hydrogen atoms, were obtained from energy-minimized structure of Acd within G2. Allowed conformations of the mutated residues were those in a rotamer library (Shapovalov, 2011). Sidechain bond angles and bond lengths used were the average values collated from the HiQ54 dataset of crystal structures (Leaver-Fay, 2013). Sidechain modeling conformations were performed with an effective inverse temperature of $\beta = 0.5$ mol/kcal. This procedure yielded a model of A9 having the most probable sidechain conformation of each mutated amino acid.

Ligand Modeling in AcdRS2b (A9). Atomic coordinates for Acd and Npf were obtained from the minimization calculations in G2. PSF files and PDB structure files with hydrogen atoms were generated using the AutomaticPSF builder tool in VMD. 10,000 steps of energy minimization of ligand atoms, mutated amino acid sidechains, and all hydrogen atoms were carried out using the conjugate gradient method in NAMD. The CHARMM22 force field was used for all protein atoms, and the CGenFF force field for all interactions involving the ligands.



Supplemental Figure 9. Measuring the Kinetics of LexA-DNA Binding. LexA-Acd₆₀ was rapidly mixed with double-stranded 44-mer DNA substrates that contained either the “consensus” LexA binding site or a “scrambled” LexA binding site using a KinTek Auto SF-120 stopped-flow instrument to monitor Acd fluorescence anisotropy. An increase in anisotropy was observed upon specific binding of LexA-Acd₆₀ to consensus operator DNA (but not scrambled DNA). Polarized light at 386 nm was used to excite the mixed sample, and emitted signals in parallel and perpendicular polarized directions were measured at 440 nm. Anisotropy values were calculated automatically using the KinTek Auto SF-120 instrument software. Five replicates were collected, timepoints were averaged, and the data were fit to single exponential (for the consensus DNA) or straight line (for the scrambled DNA) models.

Supporting Information for Chapter 3

The contents below represent a condensed version of the Supporting Information for Chapter 3, for sake of brevity. The complete Supporting Information is available free of charge at DOI: 10.1021/acscchembio.8b00696.

Experimental Methods

Amber stop codon mutagenesis in LexA and RecA overexpression plasmids. Previously-described pET41 overexpression plasmids encoding either catalytically-inactive LexA with a C-terminal HIS tag (Sungwienwong, 2017) or wildtype RecA with an N-terminal HIS tag (Mo, 2018) were used as the template sequences for site-directed mutagenesis with Phusion polymerase (NEB) and pairs of synthetic oligonucleotides (IDT) designed to incorporate the 5'-TAG-3' amber stop codon. Successful mutagenesis was confirmed by sequencing (GeneWiz).

Parallel overexpression of LexA or RecA mutants. Overexpression plasmids were transformed into BL21(DE3) cells harboring the pDule2-Acd plasmid, which encodes a tRNA/tRNA synthetase evolved for specific incorporation of Acd (Speight, 2013; Sungwienwong, 2017), and grown on MDAG-11 non-inducing plates (Studier, 2014) with 50 µg/mL spectinomycin and 120 µg/mL kanamycin. For each replicate, an individual colony was seeded into 1 mL of MDAG-135 non-inducing broth (Studier, 2014) with selective antibiotics and grown at 30°C. Cell densities of overnight cultures were adjusted so that each 1:1000 inoculation of 1 mL of MDA-5052 autoinduction media (Studier, 2014) with selective antibiotics transferred an equivalent amount of cells. To autoinduction media, solubilized Acd was added to a final concentration of 0.5 mM. After 24 hours of growth at 30°C, cells were harvested and stored at -20°C.

Cell lysis and soluble protein fractionation. LexA lysis buffer contained 20 mM sodium phosphate pH 6.9, 500 mM NaCl, 0.25 mg/mL lysozyme (Sigma), 25 U/mL benzonase (Sigma), and 1x BugBuster protein extraction reagent (EMD Millipore). Cell pellets from the LexA experiment were lysed by resuspending in 15 µL of LexA lysis buffer per milligram of cell pellet to normalize the measurements and incubating at room temperature for 30 minutes. Cell pellets containing RecA were lysed following established protocol, again normalizing the amount of lysis buffer against cell pellet weight (Shibata, 1983). The soluble fractions of total cell lysates for LexA or RecA were obtained by centrifuging samples for 15 min at 13,000 rpm in a microcentrifuge at 4°C.

Determination of properties from sequence and structure files. The DNA sequence from the LexA overexpression plasmid was used to determine the effect of 3' nucleotides on nonsense codon suppression efficiencies. Primary amino acid sequences for LexA and RecA were used to calculate the following position-based metrics: Blosum62 substitution matrix similarity scores for Trp, Tyr, or Phe (Henikoff, 1992), residue volumes and surface areas (Heinig, 2004; Tien, 2013; Zamyatnin, 1972), residue hydrophobicity scores (Monera, 1995; Wimley, 1996; Wolfenden, 2007), and evolutionary tolerances to Trp, Tyr, or Phe (Sim, 2012). LexA and RecA PDB codes (1JHH or 2REB, respectively) were used as inputs for either the ConSurf database for conservation scores (Celniker, 2013; Goldenberg, 2009) or the STRIDE database for secondary structure classifications (Heinig, 2004). Remaining position-based metrics for LexA

(PDB code 1JHH) were retrieved from the PDBparam server (Nagarajan, 2016). We note that the PDBparam server was intermittently unavailable, and we were unable to retrieve the same set of PDBparam properties for RecA for this analysis.

Amino acid properties were examined using R (R core team, 2017; Wickham, 2016). Numerical parameters assigned to the chosen LexA residues whose distributions were approximately uniformly or normally distributed were maintained as continuous factors (solvent accessible area, average hydrophobic gain/ratio, C α or C β within 8 or 14 Å, conservation, fractional loss of accessible area, hydrophobicity, surrounding hydrophobicity, surrounding residues, and residue volume), whereas remaining numerical parameters with obvious skew were simplified to categorical factors. The degree to which each property was sampled by the chosen positions in LexA was assessed by plotting individual histograms or bar charts (Supplemental Figure 10 and Supplemental Figure 11). A more rigorous assessment of the variability of the chosen positions was accomplished through a principal component analysis. From the above continuous factors, highly-correlated parameters were dropped; the remaining continuous factors (solvent accessible area, average hydrophobic ratio, alpha carbons within 14 Å, conservation, hydrophobicity, surrounding hydrophobic residues, surrounding residues, and residue volume) were used to generate principal components using the base “pca()” function in R.

Specific detection of Acd fluorescence. To specifically detect Acd-labeled LexA or RecA, total cell lysate and soluble fraction samples were mixed with equivalent volumes of 2x Laemmli buffer and 8 μ L were run on 15% SDS-PAGE gels. On each gel, three dilutions of previously-purified Acd-labeled LexA were also run as standards (Sungwienwong, 2017). Acd fluorescence was visualized by illuminating the gels in the dark with an Entela UL3101-1 handheld UV lamp and exposing with a Sony ILCE-6000 camera with E 35 mm F1.8 OSS lens outfitted with a 440 nm fluorescence bandpass filter (Edmund Optics). Red and green channels were removed from raw images, and fluorescence intensities were quantified using ImageJ (Schneider, 2012). A standard curve for each set of purified LexA standards was used to transform raw fluorescence readings to protein concentrations. To facilitate comparison between total and soluble measurements, fluorescent protein concentrations were logarithmically-transformed, i.e. $y = \log_{10}(x/x_0)$, where y is the transformed value, x is the measured value, and x_0 is equal to 1 unit of fluorescent protein (in nM). To compare differences in protein solubilities between samples, a ratio of the measured soluble fluorescent protein was divided by the measured total fluorescent protein.

Simulation of Acd incorporation into LexA or RecA with Rosetta. Prior to performing simulations, a parameter file and rotamer library were produced for Acd following a previously described method (Drew, 2013). Starting structures for the LexA simulations were prepared from PDB 1JHE and PDB 1JHF by adding the missing residues using the remodel application in Rosetta (Huang, 2011). A blueprint file was prepared from each monomer and the primary sequence was modified to match that of the LexA expression construct. After adding the missing residues to each monomer, the dimer was reconstructed by merging the two PDB files and the resultant structure was minimized using the Relax application. The Relax application was run by setting the jump_move, bb_move, and chi_move flags to False and using the relax:fast flag. The starting structure was selected as the lowest energy structure of 10 outputs. The same protocol was followed to produce the RecA starting structures from PDB 3CMW, omitting the remodel application step as all residues

were present. For the Backrub-based method, a total 2,500 structures were produced from each starting structure. This was done by running the Backrub application in Rosetta performing 10,000 trials at 0.6 kT to generate each output structure. The total energy was computed for each member of the ensemble following the single-site mutation to Acd and global repacking in PyRosetta. For RecA, all mutations were performed and assessed within a single monomeric unit (residues 967-1299) within the multimer. The total energy was averaged across all members of the single ensemble for RecA and across all members of both ensembles for LexA. LexA simulations based on the relax-based algorithm were performed in PyRosetta using the same initial structures as starting points. The method consisted exclusively of the FastRelax mover constrained to the starting coordinates using the 'lbfgs_armiho_nonmonotone' min_type and a maximum of 200 iterations. A total of five output were produced for each mutation and the energy was averaged across all outputs for both starting structures for a given site. All methods were run using the 'beta_nov15' score function weights.

Exploring amino acid properties and levels of Acd-labeled proteins. The calculated soluble fractions for LexA or RecA were fit to individual linear regression models for each categorical or numerical factor using the base "lm()" function in R. Data fitted to the models were evaluated using the base "summary()" function, which provide summary statistics for the fits. Models with single explanatory factors were as follows:

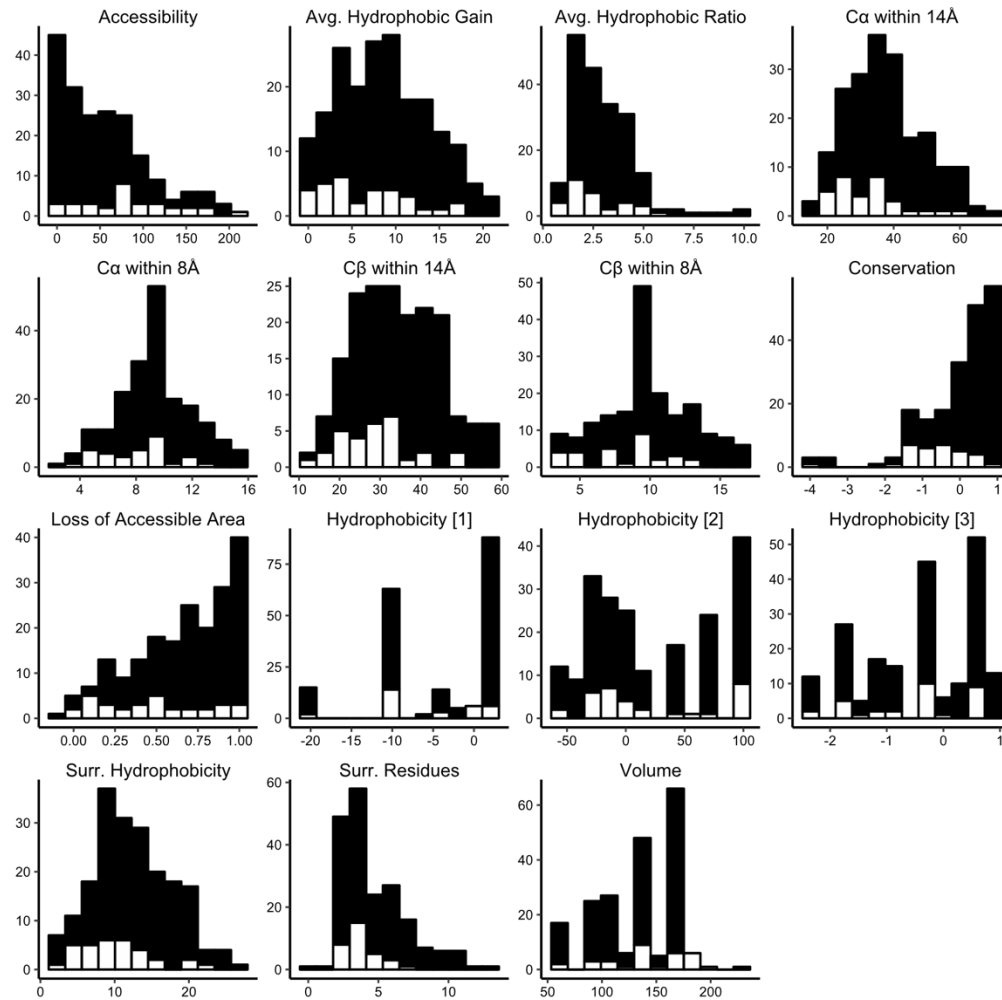
$$y_i = \alpha + \beta_a x_i + \varepsilon_i$$

where, y is the fraction of soluble protein, β is the coefficient for a given property "a", α is the intercept, ε is the error term, and i represents each individual observation. Summary statistics describing the quality of each fit, including adjusted R², are provided in Supplemental Table 4 and Supplemental Table 5. Models with protein domain and an individual property as two explanatory factors were modified from the above single-factor model, now explicitly including the term $\beta_{\text{domain}x_i}$ for the protein domain factor:

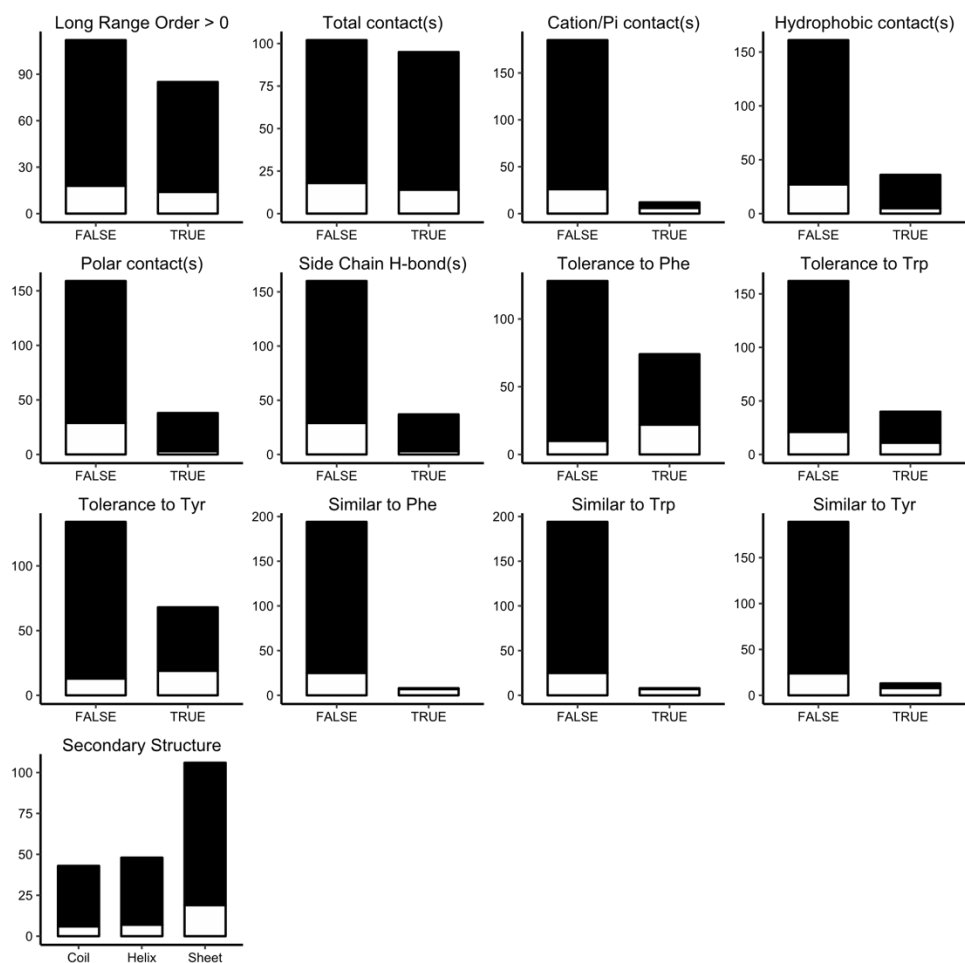
$$y_i = \alpha + \beta_{\text{domain}x_i} + \beta_a x_i + \varepsilon_i$$

For the two-factor models, the coefficient estimate and standard error for each $\beta_{\text{domain}x_i}$ term were reported in Supplemental Table 6 and Supplemental Table 7. In cases where there were too few observations for a given domain and individual property, the model was excluded from analysis. Between-group comparisons for the "domain" and "secondary structure" factors were performed with Tukey's HSD test using the base "TukeyHSD()" function in R.

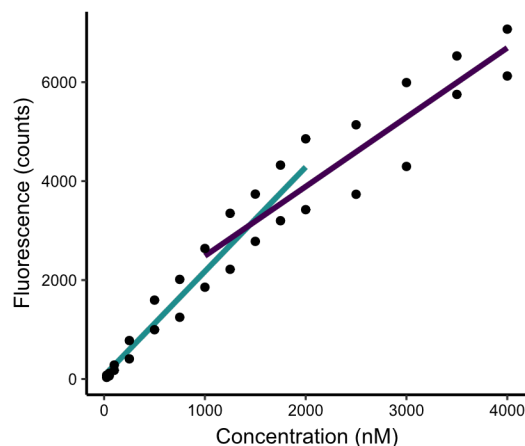
Supplemental Figures



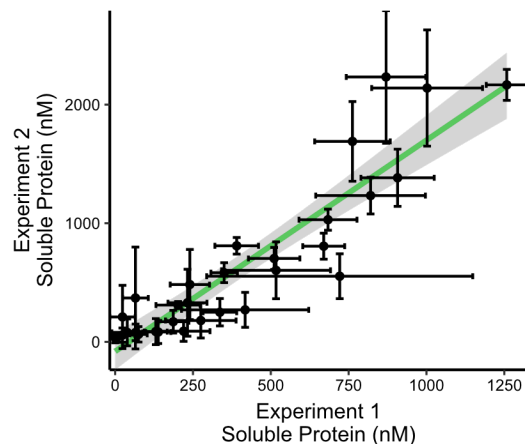
Supplemental Figure 10. Sampling of Numerical Properties by Chosen Positions in LexA. Histograms for each individual numerical structural, evolutionary, or physicochemical metric illustrate the frequency distribution of all positions in LexA. Positions that were advanced for unnatural amino acid mutagenesis are colored white, and the remaining positions in LexA are colored black.



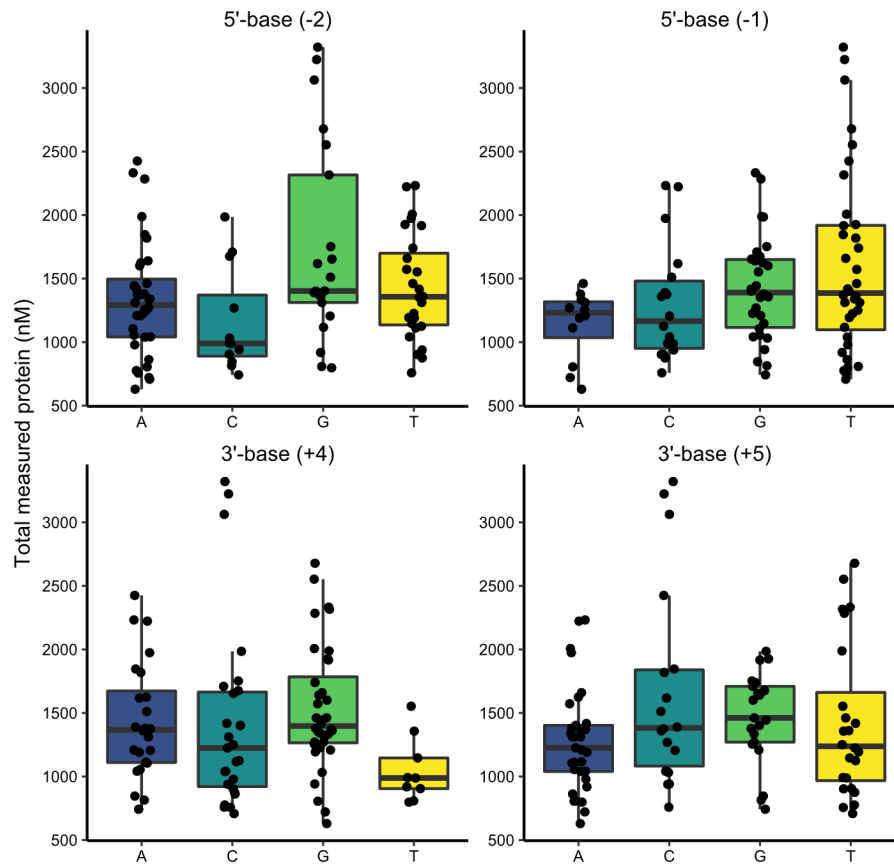
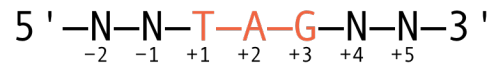
Supplemental Figure 11. Sampling of Categorical Properties by Chosen Positions in LexA. Bar graphs for each non-numeric structural, evolutionary, or physicochemical metric illustrate the categorization of all positions in LexA. Positions that were advanced for unnatural amino acid mutagenesis are colored white, and the remaining positions in LexA are colored black.



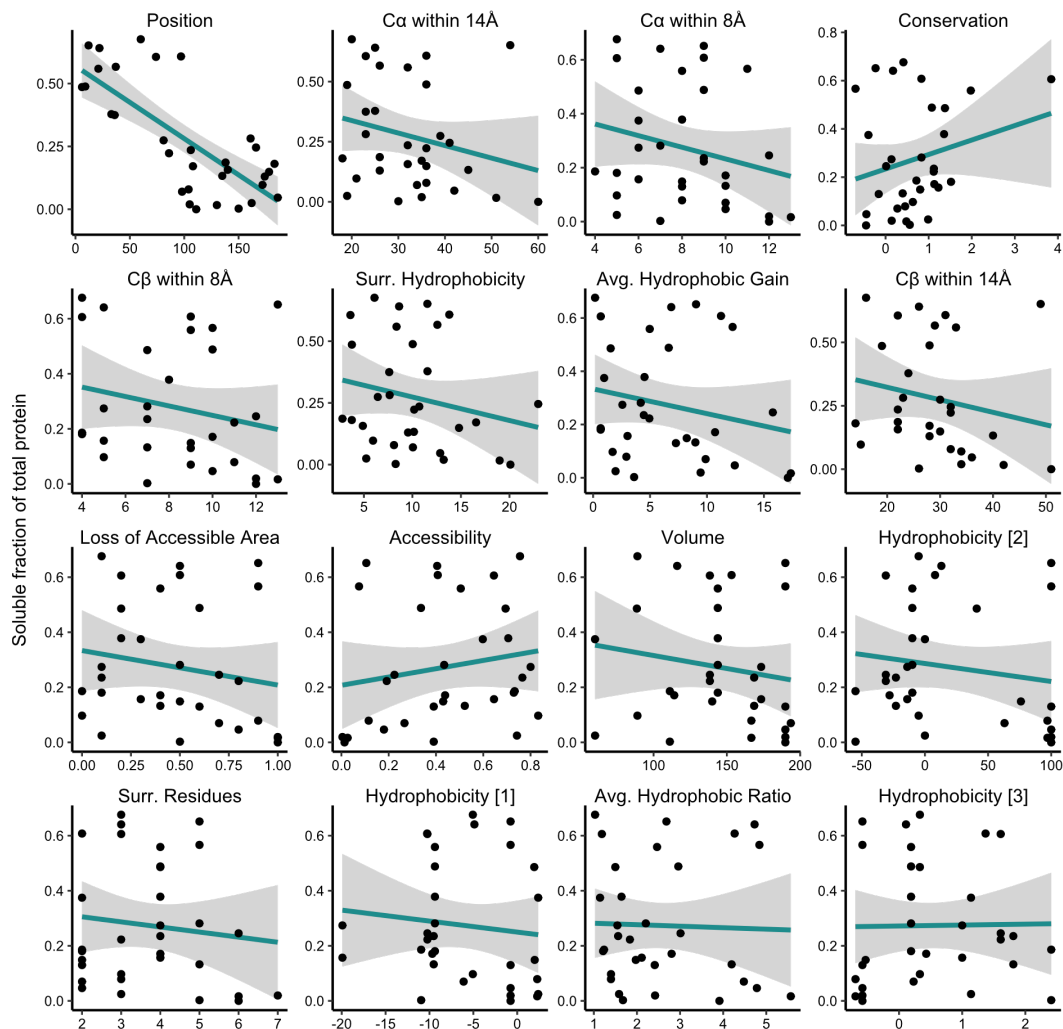
Supplemental Figure 12. Dynamic Range Determination from Purified LexA Standards. Dilutions of purified Acd-labeled LexA were run on 15% SDS-PAGE gels and Acd fluorescence was visualized and quantitated. The band intensities were plotted as a function of known concentration for each protein standard, revealing a nearly 100-fold dynamic range. Two separate linear fits show the concentrations from which purified LexA standards were used: standards from the turquoise curve (from 25 to 2000 nM) were used for quantifying LexA samples, whereas standards from the purple curve (from 1000 to 4000 nM) were used for quantifying RecA samples.



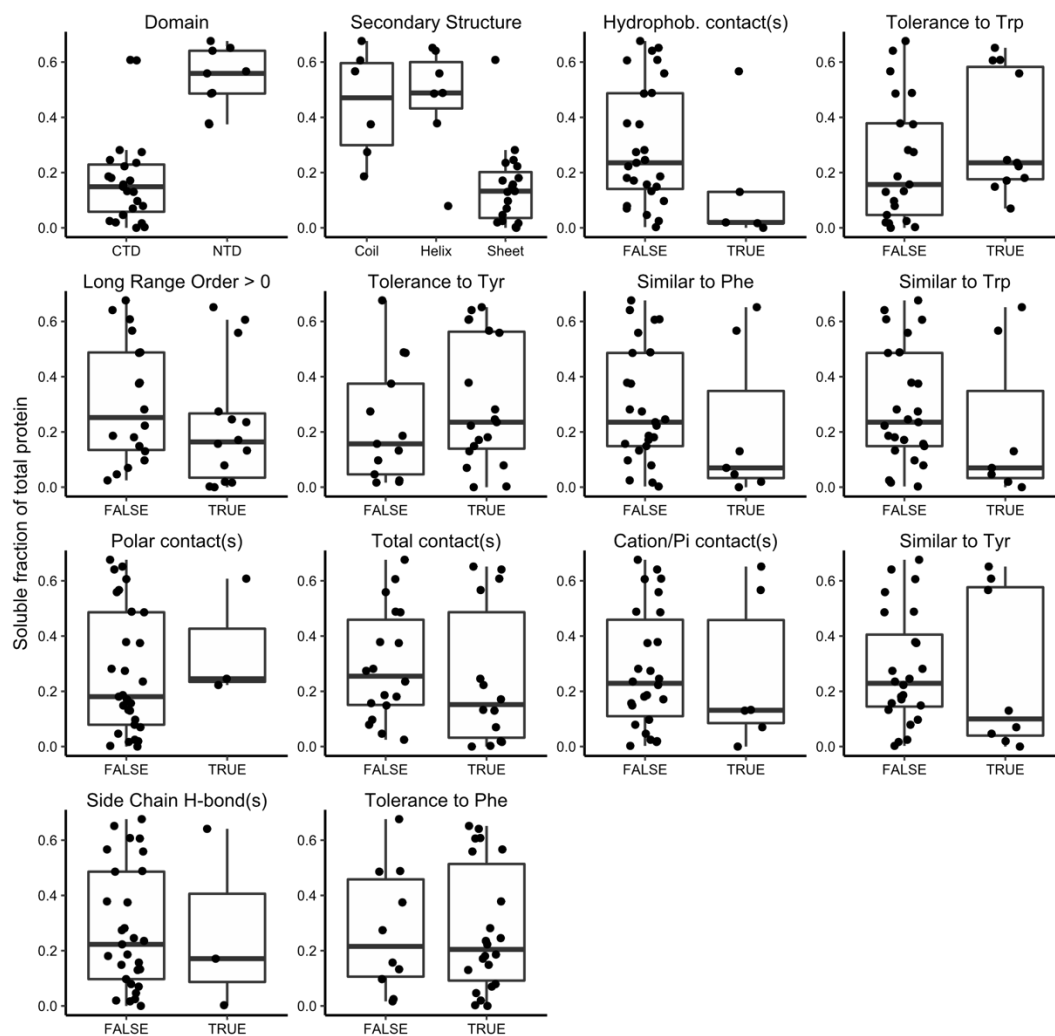
Supplemental Figure 13. Reproducibility of Experimental Approach. Plot of soluble protein measurements from two separate overexpression experiments in which Acd was incorporated into each of the 32 chosen positions in LexA. Each set of samples were overexpressed, processed, and measured on different days. Data points represent the average amount of soluble protein for each sample across the two separate experiments. Error bars represent the standard deviation of three replicates for each sample. A linear fit of the data (green line) shows good correlation (Pearson coefficient = 0.91) of the measured values, with a 95% confidence interval shown in gray.



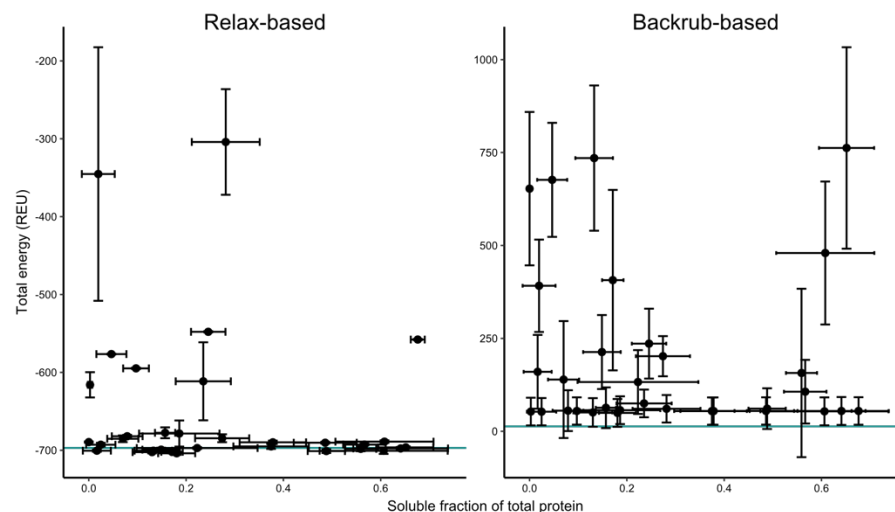
Supplemental Figure 14. Effect of Neighboring Nucleotides on Amber Suppression Efficiency. (Top) Schematic of the 5' and 3' nucleotide context surrounding the amber stop codon. (Bottom) Boxplots illustrating the relationship between total expressed protein and the surrounding nucleotide context either upstream, with the (-2) or (-1) 5'-base, or downstream, with the (+4) or (+5) 3'-base, of the amber stop codon in each mutant. Data points represent measurements of individual replicates of total expressed protein.



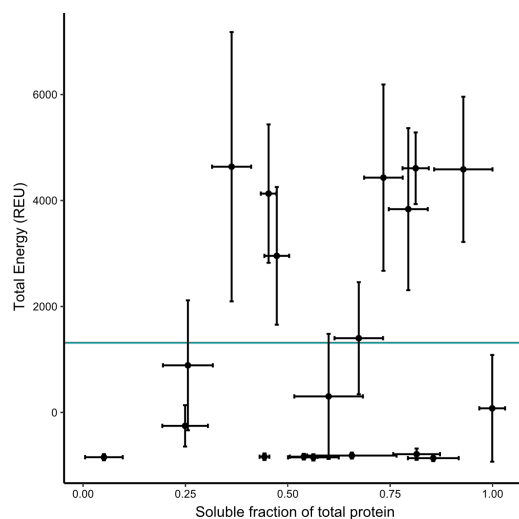
Supplemental Figure 15. Effect of Individual Numerical Properties on LexA Solubility. Scatterplots illustrating the relationships between the soluble fraction of total protein as a function of each of the numerical structural, evolutionary, or physicochemical properties. Data points represent the average soluble fraction of total protein for each sample in LexA. Linear fits of the data (turquoise) with 95% confidence intervals (gray) for each property are shown.



Supplemental Figure 16. Effect of Individual Categorical Properties on LexA Solubility. Boxplots illustrate the relationships between the fraction of soluble protein produced across each of the categorical structural, evolutionary, or physicochemical properties. Data points represent the average soluble fraction of total protein for each sample in LexA.



Supplemental Figure 17. Predicting Protein Solubility through Simulation of Acd Incorporation in LexA. Scatterplots of the total energies in Rosetta Energy Units (REU) from simulating Acd incorporation in LexA as a function of the soluble fraction of total protein. Rosetta energies were obtained by performing each single mutation on a relaxed structure of LexA derived from one of two previously published structures (PDB: 1JHE or 1JHF), using either a Relax-based (left) or Backrub-based (right) method. The total energy of each LexA mutant was computed following mutation of the residue of interest to Acd either by minimizing of the energy using a relax-based protocol or following repacking of all residues for each member of an ensemble of LexA structures. Each point represents the average of the two different simulations, with vertical error bars representing standard deviations. The solid turquoise line represents the average energy of energy-minimized LexA without any Acd mutation.



Supplemental Figure 18. Predicting Protein Solubility through Simulation of Acd Incorporation in RecA. Scatterplot of the total energies in Rosetta Energy Units (REU) from simulating Acd incorporation in RecA as a function of the soluble fraction of total protein. Rosetta energies were obtained by performing each single mutation on each member of a 2,500 structure RecA ensemble generated using the Backrub application. Separate ensembles were generated from the previously published structure (PDB: 3CMW). The total energy of each RecA mutant was computed after mutating the residue of interest to Acd and repacking all residues in RecA. Each point represents the average energy computed across all members of the different simulations, with vertical error bars representing standard deviations. The solid turquoise line represents the average energy of energy-minimized RecA without any Acd mutation.

Supplemental Tables

Supplemental Table 3: Measured Total and Soluble Amounts of Fluorescent LexA

	Total fluorescent protein (nM)		Soluble fluorescent protein (nM)		Soluble fraction of total protein	
Sample	Average	SD	Average	SD	Average	SD
A6	8.0×10^2	5.3×10^1	3.9×10^2	7.0×10^1	0.49	0.07
Q9	7.2×10^2	8.8×10^1	3.5×10^2	4.2×10^1	0.49	0.04
F12	1.3×10^3	9.7×10^1	8.7×10^2	1.3×10^2	0.65	0.06
Q21	1.4×10^3	1.5×10^2	7.6×10^2	1.2×10^2	0.56	0.03
T22	1.6×10^3	1.0×10^2	1.0×10^3	1.8×10^2	0.64	0.10
Q33	1.8×10^3	1.7×10^2	6.7×10^2	6.7×10^1	0.38	0.07
G36	1.4×10^3	2.0×10^2	5.2×10^2	1.7×10^2	0.37	0.08
F37	1.6×10^3	1.8×10^2	9.1×10^2	1.2×10^2	0.57	0.04
S60	1.9×10^3	1.0×10^2	1.3×10^3	6.6×10^1	0.68	0.01
E74	1.4×10^3	9.7×10^1	8.2×10^2	1.8×10^2	0.61	0.13
R81	2.5×10^3	1.8×10^2	6.8×10^2	9.3×10^1	0.27	0.06
E86	3.2×10^3	1.3×10^2	7.2×10^2	4.3×10^2	0.22	0.12
H97	8.4×10^2	6.7×10^1	5.1×10^2	8.3×10^1	0.61	0.10
Y98	9.6×10^2	9.0×10^1	6.6×10^1	2.8×10^1	0.07	0.03
L104	9.6×10^2	4.9×10^1	7.5×10^1	2.8×10^1	0.08	0.03
F105	1.2×10^3	1.0×10^2	2.3×10^1	4.0×10^1	0.02	0.03
K106	9.1×10^2	1.4×10^2	2.2×10^2	8.5×10^1	0.24	0.06
N108	1.1×10^3	1.7×10^2	1.9×10^2	5.0×10^1	0.17	0.02
F111	7.5×10^2	3.5×10^1	0.0×10^0	0.0×10^0	0.00	0.00
L130	1.3×10^3	1.1×10^2	2.4×10^1	4.1×10^1	0.02	0.03
K135	2.0×10^3	3.4×10^2	2.8×10^2	1.1×10^2	0.13	0.04
D138	2.2×10^3	1.9×10^2	4.2×10^2	2.0×10^2	0.19	0.08
R140	2.1×10^3	1.5×10^2	3.4×10^2	5.4×10^1	0.16	0.02
D150	1.7×10^3	2.3×10^2	5.1×10^0	8.9×10^0	0.00	0.00
Q161	1.3×10^3	8.8×10^1	3.6×10^2	1.2×10^2	0.28	0.07
G162	1.4×10^3	2.1×10^2	3.9×10^1	4.9×10^1	0.03	0.03
E166	9.7×10^2	1.4×10^2	2.4×10^2	6.3×10^1	0.25	0.04
S172	1.3×10^3	2.9×10^1	1.3×10^2	3.7×10^1	0.10	0.03
F174	1.1×10^3	3.4×10^1	1.4×10^2	4.1×10^1	0.13	0.04
V178	1.3×10^3	1.4×10^2	2.0×10^2	7.1×10^1	0.15	0.04
Q183	1.3×10^3	1.5×10^2	2.3×10^2	6.3×10^1	0.18	0.04
F186	1.5×10^3	1.4×10^2	6.5×10^1	4.1×10^1	0.05	0.03

Supplemental Table 4: Summary of Linear Regression Fits for Categorical Properties With LexA

Parameter	R ²	Adj R ² , ¹	F-statistic ²	DF	DF residuals	p-value ³
Domain	0.53	0.53	106.67	5	91	0.00
Secondary Structure	0.47	0.45	40.54	2	94	0.00
Hydrophobic contact(s)	0.06	0.05	5.91	3	93	0.02
Tolerance to Trp	0.04	0.03	3.97	2	94	0.05
Long Range Order > 0	0.04	0.03	3.55	2	94	0.06
Tolerance to Tyr	0.03	0.02	2.58	2	94	0.11
Similar to Trp	0.02	0.01	2.02	2	94	0.16
Similar to Phe	0.02	0.01	2.02	2	94	0.16
Polar contact(s)	0.01	0.00	1.41	2	94	0.24
Total contact(s)	0.01	0.00	0.93	2	94	0.34
Cation/Pi contact(s)	0.00	-0.01	0.10	2	94	0.75
Similar to Tyr	0.00	-0.01	0.09	2	94	0.76
Side chain H-bond(s)	0.00	-0.01	0.00	2	94	0.98
Tolerance to Phe	0.00	-0.01	0.00	2	94	0.98

¹ Adj R² = adjusted R², which is the R² value adjusted for the number of parameters in the model

² F-statistic = ratio of variance explained by model to the variance explained by residuals

³ Probability of F-statistic for an F-distribution with indicated degrees of freedom (DF)

Supplemental Table 5: Summary of Linear Regression Fits for Numerical Properties with LexA

Parameter	R ²	Adj R ² , ⁴	F-statistic ⁵	DF	DF residuals	p-value ⁶
Position	0.53	0.53	106.60	2	94	0.00
C_α within 14 Å	0.06	0.05	5.62	2	94	0.02
C_α within 8 Å	0.05	0.04	5.21	2	94	0.02
Conservation	0.05	0.04	5.16	2	94	0.03
C_β within 8 Å	0.05	0.03	4.20	2	88	0.04
Surrounding Hydrophobicity	0.04	0.03	4.10	2	94	0.05
Avg. Hydrophobic Gain	0.04	0.03	4.02	2	94	0.05
C_β within 14 Å	0.04	0.03	3.31	2	88	0.07
Fractional Loss of Accessible Area	0.03	0.02	2.98	2	94	0.09
Accessibility	0.03	0.02	2.87	2	94	0.09
Volume	0.03	0.02	2.47	2	94	0.12
Hydrophobicity [2]⁷	0.02	0.01	2.28	2	94	0.13
Surrounding Residues	0.01	0.00	1.26	2	94	0.26
Hydrophobicity [1]⁸	0.01	0.00	1.10	2	94	0.30
Avg. Hydrophobic Ratio	0.00	-0.01	0.09	2	94	0.76
Hydrophobicity [3]⁹	0.00	-0.01	0.02	2	94	0.89

⁴ Adj R² = adjusted R², which is the R² value adjusted for the number of parameters in the model

⁵ F-statistic = ratio of variance explained by model to the variance explained by residuals

⁶ Probability of F-statistic for an F-distribution with indicated degrees of freedom (DF)

⁷ Hydrophobicity index (Monera, 1995)

⁸ Hydrophobicity index (Wolfenden, 2007)

⁹ Hydrophobicity index (Wimley, 1996)

Supplemental Table 6: Two-Factor Linear Regression Fits for Categorical Properties with LexA

Parameter	Coefficient¹⁰	Std. Error	NTD samples¹¹	CTD samples¹²	p-value¹³
Tolerance to Trp	0.15	0.03	2	9	0.00
Polar contact(s)	0.22	0.05	0	3	0.00
Tolerance to Tyr	0.09	0.03	5	14	0.00
Hydrophobic contacts(s)	-0.12	0.04	1	4	0.01
Similar to Trp	-0.08	0.04	2	5	0.03
Similar to Phe	-0.08	0.04	2	5	0.03
Tolerance to Phe	0.07	0.03	5	17	0.06
Cation-π contact(s)	-0.04	0.04	2	4	0.30
Side chain H-bond(s)	-0.02	0.05	1	2	0.67
Long Range Order > 0	0.00	0.03	2	12	0.92
Total contact(s)	0.00	0.03	3	11	0.96
Similar to Tyr	0.00	0.04	2	6	0.98

¹⁰ Estimated coefficient for indicated parameter in two-factor linear regression model

¹¹ Number of samples in NTD for which the value of the indicated parameter is TRUE

¹² Number of samples in CTD for which the value of the indicated parameter is TRUE

¹³ Probability of rejecting null hypothesis using t-distribution (parameters not shown)

Supplemental Table 7: Two-Factor Linear Regression Fits for Numerical Properties with LexA

Parameter	Coefficient ¹⁴	Std. Error	p-value ¹⁵
Conservation	0.07	0.02	0.00
Hydrophobicity [1] ¹⁶	-0.01	0.00	0.00
Position	0.00	0.00	0.00
Accessibility	0.00	0.00	0.00
C β within 8 Å	-0.02	0.01	0.00
Hydrophobicity [3] ¹⁷	0.05	0.02	0.01
Hydrophobicity [2] ¹⁸	0.00	0.00	0.01
C α within 8 Å	-0.01	0.01	0.03
Fractional Loss of Accessible Area	-0.10	0.05	0.04
C β within 14 Å	0.00	0.00	0.05
Surrounding Residues	-0.02	0.01	0.06
Surrounding Hydrophobic Residues	0.00	0.00	0.16
Avg. Hydrophobic Gain	0.00	0.00	0.21
C α within 14 Å	0.00	0.00	0.23
Avg. Hydrophobic Ratio	-0.01	0.01	0.66
Volume	0.00	0.00	0.79

¹⁴ Estimated coefficient for indicated parameter in two-factor linear regression model

¹⁵ Probability of rejecting null hypothesis using t-distribution (parameters not shown)

¹⁶ Hydrophobicity index (Wolfenden, 2007)

¹⁷ Hydrophobicity index (Wimley, 1996)

¹⁸ Hydrophobicity index (Monera, 1995)

Supplemental Table 8: Measured Total and Soluble Amounts of Fluorescent RecA

Sample	Total fluorescent protein (nM)		Soluble fluorescent protein (nM)		Soluble fraction of total protein	
	Average	SD	Average	SD	Average	SD
E4	9.7 x 10 ³	1.2 x 10 ³	2.4 x 10 ³	2.7 x 10 ²	0.25	0.06
R33	7.4 x 10 ³	9.8 x 10 ²	7.4 x 10 ³	1.0 x 10 ³	1.00	0.03
Y65	6.3 x 10 ³	1.1 x 10 ³	3.7 x 10 ³	4.2 x 10 ²	0.60	0.08
R85	7.2 x 10 ³	1.7 x 10 ³	6.1 x 10 ³	1.6 x 10 ³	0.86	0.06
E86	6.6 x 10 ³	1.4 x 10 ³	5.4 x 10 ³	1.3 x 10 ³	0.81	0.06
I102	4.0 x 10 ³	6.4 x 10 ²	2.7 x 10 ³	2.0 x 10 ²	0.67	0.06
T121	7.0 x 10 ³	9.7 x 10 ²	5.5 x 10 ³	5.7 x 10 ²	0.79	0.05
Q124	7.4 x 10 ³	1.2 x 10 ³	6.0 x 10 ³	9.6 x 10 ²	0.81	0.03
R134	4.5 x 10 ³	6.7 x 10 ²	2.5 x 10 ³	2.3 x 10 ²	0.56	0.06
T150	5.8 x 10 ³	1.0 x 10 ³	2.1 x 10 ³	3.6 x 10 ²	0.36	0.05
E156	6.2 x 10 ³	1.6 x 10 ³	3.4 x 10 ³	1.0 x 10 ³	0.54	0.03
M197	5.8 x 10 ³	1.5 x 10 ³	5.4 x 10 ³	1.7 x 10 ³	0.93	0.07
P206	6.0 x 10 ³	5.8 x 10 ²	4.4 x 10 ³	6.9 x 10 ²	0.73	0.05
N213	5.2 x 10 ³	7.4 x 10 ²	2.5 x 10 ³	4.7 x 10 ²	0.47	0.03
E233	1.2 x 10 ³	2.2 x 10 ²	7.7 x 10 ²	1.5 x 10 ²	0.66	0.11
E266	4.1 x 10 ³	4.5 x 10 ²	2.2 x 10 ²	2.2 x 10 ²	0.05	0.05
L277	4.9 x 10 ³	6.4 x 10 ²	2.2 x 10 ³	2.7 x 10 ²	0.45	0.02
D311	3.7 x 10 ³	4.3 x 10 ²	1.6 x 10 ³	1.5 x 10 ²	0.44	0.01
K321	5.4 x 10 ³	9.5 x 10 ²	1.4 x 10 ³	2.2 x 10 ²	0.26	0.06

Supplemental Table 9: Summary Statistics of Linear Regression Models with RecA

Parameter	R ²	Adj R ² , ¹⁹	F-statistic ²⁰	DF	DF residuals	p-value ²¹
Domain	0.26	0.23	9.51	3	54	0.00
Position	0.19	0.17	12.80	2	55	0.00
Tolerance to Trp	0.17	0.15	11.00	2	55	0.00
Hydrophobicity [3]²²	0.12	0.11	7.77	2	55	0.01
Tolerance to Phe	0.11	0.09	6.79	2	55	0.01
Secondary Structure	0.13	0.09	3.51	3	48	0.04
Accessibility	0.09	0.07	5.02	2	49	0.03
Volume	0.04	0.03	2.47	2	55	0.12
Conservation	0.04	0.02	2.02	2	55	0.16
Hydrophobicity [2]²³	0.04	0.02	1.99	2	52	0.16
Hydrophobicity [1]²⁴	0.02	0.00	0.99	2	55	0.33
Tolerance to Tyr	0.00	-0.02	0.09	2	55	0.77
Similar to Trp	0.00	-0.02	0.00	2	55	0.96
Similar to Phe	0.00	-0.02	0.00	2	55	0.96
Similar to Tyr	0.00	-0.02	0.00	2	55	0.96

¹⁹ Adj R² = adjusted R², which is the R² value adjusted for the number of parameters in the model

²⁰ F-statistic = ratio of variance explained by model to the variance explained by residuals

²¹ Probability of F-statistic for an F-distribution with indicated degrees of freedom (DF)

²² Hydrophobicity index (Wimley, 1996)

²³ Hydrophobicity index (Monera, 1995)

²⁴ Hydrophobicity index (Wolfenden, 2007)

BIBLIOGRAPHY

- Adikesavan, A. K., Katsonis, P., Marciano, D. C., Lua, R., Herman, C., & Lichtarge, O. (2011). Separation of recombination and SOS response in *Escherichia coli* RecA suggests LexA interaction sites. *PLoS Genetics*, 7(9), e1002244.
- Albayrak, C., & Swartz, J. R. (2013). Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Research*, 41(11), 5949–5963.
- Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., ... Gray, J. J. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, 13(6), 3031–3048.
- Arpino, J. A. J., Baldwin, A. J., McGarrity, A. R., Tippmann, E. M., & Jones, D. D. (2015). In-frame amber stop codon replacement mutagenesis for the directed evolution of proteins containing non-canonical amino acids: identification of residues open to bio-orthogonal modification. *PloS One*, 10(5), e0127504.
- Bai, X. chen, McMullan, G., & Scheres, S. H. W. (2015). How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1), 49–57.
- Bain, J. D., Glabe, C. G., Dix, T. A., Chamberlin, A. R., & Diala, E. S. (1989). Biosynthetic Site-Specific Incorporation of a Non-Natural Amino Acid into a Polypeptide. *Journal of the American Chemical Society*, 111(20), 8013–8014.
- Bell, J. C., & Kowalczykowski, S. C. (2016). RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends in Biochemical Sciences*, 41(6), 491–507.
- Belmont, P., Bosson, J., Godet, T., & Tiano, M. (2007). Acridine and acridone derivatives, anticancer properties and synthetic methods: where are we now? *Anti-Cancer Agents in Medicinal Chemistry*, 7(2), 139–169.
- Bender, G. M., Lehmann, A., Zou, H., Cheng, H., Fry, H. C., Engel, D., ... DeGrado, W. F. (2007). De novo design of a single-chain diphenylporphyrin metalloprotein. *Journal of the American Chemical Society*, 129(35), 10732–10740.
- Berger, M. D., Lee, A. M., Simonette, R. A., Jackson, B. E., Roca, A. I., & Singleton, S. F. (2001). Design and evaluation of a tryptophanless RecA protein with wild type activity. *Biochemical and Biophysical Research Communications*, 286(5), 1195–1203.
- Bernier, S. P., Lebeaux, D., DeFrancesco, A. S., Valomon, A., Soubigou, G., Coppée, J.-Y., ... Beloin, C. (2013). Starvation, together with the SOS response, mediates high biofilm-specific tolerance to the fluoroquinolone ofloxacin. *PLoS Genetics*, 9(1), e1003144.
- Blizzard, R. J., Backus, D. R., Brown, W., Bazewicz, C. G., Li, Y., & Mehl, R. A. (2015). Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid. *Journal of the American Chemical Society*, 137(32), 10044–10047.
- Brabham, R., & Fascione, M. A. (2017). Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *ChemBioChem*, 18(20), 1973–1983.
- Brent, R., & Ptashne, M. (1981). Mechanism of action of the *lexA* gene product. *Proceedings of the National Academy of Sciences of the United States of America*, 78(7), 4204–4208.
- Brown, E. D., & Wright, G. D. (2016). Antibacterial drug discovery in the resistance era. *Nature*, 529(7586), 336–343.

- Brun, M.-P., Bischoff, L., & Garbay, C. (2004). A very short route to enantiomerically pure coumarin-bearing fluorescent amino acids. *Angewandte Chemie (International Ed. in English)*, 43(26), 3432–3436.
- Butala, M., Klose, D., Hodnik, V., Rems, A., Podlesek, Z., Klare, J. P., ... Zgur-Bertok, D. (2011). Interconversion between bound and free conformations of LexA orchestrates the bacterial SOS response. *Nucleic Acids Research*, 39(15), 6546–6557.
- Butala, M., Žgur-Bertok, D., & Busby, S. J. W. (2009). The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences*, 66(1), 82–93.
- Calhoun, J. R., Kono, H., Lahr, S., Wang, W., DeGrado, W. F., & Saven, J. G. (2003). Computational design and characterization of a monomeric helical dinuclear metalloprotein. *Journal of Molecular Biology*, 334(5), 1101–1115.
- Campbell-Valois, F.-X., Tarassov, K., & Michnick, S. W. (2005). Massive sequence perturbation of a small protein. *Proceedings of the National Academy of Sciences of the United States of America*, 102(42), 14988–14993.
- Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., ... Ben-Tal, N. (2013). ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal of Chemistry*, 53(3–4), 199–206.
- Cestari, I., & Stuart, K. (2013). A spectrophotometric assay for quantitative measurement of aminoacyl-tRNA synthetase activity. *Journal of Biomolecular Screening*, 18(4), 490–497.
- Charbon, G., Brustad, E., Scott, K. A., Wang, J., Løbner-Olesen, A., Schultz, P. G., ... Chapman, E. (2011). Subcellular protein localization by using a genetically encoded fluorescent amino acid. *Chembiochem: A European Journal of Chemical Biology*, 12(12), 1818–1821.
- Chatterjee, A., Guo, J., Lee, H. S., & Schultz, P. G. (2013). A genetically encoded fluorescent probe in mammalian cells. *Journal of the American Chemical Society*, 135(34), 12540–12543.
- Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., & Schultz, P. G. (2013). A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry*, 52(10), 1828–1837.
- Chen, S., Fahmi, N. E., Bhattacharya, C., Wang, L., Jin, Y., Benkovic, S. J., & Hecht, S. M. (2013). Fluorescent biphenyl derivatives of phenylalanine suitable for protein modification. *Biochemistry*, 52(47), 8580–8589.
- Chen, S., Fahmi, N. E., Wang, L., Bhattacharya, C., Benkovic, S. J., & Hecht, S. M. (2013). Detection of dihydrofolate reductase conformational change by FRET using two fluorescent amino acids. *Journal of the American Chemical Society*, 135(35), 12924–12927.
- Chen, Z., Yang, H., & Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, 453(7194), 489–494.
- Chin, D., & Means, A. R. (2000). Calmodulin: a prototypical calcium sensor. *Trends in Cell Biology*, 10(8), 322–328.
- Cirz, R. T., Chin, J. K., Andes, D. R., de Crécy-Lagard, V., Craig, W. A., & Romesberg, F. E. (2005). Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biology*, 3(6), e176.
- Cirz, R. T., & Romesberg, F. E. (2007). Controlling mutation: intervening in evolution as a therapeutic strategy. *Critical Reviews in Biochemistry and Molecular Biology*, 42(5), 341–354.

- Cooley, R. B., Karplus, P. A., & Mehl, R. A. (2014). Gleaning unexpected fruits from hard-won synthetases: probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *Chembiochem: A European Journal of Chemical Biology*, 15(12), 1810–1819.
- Courcelle, J., Khodursky, A., Peter, B., Brown, P. O., & Hanawalt, P. C. (2001). Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics*, 158(1), 41–64.
- Cox, M. M. (2007). Regulation of bacterial RecA protein function. *Critical Reviews in Biochemistry and Molecular Biology*, 42(1), 41–63.
- Culyba, M. J., Kubiak, J. M., Mo, C. Y., Goulian, M., & Kohli, R. M. (2018). Non-equilibrium repressor binding kinetics link DNA damage dose to transcriptional timing within the SOS gene network. *PLoS Genetics*, 14(6), e1007405.
- Culyba, M. J., Mo, C. Y., & Kohli, R. M. (2015). Targets for Combating the Evolution of Acquired Antibiotic Resistance. *Biochemistry*, 54(23), 3573–3582.
- Da Re, S., Garnier, F., Guérin, E., Campoy, S., Denis, F., & Ploy, M.-C. (2009). The SOS response promotes qnrB quinolone-resistance determinant expression. *EMBO Reports*, 10(8), 929–933.
- Díaz, E., Ferrández, A., Prieto, M. A., & García, J. L. (2001). Biodegradation of aromatic compounds by *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, 65(4), 523–569.
- Dixon, S. J., & Stockwell, B. R. (2014). The role of iron and reactive oxygen species in cell death. *Nature Chemical Biology*, 10(1), 9–17.
- Dörr, T., Lewis, K., & Vulić, M. (2009). SOS response induces persistence to fluoroquinolones in *Escherichia coli*. *PLoS Genetics*, 5(12), e1000760.
- Drew, K., Renfrew, P. D., Craven, T. W., Butterfoss, G. L., Chou, F.-C., Lyskov, S., ... Bonneau, R. (2013). Adding diverse noncanonical backbones to rosetta: enabling peptidomimetic design. *PloS One*, 8(7), e67051.
- Dumas, A., Lercher, L., Spicer, C. D., & Davis, B. G. (2015). Designing logical codon reassignment - Expanding the chemistry in biology. *Chemical Science*, 6(1), 50–69.
- Erill, I., Campoy, S., & Barbé, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiology Reviews*, 31(6), 637–656.
- Ferrie, J. J., Ieda, N., Haney, C. M., Walters, C. R., Sungwienwong, I., Yoon, J., & Petersson, E. J. (2017). Multicolor protein FRET with tryptophan, selective coumarin-cysteine labeling, and genetic acridonylalanine encoding. *Chemical Communications (Cambridge, England)*, 53(80), 11072–11075.
- Foster, P. L. (2007). Stress-induced mutagenesis in bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, 42(5), 373–397.
- Francklyn, C. S., First, E. A., Perona, J. J., & Hou, Y. M. (2008). Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods*, 44(2), 100–118.
- Friedman, N. D., Temkin, E., & Carmeli, Y. (2016). The negative impact of antibiotic resistance. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 22(5), 416–422.
- Friedman, N., Vardi, S., Ronen, M., Alon, U., & Stavans, J. (2005). Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria. *PLoS Biology*, 3(7), 1261–1268.

- Fry, H. C., Lehmann, A., Sinks, L. E., Asselberghs, I., Tronin, A., Krishnan, V., ... Therien, M. J. (2013). Computational de novo design and characterization of a protein that selectively binds a highly hyperpolarizable abiological chromophore. *Journal of the American Chemical Society*, 135(37), 13914–13926.
- Galkin, V. E., Britt, R. L., Bane, L. B., Yu, X., Cox, M. M., & Egelman, E. H. (2011). Two modes of binding of DinI to RecA filament provide a new insight into the regulation of SOS response by DinI protein. *Journal of Molecular Biology*, 408(5), 815–824.
- Giepmans, B. N. G., Adams, S. R., Ellisman, M. H., & Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *Science (New York, N.Y.)*, 312(5771), 217–224.
- Giese, K. C., Michalowski, C. B., & Little, J. W. (2008). RecA-dependent cleavage of LexA dimers. *Journal of Molecular Biology*, 377(1), 148–161.
- Goerke, A. R., & Swartz, J. R. (2009). High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnology and Bioengineering*, 102(2), 400–416.
- Goldberg, J. M., Batjargal, S., Chen, B. S., & Petersson, E. J. (2013). Thioamide quenching of fluorescent probes through photoinduced electron transfer: mechanistic studies and applications. *Journal of the American Chemical Society*, 135(49), 18651–18658.
- Goldberg, J. M., Speight, L. C., Fegley, M. W., & Petersson, E. J. (2012). Minimalist probes for studying protein dynamics: Thioamide quenching of selectively excitable fluorescent amino acids. *Journal of the American Chemical Society*, 134(14), 6088–6091.
- Goldenberg, O., Erez, E., Nimrod, G., & Ben-Tal, N. (2009). The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Research*, 37(Database issue), D323–D327.
- Griffin, B. A., Adams, S. R., & Tsien, R. Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. *Science (New York, N.Y.)*, 281(5374), 269–272.
- Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., ... Mazel, D. (2009). The SOS response controls integron recombination. *Science (New York, N.Y.)*, 324(5930), 1034.
- Hamada, H., Kameshima, N., Szymańska, A., Wegner, K., Lankiewicz, Ł., Shinohara, H., ... Sisido, M. (2005). Position-specific incorporation of a highly photodurable and blue-laser excitable fluorescent amino acid into proteins for fluorescence sensing. *Bioorganic & Medicinal Chemistry*, 13(10), 3379–3384.
- Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C., & Mehl, R. A. (2007). Preparation of site-specifically labeled fluorinated proteins for ¹⁹F-NMR structural characterization. *Nature Protocols*, 2(10), 2601–2607.
- Haney, C. M., Wissner, R. F., & Petersson, E. J. (2015). Multiply labeling proteins for studies of folding and stability. *Current Opinion in Chemical Biology*, 28, 123–130.
- Harkiss, A. H., & Sutherland, A. (2016). Recent advances in the synthesis and application of fluorescent α -amino acids. *Organic & Biomolecular Chemistry*, 14(38), 8911–8921.
- Heckler, T. G., Chang, L. H., Zama, Y., Naka, T., Chorghade, M. S., & Hecht, S. M. (1984). T4 RNA Ligase Mediated Preparation of Novel "Chemically Misacylated" tRNAs. *Biochemistry*, 23(7), 1468–1473.
- Heinig, M., & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Research*, 32(Web Server issue), W500–W502.

- Hendrickson, T. L., de Crécy-Lagard, V., & Schimmel, P. (2004). Incorporation of nonnatural amino acids into proteins. *Annual Review of Biochemistry*, 73(1), 147–176.
- Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22), 10915–10919.
- Hino, N., Hayashi, A., Sakamoto, K., & Yokoyama, S. (2006). Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded genetic code. *Nature Protocols*, 1(6), 2957–2962.
- Hocquet, D., Llanes, C., Thouverez, M., Kulasekara, H. D., Bertrand, X., Plésiat, P., ... Miller, S. I. (2012). Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS Pathogens*, 8(6), e1002778.
- Hostetler, Z. M., Ferrie, J. J., Bornstein, M. R., Sungwienwong, I., Petersson, E. J., & Kohli, R. M. (2018). Systematic Evaluation of Soluble Protein Expression Using a Fluorescent Unnatural Amino Acid Reveals No Reliable Predictors of Tolerability. *ACS Chemical Biology*, 13(10), 2855–2861.
- Hover, B. M., Kim, S.-H., Katz, M., Charlop-Powers, Z., Owen, J. G., Ternei, M. A., ... Brady, S. F. (2018). Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nature Microbiology*, 3(4), 415–422.
- Huang, P.-S., Ban, Y.-E. A., Richter, F., Andre, I., Vernon, R., Schief, W. R., & Baker, D. (2011). RosettaRemodel: a generalized framework for flexible backbone protein design. *PloS One*, 6(8), e24109.
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38.
- Ibba, M., & Soll, D. (2000). Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry*, 69, 617–650.
- Johnson, K. A., Simpson, Z. B., & Blom, T. (2009a). FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Analytical Biochemistry*, 387(1), 30–41.
- Johnson, K. A., Simpson, Z. B., & Blom, T. (2009b). Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Analytical Biochemistry*, 387(1), 20–29.
- Kalstrup, T., & Blunck, R. (2013). Dynamics of internal pore opening in K(V) channels probed by a fluorescent unnatural amino acid. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20), 8272–8277.
- Kao, C., Zheng, M., & Rüdisser, S. (1999). A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA (New York, N.Y.)*, 5(9), 1268–1272.
- Kellogg, E. H., Leaver-Fay, A., & Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins*, 79(3), 830–838.
- Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A., & Collins, J. J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. *Cell*, 130(5), 797–810.

- Kono, H., & Saven, J. G. (2001). Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *Journal of Molecular Biology*, 306(3), 607–628.
- Kovačič, L., Paulič, N., Leonardi, A., Hodnik, V., Anderluh, G., Podlesek, Z., ... Butala, M. (2013). Structural insight into LexA-RecA* interaction. *Nucleic Acids Research*, 41(21), 9901–9910.
- Kowalczykowski, S. C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*, 7(11), a016410.
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., ... Isaacs, F. J. (2013). Genomically recoded organisms expand biological functions. *Science (New York, N.Y.)*, 342(6156), 357–360.
- Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy. Principles of fluorescence spectroscopy*, Springer, New York, USA, 3rd edn, 2006.
- Lampkowski, J. S., Uthappa, D. M., & Young, D. D. (2015). Site-specific incorporation of a fluorescent terphenyl unnatural amino acid. *Bioorganic and Medicinal Chemistry Letters*, 25(22), 5277–5280.
- Lang, K., & Chin, J. W. (2014). Cellular incorporation of unnatural amino acids and bioorthogonal Labeling of Proteins. *Chemical Reviews*, 114(9), 4764–4806.
- Lashuel, H. A., Overk, C. R., Oueslati, A., & Masliah, E. (2013). The many faces of α -synuclein: from structure and toxicity to therapeutic target. *Nature Reviews. Neuroscience*, 14(1), 38–48.
- Lau, K. F., & Dill, K. A. (1990). Theory for protein mutability and biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 87(2), 638–642.
- Leaver-Fay, A., O'Meara, M. J., Tyka, M., Jacak, R., Song, Y., Kellogg, E. H., ... Kuhlman, B. (2013). Scientific benchmarks for guiding macromolecular energy function improvement. *Methods in Enzymology*, 523, 109–143.
- Lee, H. S., Guo, J., Lemke, E. A., Dimla, R. D., & Schultz, P. G. (2009). Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *Journal of the American Chemical Society*, 131(36), 12921–12923.
- Lim, W. A., & Sauer, R. T. (1989). Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature*, 339(6219), 31–36.
- Lin, L. L. (1988). *A genetic and biochemical analysis of LexA repressor cleavage in Escherichia coli K-12*. The University of Arizona.
- Lin, L. L., & Little, J. W. (1988). Isolation and characterization of noncleavable (Ind-) mutants of the LexA repressor of *Escherichia coli* K-12. *Journal of Bacteriology*, 170(5), 2163–2173.
- Lin, L. L., & Little, J. W. (1989). Autodigestion and RecA-dependent cleavage of Ind- mutant LexA proteins. *Journal of Molecular Biology*, 210(3), 439–452.
- Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., ... Lewis, K. (2015). A new antibiotic kills pathogens without detectable resistance. *Nature*, 517(7535), 455–459.
- Little, J. W. (1984). Autodigestion of lexA and phage lambda repressors. *Proceedings of the National Academy of Sciences of the United States of America*, 81(5), 1375–1379.
- Little, J. W., Kim, B., Roland, K. L., Smith, M. H., Lin, L. L., & Slilaty, S. N. (1994). Cleavage of LexA repressor. *Methods in Enzymology*, 244(C), 266–284.

- Lovett, C. M., O’Gara, T. M., & Woodruff, J. N. (1994). Analysis of the SOS inducing signal in *Bacillus subtilis* using *Escherichia coli* LexA as a probe. *Journal of Bacteriology*, 176(16), 4914–4923.
- Lu, T. K., & Collins, J. J. (2009). Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proceedings of the National Academy of Sciences of the United States of America*, 106(12), 4629–4634.
- Luo, J., Upreti, R., Naro, Y., Chou, C., Nguyen, D. P., Chin, J. W., & Deiters, A. (2014). Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *Journal of the American Chemical Society*, 136(44), 15551–15558.
- Luo, Y., Pfuetzner, R. A., Mosimann, S., Paetzel, M., Frey, E. A., Cherney, M., ... Strynadka, N. C. J. (2001). Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell*, 106(5), 585–594.
- Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J., & McAllister, W. T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. *Journal of Molecular Biology*, 280(2), 201–213.
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., ... Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry. B*, 102(18), 3586–3616.
- Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., & Miller, J. H. (1994). Genetic studies of the lac repressor. XIV. Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as “spacers” which do not require a specific sequence. *Journal of Molecular Biology*, 240(5), 421–433.
- McGann, R. G., & Deutscher, M. P. (1980). Purification and characterization of a mutant tRNA nucleotidyltransferase. *European Journal of Biochemistry*, 106(1), 321–328.
- McGhee, J. D., & von Hippel, P. H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *Journal of Molecular Biology*, 86(2), 469–489.
- McGrew, D. A., & Knight, K. L. (2003). Molecular design and functional organization of the RecA protein. *Critical Reviews in Biochemistry and Molecular Biology*, 38(5), 385–432.
- Miller, J. H., & Albertini, A. M. (1983). Effects of surrounding sequence on the suppression of nonsense codons. *Journal of Molecular Biology*, 164(1), 59–71.
- Mitchell, A. L., Addy, P. S., Chin, M. A., & Chatterjee, A. (2017). A Unique Genetically Encoded FRET Pair in Mammalian Cells. *ChemBioChem*, 18(6), 511–514.
- Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., & Brewer, S. H. (2009). Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry*, 48(25), 5953–5962.
- Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A., & Mehl, R. A. (2010). Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry*, 49(8), 1667–1677.
- Mo, C. Y., Birdwell, L. D., & Kohli, R. M. (2014). Specificity determinants for autoproteolysis of LexA, a key regulator of bacterial SOS mutagenesis. *Biochemistry*, 53(19), 3158–3168.

- Mo, C. Y., Culyba, M. J., Selwood, T., Kubiak, J. M., Hostetler, Z. M., Jurewicz, A. J., ... Kohli, R. M. (2018). Inhibitors of LexA Autoproteolysis and the Bacterial SOS Response Discovered by an Academic-Industry Partnership. *ACS Infectious Diseases*, 4(3), 349–359.
- Mo, C. Y., Manning, S. A., Roggiani, M., Culyba, M. J., Samuels, A. N., Sniegowski, P. D., ... Kohli, R. M. (2016). Systematically Altering Bacterial SOS Activity under Stress Reveals Therapeutic Strategies for Potentiating Antibiotics. *MSphere*, 1(4), e00163-16.
- Mohana-Borges, R., Pacheco, A. B. F., Sousa, F. J. R., Foguel, D., Almeida, D. F., & Silva, J. L. (2000). LexA repressor forms stable dimers in solution. The role of specific DNA in tightening protein-protein interactions. *Journal of Biological Chemistry*, 275(7), 4708–4712.
- Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M., & Hodges, R. S. (1995). Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *Journal of Peptide Science : An Official Publication of the European Peptide Society*, 1(5), 319–329.
- Mustard, J. A., & Little, J. W. (2000). Analysis of Escherichia coli RecA interactions with LexA, lambda CI, and UmuD by site-directed mutagenesis of recA. *Journal of Bacteriology*, 182(6), 1659–1670.
- Nagarajan, R., Archana, A., Thangakani, A. M., Jemimah, S., Velmurugan, D., & Gromiha, M. M. (2016). PDBparam: Online Resource for Computing Structural Parameters of Proteins. *Bioinformatics and Biology Insights*, 10, 73–80.
- Namsaraev, E. A., Baitin, D., Bakhlanova, I. V., Alexseyev, A. A., Ogawa, H., & Lanzov, V. A. (1998). Biochemical basis of hyper-recombinogenic activity of Pseudomonas aeruginosa RecA protein in Escherichia coli cells. *Mol Microbiol*, 27(4), 727–738.
- Neher, S. B., Flynn, J. M., Sauer, R. T., & Baker, T. A. (2003). Latent ClpX-recognition signals ensure LexA destruction after DNA damage. *Genes & Development*, 17(9), 1084–1089.
- Nehring, S., Budisa, N., & Wiltschi, B. (2012). Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code. *PloS One*, 7(4), e31992.
- Neumann-Staubitz, P., & Neumann, H. (2016). The use of unnatural amino acids to study and engineer protein function. *Current Opinion in Structural Biology*, 38, 119–128.
- Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., & Schultz, P. G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, 244(4901), 182–188.
- Peeler, J. C., Woodman, B. F., Averick, S., Miyake-Stoner, S. J., Stokes, A. L., Hess, K. R., ... Mehl, R. A. (2010). Genetically encoded initiator for polymer growth from proteins. *Journal of the American Chemical Society*, 132(39), 13575–13577.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781–1802.
- Pope, A. J., Haupts, U. M., & Moore, K. J. (1999). Homogeneous fluorescence readouts for miniaturized high-throughput screening: Theory and practice. *Drug Discovery Today*, 4(8), 350–362.
- Pott, M., Schmidt, M. J., & Summerer, D. (2014). Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS Chemical Biology*, 9(12), 2815–2822.

- Price, S. R., Ito, N., Oubridge, C., Avis, J. M., & Nagai, K. (1995). Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *Journal of Molecular Biology*, 249(2), 398–408.
- Quinones, M., Kimsey, H. H., & Waldor, M. K. (2005). LexA cleavage is required for CTX prophage induction. *Molecular Cell*, 17(2), 291–300.
- R core team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.
- Radman, M. (1975). SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sciences*, 5A, 355–367.
- Rauch, B. J., Porter, J. J., Mehl, R. A., & Perona, J. J. (2016). Improved Incorporation of Noncanonical Amino Acids by an Engineered tRNA(Tyr) Suppressor. *Biochemistry*, 55(3), 618–628.
- Reddington, S. C., Baldwin, A. J., Thompson, R., Brancale, A., Tippmann, E. M., & Jones, D. D. (2015). Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chemical Science*, 6(2), 1159–1166.
- Review on Antimicrobial Resistance. (2016). *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*.
- Richter, M. F., Drown, B. S., Riley, A. P., Garcia, A., Shirai, T., Svec, R. L., & Hergenrother, P. J. (2017). Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature*, 545(7654), 299–304.
- Ries, L. K., Schmid, F. X., & Schmidpeter, P. A. M. (2016). Incorporation of an Unnatural Amino Acid as a Domain-Specific Fluorescence Probe in a Two-Domain Protein. *Biochemistry*, 55(49), 6739–6742.
- Rodríguez-Hernández, A., & Perona, J. J. (2011). Heat maps for intramolecular communication in an RNP enzyme encoding glutamine. *Structure (London, England : 1993)*, 19(3), 386–396.
- Rodríguez-Hernandez, A., Spears, J. L., Gaston, K. W., Limbach, P. A., Gamper, H., Hou, Y.-M., ... Perona, J. J. (2013). Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position. *Journal of Molecular Biology*, 425(20), 3888–3906.
- Rodríguez-Rojas, A., Rodríguez-Beltrán, J., Couce, A., & Blázquez, J. (2013). Antibiotics and antibiotic resistance: a bitter fight against evolution. *International Journal of Medical Microbiology : IJMM*, 303(6–7), 293–297.
- Roland, K. L., Smith, M. H., Rupley, J. A., & Little, J. W. (1992). In vitro analysis of mutant LexA proteins with an increased rate of specific cleavage. *Journal of Molecular Biology*, 228(2), 395–408.
- Romero, P. A., Tran, T. M., & Abate, A. R. (2015). Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proceedings of the National Academy of Sciences of the United States of America*, 112(23), 7159–7164.
- Ronen, M., Rosenberg, R., Shraiman, B. I., & Alon, U. (2002). Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), 10555–10560.
- Sassanfar, M., & Roberts, J. W. (1990). Nature of the SOS-inducing signal in Escherichia coli. The involvement of DNA replication. *Journal of Molecular Biology*, 212(1), 79–96.

- Schlacher, K., & Goodman, M. F. (2007). Lessons from 50 years of SOS DNA-damage-induced mutagenesis. *Nature Reviews. Molecular Cell Biology*, 8(7), 587–594.
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675.
- Schultz, K. C., Supekova, L., Ryu, Y., Xie, J., Perera, R., & Schultz, P. G. (2006). A genetically encoded infrared probe. *Journal of the American Chemical Society*, 128(43), 13984–13985.
- Schüttelkopf, A. W., & van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica. Section D, Biological Crystallography*, 60(8), 1355–1363.
- Serrano, A. L., Troxler, T., Tucker, M. J., & Gai, F. (2010). Photophysics of a Fluorescent Non-natural Amino Acid: p-Cyanophenylalanine. *Chemical Physics Letters*, 487(4–6), 303–306.
- Shapovalov, M. V., & Dunbrack, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure (London, England : 1993)*, 19(6), 844–858.
- Shepley, D. P., & Little, J. W. (1996). Mutant LexA proteins with specific defects in autodigestion. *Proceedings of the National Academy of Sciences of the United States of America*, 93(21), 11528–11533.
- Sherlin, L. D., Bullock, T. L., Nissan, T. A., Perona, J. J., Lariviere, F. J., Uhlenbeck, O. C., & Scaringe, S. A. (2001). Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA (New York, N.Y.)*, 7(11), 1671–1678.
- Shibata, T., Osber, L., & Radding, C. M. (1983). Purification of recA protein from Escherichia coli. *Methods in Enzymology*, 100(1980), 197–209.
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(Web Server issue), W452–W457.
- Slilaty, S. N., & Little, J. W. (1987). Lysine-156 and serine-119 are required for LexA repressor cleavage: a possible mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 84(12), 3987–3991.
- Smith, C. A., & Kortemme, T. (2011). Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible backbone design. *PloS One*, 6(7), e20451.
- Smith, M. H., Cavenagh, M. M., & Little, J. W. (1991). Mutant LexA proteins with an increased rate of in vivo cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), 7356–7360.
- Smith, P. A., & Romesberg, F. E. (2007). Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nature Chemical Biology*, 3(9), 549–556.
- Speight, L. C., Muthusamy, A. K., Goldberg, J. M., Warner, J. B., Wissner, R. F., Willi, T. S., ... Petersson, E. J. (2013). Efficient synthesis and in vivo incorporation of acridon-2-ylalanine, a fluorescent amino acid for lifetime and Förster resonance energy transfer/luminescence resonance energy transfer studies. *Journal of the American Chemical Society*, 135(50), 18806–18814.
- Speight, L. C., Samanta, M., & Petersson, E. J. (2014). Minimalist Approaches to Protein Labelling: Getting the Most Fluorescent Bang for Your Steric Buck. *Australian Journal of Chemistry*, 67(5), 686.

- Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., & Goedert, M. (1997). Alpha-synuclein in Lewy bodies. *Nature*, 388(6645), 839–840.
- Stokes, A. L., Miyake-Stoner, S. J., Peeler, J. C., Nguyen, D. P., Hammer, R. P., & Mehl, R. A. (2009). Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Molecular BioSystems*, 5(9), 1032–1038.
- Studier, F. W. (2014). Stable expression clones and auto-induction for protein production in *E. coli*. *Methods in Molecular Biology (Clifton, N.J.)*, 1091, 17–32.
- Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J. W., & Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *Proceedings of the National Academy of Sciences of the United States of America*, 103(26), 9785–9789.
- Sungwienwong, I., Ferrie, J. J., Jun, J. V., Liu, C., Barrett, T. M., Hostetler, Z. M., ... Petersson, E. J. (2018). Improving the fluorescent probe acridonylalanine through a combination of theory and experiment. *Journal of Physical Organic Chemistry*, 31(8), e3813.
- Sungwienwong, I., Hostetler, Z. M., Blizzard, R. J., Porter, J. J., Driggers, C. M., Mbengi, L. Z., ... Petersson, E. J. (2017). Improving target amino acid selectivity in a permissive aminoacyl tRNA synthetase through counter-selection. *Organic & Biomolecular Chemistry*, 15(17), 3603–3610.
- Szymańska, A., Wegner, K., & Łankiewicz, L. (2003). Synthesis of N-[(tert-Butoxy)carbonyl]-3-(9,10-dihydro-9-oxoacridin-2-yl)-L-alanine, a New Fluorescent Amino Acid Derivative. *Helvetica Chimica Acta*, 86(10), 3326–3331.
- Takahashi, M., Daune, M., & Schnarr, M. (1986). Fluorescence study of the RecA-dependent proteolysis of LexA, the repressor of the SOS system in *Escherichia coli*. *FEBS Letters*, 196(2), 215–218.
- Taki, M., Yamazaki, Y., Suzuki, Y., & Sisido, M. (2010). Introduction of a Highly Photodurable and Common-laser Excitable Fluorescent Amino Acid into a Peptide as a FRET Acceptor for Protease Cleavage Detection. *Chemistry Letters*, 39(8), 818–819.
- Tang, J., Yin, H., Qiu, J., Tucker, M. J., DeGrado, W. F., & Gai, F. (2009). Using two fluorescent probes to dissect the binding, insertion, and dimerization kinetics of a model membrane peptide. *Journal of the American Chemical Society*, 131(11), 3816–3817.
- The Pew Charitable Trusts. (2016). A Scientific Roadmap for Antibiotic Discovery. *Report*, May(May), 1–47.
- Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J., & Wilke, C. O. (2013). Maximum allowed solvent accessibilities of residues in proteins. *PloS One*, 8(11), e80635.
- Tippin, B., Pham, P., & Goodman, M. F. (2004). Error-prone replication for better or worse. *Trends in Microbiology*, 12(6), 288–295.
- Truglio, J. J., Croteau, D. L., Van Houten, B., & Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical Reviews*, 106(2), 233–252.
- Uttamapinant, C., Sanchez, M. I., Liu, D. S., Yao, J. Z., & Ting, A. Y. (2013). Site-specific protein labeling using PRIME and chelation-assisted click chemistry. *Nature Protocols*, 8(8), 1620–1634.
- Van Den Bedem, H., & Fraser, J. S. (2015). Integrative, dynamic structural biology at atomic resolution - It's about time. *Nature Methods*, 12(4), 307–318.

- VanLoock, M. S., Yu, X., Yang, S., Galkin, V. E., Huang, H., Rajan, S. S., ... Egelman, E. H. (2003). Complexes of RecA with LexA and RecX differentiate between active and inactive RecA nucleoprotein filaments. *Journal of Molecular Biology*, 333(2), 345–354.
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., ... Mackerell, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4), 671–690.
- Vanommeslaeghe, K., & MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *Journal of Chemical Information and Modeling*, 52(12), 3144–3154.
- Vanommeslaeghe, K., Raman, E. P., & MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling*, 52(12), 3155–3168.
- Vetter, S. W., & Leclerc, E. (2003). Novel aspects of calmodulin target recognition and activation. *European Journal of Biochemistry*, 270(3), 404–414.
- Walsh, C. (2003). Where will new antibiotics come from? *Nature Reviews. Microbiology*, 1(1), 65–70.
- Wang, J., Xie, J., & Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *Journal of the American Chemical Society*, 128(27), 8738–8739.
- Wang, L., Brock, A., Herberich, B., & Schultz, P. G. (2001). Expanding the genetic code of *Escherichia coli*. *Science (New York, N.Y.)*, 292(5516), 498–500.
- Wang, L., & Schultz, P. G. (2004). Expanding the genetic code. *Angewandte Chemie (International Ed. in English)*, 44(1), 34–66.
- Wickham, H. (2016). *tidyverse: Easily Install and Load "Tidyverse" Packages*. R package version 1.0.0.
- Wimley, W. C., & White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*, 3(10), 842–848.
- Witkin, E. M. (1976). Ultraviolet mutagenesis and inducible DNA repair in *Escherichia coli*. *Bacteriological Reviews*, 40(4), 869–907.
- Wolfenden, R. (2007). Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. *The Journal of General Physiology*, 129(5), 357–362.
- Wolfson, A. D., & Uhlenbeck, O. C. (2002). Modulation of tRNA^{Ala} identity by inorganic pyrophosphatase. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), 5965–5970.
- World Economic Forum. (2018). *The Global Risks Report 2018, 13th Edition*.
- Xiao, H., & Schultz, P. G. (2016). At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harbor Perspectives in Biology*, 8(9), a023945.
- Xu, H., Wang, Y., Lu, J., Zhang, B., Zhang, Z., Si, L., ... Zhou, D. (2016). Re-exploration of the Codon Context Effect on Amber Codon-Guided Incorporation of Noncanonical Amino Acids in *Escherichia coli* by the Blue-White Screening Assay. *Chembiochem: A European Journal of Chemical Biology*, 17(13), 1250–1256.
- Young, T. S., & Schultz, P. G. (2010). Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry*, 285(15), 11039–11044.

- Yu, X., & Egelman, E. H. (1993). The LexA repressor binds within the deep helical groove of the activated RecA filament. *Journal of Molecular Biology*, 231(1), 29–40.
- Zamyatnin, A. A. (1972). Protein volume in solution. *Progress in Biophysics and Molecular Biology*, 24, 107–123.
- Zhang, A. P. P., Pigli, Y. Z., & Rice, P. A. (2010). Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature*, 466(7308), 883–886.
- Zhang, B., Xu, H., Chen, J., Zheng, Y., Wu, Y., Si, L., ... Zhou, D. (2015). Development of next generation of therapeutic IFN- α 2b via genetic code expansion. *Acta Biomaterialia*, 19, 100–111.
- Zheng, L., Baumann, U., & Reymond, J.-L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, 32(14), e115.
- Zheng, Y., Yu, F., Wu, Y., Si, L., Xu, H., Zhang, C., ... Zhou, D. (2015). Broadening the versatility of lentiviral vectors as a tool in nucleic acid research via genetic code expansion. *Nucleic Acids Research*, 43(11), e73.