

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 10-01-2018		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 13-May-2013 - 12-May-2017	
4. TITLE AND SUBTITLE Final Report: The De Novo Design of Protein-Protein Interfaces			5a. CONTRACT NUMBER W911NF-13-1-0155		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 206022		
6. AUTHORS			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES San Diego State University Foundation 5250 Campanile Drive San Diego, CA 92182 -1931			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 62695-LS-REP.5		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			John Love
					19b. TELEPHONE NUMBER 619-594-2063

RPPR Final Report

as of 28-Feb-2018

Agency Code:

Proposal Number: 62695LSREP

Agreement Number: W911NF-13-1-0155

INVESTIGATOR(S):

Name: PhD John Love

Email: jlove@mail.sdsu.edu

Phone Number: 6195942063

Principal: Y

Organization: **San Diego State University Foundation**

Address: 5250 Campanile Drive, San Diego, CA 921821931

Country: USA

DUNS Number: 073371346

EIN: 956042721

Report Date: 12-Aug-2017

Date Received: 10-Jan-2018

Final Report for Period Beginning 13-May-2013 and Ending 12-May-2017

Title: The De Novo Design of Protein-Protein Interfaces

Begin Performance Period: 13-May-2013

End Performance Period: 12-May-2017

Report Term: 0-Other

Submitted By: PhD John Love

Email: jlove@mail.sdsu.edu

Phone: (619) 594-2063

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 8

STEM Participants: 5

Major Goals: The primary goal of this research project was to explore the biophysical forces that drive biomolecular complex formation and further elucidate the parameters that give rise to specific protein/protein interactions. Thus far a large body of data has been collected on natural protein complexes and it was our intention to add to this body by engineering de novo (from scratch) protein/protein complexes. Using this 'inverse' approach we have furthered the understanding of biomolecular specificity by engineering protein complexes from previously monomeric 'test' proteins. The goal is to elucidate the key physical features needed to drive specific protein/protein interactions. It is considered 'inverse' because, instead of studying natural complexes, we are exploring key physical parameters by designing novel protein/protein interfaces. To achieve this we initially utilized robust computational algorithms but then ultimately succeeded by combining these with rational design methods based primarily on the X-ray crystal structures of early imperfect results. Using these methods we recently succeeded at driving small test proteins to form complexes of specific structure. After a number of tantalizing yet imperfect results we ultimately achieved success by designing metal binding sites at the interface of the target dimer complexes. These relatively small complexes were generated using the small test protein, protein-G (56 amino acids), as a design scaffold. The subunits of our designed complexes are perfectly monomeric in the absence of metal, yet form very clean dimers in the presence of zinc ions. We are in the process of designing and testing additional mutants with the goal of further exploring the important biophysical features that drive these novel biomolecular dimer complexes.

[There is a report below that describes the Goals in greater detail.]

Accomplishments: Thus far we have successfully accomplished our original goal of designing a number of highly specific protein/protein complexes. The overall goal of this project was to provide additional insights into important biophysical features of biomolecular interactions. Previously we successfully designed a number of protein/protein complexes, albeit of uncertain molecular complexity (i.e., it was uncertain as to whether the complexes were dimer, trimer, tetramer, etc.). Initially we attempted to use size exclusion chromatography (SEC) and multiangle light scattering (MALS) to better characterize the extent to which these poly-dispersive designed proteins formed complexes. This pursuit hit a number of technical roadblocks but we persevered and learned that the MALS instrument is highly sensitive to mixtures of complexes and thus cannot be used to determine the degree of multimer complexity for these poly-dispersive complexes. To design clean mono-dispersive complexes we started with the structure of a protein-G dimer reported in the literature. In addition we also used X-ray crystallography to solve the structures of two complexes from this first set of designs. The resulting structures were used directly as model guides in a second round of complex design, which recently resulted in the successful achievement of the original goal. We now have engineered at least four complex designs that are perfectly monomeric in the absence of metal and form precise dimers in the presence of metal (1 mM zinc). Therefore we have achieved our goal of

RPPR Final Report as of 28-Feb-2018

designing metal-mediated dimer formation and have used these successful designs to explore important biophysical contributions to protein complex formation. This work is complete and we are currently preparing a manuscript that will be submitted as an article to the Journal of the American Chemical Society.

In addition to the great success we had in generating clean metal-mediated complexes we also used our growing knowledge of protein/protein interactions to engineer chimeric enzymes that possesses enhanced ability to use ethanol as an alkyl source for the esterification of fatty acids. The parent enzyme was isolated from MRSA bacteria and rendered chimeric with the LUSH ethanol-binding domain from *Drosophila*. This work resulted in the publication of a manuscript in the Journal of Molecular Catalysis B: Enzymatic.

Another significant accomplishment entailed the use of unique secondary structure elements to enhance the production of proteins recalcitrant to expression in standard bacterial expression systems. To achieve this, β -hairpin structures were appended to the N- and C- termini of small test proteins and demonstrated to greatly enhance the extent of protein production in *E. coli*. Microscale Thermophoresis was then used to demonstrate that the presence of the β -hairpin structures reduced the binding affinity of the protein/protein interactions that necessarily takes place between elements of the degradation machinery (i.e., ClpS) and the test proteins that contained the β -hairpin structures. This work is approximately 95% complete and will be published following a small number of control experiments.

Finally to further enhance our desire to generate novel protein/protein complexes we successfully engineered a highly effective Bacterial Surface Display (BSD) system. This system is unique in that it enables a relatively high level of protein expression on the bacterial surface, optimized via the incorporation of the red fluorescent protein mCherry. The system is composed of series of proteins that are expressed from a single gene and, upon expression, results in a protein-of-interest displayed on the outer surface of *E. coli*. Therefore all forms of chemical analysis (such as protein/protein binding assessment) can be carried out directly on the surface of bacteria. This will greatly enhance our ability to generate libraries of potential binding partners and assess and isolate binding events using methods such as Fluorescent Activated Cell Sorting (FACS). The work describing the BSD system is complete and we are currently preparing a manuscript describing the design and implementation of the system.

[There is a report below that describes the Accomplishments in greater detail.]

Training Opportunities: The funding from this project enabled the direct financial buyout for one graduate student over a three-year period. In addition, this funding enabled the purchase of supplies and equipment that allowed the training of approximately seven additional graduate students (both M.S. and Ph.D. students), approximately 5 undergraduate students, and one high school student. The Ph.D. graduate student who is still working on this project, Brian Maniaci, is in his 4th year and anticipates defending his Ph.D. dissertation next year. He then plans to pursue a career in Biotechnology. Melissa Lokensgard defended her Ph.D. dissertation last year and is now employed as a Staff Scientist at InhibRx (a Biotherapeutic company). Benjamin Saylor defended his Ph.D. thesis in 2016 and is now a Postdoctoral Scholar at the Division of Endocrinology and Metabolism in the UCSD Department of Medicine. Peter Suon defended his M.S. thesis approximately 1.5 years ago and is now employed as a Scientist at Xip Biotechnology. Catrin Law was a high school student who was funded by the Army AEOP Program to perform research during the summer of 2015. Catrin is now an undergraduate student here at SDSU and is majoring in chemistry. The highly technological training that all of these bright students received would not have been possible without the funding afforded by this DOD ARO grant.

RPPR Final Report as of 28-Feb-2018

Results Dissemination: Thus far this project has resulted in the publication of one manuscript in the Journal of Molecular Catalysis B: Enzymatic. In addition will be in the final stages of research and preparing a manuscript that describes the main body of work that this grant funded. This work will be submitted as an article to the Journal of the American Chemical Society. This work has also generated enough data for a second publication of which the research and analysis is almost complete. There are also two additional sub-projects that are almost complete and will lead to publications that we hope to have submitted in the next year.

In addition to the above manuscripts my students and I have presented our results at a number of local, national, and international conferences. Titles of the presentations and posters for these work described at these meetings are listed below.

B. Maniaci, J. J. Love, "Metal-Mediated Dimerization of a Designed Protein-Protein Interface" Southern California Regional Users Meeting-Wyatt, Green Acre Campus Pointe, 10300 Campus Pointe Dr. San Diego, CA 92121 June 7th, 2016.

B. Maniaci, J. J. Love, "Metal-Mediated Dimerization of a Designed Protein-Protein Interface" San Diego State University Student Research Symposium, San Diego, CA 92182 (March 10th-11th, 2017)

A. C. Prem, P. Suon, M. Lokensgard, J. J. Love. "Computational Study of the Effects of Variable Temperatures on Protein Secondary Structure and Stability." San Diego State University Student Research Symposium, San Diego, CA 92182 (March 10th-11th, 2017)

A. C. Prem, P. Suon, M. Lokensgard, J. J. Love. "Computational Study of the Effects of Variable Temperatures on Protein Secondary Structure and Stability." 28th Annual CSU Biotechnology Symposium, Santa Clara Marriott, 2700 Mission College Blvd, Santa Clara, CA 95054 (January 5th-7th, 2017).

M. Lokensgard and J. J. Love. "Beta-Hairpins: Molecular Accessories for Helical Peptide Expression." The 29th Annual Symposium of the Protein Society, Fira de Barcelona, Barcelona, Spain (July 23, 2015).

M. Lokensgard and J. J. Love. "Beta-Hairpins: Molecular Accessories for Helical Peptide Expression." American Chemical Society Western Regional Meeting 2015, California State University San Marcos, San Marcos, CA (November 7, 2015).

M. Lokensgard and J. J. Love. "Beta-Hairpins: Molecular Accessories for Helical Peptide Expression." Student Research Symposium 2016, San Diego State University, San Diego, CA (March 4, 2016).

B. Saylor and J.J. Love. "Identification, Characterization, and Modification of Fatty Acid Modifying Enzyme from *Staphylococcus aureus*." The 29th Annual Symposium of the Protein Society, July 22-25, 2015, Fira de Barcelona, Barcelona, Spain.

B. Saylor and J.J. Love. "Identification, Characterization, and Modification of Fatty Acid Modifying Enzyme from *Staphylococcus aureus*." 2015 ACS Western Regional Meeting, CSU San Marcos, San Marcos, CA (November 2015).

B. Maniaci, C. Lipper, J.J. Love, "Optimization of a Designed Protein-Protein Interface" The 29th Annual Symposium of the Protein Society, July 22-25, 2015, Fira de Barcelona, Spain.

B. Maniaci, C Lipper, J.J. Love, "Optimization of a Designed Protein-Protein Interface" 2015 ACS Western Regional Meeting, CSU San Marcos, San Marcos, CA (November 2015).

P. Suon, M Navarro, L. Carmody, O. Romo, and J. J. Love. "Assessment of UCH-L3 Substrate Hydrolysis Using Ubiquitin Fusions." The 29th Annual Symposium of the Protein Society, Fira de Barcelona, Barcelona, Spain (July 23, 2015).

P. Suon, M Navarro, L. Carmody, O. Romo, and J. J. Love. "Assessment of UCH-L3 Substrate Hydrolysis Using Ubiquitin Fusions." American Chemical Society Western Regional Meeting 2015, California State University San Marcos, San Marcos, CA (November 7, 2015).

RPPR Final Report as of 28-Feb-2018

M. Acevedo, C. Scholl, C. Law, P. Suon, and J.J. Love. "Using Protein Design to Engineer the CIF Epoxide Hydrolase Enzyme for Neutralization of Mycotoxins." 2015 ACS Western Regional Meeting, CSU San Marcos, San Marcos, CA (November 2015).

M. Acevedo, C. Scholl, C. Law, P. Suon, and J.J. Love. "Using Protein Design to Engineer the CIF Epoxide Hydrolase Enzyme for Neutralization of Mycotoxins." The 9th Annual Student Research Symposium, San Diego, CA (March 2016).

P. Suon, M Navarro, L. Carmody, O. Romo, and J. J. Love. "Assessment of UCH-L3 Substrate Hydrolysis Using Engineered Ubiquitin Fusions with Variable Linker Lengths." The 28th Symposium of the Protein Society, The Manchester Grand Hyatt, San Diego, California, July 27 – July 30, 2014

P. Suon, M. Navarro, L. Carmody, O. Romo, and J. J. Love. "Assessment of UCH-L3 Substrate Hydrolysis Using Engineered Ubiquitin Fusions with Variable Linker Lengths." Student Research Symposium, San Diego State University, California, March 6-7, 2015.

M. Lokensgard and J. J. Love. "α-Hairpins as stability tags for recombinant expression of recalcitrant small helical proteins." Experimental Biology 2014. San Diego Convention Center, San Diego, California, April 26 - 30, 2014.

M. Lokensgard and J. J. Love. "Beta-hairpin tags for increased expression of helical peptides." The 28th Annual Symposium of The Protein Society. Manchester Grand Hyatt, San Diego, California, July 27 - 30, 2014

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)

Participant: Brian Maniaci

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Melissa Lokensgard

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Benjamin Saylor

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

RPPR Final Report
as of 28-Feb-2018

Other Collaborators:

Participant Type: High School Student

Participant: Catrin Law

Person Months Worked: 3.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Undergraduate Student

Participant: Maileen Acevedo

Person Months Worked: 3.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Markus Ruetsche

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Peter Suon

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Ariana Pientka

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Aishani Prem

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

RPPR Final Report
as of 28-Feb-2018

National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)

Participant: Myung Soo Ko

Person Months Worked: 9.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Undergraduate Student

Participant: Steven Phi

Person Months Worked: 3.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Undergraduate Student

Participant: Candace Lei

Person Months Worked: 3.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Undergraduate Student

Participant: Alissa Calderon

Person Months Worked: 3.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

Participant Type: Undergraduate Student

Participant: Gabriela Bremer

Person Months Worked: 3.00

Funding Support:

Project Contribution:

International Collaboration:

International Travel:

National Academy Member: N

Other Collaborators:

RPPR Final Report
as of 28-Feb-2018

DISSERTATIONS:

Publication Type: Thesis or Dissertation

Institution: San Diego State University

Date Received: 10-Jan-2018

Completion Date: 11/19/16 9:27AM

Title: Beta-Hairpin Fusion to Alpha-Helical Domains for Exploring Recombinant Protein Stability in E. coli

Authors: Melissa Lokensgard, John Love

Acknowledged Federal Support: **Y**

Publication Type: Thesis or Dissertation

Institution: San Diego State University

Date Received: 10-Jan-2018

Completion Date: 9/1/16 2:00PM

Title: Identification and Modification of Fatty Acid Modifying Enzyme From Staphylococcus aureus

Authors: Benjamin Saylor, John Love

Acknowledged Federal Support: **Y**

Publication Type: Thesis or Dissertation

Institution: San Diego State University

Date Received: 10-Jan-2018

Completion Date: 5/4/17 7:00AM

Title: Assessment of UCH-L3 Substrate Selectivity using Engineered Ubiquitin Fusions

Authors: Peter Suon, John Love

Acknowledged Federal Support: **Y**

Title: The De Novo Design of Protein-Protein Interfaces

John J. Love San Diego State University
Department of Chemistry and Biochemistry

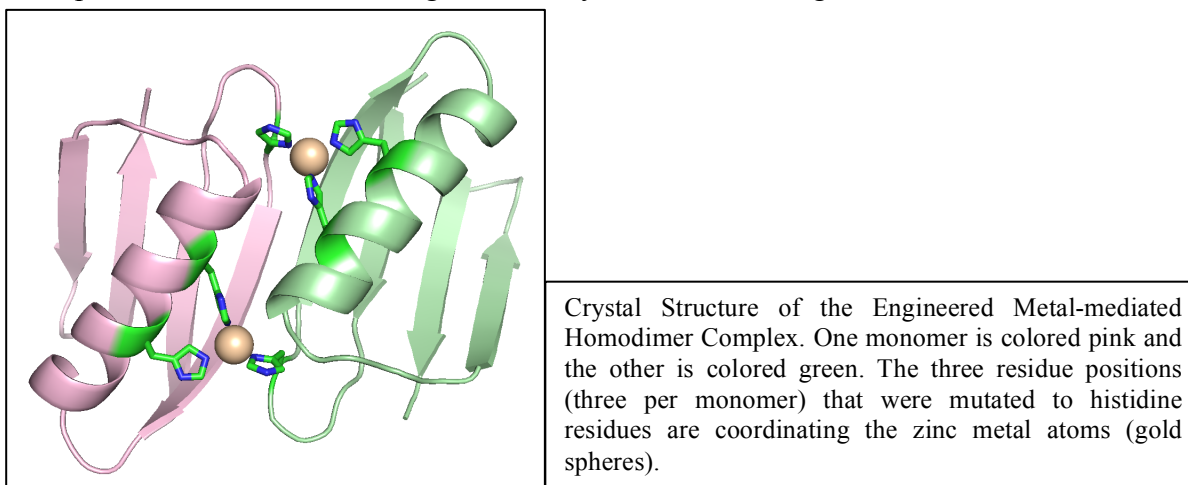
Contract Number: W911NF1310155 - **Dates Covered:** May, 13 2013 to May, 12 2017

Foreword

I am taking the opportunity in this ‘Foreword’ to state that the application of this grant was highly successful in 1) achieving the primary research goal of greatly extending our knowledge of the biophysical forces exploited to drive designed protein/protein interactions, and 2) training graduate, undergraduate, and even high school students in all of the molecular biology, biochemical, and protein design techniques that were brought to bear to drive specific protein complex formation.

The primary goal of the research funded by this DOD ARO grant was to reengineer a previously monomeric test protein such that it would form a specific homodimer complex. For this goal I can proudly state that we were highly successful (figure below). We have engineered at least three different homodimer complexes that are completely monomeric in the absence of metal yet form very clean dimer complexes in the presence of zinc. The DOD ARO funds that supported this research was absolutely crucial to this success, especially the funding that enabled the purchase of the Dynamic Light Scattering, Multiangle Light Scattering, and Circular Dichroism Spectropolarimeter instrumentation.

In regards to student training, the Ph.D. graduate student who is still working on this project, Brian Maniaci, is in his 4th year and anticipates defending his Ph.D. dissertation next year. He then plans to pursue a career in Biotechnology. Melissa Lokensgard defended her Ph.D. dissertation last year and is now employed as a Staff Scientist at InhibRx (a Biotherapeutic company). Benjamin Saylor defended his Ph.D. thesis in 2016 and is now a Postdoctoral Scholar at the Division of Endocrinology and Metabolism in the UCSD Department of Medicine. Peter Suon defended his M.S. thesis approximately 1.5 years ago and is now employed as a Scientist at Xip Biotechnology. Catrin Law was a high school student who was funded by the Army AEOP Program to perform research during the summer of 2015. Catrin is now an undergraduate student here at SDSU and is majoring in chemistry. The highly technological training that all of these bright students received would not have been possible without the funding afforded by this DOD ARO grant.



Abstract: The primary focus of this research project was to explore the biophysical forces that drive biomolecular complex formation and further elucidate the parameters that give rise to specific protein/protein interactions. Thus far a large body of data has been collected and analyzed on important physical properties of natural protein complexes¹⁻⁴ and it was our intention to add to this body by engineering *de novo* (completely new) protein/protein complexes. Using this ‘inverse’ approach we furthered the understanding of biomolecular specificity by generating protein complexes from previously monomeric ‘test’ proteins. The goal was to elucidate the key physical features needed to drive specific protein/protein interactions. To achieve this we utilized robust computational algorithms in combination with rational design methods to drive small test proteins to form complexes of specific structure. After a number of tantalizing yet imperfect results we recently had great success in generating metal-mediated specific dimer complexes.

Summary of Important Results:

This final report describes Important Results made thus far on the following four distinct but related projects. A number of these projects are still being pursued in the Love Laboratory:

- 1) The successful *de novo* design of three metal-mediated dimer complexes was achieved using as the test protein scaffold the β 1 domain of streptococcal protein-G. The three protein-G mutants are completely monomeric in the absence of metal yet form clean dimers in the presence of zinc. Successful dimer formation was confirmed using size-exclusion chromatography combined with multiangle light scattering (SEC-MALS) (instrumentation purchased with funds from this grant). The successful design target structure was confirmed with X-ray crystallography that was used to solve the high-resolution structure of two metal-mediated complexes. The metal-mediated complexes were also biophysically characterized using Fluorescence Polarization, Isothermal Titration Calorimetry, and Circular Dichroism (instrumentation purchased from this grant). We recently were able to measure the binding affinity of the complexes using MicroScale Thermophoresis (instrumentation purchased from a DOD grant of my colleague Tom Huxford). We are in the process of preparing two manuscripts describing these results. The primary data (described above) will be submitted to the *Journal of the American Chemical Society* as a Research Article.
- 2) The design and engineering of a chimeric protein that consisted of a staphylococcal lipase combined with a *Drosophila* ethanol binding protein. The design of this protein/protein chimera resulted in a novel enzyme that is capable of catalyzing the formation of fatty acid alkyl esters (biodiesel) using ethanol as an alkyl donor. The natural substrates for the staphylococcal lipase are large alcohols, which are not viable for large-scale biodiesel production. The design of this protein/protein chimera resulted in an engineered enzyme with altered substrate selectivity that provides potential leads for enhanced biodiesel production. This work was published in 2016 (Saylor, B. D., Love, J. J., A secreted *Staphylococcus aureus* lipase engineered for enhanced alcohol affinity for fatty acid esterification. *Journal of Molecular Catalysis B: Enzymatic* 133 (2016) S44-S52 <http://dx.doi.org/10.1016/j.molcatb.2016.11.013>)

- 3) Exploring the role that small, relatively well-folded secondary structure elements (β -hairpins) play in protein structure and stability. The β 1 domain of streptococcal protein-G contains two β -hairpin secondary structure elements that are understood to contribute to the enhanced stability observed for this 56 amino acid protein. To explore if these β -hairpin elements could enhance the stability and bacterial expression of other proteins they were incorporated into proteins that previously did not express well in standard *E. coli* expression systems. This design approach proved to be highly successful and demonstrated that the incorporation of the protein-G β -hairpin elements at either the N- or C-terminus of proteins recalcitrant to bacterial expression greatly improved the amount of produced protein. This was the thesis project for Melissa Lokensgard who successfully defended her thesis in 2016 and we are currently preparing a manuscript describing her thesis research.
- 4) A fourth project that funding from this DOD grant greatly advanced was the development of a highly effective bacterial surface display (BSD) system. The main purpose of this system is to enhance the ability to engineer novel protein/protein interactions. The BSD system enables the display a large library of mutated protein variants on the outer surface of *E. coli*. The system is similar in concept to phage display except that, since library expression is directed to the bacterial surface, it is simpler as it only entails the use of one organism (*E. coli*). The system consists of a fusion construct that contains 1) a leader sequence that directs the fusion to the periplasmic space, 2) the red fluorescent protein mCherry, for expression optimization, and 3) an autotransporter esterase that anchors the construct in the outer membrane. The system was benchmarked with a series of test proteins (described below) in addition to the incorporation of the metal-mediated dimers described in section 1 above. We are currently using the display of the metal-mediated dimer on the surface of *E. coli*, in conjunction with a free fluorescent metal-mediated dimer variant, to ascertain its effective as a probe in a Fluorescent Activated Cell Sorting (FACS) assay.

Full Descriptions of the Results and Status of the Four-Main Projects Described Above

1) *De novo* Docking – the small (56 amino acid) β 1 domain of streptococcal protein-G was successfully utilized to engineer novel metal-mediated protein-protein interactions.

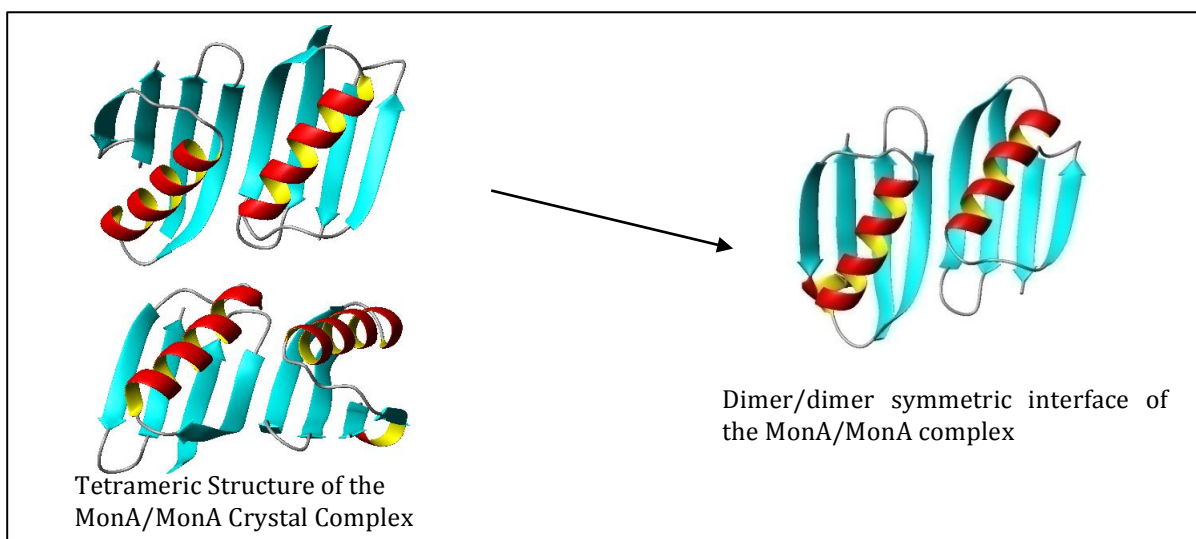
Title: Designed Metal-Mediated Protein Dimerization

Background and Previous Results: This project originated with the computational docking of protein-G to itself followed by the use of the ORBIT suite of protein design algorithms to mutate residues at the newly created interface⁵⁻⁷. The resulting heterodimer, referred to as monomer-A (MonA) and monomer-B (MonB), was physically expressed within *E. coli*, purified, and tested for complex formation. NMR and analytical ultracentrifugation was used to demonstrate that the complex did form but with relatively weak binding affinity. A former M.S. student used the model structure of the heterodimer in combination with rationale design to re-engineer the original MonA/MonB complex. This

resulted in 24 additional MonA variants and 24 additional MonB variants. For a number of these variants metal-coordination binding sites were built into the interface with the goal of increasing binding affinity in the presence of specific metals. Many of the mutants had little to no effect on binding affinity of the MonA/MonB complex except for three of the MonA mutants. These formed higher-order homo-complexes (*e.g.*, dimers, trimers) in the presence of the metal yet remained completely monomeric when no metal was present. The degree of complex formation, for the three MonA variants that formed complexes, was assessed using NMR and size-exclusion chromatography (SEC). The results indicate that these three variants do form higher order homo-complexes but it was not possible to definitively determine if the complexes were dimers, trimers, or higher-order complexes. Based on these successful results we solved the 3-dimensional structures of these complexes using X-ray crystallography.

In collaboration with an SDSU crystallographer, Tom Huxford, we used the crystal structure of wild-type G β 1⁸ with molecular replacement to solve the structure of one of the self-assembled MonA mutant (E21H, L25H) complexes (Figure 1 below). The structure was refined to 1.48Å resolution. The complex forms a tetramer within the asymmetric unit (actually a dimer of two dimers) and it appears that the zinc atoms function as catalysts for crystallization in addition to driving self-assembly. The dimeric complex is quite exciting as it provides us with an experimentally determined complex structure that we have used to re-engineer interfacial amino acid positions.

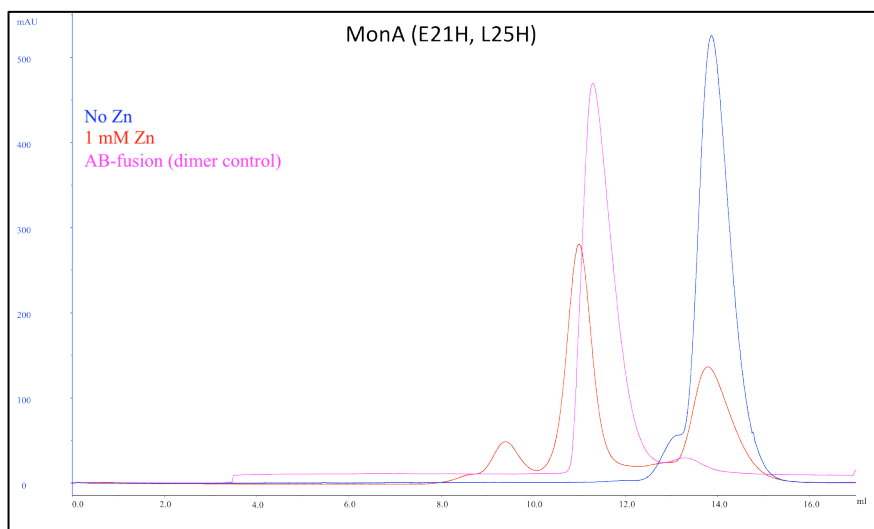
The X-ray crystal structure of a second MonA mutant complex (A32H, A36H) was solved by single wavelength anomalous dispersion (SAD) phasing using the absorption edge of zinc and refined to 1.49Å resolution. This allowed for the direct calculation of zinc ion positions within the protein unit cell and calculation of experimental electron density maps from these initial zinc positions. The overall features of the structure of this MonA complex is highly similar to that of the structure described above for the E21H, L25H mutant (data not shown). Both the refined structure of this mutant, and the one described above, reveal that zinc ions are, in fact, bound by the engineered histidine pairs. However, it was not clear from the structures alone how this metal binding contributes to higher order oligomerization of the MonA multimers in solution.



X-ray crystal structures of the E21H, L25H MonA mutant variant.

Status as of 2015: In a previous annual report (2014) it was stated - “This is a very exciting time for this project as we are about to begin a more detailed biophysical analysis (e.g., SEC-MALS, Circular Dichroism, etc.) on a number of previously engineered complexes plus two newly engineered variants.” This was followed the next year (2015) with the following statement “This was very true but I will have to state that this past year has been extremely frustrating as the MALS data (generated after running multiple standards, multiple milligram quantities of different samples from multiple bacterial preps, multiple controls, using at least two different columns, and cleaning the columns and optics frequently, etc.) has been highly inconsistent regarding the reported molecular weights determined with the MALS instrument. This experience has been frustrating yet we are persevering and coming to terms with the fact that the MALS instrument is extremely sensitive to even vanishing small amounts of larger molecular weight entities (aggregates).”

Status as of 2016: In 2016 I was very happy to report that we had persevered and worked out the issues associated with the inconsistent data generated from the Multi Angle Light Scattering (MALS) instrument and we are now generating extremely consistent and highly accurate molecular weights on newly designed metal-mediated dimers. The primary issue with the previous data was that the instrument was reporting highly variable calculated molecular weights for various peaks off of the size exclusion chromatography (SEC) column. We now have a better understanding of the origin of these problems. The complexes that were generated previously gave rise to multiple peaks from the SEC column. An example SEC trace is shown below.



Size-Exclusion Chromatography Run on the MonA (E21H, L25H) Variant. The three different colored traces correspond to the MonA (E21H, L25H) variant run in the absence of zinc (blue trace), and in the presence of zinc (1 mM, red trace). The violet/pink trace corresponds to a protein-G fused dimer control (MW standard).

As observed there is clearly evidence of multiple peaks in the red trace that corresponds to the MonA variant in the presence of 1 mM zinc [*i.e.*, there is a monomer peak under the blue trace peak, possibly a dimer peak (near the pink dimer control peak) and then also an additional smaller peak that may be a tetramer or larger molecular weight complex]. We initially assumed that the MALS instrument would be able to accurately measure the molecular weights of these different complexes but that is definitely not the case. We ultimately learned that the MALS instrument is extremely sensitive to any impurities

(aggregates) even if the impurities are at a vanishingly low concentration and thus not detectable via UV during the SEC run. In addition, the MALS instrument cannot accurately measure the molecular weights on samples that contain multiple species. So it was not possible to accurately determine the molecular weights, nor the degree of multimerization for these samples that demonstrated multiple states of complex formation (*i.e.*, they are poly-dispersive). The poly-structural nature of these previously designed complexes was also observed for the structures solved by X-ray crystallography and illustrated above (image page 4).

Biophysical Lessons Learned: Although it is definitely not possible to determine the multimeric states of these complexes we did learn a number of very important structural aspects of engineered complex formation for the small test protein, protein-G. The main lesson learned is that our initial design, consisting of protein-G packed helix-face-to-helix-face is likely not possible due to the preferred enthalpic energy associated with the regular pattern of intermolecular hydrogen bonds that occur when protein-G associates via hydrogen bonding and extension of the β -sheets observed in the crystal structures of our multimeric complexes. Therefore we decided to pursue this structural form of protein-G complex formation and we designed dimers that associate through intermolecular hydrogen bonds as a β -sheet extension. We are happy to report that this approach has been phenomenally successful.

A Well-Designed and Successful Protein Dimer: Based on the results described above we decided to pursue the goal of metal mediated protein dimer formation using the dimer orientation observed in the crystal structures of the multimeric complex formations. Therefore instead of continuing with the ‘computationally’ derived helix-face-to-helix-face orientation we opted to exploit the β -sheet-to- β -sheet-extension structure observed in the multimeric complexes. To enhance the likelihood of creating a metal-mediated dimer structure we utilized previous work from a different research group that inadvertently generated a stabilized dimer variant of protein-G using a phage-display methodology referred to as Proside⁹. In that work specific codons in the gene for protein-G were randomized and the resulting library was displayed on phage and challenged for stability through exposure to proteases. Briefly, the results of this approach generated a four-fold mutant of protein-G that achieved greater thermal stability through tight dimer association. The crystal structure of that variant was solved and, not surprisingly, the dimer interface consists of the β -sheet-to- β -sheet extended orientation that we observed in the crystal structures of the multimeric variants of the MonA protein-G variants obtained previously (images page 4).

We synthesized the gene and expressed the protein for the four-fold mutant variant (termed G β 1-M2) described in the above publication⁹. The four mutations are located at the dimer interface and consist of the following - E15V, T16L, T18I, and N37L. This well-characterized dimer is the variant that enabled us to confirm that, for accurate molecular weight determination using SEC MALS, one needs a very clean and monodispersive sample. When the monodispersive MonA samples (no zinc added) were run through SEC MALS the measured mass was generally correct (6.2 kD), but when 1 mM zinc was added, and the samples became polydisperse, the measured molecular weights were highly variable and did not match the weights of potential dimer, trimer, tetramer, or octomer structures. But when we ran the G β 1-M2 four-fold mutant, which forms a very clean dimer, the measured molecular weight from the SEC MALS instrument is very close to the calculated molecular weight (~12.2 kD). This was an exceptionally gratifying finding, especially as the SEC

MALS instrument, which was purchased using funding from this grant, is key to measuring the molecular weight of our designed dimers.

Using the crystal structure of the G β 1-M2 variant, plus the analogous crystal structures of our polydisperse complexes, we designed a series of protein-G variants that are perfectly monomeric in the absence of zinc and form very clean dimer complexes in the presence of zinc. These variants consist of variations of the above-described four ‘dimerization’ mutations plus specific interface mutations in which select positions were mutated to metal binding histidine residues. Below is a list of the names and mutations of the additional mutants for which the genes were synthesized, proteins expressed, purified, and examined using SEC MALS. The interfacial mutations that gave rise to the G β 1-M2 constituent dimer variant are highlighted in yellow.

M11H: T11H, E15V, T16L, T18I, V29H, Y33H, N37L

This variant is the G β 1-M2 variant plus three additional histidine binding sites.

M12H: L12H, E15V, T16L, T18I, V29H, Y33H, N37L

The G β 1-M2 variant plus three different histidine binding sites (one different from above).

M12I: L12H, E15V, T16V, T18I, V29H, Y33H, N37V

Positions 16, 37 were changed from Leu $\rightarrow$ Val

M13: L12H, E15V, T18I, V29H, Y33H

The G β 1-M2 variant with positions 16, 37 reverted back to the wild type amino acids from protein-G.

GB12: L12H, V29H, Y33H

Protein-G with histidine residues corresponding to the above M12H variant.

Construct 1: L12H, T16L, V29H, Y33H, N37L

Construct 1 was engineered to determine if the mutations at positions 16 and 37 are sufficient for protein dimerization. Future mutations will vary the positions 16 and 37 from leucine to valine.

Construct 2: L12H, E15V, T16A, T18I, V29H, Y33H, N37A

These mutations were made in an attempt to reduce the tendency of M12I to self-associate in the absence of zinc by mutating positions 16 and 37 from valine to alanine. Double mutation of positions 16 and 37 to alanine.

Construct 3: L12H, E15V, T16A, T18I, V29H, Y33H, N37V

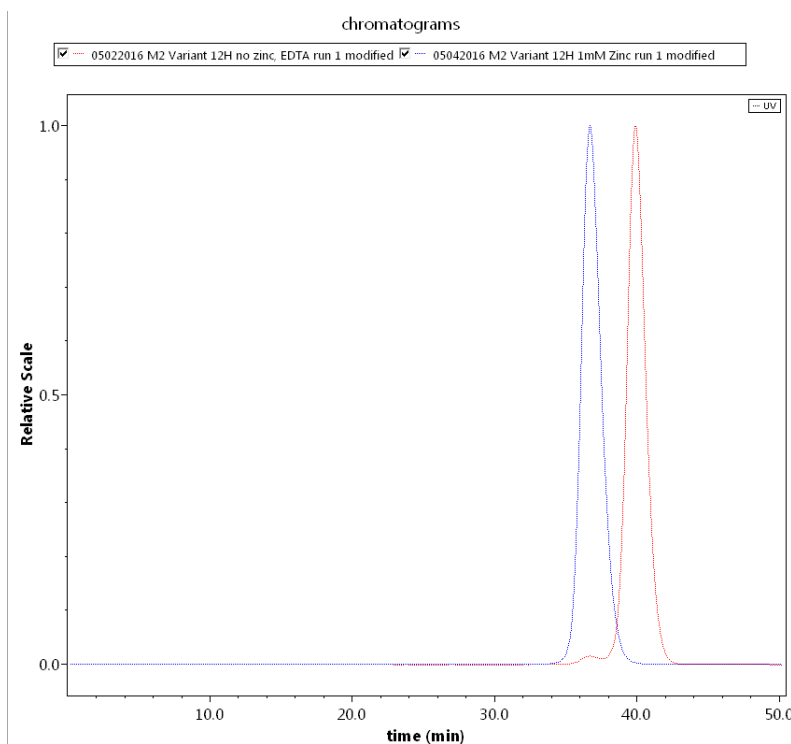
These mutations were made in an attempt to reduce the tendency of M12I to self-associate in the absence of zinc by mutating position 16 valine to alanine and position 37 to valine.

Construct 4: L12H, E15V, T16V, T18I, V29H, Y33H, N37A

These mutations were made in an attempt to reduce the tendency of M12I to self-associate in the absence of zinc by mutating positions 37 from leucine to alanine. Position 37 mutated to alanine.

Molecular Weights (all values within the table correspond to molecular weights in kD).

Protein Name	EDTA	No EDTA Additive	250uM	500uM	1mM	2mM
Protein G	6.7	8.0	6.7	6.5	6.7	6.6
M2	12.0	12.1	12.2	12.2	12.4	12.5
M11H	12.2	12.1	12.3	13.7	12.3	12.4
M12H	6.2	11.9	12.0	12.6	12.2	12.3
M12I	6.1	7.0	11.6	12.7	12.2	13.1
M13	6.1	6.4	8.4	8.5	8.8	9.5
GB12	6.0	6.1		7.7	7.6	7.3
Construct 1	6.6	6.6	12.3	12.9	12.4	12.7
Construct 2	6.5	6.5	10.9	11.6	12.0	12.0
Construct 3	6.9	6.8	7.1	7.0	6.8	6.7
Construct 4	6.1	6.4	11.6	12.0	12.1	12.3



Representative SEC-MALS Analysis Metal-Mediated Dimerization. This chromatograph is a representative SEC trace of the M12H variant that is primarily monomeric in the absence of zinc (red trace), and is transformed into a very clean dimer in the presence of 1 mM zinc (blue trace).

The MALS measured molecular weights for the monomer and dimer of M12H are 6.2 and 12.2 kD respectively. This variant shows evidence of some dimer formation in the

absence of EDTA and we believe this may be due to residual metals that originate from the bacterial expression media. In the presence of metal chelating EDTA this variant is a monodisperse monomer (kD 6.2).

We are absolutely ecstatic over these really clean results and have mutated additional interfacial residues in order to probe the thermodynamic contributions from specific interfacial residues. Some of the mutant variants generated thus far are described in the above list. We have subjected the most promising variants to X-ray structure characterization and fluorescence polarization analysis.

X-Ray Crystallographic Analysis: Structural validation of the designed variants has been undertaken using X-ray crystallography. Crystals for the M12H and Construct 1 variants have been grown in the presence and absence of the zinc dimerization-inducing metal. These variants were selected for structural validation based upon the accurate molecular weights determined using SEC-MALS.

A Mosquito LPC nanoliter crystallization robot was used for high-throughput screening of crystallization conditions. The initial screen hits were further optimized for improved crystal formation and quality. The optimal crystallography conditions for the variants were as follows:

M12H (no metal): 31% PEG 4,000, 0.1M Tris, pH 8.5, 200 mM MgCl₂
M12H (with zinc): 4M NaCl, 0.1M HEPES, pH 7.5, 50 mM MgCl₂
Construct 1 (no metal): 30% PEG 4000, 0.1M HEPES, 200 mM MgCl₂
Construct 1 (with zinc): 18.5% PEG 4000, 0.1M HEPES, pH 7.5, 50 mM MgCl₂

A cryoprotectant containing the optimized crystallization buffer and 50% glycerol was used for storage and transport in liquid nitrogen. Data was collected from two different beam sources: 1) SDSU home source (Rigaku-RU-H39, with a rotating anode generator - 50 kilowatts at 100 milliamps) and 2) the Berkley Lab Advanced Light Source (ALS-5.0.1). The data collected at SDSU was imaged using the CrystalClear program. Data processing was accomplished with the HKL2000 software to determine the symmetry and space group of the proteins. We have refined the data using molecular replacement in the CCP4 suite of programs (Phaser MR and Molrep). Refinement of the preliminary model was analyzed with the programs contained in the CCP4 suite, including Refmac5 and Molprobability.

Data Collected at the Berkley Lab Advanced Light Source ALS 5.0.1

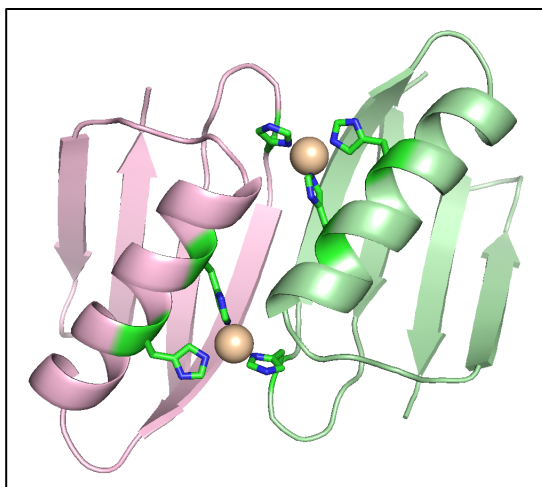
Naming convention (Puck, Position #):

37402: M12H, zinc, resolution of 1.34Å
37404: M12H, No zinc, resolution of 1.70Å
37406: Construct 1, No zinc, resolution 1.70
37408: Construct 1, zinc, resolution 1.50Å

Data Collected from home source:

M12H, no zinc: with a high resolution of 2.30Å
Construct 1, Zinc: with a high resolution of 1.70Å

We have successfully solved the structures of the above two variants (M12H and Construct 1) in both the presence (image below) and absence of zinc. Not surprisingly, the quality (resolution) of the data from the samples that contain zinc are much higher. For the samples with zinc we have preliminary structures that are not fully refined yet but fitting of the model structures into the electron density clearly indicates that the designed metal binding residues (histidines) are binding zinc at the engineered interface.



Crystal Structure of the Engineered Metal-mediated Homodimer Complex. One monomer is colored pink and the other is colored green. The three residue positions (three per monomer) that were mutated to histidine residues are coordinating the zinc metal atoms (gold spheres).

Binding Affinity Assessment using Fluorescent Polarization (previous): The variants M12H and Construct 1 were mutated additionally to contain a glycine-serine (GS) linker at the C-terminus and a cysteine residue that provides the ability to label these proteins with a fluorophore. Protein-G and all the variants thus far generated do not contain cysteine residues except for these two variants that contain modified residues at the C-terminus. Selective labeling at the C-terminal cysteine provides the ability to study fluorescence anisotropy of the variants in the monomeric form, and separately, in the dimeric forms. We attempted to use fluorescence polarization to measure monomer-dimer equilibrium by observing changes in anisotropy upon formation of the more slowly rotating dimeric species. The fluorescence polarization titration experiments were meant to provide an estimate of the dimer dissociation constant (K_d) in the presence and absence of metal (zinc). For this assessment the amount of labeled protein was kept constant, while the unlabeled protein was titrated into the solution.

Binding Affinity Assessment (update): It appears that, for each step we undertook in this project (except for the X-ray crystallography), that assessment of a biophysical aspect of the engineered dimer complexes entailed early steps of considerable frustration followed by reassessment and then success with either a revision of the biophysical method (*e.g.*, SEC-MALS) or a application of a new method. This is definitely the case with the measurement of the binding affinities for the engineered complexes using fluorescence anisotropy as it was not possible to obtain binding affinities using this method. But as with all previous methods, we persevered and ultimately overcame the previous issues/problems with measuring the binding affinities.

In brief, measurement of the binding affinities using fluorescent anisotropy did not ultimately turn out to be feasible. We believe that this is because the molecular weight differences between the metal driven dimers (~12 kD) is not sufficiently large enough,

relative to the monomers (~6 kD), to generate enough dynamic range in the measured differences in the associated anisotropy. There wasn't a large enough, or consistent enough change in the fluorescent signal upon titration of the unlabeled variants into the fluorescently labeled variants. As with the previous forms of biophysical assessment that was frustrating we systematically varied all instrument parameters, actually changed the covalently attached fluorophore (which entailed significant bench lab work), and reevaluated all aspects of the fluorescent anisotropy assessment, but to no avail.

Therefore we decided to discontinue this approach and instead pursue the assessment of the binding affinities using MicroScale Thermophoresis.

Successful Assessment of Complex Binding Affinities: We are currently in the process of using MicroScale Thermophoresis (MST) to measure the binding affinities of the various engineered metal-mediated dimer complexes. We are very happy to report that this method, as opposed to fluorescent anisotropy, is generating data that is enabling the relatively accurate assessment of the binding affinities for the engineered dimer complexes. The instrument we are using to generate the MST data is a Nano Temper Monolith NT.115⁸. A departmental colleague, Tom Huxford, used funding from a similar DOD funding source to purchase this instrument. Thus far we have generated preliminary data, which indicates that the complexes binding constants fall approximately in the sub- to low-micromolar range. These results are especially gratifying especially since we worked on the fluorescent anisotropy experiments for over 1.5 years.

Current Status: We are happy to report that the final series of MST experiments are currently being run and therefore we anticipate completing the manuscript that describes these results in the very near future. We plan to submit the manuscript as an article to the *Journal of the American Chemical Society*. In addition to publishing the main scientific success of engineering metal-mediated dimers we also plan to publish a second paper that describes the crystal structures of the first round of poly-dispersive complexes (*i.e.*, those that formed complexes that are heterogeneous in regards to the degree of multimerization). These complexes form unique crystal structures and may provide enhanced structural insights into the means by which disease associated proteins form aberrant protein/protein fibers (such as β -amyloid) through β -sheet associations.

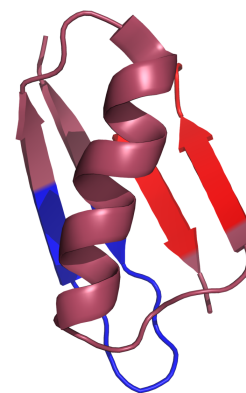
2) A secreted *Staphylococcus aureus* lipase engineered for enhanced alcohol affinity for fatty acid esterification.

This project was the Ph.D. thesis project of my former student Benjamin Saylor. Benjamin is currently a Postdoctoral Scholar at the Division of Endocrinology and Metabolism in the UCSD Department of Medicine working in the laboratory of Dr. Yury Miller. A primary goal of this research was to identify and re-engineer the fatty acid modifying enzyme (FAME) detected in solutions surrounding infections of methicillin resistant *Staphylococcus aureus* (MRSA)^{10,11}. Benjamin isolated proteins secreted by *Staphylococcus aureus* and then identified and characterized two putative genes for the FAME enzymes. FAME is an esterase, found in MRSA lesions and MRSA culture broth, which esterifies free fatty acids using a relatively broad range of alcohols. The protein responsible for this activity was isolated and its activity was characterized using various alcohol substrates that are viable candidates as alkyl donors. To achieve this, *Staphylococcus* strains were cultured, filtered, and fractionated using size-exclusion chromatography. Gas chromatography-mass spectrometry was used on the size-exclusion fractions that exhibited the ability to catalyze the formation of the fatty acid alkyl ester butyl oleate. Mass spectrometric protein sequencing revealed several proteins as the predominant protein components of the selected size-exclusion fractions. Based on predictions made by similarities in existing *Staphylococcus* genomes it was hypothesized that the formation of butyl oleate was driven by one or both of two proteins that exist at high concentrations in the selected fractions. These two highly similar enzymes fall (based on predicted structural motifs) into the glycerol ester hydrolase (geh) family. PCR was used to extract the genes for these two proteins from *S. aureus* and cloned them into *E. coli* using the pet22b vector. *E. coli* culture lysates containing either protein exhibited lipase activity, as shown by spectrometric monitoring of nitrophenyl acetate hydrolysis. The enzymes were also shown to have the ability to catalyze the formation of butyl oleate, supporting our prior hypothesis. Based on sequence similarities, both of the candidate proteins were likely to have structures highly similar to those of structurally classified lipases and phospholipases in the same family. A protein purification scheme was developed using Nickel column chromatography followed by an additional purification step using sepharose ion exchange columns. Protein expression was verified using standard SDS-PAGE. Lipase activity for the recombinantly expressed proteins was confirmed using a nitrophenyl acetate assay. Additional assays demonstrated that the wild-type enzymes exhibited a strong substrate preference for large hydrophobic alcohols such as butanol and low preference for the more convenient sources of alkyl methyl/ethyl groups such as ethanol. It was demonstrated that the enzymes bind larger alcohol substrates well, yet exhibit much lower affinity for methanol and ethanol. In an effort to improve the rate at which they can produce biodiesel with ethanol, an engineered a protein/protein chimeric variant was designed that utilized the “LUSH” ethanol-binding domain from *Drosophila* chemoreceptor organs. We hypothesized that the proximity of an alcohol-binding domain would increase the residence time of the substrate near the catalytic domain. Catalytic analysis followed by GCMS testing of the FAME/LUSH fusion proteins indicated a marked increase in the overall catalytic activity using ethanol as a substrate.

This work was completed in 2016 and published as an article in the *Journal of Molecular Catalysis B: Enzymatic*¹².

3) Exploring the role that small, relatively well-folded secondary structure elements (β -hairpins) play in protein structure and stability.

Background and Previous Results: In a previous Love Lab project, an *in vivo* screen for protein stability was devised and implemented^{13,14}. The project resulted in the creation of 20 protein-G mutant variants that span a thermal melting temperature range from as low as ~20 °C to greater than 100 °C. The screen was developed to generate a random pool of mutants and for the selection of stable protein-G variants. During screen assessment it was also necessary to prove the corollary and determine if the screen could identify low-stability variants. This test was successful as two thermally destabilized protein-G variants were isolated. One variant had a melting temperature of 28 °C and the other was so destabilized that it was not possible to determine its melting temperature (it exists as a potentially partly folded random-coil). What was quite interesting about these variants, and the original motivation for this project, was that they were successfully expressed and purified from a standard *E. coli* expression system, even though they were quite unstable and thus highly prone to degradation by proteases within the cytosol of *E. coli*. We hypothesized that the reason for this was due to the presence of N- and C-terminal β -hairpin secondary structure elements (shown in blue and red in figure at right).

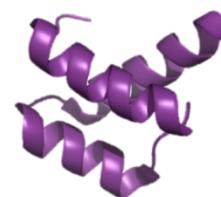


Protein-G Structure. The N- and C-terminal β -hairpins are colored blue and red respectively.

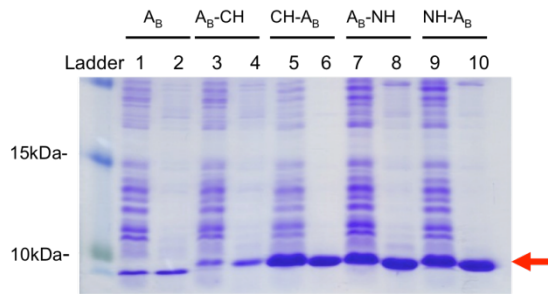
Using these β -hairpin secondary structure elements we developed a method for stabilizing proteins that previously did not express well in standard *E. coli* systems. These methods have also inspired various physical and biological questions. What changes in physical properties and structure result from the addition of this sequence? What protein-protein interactions in the biological expression system are affected by these changes?

To test if these β -hairpin sequences can increase the expression of other test proteins, a twelve amino acid N-terminal β -hairpin from *Streptococcal* Protein G (G β 1) was appended to various test proteins. The amino acid sequence of this motif is LILNGKTLKGET, which we term “NH” (N-terminal Hairpin). When the NH sequence is added to either terminus of *Staphylococcal* Protein A B domain (SpA_B), a helical protein of similar size and function and recalcitrant to expression in *E. coli*, we observe considerably higher yields of the chimeric protein. As can be seen in the gel image in the figure on the next page (on the left), the expression of protein-A (A_B) is greatly enhanced when either β -hairpin is appended to either of its termini. This is true except for when the C-terminal protein-G β -hairpin is appended to the C-terminus of protein A (lanes 3 and 4, gel on left).

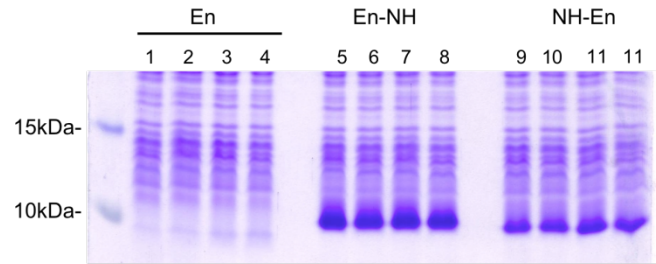
We have duplicated this result with another small helical domain, the *D. melanogaster* Engrailed homeobox DNA binding domain (image at right). This results in two new proteins, NH-En, and En-NH. As with the case of the modified protein-A variants, the expression levels of the Engrailed homeobox domain is greatly enhanced when it is expressed with the N-terminal β -hairpin of protein-G appended to either of its termini (right gel image below).



Engrailed homeobox DNA binding domain.

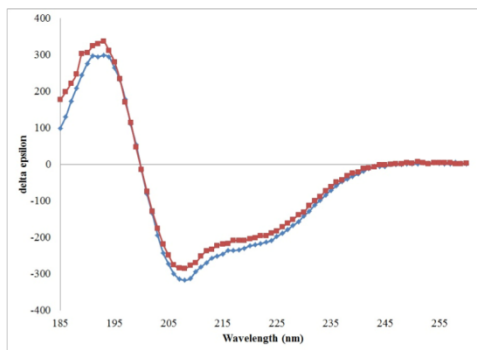


Expression Test of N- and C-terminal proG β -hairpins appended to the N- and C-termini of proA. Odd lanes – total cell lysate. Even lanes – samples post freeze-thaw. A_B – wild-type protein-A. A_B -CH – C-terminal proG β -hairpin appended to the C-terminal of proA. CH- A_B – C-terminal proG β -hairpin appended to the N-terminal of proA. A_B -NH – N-terminal proG β -hairpin appended to the C-terminal of proA. NH- A_B – N-terminal proG β -hairpin appended to the N-terminal of proA.

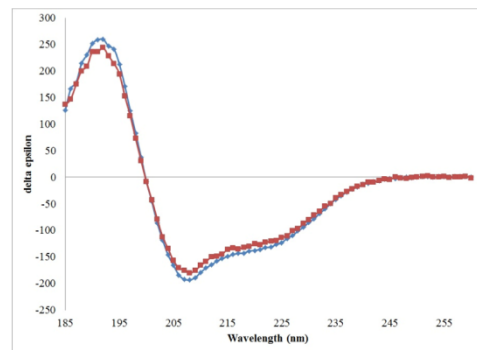


Expression Test of N- and C-terminal proG β -hairpins appended to the the Engrailed homeodomain. Lanes 1-4 – wild-type Engrailed. En-NH (lanes 5-8) – Engrailed with the N-terminal proG β -hairpin appended to the C-terminal of Engrailed. NH-En (lanes 9-11) – N-terminal proG β -hairpin appended to the N-terminal of Engrailed.

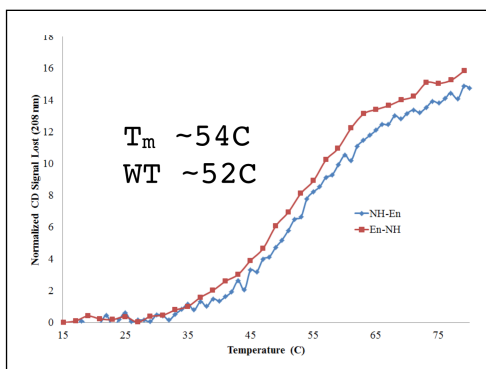
To investigate the physical properties of these newly engineered proteins, we also employed various qualitative and spectroscopic methods. Circular dichroism spectra of NH-En and En-NH (images below) show that the helicity of the parent En protein is retained, and that some beta character, presumably corresponding to the added NH sequence, can be revealed by spectral deconvolution algorithms of the CD data. These two engineered proteins also retain the relatively high thermal stability of En. Expanding on our original NH sequence, we also explored if a similar 17-amino acid hairpin sequence (the N-terminal hairpin of Ubiquitin, “Ub-Hp”) can also serve as an expression-enhancing tag when fused to either terminus of Sp A_B and En. All of these proteins were expressed and purified and the secondary structure and thermal denaturation temperatures were measured by CD.



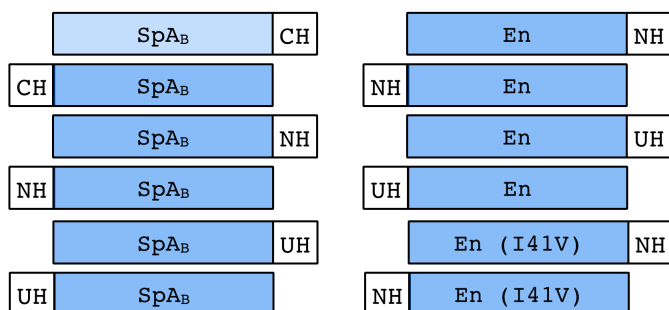
CD Spectrum of NH-Engrailed. Red curve –wild-type, blue curve NH-En. The melting temperature of this variant (55 °C) is highly similar to that of the wild-type domain (51 °C). The presence of the proG β -hairpin does not perturb helical content as illustrated by the similar spectra and also predicted by a CD spectral deconvolution algorithm.



CD Spectrum of Engrailed-NH. Red curve –wild-type, blue curve En-NH. The melting temperature of this variant is the same as the NH-En variant and highly similar to that of the wild-type domain (51 °C). The presence of the proG β -hairpin does not perturb helical content as illustrated by the similar spectra and also predicted by a CD spectral deconvolution algorithm.

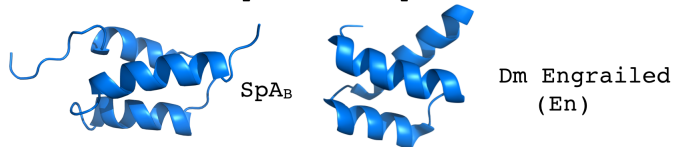


CD thermal melts of NH-Engrailed and Engrailed-NH. The melting temperatures of both variants are highly similar to each other ($\sim 54^\circ\text{C}$) and only slightly higher than the wild-type engrailed domain ($\sim 51\text{--}52^\circ\text{C}$). This implies that the presence of the N-terminal protein-G β -hairpin at either the N- or C-terminus of engrailed 1) does not perturb the overall structure the engrailed domain, and 2) that the protein-G β -hairpin does not appreciably interact with the engrailed domain.



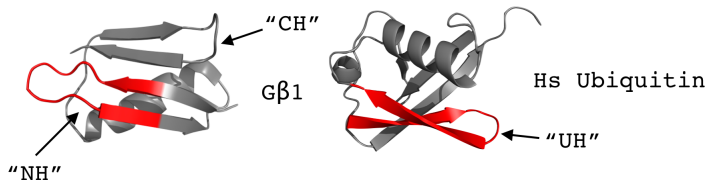
Schematic representations of the different β -hairpin chimeric fusion proteins. SpA_B – helical protein-A; CH – protein-G C-terminal β -hairpin; NH – protein-G N-terminal β -hairpin; UH – ubiquitin β -hairpin; En – Engrailed domain; En (I41V) – destabilizing mutation in the hydrophobic core of the Engrailed domain at position 41.

Recalcitrant to cytosolic expression

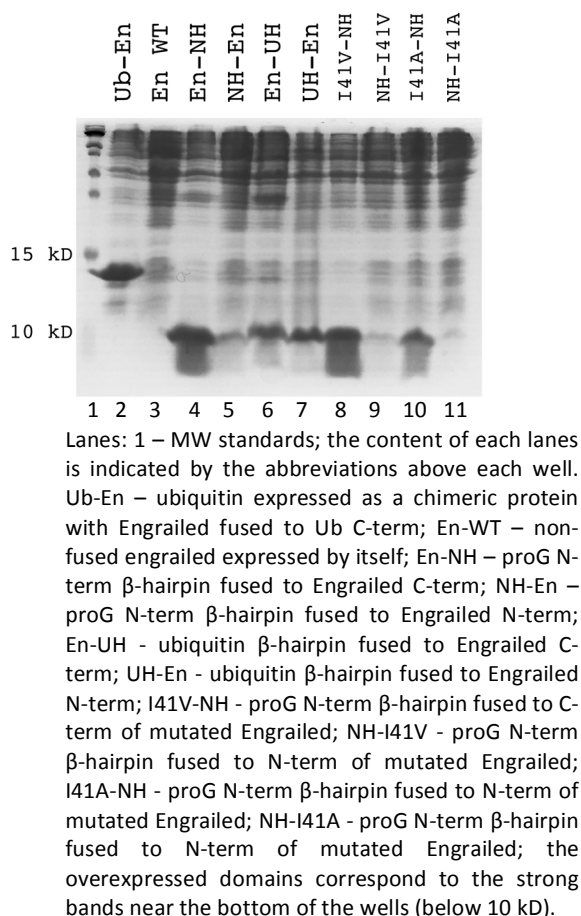
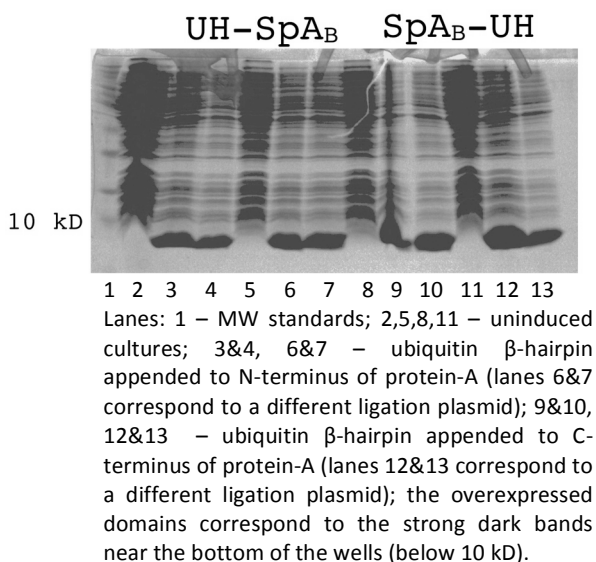


Structures of SpA_B (upper left), Engrailed domain (upper right), protein-G, G β 1 (lower left), ubiquitin (lower right)

Excellent expression, even at reduced T_m



We continued the analysis of the role that β -hairpin structures play in stabilizing proteins that are difficult to express in *E. coli* expression systems (T7). To this end the 17 amino acid β -hairpin from the protein ubiquitin was cloned onto both the N- and C-termini of the primarily α -helical small test protein - *Staphylococcal* protein A B domain (SpA_B), and also onto the N- and C-termini of the Engrailed domain.



The β-hairpin from ubiquitin was chosen as an N- and C-terminal tag to protein-A (SpA_B) because of the proven ability of ubiquitin C-terminal fusions to increase expression level of poorly expressed proteins and peptides (lane 2, gel image above right). As illustrated in the above gel image, by solely appending the 17 amino acid β-hairpin from ubiquitin, the expression levels for the Engrailed domain increase markedly (compare the poor expression shown in lane 3 of gel on the right to lanes 3, 4, 6, 7, 9, 10, 12, and 13 of the gel on the left).

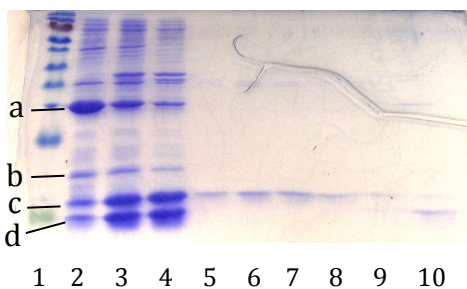
To further explore the ability of these small, stable secondary structure elements to rescue, or increase the expression levels of proteins that express poorly we have appended the N-terminal β-hairpin of protein-G to the N- and C-termini of destabilized Engrailed domains. Mutations to the core of Engrailed (I41V, I41A) were introduced with the assumption that the reduction in the size of hydrophobic amino acids in the core of the domain would result in lower thermal stability and thus lower expression levels. For these two mutations it can be seen in the above gel (right) that when the protein-G N-terminal β-hairpin is appended to the C-terminus of engrailed, expression levels are relatively high (lanes 8 and 10), whereas there appears to be only a very slight increase in expression of the destabilized variants when the protein-G N-terminal β-hairpin is appended to the N-terminus of engrailed (lanes 9 and 11).

Biological Significance: The results depicted above strongly illustrate the ability of relatively small, stable secondary structure elements to increase expression levels for proteins that are recalcitrant to bacterial expression. This raises the following questions – Why and how? Or more precisely – What is the molecular mechanism by which these secondary structure elements confer higher levels of protein expression in the cytosol of bacteria? We

presume that, since the mechanism for protein production is essentially identical for the protein with and without β -hairpins, the change in expression levels (increase in expression) results from a change (reduction) in the degradation of the peptide by host protein recycling machinery. Prokaryotes and eukaryotes have protein recycling systems (*i.e.*, proteasomes) that function to degrade and recycle damaged proteins or proteins that have exceeded their functional needs. What are the molecular recognition elements that enable cytosolic degradation of proteins destined to the proteasome? Answers to these questions are understandably highly complex. To address these issues we focused on a fairly well studied degradation pathway in *E. coli* known as the N-end rule and the ClpS/ClpAP degradation system. The ClpS adaptor delivers N-end rule substrates to ClpAP, which is an energy-dependent AAA+ protease that functions to degrade targeted proteins.

The *E. coli* BL21 (DE3) strain used for over expression is deficient in Lon, an ATP-dependent serine protease. ClpAP and ClpXP are the remaining ATP-dependent proteases. ClpX partners with ClpP for degradation of ssrA-tagged substrates, which are generated during a ribosomal stall. The fact that the β -hairpin chimeras express well, and contain the same mRNA sequence as the proteins without the β -hairpins, would seem to indicate that the ribosome likely does not stall during translation of our sequences. Additionally, the DNA is optimized to prevent rare codon use and clusters, as well as detrimental mRNA secondary structure, both of which could lead to ribosomal stalling or poor expression. This leaves the ClpAP protease, which acts in conjunction with the ClpS adapter protein to find and process proteins or peptides with unfolded segments or that have N-end rule residues at their amino terminus. The ability of the poorly expressed wild-type variants of our test proteins to bind to various elements of the ClpAP system was experimentally tested. This was contrasted against the ability of the test proteins fused to β -hairpins to physically interact with the elements of the degradation system. To this end we cloned, expressed, and purified different elements of the ClpS/ClpAP system.

ClpS was cloned as a ubiquitin fusion protein, Ub-ClpS. This protein is his-tagged and was purified using Nickel affinity chromatography. ClpS is separated from ubiquitin by processing with the hydrolytic enzyme UCH-L3 followed by a second round of chromatography, where the protein is collected in the flow-through.



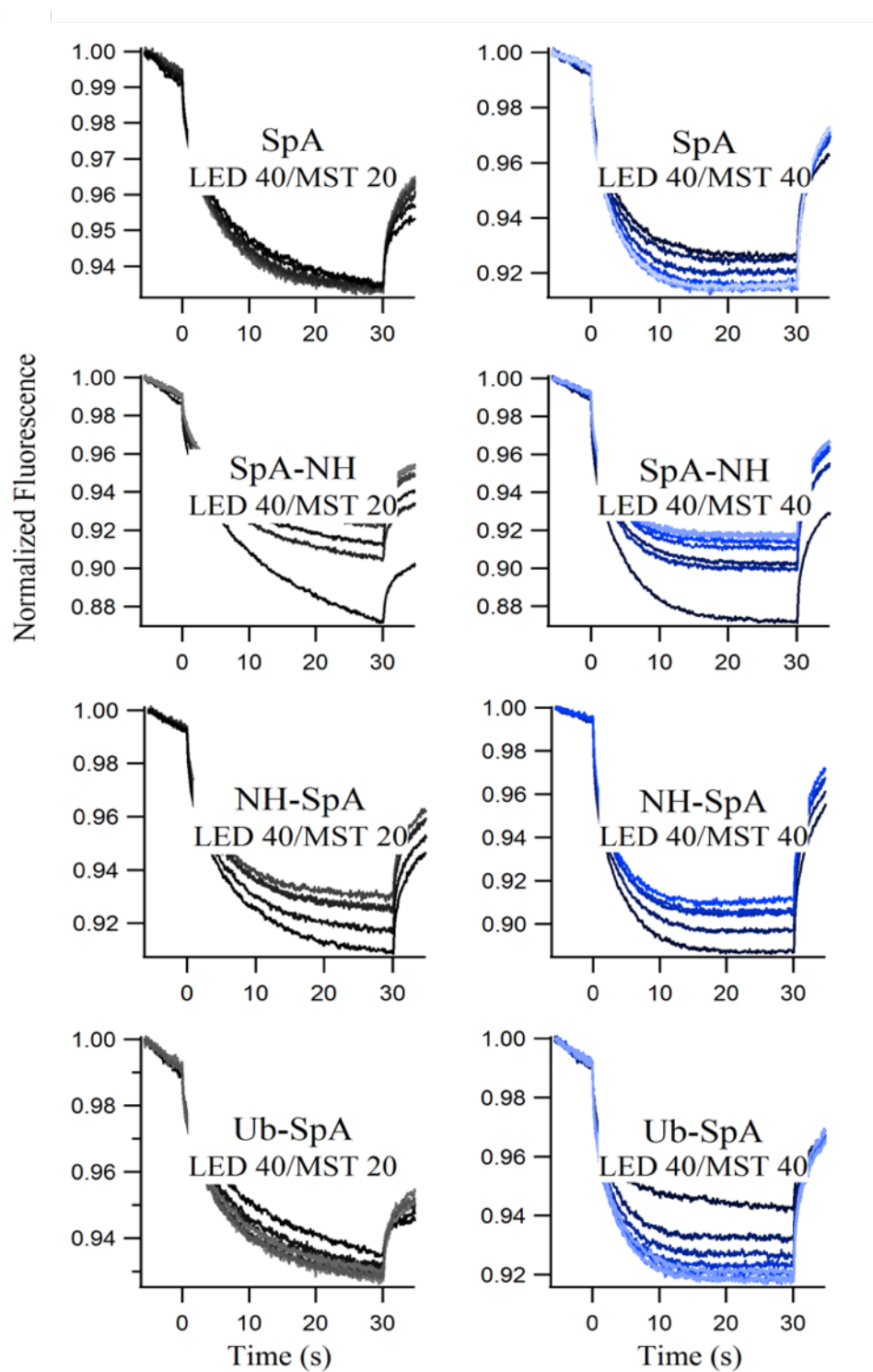
Cleavage of the Ub-ClpS construct with the enzyme Ubiquitin C-Terminal Hydrolase (UCH-L3) and subsequent nickel column purification. a) UCH-L3 b) Ub-ClpS fusion construct c) ClpS d) ubiquitin. Lane 1 – MW standards; lanes 2, 3, 4 – UCH-L3 cleavage reactions using three different sources of enzymes. Lanes 5 thru 10 – FPLC chromatography of different elution fractions off of a nickel affinity column.

ClpA (84 kDa) was cloned into pET21a alone and was purified using cation exchange chromatography. ClpA overexpresses well, but initial efforts to purify ClpA were unsuccessful. A sodium acetate buffer was initially used to resuspend the protein and perform the chromatography on a HiTrap SPHP column but ClpA appears to precipitate somewhat. ClpP (24 kDa) was cloned into pET21a with a C-terminal histidine tag. ClpP-6H overexpresses well, but buffer conditions for purification had to be optimized. A 10 %

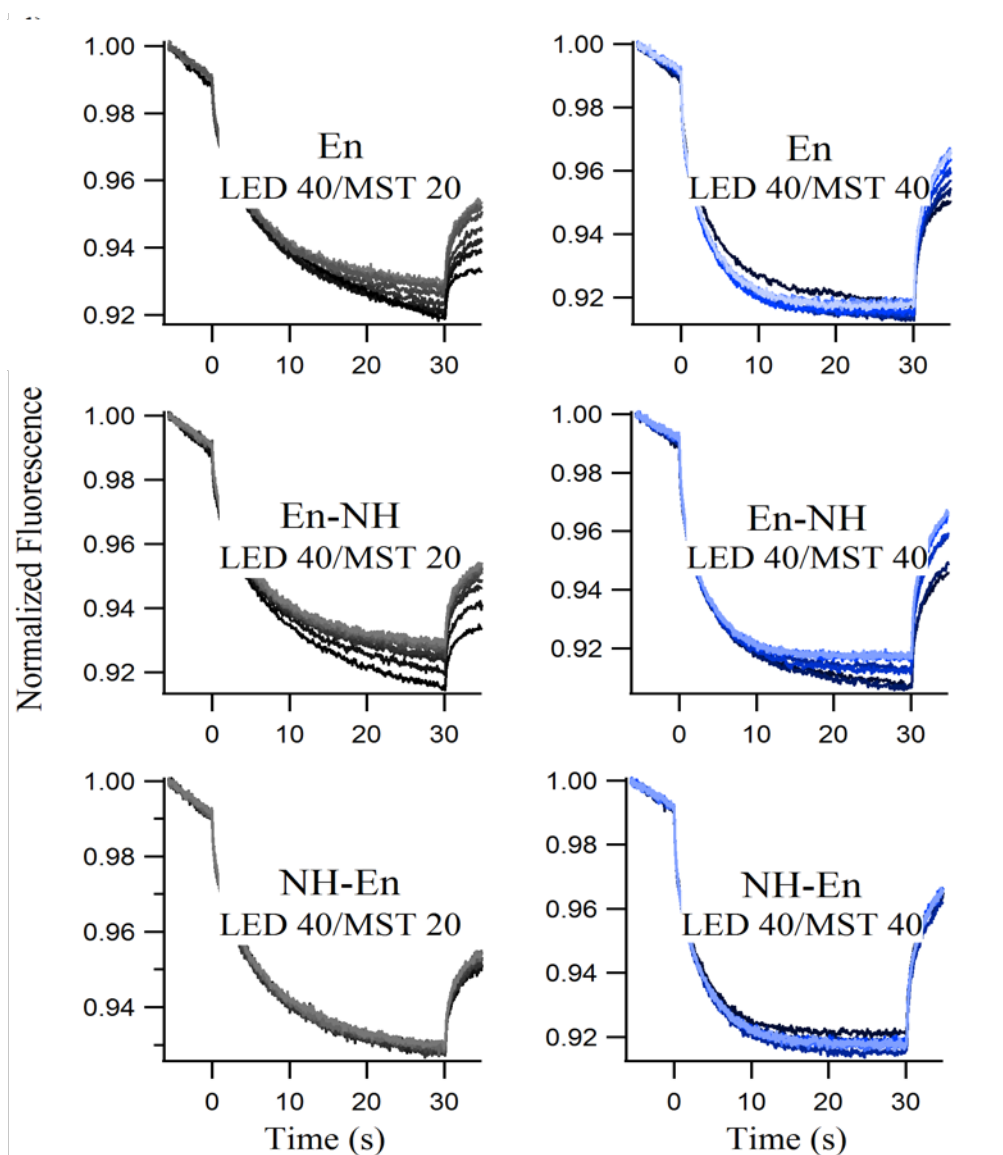
glycerol solution was included in the resuspension and chromatography buffers, and this may have interfered with the binding of ClpP to the column.

MicroScale Thermophoresis (MST) Analysis of ClpS/protein-A Interactions: A possible hypothesis as to why appended β -hairpins elements increase bacterial expression for poorly expressed proteins is because they function to reduce essential protein/protein interactions between the overexpressed proteins and the cellular degradation machinery. To test this hypothesis we developed a system in which variants of protein-A (*i.e.*, wild-type SpA, NH-SpA, SpA-NH, Ub-SpA) were tested for their ability to physically interact with a fluorescently labeled variant of ClpS. The physical protein/protein interactions between these protein-A variants and the fluorescently labeled variant of ClpS were measured using MicroScale Thermophoresis (MST).

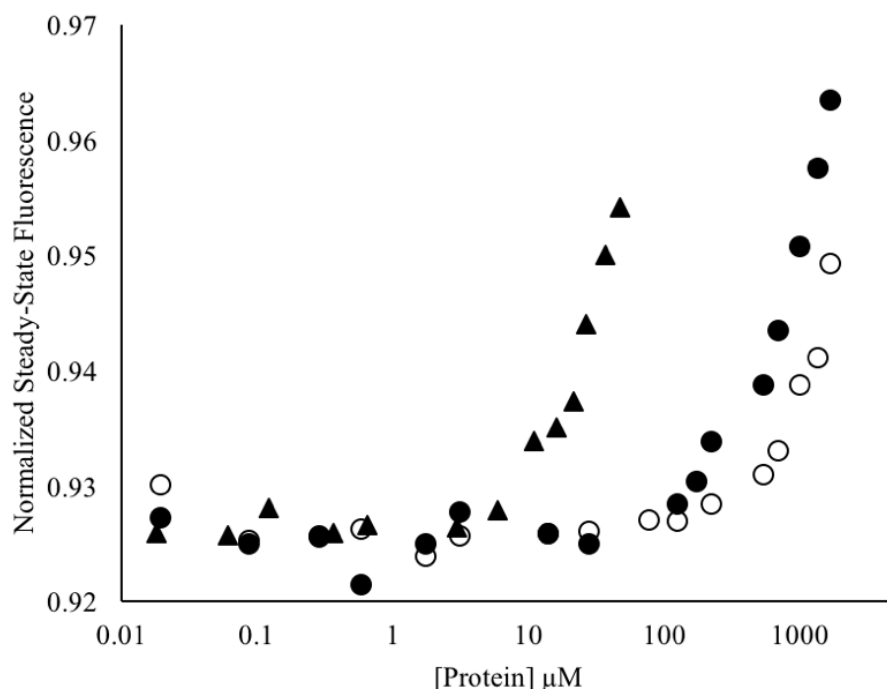
The fluorescently labeled *E. coli* ClpS variant was first engineered to remove the N-terminal extension (residues 1-26) to improve solubility. A surface cysteine was introduced via the mutation E79C using the NEB Q5® site-directed mutagenesis kit. The construct, His-Ub-ClpS(E79C) Δ NTE, was expressed in *E. coli* BL21 (DE3) and purified by immobilized metal affinity chromatography on Ni-TED resin. The N-terminal His-Ub- portion was cleaved using recombinant UCH-L3 enzyme. The resulting ClpS(E79C) Δ NTE variant was concentrated and exchanged into F5M labeling buffer. Fluorescein-5-maleimide was reacted with the purified ClpS(E79C) Δ NTE variant. The buffer used for the MST experiments contained HEPES, NaCl, MgCl₂, and Tween 20. Standard treated NanoTemper MST capillaries were loaded with the reactions and the fluorescence intensity and thermophoresis were measured as a function of time. The generated raw data and associated binding curves are illustrated on the following pages.



Normalized fluorescent time traces of indicated SpA variants titrated against 20 nM fluorescein-labeled ClpS(E79C)- Δ NTE at infrared laser powers of 20 (left) and 40% (right). Gradient colors reflect concentration of titrant.



Normalized fluorescent time traces of indicated En variants titrated against 20 nM fluorescein-labeled ClpS(E79C)- Δ NTE at infrared laser powers of 20 (left) and 40% (right). Gradient colors reflect concentration of titrant



Normalized fluorescence at steady state thermophoresis plotted against concentration of SpA (triangles), SpA-NH (closed circles), and NH-SpA (open circles).

From the preliminary data illustrated above, the MST results demonstrate that the binding affinity of the different protein-A variants (SpA-NH, NH-SpA) to ClpS is different and the presence of the β -hairpin elements appended to the N- and C-termini of protein-A causes a reduction in binding. The binding constant for wild-type protein A (triangles) appears to be in the range of 10-50 μM whereas the binding constant for the two variants with β -hairpin elements appended to the N- and C-termini (open and closed circles) is in the range of 900-1000 μM . This represents a reduction in the binding affinity of approximately 100 fold.

These results imply that the increase in expression observed for the variants that contain the β -hairpin elements may be due to the fact that a key element of the degradation machinery, ClpS, cannot bind these proteins as well as it binds the wild-type variant. This may be due to the fact that the hydrogen bond donor and acceptor atoms in the β -hairpin elements are already occupied and thus not available for binding and recognition by elements of ClpS. Essentially, the β -hairpin containing variants do not provide a convenient 'handle' for ClpS to effectively bind to.

These initial results provide promising evidence that appending β -hairpin elements to the N- and C-termini of protein-A reduces the physical interactions between ClpS and these engineered proteins. Therefore these engineered proteins may act as effective probes of the degradation machinery within cells. The series of MST experiments described above correspond to the first set experiments in which useful binding data was generated for the interactions between ClpS and the β -hairpin containing engineered proteins. Therefore these experiments will be repeated using additional controls and varied conditions in order to further confirm the above-described hypothesis.

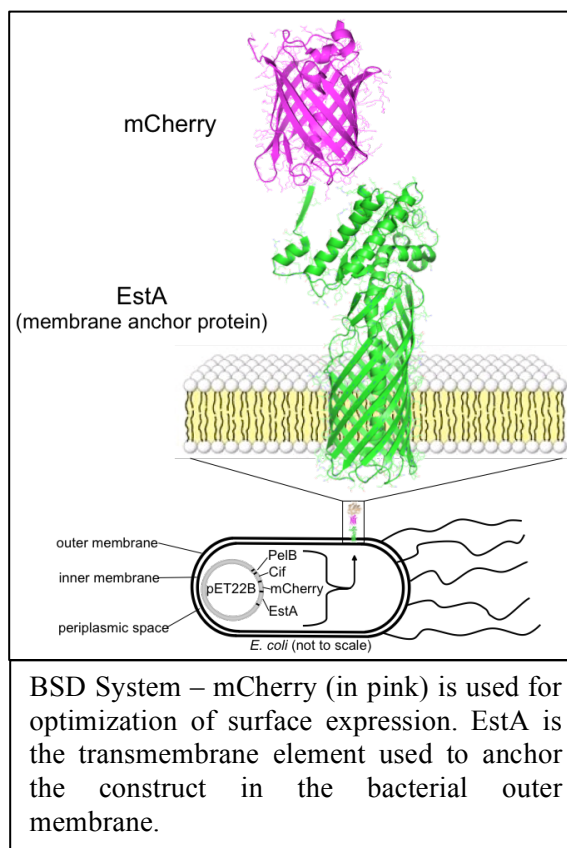
4) A Fluorescence-Based Bacterial Display System Enhanced for High-Level Protein Surface Expression

As detailed above a major focus of the Love Laboratory is protein design and engineering, in particular the design of novel protein/protein interfaces. Additional goals include the development and implementation of protein design tools used to explore biophysical aspects of protein biochemistry such as the *in vivo* roles that thermal stability plays in protein function, and the unique functions of specific enzymes.

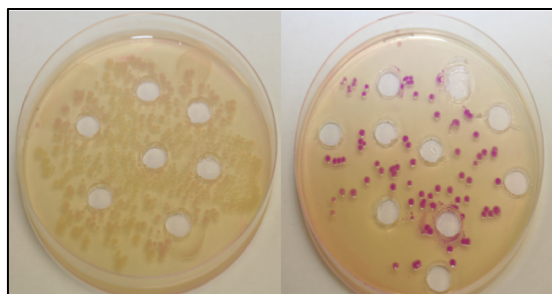
We have enhanced our protein design tools through the engineering of a fluorescence-based bacterial surface display (BSD) system. Standard protein expression systems enable the production of large quantities of natural and synthetic proteins within the cytosol of *E. coli*. To isolate expressed proteins the bacterial cells must be ruptured and the desired protein purified away from the estimated 4000 bacterial proteins. During this process significant care must be taken to insure that the desired protein does not degrade. This is especially important for protein enzymes as deleterious changes in solution conditions can easily degrade the catalytic properties of the expressed enzyme. The BSD system developed in our laboratory streamlines protein expression and eliminates many of the above concerns.

This BSD system (image at right) was developed using readily available protein components that were combined to engineer a red-fluorescence based display system using *E. coli* as the host organism. Fluorescence enables straightforward expression optimization and results in high expression levels of the displayed chimeric construct. In brief, the system is composed of series of proteins that are expressed from a single gene. Beginning at the N-terminus, the PelB signal sequence is used for delivery of the fusion construct to the bacterial periplasmic space. Next is the red-fluorescent protein mCherry for enhanced expression optimization, and finally a truncated variant of the autotransporter EstA from *Pseudomonas aeruginosa*. EstA enables the proper insertion, anchoring, and orientation of the expressed multi-domain construct in the outer membrane of *E. coli*. To display a protein-of-interest (POI) on the surface of *E. coli* its gene is inserted between the genes for the PelB leader sequence and mCherry. The PelB leader peptide is ultimately cleaved resulting in full N-terminal display of the intact POI.

Especially related to this project, this system is ideal for enhanced production and assessment of novel protein/protein interactions. The BSD system is analogous to phage and ribosome display in that it enables the display of proteins on a surface in combination with direct connection to the sequence of the protein, which is encoded by the DNA plasmid

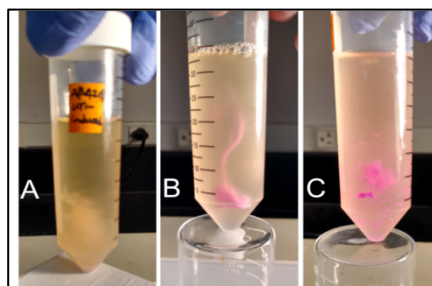


contained within the displaying bacteria. This enables the potential selection of a particular protein variant, from an engineered library of displayed proteins, which has the desired property. The property could be, for instance, a unique catalytic property or, pertinent to this project, the ability of the displayed protein to bind to a specific target protein. The advantage of this system over phage and ribosome display is the simplicity associated with cloning a single gene into the system as the genes for the resulting chimera (described above) are all contained on one standard bacterial plasmid. Since the BSD system results in the fusion protein being directly expressed on its surface, the need for multiple plasmids (or helper phage), or complicated multi-step processes (needed for ribosome display) the system enables rapid cloning and assessment in a relatively short period of time. Thus far we have 1) fully optimized and tested the system, 2) cloned a series of test proteins into the system to benchmark the system, and 3) cloned into the system the gene for the metal-mediated dimer described in section one above (Figure 1, page 1).



Expression Test of the BSD System.
Left plate – standard BL21(DE3) cells induced with IPTG. Right plate – BSD expressing cells induced with IPTG.

Below are additional examples of active enzymes/proteins/peptides that we have thus far successfully expressed on the surface of *E. coli* using the BSD system. The gene for the 42 amino acid Amyloid- β peptide, which is a key molecular component of the plaques found in brains of Alzheimer's patients, was cloned into the BSD system. Upon induction these cells have a distinct pink/red color and after approximately 10-12 hours the cells significantly aggregate due to expression of the strongly self-associating A β -42 peptide displayed on the bacterial surface (panel C, image at right). We have engineered this system to demonstrate the ability of surface expressed proteins to bind one another and induce cellular aggregation. We have also cloned the gene for α -synuclein, which is the aggregated protein found within Lewy bodies of Parkinson patients, into the BSD system and also tested its ability to cause bacteria to self-associate. In addition we have cloned an epoxide hydrolase enzyme from *P. aeruginosa* into the system. This enzyme, known as the CFTR inhibitory factor (Cif), promotes the ubiquitin-mediated degradation of CFTR and thus is able to emulate cystic fibrosis at the cellular level. The gene for Cif has been cloned into the BSD system and the expressed enzyme is fully active on the surface of *E. coli*. We plan to re-engineer the active site of this enzyme such that it can hydrolyze and degrade molecules used as chemical weapons such as the mycotoxin deoxynivalenol (AKA – vomitoxin). The BSD system will be combined with directed evolution methods to produce a novel enzyme that will degrade mycotoxins and potentially protect American service people deployed in warzones and hazardous regions around the globe.



BSD Expression of aggregation prone proteins. (A) A β -42 uninduced (B) α -synuclein induced, (C) A β -42 induced.

The research associated with the preliminary test of the BSD system is 95% complete and we are in the process of preparing a manuscript that fully describes the design, implementation, and verification of the efficacy of the system. Future goals associated with this overall project include the engineering of a heterodimeric metal-mediated dimer complex. To ascertain if the BSD system can be used for directed evolution of such a complex we have cloned the gene for the metal-mediated homodimer complex into the BSD system. This successfully resulted in the construct being displayed on the outer surface of *E. coli*. Upon expression and exposure to 1 mM zinc the cells were shown to self-associate in a manner similar to that observed for the cells that displayed the A β -42 peptide (figure previous page, panel C).

For the goal of using directed evolution to engineer a metal-mediated heterodimer the two different protein components we be generated as follows. A unique set of amino acids will be incorporated at the model interface for one of the binding partners. This variant will be expressed by itself and contain a short C-terminal linker sequence and a terminal cysteine residue. After expression and purification this variant will be labeled with a fluorescent moiety such as fluorescein (using the same method as for the fluorescent anisotropy and MST experiments described above). For the second binding partner, its gene will be incorporated into the BSD system except that during the PCR-based synthesis of this variant, codons associated with amino acid positions at the putative interface will be randomized to generate a library of variants that differ at interfacial positions. This will result in a library of proteins, displayed on the BSD system, that are structurally the same yet differ at positions at the targeted interface. After successful surface expression of the library is verified the fluorescently labeled binding partner will be incubated with the BSD cells and successful binding will be assessed using Fluorescent Activated Cell Sorting (FACS). Binding of the fluorescently labeled binding partner will impart green fluorescence on these BSD cells, and therefore variants that contain optimal amino acids for driving complex formation will be selected using FACS sorting. As described above, we have already incorporated one of the rationally designed metal-mediated dimers (Figure 1, page 1) into the BSD system and have verified binding via cell self-association. We are currently in the process of testing the binding affinity using MST and a fluorescein labeled variant (as described above). We plan to use this fluorescein labeled variant to test and verify the ability to detect binding using FACS sorting. Once these tests are complete we intend to use the BSD system for directed evolution of metal-mediated dimers using the success we have demonstrated with the protein-G scaffold. Once successful heterodimers are generated we will then attempt to use other protein scaffolds to generate small antibody like proteins that bind specific target proteins.

- (1) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5772–5777 DOI: 10.1073/pnas.1030237100.
- (2) Nooren, I. M. A.; Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **2003**, *22*, 3486–3492 DOI: 10.1093/emboj/cdg359.
- (3) Jones, S.; Thornton, J. M. Protein-protein interactions: A review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **1995**, *63*, 31–65 DOI: 10.1016/0079-6107(94)00008-W.
- (4) Gromiha, M. M.; Yugandhar, K.; Jemimah, S. Protein–protein interactions: scoring schemes and binding affinity. *Curr. Opin. Struct. Biol.* **2017**, *44*, 31–38 DOI: 10.1016/J.SBI.2016.10.016.
- (5) Shukla, U. J.; Marino, H.; Huang, P.-S.; Mayo, S. L.; Love, J. J. A designed protein interface that blocks fibril formation. *J. Am. Chem. Soc.* **2004**, *126* DOI: 10.1021/ja0456858.
- (6) Huang, P.-S.; Love, J. J.; Mayo, S. L. Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J. Comput. Chem.* **2005**, *26* DOI: 10.1002/jcc.20252.
- (7) Huang, P.-S.; Love, J. J.; Mayo, S. L. A de novo designed protein-protein interface. *Protein Sci.* **2007**, *16* DOI: 10.1110/ps.073125207.
- (8) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **1994**, *33*, 4721–4729.
- (9) Thoms, S.; Max, K. E. A.; Wunderlich, M.; Jacso, T.; Lilie, H.; Reif, B.; Heinemann, U.; Schmid, F. X. Dimer Formation of a Stabilized Gβ1 Variant: A Structural and Energetic Analysis. *J. Mol. Biol.* **2009**, *391*, 918–932 DOI: 10.1016/j.jmb.2009.06.031.
- (10) Lu, T.; Park, J. Y.; Parnell, K.; Fox, L. K.; McGuire, M. A. Characterization of fatty acid modifying enzyme activity in staphylococcal mastitis isolates and other bacteria. *BMC Res. Notes* **2012**, *5*, 323 DOI: 10.1186/1756-0500-5-323.
- (11) Mirani, Z. A.; Naz, S.; Khan, F.; Aziz, M.; Asadullah; Khan, M. N.; Khan, S. I. Antibacterial fatty acids destabilize hydrophobic and multicellular aggregates of biofilm in *S. aureus*. *J. Antibiot. (Tokyo)*. **2017**, *70*, 115–121 DOI: 10.1038/ja.2016.76.
- (12) Saylor, B. D.; Love, J. J. A secreted *Staphylococcus aureus* lipase engineered for enhanced alcohol affinity for fatty acid esterification. *J. Mol. Catal. B Enzym.* **2016**, *133*, 44–52 DOI: 10.1016/j.molcatb.2016.11.013.
- (13) Barakat, N. H.; Barakat, N. H.; Carmody, L. J.; Love, J. J. Exploiting elements of transcriptional machinery to enhance protein stability. *J. Mol. Biol.* **2007**, *366*, 103–116.
- (14) Barakat, N. H.; Love, J. J. Molecular diversity in engineered protein libraries. *Curr. Opin. Chem. Biol.* **2007**, *11*, 335–341.