Reflective Statement Summary: I arrived at Michigan State University in August 2011 with a vision, and ***long-term research goal***, to design, engineer, and build the biomolecular components of life *de novo*. This overarching goal not only satisfies one of the grand challenges of biology – "is our understanding of living matter sufficient to redesign cellular life?" - but also has implications in several National Academy of Engineering grand challenges in the 21st century, including engineering better medicines and engineering the tools of scientific discovery. As a ***near-term research goal***, my group is building on our vision of engineering biology by developing cutting edge methods to map protein sequence to function for medical and industrial biotechnology. The essay below reflects my progress in sharpening the focus of this vision as instantiated by my research program, while also detailing the impact of my presence on campus through my integration of research, teaching, and outreach.

Research summary: I have a strong foundation of research excellence in my independent laboratory as shown by several metrics. I am or have been a PI (12) or co-PI (2) on **14 funded projects, including an NSF CAREER** award (my contribution $1.4 million; 8 federally-funded projects). I have submitted or published **34 peer-reviewed journal articles in total, with 20 since arriving at MSU including 16 as corresponding author** (>1030 total citations, h-index 15 by Google Scholar). I have been invited for 10 talks at workshops and conferences[1], including the prestigious 2017 Young Scientist Keynote Presentation at PEGS. My pre- and postdoctoral students have won DOE, USDA, and NIH graduate and postdoctoral fellowships and multiple College and International conference presentation awards. I have graduated two PhD students, with one employed as an antibody engineer/scientist at AdiMab and another as a postdoctoral researcher in the Chem. Eng. Dept. at U. Minnesota. A former postdoc is now a research professor at U. Oregon, and my current lab comprises 4 graduate students.

Teaching and Service summary: I have developed and taught a new Synthetic Biology graduate course along with several undergraduate Chemical Engineering courses. I have participated in several experiential research experiences for undergraduates[2] including founding and leading the first iGEM synthetic biology team on the MSU campus. My service encompasses K-12 through the professional level, including (externally) co-chairing the 2014 International RosettaCon[3] and being an executive committee member for an MSU NIH T32 training grant.

Research Activities: Proteins are central to a diverse set of fundamental activities necessary for cellular life. This same versatility leads to tremendous potential for proteins as designable agents in medicine (e.g. monoclonal antibodies like Rituximab that can target and destroy non-Hodgkin lymphomas) and in industry (e.g. enzymes that can deconstruct cellulosic biomass to simple carbohydrates). Yet natural proteins are not always optimal for such applications, and current methods of improving proteins can be expensive and laborious. My ***medium-term goal*** (10-20 years) is to design or engineer proteins for new/enhanced functions at will. My ***short-term goals*** (3-8 years) are to develop methods in support of our medium- and long-term goals.

My current research interests lie in three overlapping yet distinct avenues: (I.) design and engineering of proteins as diagnostics, therapeutics, and vaccines; (II.) design and engineering of enzymes to deconstruct renewable biomass to value-added fuels/chemicals; and (III.) new methods development to enable large-scale and comprehensive analysis of protein libraries.

(I.) Design and engineering of proteins as diagnostics, therapeutics, and vaccines.
Proteins that bind other proteins (like antibodies) can be used as diagnostic probes, therapeutics, or prophylactics. While I was a postdoc, I developed small proteins that broadly neutralize Influenza group I viruses by targeting a conserved stem region of Hemagglutinin (HA). I was able to determine the binding and stability of nearly every single point mutant of these yeast-displayed protein-binding variants by coupling deep sequencing to flow cytometry-activated sorting. I integrated these sequence-function maps with computational protein design to improve the affinity, specificity, and function of these proteins[4].

We realized that the same strategy of comprehensively mapping a protein sequence to its function could be used to identify conformational epitopes for antibody-antigen interactions. Our recent papers demonstrate that this conformational epitope mapping strategy, confirmed using antibody panels against diverse antigens, is much more powerful than competing methods[5]. We have also used deep sequencing to quantitatively determine the effects of binding affinity upon mutation for nearly every possible single point mutant in two different protein sequences[6]. Combined, these methods allow the routine determination of affinity and specificity for all single point mutants for a given protein-protein binder in a massively parallel fashion, thus enabling more efficient engineering of proteins as therapies and vaccines. As one striking supporting example, a recent report detailed use of our methodological advances to design and engineer an immunogen for a potential HIV vaccine[7].

We are currently applying these methods to develop specific diagnostic reagents for Zika and related Flaviviruses, to map neutralizing antibodies for Zika and Dengue in order to create pan-Flavivirus prophylactics and vaccines, and to map epitopes for therapeutic antibodies for multiple indications. NIH, NSF, and industrial sponsors fund this work.

(II.) Enzyme engineering by deep sequencing and computational design.
We have extended high-resolution sequence function mapping to evaluate the comprehensive sequence determinants to pathway flux in a pyrolysis oil catabolic pathway in bacteria[8]. We identified the function of over 8,000 single point mutants in a single experiment and found that most mutations that improved pathway productivity worked by improving the active concentration of enzyme *in vivo*. Nevertheless, of the hundreds of mutations that improved function over 50 increased the catalytic efficiency of a key catabolic enzyme. We then integrated deep sequencing with computational design to support a 15-fold improvement in growth rate for bacteria grown on a pyrolysis oil anhydrosugar as the sole carbon source. We have also used this deep sequencing pipeline to understand the comprehensive determinants to substrate specificity for an enzyme, revealing that globally beneficial mutations are rare and that enzyme specificity is globally encoded in protein sequence and structure space[9].

We remain interested in applying *computational design* to uncover rational design rules for biomolecular engineering. As one example, lignocellulosic biomass can be converted to biofuels

[redacted]

by enzymatic deconstruction of polysaccharide to fermentable sugars, followed by microbial fermentation. However, the enzymes (cellulases) needed in this process are expensive in part because they bind non-productively to and inactivate in the presence of lignin. To understand the basis of this protein-lignin inactivation, we used computation to design proteins a wide range of hydrophobicity, net charge, and charge density. We found that negative net charge is the single largest determinant to protein-lignin binding, and we are currently using that information to redesign more stable, active, and cheaper cellulases[10]. As another example, we computationally redesigned bacterial outer membrane proteins to reveal thermodynamic design principles behind membrane protein assembly and folding[11].

(III.) Improving deep sequencing methods for protein engineering and biotechnology. Methodological advances in deep sequencing developed by my laboratory, along with other labs worldwide, facilitated the above examples. With commercial next generation (deep) sequencers it is now possible to read sequences of a million DNA nucleotides for pennies, enabling one to sequence entire populations of similar biomolecules before and after a selection for function. The frequency change of each member of a population can be converted to a relative function. However, before integrating this technology with protein engineering there were a number of technical challenges that needed to be overcome. To that end, my group invented new methods to allow the technique to be performed easier for full-length proteins (on the order of 400 amino acids), improved several steps to decrease costs, and developed rigorous analytical normalization equations to correlate sequencing counts to binding dissociation constants (for protein binders) or relative growth rates (for enzymes)[12]. Additionally, a robust and accessible method for the construction of high quality, user-defined mutational libraries was lacking. Commonly used mutagenesis methods such as error-prone PCR suffer from limited codon sampling and imprecise control over the number of mutations introduced. To solve this challenge, we invented a new mutagenesis method to allow user-defined, comprehensive mutagenesis libraries from routinely prepped plasmid dsDNA in a single day and single pot[13].

Finally, we overcame a technical challenge for counting the frequency of library members in a population. In our specific applications we need the ability to identify sequence variants with accuracy. However, the short-read assembly paradigm currently dominates genomics, and the loss of linkage information during the generation of short reads limits their utility. We developed a library preparation method that enable long "synthetic" reads up to 11.6 kilobases in length to be constructed from conventional short (150-bp) reads, providing a general platform for synthetic read generation from a wide range of input nucleic acid types. We demonstrated that this method could resolve multiple splice junctions of individual RNA molecules, differentiate between distinct HIV *Env* variants, and improve the genome assembly of different organisms[14]. Based on the totality of my current research program ***I am a recognized expert*** in integrating deep sequencing with protein engineering and design[15].

Future Research Program. My research group is applying our core technologies to a number of pressing applications. For example, we are beginning work on the structure-based design of

effective vaccines for human and livestock infectious diseases (like Zika), as immunogen design is at heart a protein engineering challenge. We are also excited about possibilities in remodeling protein-protein interactions in areas like T-cell immunotherapy and rational modification of the gut microbiome. In the area of industrial biotechnology, we anticipate extending our techniques to optimize entire synthetic metabolic pathways concurrently. For example, we are working on general approaches to simultaneously construct highly efficient, active enzymes for *in vivo* biomanufacturing using the recently discovered tropane alkaloid biosynthetic pathway as a model system. As another example, we are rationally designing complete cellulase and hemicellulase cocktails to saccharify commercially relevant pretreated lignocellulosic biomass.

Teaching

My teaching duties range from advising students[16] and directing student research in my laboratory[17] to more formal class-based instructions.

Undergraduate instruction: I have taught a Senior-level Biochemical Engineering Laboratory (CHE481) and CHE 201, a Sophomore-level Material and Energy Balances course. To prepare I sat in on the previous instructor's lectures[18] and have continually taken workshops in improving instructional course objectives[19]. Using student performance on the on-line homework, I can better evaluate how well students master the learning objectives set for the course. For example, I noticed that there were student deficiencies in drawing process flow diagrams and setting up degree of freedom analyses. Thus, I restructured the course to include written homework sets in which students had to explicitly set up process flow diagrams. Perhaps because of the above efforts efforts, I have received excellent teaching evaluations for the sections I have taught thus far with one exception[20,21]. To correct this subpar evaluation (occurring FS2015) I took several Faculty& Organizational Development classes in the Spring semester and retooled my class to focus on engineering fundmentals.

Graduate instruction: I developed and taught a new graduate-level course on Synthetic Biology in SS2015[22]. The next generation of practicing biochemical engineers needs to be familiar with modern tools of genetic engineering as it applies to microbial strain development, genome editing, antibody engineering, and industrial biotechnology; this course is intended to provide a hands-on overview. This course was needed, as no course on bio-molecular engineering currently exists in either curriculum. We enrolled undergraduate and graduate students from Engineering, Molecular Biology, Zoology, and Microbiology[23]. There were two seminar-style lectures per week and one class period per week involved a student-led journal club describing 1-2 key papers describing the main advances of the field. Key enabling technologies were described including methods for genome assembly, genome editing like CRISPR-Cas, and

protein and RNA engineering and design using tools like Rosetta. For the capstone project, students developed and proposed a novel use of synthetic biology (akin to the iGEM competition).

Outreach & Service

External service. I have also participated in service to professional organizations off-campus. I have been an ad hoc or panel reviewer for NSF 5 times, have served on the Zika special emphasis panel at the NIH, have been a reviewer for over a dozen journals[24], and have served as chair or co-chair for AIChE and ACS annual meetings since 2013. I am also a member of the Rosetta macromolecular software community, and as a member have co-organized the annual RosettaCON meeting, was a co-editor of bi-annual Rosetta Special Collection in PLoS journals[25], and currently serve on the executive committee.

MSU service. I have structured my outreach & service responsibilities in order to enrich and enhance the culture of research excellence at MSU. To increase retention of STEM students I have utilized existing MSU programs to place high school and undergraduate students in research projects in my lab and have prepared and taught a series of workshops on biochemical engineering fundamentals for high school students[26]. I have delivered over 20 lectures to high school and freshmen students and am an active connector faculty member to freshmen engineers. Finally, I have recently won funding for and formed the first MSU team for the international genetically engineered machines (iGEM) competition, in which undergraduates perform self-guided research projects and present results at an international meeting[27].

Plant Biology is one of the crown jewels of the MSU research portfolio, and I have worked to strengthen their ties to Engineering. To that end, I have also been involved with assorted Plant Biology faculty members in writing and winning an NIH Training Grant proposal to support Plant Biotechnology training for pre-doctoral students[28], have been a co-organizer of three annual symposiums on Plant Biotechnology held between 2012-2016 on the MSU campus[29], and am the engineering representative on the executive committee. Additionally, I have been serving as the CHEMS representative on the Engineering Research Committee since 2013.

Final statements

Successful scholars integrate their scholarship across the multiple missions of the university – teaching, research, and outreach & service. As detailed above, I have interweaved my scholarship into my formal teaching and outreach in order to enrich the culture of research and academic excellence on campus.

My research program involves big, ambitious projects that are expected to significantly advance scientific and engineering knowledge of constructing biological components of life *de novo*. From our initial success, my peers understand the groundbreaking ***impact*** of several of our ideas, and I am grateful to have the freedom thus far to pursue my vision.