**Doctoral Dissertation and Research Experience**

**Predoctoral research**

**1 Neitz lab**

Mentor Dr. Jay Neitz

Role Summer intern, Neitz lab, University of Washington medical school, 2011-2012

Description My first research experiences were summer internships in the Neitz lab during high school. I worked with postdoctoral fellows and graduate students in the lab to validate a mouse model for introduction of new photo-receptors in adult animals to study color vision neural circuitry function. As part of this project I worked on a color discrimination test for mice, took confocal images of mouse retinas, and measured electroretinograms of mice after pilot experiments. I presented this work at two local competitions and a regional competition.

**Desmarais, J. J.** Photopigment expression and function using targeted knock-in/knock-out mice: Intravitreal injections. Junior Science & Humanities Symposia, Westchester; 2012; Katonah, NY.

**Desmarais, J. J.** Photopigment expression and function using targeted knock-in/knock-out mice: Intravitreal injections. Junior Science & Humanities Symposia, Upstate New York; 2012; Albany, NY.

**Desmarais, J. J.** Photopigment expression and function using targeted knock-in/knock-out mice: Intravitreal injections. Westchester Science and Engineering Fair; 2012; Tarrytown, NY.

**2 STEM innovation program**

Mentor Drs. Noah Graham, Frank Swenton, Jeremy Ward

Role Researcher, STEM innovation program, Middlebury College, 2013

Description During my freshman year at college, I applied to and joined the STEM innovation program, where a group of 9 undergraduates designed and executed a synthetic biology project. We set out to create a biosensor for detecting aromatic hydrocarbons in water samples, focusing on benzene, toluene, ethylbenzene and xylene (BTEX) as they have been detected in groundwater near fracking wells. We designed and built constructs using a transcriptional regulator from P. putida, tested lyophilization/rehydration protocols, and built and tested a portable fluorescence reader for testing samples.

**3 Gibson lab**

Mentor Drs. Matthew C. Gibson, Aissam Ikmi

Role Stowers Summer Scholar, Gibson lab, Stowers Institute for Medical Research, 2014

Description During the summer after my sophomore year, I was a Stowers summer scholar in Matthew Gibson’s lab. I worked with the postdoc Dr. Aissam Ikmi, studying how the size on the embryo affects the early development of N. vectensis sea anemones. N. vectensis grows from an egg into a polyp with 4 tentacles and then begins to eat and grow. At the 4 cell stage, all 4 cells still retain their ability to produce a viable polyp. By subdividing embryos and observing their development, we hoped to observe the effects of size on developmental processes like tentacle patterning. Results Reducing embryo size also reduced polyp size, reducing length more than width, and tentacle number. Mesentery number changed with tentacle number maintaining a ratio of two mesenteries for each tentacle. Regardless of initial size, polyps grew to similar sizes before developing the first pair of additional tentacles. All together, this data suggested that tentacles are patterned in a size dependent manner.

**4 Keasling lab**

Mentor Drs. Jay Keasling, Victor Chubukov

Role Amgen Scholar, Keasling lab, Joint Bioenergy Institute, University of California, Berkeley, 2015

Description After my junior year of college, I joined the Keasling lab for the summer as an Amgen Scholar. Under the mentorship of postdoc Dr. Victor Chubukov, I worked on developing chassis strains of E. coli that could be used to improve yields in bio-production of fuels and chemicals. By developing chassis strains, we hoped to provide broadly applicable methods that applied to a variety of different target chemicals. One major issue encountered by metabolic engineers in producing chemicals is shunting carbon and energy towards growth not production. In order to avoid this growth and production can be separated by placing the cells in growth limiting conditions while inducing the production pathway. However, many chassis strains will go dormant under these conditions. We tested the hypothesis that increasing glucose uptake during nitrogen starvation would increase the amount of carbon and energy that could be directed to production. Results In nitrogen starvation, α-ketoglutarate levels rise, inhibiting the enzyme PtsI. This blocks glucose phosphorylation and therefore uptake. We trialed 3 methods to overcome this regulation. The first was over-expressing PtsI. The second was a PtsI-PtsP chimera that was hypothesized to avoid inhibition. The third was GalP and Glk a permease and kinase that take up glucose through an orthogonal pathway. We found that all of these strategies increase glucose uptake during nitrogen limitation, but PtsI over-expression is more effective than chimera over-expression, and Galp/Glk over-expression can cause cell death. The PtsI over-expression strain consumed 4x more glucose than WT during nitrogen starvation despite optical density staying constant and a lack of fermentation byproducts being secreted. This suggests that the glucose was converted all the way to CO2 by the TCA cycle. We then tested if this strategy improved yield in a fatty alcohol production experiment. We saw that while nitrogen staring the cells increased carbon use efficiency, increasing their metabolic activity with PtsI did not improve yield. We published these finding in NPJ systems and synthetic biology.

Chubukov, V., **Desmarais, J. J**., Wang, G., Chan, L. J. G., Baidoo, E. E. K., Petzold, C. J., Keasling, J. D. & Mukhopadhyay, A. Engineering glucose metabolism of Escherichia coli under nitrogen starvation. *NPJ Syst Biol Appl* **3,** 16035 (2017).

**5 Ward lab**

Mentor Dr. Jeremy Ward

Role Researcher, Ward lab, Middlebury College, 2014-2016

Description In my Junior and Senior years of College I worked in the lab of Jeremy Ward. During my Junior year, I studied meiotic crossover in Mouse spermatogenesis looking for co-localization of HEI10 and POLB in meiotic chromosome spreads. During my senior year, I worked on developing biosensors for *B. burgdorferi* the pathogen that causes Lyme disease. These biosensors were designed to activate gene expression in the presence of high expression *B. burgdorferi* mRNAs by unwinding a toehold riboswitch that occluded the biosensor’s ribosome binding site. I designed riboswitch constructs using custom scripts and RNA folding software and built designs to test in the lab. This work formed the basis of my thesis and was awarded high honors by the department.

**6 Connection to fellowship** These experiences helped me to decide to pursue a career as a scientist and launched my scientific career. These early projects exposed me to a wide variety of research areas and helped to guide me towards my eventual area of focus. These early research projects also gave me a wide breadth of experience with different techniques and organisms. Finally, they allowed me to build my skills planning and managing research projects, especially through my time running independent projects in the Ward lab and the STEM Innovation Project.

**Doctoral research**

**1 Savage lab - Studies of the CO2 concentrating mechanism**

Mentor Dr. David Savage

Role Graduate Student Researcher, Savage lab, University of California, Berkeley, 2016-2022

Description As part of my doctoral research, I studied the mechanism and evolution of HCO3- pumps that power the CO2 concentrating mechanism. Many autotrophic bacteria use CO2 concentrating mechanisms in order to increase the concentration of the CO2 they use as a carbon source. These systems work by pumping HCO3- into the cytosol, this HCO3- then enters a compartment where it is converted into a high local concentration of CO2. While the other components of the system were well characterized, the mechanism of HCO3- pumping was unknown in chemotrophic bacteria. I set out to characterize the mechanism of HCO3- pumping in the model chemotroph *H. neapolitanus*. My first step was to use transposon mutagenesis to produce a pooled library of ~100,000 mutant strains. I measured growth rates for each of these strains as a function of CO2 concentration with a massively parallel growth assay, and used this information to identify putative HCO3- pumps. I identified two putative transporter operons and set out to characterize them mechanistically. I first demonstrated that they were in fact sufficient for HCO3- pumping by cloning them into *E. coli* and using reporter strain assays and C14 uptake assays to demonstrate pumping. I then launched a series of mechanistic experiments with one of the two operons. I demonstrated that this operon produced a protein complex with a predicted zinc binding carbonic anhydrase active site. I confirmed the active site residues with point mutants and the presence of zinc with x-ray fluorescence. This suggested that the mechanism of pumping was not direct HCO3- import but instead energy coupled carbonic anhydrase activity. This causes pumping by converting membrane permeable CO2 into membrane impermeable HCO3- trapping it inside the cell. I used pH response experiments to confirm this mechanism. I also identified that this activity was coupled to the electrochemical membrane gradient with drug treatments. This is only the second family of energy coupled carbonic anhydrases ever discovered. I looked for homologous operons in other organisms, and found operons in a wide variety of other bacterial phyla including both other autotrophs and heterotrophs. I cloned the operons from two heterotrophic human pathogens *V. cholerae* and *B. anthracis* and confirmed that they were also functional HCO3- pumps. I presented posters of this work at the Western Photosynthesis conference and the Photosynthesis Gordon Research Conference. Most of the work was published in a Nature Microbiology paper describing the discovery of the HCO3- pumps, but some of was part of a study on the evolution of CO2 concentrating mechanisms that was published in PNAS.

Connection to fellowship This research helped me gain important skills in massively parallel assays for quantitative phenotypes and mechanistic biochemistry experiments that I will use in this proposal. This research also helped me to decide that massively parallel assays for quantitative phenotypes was a field that I was interested in focusing on for my career. This proposal will give me an opportunity to train in applications of this technique to a new field.

**Desmarais, J. J.**, Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., Chen, A. W., Wetmore, K., Diamond, S., Wang, J. Y. & Savage, D. F. DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nat Microbiol* **4,** 2204–2215 (2019).

Flamholz, A. I., Dugan, E., Panich, J., **Desmarais, J. J.**, Oltrogge, L. M., Fischer, W. W., Singer, S. W. & Savage, D. F. Trajectories for the evolution of bacterial CO2-concentrating mechanisms. *Proceedings of the National Academy of Sciences* **119,** e2210539119 (2022).

**Desmarais, J. J.**, Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., Chen, A. W., Wetmore, K., Diamond, S., Wang, J. Y. & Savage, D. F. DABs Accumulate Bicarbonate. Gordon Research Conference - Photosynthesis; 2019; Sunday River Resort, Maine.

**Desmarais, J. J.**, Chen, A. W., Savage, D. F. The essential gene set for bacterial carbon concentration. Western Photosynthesis Conference; 2018; Biosphere 2, Oracle, Arizona.

**2 Savage lab - Protein fitness landscape mapping**

Mentor Dr. David Savage

Role Graduate Student Researcher, Savage lab, University of California, Berkeley, 2016-2022

Description In the second half of graduate school, I began working on applying my interest in massively parallel assays and quantitative phenotypes to understanding the fitness landscape of proteins.Most efforts to map the fitness landscape of proteins have focused on mutational scans. These efforts rely on generating all single amino acid changes in a protein and evaluating the effect on a phenotype of interest in a massively parallel assay. However, single mutant libraries can be difficult to generate. Further, only considering single mutants restricts the data to the effect of the mutant in the wild type sequence context effects which can change as sequence context changes. In order to overcome these issues, I set out to develop a method for mapping the protein fitness landscape using random mutant libraries. Random libraries are easier to generate and each can mutant be seen in multiple contexts providing more information on the mutations effect. In this work, I focused on Dihydrofolate reductase an enzyme that catalyzes an essential step of nucleotide and amino acid synthesis and a target of antimalarials, antibacterials, and antifungals. I generated libraries of random mutants with error-prone PCR and measured kcat/km with a massively parallel growth assay that I calibrated using enzymes with known kinetic parameters. I then tested different models for their ability to learn mutant effects using cross validation. I focused on comparisons between linear models, general epistatic models, and neural networks. I also tested adding phylogenetic information about specific epistasis by evaluating the effect of including a potts model energy as an additional feature for each mutant. I found that general epistasis models performed similarly to neural nets even with far fewer parameters and they both showed marginal benefits from the inclusion of phylogenetic information. In order to test how well each model type extrapolated, I used simulated annealing, greedy optimization, and genetic optimization to design new mutants for each model. I selected mutants predicted to have high growth rates across a wide range of mutation numbers from each model and optimization strategy and built a set of 12,000 to test in a massively parallel growth assay. Initial results from the experiment look good, but we are still waiting for the sequencing results back from the final experiment to draw any firm conclusions about how far away from wild type each model and optimization strategy is able to generalize. This work will help develop simple and effective methods to map the fitness landscape of proteins. We hope to publish the work once this data comes in and is analyzed.

Connection to fellowship This work allowed me to apply my interest in massively parallel assays, quantitative phenotypes, and modeling in a new field. This helped me strengthen the skills I will use for my proposed project and deepened my interest in these methods.

**3 Savage lab - CRISPR tool development**

Mentor Dr. David Savage

Role Graduate Student Researcher, Savage lab, University of California, Berkeley, 2016-2022

Description During my graduate work I also worked on several large collaborative projects developing CRISPR tools. In the first of these projects, I used x ray fluorescence techniques to confirm the presence of zinc in CasX as part of the effort that characterized it a new RNA guided nuclease. This work was published in Nature. In the second effort I developed kinetic modeling strategies, data analysis software, and statistical pipelines for an effort to create more sensitive CRISPR diagnostics for detecting RNA in clinical samples. We ended up designing a new Cas13 based diagnostic that used Csm6 to amplify the signal and was able to robustly detect SARS-COV2 in patient samples. My modeling was an important part of identifying sensitivity limiting processes in the reactions and guided our decision to switch to a fluoro-modified Csm6 activator, which greatly improved sensitivity. Additionally, my data analysis software was used in the evaluation of all of the diagnostic designs we trialed. My statistical methods were used to distinguish patient samples from control samples from noisy one-off measurements in a microfluidic device. This work was published in Nature Chemical Biology. The final CRISPR tool development project I worked on was to develop and optimize a massively parallel assay for CRISPR cutting and homologous recombination in yeast. I did this work as part of an effort to apply my protein fitness landscape mapping techniques to characterize CRISPR effectors from metagenomics data and design improved CRISPR effectors. This project is still in progress and is continuing under the direction of other graduate students in the Savage lab now that I have graduated.

Connection to fellowship This work allowed me to get experience developing tools, methods, and software for others to use. I found that I very much enjoyed this approach to science, and I will be incorporating it into this proposal through my second aim in which I will be developing MPSA analysis software.

Liu, T. Y., Knott, G. J., Smock, D. C. J., **Desmarais, J. J.**, Son, S., Bhuiya, A., Jakhanwal, S., Prywes, N., Agrawal, S., de León Derby, M. D., Switz, N. A., Armstrong, M., Harris, A. R., Charles, E. J., Thornton, B. W., Fozouni, P., Shu, J., Stephens, S. I., Kumar, G. R., Zhao, C., Mok, A., Iavarone, A. T., Escajeda, A. M., McIntosh, R., Kim, S. E., Dugan, E. J., IGI Testing Consortium, Pollard, K. S., Tan, M. X., Ott, M., Fletcher, D. A., Lareau, L. F., Hsu, P. D., Savage, D. F. & Doudna, J. A. Accelerated RNA detection using tandem CRISPR nucleases. *Nat. Chem. Biol.* 1–7 (2021). doi:10.1101/2021.03.19.21253328

Liu, J. J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L. M., Chuck, J., Tan, D., Knott, G. J., Harrington, L. B., Al-Shayeb, B., Wagner, A., Brötzmann, J., Staahl, B. T., Taylor, K. L., **Desmarais, J.**, Nogales, E. & Doudna, J. A. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 1 (2019). doi:10.1038/s41586-019-0908-x

**Goals for the Fellowship and Training**

My goal for this fellowship is to prepare myself for a career as an independent researcher. In my eventual lab, I aim to focus on applying massively parallel assays and computational methods to questions with both fundamental biological relevance and real applications. In order to prepare for this career, I hope to learn new applications of massively parallel assays and computational techniques in the field of RNA biology, further develop my expertise in computational methods, and strengthen skills I will need as an independent researcher like grant writing and teaching. The project I propose here will give me the opportunity to accomplish these goals, helping me to develop as a scientist and to launch my independent career.

My proposed project provides an opportunity to learn new applications massively parallel assays and computational techniques in the field of RNA biology by focusing on developing exactly these methods for the application of RNA splicing. Further, as part of this project, I will be optimally placed to train my skills in RNA biology, massively parallel assays, and computational methods. During this project I will attend lab meetings with both the Kinney and Krainer labs. The Kinney lab is focused on using massively parallel assays and quantitative modeling to understand the biology of transcription and splicing so this is an environment that will enable me to sharpen these skills. Further, Kinney lab meetings are held as joint meetings with the Koo and McCandlish labs. The Koo lab use neural nets and interpretability methods to understand genomics data while the McCandlish lab builds theoretical understandings of fitness landscapes. Sharing lab meeting with these groups will allow me to draw on their expertise in computational methods and modeling for training. I am also attending the Koo lab machine learning in genomics weekly journal club to help me keep my knowledge of neural net methods strong. I will also attend every other week machine learning for protein engineering journal clubs. The Krainer lab on the other hand is a lab with a long history studying RNA splicing and developing splice modifying drugs for human diseases, attending their lab meetings will help me build my expertise in RNA biology. I will also gain training in RNA techniques from the members of the Krainer lab. Finally, during this project, I will train my skill at grant writing by applying for a K99 grant during the third year of my posdoc.

I will also take advantage of the advantages of my location at Cold Spring Harbor for training. Cold Spring Harbor is famous for hosting scientific meetings, I will attend biennial meetings focusing on RNA and quantitative biology. This includes the Post-Transcriptional Gene Regulation Gordon conference and the eukaryotic mRNA processing, probabilistic modeling in genomics, systems biology, biology of genomes, and biological data science Cold Spring Harbor meetings. Since each meeting is biennial, this will be one RNA meeting and two quantitative biology meetings a year. Attending these meetings will both help me to expand my skill set in these topics and build my professional network in the fields I hope to study in my independent career. They will also help me to sharpen my ability to present my work to scientists in a wide array of disciplines an essential skill for an independent investigator. I will also attend seminar series including a weekly quantitative biology seminar as well as 2 weekly Cold Spring Harbor lab wide seminars. These seminars will help me maintain a broad understanding of topics in quantitative biology. Cold Spring Harbor also offers professional development courses for postdoctoral fellows. These include grant writing and academic job search courses. I will utilize these offerings to strengthen my grant writing skill and prepare for my job search. Finally, In order to strengthen my skill teaching I will TA the graduate level quantitative biology course at CSHL.

Together, these plans will prepare me for a career as independent researcher by helping me to learn applications of my expertise to new fields, deepening my knowledge of computational methods, building my network of mentors and potential collaborators, and strengthening my skills at essential nonscience tasks like teaching and grant writing.

**Activities Planned Under This Award**

Year 1:

*Research: 85% effort*

* Aim 1:
  + Generate SMN2 test library
  + Optimize MPSA sequencing library preparation
* Aim 2:
  + Write data simulation pipeline
* Aim 3:
  + Perform course mapping of PKM minigene introns
  + Design and order oligo library encoding single mutants
  + Construct variant library

*Meetings, seminars, journal clubs: 10% effort*

* Machine learning in genomics weekly journal club
* Machine learning for protein engineering biweekly journal club
* Quantitative biology weekly seminar
* Cold Spring Harbor lab wide twice weekly seminars
* Attend weekly Lab meeting in the Krainer lab
* Attend weekly joint lab meeting Kinney, Koo, and McCandlish labs
* Post-Transcriptional Gene Regulation Gordon Conference
* Biological data science Cold Spring Harbor meeting
* Systems biology Cold Spring Harbor meeting

*Other opportunities: 5% effort*

* Teach quantitative biology course
* Attend professional development courses

Year 2:

*Research: 85% effort*

* Aim 1:
  + Benchmark LR-MPSA against other methods
* Aim 2:
  + Test analysis software on simulated data
  + Package and distribute software
* Aim 3:
  + Perform LR-MPSA on PKM variant library
  + Biochemical follow up on motifs of interest

*Meetings, seminars, journal clubs: 10% effort*

* Machine learning in genomics weekly journal club
* Machine learning for protein engineering biweekly journal club
* Quantitative biology weekly seminar
* Cold Spring Harbor lab wide twice weekly seminars
* Attend weekly Lab meeting in the Krainer lab
* Attend weekly joint lab meeting Kinney, Koo, and McCandlish labs
* Eukaryotic mRNA Processing Cold Spring Harbor meeting
* Biology of Genomes Cold Spring Harbor meeting
* Probabilistic Modeling in Genomics Cold Spring Harbor meeting

*Other opportunities: 5% effort*

* Write K99 application
* Attend professional development courses