OMB No. 0925-0001 and 0925-0002 (Rev. 10/2021 Approved Through 09/30/2024)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.

Follow this format for each person. DO NOT EXCEED FIVE PAGES.

|  |
| --- |
| NAME: Desmarais, John |
| eRA COMMONS USER NAME (credential, e.g., agency login): jdesmarais |
| POSITION TITLE: Computational Postdoctoral Fellow |

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| INSTITUTION AND LOCATION | DEGREE (if applicable) | START DATE MM/YYYY | COMPLETION DATE MM/YYYY | FIELD OF STUDY |
| Middlebury College, Molecular Biology and Biochemistry, Middlebury, VT | B.A. | 09/2012 | 05/2016 | Molecular Biology and Biochemistry |
| University of California Berkeley, Molecular and Cell Biology, Berkeley, CA | PHD | 08/2016 | 08/2022 | Molecular and Cell Biology |
| University of California, Berkeley, Berkeley, CA | Postdoctoral Fellow | 10/2022 | 12/2022 | Postdoctoral Fellow in the Savage lab |
| Cold Spring Harbor Laboratory, Cold Spring Harbor, NY | Postdoctoral Fellow | 01/2023 | present | Computational Postdoctoral Fellow in the Kinney Lab |

### A. Personal Statement

I am a perfect fit for this project because it leverages my past experience yet provides training that matches my career goals. I hope to lead a lab that uses massively parallel assays and modeling to understand how sequence information encodes phenotype. I aim to study fundamental problems with an eye for impactful applications. My proposed project aligns perfectly with my goals, allowing me to explore new applications of these techniques through fundamental questions with deep relevance for human health. The central question I ask in this proposal is, how does the sequence of an RNA molecule create anti-correlation between splicing decisions. I will develop massively parallel splicing assays to identify sequences essential for mutually exclusive splicing and characterize them with targeted experiments and modeling. My model system will be exons 9 and 10 in pyruvate kinase M, a splicing switch that drives the Warburg effect in cancer, however the tools I develop will be useful for understanding correlated splicing in many proteins and disease systems. This project perfectly leverages skills I built through my time in graduate school. During my Ph.D., I performed massively parallel growth assays to identify genes essential for CO concentration in a chemoautotroph and measured their phenotypes across CO concentrations. I then illuminated mechanistic details with biochemical and genetic experiments. In a second currently ongoing project, I used a massively parallel growth assays and modeling to map the fitness landscape of dihydrofolate reductase and design novel variants. In collaboration with the Doudna lab, I demonstrated my skill in biochemistry by showing zinc binding in the newly discovered CasX, and applied my modeling, data analysis, and statistical expertise to signal amplification in CRISPR diagnostics. These experiences prepared me to use massively parallel assays, genetic and biochemical experiments, and modeling to characterize splicing. My proposed project will allow me to further expand my skills and explore applications to RNA biology, in the human-cell context, and using long read sequencing, skills that will aid me in my independent research career. Further, this experience will allow me to strengthen my modeling abilities by working with expert teams in the Kinney lab and the Simons Center for Quantitative Biology. Finally, at Cold Spring Harbor I will have access to a wide variety of training resources including attending the on campus meetings like Eukaryotic RNA processing and the Probabilistic Modeling in Genomics.

### B. Positions and Honors

Positions and Scientific Appointments

|  |  |
| --- | --- |
| 2023 - | Computational Postdoctoral Fellow, Cold Spring Harbor Laboratory, Kinney Lab, Cold Spring Harbor, NY |
| 2022 - 2022 | Postdoctoral Fellow, University of California, Berkeley, Savage Lab, BERKELEY, CA |
| 2017 - 2019 | Graduate student instructor, University of California, Berkeley, BERKELEY, CA |
| 2016 - | Member, Phi Beta Kappa honor society, Middlebury, VT |
| 2016 - 2022 | Graduate student researcher, University of California, Berkeley, Savage Lab, BERKELEY, CA |
| 2015 - 2015 | Amgen Scholar, University of California, Berkeley, Joint Bioenergy Institute, Keasling Lab, BERKELEY, CA |
| 2014 - 2016 | Researcher, Middlebury College, Ward Lab, Middlebury, VT |
| 2014 - 2014 | Stowers Summer Scholar, Stowers Institute for Medical Research, Matt Gibson Lab, Kansas City, MO |
| 2013 - 2013 | Researcher, Middlebury College, 2013 STEM Innovation Program, Middlebury, VT |
| 2011 - 2012 | Intern, University of Washington Medical School, Neitz Color Vision Lab, Seattle, WA |

Honors

|  |  |
| --- | --- |
| 2013 - 2016 | College Scholar, 6 semesters, Middlebury college |
| 2017 | The 27th Annual Western Photosynthesis Conference travel award , Western Photosynthesis conference |
| 2016 | Elbert C. Cole ’15 Memorial Fund Prize, Middlebury College, Department of Biology |
| 2016 | Summa cum laude, Middlebury College |
| 2016 | High Honors, Middlebury College Department of Molecular Biology and Biochemistry |
| 2016 | Inducted Phi Beta Kappa honor society, Middlebury College |
| 2015 | Dean’s List, spring semester, Middlebury College |

### C. Contribution to Science

**Contribution 1 A new type of inorganic carbon pump that drives CO concentration**

1.1 Historical background The enzyme rubisco fixes 99.5% of the CO entering the biosphere each year and is essential in plants, algae, and most autotrophic bacteria. However, Rubisco is inhibited by O, a problem in the the modern atmosphere with its 20% O and only 0.04% CO. Many bacteria overcome this using an -carboxysome based CO concentrating mechanism (-CCM). These systems rely on HCO pumping, however, the mechanism of HCO pumping was unknown in chemotrophs.

1.2 Central finding To identify HCO pumps, I screened for -CCM genes in the model chemotroph *H. neapolitanus*. I identified two putative transporter operons, then showed sufficiency for pumping in *E. coli*. Unexpectedly, the data were consistent with energy coupled carbonic anhydrase (CA) activity not direct pumping. This causes HCO flux by converting membrane permeable CO into membrane impermeable HCO trapping it in the cell. I showed homologs in the pathogens *V. Cholera* and *B. anthracis* had the same activity.

1.3 My role I conceived and designed the experiments, performed the genetic screens, analyzed sequencing data, performed the mechanistic experiments, and purified protein.

1.4 Influence/Application This work identified a new family of energy coupled CAs, only the second such family known. This work enabled reconstitution of a functional -CCM in *E. coli*. A homolog found in *S. aureus* has the same function and is essential for growth in air. Factoring in energy coupled CAs aided interpretation of carbon isotope fractionation in rock strata. There are proposed applications of these pumps in engineering crop plants and autotrophic bio-fuel production hosts.

**Contribution 2 Potential evolutionary paths of carbon dioxide concentrating mechanisms**

2.1 Historical background The -carboxysome based CO concentrating mechanism (-CCM) required several major evolutionary steps to evolve. However, none of the potential intermediates are expected to provide a fitness benefit in modern conditions so it is not clear how it evolved. The atmosphere was very different when the -CCM evolved, with much higher levels of CO and much lower levels of O. This lead us to hypothesize that evolutionary intermediates of the CCM may have provided fitness benefits at intermediate atmospheric compositions.

2.2 Central finding Evolving an -CCM required acquiring a CA, gaining a HCO pump, and co-encapsulating CA and rubisco. Removing any of these stops -CCM function in normal atmosphere. We measured the effect of -CCM gene knockouts in *H. neapolitanus* across different CO concentrations. We also measured the CO dependent phenotypes of potential evolutionary intermediates in reporter strains of *E. coli* and *C. necator* that we constructed. We modeled carbon fluxes as a function of growth rate and CO concentration. This data suggested that as CO concentrations fall, HCO becomes limiting before CO. This suggested that as CO started to fall either a pump or CA can help. As levels fall further CO and HCO become co-limiting and having both a CA and a pump provides a benefit despite the potential for producing a futile cycle Eventually, only a full -CCM will work. This provides a potential path for the evolution of a -CCM.

2.3 My role I performed a massively parallel growth assay of gene knockouts in *H. neapolitanus* across intermediate CO concentrations to identify CCM genes needed at intermediate CO concentrations.

2.4 Influence/Application This work provides insight into the evolution of the -CCM and into possible life strategies of modern organisms living in high CO environments. These advances might also be useful for improving the growth of industrial autotrophs. Further, showing expression of the pumps in *C. necator* offers the potential to improve bio-plastics production.

**Contribution 3 General epistasis protein fitness landscape mapping and design**

3.1 Historical background Mutational scanning maps a protein’s fitness landscape by measuring the fitness of all single mutants. This information is used for variant effect prediction and design. However, mutational scanning experiments require production of all single mutants and only provide information on fitness in the protein’s local context, limiting their utility for divergent proteins. Being able to learn from data sets that are easier to generate and provide information over a wider area would be greatly beneficial. Current efforts to learn from random mutagenesis have relied on neural networks or linear methods. However, linear methods miss nonlinearities in the data’s true structure and neural networks cannot be inspected to gain insight into function. General epistatic models capture nonlinear genotype-phenotype relationships without sacrificing interpretability but have not been applied to protein design. Phylogentic data also provide insight into fitness landscapes but inspectable models that are able to leverage both phylogenetic and experimental data have been rare.

3.2 Central finding We performed random mutagenesis on dihydrofolate reductase and measured enzyme activity in a massively parallel growth assay. I trained models including linear models, general epistatic models, and large neural nets with and without access to phylogenetic data. I showed that the general epistatic models had simmilar performance to the neural net on held out test data. Both these models showed improvement when provided with phylogenetic data but out performed the phylogenetic data alone or with the linear model. I also used each model to design new variants while systematically varying both number of mutations and optimization strategy to evaluate how well each model is able to extrapolate. I used a massively parallel growth assay to validate 12,000 designed proteins. This work is currently awaiting sequencing results from the final experiment before submission.

3.3 Influence/Application This work provides new methods for mapping the genotype-phenotype landscape of proteins and designing new variants. It also provides simple directly inspectable models that produce performance on par with neural net approaches for some prediction and design tasks.

3.4 My role I am the primary author on this work. I designed and conceived of the experiments, produced mutant libraries, performed massively parallel growth assays, wrote analysis code, wrote model code, trained models, tested model performance, and evaluated optimized sequence behavior.

**Contribution 4 Development of new CRISPR tools**

4.1 Historical background The ever expanding suite of CRISPR tools has helped drive a biotech revolution. I have been part of two tool development projects, identifying CasX as a new RNA guided DNA nuclease and the development of Cas13/Csm6 RNA diagnostics.

4.2 CasX genome editing RNA guided DNA nucleases launched the CRISPR field. However, nuclease size has hindered therapeutic applications. CasX is a 1,000 amino acid RuvC containing protein from CRISPR loci. We demonstrated guide directed cutting activity *in vitro* as well as cutting and CRISPRi *in vivo*. We solved a structure of the complex with target DNA and identified two new domains. We also detected a zinc binding motif and showed that CasX binds zinc. My role was to use x-ray fluorescence to detect zinc bound to the purified protein. CasX provides a new modality for genome editing that is proving useful in a variety of applications. Scribe Therapeutics is pursuing casX based therapies.

4.3 Csm6 boosted Cas13 RNA detection CRISPR diagnostics can detect nucleic acids in one-pot isothermal reactions. This makes them attractive for at-home or point-of-care diagnostics. However, poor sensitivity meant pre-amplification was required to detect SARS-COV2 in patient samples. Class III CRISPR systems include a cyclic-oligo-A activated nuclease, Csm6. We hypothesized that linking cas13 and csm6 using cas13 targets that release Csm6 activator upon cleavage would improve sensitivity. We used kinetic modeling and spike in assays to show that secondary activator cleavage was limiting sensitivity. Using an activator resistant to secondary cleavage allowed detection of SARS-COV2 in clinical samples. My role in this work was kinetic modeling and writing the data analysis and statistical pipelines. This work has improved time to detection and sensitivity in CRISPR diagnostics.

**Contribution 5 Driving carbon flux toward chemical production by engineering glucose uptake during nitrogen starvation**

5.1 Historical background Using microbial hosts to produce chemicals offers the potential to produce a variety of compounds from renewable feed-stock. However, production hosts frequently loose production efficiency as natural selection drives evolution towards redirecting carbon and energy flux towards growth not production. This can be overcome by coupling production of the desired chemical to cell fitness, but in many cases this is not possible. An alternative strategy is growth decoupling, in which production hosts are grown up, then growth is stopped and all metabolic flux is directed towards production. However, when growth is stopped, many chassis organisms including *E. coli* slow and eventually halt their metabolism stopping production. A general strategy for enhancing metabolic rate during growth decoupling would dramatically improve prospects for engineered chemical production in biological hosts.

5.2 Central finding We found that by over-expressing PstI we were able to increase glucose uptake after growth was stopped by nitrogen limitation. However we did not find this increased glucose consumption increased yield significantly, and it is likely additional work will be needed to direct this increased flux towards chemical production.

5.3 Influence/Application This work provides another tool that metabolic engineers can use to optimize the production of their compound of interest.

5.4 My role My role in this project was to perform growth and chemical production assays with modified strains and to prepare samples for mass spectrometry.

### D. Scholastic Performance

Scholastic Performance

|  |  |  |
| --- | --- | --- |
| YEAR | COURSE TITLE | GRADE |
| Middlebury College | | |
| University of California Berkeley | | |
| 2016 | MCELLBI 200A - Fundamentals of Molecular and Cell Biology | A |
| 2016 | MCELLBI 200B - Fundamentals of Molecular and Cell Biology | A |
| 2016 | MCELLBI 280A - Selected Topics in Molecular and Cell Biology | S |
| 2016 | MCELLBI 291A - Introduction to Research | A |
| 2016 | MCELLBI 293A - Research Seminar | S |
| 2017 | MCELLBI 206 - Physical Biochemistry | A |
| 2017 | MCELLBI C212A - Chemical Biology I - Structure, Synthesis and Function of Biomolecules | A+ |
| 2017 | MCELLBI C212B - Chemical Biology II - Enzyme Reaction Mechanisms | A |
| 2017 | MCELLBI C212C - Chemical Biology III - Contemporary Topics in Chemical Biology | A- |
| 2017 | MCELLBI 291B - Introduction to Research | A |
| 2017 | MCELLBI 293C - Responsible Conduct, Rigor and Reproducibility in Research | S |
| 2017 | MCELLBI 295 - Careers for Life Sciences Ph.D's | S |
| 2017 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2017 | MCELLBI 292 - Research | A |
| 2017 | MCELLBI 380 - Teaching of Molecular and Cell Biology | S |
| 2018 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2018 | MCELLBI 292 - Research | A |
| 2018 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2018 | MCELLBI 292 - Research | S |
| 2018 | MCELLBI 294 - Current Topics in Biomedical Sciences | S |
| 2019 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2019 | MCELLBI 292 - Research | A |
| 2019 | MCELLBI 380 - Teaching of Molecular and Cell Biology | S |
| 2019 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2019 | MCELLBI 290 SEM A02 - Graduate Seminar | A+ |
| 2019 | MCELLBI 290 SEM D01 - Graduate Seminar | A+ |
| 2019 | MCELLBI 292 - Research | A |
| 2019 | MCELLBI 294 - Current Topics in Biomedical Sciences | S |
| 2020 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2020 | MCELLBI 290 - Graduate Seminar | A |
| 2020 | MCELLBI 292 - Research | A |
| 2020 | MCELLBI 293R - Responsible Conduct of Research Refresher | S |
| 2020 | MCELLBI 295 - Careers for Life Sciences Ph.D's | S |
| 2020 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2020 | MCELLBI 292 - Research | A |
| 2021 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2021 | MCELLBI 292 - Research | A |
| 2021 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2021 | MCELLBI 290 - Graduate Seminar | A+ |
| 2021 | MCELLBI 292 - Research | A |
| 2022 | MCELLBI 218X - Research Review in Biochemistry and Molecular Biology: Chemical Reactions of Metabolism | S |
| 2022 | MCELLBI 292 - Research | A |
| 2022 | MCELLBI 294 - Current Topics in Biomedical Sciences | S |

For all University of California Berkeley graduate level courses, the scale is from A to F (A+ is awarded as a mark of achievement but both A+ and A are counted as 4.0 for GPA calculations) and passing grades are B- or higher. In this scale A+ is a 4.0, A is also 4.0, and an A- is a 3.7. S indicates a passing grade (B- or higher) in a course graded on a Satisfactory/Not Satisfactory grading scheme, courses graded on this scheme are not included in GPA calculations.