

APPLICANT'S BACKGROUND AND GOALS FOR FELLOWSHIP TRAINING

Doctoral Dissertation and Research Experience

My undergraduate research focused on computational biochemistry, forming a robust foundation for the proposed work. I began in synthetic chemistry which provided me with strong skills at the bench, but after coursework in bioorganic chemistry and research in structural biophysics I transitioned to a more computational focus. I defended an Honors thesis in Biochemistry and left Oberlin as a well-rounded scientist. My early emphasis on computational biology led me to a position in the lab of Dr. Blanton Tolbert (see letter of reference and biosketch), where I conducted molecular dynamics simulations for RNA structure calculation, ultimately contributing to 4 publications. I built my technical background by excelling in both the preclinical medical school curriculum and Bioinformatics and Genomics graduate coursework at Penn State. I continued my research productivity into graduate school, co-authoring multiple papers and presenting preliminary data for the proposed work at national meetings. Today I work within an interdisciplinary team that we believe will optimally prepare me for a career as a physician-scientist investigating technical questions in cancer biology.

Undergraduate Research: I began my research career in the lab of Dr. Albert Matlin midway through my sophomore year at Oberlin College. I had decided to major in Biochemistry and had just completed organic chemistry. The course had been challenging but I became fascinated with the subject matter, especially mechanisms and synthesis. This prompted me to seek out a winter term project in the Matlin Lab and to take advanced organic chemistry courses in the spring semester of my sophomore year (CHEM 254 and CHEM 325). I was fortunate to learn advanced techniques in organic synthesis, enantioselective catalysis, chromatography, and NMR.

CHEM 254 was taught by Dr. Duy Hua, a structural biologist. Her course connected organic chemistry and structural biology, emphasizing the biophysics of protein-protein and protein-drug interactions. My interest in this subject matter drove me to turn a corner in my academic career, and I scored an A in both CHEM 254 and 325. Dr. Hua was a new professor seeking to mentor students in structural biology research with a focus on molecular dynamics and docking simulations of protein structures. I seized the opportunity and thus began my journey into computational research. I spent the entirety of my Junior year at Oberlin working in Dr. Hua's lab, taking the leap from chemistry and chromatography to research primarily conducted on the command line. My research involved molecular dynamics (MD) simulations of phosphorylated and dephosphorylated human spleen tyrosine kinase (SYK) structures acquired via x-ray crystallography. At Oberlin, we had remote access to a cluster computer with many CPU and GPU nodes that I used to conduct MD simulations for the Hua lab. In addition to developing technical skills, I deepened my understanding of protein structural change and signaling. Through the opportunity to conduct research in her lab, I became fascinated with molecular dynamics simulations. I sought opportunities to dive deeper into the mysteries of how biomolecular structure governed the biology of disease. Dr. Hua encouraged me to pursue molecular dynamics research and ultimately to apply to MD/PhD programs, setting me on the path to a career in computational biology.

Throughout my junior year, I applied to labs that performed similar research and ultimately chose a position to work for Dr. Blanton Tolbert (see letter of reference and biosketch) in the Department of Chemistry at Case Western Reserve University. I began my work in Dr. Tolbert's lab during the summer after my junior year at Oberlin and started on several projects performing MD simulations of non-coding RNAs and RNA-protein complexes. In my first summer in the lab, I worked to integrate NMR and small-angle x-ray scattering (SAXS) data into MD simulations to predict 3D solution structures of key regulatory RNAs in HIV-1. I presented my work that summer at the Meeting for Structural Biology Related to HIV/AIDS at the NIH, and I gave an oral presentation titled *Hacking the Viral Mechanisms of HIV with Molecular Dynamics Simulations* at Oberlin my senior year.

I successfully defended an Honors thesis in Biochemistry during my senior year at Oberlin. My project focused on modulating the self-polymerization of polydopamine with a variety of small molecules in the lab of Dr. Jason Belitsky. This organic synthesis/materials chemistry project honed my skills in the wet lab and aptly supplemented my focus in computational biochemistry. The challenging process of writing and defending an undergraduate thesis helped shape me as an academic scientist and cemented my desire to work toward a PhD. I presented my Honors thesis research in an oral presentation at the ACS Meeting in Miniature (MiM) at John Carroll University and in two seminar talks at Oberlin College. My thesis defense was difficult, but the rewarding process of seeing the project to completion and learning from my mistakes made me certain in my desire to pursue a career in science.

Postgraduate Research: I joined the Tolbert Lab as a Research Assistant upon graduation from Oberlin and continued my research on RNA 3D structure and its role in infectious diseases. In addition to continuing my work conducting MD simulations of non-coding RNAs in HIV, I also joined a project investigating RNA-drug interactions in enterovirus-71 (EV-71), a causative agent of hand, foot, and mouth disease. Our team collaborated with Dr. Amanda Hargrove's Lab at Duke and together discovered that an amiloride-based small molecule could bind to a stem-loop structure within the virus and prevents transcription. I performed MD simulations that integrated NMR and SAXS data to help reveal the altered structure of a binding site for a heterogenous nuclear ribonucleoprotein which blocked transcription of the virus. This work culminated in a paper published in Nature Communications, on which I am an author.

I also spent significant time in the Tolbert lab conducting MD simulations of other non-coding RNAs that govern transcription in HIV-1. I translated NMR data such as residual dipolar couplings (RDCs) to constrain simulations and worked with other members of the lab to validate structures in the context of the molecular pathways they participate in. This work resulted in two other publications on which I am an author - one in the Journal of Molecular Biology and another in the Journal of Biological Chemistry.

Conducting research on RNA viruses and RNA drug interactions affirmed my desire to become a physician scientist. In the Tolbert lab, we constantly investigated dynamic structures for which we held only clues derived from experimental data. I became fascinated with the inference of conformational states and 3D structure, and it was through this work I was first introduced to challenge of statistical uncertainty. This curiosity laid the groundwork for my interest in the Silverman lab, and the familiarity of 3D structural research made the Cheng lab a great partner in my efforts to contribute to puzzle-solving in biology.

Graduate and Medical Coursework: From the Tolbert lab, I joined the MSTP program at Penn State College of Medicine to pursue training at a top research institution with a premier cancer institute. I have built on my foundation in computational biology by taking on a challenging courseload in the Bioinformatics and Genomics program at University Park, especially in BMMB 802 (Applied Bioinformatics), MCIBS 554 (Bioinformatics 1), BIOL 428 (Population Genetics), and STAT 555 (Statistical Genomics). These courses required programming skills in python, unix, and R (802, 554, and 555), and provided thorough instruction in both applied frequentist and Bayesian statistics (STAT 555 and BIOL 428). I received an A in each of these and completed my graduate coursework requirements with a 3.93 GPA. I have also been engaged in an ongoing self-study in statistical methods and machine learning under the supervision of Dr. Silverman. As a part of this self-study, I have worked through the textbook *Statistical Inference* by Casella and Berger to address gaps in my knowledge of probability theory in support of aim 2 and 3. In addition to building my knowledge in statistical theory, I have performed an intensive study of the fundamentals of Topological Data Analysis (TDA), conducting a literature review centered around its applications to 3D data and imaging. I have built my knowledge base in TDA by working through the textbook *Computational Topology* by Herbert Edelsbrunner and John Harer as a reference. Throughout this self-study I have referred to the book *Matrix Algebra from a Statistician's Perspective* by David A. Harville to fill my knowledge gaps in Linear Algebra.

I have completed the BMS 591 course in Biomedical Research Ethics at Penn State College of Medicine which has supplemented the training in medical ethics and humanities I received through the first two years of medical school. Penn State emphasized the Science of Health Systems (SHS) courses during the medical years of our training, during which we also discussed ethical problems in the setting of hospital medicine and biomedical research. In SHS we received in-depth instruction in other topics such as quality improvement research and biostatistics. Discussion of high-value care and systems approaches to patient-facing problems have influenced my research as we strive to build translational methods that positively supplement the work of clinicians and scientists alike. This coursework has also supplemented my studies in data science and statistics with an emphasis on the analysis of EMR and the interpretation of patient data.

Doctoral Dissertation Research: My rotations in Dr. Silverman and Dr. Cheng's labs advanced both my expertise in machine learning and micro-CT imaging, forming the basis for my research into tumor phenotype quantification. I sought out a pre-matriculation rotation with Dr. Cheng because of the unique imaging experiments conducted in the lab as well as the novel computational problems that arise in the analysis of these images. My project during the rotation focused building a pipeline for the automated segmentation of blood cells in micro-CT scans of zebrafish. This project ultimately enabled researchers who study zebrafish models of disease genetics to computationally and quantitatively phenotype whole organisms based on the shape characteristics of their blood

cells. This rotation in the Cheng Lab provided a strong foundation for my interest in micro-CT and continues to compel my affinity for computational problems. This project resulted in a paper published in *eLife*, of which I am one of the authors. I also presented aspects of this work in an oral presentation at the 2021 MD/PhD National Conference and in a poster at the 2023 Mid-Atlantic Regional Zebrafish meeting at the NIH.

I completed a lab rotation in Dr. Silverman's lab to gain exposure to a theory-driven approach to machine learning. During my rotation, I searched for methods from topological data analysis and applied them to the classification of complex shapes in biomedical data. Dr. Silverman and I met a minimum of once per week and studied the methods I was learning at the whiteboard often for multiple hours at a time. I first applied methods from persistent homology during this rotation, where I used the Vietoris-rips filtration to classify wild-type and mutant zebrafish blood cells based on shape. I believe that TDA has further untapped potential in image analysis, whereby it will contribute to the goal of quantifying phenotype based on cellular shape and location.

After rotating in the Cheng and Silverman labs, I established a co-mentorship that will provide unmatched training for a career at the technical cutting-edge of cancer research. Both co-sponsors of this grant hold MD and PhD degrees, and will have significant roles in shaping my development into a physician-scientist. Dr. Silverman and Dr. Cheng each have a unique skillset that will both challenge and support me along my journey towards becoming an independent investigator. Their areas of research are independent but also complementary: Dr. Silverman is an expert in statistical methods and the analysis of biomedical data, while Dr. Cheng pathologist by training who has an extensive background in both imaging and genetics research. The first two years of this co-mentorship have been productive, resulting in significant preliminary data and multiple collaborations. My project requires the collection of high-resolution, volumetric data, and I have conducted a series of advanced experiments at synchrotron beamlines in pursuit of this goal. After initial attempts to stain biopsies with contrast-enhancing heavily metals failed, I explored phase-contrast imaging with unstained samples. I attended 5 synchrotron trips in total, and I assumed the role of experimental lead for the two most recent trips where the focus was phase-contrast imaging of unstained tumor biopsies. These trips were successful, and I have presented these results in my MSTP program seminar and in a poster at the American Society for Investigative Pathology (ASIP) Pathobiology conference in 2024.

B. Training Goals and Objectives

1. Career Goals and Research Interests: My long-term goal is to become a physician-scientist leading a translational research lab that develops and applies novel imaging and computational methods to advance understanding of phenotypic heterogeneity in cancer. I became interested in cancer research after shadowing the bone marrow transplant wards at John's Hopkins as an undergraduate, and my desire to dedicate my career to the field was magnified by my experiences working with thesis committee members Dr. Hohl and Dr. Warrick (see reference letter and biosketch) to date. I plan to apply to a research-track internal medicine or pathology residency-fellowship program with the aim of becoming a tenure-track faculty member at a top academic institution. I will work towards this goal in the short term by continuing to hone my clinical skills throughout my graduate years and by anchoring my thesis research in translational problems.

I have two goals for this phase of my research training. First, I aim to develop and optimize algorithms that will contribute to the analysis of complex 3D images, specifically aimed at methods applied to cancer biopsies. Aim 2 and especially Aim 3 of this proposal will provide me with a unique opportunity to do so. Second, I strive to cultivate a skillset and understanding of data science and statistical methods that I will use throughout every facet of my career as a physician-scientist.

3. Clinical Training Goals: I aim to improve my clinical skills and specialty exposure throughout the duration of this award to prepare me both for success in medical school and my future career as a physician-scientist. Within the proposed work, I plan to invest time in both medical oncology and pathology to identify the optimal career fit for the clinical aspect of my career as a physician-scientist. The questions explored by the proposed research lie at the intersection of medical oncology and pathology, and thus the key goal of my clinical training under this award is to cultivate the expertise and experience necessary to communicate with experts in translational problems across these fields.

4. Professional Development Goals: By leading the proposed work, I plan to improve my scientific writing, communication, and interpersonal skills under this award such that I will be competitive for a future career grant K99/R00 award. I aim to leverage the unique opportunities afforded to me under this award to present my research to and collaborate with leaders in the fields of cancer biology and statistics.

C. Activities Planned Under This Award

Research Training Activities: Through the experiments planned under this proposal and the guidance of my co-sponsors and broader mentorship team, I will accomplish my research goals and build on the research experience I have acquired to date.

- **Investigation of Tumor Heterogeneity:** Prostate cancer and other solid tumors exhibit multiple phenotypes that evolve under the selective pressure of chemotherapy, leading to treatment failure and disease progression. Improved detection of these phenotypes will directly advance cancer biology research and ultimately patient care. Thus far in my PhD I have worked to contribute to this field of cancer research by beginning with the first principles of histology and building towards a rigorous and reproducible 3D histopathological workflow. With a team that includes Dr. Cheng and Dr. Warrick we will be well-positioned to contribute advances in this space in the near future.
- **Applied Statistics:** Concomitant with imaging experiments, I have invested significant time in developing the ability to analyze complex data. Dr. Silverman's lab consists of several students working on difficult but related problems, including two from the Bioinformatics and Genomics program and two from the statistics program. Lab meetings in which we present challenging problems to one another as well as one-on-one discussions with Dr. Silverman have honed my skillset in applied statistics. This will continue to accelerate under the proposed award as I complete additional coursework and allocate further time to data analysis upon completion of A1.
- **Synchrotron Micro-CT Imaging:** Imaging experiments in the Cheng Lab involve using micro-CT resources both on campus at Hershey Medical Center (HMC) and at Lawrence Berkeley National Laboratory (LBNL). I will gain experience with x-ray imaging in each setting. In service of aims 1 and 2 of this proposal I will lead one experimental imaging trips to beamline 8.3.2 of the Advanced Light Source at the LBNL and one to the Advanced Photon Source of Argonne National Labs.
- **Image Reconstruction:** Throughout my training under this award I will build on the programming skills I have already acquired by continuing to perform and refine 3D image reconstructions in python. Additionally, I will work closely with Dr. Patrick La Riviere (see letter of reference, biosketch) from the University of Chicago in collaboration on phase-contrast experiments. Through this collaboration I will deepen my understanding of medical physics and the mathematics that govern x-ray imaging.
- **Statistical Methods and Study Design:** Under this proposal, I will also study non-inferiority trials, ANOVA, and mixed effects models in support of aim 2. This training will improve my literacy in statistical methods that will translate beyond basic science research. Completion of these aspects of my training will result in a robust foundation in statistical theory that will mold me into a well-rounded physician-scientist capable of contributing to both basic and clinical research.
- **Topological Data Analysis:** The execution of aim 3 will leave me with an expertise in the applications of persistent homology and a skillset that will transfer to my career as an independent researcher. I will complete the textbooks *Computational Topology* by Edelsbrunner and Harer and *Computational Topology for Data Analysis* by Tamal Krishna Dey and Yusu Wang. I will take a formal course in Topology within the mathematics or statistics department. These activities will directly support aim 3 and my future career - morphology is an oft understudied characteristic of tumor phenotype and TDA represents a means of quantifying it, potentially contributing a wealth of untapped insight, especially in complex cases.

Clinical Training Activities:

- **MSTP Clinical Exposure Program (CEP):** All MSTP students at PSCOM are required to participate in CEP. The CEP requires that these students join an advisor in clinic for a minimum of 6 half days each semester. Through this program, I am working with Dr. Lilia Reyes, a Pediatric Emergency Medicine Physician at Penn State Health. She will evaluate my clinical skills and note writing in preparation for a return to medical school.
- **MSTP Clinical Research Conference:** The MSTP program hosts CRC once every two months, during which students are guided by a physician scientist through a presentation of a challenging clinical case, including detailed discussion of the pathophysiology and a pertinent paper. In 2023 I presented the pathophysiology of pre B-cell ALL. In 2024, I will present the paper related to a different clinical case.
- **Objective Structured Clinical Exams (OSCEs):** OSCEs are used to adjudicate the clinical skills of medical students throughout training at PSCOM. Students have the opportunity to enroll in OSCEs during the PhD portion of training, and I will participate in one OSCE per semester throughout the duration of this award.
- **Lioncare Volunteering:** To supplement formal clinical training, I will continue my participation in the student-

run Lioncare free clinic treating underserved populations at the Bethesda Mission in Harrisburg a minimum of once every two months.

- **Subspecialty Shadowing:** I will join Dr. Raymond Hohl (thesis committee member) for one full day in outpatient Hematology/Oncology clinic per semester. I will also shadow Dr. Joshua Warrick (thesis committee member, see reference letter and biosketch) in pathology to support my connection of the proposed work to clinical need and to explore the specialty of pathology as a career path.

Goal	Year 1	Year 2	Year 3	Year 4
Micro-CT Imaging	Lead additional imaging trip to Advanced Light Source at LBNL Lead experimental imaging trip to Argonne National Labs	Present imaging data to pathologists Present at SPIE medical imaging conference Conduct non-inferiority trial	Develop image analysis pipelines during dedicated research time Draft, submit, and edit additional manuscripts	Revise, edit, and submit additional manuscripts
Statistical Methods	Complete coursework in machine learning (IST557) Longitudinal self-study in Linear Algebra and FDA	Present applications of self-study and Data Mining to Silverman Lab Prepare for submission of manuscript based on A3 and thesis defense	Apply methods from thesis to data analysis of translational data in protected research clerkships	Resume self-study and conduct literature review to prepare for transition to post-doctoral research
Topological Data Analysis	Begin self-study in TDA Conduct progress meetings with Dr. Silverman covering self-study and project	Present results of self-designed project Complete coursework in TDA	Present results of A3 at JSM 2026 conference Explore translational applications of TDA during research rotations	Conduct literature review to prepare for transition to post-doctoral research
Professional Development	Present at ALS users meeting Present clinical research conference to MSTP program	Present at USCAP conference Present MSTP seminar ahead of PhD defense	Gain experience in clinical and translational applications of thesis work within medical school research	Present thesis work at MD/PhD National conference ahead of residency applications
Clinical Training	Longitudinal clinical training (BMS802) in Pediatric Emergency Medicine with Dr. Lilia Reyes	Complete OSCE exams and pre-clerkship	Complete clinical clerkships and research rotations Sit for shelf exams and USMLE step 2	Complete acting internships Apply to physician-scientist training programs (PSTPs) for residency and fellowship

Table 1: Roadmap of training experiences planned under this award

Research Training: On behalf of the Cheng lab and in close collaboration with Dr. La Riviere (see letter of reference), I will continue to conduct highly specialized research in x-ray physics and interact with beamline scientists at Lawrence Berkeley National Laboratory and Argonne National Laboratory. With the guidance of Dr. Cheng and through collaboration with Dr. Warrick I will present my work to clinical pathologists who, in their daily practice, directly encounter the problems defined in this proposal, and I will integrate their feedback into my future experiments. To further pursue the input of pathologists I will apply to present this work at conferences such as the United States and Canadian Academy of Pathology (USCAP) Annual Meeting. Additionally I will present my work to an audience of computational researchers and statisticians. I will apply to present at meetings such as the SPIE medical imaging conference to improve my understanding of computational research relevant to this project. I will also apply to present my work on topological data analysis (A3) at JSM in 2026.

Training in Statistical Methods and Topological Data Analysis: Under the proposed work, I will bolster my background in statistical theory and mathematics to achieve the level of expertise I need to contribute from a data science perspective. I will build on my already strong background in machine learning and applied data science by taking Data Mining (IST557) this fall. IST557 takes material from *Pattern recognition and machine learning* by Christopher Bishop and *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. For functional data analysis, I will complete the chapters and practice exercises of *Introduction to Functional Data Analysis* by Piotr Kokoszka and Matthew Reimherr, reviewing difficult problems with Dr. Silverman weekly. Completion of this course and self study will give me expertise in complex algorithms across regression, classification, and clustering, expanding my toolbox and enabling me to carry out the FDA elements of aim 3.

Topological data analysis (TDA) requires a greater understanding of linear algebra (LA) than I currently have. Although IST 557 involves significant content review of LA I will supplement this instruction by completing a guided self-study with Dr. Silverman. During this self-study I will refer to *All the Mathematics You Missed* by David Garrity to bridge any remaining gaps in understanding. To further support my knowledge of TDA and thus

aim 3, I will complete the *Computational Topology* textbook as well as the book *Computational Topology for Data Analysis* by Tamal Krishna Dey and Yusu Wang. To certify my understanding of these topics I will complete and present a self-designed project applying TDA to 2D histopathology images, transforming the results into a continuous variable representing severity of malignancy and thus building on the work done by Lawson et. al. I will present this project to the Silverman lab and incorporate their feedback as I begin aim 3. Upon completion of this material and presentation, I will take the Topology course in the statistics/math department at PSU. This will not only further support Aim 3 of this proposal, but will also advance my skills as a data scientist and build a foundation for critical analysis of other biomedical research problems. To supplement our learning and improve student writing skills, students in Dr. Silverman's lab also conduct a self-study reviewing *The Sense of Structure: Writing from the Reader's Perspective* by George Gopen.

Professional Development Activities: My co-mentorship will provide a strong professional development experience that will uniquely prepare me for a career as a physician-scientist capable of working across disciplines. The Silverman Lab meets weekly, alternating between professional presentations given by students and open-discussion journal club. The lab has an expertise in statistical theory and methods applied to biological problems, and I will benefit not only from hands-on training from Dr. Silverman, but also from regular lunch meetings and close working relationships with other students in the lab. I meet with Dr. Silverman individually on a weekly basis and additionally when needed. The Cheng Lab also meets weekly, with each student presenting 45-min individual updates alternating with student-led journal clubs. I also meet weekly with Dr. Cheng to plan projects and receive feedback. I have been fortunate to form a close relationship with both of my co-sponsors and have benefited greatly from their mentorship.

To support this co-mentorship efficiently, I will continue to commute to University Park (Penn State University main campus) on Wednesdays where I have a dedicated workspace in the Silverman Lab. The Cheng Lab meets every Wednesday morning at 9am, and then the Silverman Lab meets in person every Wednesday at 1:30pm. Dr. Silverman and I then typically meet in person at 4pm before I return to Hershey. To supplement our regular one-on-one meetings, Dr. Silverman also meets with me before and after each presentation I deliver, as he does with all of his mentees. Dr. Silverman and Dr. Cheng also continue to have regular meetings for their ongoing collaborations and the joint supervision of my project. We meet monthly via zoom and additionally as needed. This has benefitted me as a student and supported both the publication of a paper on which I am an author as well as my successful completion of the PhD comprehensive exam in the Bioinformatics and Genomics program.

Clinical Training and Medical School Clerkships: In addition to clinical exposure through the BMS 802 course and shadowing, I will also continue to volunteer in monthly free general medicine clinics at the Bethesda Mission in Harrisburg through the Lioncare student group at PSCOM. This will supplement the clinical training I receive in the pediatric emergency department with additional repetitions of history taking, physical exams, and note writing in the setting of a different patient population. Both of my co-mentors have a background in clinical medicine and are familiar with the training and experience required for effective medical training.

After successfully defending my PhD thesis, I will re-enter medical school for the third and fourth years of clinical training and this award. I will focus during this period of training on my clinical skills and prioritize my development into a well-rounded physician. I will continue to develop my research, leadership, and professional skills in addition to my clinical training throughout this award. During the third year of medical school (M3 in table 1) I will complete clerkships in all major specialties and select elective rotations in pathology and hematology/oncology. Upon completion of M3 I will sit for the USMLE Step 2 medical board exam.

During the fourth year of medical school (M4 in table 1) I will utilize at least 3 of my 12 4-week blocks to conduct research exploring the applications of the proposed work in digital pathology and malignant hematology. The expected balance of longitudinal professional development and research experiences under the award is summarized in Figure 2. The training the proposed work would provide will prepare me for a career as an independent physician scientist. I am fortunate to collaborate with experts in biomedical imaging, histopathology, and statistical methods who will support this interdisciplinary project and shape me into a versatile early-career scientist. As I complete the final year of this award and complete medical school I will apply to either a research-track internal medicine or research-track pathology residency-fellowship program.

Year	Research	Coursework	Prof. Development	Clinical Training
G3	80%	10%	5%	5%
G4	80%	1%	5%	14%
M3	5%	1%	1%	93%
M4	15%	5%	10%	70%

Table 2: Distribution of activities across stages of proposal