A. Rationale, Mission, and Objectives

The Computational and Mathematical Modeling of Biological Systems (CMMBS) training program is designed to meet a critical and under-addressed need in graduate education: the development of expertise in computational and mathematical modeling to understand and predict the behavior of complex biological systems. Essential applications of such modeling include disease understanding, drug discovery, and personalized medicine. While the importance of quantitative approaches is widely recognized, few graduate programs provide explicit and comprehensive training in this area, and fewer still do so in a way that is accessible to biology majors who graduate with only the typical amount of mathematical coursework. By integrating cross-disciplinary coursework and activities with the requirements of students' home graduate programs, CMMBS fosters both deep domain expertise and the ability to communicate and collaborate across disciplinary boundaries.

A hallmark of the program is its intentional recruitment of trainees from multiple Ph.D. programs, including those oriented toward theory as well as those oriented toward biology. These programs include Applied Mathematics, Molecular and Cellular Biology, and Biomedical Engineering. This allows students to learn from and with peers whose skills complement their own. We have designed the CMMBS program to support five cohorts of eight interdisciplinary trainees each year.

Central to the program is the Quantitative Biology Colloquium (QBC), which has been running successfully for more than 25 years, connecting students across disciplines and enhanced to address key objectives of the revised training program. Two additional core courses equip trainees with foundational modeling skills, and customized workshops equip them with essential competencies in open science.

The training program combines the requirements of the individual trainee's own graduate program with courses and activities specifically designed to develop trainees' cross-disciplinary knowledge and their ability to communicate and collaborate across traditional disciplinary boundaries. Students can complete the program requirements without extending the time needed to earn their graduate degrees, making this a highly accessible and impactful opportunity. The University of Arizona's (U of A) longstanding commitment to interdisciplinary graduate education provides an ideal foundation for this initiative.

A1. Rationale

The goal of the Computational and Mathematical Modeling of Biological Systems (CMMBS) training program is to prepare predoctoral scientists for research careers that incorporate mathematical and computational modeling of medically significant biological systems through interdisciplinary training. Several trends have combined in recent decades to make computational and mathematical approaches increasingly important and even essential for the further progress of biomedical research. The techniques of molecular biology and advances in information technology have made available vast amounts of data, from genomes and proteomes to images and patient medical records. The analysis of such data poses enormous challenges. With the availability of such data has come the recognition that biological processes and diseases are not related in a simple way to genomic data but result from a multi-layered web of interacting processes extending from molecular to cellular to tissue to whole-body scales. Mathematical and computational models are essential theoretical tools for integrating such diverse information and linking biological processes that occur on disparate scales. The computer power available to investigators has increased dramatically, opening possibilities for data analysis and biological simulations that were previously impractical. Applications of modeling in biology are offering new insights into disease mechanisms and informing drug discovery and clinical decision-making through personalized medicine.

A2. Mission and Objectives

The mission of CMMBS is to prepare predoctoral scientists for research careers at the interface between biology and quantitative modeling, encompassing computational, mathematical, and statistical approaches to medically significant biological systems. This mission directly addresses the stated objective of the funding announcement to "encourage changes in biomedical graduate training to keep pace with the rapid evolution of the biomedical research enterprise that is increasingly complex, interdisciplinary, and collaborative."

The overall objectives of the proposed program are to:

- develop a pool of scientists who are trained to apply computational and mathematical modeling approaches to biomedical systems, as demonstrated by successful completion of CMMBS coursework;
- achieve a 90% or greater Ph.D. completion rate among trainees;
- ensure median time to degree does not exceed six years.

For each individual trainee, training objectives are to:

- broaden knowledge of multiple areas of biomedical science in which modeling approaches play a substantial role;
- develop critical thinking and creativity sufficient for independent research;
- build the ability to work and communicate effectively across disciplinary boundaries;
- develop skills as communicators of science to both specialist and general audiences;
- ensure commitment to rigorous and reproducible research and responsible conduct;
- instill the ability to work effectively in cross-disciplinary scientific teams and to promote supportive research environments;
- prepare trainees for careers across academia, industry, and government, and ensure awareness of the range of available opportunities.

History: Interdisciplinary training at the interface of biology and quantitative modeling at the University of Arizona dates to 1995, when a Flinn Foundation award established the campus-wide Biological Mathematics and Dynamics initiative. In 1998, that initiative was selected as one of 17 NSF IGERT recipients from a field of 620 competing proposals, providing external validation of the program's quality and expanding its scope through 2005. The BIO5 Institute provided bridging support until NIGMS funding established CMMBS in 2009, which has operated continuously ever since.

A3. Key changes in this renewal application

This renewal reflects programmatic development since the prior funding period across five areas: leadership, curriculum, mentoring and oversight, career development, and evaluation.

- The PI team has expanded from sole-PI Tim Secomb to a multi-PI structure led by Ryan Gutenkunst, Joanna Masel, and Travis Wheeler. Both Gutenkunst and Masel have served on the CMMBS Executive Committee for years, Masel since the program's founding in 2009, Gutenkunst since 2017, and Secomb remains on the Executive Committee. This structural change deliberately distributes leadership responsibilities, broadens the program's disciplinary reach, and reduces organizational fragility, while preserving full continuity of program expertise and institutional memory.
- The core coursework has been completely revised to be accessible to those whose prior mathematical training is limited to contemporary biology undergraduate degree requirements (introductory statistics and first-semester calculus) and to explicitly incorporate statistical and machine learning approaches to modeling. (The prior core coursework required second-semester calculus and recommended differential equations, which is beyond the mathematical training of many new biology PhD students.) A writing accountability group has been added to enhance trainee productivity and promote cohort cohesion.
- At onboarding, both trainees and their advisors now complete a two-hour workshop through the University of Arizona's MENTOR Institute. The IDP review cycle has been formalized to three reviews per year. The structured annual review now follows a three-step sequence in which PI Masel solicits input from course instructors, meets separately with the trainee, and meets separately with the advisor. Explicit conflict resolution procedures, with a staged escalation pathway from PI Masel through departmental leadership, are communicated to all trainees at onboarding and revisited at each annual review.
- The Spring QBC semester now opens with a professional networking session that includes required LinkedIn profile setup and structured preparation for informational interviews with visiting alumni. A formal industry advisory board, with initial membership drawn from CMMBS alumni and contacts at organizations including Thermo Fisher Scientific, Bristol-Myers Squibb, and the Critical Path Institute, is being established to provide annual feedback on curriculum and emerging workforce needs.
- The program has partnered with CIMER to conduct external evaluation using a longitudinal, comparison-group design. Students in BMCB and Applied Mathematics who did not receive the CMMBS intervention serve as local controls, and outcomes are benchmarked against CIMER's national

training program data. This design supports attribution of observed outcomes to CMMBS specifically rather than to ordinary Ph.D. development. Full details are provided in Section J.

A4. Relationship to Other Training Programs

The proposed program has several distinctive features compared to other existing NIGMS Predoctoral Training Programs in the area of Bioinformatics and Computational Biology:

1. It is strongly linked to a leading interdisciplinary graduate Program in Applied Mathematics and emphasizes modeling (both mathematical and machine learning) across all scales of biology.
2. It builds on more than 25 years of experience in research training at the interface between mathematics and biology.
3. It exploits an academic environment at the U of A where collaborations across disciplinary boundaries are fostered by many mechanisms, including a well-developed system of interdisciplinary graduate programs, the close proximity of the College of Medicine to the main campus, and a campus-wide atmosphere of cooperation and collegiality.
4. It employs novel training paradigms that develop the ability to communicate and collaborate across traditionally separate realms of biological and mathematical graduate programs.

Table PP.1 Trainees over entire CMMBS history

Graduate Program	Trainee-Years
Applied Mathematics	22
Biomedical Engineering	16
Biochemistry and Molecular & Cellular Biology	10
Physics	8
Ecology & Evolutionary Biology	8
Genetics	4
Statistics	4
Optical Sciences	3
Chemistry and Biochemistry	3
Physiological Sciences	2
Biosystems Engineering	2
Mathematics	2
School of Information	1

The CMMBS program draws from multiple Ph.D. programs at the U of A, including Applied Mathematics and Biochemistry and Molecular & Cellular Biology (**Table PP.1**), along with Physics, Computer Science, Pharmacy Practice and Science, and Ecology and Evolutionary Biology. Together, well over 100 training-grant eligible students matriculate to these programs each year (**Table 6A**, note that Biochemistry and Molecular & Cellular Biology recruits students through the larger Arizona Biomedical and Biological Sciences umbrella program). Of these students, at least 20% have interests in modeling approaches to biological problems, yielding at least 20 potential applicants. In the previous 15 years of the program, all available positions were filled by highly qualified applicants every year, and we have consistently had to turn away other well-qualified applicants due to the limited number of positions available. In the next grant period, we will vigorously expand recruitment into the medical campus at the U of A, which we expect to yield even more qualified applicants.

There are currently two other NIGMS-T32 programs at the U of A:

- The **Graduate Training in Biochemistry and Molecular Biology** T32 (GM136536; PI: Dr Andrew Capaldi) trains predoctoral students, equipping them with a broad understanding of modern approaches in Biochemistry and Molecular, Cellular, and Systems Biology. Ph.D. students in the Genetics and BMCB graduate programs are able to apply to both CMMBS and this program, but since both target funding in Years 2 and 3 of their Ph.Ds, they can accept only one. We expect students with more quantitative research interests to choose CMMBS, as has happened for the predecessor programs of both. Overlapping eligibility does not arise for Ph.D. students in the other 11 programs listed in Table PP.1, which CMMBS has historically drawn from. Nor is there overlap for other Ph.D. programs within the medical and public health schools that we have recruited training faculty from, with the aim of broadening the disciplinary backgrounds of future trainees.

- The **Interdisciplinary Training of Future Physician Scientists** T32 (GM141830; PI: Dr Todd Vanderah) trains MD-PhD students. This T32 is oriented towards training physician-scientists who draw from clinical experience in developing biomedical hypotheses. Its themes do not overlap with those of CMMBS.

A5. Administrative Structure

CMMBS is led by three co-PIs, Ryan Gutenkunst, Joanna Masel, and Travis Wheeler, who each commit 5% effort, supported by an Executive Committee of four additional training faculty. The leadership team is intentionally small so that every responsibility has a named owner and accountability is clear.

Responsibilities are distributed by area of expertise. PI Gutenkunst manages the administrative aspects of the program, including trainee appointments, NIH reporting, and institutional liaison communications. PI Masel oversees trainee mentoring and progress, conducting the structured IDP review cycle and annual mentor evaluations, as described in Program Oversight of Training and Mentoring. PI Wheeler leads recruitment from doctoral programs within the University of Arizona Health Sciences campus, expanding the program's disciplinary reach. All three PIs share responsibility for curriculum development, running the Quantitative Biology Colloquium, and student recruitment. Disagreements among PIs are resolved by majority vote; matters of broader program significance are brought to the full Executive Committee.

The Executive Committee consists of four additional training faculty with complementary expertise: Prof. Timothy Secomb, who served as CMMBS Program Director for 15 years and provides continuity of institutional memory; Prof. Joseph Watkins, founding coordinator of the Quantitative Biology Colloquium; Prof. Vignesh Subbian, who brings expertise in training program administration and currently is Interim Director of the BIO5 Institute; and Prof. Megha Padi, who contributes expertise in cancer bioinformatics and co-teaches a core curriculum course. The Executive Committee meets at least annually to review program composition, trainee progress, and evaluation outcomes, and is convened as needed to address escalated mentoring concerns or program policy decisions.

Should any PI become unable to continue in their leadership role, the remaining PIs and the Executive Committee will convene to identify a replacement from among the training faculty. Several training faculty have expressed interest in a leadership role should a vacancy arise, including Professor Laura Miller, an established CMMBS faculty member with relevant administrative and mentoring experience. Full descriptions of each PI's qualifications, time commitment, and mentoring background are provided in the Training Program Directors section.

B.Curriculum and Overall Training Plan

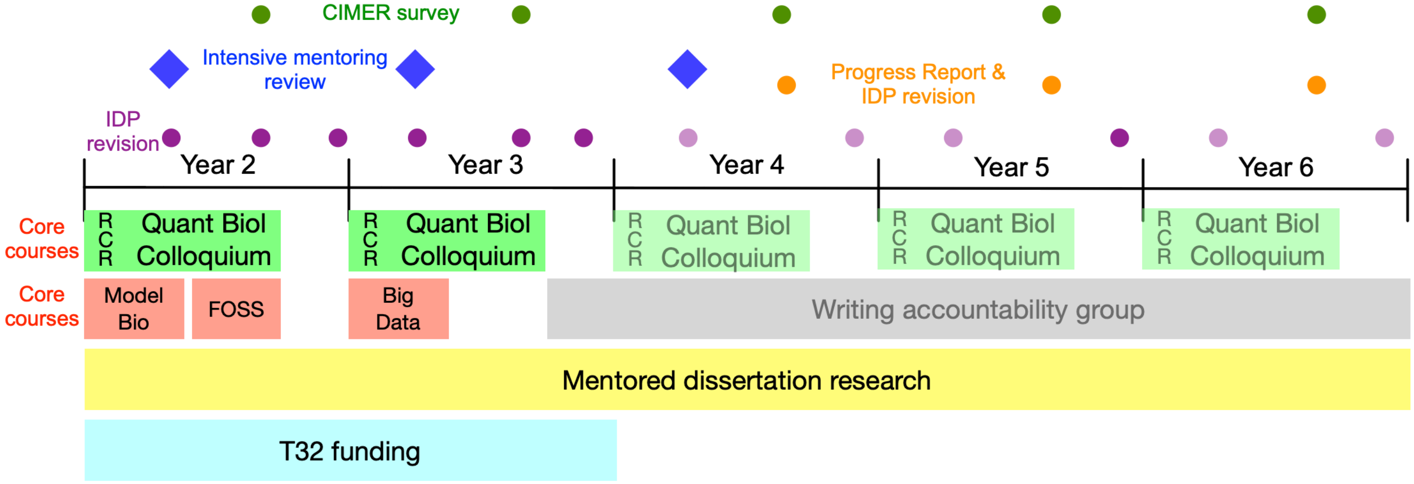

Figure PP.1: Timeline of training program and student monitoring. Outside the funding period, components in lighter shades are encouraged but not required. Didactic training in years 2 and 3 includes the Quantitative Biology Colloquium, Introduction to Modeling in Biology, Foundational Open Science Skills (FOSS), and Big Data in Molecular Biology and Biomedicine.

The CMMBS curriculum is designed to develop all three categories of skills that NIGMS identifies as essential for careers in the biomedical research workforce (**Figure PP.1**). Trainees are funded in years 2 and 3 or years 3 and 4 of their predoctoral studies, when participation in all training components is required. Technical skills, including mathematical modeling, statistical inference, machine learning, and computational implementation, are developed through the Introduction to Modeling in Biology course, the Big Data in Molecular Biology and Biomedicine course, and the Foundational Open Science Skills workshop. Operational skills, including rigorous experimental design, critical reading of the literature, data interpretation, reproducible analysis, and open science practices, are built through these same courses and workshop. Professional skills, including science communication, teamwork across disciplinary boundaries, career planning, time and project management, and responsible conduct of research, are developed through the structured elements of the Quantitative Biology Colloquium, including professional development sessions run by the U of A Graduate Center and alumni engagement activities. Following the funded period, scholars are encouraged, but not required, to continue attending the Quantitative Biology Colloquium and to join the writing accountability group, which maintains cohort identity and provides structured peer support that improves writing productivity through regular social commitment, shared strategies, and mutual feedback. Students engage in dissertation research with a CMMBS mentor throughout and after the funded period.

The academic activities specific to this training program complement those of the home graduate program, with emphasis on the specific objectives listed above. They are:

- The weekly **Quantitative Biology Colloquium** that has been running for more than 25 years and presents research at the interface between theoretical and biological disciplines.
- The one-semester course **Introduction to Modeling in Biology**.
- The one-semester course **Big Data in Molecular Biology and Biomedicine**.
- The 12-week workshop **Foundational Open Science Skills**.
- The **Writing Accountability Group**.

To avoid extending the time to completion of the doctoral degree, the number of required courses for the training program has been kept to a minimum. All the participating graduate programs require elective courses, which students can fulfill through the courses listed above.

Trainee Prerequisites and Accommodation of Varied Preparation. Eligibility for CMMBS requires successful completion of one semester of calculus and one semester of introductory statistics. These prerequisites are met by students entering biology-oriented doctoral programs at the University of Arizona, ensuring that CMMBS is accessible to trainees who arrive with standard biology undergraduate preparation. Students entering from quantitatively-oriented programs (Applied Mathematics, Physics, Statistics) will exceed these requirements substantially.

The curriculum is designed for the range of mathematics and computation preparation among trainees. In the Introduction to Modeling in Biology course, necessary mathematics is taught just-in-time as each modeling paper is encountered, so that students without differential equations or linear algebra can engage with the biological content without first completing formal prerequisites. The course's emphasis on high-level "input-output" analysis, understanding what a model does and what biological insight it provides, independent of implementation details, allows biology students to participate fully while quantitative students develop a mode of thinking about models that differs from their prior training. In the Big Data course, the AI-first programming approach (in which students are encouraged to use large language model tools to generate code and focus their own effort on testing and interpretation) substantially lowers the programming barrier for trainees without a strong computational background. Quantitative trainees, for whom machine learning methods may be familiar, engage with biological applications and data types that are new to them. The QBC's interdisciplinary group presentations are explicitly structured so that trainees with biological and quantitative backgrounds must collaborate and each contribute from their area of strength, reinforcing cross-disciplinary confidence in both directions. The net effect is a program in which preparation asymmetry between trainees is a feature rather than a problem: the two populations teach each other.

B1. Quantitative Biology Colloquium

The QBC is the centerpiece of the training program, taken throughout all four funded semesters, and includes several innovative features. It is coordinated by Dr. Joseph Watkins, a member of the training faculty and executive committee, along with the three PIs. A 2-unit course, it is required during each semester of funding. With presentations by students, faculty and visiting speakers, it promotes dialog between trainees and faculty with primarily mathematical or computational backgrounds and those with strong biological training. To address the objectives above, several elements comprise each two-semester QBC sequence:

- **Responsible conduct of research:** Each Fall semester begins with four weeks devoted to responsible conduct of research (RCR), led by PI Gutenkunst. For the first three sessions, trainees work in small groups to investigate and present prior well-documented cases of research misconduct, followed by discussion of the ethical issues involved. For the fourth session, trainees work with their mentors to understand responsible publication and authorship, which is pertinent for CMMBS because norms differ between biological and quantitative fields. For full details on these sessions and other aspects of RCR training, see the Plan for Instruction in the Responsible Conduct of Research.
- **Professional development:** Two sessions (one in Fall and one in Spring) are led by the University of Arizona Graduate Center's team of specialists, focusing on professional management and career exploration. The Fall management session covers time management, stress management, team dynamics and conflict management, project management, and financial management. Financial management includes identifying relevant grants and fellowships, as well as managing personal finances and educational debt. The Spring career exploration session guides trainees in exploring multiple career options aligned with their values, skills, interests, identities, preferences, and goals; developing skills and competencies needed for those career pathways; and articulating transferable skills and competencies. This session also includes skills in informational interviewing, to prepare students to interact with invited CMMBS alumni.
- **Deeper exploration of topics**: For the remainder of the Fall semester, trainees are divided into three groups, each devoted to a different topic over a block of three weeks. These groups deliberately mix students with primarily biological and those with primarily quantitative training. They work as a group with a faculty expert in that area to present a sequence of talks covering the background and current status of the topic and providing a critique of work in the field. Over the course of the three-week block, **students with different scientific backgrounds are required to communicate and collaborate in developing and coordinating their presentations.** We consider this one of the most valuable aspects of the training program. Previous topics covered include stochastic modeling of biochemical kinetics, approaches for modeling epidemics, and an introduction to neurobiology and associated models.
- **Scientific presentations on a wide range of topics**: During the Spring semester, CMMBS faculty mentors, student researchers, and visiting speakers present on their current research or more broadly on emerging topics that span modeling and biology. This colloquium format addresses the objective of broadening trainees' knowledge of multiple areas of biomedical science in which theoretical approaches play a significant role and helps inform trainees of new techniques and theoretical issues emerging across a range of scientific domains.
- **Connections with CMMBS alumni:** Also during the Spring semester, former program alumni (dating back to 1995) are invited to visit and give presentations in the QBC, and to meet informally with the current trainees to discuss their careers and experiences. Former trainees represent a wide range of career paths across academia and industry. For example in academia, Kathleen Lasick is an Assistant Professor of Biology and Brendan Fry and Nicholas Lytal are Professor and Assistant Professor of Mathematics and Statistics. Now working in industry, David Lyttle is Principal Bioinformatics Scientist at Evvy, Brian Anderson is Senior Principal Scientist at Bristol Myers Squibb, and Brian Hong is a Software Engineer at Mathworks. Interactions between alumni such as these and current trainees help to ensure that trainees are exposed to a variety of research careers and are aware of the range of opportunities available.

Cohort Building. The QBC develops trainees' cross-disciplinary expertise and career readiness while fostering a strong, collaborative cohort community. The QBC meets weekly throughout the funded period, giving trainees sustained contact with peers from complementary disciplinary backgrounds. The Fall group presentations, in which trainees with biological and quantitative training must collaborate to prepare and deliver a coordinated sequence of talks, are particularly effective at building mutual respect across disciplinary lines and developing each trainee's confidence in contributing outside their primary area of expertise. Professional development sessions address science identity and research self-efficacy directly, including attention to imposter syndrome and the particular challenges faced by trainees who enter quantitative programs from biology backgrounds, or biology programs from mathematics backgrounds. The Spring alumni visits reinforce a sense of belonging in the broader CMMBS community by demonstrating that prior trainees, including those who felt uncertain about the quantitative or biological aspects of their training, went on to productive careers. CIMER survey data, collected annually, tracks trainee self-perceptions of research self-efficacy, science identity, and psychological safety, allowing the program to identify and respond to cohort-level concerns as they develop.

The flexible format of the QBC enables us to tailor it to students' interests and needs. For instance, if a student is leaning towards a particular research direction, we frequently schedule that as one of the topics for the semester. In many cases, interactions initiated in the QBC have led to students developing a new focus in their research that has become a dissertation topic.

B2. Introduction to Modeling in Biology

Introduction to Modeling in Biology is a 3-credit course that trains PhD students to read modeling papers effectively and critically, assuming only familiarity with calculus concepts. The focus is on high level "input→black box→output" analysis approaches that enable students to grasp the key insights from and limitations of a model without getting lost in technical details. Biology students learn how biological questions can be addressed by models, even if implementation details in those models exceed their mathematical expertise. Quantitative students learn to look beyond implementation details to understand how models are applied in particular ways to a variety of biological problems.

The course covers seminal papers that are carefully chosen to span diverse biological topics, modeling philosophies, and mathematical techniques. Students receive reading guides that also structure in-class discussion. Necessary math is taught just-in-time. Assessment includes in-class quizzes after class review of each paper, Mathematica programming homework assignments that reproduce key figures, and a final take-home exam to assess student engagement with a paper without the benefit of in-class discussions.

The course strongly emphasizes the scientific method, through the lens of the wide range of roles that modeling can play within the practice of science. Modeling aims are not limited to generating testable predictions capable of distinguishing among competing hypotheses (e.g. Luria & Delbruck 1943), which is the dominant modality for models in physics but not in biology. Modeling aims also include disproof of principle, i.e. formally proving that a verbal hypothesis is either internally inconsistent or inconsistent with existing stylized facts. A great example of this is Hardy 1908, the original but greatly misunderstood "Hardy-Weinberg" paper, whose model is taught in introductory biology as though its purpose were to make testable predictions, rather than in its historical context as a refutation of Yule's arguments about the dynamics of dominant alleles. Other modeling papers provide the converse result, i.e. an unexpected proof of a surprising principle. Many modeling projects might be initiated with the aim of exploration, but are ultimately written in terms of generating prediction, proof of principle, or disproof of principle. Students learn to categorize models both in terms of the mathematical objects involved, and in terms of the kind of biological insight obtained.

Students also learn basic programming in Mathematica (with access provided through U of A's High Performance Computing [HPC] resources in the Research Data Center) to reproduce key figures. Students learn how to program in ways that bypass technical/procedural mathematical training, emphasizing instead conceptual grasp of the diversity of mathematical objects and what can be done with them. This prepares students well for the next step: designing and implementing their own models as part of their research.

Co-taught by Joanna Masel and Anna Dornhaus, this course has been offered by the Department of Ecology & Evolutionary Biology since 2014. Table PP.2 below shows how the course uses seminal papers to cover multiple topics in biology, mathematics, and diverse roles for the models within the scientific method.

Table PP.2: Seminal Readings, Introduction to Modeling in Biology

Paper	Biology	Scientific Method	Math/Model
Hardy 1908	Hardy-Weinberg equilibrium	Disproof of principle	Algebra
Schmid-Hempel et al. 1985	Optimal foraging	Generate predictions	Algebra
Masel et al. 1999	Prions, r vs R_0	Failure to disprove principle	System ODEs, exponentials
Huang & Ferrell 1996	Cooperative reactions	Proof of principle	Hill coefficients
Barabasi & Albert 1999	Preferential attachment	Proof of principle	Networks
Maynard Smith & Price 1973	Limited war	Proof of principle	ESS, simulation
Stainforth et al. 2005	Climate	Forecasting	Parameter uncertainty
Luria & Delbruck 1943	Spontaneous mutation, fluctuation test	Generate predictions	Probability distributions

Armitage & Doll 1954	Multi-hit model of cancer	Generate predictions	Log-log plots
Fussmann & Blasius 2004	Lotka-Volterra cycles, paradox of enrichment	Proof of principle	Dynamical systems, sensitivity to model structure

B3. Big Data in Molecular Biology and Biomedicine

Big Data in Molecular Biology and Biomedicine is a 3-credit course that gives students a broad and practical introduction to statistical and machine learning models applicable to modern large-scale biomedical data sets. The course assumes familiarity with introductory statistics and calculus. Throughout the course, students analyze a wide variety of real-world data, including human genotypes, RNA-seq counts, biomedical images, and electronic medical records. The course emphasizes not the bioinformatics necessary to generate each type of data, but rather **analysis approaches that can be applied to many types of data**. As a mathematical foundation, the course emphasizes the concept of a probability distribution, be it the high-dimensional distribution of data or the one-dimensional distribution of a test statistic.

Through both extensive in-class exercises and homework assignments, students use R in the RStudio environment to implement their analyses. To teach a key element of reproducible analysis, students are instructed how to write R Markdown notebooks, and they must return all assignments as notebooks. In teaching R and R Markdown, an “AI first” programming paradigm is adopted, in which students are encouraged to use the power of large language model tools to write their code. This alleviates many of the syntactical challenges of coding and drives students to focus on testing to assess code correctness.

An abbreviated schedule of topics includes:

- Dimension Reduction: Principal Component Analysis, tSNE, and UMAP
- Hierarchical Clustering
- Maximum Likelihood: Mixture Models
- Hypothesis Testing: Gene Set Enrichment Analysis, Multiple Testing, False Discovery Rate
- Introduction to Machine Learning: k-Nearest-Neighbors, Cross Validation
- Regression: Linear, Logistic, Penalized
- Classification: Decision Trees, Random Forests
- Neural Networks
- Convolutional Neural Networks: Transfer Learning, Interpretability
- Deep Learning: Autoencoders, Embedding, Transformers, AlphaFold2, Protein Language Models
- Big Data Ethics

Two application projects serve as capstones for student learning. As a mid-semester project, students analyze publicly available cancer genome data, using them to cluster patients into tumor subtypes and predict prognosis and drug sensitivity. As an end-of-semester project, students use ClinVar training data to build a variant effect predictor and assess the utility of including embeddings from state-of-the-art protein language models as features.

The course was originally developed by PI Gutenkunst and has subsequently been developed and co-taught by CMMBS faculty mentors Megha Padi (an expert in cancer ‘omics and network modeling) and Claire McWhite (an expert in protein language models). The course has proven accessible and impactful for both Ph.D. students with biology backgrounds, for whom the machine learning is new, and for students with quantitative backgrounds, for whom the biological applications are new.

B4. Foundational Open Science Skills

The 12-week Foundational Open Science Skills (FOSS) workshop is a hands-on modular training program that teaches core competencies in open science, including command-line computing, version control with Git and GitHub, FAIR data principles, data hygiene and governance, reproducible workflows, and cloud-based analysis using CyVerse’s national cyberinfrastructure. While technically rigorous, the curriculum is designed for accessibility, welcoming participants with little to no coding experience and guiding them toward proficiency in modern cloud-based, computational research practices. The curriculum was developed by CyVerse and the U of A Data Science Institute to equip early-career scientists with the practical skills required to thrive in today’s collaborative, computational research environment. FOSS has been offered since 2019, recently completing its 10th offering. More than 500 new PIs, researchers, educators, students, and others have gained these foundational skills.

The schedule of topics includes:

- Data management, including data management plans
- Documentation and communication, including GitHub Pages websites
- Command-line computing
- Version control with Git and GitHub
- Software environments
- Running and building containers
- Reproducible research with AI
- Remote computing with CyVerse and HPC

These topics are presented in weekly 1.5-hour sessions, including hands-on activities and live instructors. For CMMBS trainees, the lessons from each week are reinforced in discussions during the Quantitative Biology Colloquium. During the FOSS workshop students work on a capstone project, applying these concepts to their own research. For example, a student might move their existing code into a documented GitHub repository or develop a tutorial for a tool they use in their own work. The FOSS workshop culminates in 10-minute presentations about these capstone projects.

B5. Writing accountability group

Beginning in the summer of the second funded year, students join a CMMBS writing accountability group that meets every two weeks, led by PI Gutenkunst. During each group meeting, all attendees publicly set writing goals and share progress on the previous meeting’s goals. Example goals include drafting an introduction to an article, preparing a figure, outlining a methods section, or drafting the conclusion of a dissertation. The group provides social accountability for a regular and consistent writing process, reflecting the principle that writing is a critical form of structured thinking that should occur throughout a research project, not just at its conclusion. Because these groups encompass all active CMMBS trainees, students in the middle of their graduate studies engage with those in the final stages of dissertation writing, allowing early-stage trainees to learn hard-won practical lessons about the writing and publication process. Beyond research productivity, these groups preserve CMMBS cohort identity and professional networks that persist beyond the funded period.

B6. Responsible conduct and reproducibility training

Responsible conduct of research and methods for enhancing reproducibility are not confined to a single curricular moment but are reinforced across multiple program components. The QBC devotes four sessions each fall to RCR, with case-study discussions of prior research misconduct followed by a session on responsible publication and authorship norms, the latter addressed specifically because norms differ between biological and quantitative fields and trainees must be prepared to navigate both. The Introduction to Modeling in Biology course addresses rigor through its emphasis on understanding model assumptions and limitations, and on the distinctive roles models play within the scientific method (prediction, disproof of principle, proof of principle), training students to evaluate claims made by models critically rather than accepting outputs uncritically. The Big Data course requires all analyses to be submitted as R Markdown notebooks, directly instantiating reproducible workflow practices, and includes a dedicated unit on big data ethics. The Foundational Open Science Skills workshop covers FAIR data principles, data governance, version control, and containerized reproducible workflows. Faculty mentors are expected to reinforce these standards in their own laboratories throughout the trainee's graduate career, as described in Program Oversight of Training and Mentoring. Full details of the RCR curriculum are provided in the Plan for Instruction in the Responsible Conduct of Research, and the reproducibility curriculum is detailed in the Plan for Instruction in Methods for Enhancing Reproducibility.

B7. Mentor-supervised research

All CMMBS trainees conduct dissertation research under the primary supervision of a CMMBS faculty mentor throughout and beyond the funded period. All dissertation research in CMMBS includes mathematical or computational components. The U of A Research Computing environment, described in Resources, provides adequate capacity for the simulation, inference, and machine learning workloads typical of CMMBS dissertations.

Research progress is monitored through the IDP and IMP framework described in Section D. The three-times-per-year IDP review cycle tracks dissertation milestones, development of independent modeling skills, and progress toward publication. The annual review process supplements advisor reporting by soliciting input from QBC and core course instructors, who observe each trainee's technical development from a perspective that complements the primary mentor's view.

B8. Training resources beyond the funded period

With support from the CMMBS program, the University of Arizona Graduate Center’s team of specialists provides a variety of resources to CMMBS trainees both during and after their funded period is complete. Examples include the Summer Fellowship Application Development Program, customized support for NIH F30/F31 applicants, and the Grad Slam, a presentation competition that builds public speaking proficiency. These are in addition to any resources provided by trainees’ home graduate programs and advisors.

B9. Sample Training Programs

The proposed program draws from multiple Ph.D. programs at the U of A, including Applied Mathematics, Biochemistry and Molecular & Cellular Biology, Statistics, Physics, Chemistry and Biochemistry, Biomedical Engineering, and Ecology and Evolutionary Biology (Table 1). This breadth of contributing graduate programs is a strength of the program, as it exposes trainees to varied and complementary areas of expertise they might not otherwise encounter. This is essential for equipping trainees with the skills to operate at the intersection of biology and quantitative theoretical methods and to communicate and collaborate effectively across traditional disciplinary boundaries. Space does not allow for a detailed description here of all the participating Ph.D. programs. Therefore, to illustrate our two core areas, biology and mathematics, below we provide an overview and general plan for study for two programs: Biochemistry and Molecular & Cellular Biology and Applied Mathematics, which exemplify graduate training at the U of A.

Biochemistry and Molecular & Cellular Biology: Overview and Plan of Study

The BMCB Ph.D. program is representative of biology-oriented doctoral programs at the University of Arizona, including Genetics, Ecology and Evolutionary Biology, and Physiological Sciences. We describe it here to illustrate the preparation and timeline typical of CMMBS trainees from biological backgrounds. It draws from 50 faculty across the Colleges of Science and Medicine, with an explicit systems biology emphasis added in 2010. The program is designed to develop both deep expertise in a specific area of molecular and cellular biology and broad understanding of how molecular components assemble into functional biological systems.

Core coursework in the first two years builds quantitative and analytical foundations progressively. In the first year, students take Cell Systems (MCB 572), which examines information flow through signaling networks and how cellular networks evolve under mutation, and Genetic and Molecular Networks (MCB 546), which approaches cell signaling and gene regulation from integrated biochemical, genetic, and systems biology perspectives. In the second year, students take a scientific communication course (MCB 575), and a computational methods course (MCB 585), which covers R-based statistical analysis, quantitative image analysis, RNA-seq, basic machine learning, and emerging multi-omic approaches. Foundational skills are further developed through participation in the BMCB journal club and seminar series.

Students complete at least three laboratory rotations in their first year before selecting a dissertation advisor and choose elective courses each semester from areas including cancer biology, immunology, plant biology, and computational biology. Computational biology is offered as a structured elective area, and we anticipate that most BMCB students joining CMMBS will follow it. Students typically select their dissertation advisor and supervisory committee by the end of their second year and complete written and oral comprehensive examinations before beginning focused dissertation work. This structure of first-year rotations, advisor selection by the end of year two, and comprehensive examinations before focused dissertation work is common across biology-oriented Ph.D. programs at the University of Arizona. CMMBS trainees from these programs therefore arrive with a consistent and predictable preparation level, regardless of their specific home program. All entering Ph.D. students receive a five-year financial support commitment, independent of training grant funding.

BMCB students typically enter CMMBS in their second year, after rotations and advisor selection are complete but before comprehensive examinations, making this an ideal point to layer in the cross-disciplinary training that CMMBS provides.

Applied Mathematics: Overview and Student Plan of Study

The Program in Applied Mathematics is representative of quantitatively-oriented doctoral programs at the University of Arizona, including statistics and physics, and we describe it here to illustrate the preparation and timeline typical of CMMBS trainees from mathematical and computational backgrounds. Founded in 1976, the program draws faculty from departments across the mathematical, physical, biological, and engineering sciences and has established a national reputation for rigorous interdisciplinary training.

Core coursework in the first year covers Methods of Applied Mathematics (APPL 581), Advanced Applied Mathematics Theory (APPL 584), and Computational and Algorithmic Applied Mathematics (APPL 589), providing integrated training in analytical, theoretical, and numerical methods. Each course is taught by a rotating group of three to five faculty, ensuring breadth of perspective. Concurrently, students develop research and professional skills through a structured seminar and journal club sequence (Math 586a/b, Math 599) spanning the first year and a half, which includes faculty research presentations, independent study projects, and progressively more substantial written and oral research reports. Students pass a written qualifying examination at the end of their first year before selecting a research advisor, typically in their second year after extensive faculty exposure through coursework and seminars.

After the first year, students choose advanced electives from within mathematics and across other departments, including the CMMBS curriculum. This structure -- intensive first-year coursework, qualifying examination, and advisor selection in the second year -- is common across quantitatively-oriented Ph.D. programs at the University of Arizona. CMMBS trainees from these programs therefore arrive with strong formal mathematical foundations and meaningful exposure to biological applications before entering the program.

Applied Mathematics students typically enter CMMBS in their second or third year, after qualifying examinations and initial advisor identification, making this a natural point to formalize their cross-disciplinary training in biological modeling.

B10. Illustrative Trainee Trajectories

To illustrate how CMMBS activities integrate with home program requirements, consider two representative cases.

A student entering from BMCB completes three laboratory rotations in his first year and selects a dissertation advisor working on quantitative imaging of single-cell signaling dynamics. He joins CMMBS in Year 2, after advisor selection and before his comprehensive examinations. The Introduction to Modeling in Biology course, taken in the Fall semester of Year 2, satisfies a BMCB elective requirement and prepares him to read and critically evaluate the mathematical modeling literature central to his dissertation. The Big Data course, taken in Spring of Year 2, introduces statistical and machine learning approaches he applies directly to image analysis in his research. In the QBC, he is paired in a Fall group presentation with Applied Mathematics and Physics students, an experience that builds his confidence explaining biological systems to quantitative audiences. His Individualized Mentorship Plan (IMP) identifies his primary advisor as mentor and a CMMBS faculty member in Applied Mathematics as co-mentor for quantitative methods development. PI Masel's annual review after his first funded year, incorporating feedback from QBC instructors and his advisor, confirms that his modeling skills have advanced to the point that he is leading quantitative analysis independently in his dissertation work.

A student entering from Applied Mathematics completes the program's first-year core sequence and passes the written qualifying examination before identifying a dissertation advisor in Year 2 whose work applies dynamical systems models to cell signaling. She joins CMMBS in Year 3, after advisor selection and initial dissertation scoping are complete. Her CMMBS courses, Introduction to Modeling in Biology and Big Data in Molecular Biology and Biomedicine, count toward her Applied Mathematics elective requirement and require no additional time to degree. She participates in QBC throughout her two funded years, completing an RCR sequence in the Fall and presenting in a cross-disciplinary group alongside BMCB and Biomedical Engineering peers on stochastic modeling of biochemical kinetics, an area directly adjacent to her dissertation. Her IMP, drafted with PI Masel and her dissertation advisor, identifies a CMMBS faculty member in Molecular and Cellular Biology as a co-mentor who provides biological interpretation guidance for her modeling work. Annual

IDP reviews track her progress toward modeling independence and her growing ability to communicate results to biological audiences, a transferable skill she documents through a QBC research presentation in Year 4.

In both cases, CMMBS activities are layered onto existing program requirements rather than added on top of them, and co-mentorship arrangements are used to bridge the disciplinary gap that each trainee's home program alone cannot close.

C. Career Development

CMMBS makes career outcome data available to prospective applicants before they apply. The program website lists current and former trainees with links to their LinkedIn profiles, and includes aggregate data on PhD completion rates, time to degree, and post-degree positions. Prospective applicants who contact the program are directed to this resource and are encouraged to reach out directly to recent alumni. The QBC's Spring alumni visit series is also open to non-funded students, including prospective applicants who have not yet been selected for the program, giving them direct access to alumni who can speak candidly about their training experience and career trajectories. This transparency about outcomes is consistent with NIGMS expectations and helps ensure that applicants self-select based on realistic understanding of where the program leads.

Alumni connections. The Quantitative Biology Colloquium (QBC) is a key tool in preparing trainees for industry, academic, and government careers. During the Spring semester, training grant funding is used to focus the QBC on in-person visits of program alumni and professionals blending modeling and biology in their careers.

To enable trainees to take maximum advantage of these campus visits, in the revised program the Spring semester of the QBC begins with a session on professional networking strategies. Trainees are required to set up a LinkedIn profile, and the program maintains a webpage that lists program alumni with links to their LinkedIn pages, facilitating connection. During each campus visit, trainees conduct informational interviews with the speakers, broadening their career horizons and opening doors toward potential later internships, which are encouraged but not a required component of CMMBS. If an industry internship occurs during the funded period, the CMMBS PIs will ensure oversight by a U of A faculty member, who will be responsible for defining expectations, support structures, responsibilities, and time limits, while PI Gutenkunst will ensure compliance with NRSA and funding announcement appointment requirements. By targeting students in their second and third years, we are planting seeds for building relationships with industry; the fruition is likely to come in trainees’ final year of studies, when they more aggressively network with and make pitches to industry contacts.

Alumni from the predecessor programs now work for employers ranging from Thermo Fisher Scientific, Accelerate Diagnostics, Bristol-Myers Squibb, Hologic, Amgen, Applied Biomathematics, and a stealth-mode startup company. We have also identified prospective speakers from among additional U of A alumni, currently employed at organizations such as the Critical Path Institute and Raytheon, both located in Tucson, as well as at Empress Therapeutics, GSK, Regeneron, Illumina, Bayer Crop Science, Delfi Diagnostics, and Ancestry.com. PI Masel has experience in the startup sector, which informs the mentorship that she offers CMMBS trainees on their career paths.

In addition, we leverage the strong connections between the University of Arizona and National Laboratories. PI Gutenkunst and the former Director of the Applied Mathematics program are both alumni of Los Alamos National Laboratory (LANL). The annual Arizona-Los Alamos Days conference reinforces that connection, especially to the strong theoretical biology group at LANL.

Feedback from employers. The program actively engages a range of potential employers to ensure that trainees develop the skills, knowledge, and professional networks needed to transition into their chosen sectors. The Spring QBC alumni visits are structured as a two-way exchange: alumni provide trainees with career perspective, and the PIs use these interactions to solicit feedback from employers about the skills they value most and the gaps they observe in candidates from quantitative biology programs. This feedback directly informs program planning. Building on these relationships, the PIs are developing a formal industry advisory board, with initial membership drawn from CMMBS alumni and contacts at organizations including Thermo Fisher Scientific, Bristol-Myers Squibb, Amgen, the Critical Path Institute, and Los Alamos National Laboratory.

The advisory board will meet annually with the PIs to review training outcomes, identify emerging skill needs, and suggest adjustments to curriculum or career development activities.

Preparation for academic careers is primarily provided by each trainee’s advisor. To supplement that training, the QBC includes a session on grant and fellowship preparation, particularly opportunities that bridge mathematics, computation, and biology.

D. Program Oversight of Training and Mentoring

Effective mentorship requires both proactive support and structured accountability. The revised CMMBS program implements a tiered oversight system that provides intensive monitoring during the funded period, continued monitoring through degree completion, and clear escalation pathways when concerns arise (**Figure PP.1**).

Students are matched with their primary research advisor by their graduate programs before applying to CMMBS. (See section B.9 for examples of the matching process for a biological and a quantitative program.) In the event an applicant’s primary advisor is not a CMMBS mentor, the advisor will be evaluated for membership in CMMBS simultaneously with the applicant’s application.

Onboarding. Upon appointment to the program, both trainees and their faculty advisors/mentors complete a two-hour workshop through the University of Arizona's MENTOR Institute, which is based on the National Research Mentor Network and CIMER training frameworks. The workshop focuses on building effective mentoring skills for both parties, including individualized mentoring plans, setting expectations, and reviewing the mentoring relationship. Faculty who have recently completed an equivalent workshop may substitute the asynchronous online version. Following the workshop, each trainee drafts an Individualized Mentorship Plan (IMP), articulating expectations for the format and nature of the mentoring relationship, and an Individual Development Plan (IDP), defining concrete priorities for training customized to the trainee's research and career goals, including both a primary and an alternate career plan.

During the funded period (Years 1-2 of support). PI Masel oversees an intensive three-times-per-year IDP review cycle. Reviews occur at the end of the Fall semester, the end of the Spring semester, and during the summer. Each review requires at least one dedicated meeting between the trainee and their advisor, culminating in an updated IDP filed with PI Masel. Each Fall, PI Masel conducts a more intensive annual review consisting of three sequential steps:

- Step 1: PI Masel verbally solicits observations about the trainee from QBC and core course instructors and solicits a brief written update from the training faculty member.
- Step 2: PI Masel meets individually with the trainee to discuss progress and the mentoring relationship from the trainee's perspective.
- Step 3: PI Masel meets individually with the advisor to discuss current mentoring practices and reflect on opportunities for improvement.

This three-step process occurs once each Fall during the funded period, and once in the Fall immediately following the cessation of funding, for a total of three annual reviews per trainee funded for the two-year maximum.

After the funded period (through degree completion). Oversight continues at a reduced intensity for the remainder of each trainee's graduate career. Trainees and advisors are encouraged to maintain the three-times-per-year IDP review cycle, but only annual documentation is required and reviewed by the Executive Committee: a revised IDP, an optional IMP revision, and a projected timeline to graduation. For trainees whose projected time to degree exceeds six years, PI Masel resumes the full intensive annual review process described above.

CMMBS operates within a broader institutional structure that provides continuity of support for the duration of each trainee's graduate career, not only during the funded period. All participating graduate programs commit five years of financial support to incoming doctoral students through a combination of Teaching Assistantships, Research Assistantships, and Graduate College Fellowships. This commitment is independent of training grant funding and ensures that no trainee loses support at the conclusion of their CMMBS appointment. PI Gutenkunst, in his capacity as Department Head of Molecular and Cellular Biology, maintains direct institutional relationships with the graduate program directors from which most CMMBS trainees are drawn, providing a channel to identify and address cases where a department or program is not adequately supporting a trainee's continued progress. For trainees whose home program raises concerns about unsatisfactory progress or inadequate departmental support, PI Masel engages directly with the program director to understand the situation and, where appropriate, advocates for the trainee within the institutional structure. This combination of guaranteed institutional support and active PI oversight ensures that CMMBS trainees are not left without a support structure when their training grant appointment ends.

Conflict resolution. The program maintains explicit conflict resolution procedures for concerns arising within the training environment, including mentor-mentee, peer-peer, and faculty-trainee conflicts. The first point of contact for any trainee experiencing difficulty is PI Masel, who conducts the regular IDP reviews and annual mentoring assessments and is therefore positioned to detect concerns early. When a trainee raises a concern or when PI Masel observes one through the review process, the response follows a staged approach. For minor concerns, such as miscommunication about expectations or temporary disruptions to research progress, PI Masel works directly with the trainee and advisor to identify the issue and develop a remediation plan, documented in writing. For more serious concerns, including persistent mentoring deficiencies, interpersonal conflict, or a trainee's reluctance to escalate through normal channels, PI Masel convenes a meeting among all three PIs. If the concern involves a faculty mentor and is not resolved through direct discussion, PI Masel coordinates with the head of the trainee's home graduate program and the head of the mentor's department. Faculty who demonstrate persistent unacceptable mentoring practices are removed from the program following discussion among the PIs and Executive Committee. Trainees are informed of these procedures at onboarding and are reminded of them during each annual review.

E. Training Program Directors

Ryan Gutenkunst (PI), Ph.D. is a tenured full Professor and Department Head of Molecular and Cellular Biology, with 15 years of experience at the U of A. His undergraduate and graduate training was in physics, and his Ph.D. dissertation focused on analyzing differential equation models of complex systems biology networks. As a postdoctoral scholar, he trained in population genetics, using numerical solutions of partial differential equations to fit models of demographic history to contemporary patterns of genetic variation. The software he developed, dadi, is widely used by the population genetics community, and his group continues to develop and enhance it, emphasizing rigorous software development and enabling reproducible research. His research group focuses on computational and mathematical population genetics, with particular interest in inferring demographic history and natural selection and including recent applications to cancer. Methods developed in his group employ both classical models optimized by maximum likelihood and machine learning models. Prof. Gutenkunst's research has been funded by the NIH, NSF, and DARPA.

Dr. Gutenkunst has been involved with the CMMBS program since arriving at the U of A, and it was key to attracting him there. Soon after arriving, he taught a mini-course in stochastic simulation in the Quantitative Biology Colloquium. He has served on the Executive Committee for the CMMBS program since 2017, and he has served on the Executive Committee of the Genetics Graduate Interdisciplinary Program since 2020. He has had PhD students from both quantitative and biological programs: two from Applied Mathematics, one from Biostatistics, one from Genetics, and one from Ecology & Evolutionary Biology. Of his graduated PhD trainees, two are now faculty in biology departments, one is a Senior Scientist at Foundation Medicine, and one is now a research specialist at Canada's Michael Smith Genome Sciences Centre. In 2013 he was awarded the College of Science Distinguished Early-Career Teaching award for his development of an undergraduate course in quantitative biology that incorporated extensive active learning. In addition to teaching in computational biology, Gutenkunst has also taught a graduate course in scientific communication and run multiple journal clubs. Through his roles as Professor and Department Head, he has completed extensive mentorship and conflict resolution training at the U of A. He has served as a panelist evaluating NSF Graduate Research Fellowships applications five times and has mentored two winners. Through his work as Department Head, Prof. Gutenkunst has developed the contacts with U of A leadership necessary to assure support of the CMMBS program. The 5% time he devotes to the CMMBS program will count toward his university service commitment.

Joanna Masel (PI), D.Phil. is a tenured full Professor in the Department of Ecology & Evolutionary Biology, with 21 years of experience at the U of A. Her undergraduate training was in both mathematics and genetics, and her doctoral dissertation modeled the nucleic-acid free mechanism of prion replication and its implications for treatment strategies. As a postdoctoral scholar, she trained in population genetics and has since applied these techniques to understanding the robustness and evolvability of biological systems, improving methods of phylogenetic inference, examining somatic evolution during aging, characterizing forms of competition, and understanding the consequences of the surprisingly high rates of deleterious mutation experienced by human populations. In addition to mathematical modeling, her research group conducts bioinformatic data analyses on protein evolution and molecular error rates. She also pivoted to risk analysis research in support of digital contact tracing during the pandemic. Dr. Masel's research has been funded by the NIH, NSF, NASA, the John Templeton Foundation, and the Pew Charitable Trusts.

Dr. Masel has served on the Executive Committee of the CMMBS program since it began in 2009. She has taught two different mini-courses within the Quantitative Biology Colloquium, and three of her Ph.D. students have been CMMBS trainees. She also served on the Executive Committee of the Applied Math Graduate Interdisciplinary Program throughout 2009-2017 and has served since 2019 on the Executive Committee of the Genetics Graduate Interdisciplinary Program. She has had one Ph.D. student from Applied Mathematics, two from Genetics, two from Molecular & Cellular Biology, and two from Ecology & Evolutionary Biology. Of her four graduated PhD trainees, one is a bioinformatics scientist at Regeneron, one is scientific director for quantitative medicine at the Critical Path Institute, one is an engineer at Raytheon Missile Systems, and one is currently a postdoc. She has been the de facto advisor and mentor to three other Ph.D. students not at her institution, with one (Ph.D in Earth System Science at Stanford) now on an NSF PRFB funded postdoc, one (Ph.D. in Computer Science at University College London) a startup founder, and one (Ph.D in Applied Mathematics at University of Waterloo) employed by a thinktank for AI governance. Her teaching load includes the core Introduction to Modeling in Biology class. She has been a campus leader in the adoption of active learning techniques, has been involved in a variety of education initiatives, and was a finalist for an HHMI Professorship. The 5% time she devotes to the CMMBS program will count toward her university service commitment.

Travis Wheeler (PI), PhD, is a tenured Associate Professor in the Department of Pharmacy Practice and Science, with affiliations in Computer Science, the School of Information, Applied Math, and others. His undergraduate training was in evolutionary biology, and his doctorate was in Computer Science, with a dissertation focused on development of algorithms for biological sequence comparison and phylogeny inference. His postdoctoral research (with Sean Eddy) emphasized development of statistical models and algorithms for sequence homology detection and discovery/classification of transposable elements. His research group at Arizona designs algorithms, statistical models, and AI methods for problems ranging from annotation of biological sequences to drug discovery, precision medicine, and tracking animals in recorded video.

Dr. Wheeler served as Graduate Program Coordinator for the University of Montana Computer Science department from 2018 until moving to Arizona in 2022. At Montana he led expansion from a Master's-only program with three graduate students to a PhD-granting department with 18 graduate students (5 PhD). He has mentored seven PhD students (3 graduated), and 13 Master's students (11 graduated). His graduate students have moved into professional and academic roles at Fred Hutchinson Cancer Center, University of San Diego, University of Montana, the College of Wooster, the Genomics Institute at UC Santa Cruz, and Lawrence Berkeley National Labs. He co-leads an annual two-week computational biology workshop (CompBioAsia) in which nearly 40 graduate students are introduced to analysis of sequence data, protein structure prediction, protein-ligand affinity prediction, and molecular dynamics simulation. Dr. Wheeler guided the U of A's Health Sciences Bioinformatics Group since 2023 (shortly after arriving at Arizona) and his team provides methodological support for bioinformatics researchers developing precision medicine methods in the NIH-supported All of Us Research Program platform. He also acts as Director of two international community resources emphasizing acquisition, quality assurance, and analysis of key biological data: (1) TE Hub, a platform for international collaboration and communication among transposable element researchers; and (2) MDRepo, a petabyte-scale data warehouse for community-contributed molecular dynamics simulation data. The 5% time he devotes to the CMMBS program will count toward his university service commitment.

An Executive Committee, comprising four additional faculty with complementary expertise, provides specialized oversight that complements PI leadership by reviewing program composition, trainee progress, and evaluation outcomes at least annually, and convening as needed to address escalated mentoring concerns or program policy decisions.

Timothy Secomb (Executive Committee), PhD, is a tenured Professor in the Department of Physiology. Trained as an applied mathematician, his research focuses on the circulatory system, including blood flow and mass transport in the microcirculation, angiogenesis and structural remodeling of blood vessels, regulation of blood flow, cancer drug transport and pharmacodynamics, and cardiac mechanics. He has published more than 240 peer-reviewed journal articles and book chapters and mentored 18 PhD students. For 15 years, he was Program Director of CMMBS. He will continue to serve on the CMMBS Executive Committee, lending his extensive experience in mentoring and program administration.

Joseph Watkins (Executive Committee), PhD, is a tenured Professor in the Department of Mathematics. His doctoral training was in the theory of probability and stochastic processes, which he has applied extensively to population and clinical genetics. He has served as Chair of the Interdisciplinary Program in Statistics and Data Science and is now the founding Chair of the Data Sciences Academy. Established in 2020, the Data Sciences Academy is an umbrella organization designed to coordinate and facilitate activities across campus and in the community. He has served on the executive committee for earlier cycles of the CMMBS training grant and has played a leading role in organizing the Quantitative Biology Colloquium. Due to the centrality of probability, stochastic processes, and modern statistics, he is a regular collaborator across many fields, including bacteriology, biochemistry, biophysics, cancer biology, computer science, entomology, linguistics, and medicine. Thus, he frequently serves on comprehensive exam and final defense committees for the CMMBS trainees.

Vignesh Subbian (Executive Committee), PhD, is a tenured Associate Professor in the Department of Biomedical Engineering. His undergraduate training was in Electrical Engineering, and his PhD training was in Computer Science and Engineering. His research focuses on advancing the science of health systems through interdisciplinary research that integrates theories and methods from informatics and cognitive engineering. He studies clinical decision-making with emphasis on phenotyping, interpretability, and algorithmic fairness. He is the co-lead and MPI for the informatics services core within the NIH All of Us Research Program at UA & Banner Health. He is also program co-director for two undergraduate R25 research training programs: Place-based Health Informatics Research Education (PHIRE) and Strengthening Arizona's Biomedical Engineering Research Workforce (SABER). He was also recently named Interim Director of the Bio5 Institute.

Megha Padi (Executive Committee), PhD, is a tenured Associate Professor in the Department of Molecular and Cellular Biology. Her undergraduate and PhD training was in physics, and she shifted to biology as a postdoc working on network models of how disease can emerge from the complex interactions between genes and proteins. Her lab creates new network-based algorithms and performs wet-lab validation of their hypotheses in early-onset colorectal cancer and Merkel cell carcinoma. She regularly collaborates with cancer investigators to perform multi-omics data integration, and she formerly directed the Bioinformatics core service at the U of A Cancer Center.

F. Preceptors/Mentors

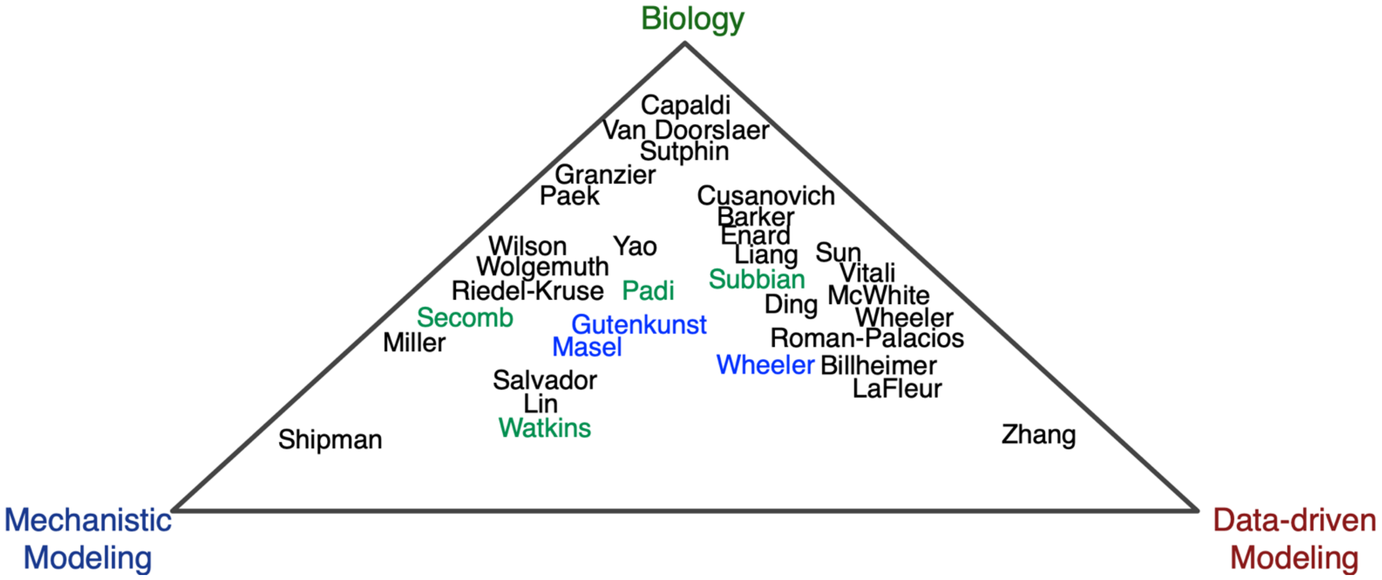

Figure PP.2: CMMBS mentors are qualitatively placed to illustrate the degree to which their research focuses on experimental biology, mechanistic mathematical modeling, or data-driven computational modeling. Program PIs are highlighted in blue and executive committee members in green.

The program faculty consists of 32 members (**Table 2; Figure PP.2**). All are highly qualified and committed to PhD education, as demonstrated in their biosketches. They range from primarily experimental biologists who incorporate modeling in their work to dedicated mathematicians and computer scientists who apply their expertise to biological problems (**Figure PP.2**). The problems and phenomena that they study cover a wide range of scales, from molecular dynamics and sequence analysis to the function of the circulatory system and the theory of evolution (**Table 2** and **Table PP.4**). Faculty mentors hold primary appointments in multiple departments, including Mathematics, Molecular and Cellular Biology, Physiology, Physics, and Ecology and Evolutionary Biology. Many hold joint appointments in additional departments and graduate programs. For example, nine are members of BMCB and ten of Applied Mathematics, so Table 2 understates the breadth of their disciplinary connections.

Mentor financial resources. Research support for all program faculty is documented in **Table 4**. Of 32 faculty mentors, 14 are PI of NIH research grants, 6 are PI of NSF grants, 2 are early-career faculty with substantial start-up funds, and several faculty are funded by other sources (USDA, USGS, industry). Support averaged across all faculty is $295,350. All faculty have adequate resources to serve as mentors of trainees. Several features of the CMMBS program are relevant when assessing faculty research support. First, program faculty have highly collaborative research programs; faculty in mathematics and computer science routinely collaborate with funded experimental colleagues. Second, trainee research projects are computational or theoretical in nature, so the relevant infrastructure is access to computing rather than wet-laboratory resources. All students and all faculty have access at no charge to powerful supercomputer systems (see Resources). Third, non-federal grants, while typically smaller than NIH R01 awards, provide adequate infrastructure for theoretical research. Fourth, restricting the mentor pool to faculty with NIH support as PI would systematically underrepresent theoreticians, undermining the interdisciplinary character that defines CMMBS. Fifth, for faculty whose external funding does not cover student support directly, the participating graduate programs commit five years of financial support to all incoming doctoral students through Teaching Assistantships, Research Assistantships, and Graduate College Fellowships. These institutional resources complement mentor funding and ensure that all trainees are fully supported regardless of their mentor's grant portfolio.

Mentor pool selection. Faculty mentors are selected in part to ensure that research group sizes permit effective individual mentoring. The Executive Committee reviews each mentor's current trainee load as part of the faculty evaluation process; mentors whose groups have grown to a scale that would compromise the attention available to a CMMBS trainee are not accepted or are asked to defer participation until capacity allows. Faculty mentors are expected to commit sufficient time to trainee supervision given their other professional obligations, and this expectation is made explicit at onboarding. The interdisciplinary structure of CMMBS creates natural opportunities for faculty collaboration: trainees frequently work at the boundary of two mentors' expertise, and the program actively encourages co-mentorship arrangements that bring a biologist and a quantitative scientist into joint supervision of a single student. Finally, all mentors are expected to structure trainee research experiences to promote increasing self-direction over time. This expectation is operationalized through the IDP process, in which each annual revision documents goals the trainee has met independently and establishes new ones of greater scope, providing a longitudinal record of growing research autonomy.

Faculty are added to the program following Executive Committee review of each candidate's curriculum vitae and statement of research interests. Selection criteria include suitability of research area, ability to provide an appropriate research environment for trainees, and availability of adequate resources. All current mentors are personally known to Executive Committee members and were selected for their high standards of scientific rigor, insistence on responsible conduct of research, and demonstrated responsiveness to trainee needs. Faculty may be removed from the program if they leave the institution, retire, change research interests, demonstrate unacceptable mentoring practices, or otherwise prove disruptive to the CMMBS program.

Faculty at all tenure-track levels are included (**Table PP.3**). Women are dramatically underrepresented in the quantitative sciences, comprising only 18% of faculty at

Table PP.3: Composition of CMMBS faculty

Rank	**Men**	**Women**
Professor	11	4
Associate Professor	6	0
Assistant Professor	5	6
All ranks	**22**	**10**

doctoral institutions in mathematics and statistical sciences and 16.7% in computer science. The CMMBS training faculty substantially exceeds these national baselines, with women comprising 30% of the mentor pool (**Table PP.3**).

Table PP.4: Research interests of CMMBS faculty.

Faculty Mentor	Research Description
Barker, Mike	Develops computational genomic and transcriptomic approaches to study patterns of genome evolution, including whole-genome duplication and diversification across plant lineages.
Billheimer, Dean	Develops Bayesian statistical frameworks and compositional data analysis methods for inference in complex biological and environmental health datasets.
Capaldi, Andrew	Uses quantitative genetics and systems biology approaches to dissect signaling network architecture controlling nutrient sensing and homeostasis via the TORC1 complex.
Cusanovich, Darren	Develops single-cell genomics algorithms and computational tools to infer temporal dynamics of gene regulatory networks in respiratory development and disease.
Ding, Hongxu	Develops machine/deep learning and statistical approaches to interpret nanopore sequencing readouts.
Enard, David	Uses population genomic statistical frameworks to detect signatures of positive selection in human genomes driven by ancient viral pathogens.
Granzier, Henk	Applies quantitative structural and mechanical assays alongside computational modeling to understand the contributions of giant sarcomeric proteins titin and nebulin to striated muscle function.
Gutenkunst, Ryan	Develops and applies diffusion approximation PDE models and machine learning to infer demographic history and natural selection from population genomic data.
LaFleur, Bonnie	Applies statistical and machine learning methods for biomarker discovery, validation, and clinical translation in large-scale biomedical datasets.
Liang, Jingjing	Uses multi-omics and computational genomics approaches to identify genetic and epigenetic drivers of neurodegenerative diseases.
Lin, Kevin	Develops machine learning methods to analyze neural population activity and decode latent brain states from electrophysiological time series, and constructs computational models of cortical circuitry to elucidate neurophysiology and dynamical mechanisms.
Masel, Joanna	Develops formal mathematical and simulation-based models of evolutionary theory, including protein evolution, evolvability, and the population genetics of deleterious mutation accumulation.
McWhite, Claire	Develops and applies protein language models and deep learning to predict protein function and interactions at proteome scale.
Miller, Laura	Develops immersed boundary and other computational fluid dynamics methods to study the biomechanics of swimming, filtering, and pumping in biological systems.
Padi, Megha	Develops network-based algorithms and multi-omics integration methods to identify regulatory circuits driving cancer, with a focus on Merkel cell carcinoma.
Paek, Andrew	Combines live-cell imaging with computational image analysis and mathematical modeling to quantify single-cell signaling dynamics in response to cellular stress.
Riedel-Kruse, Ingmar	Uses quantitative computational modeling alongside synthetic biology to engineer and understand programmable living materials based on bacterial biofilms and synthetic adhesins.
Roman-Palacios, Cristian	Develops biological data science and machine learning approaches for macroevolutionary inference and modeling of biodiversity and disease vector ecology.
Salvador, Liliana	Develops computational and mathematical models of epidemiological and evolutionary dynamics to understand the ecology and spillover risk of zoonotic infectious diseases.
Secomb, Tim	Develops theoretical and computational models of the circulatory system, including blood flow, mass transport, and cardiac function.
Shipman, Patrick	Applies PDE and dynamical-systems theory to biological pattern formation, including tissue organization and phase-change patterning in biological materials.

Faculty Mentor	Research Description
Subbian, Vignesh	Develops and applies machine learning, electronic phenotyping, and explainability algorithms to address clinical decision-making and patient safety problems in learning health systems
Sun, Xiaoxiao	Develops novel statistical and machine learning methods for longitudinal epigenetic, spatial transcriptomics, and single-cell omics data.
Sutphin, George	Uses comparative genomics and systems genetics across model organisms, combined with high-content imaging and experimental validation, to identify genetic mechanisms of aging and stress response.
Van Doorslaer, Koenraad	Integrates genomics, bioinformatics, and evolutionary analysis to study how papillomaviruses remodel host cell biology and how evolutionary history predicts oncogenic potential.
Vitali, Francesca	Develops network-based bioinformatics and matrix factorization methods for multi-omics data integration, drug repurposing, and single-subject precision medicine in Alzheimer's disease.
Watkins, Joseph	Applies stochastic process theory and Bayesian statistical models to problems in population genetics, archaic admixture detection, genetic epilepsy classification, and epidemic dynamics.
Wheeler, Travis	Develops profile hidden Markov models, neural embedding architectures, and high-performance algorithms for biological sequence annotation, homology detection, and drug discovery.
Wilson, Amanda	Develops stochastic and agent-based mathematical models integrating behavioral and laboratory data to quantify microbial infection risks and evaluate intervention efficacies in healthcare environments.
Wolgemuth, Charles	Develops continuum mechanics and reaction-diffusion PDE models to understand the biophysics of bacterial and eukaryotic cell motility, shape generation, and collective migration.
Yao, Guang	Combines ODE-based mathematical modeling and machine learning-based statistical modeling with single-cell experiments to dissect the control mechanisms underlying cellular quiescence depth, senescence, and cancer dormancy
Zhang, Hao (Helen)	Develops penalized regression, support vector machines, and nonparametric smoothing methods for high-dimensional statistical machine learning in genomics, medical imaging, and cancer research.

G. Candidate Review Process

New trainees in the CMMBS program are second- or third-year University of Arizona PhD students who have already chosen a primary advisor and who expect modeling in biology to be a substantial component of their dissertation research. This timeline accommodates the diversity of graduate programs from which students enter CMMBS. For example, BMCB students choose an advisor and broad dissertation topic late in their first year after rotations, but Applied Mathematics students choose in the middle of their second year after extensive coursework. Regardless of their home graduate program, all students must have successfully completed one semester of calculus and one semester of statistics to be eligible for the revised CMMBS program. (Previously, two semesters of calculus were required.)

In early March of each year, applications for trainee positions are solicited broadly within appropriate Ph.D. programs. The desired qualifications of applicants are as follows:

> “Applicants for training grant support must be currently enrolled in a Ph.D. program at the University of Arizona and must be US citizens or permanent residents. ... Suitable qualifications include a biological sciences background and demonstrated interest in computational or mathematical modeling approaches, OR a computational or mathematical background and demonstrated interest in biological applications.”

The availability of training grant positions is advertised through the training faculty and through their graduate programs. However, suitable candidates from other programs are also eligible to apply. **The selection of trainees from a wide range of biologically and quantitatively oriented graduate programs is a distinctive aspect of CMMBS, because it helps to train students to communicate and collaborate with others having complementary expertise.**

Applicants' home graduate programs provide copies of the students' original application files, including letters of recommendation. Students submit a statement of their research plans and directions, and the students' current Ph.D. mentor submits a letter of recommendation.

The Executive Committee meets in early April to select eight trainees for the next year of the training program. The main criteria used to select trainees are:

- Academic record, including mathematical and computational background
- Letters of recommendation
- Statement of interest and goals
- Area of research interest: must include application of modeling approaches to biomedical problems
- Training environment provided by mentor
- Year in graduate program (second or third year preferred)
- Balance of quantitative and biological backgrounds among trainees

No order of priority in criteria is implied by this list. Although sex is not a criterion for selection, 43% of CMMBS trainees are women, much higher than is typical in quantitative sciences. Students offered positions receive a formal letter describing program benefits and obligations. A signed acceptance is required before students are appointed to the program.

CMMBS does not conduct independent recruitment and admissions of PhD students to the University of Arizona. Instead, it appoints students who have already been admitted to and are enrolled in existing U of A doctoral programs. This approach is deliberate and appropriate for several reasons. First, the program's central educational goal, developing the ability to communicate and collaborate across disciplinary boundaries, requires trainees who are genuinely embedded in distinct disciplinary programs, not students recruited into a single unified CMMBS degree program. The cross-disciplinary cohort that results from drawing trainees from Applied Mathematics, BMCB, Biomedical Engineering, and other programs cannot be replicated by a self-contained admissions process. Second, the programs from which CMMBS draws already conduct rigorous, discipline-specific admissions that identify students with the research potential and academic background that CMMBS requires. Conducting a parallel CMMBS admissions process would duplicate this effort without improving trainee quality or diversity. Third, by appointing second- and third-year students, CMMBS identifies trainees whose modeling interests have been confirmed by actual research experience, not only by stated intent at time of application; this is a more reliable indicator of fit than admissions-stage self-identification. The CMMBS selection process then applies its own additional criteria, described above, to identify among the pool of eligible enrolled students those who will most benefit from and contribute to the program.

Eligible students who have completed one year of support are invited to apply for a second year. They are asked to provide a current CV, a brief statement of research activities during the current year and plans for the next year, and a letter of recommendation from their advisor. Reappointment is not automatic and is dependent on satisfactory progress during the current year, including active participation in the Quantitative Biology Colloquium, and subject to availability of trainee positions.

H. Trainee Appointments and Financial Support

The CMMBS program is recognized throughout the U of A as an innovative and important interdisciplinary graduate training program. Consequently, the Graduate College and the Department of Molecular and Cellular Biology have agreed to provide generous financial and administrative support to the CMMBS program. The Graduate College will provide tuition waivers for eight students appointed to the training grant at a cost of up to $792,000 over the next five years. Because the CMMBS program and PI Gutenkunst are closely connected to the MCB Department, the department has committed to funding the administrative assistant for the program. Eight CMMBS scholars will receive NRSA-funded stipends over two years (typically years two and three of their doctoral program), in addition to travel funds.

The U of A has a longstanding commitment to graduate programs, and in particular, interdisciplinary programs. The UA Graduate College offers almost 100 doctoral programs, over 100 Master’s programs, as well as several specialist programs. The UA's 21 Graduate Interdisciplinary Programs (including Applied Mathematics) and more than 80 research centers and institutes unite researchers and students from diverse disciplines and provide exciting cross-disciplinary opportunities for research and training. Graduate students number approximately 11,500 of a total of about 45,000 students at the University.

The level of commitment of the participating graduate programs to their students is indicated by two key shared characteristics:

- all entering Ph.D. students receive a commitment of five years of financial support (from several sources) subject to satisfactory progress in their program;
- all entering Ph.D. students have an opportunity to spend at least the first year doing rotations with several potential advisors, before they commit to a specific research advisor.

These characteristics help to provide an optimal environment for success in Ph.D. training. The financial support for all students is derived mainly from a combination of Graduate College Fellowships, teaching Assistantships, Research Assistantships, individual fellowships (including NSF, NIH and AHA), and NSF and NIH training grants.

I. Trainee Outcomes

Table 8A documents the progress and outcomes of all 23 trainees appointed during the 2019-2024 funding period. **Table 5A** documents their scholarly productivity. The narrative below summarizes outcomes across the three dimensions the solicitation identifies: scholarly productivity, degree attainment, and career placement.

PhD Completion and Time to Degree. Of 23 appointed trainees, 16 have completed doctoral degrees, 5 remain in active training on track for graduation, and 2 left the program with Master's degrees. No trainee withdrew without a degree. The 91% rate of PhD completion or continued active enrollment meets the program's stated objective of 90% or greater. The two trainees who completed Master's degrees both remain employed in technically demanding roles: Matthew Miller as a Solutions Architect at the Chemical Abstracts Service and Zachary Frankman as a Research Associate at Roche. Miller's departure from the PhD program was directly attributable to personal circumstances precipitated by the COVID-19 pandemic rather than academic deficiency.

Time to degree was calculated from the semester of first enrollment in the trainee's home doctoral program to the semester in which the PhD was conferred. Among the 16 PhD completers, time to degree ranged from 3.5 to 6.5 years, with a median of 5 years. No trainee exceeded 7 years. These figures meet the program's stated objective of a median time to degree under 6 years and reflect the effectiveness of the program's IDP-based monitoring structure in identifying and resolving obstacles to progress before they compound.

Scholarly Productivity. Every trainee who completed a doctoral degree published peer-reviewed work before or shortly after graduation, as documented in Table 5A. Among the 16 PhD completers, first-author publications are well represented: Ryan (3 first-author papers), Knapp (4), Pungitore (5), Lasick (1), Lopez-Pier (1), Schackart (1), Preston (1), Aucoin (1), Harrison (2, with one under review), and Kinney (1, with one under review). Trainees with co-authored but not first-authored publications, including Sosnowski (5 co-authored papers), reflect productive collaborative research environments typical of experimental and engineering-adjacent dissertation projects. Several trainees also presented their work at national and international conferences, including the Biophysical Society annual meeting, IEEE International Ultrasonics Symposia, the NIH BRAIN Initiative Meeting, and the SIAM Conference on the Life Sciences.

Career Outcomes. Trainees who have completed their degrees are employed across the full range of biomedical research workforce sectors. Academic placements include Kathleen Lasick as Assistant Professor of Biology at Shenandoah University, and four trainees currently in postdoctoral positions: Chet Preston and Margaret Allard at the University of Arizona, Adrienne Kinney at Bowdoin College, and Alexa Aucoin at the New Jersey Institute of Technology. Industry placements span pharmaceutical research (Marissa Lopez-Pier and Tom Knapp as Scientists at Roche, Katie Sosnowski as Research Scientist at Ceria Therapeutics), defense and advanced technology (Paul Ryan as Advanced Technology Research Scientist at Honeywell, Bill Fries as Senior Principal Engineer at Northrop Grumman), data science (Sarah Pungitore as Data Scientist and AI Lead at Republic Services, Addie Harrison as Specialist in Data Analytics at Lincoln Financial), and software and consulting (Kenneth Schackart at Booz Allen Hamilton, Andres Nuncio Zuniga at Accelerate Diagnostics). This distribution across academia, biotech and pharma, defense research, and data-intensive industries is consistent with the program's goal of preparing trainees for the breadth of careers in which quantitative biology training is valuable.

Disciplinary Breadth. The 23 trainees in the prior funding period were drawn from eight graduate programs: Applied Mathematics (7), Biomedical Engineering (7), Molecular and Cellular Biology (3), Optical Sciences (2), Physics (1), Biosystems Engineering (1), Genetics (1), and Information (1). This cross-disciplinary composition is a defining feature of CMMBS and a direct contributor to the career breadth described above: trainees with mathematics training have moved into pharmaceutical research and academic biology, while trainees with engineering backgrounds have moved into computational and data science roles. The program's deliberate mixing of trainees from quantitative and biological programs produced the cross-disciplinary competency that employers in all sectors have recognized.

J. Program Evaluation Plan, Dissemination, and Data Storage

The long-term measures of success for CMMBS are the number of students who complete the program and graduate with doctoral degrees and the knowledge and skills that they bring to their future careers, particularly regarding the interface between biology and modeling. We have an established record, as detailed in **Table 8A**. We also make information about the subsequent careers of trainees available on the program website.

These long-term measures necessarily lag our training activities by several years. Shorter-term measures of progress are also monitored closely, including

- number and quality of program applicants;
- distribution of trainees with regard to their primary areas of training (biology vs. modeling);
- areas of dissertation research in relation to aims of this program;
- signs of trainee growth as evidenced from regular IDP revisions, in which old goals are met and thus replaced with new goals;
- signs of trainee flourishing as subjectively viewed by the instructors of core courses;
- time to degree;
- trainee publications and their quality;
- career placement.

The Executive Committee meets at least once a year. At each meeting, it reviews the composition of the student group and their progress in relation to the goals of the program. The trends in these parameters guide the selection of trainees for the next year of the program, as well as suggesting any adjustments in coursework or mentoring plans for current trainees. Any changes in trainee activities are coordinated with the responsible faculty members in the home graduate program of the student.

The present proposal reflects the feedback that we received from peer reviews, from program officers (e.g., a robust plan for training in responsible conduct of research), and from surveys of program alumni. As an improvement going forward, we are conducting assessment that is both more rigorous and completely separate from the training and oversight processes described above. We are committed to being responsive to advice and evaluations, whether from internal or external sources.

J1. External evaluation

We have partnered with the Center for the Improvement of Mentored Experiences in Research (CIMER) to assess the mentored research experience. CIMER's mission is to improve mentoring relationships for both mentees and mentors across career stages and disciplines through the development, implementation, and study of evidence-based and trainee-responsive interventions.

Each Spring, CIMER will survey past and present trainees and mentors for their assessment of the program's impact on their scientific development and subsequent career. Survey results will be tracked longitudinally, beginning when trainees are selected at the end of Spring semester. As controls, they will also survey students within our two most important contributing graduate programs, BMCB and Applied Mathematics. By surveying students within BMCB and Applied Mathematics who did not receive the CMMBS intervention, the design enables a quasi-experimental attribution of outcomes to program activities specifically. Students are not randomly assigned to CMMBS, so self-selection remains a factor. To address this, the CIMER survey protocol collects validated baseline measures of research self-efficacy, science identity, and interdisciplinary confidence at the start of longitudinal tracking, before CMMBS activities would be expected to produce observable effects. Growth trajectories across the funded period, rather than endpoint comparisons alone, are the primary outcome measure.

Surveys developed with CIMER are customized to the goals of CMMBS and include validated measures that can be compared nationally. CIMER has expertise in survey measures used to evaluate mentorship interventions and mentored research experiences, encapsulated in their CIMER Mentorship Measurement Library. Trainee survey items assess self-perceptions of skills in general Ph.D. competencies, communication, interdisciplinarity, and domain-specific expertise in both biology and modeling. They also assess psychosocial factors, both general factors regarding research self-efficacy and science identity, and items specific to the mentor relationship. Lastly, they assess program structure, training climate, and psychological safety. Mentor survey items assess mentors' perceptions of trainee skills, mentors' self-perceptions of their own mentoring skills, and the specific mentoring relationship. CIMER uses validated and published scales to assess each survey item.

CIMER will provide aggregate results of trainee and faculty survey data in an annual report to the Executive Committee. This report covers growth in research skill gains, overall mentored research experience, mentoring relationships for both mentees and mentors as well as overall program satisfaction. The report also includes measures of 'soft' outcomes, such as research self-efficacy, science identity, and perception of safety, which have been shown to be integral to long term mentee outcomes such as productivity and persistence. Highlighted in this report are also questions around the mentor's effectiveness and skills, as well as the mentoring relationship, which has also been shown to be a key element of successful mentored research experiences.

Partnership with CIMER adds a second comparison axis independent of the local control group. The validated instruments used in this evaluation have been administered across dozens of training programs nationally and are archived in the CIMER Mentorship Data Repository. This allows CMMBS outcomes to be evaluated against national distributions rather than only against local comparison students. The Executive Committee will receive annual CIMER reports summarizing both comparisons, allowing the program to identify specific areas of strength and areas where additional investment is warranted.

Data security. Data will be collected through the electronic CIMER Assessment Platform and will exist under the CMMBS project group. All data will belong to the U of A and PIs can access data at all times. From CIMER, only the Research and Evaluation team members will have access to the data to allow for deliverables to be created. All data will be archived in the CIMER Assessment Platform behind a secure firewall at the Wisconsin Center for Educational Research, at the University of Wisconsin-Madison.

Program participants also have the option to consent into having their survey data added to the CIMER Mentorship Data Repository. This is covered under UW-Madison IRB #2022-0782. All data shared with the CIMER Mentorship Data Repository will be de-identified.

All data in reports will be de-identified and aggregated to help protect participants' confidentiality.

J2. Dissemination

The program makes training and career outcome data publicly available on the CMMBS website, including aggregate, de-identified information on trainee PhD completion rates, time to degree, and post-degree employment. Alumni LinkedIn profiles are linked from the program webpage, providing a transparent and current record of career trajectories for prospective applicants, current trainees, and the broader community. The PIs disseminate effective practices from the program to the training community through presentations at relevant venues including the annual NIGMS T32 program meeting and, where appropriate, through publication. The CIMER partnership provides an additional dissemination pathway: survey data contributed to the CIMER Mentorship Data Repository, under de-identified conditions, supports the broader research base on evidence-informed mentoring practices in biomedical training programs. These activities ensure that CMMBS contributes not only to the development of its own trainees but to the improvement of graduate training practices more broadly.