COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 04-2	FOR NSF USE ONLY
NSF 04-546 **04/26/04**	**NSF PROPOSAL NUMBER**

FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.)

DUE - INTERDISC TRNG IN BIO & MATH

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
				020657151	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS ☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)
141368361		

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE
SUNY College at Geneseo	**SUNY College at Geneseo** **1 College Circle** **Geneseo, NY. 14454-146**

AWARDEE ORGANIZATION CODE (IF KNOWN)

0028456000

NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE	ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE

PERFORMING ORGANIZATION CODE (IF KNOWN)

IS AWARDEE ORGANIZATION (Check All That Apply)   ☐ SMALL BUSINESS   ☐ MINORITY BUSINESS   ☐ IF THIS IS A PRELIMINARY PROPOSAL
(See GPG II.C For Definitions)   ☐ FOR-PROFIT ORGANIZATION   ☐ WOMAN-OWNED BUSINESS   THEN CHECK HERE

TITLE OF PROPOSED PROJECT   **RUI: Undergraduate Biomathematical Research Career Initiative at SUNY-Geneseo**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ **1,151,241**	**60** months	**10/01/04**	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW
☐ BEGINNING INVESTIGATOR (GPG I.A)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.B, II.C.1.d)
☐ HISTORIC PLACES (GPG II.C.2.j)
☐ SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.1)
☐ VERTEBRATE ANIMALS (GPG II.D.5) IACUC App. Date _____

☐ HUMAN SUBJECTS (GPG II.D.6)
  Exemption Subsection _____ or IRB App. Date _____
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED
  (GPG II.C.2.j)
  _____
☐ HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.E.1)

PI/PD DEPARTMENT **Mathematics**	PI/PD POSTAL ADDRESS **1 College Circle**
PI/PD FAX NUMBER **585-245-5128**	**Geneseo, NY 14454** **United States**

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME **Anthony J Macula**	**PhD**	**1989**	**585-245-5482**	**macula@geneseo.edu**
CO-PI/PD **Gregg Hartvigsen**	**PhD**	**1995**	**585-245-5448**	**hartvig@geneseo.edu**
CO-PI/PD **Christopher Leary**	**PhD**	**1985**	**585-245-5383**	**leary@geneseo.edu**
CO-PI/PD **Wendy K Pogozelski**	**PhD**	**1994**	**585-245-5453**	**pogozels@geneseo.edu**
CO-PI/PD				

Electronic Signature

Project Summary

Abstract: The *Undergraduate Biomathematical Research Career Initiative at SUNY Geneseo* project will support a research program at Geneseo that will bring the continuing commitment of sixteen students in two related projects and will institute broad, integrated curricular improvement in the area of biomathematics. The project will build upon the current strengths of the undergraduate program at Geneseo and the expertise of the four co-PI's. It will increase the number of students trained in mathematics and biology, and it will increase the number of students and faculty performing research at the intersection of the two fields. It will increase the visibility of biomathematics research at Geneseo and will promote undergraduate biomathematics research in the western New York region and nationally.

Intellectual Merit: The two major research areas, DNA hybridization technologies and biological networks, are fields that share several characteristics that make them ideal for a UMB proposal. First, they are both active fields at the cutting edge of research. The amount of work being done in these areas is large, and there are many important questions that are both accessible and exciting to undergraduates. Both fields are prime examples of biomathematical research, with biological questions leading naturally to mathematical investigations and mathematical models and theories being proofed in the world of biology. The co-PI's have worked extensively in these fields and have mentored many undergraduate research students in these areas.

One of the key features of the recruitment plan is to bring students into the research setting as early as possible in their undergraduate career. Both of the research projects have different levels of questions that can be investigated, so first-year students can join the research teams and contribute to the ongoing research in a meaningful way at the same time as they learn the more advanced mathematics and biology that will be necessary for their work later in their undergraduate careers.

To aid the students' maturation process, the proposal outlines a linked curriculum that will support both the students in the lab and the general student population at Geneseo. The curricular changes that we propose will help bring research at the intersection of the biological and mathematical sciences to the forefront at Geneseo, and will make more Geneseo students aware of the opportunities that are available to researchers who are well-trained in both areas.

Broader Impact: Undergraduate research has a long tradition at Geneseo, and that tradition includes the dissemination of research results by undergraduates, both by publication in appropriate journals and by presentations at regional and national meetings. Beyond these obvious steps, we will sponsor a yearly undergraduate conference focused on research in biomathematics, bring outside experts to Geneseo both to run workshops and to observe what is being accomplished at Geneseo. The co-PIs will disseminate the results of the project through presentation and through publication.

The ideas and structure of the proposal are fully exportable and could be used to fashion a similar research program at any four-year institution. The assessment program that is proposed will also be useful for other institutions that would be considering a similar project.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.C.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	
Table of Contents	1	
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) **(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**	15	
References Cited	5	
Biographical Sketches (Not to exceed 2 pages each)	8	
Budget (Plus up to 3 pages of budget justification)	13	
Current and Pending Support	4	
Facilities, Equipment and Other Resources	1	
Special Information/Supplementary Documentation	4	
Appendix (List below.) **(Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**		
Appendix Items:		

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

A PRIOR NSF SUPPORT

A.1 Anthony J. Macula

A. Macula, co-PI, was supported by the *NSF-9973252* ($54K, 6/99-7/01) "Applied nonadaptive group testing and superimposed codes," *NSF-0107179* ($60K, 7/01-6/03) "Group Testing for Complexes," and is currently supported by *NSF-0107179 UBM Supplement,* ($34K, 9/03-8/04) "Analysis of Biological Networks." The relationship between the first two funded projects and this proposed project is that the group testing and data mining methods which were developed in those projects will be applied to advance DNA hybridization technology and to find important structures in biological networks. (See Sections C.2 and C.3.) The relationship between the third project and this proposal is discussed in Section A.1.1 immediately below.

A.1.1 Current Projects

The *NSF-0107179 UBM Supplement* project is ongoing. Six undergraduates (3 biology, 2 math, 1 math/bio) are being funded under this proposal as is co-PI G. Hartvigsen. See (Section B.7.) The goals and objectives of the supplement are: 1) To introduce the analysis of biological networks to our students in mathematics and biology in an interdisciplinary, inquiry-based learning approach. 2) To apply the mathematical and algorithmic methods of group testing for complexes to identify critical structures in the underlying networks. 3) To involve students in research at the intersection of mathematics and biology.

Five out of these six students will present their results this spring at either our local *2004 SUNY Geneseo Scholars Symposium* or the *2004 Council on Undergraduate Research (CUR) Posters on the Hill Conference* in Washington, DC. The first and the third goals have been achieved. In regard to the second goal, we have developed some preliminary ideas for group tests to detect shortcut edges in networks. Our students will test these conjectures this summer.

A.1.2 Completed Projects

The projects *NSF-9973252* and *NSF-0107179* have been completed. The goals of *NSF-9973252* were: 1) To find new and improved constructions of superimposed codes and group testing designs. 2) To apply these designs to practical problems in bioinformatics and other areas of search theory. 3) To develop methods that identify positive combinations of objects. The goals of *NSF-0107179* were: 4) To find new group testing for complexes designs.   5) To develop new group testing decoding methods. 6) To develop group testing designs that are error-correcting.   7) To apply group testing for complexes designs to data mining with particular attention to biotechnology and information assurance.

The goals of the proposals were achieved. (See publications below.) Along with the PI, eight undergraduates (2 sophomores, 3 juniors and 3 seniors) worked on these projects. Four of the fifteen publications from these projects have student co-authors. All student participants gained valuable research and programming experience in the process. The students learned advanced topics in algebra, combinatorics, coding theory, and probability. They observed first-hand how mathematics can be applied to real world problems. They also learned how to implement these applications through computer programming. This research experience has encouraged them to continue their research endeavors in graduate school. Of the five students that have since graduated, four have gone on to graduate school and three of these have received fellowships.

A.1.3 Project Publications

Fifteen publications arose from these projects. *NSF-9973252* publications [33], [60], [61], [62], [42], [63], [40], [76] and *NSF-0107179* publications [34], [64], [65], [39], [35], [66], [36] are cited.

A.2   Wendy K. Pogozelski

W. Pogozelski, co-PI, is currently supported by *NSF-0126642 ($78K, 1/02 - 6/04)* "Quantitative Biochemistry Laboratory Experiments: Introduction of Isothermal Titration Calorimetry and Real-Time PCR in the Undergraduate Curriculum." This was funded under the NSF-CCLI-Adaptation and Implementation Program. This funded project is connected to both the research and curricular parts of this proposal. The instrument will be used for DNA quantification in the *Introduction to Nucleic Acids and Biology* and *Computational Molecular Biology* courses which are part of this proposal. It will be used in research to validate DNA codes. (For other applications, see Sections D and C.3.)

A.2.1   Current Projects

The objectives of W. Pogozelski's portion of *NSF-0126642* were met ahead of schedule. A real-time PCR thermalcycler was purchased, set up, and tested in 2002. Since then, it has been used in twelve undergraduate research projects leading to eleven presentations. The PI also presented two posters (one education track, one research track) at the ASBMB meeting (Experimental Biology 2003), as well as a poster at Mitochondria 2003 and an invited talk at the fall 2003 American Chemical Society national meeting. A student coauthored publication, "Determination of Mitochondrial DNA Copy Number Using Real-time PCR and the Fixed-Slope Method," is in preparation. In the fall of 2002, the unit was introduced for a new inquiry-based experiment in the Biochemistry Laboratory course and the activity was linked to the introduction of real-time PCR theory in Biochemistry I. The instrument has also been useful in attracting additional funding from the *AFOSR (F30602-03-2-0086, 38K, 6/03-12/03, Pogozelski)*

Outside evaluation of the project was performed by Professor S. Gal of Binghamton University Department of Biology. Results showed excellent progress in reaching the grant's objectives and in assessing student learning based on the instrument. The evaluator did suggest that the PI could go even further in making the Biochemistry Laboratory course inquiry-based. Many fruitful discussions have resulted from the evaluation, and the Biochemistry Lab course is being modified to include additional inquiry-based exercises while retaining the emphasis on quantitative biochemistry.

B   OVERVIEW

B.1   Main Goals

- *To enable, support and encourage students to do research at the interface of mathematics and biology. This goal will be achieved by long-term immersion of students in cutting-edge research activities, supported through the close integration of research-based mentorship and a coupled, co-disciplinary curriculum.*

- *To strengthen recently formed research and curricular ties between the Mathematics and Biology Departments at SUNY Geneseo.*

- *To formalize an undergraduate biomathematical research career program and to make permanent and significant biomathematical curricular changes at SUNY Geneseo.*

- *To recruit students into our research program early in their undergraduate career and to maintain their participation in mentored research throughout their tenure at SUNY Geneseo.*

- *To bring biomathematical undergraduate research more deeply into the thriving undergraduate research culture of SUNY Geneseo.*

B.2   Brief Description of our Institution and Departments

The State University of New York (SUNY) College at Geneseo is a highly selective public undergraduate college of 5000 students, over 1000 of whom are math and science majors. SAT scores for the current freshman class average 1260. The college has a strong tradition of undergraduate research and a successful record of sending students to graduate and professional schools. The Biology Department has over 500 majors, the Math Department has over 225 majors and there are over 100 biochemistry majors. Our activities will draw upon the strengths of the institution, its population of hard-working and ambitious students and its successful involvement in undergraduate research.

B.3   Research Themes

The central themes of the research projects are DNA hybridization technology and biological networks. These themes are united by the computational methods used to study them. Curricular modules associated with the research will be developed for incorporation into biomathematics curricula. Students will take classes, attend seminars, engage in mentored research, attend professional conferences, make presentations and write papers. We believe that our nation's future technological infrastructure depends on successful interactions between increasingly less distinct scientific disciplines. Our students will be trained to communicate ideas to people outside their immediate research area. This latter skill is critically important in modern academic circles and industry.

B.4   Targeted Student Participants

We will continue to recruit both mathematics and biology majors. Currently we have eight students participating in jointly mentored research. Three of these students are math majors, four are biology majors and one is a double biology and math major. Five of these students will return to SUNY Geneseo in the fall of 2004. These returning students are strong candidates for this program. In the first year, we will add at least 11 additional students of various levels to our fledgling biomathematical research program. Below is a list of candidates:

Biology: Adrian Miszczak, Amy Zielinski, Ashley Knight, Carrie Gregory, Chris Mayack, Danielle Foti, Danielle Mendez, Dewi Hartari, Francesca Magri, Holly Barcelo, Kelly Van Slyke, Laura Titus, Maiko Funayama, Matthew Turbush, Sadia Sahabi, Sandra Frandina, Sanjar Enam, Sarah Olscamp, Shuya Kyu, Stephen Kolb, Sze man Tsang, Victoria Kuzara

Mathematics: Neils Hansen, Jackie Dresch, Erica Scott, Dori Freeman, Joseph Hughto, Erin Egerbrecht, Craig Chamberlain, Christina Callear, Scott Meckler, Tonya Smith, Cynthia Stallard, Dawn Owczarzak, Dave Pazamickas, Mark Vittore, Jonathon Rado, Ryan Grover ,Gary Surman, Minda Wesley

Biochemistry: Michelle Amsler , Danielle Foti, Carrie Gregory, Matt Takos, Janhvi Patel, Dan Welchons, Nicholas Arpaia, Svend Rommel, James Keyes

Biology/Mathematics: Kevin Pinto, Melissa Pouliot, Sal Priore, Jacqueline Lex

We plan to fund 12 students each semester and 12 students each summer. (See Table 1 below for the five-year schedule.) We will use a model in which students are brought into research groups at an early stage of their career in order to give them time to acquire both the biology and math skills they need to successfully conduct biomathematical research. By supporting, enabling and encouraging students throughout their academic career, we believe that we can significantly improve their experience as biomathematical research scientists and prepare them to further their academic careers after they leave our institution.

B.5   Intellectual Focus

Over the next five years our undergraduates will embark on a research-based program that will combine new and revised problem-based curricular components (e.g., calculus for biologists, modeling biological systems, biological data analysis, discrete mathematics, graph theory, probability, statistics and computational molecular analysis) with research activities (ranging from biomolecular computing to individual-based, evolutionary modeling of food webs) that push the boundaries of modern techniques in understanding complex problems. Students at all levels will participate in this training program as they blend our evolving and innovative curriculum with research projects at the frontier of biomathematics. Jointly-mentored biomathematical research projects will be at the core of our program and these research projects will serve as a stimulus for faculty development and curricular innovation. The opportunity to directly participate in these research projects will serve as an incentive for students and faculty to excel in, and learn about, the discipline that is not their major.

B.6   Innovative Strategies

We feel that one of the best ways for students to learn to use the principles of both biology and math is to actually engage in studies that involve both of these disciplines. The PIs for this project have worked at the interface of biology and mathematics (hereafter referred to as *biomathematics*) for several years and are currently mentoring undergraduate biomathematical research projects in DNA hybridization assays and in the study and analysis of biological networks. These research endeavors are on the frontier of biomathematics. (See Section C.) We plan to continue and extend this interdisciplinary work.

We are aware of important studies that point to the need for better preparation in mathematics for biology students, and increased knowledge of biology for mathematics students. Our plan draws upon the recommendations of the Bio 2010 study, Bialek and Botstein and May and others ([10], [9], [68]) who emphasize that real mastery of ideas in biomathematical areas requires reinforcement throughout the curriculum and access to practical problem-solving. Using these ideas, the PIs are at the forefront of recent biomathematical curricular reform at SUNY Geneseo. An important goal of this proposal is to significantly extend this reform and to foster an increased sense of the importance and feasibility of cross-disciplinary biomathematical undergraduate research at SUNY Geneseo. To be successful in this goal, we recognize that certain challenges need to be addressed. Some examples are:

- Both mathematics and biology students have difficulty in understanding abstract models and their application to real world problems. They do not easily see the connections between modeling, mathematics and biology.

- Biology students need to know more mathematics and need to have more confidence in their mathematics skills.

- Mathematics students need more exposure to, and preparation in, working with biological systems and they need more confidence in their ability to conduct meaningful experiments.

We propose to meet these challenges in two ways: 1) Curricular initiatives will extend the recent commitment of SUNY Geneseo's Mathematics and Biology Departments to cross-disciplinary courses and research. (See Sections C and D.) 2) Our recently formed biomathematical research lab will be extended and its scope will be broadened. See below.

Our biomathematical research lab will be largely (but not exclusively) based on the highly successful approach used in our Physics Department, in which undergraduates begin research in their first year and continue for four years, honing skills, generating useful data, and training

incoming students. Our experience indicates that it takes time for students to become comfortable in applied mathematics and this model will enable us to involve more students more deeply in our biomathematical projects. Funding will allow us to bring students into our program earlier than might otherwise be reasonably expected and will provide these students with an intense, varied and contiguous research experience that lasts through their entire undergraduate career. We have begun to devise level appropriate research tasks. (See Section C.) The initial research tasks, while connected to the primary research activities, are relatively straightforward data analysis, experimental and modeling projects designed to introduce students to the research atmosphere. This will allow us to form a cohort of student researchers with different levels of ability where everyone can contribute in the research efforts and discussions. Students who are brought into the research project as early as possible, who are exposed to a diverse set of skills and a variety of research settings, and who are involved with research questions for 2-3.5 years will be extremely well prepared for entry into graduate programs at the interface of biology and mathematics. We also believe that by bringing students on board very early in their academic career and by securing their long-term commitment to the project, we will be in a position to provide them with effective and directed mentoring as regards their academic schedule and their academic focus throughout their undergraduate years, leading to increased academic success for all of the students involved in the project. (See Section E.)

B.7 Organizational Endorsement of the UMB Activity

The objectives of this proposal are supported by Geneseo's administration. We have obtained a commitments for 50% matching funds for the instrument we request (the Varian Cary 100 UV/VIS spectrophotometer.) and, each semester, one PI with be allowed to dedicate 1/3 of their teaching load solely to this project. Moreover, much of our normal teaching load will support the curricular initiatives of this proposal.

Research with students is part of the culture at Geneseo and is valued by the administration as evidenced by the allocation of funds. The university is currently building a $53 million integrated science building, to be called the Center for Excellence in Undergraduate Science. The building (to be completed in the fall 2006) will include new laboratory and teaching space for the projects described here. Nearly 100% of external funding at Geneseo (which last year amounted to $1 million) includes funds for students. The Geneseo foundation provides over $33,000 per year directly to students for travel to conferences and for supplies for student research projects. Each year the college holds an Undergraduate Scholars Symposium, in which students present the results of their work. This year seven of the eight current UBM students will present at this symposium. This event is attended by the administration and the President personally congratulates each student. The official organizational endorsement is a supplemental document.

B.8 Organizational Structure and Timetable

At least 16 students (ideally, 4 freshman, 4 sophomores, 4 juniors, 4 seniors) will be directly involved in the research lab during the academic year. Each student will be jointly mentored by a mathematician and a biologist with one of the two mentors taking a leading role. Each PI will be the lead mentor for 4 research students (ideally 2 beginning and 2 advanced) during each academic year. (See the discussion of a graduated approach to the research projects in Sections C.2 and C. 3.) In each summer, each PI will be the lead mentor for three research students. During the academic year, sophomores, juniors and seniors will be paid a stipend. Freshman students will not be paid during the academic year (as their activities will be focused on training,) yet a large portion of our effort is directed toward attracting freshman students. From the academic year group of 16 students, the freshmen, sophomores and juniors would be paid a stipend in the following summer. We expect these summer students to work nine weeks. See Table 1 for a detailed schedule. See

the remainder of the proposal for the description of activities listed in Table 1, but not discussed above.

Activities	Fa 04	Sp 05	Su 05	Fa 05	Sp 06	Su 06	Fa 06	Sp 07	Su 07	Fa 07	Sp 08	Su 08	Fa 08	Sp 09	Su 09
Recruitment	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Mentored Research (students, paid students)	16,6	16,12	12,12	16,12	16,12	12,12	16,12	16,12	12,12	16,12	16,12	12,12	16,12	16,12	12,12
Curriculum(mods) Development	X	X	X	X	X	X			X			X			
Curriculum(course) Development			X			X			X						
Student Workshop (5 days) (12 students)			X						X						X
Host Regional Student Conference (60 people)		X			X			X			X			X	
Student Conference Travel (12-16 Students)		X			X	X		X	X		X	X		X	X
Student Off-Site Lab Experience						X						X			
Student Assessment		X	X		X	X		X	X		X	X		X	X
External Reviewers Site Visit		X							X						X
NSF UBM Meeting			X			X			X			X		X	
External Assessment Report										X					X
Faculty Conference Travel			X			X			X			X			X
Faculty Workshop Travel			X			X			X			X			X

TABLE 1: Five-year schedule

The following is a list of the senior personnel and their responsibilities.

G. Hartvigsen, co-PI, is a quantitative biologist. He uses tools such as genetic algorithms to investigate how individual-level interactions influence large-scale phenomena such as evolution and community dynamics. He teaches ecology, biostatistics, and a yearly interdisciplinary course on networks, called, *Six Degrees of Separation*, which introduces first-year students to a variety of networks, including the building and analysis of their own developing social network. In addition, he teaches the upper-division, cross-disciplinary course entitled, *Modeling Biological Systems*, with co-PI C. Leary. He has mentored 18 undergraduates in biomathematical research over the past five years.

C. Leary, co-PI, is a mathematician with interest in biological systems. Currently serving as the Associate Chair of the Department of Mathematics, he teaches the Calculus II for Biologists course, co-teaches Modeling Biological Systems with co-PI G. Hartvigsen and has invested much time and creativity in the areas of curricular reform. He has been awarded an internal fellowship for the summer of 2004 in order to advance his curricular interests in biomathematics toward more research activity. He will partner with G. Hartvigsen in mentoring students doing research in the areas of networks, modeling and disease transmission.

A. Macula, co-PI, is a mathematician with an undergraduate degree in biochemistry. He has published in the areas of group testing, data mining, DNA library screening, computer security, medical genetics, biomolecular computing, and topology. He will direct the biomolecular computing/DNA hybridization project with W. Pogozelski, he will co-mentor students in the biological

networks project and he will be the project manager for this grant. As project manager, he will be responsible for reports and communications with NSF. He will also develop and teach the computational molecular biology course. A. Macula has mentored 33 students in research projects, has coauthored eight published papers with a total of eight students and has been the advisor for three undergraduate honors theses.

W. Pogozelski, co-PI, is a biochemist with training in mathematics. She is the co-coordinator of the Biochemistry major that is jointly sponsored by the Biology and Chemistry Departments. Her previous work has included the development of mathematical models, the design and implementation of computer algorithms, and the quantitative assessment of biological phenomenon using tools such as real-time PCR. She will work with A. Macula on the biomolecular computing/DNA hybridization project. She will also teach the Introduction to Nucleic Acids course. She has mentored 20 students resulting in 30 student presentations in the last eight years. She has coauthored one published paper with a student and has two others in preparation.

There are other members (from junior to senior) of the SUNY Geneseo Mathematics and Biology Departments who share our goals and are willing to participate in curricular initiatives described in Section D. These include: R. Simon, K. Hannam (Biology,) Jeff Johanes (Mathematics) and Ruel McKnight (Biochemistry.) It is our intent to include as many faculty members as possible in this program. Moreover, we hope to allocate some of the faculty travel money budgeted in this proposal for non-PI junior faculty member to travel to biomathematics workshops (e.g., IMA, DIMACS, BIOQUEST, NSF Chautauqua Short Courses for College Teachers.) In this way, we could also begin to incorporate our junior colleagues in the research portion of this proposal as well as the curricular portion.

C NATURE OF STUDENT ACTIVITIES

C.1 Themes, Approach and Schedule

Students who are involved in the research group will join biomathematical projects of their choice. The central themes of the initial research projects are DNA hybridization technology and biological networks. We currently have several students participating (six of eight supported by NSF #0107179 UBM Supplement) in these frontier research areas. A broader range of students will benefit from the proposed curricular modifications designed to support the proposed research. To achieve our fundamental goal of enabling undergraduates to do research at the interface of mathematics and biology, our students need three training components:

(1) Students need to build a foundation based on core topics in biology and mathematics as well as have access to targeted, topical knowledge to understand current problems in biomathematics in general and in our research activities in particular.

Here at SUNY Geneseo we have several cross-disciplinary courses (e.g., *Calculus for Biologists*, *Biological Data Analysis*, and *Modeling Biological Systems.*) We propose to develop new courses and course modifications that will serve students throughout this program. These courses include procedural programming for biomathematics, computation molecular biology, applied graph theory, introduction to nucleic acid techniques and a team-taught topical biomathematics seminar course. (See Section D.)

(2) Students need hands-on experience with cutting-edge biomathematical research.

Beginning early in their program, students will develop and test hypotheses and conduct research in biomathematics. This exploration can only happen under the guidance of multiple mentors. (See Section C.2 and C.3.)

(3) Students need exposure to a variety of additional cutting-edge research experiences through broad interactions.

This will occur through participation in other learning experiences such as conferences, workshops and travel to other institutions and research facilities. (See Tables 1 and 2 and Section D.) Each year, we plan to host an annual spring regional conference spotlighting undergraduate research in biomathematics. This activity is a significant component of our recruitment efforts. In alternate summers, we will either hold a week-long biomathematical workshop for our summer students or our students will travel to an outside biomathematical lab. Some candidates for featured workshop instructors are David Torney (Los Alamos National Labs, Theoretical Biology and Biophysics) and Douglas Deutchman (San Diego State Univ., Biology Department.) Some candidate labs for off-site experiences are Air Force Research Lab (Rome, NY) Los Alamos National Lab (Los Alamos, NM), Simon Levin's Lab (Princeton, NJ)

C.2 Biological Networks

Students will be introduced to a variety of biomathematical problems that involve the integration of graph theory and complexity in biological systems. Recent work demonstrates that a large number of complex systems exhibit network structures ([79], [6], [23], [28].) Currently, our students are investigating the network structures of model food webs and the transmission of disease epidemics through populations. These approaches to understanding and predicting the dynamics of complex systems are being widely investigated, are yielding valuable results ([80],[51] [38]) and are amenable to student researchers. We currently have several students, ranging from freshman to seniors, working on two of these problems (food web structure and disease spread in realistic networks.)

We believe that a graduated approach to research participation is essential. First-level projects are for new students while second level projects are for advanced students. The first level is the gateway for students to begin their research. We have found that the first level is open to a wide diversity of students who are eager to learn and willing to work hard. This level also provides opportunities for students to uncover new relationships that can only be discovered through blending mathematics and biology. Advanced students will dig deeper by modifying existing models and developing their own models to test hypotheses. These students also will be jointly mentored as they work to devise computational methods and algorithms to test their quantitatively-based hypotheses, find structures and patterns in networks and overcome inherent uncertainty and noise in these systems. This second level of research requires more extensive experience and training in mathematics, biology, and computer science. In addition, new models developed by second-level students will become available to first-level students for both research activities and to be used as tools for recruitment.

Below we provide three examples of investigative biomathematical experiences designed for our undergraduates. All three projects are based on existing models that are easily tested by first-level students who seek to analyze their dynamics. The models also are modifiable by second-level students who will build extensions to address additional hypotheses.

1) Cooperation: There continues to be much interest in understanding the conditions under which cooperation can arise and be maintained in evolutionary systems ([5], [27]) and how network structures influence this evolution ([37].) We have developed a simple, spatially-explicit model to investigate the dynamics of the change in the propensity of individuals to cooperate ([46].) Individuals in this simple world become more likely to cooperate when their partner is cooperative and less likely to cooperate when partners defect. Students in the first level will investigate how network topologies and population sizes interact to influence the likelihood of cooperation to emerge. Second-level students will develop evolutionarily-based algorithms to investigate the evolution of systems as diverse as bacterial biofilms, nitrogen-fixing cyanobacteria, and, potentially, the evolution of metazoans ([46].)

2) Epidemiology: Most, if not all, organisms are at risk of attack by pathogens. This fact creates

serious environmental, ecological, and economical problems throughout most biological systems. Although there is impressive diversity in the modes and extents of pathogen transmission the underlying processes are similar among diseases. These systems generally involve host individuals that are connected within a network structure to each other in which a disease agent moves from host to host. The structure of these host networks greatly influence the emergence and dynamics of disease outbreaks ([79], [70], [26], [48].) Students are currently testing a working model that allows realistic topologies to investigate the spread of influenza through a human population.

3) Food web dynamics: There currently is a large amount of research aimed at understanding how the dynamics and stability of food webs is affected by the network relationships among individuals and among interacting species ([80], [69], [74].) Our evolutionary model ([47]) is flexible and allows students to alter the number of genes involved in various types of predator-prey interactions. Students will get hands-on experience with the evolutionary programming techniques based on genetic algorithms ([44],) an important optimization programming tool, while gaining insight into how evolution and the role of network relationships influence the dynamics of natural systems.

These research activities will involve cooperative interactions between both students and faculty of mathematics and biology. Students at different levels of experience and background (major) will study network structures (graphs) and work to develop algorithms that alter the underlying network topologies of the systems, analyze the structures of the systems and assess the outcome of these changes in biological terms. Students will apply the mathematical, computational and algorithmic methods of group testing for complexes to identify critical structures (e.g., short cut edges, clusters, cut sets in the underlying networks) (See [34], [36], [39], [42], [61], [64], [65].) In analyzing these networks, students will learn how computational mathematics helps us understand biology and how biological systems can help us understand and design information systems.

Through the current *NSF-0107179 (UBM suppl., Macula)* students from biology and mathematics are investigating how connections between individuals in complex communities influence large-scale system dynamics. For example, four students (jointly mentored by co-PIs Hartvigsen, Leary and Macula) are currently exploring the rate and extent of influenza spread through a variety of network structures in a host population while a fifth student is investigating the dynamics of plant-herbivore community. To accomplish this, the group built computer models of these systems where the underlying network topologies are small-world networks. Students, for instance, now have a model that allows them to explore how influenza dynamics are influenced by factors such as the size of the host population and the structure of the underlying network, which ranges continuously from regular, through small-world, to random ([79]). In addition, students are currently investigating how epidemics can be controlled through different vaccination strategies (e.g., [51]) how the ability of diseases agents to mutate into new strains affects the dynamics of the infected host population (e.g., [21]),and how the host population's ability to benefit from cross-immunity influences disease epidemics ([13]). These students will present their results at the *2004 SUNY Geneseo Scholars Symposium* and *2004 Society of Mathematical Biology National Meeting.* These presentations include 1) Michelle Amsler's and Katrina Button's project entitled "Modeling Cross-Immunity in Influenza," 2) Jacqueline Dresch's and Amy Zielinski's project entitled "Disrupting Influenza Dynamics in a Network Model Through Different Vaccination Strategies," and Shuya Kyu's project entitled "Assessing the Role of Herbivores in Altering the Network of Plant Mating Types in a Food-Web Model."

C.3   DNA Hybridization Technologies

The primary goal of this research is the development of an enabling technology for DNA hybridization assay applications. This research combines mathematics, molecular biology and computer science. It is focused on the construction of a biomolecular architecture designed to employ algorithmic paradigms based on the massively parallel power of DNA hybridization. This biomath-

ematical technology has applications in many developing areas of biotechnology such as universal DNA chip technology ([8], [15],[16]) Single Nucleotide Polymorphism (SNP) multiplexing ([50],[18]) DNA based nanotechnology ([52],[81]) biomolecular computing ([1],[7],[2],[75]) and biomolecular information storage media ([19].) It has also produced new mathematics (e.g., block isomorphic subsequences in [32].)

Whenever two, not necessarily complementary, oppositely directed DNA strands "mirror" one another sufficiently, they are capable of coalescing into a DNA duplex. The process of forming DNA duplexes from single strands is referred to as *DNA hybridization.* The most energetically favorable duplexes are obtained when the two sequences are reverse complements of one another and the DNA duplex formed is a Waston-Crick (WC) duplex. However, there are many instances when the formation non-WC duplexes are energetically favorable. In this proposal, a non-WC duplex is referred to as a cross-hybridization (CH) duplex.

All of the DNA hybridization technologies listed above require large collections of non-cross-hybridizing oligonucleotides and there is a need to efficiently create these oligonucleotide collections. In DNA hybridization assays, the general rule is that the formation of WC duplexes is good, while the formation of CH duplexes is bad. The process of designing sets of non-cross-hybridizing oligonucleotides is known as *DNA word, DNA code*, or *TAT (tag-antitag)* design. A primary goal of DNA code design is be assured that a fixed temperature can be found that is (simultaneously) *well above the melting point of all CH and well below the melting point of all WC duplexes* that can form from strands in the code. Thus the formation of any WC duplex must be significantly more energetically favorable then all possible CH duplexes. A DNA code with this property is said to have *high binding specificity.* High binding specificity is akin to a high signal-to-noise ratio. (See [17], [24], [25], [29], [30], [31], [58], [59], [76] for more comprehensive discussions of DNA codes.) Both combinatorial and biological methods have been suggested as a means by which DNA codes can be found and programs exist that generate DNA codes. (See [59].)

The DNA code design project, briefly described below, will be used as a hub to an interdisciplinary program that addresses DNA methodology. Undergraduate participation is an integral part of the project, and thus undergraduates will be trained in the interdisciplinary knowledge and methods required of DNA technology research. Initial aspects of this project were partially supported under *AFOSR F3062-03-C-0059 (75K, 5/03-1/04, Macula)* and *AFOSR F30602-03-2-0086 (38K, 6/03-12/03, Pogozelski.)* In this new project, we introduce a new metric(s) for DNA code design that captures a key aspect of the *nearest-neighbor thermodynamic* model for hybridized DNA duplexes ([72].) Undergraduate students will be directly involved in performing quantification experiments, validation experiments, calibration experiments, data analysis, data interpretation, computational methods, experimental models and writing and implementing code.

Three students are already working in this area: 1) Matt Bernard's projects is entitled, "Exploring DNA Hybridization for DNA-Based Computing Using SYBR Green Fluorescence," and he will present at both the *2004 Geneseo Scholars Symposium* and the 2004 *National Council on Undergraduate Research (CUR) Posters on the Hill Program* in Washington DC. 2) Morgan Bishop's project is entitled, "Distributed Dynamic Programming for DNA Sequence Comparison," and he will present his findings at our local 2004 SUNY Geneseo Scholars Symposium. 3) Gavin Page is a coauthor with co-PI Macula (and others, [59].) Our work will soon become publicly available at the BIOSPICE web page. BIOSPICE is and open source framework and toolset for modeling dynamic cellular network functions. (See https://community.biospice.org.)

By emphasizing DNA hybridization, many more general topics in DNA analysis and methodology naturally arise and will be addressed in DNA curricular modules, e.g., sequence alignment, thermodynamics, PCR, secondary structure, SNP, microarrays, physical mapping, medical genetics, library screening, biomolecular computing, data visualization, distributed computing, data mining, etc. The supporting mathematics and biology will be delivered to the students through the proposed curriculum initiatives and will also be presented synchronously as the research is being

conducted. (See Section D.)

First-level student research projects in this area include: curve fitting to understand DNA hybridization reactions, analysis of simulated SNP data, benchmarking of DNA code generating software, nucleic acid experiment preparation, hybridization screening assays and melting point determination using the requested UV spectrophotometer equipped with a programmable heat source, fluorescence and quenching assays, experiments to probe the limits of our fluorescence assays, studying hysteresis and error in these measurements. Second-level student research projects include: using biomolecular inspired logic to solve discrete math questions, the construction of group testing algorithms ([34], [36], [39], [42], [61], [64], [65]) that could be used a the basis for a pooling experiment used to decide the binding efficiency of a DNA code, improvement of computer code used to generate DNA codes, development of biological methods of DNA code generation, developing extensions of the "stack pair" metric in [32], deterministic constructions of DNA codes and the computation of theoretical bounds on the size of DNA codes.

In specific regard to the equipment purchase request for the Varian Cary 100 Biomelt UV-Vis spectrophotometer (see Budget Justification,) students will use the UV-Vis spectrophotometer to screen DNA sequences and pools of sequences for CH duplexes. This will supplement our fluorescence-based assays that have some limitations. This instrument is ideal for DNA melting temperature analysis because it has the greatest accuracy on the market and permits flexibility in temperature ramp rates. Students will also use the UV data to find thermodynamic parameters under different conditions, thus determining the optimum hybridization conditions for future computing experiments. This instrument is also important to our *Introduction to Nucleic Acids* course.(See Section D.). Students will also use the UV-Vis spectrophotometer to screen DNA codes and they will also employ several different approaches in conducting the DNA hybridization biocomputation experiments.

D   Connection to Regular Academic Studies

In order to bring students into the research projects quickly and appropriately, we plan several curricular innovations. These innovations would: 1) be available to all students at Geneseo; 2) support the research projects that are at the core of our proposal; 3) help to change the campus culture in a way that will increase the visibility and importance of interdepartmental biomathematical research and curricular activities.

Integration of mathematics and biology has been a recent focus of curricular reform at Geneseo and the PIs have been at the forefront of this reform For example, BIO/MA 340 *Modeling Biological Systems* and, MA 228 *Calculus for Biologists* are regularly offered courses. (See Section B.8.) However, we recognize that we need to expand biomathematical curricular reform if we are to address the challenges we have observed in our current undergraduate biomathematical research efforts. (See Section B.6.)

Students need to build a foundation based on core topics in biology and mathematics as well as have access to targeted, topical knowledge to understand current problems in biomathematics in general and in our research activities in particular. For instance, in our program both biology and mathematics students need to be comfortable with subsets of the following topics: combinatorial and graph theoretic ideas, DNA chemistry, principles of ecology, genetics, differential equations, experimental design, instrumentation , probability and statistics, data analysis, descriptive statistics, discrete mathematical models, evolution and computer programming.

To address the challenges listed in Section B.6, support the specific research objectives of this proposal and to foster a sense that biomathematical undergraduate research is important and feasible, we propose the development of new courses and the modification of existing courses. These modifications will not occur overnight and they may not even occur across all sections of the courses. However, the successful implementation of biomathematical curricular modifications

in some sections of courses are important to the success of this project and we believe that this success will lead to deeper and more permanent curricular changes.

The proposed new courses are:

(1) A seminar course that will expose biology and mathematics students and faculty to projects and topics in biomathematics.

(2) An upper division course in computational molecular biology, designed for both biology and mathematics students, which would include topics in bioinformatics.

(3) A hands-on molecular biology course expressly designed for mathematics, computer science and other non-natural science majors who want to acquire familiarity with nucleic acids and molecular biology techniques so that they can work on problems at the interface of biology and mathematics. The only prerequisite will be a semester of college-level chemistry.

(4) A course in procedural programming more relevant for our biomathematical students and to include instruction in Computer Algebra System (CAS)technology.

	Fall	**Spring**	**Summer**
Fresh.	★*MA221 CALCULUS* ★BIO117 INTRO. TO BIOLOGY I ★*BIO-MA BIOMATH. SEM.*■	MA 222 CALCULUS II MA233 LINEAR ALGEBRA BIO 203 PRINC. OF ECOLOGY ★*BIO-MA BIOMATH.SEM.*■ *MA228 Calculus for Biologists* *BIO119 Intro. to Biology II* ●*MENTORED RESEARCH* ●*BIOMATH.STUDENT.CONF.*	●*MENTORED RESEARCH* ●*BIOMATH WORKSHOP*
Soph.	MA223 CALCULUS III MA326 DIFF. EQ. CHEM 116 CHEMISTRY I ★*BIO-MA-CSPROC.PROG.*■ ★*BIO-MA BIOMATH. SEM.*■ *MA260 Applied Statistics* *BIO250 Biological Data Analysis* *BIO203 Princ. of Ecology* ●*MENTORED RESEARCH*	MA360 PROB &STATS I *BIO222 Genetics* *BIO215 Plant Diversity* ★*MA237 DISCRETE MATH.* ★*BIO-MA BIOMATH. SEM.*■ ●*MENTORED RESEARCH* ●*BIOMATH.STUDENT CONF.*	●*MENTORED RESEARCH* ●*OFF-SITE LAB EXP.*
Jr.	MATH ELECTIVE INTRO. NUCLEIC ACIDS TECH■ ★*BIO-MA340 MODEL.BIO.SYS.* ★*BIO-MA BIOMATH. SEM.*■ *BIO300 Cell Biology* *BIO330 Molecular Biology* ●*MENTORED RESEARCH*	MA315 COMBINAT. & GRAPH TH. MA360 PROB. STATS II *BIO303 Community Ecology* *BIO390 Molecular Tech.* ★*BIO-MA BIOMATH. SEM.*■ ●*MENTORED RESEARCH* ●*BIOMATH.STUDENT CONF.*	●*MENTORED RESEARCH* ●*BIOMATH WORKSHOP*
Sr.	MA324 ANALYSIS ★COMP. MOLECULAR BIO■ ★*BIO-MA BIOMATH. SEM.*■ *BIO364 Animal Phys.* ●*MENTORED RESEARCH*	MATH ELECTIVE *BI0 Elective* ★*BIO-MA BIOMATH. SEM.*■ ●*MENTORED RESEARCH* ●*BIOMATH.STUDENT CONF.*	●*MENTORED RESEARCH* ●*OFF-SITE LAB EXP.*

TABLE 2: Courses marked with ★ are taken jointly by mathematics and biology majors. Non-credit research activities are marked with ●. Courses in small caps are mainly taken by the mathematics majors and courses in italics are mainly taken by the biology majors. New courses are marked with ■.

Curricular modules (of varying scope)associated with core topics in biomathematics and topics associated with the research activities will be developed for incorporation into both the mathematical and biological and curricula in the following courses or sections of the following courses: 1) MA228 *Calculus II for Biologists* 2) MA221 *Calculus I* 3) MA222 *Calculus II* 4) BIO/MA340 *Modeling Biological Systems* 5) BIO250 *Biological Data Analysis* 6) BIO203 *Principles of Ecology* 7) MA237 I*ntroduction to Discrete Mathematics* 8) BIO222 *Genetics* 9) MA260 *Applied Statistics* 10) MA324 *Differential Equations* 10) MA315 *Combinatorics & Graph Theory*. These courses appear in Table 2 above.

To demonstrate the connection of our proposed activity to the regular curriculum, in Tables 2 contains examples of the "biomathematical careers" two hypothetical students who are recruited to join our project in their first semester. Note, the courses listed in Table 2 are a *proper subset* of the requirements for the respective majors

E   Research Environment and Mentoring Activities

One goal of this proposal is to bring biomathematical undergraduate research more deeply into the thriving undergraduate research culture of SUNY Geneseo. The PIs have made significant contributions to SUNY Geneseo's long and distinguished track record in undergraduate research and in mentoring students in the sciences. Together the PIs have mentored over 90 undergraduates in research projects and/or presentations in the last eight years. These have led to nine publications coauthored with students. Through participating in research, our students have become more engaged in their education and have developed confidence in their own abilities. Almost all of our research students go on to graduate school. We are already involved in undergraduate research projects in biomathematics and our goal is to formalize an undergraduate biomathematical career program at SUNY Geneseo.

By beginning our mentoring in these students in their first year at Geneseo we can support and advise them throughout their academic career and we can significantly improve their experience as research scientists. In this way, these students will be well-trained in biomathematics and well-prepared to further their academic careers after they leave our institution.

In addition to the biomathematics weekly seminar, we will continue our practice of weekly meetings for each of the individual research team(s.) These meetings have been very effective in providing unit cohesion and new insights into the research problems. We will also meet as a larger group to help students prepare for the presentation of their results. A real sense of camaraderie and mutual respect is fostered in these meetings.

We will also meet as a group to hear lectures on topics of interest, outside speakers and students and faculty presentations in Geneseo's regularly scheduled colloquia in both the Mathematics and Biology Departments.

F   Student Recruitment

As the core of our proposal is long-term involvement of students in research projects throughout their tenure at Geneseo, our emphasis will be on recruiting students during their first year at Geneseo. All of the PIs teach freshman courses and the culture of the college is such that we get to know our freshmen immediately. We plan to widely advertise our project and hold public seminars each fall where returning students involved in the research and their mentors will present their research and recruit new students to apply to be a part of the program. We will also invite advanced research students to visit our freshman classes to talk to the class about their research. The regional conference on mathematics and biology that is part of this proposal will also increase student interest in applying to our program by highlighting undergraduate research in this

area. Furthermore, our level-appropriate approach to research mentoring activities will present the opportunities for biomathematical research to first year students. (See Section C.)

> *Although we will emphasize recruitment of students during their first semester at Geneseo, we will recruit and interview students at all stages of their academic career. Students who show promise and commitment to the projects will be encouraged to join.*

Students who apply to the program will be selected based on their high school record, their collegiate and research potential as viewed by their Geneseo professors, their commitment to completing four years in the program, and their future academic plans. We want to encourage our students to go to graduate school.

Geneseo has a strong commitment to diversity and we plan to tailor our recruiting to seek out promising students from diverse backgrounds to become participants in the research projects. We will work with the Department of Mathematics, the Department of Biology, and the Office of Enrollment Planning to identify promising minority students who will be encouraged to apply to the program. Currently, Geneseo's minority population in the sciences is 15% and we are working to increase that number. SUNY Geneseo has outreach programs with local high schools and with institutions such as Brooklyn Tech in metropolitan New York. We also plan to recruit strong minority students from community colleges to matriculate at SUNY Geneseo and to work in our research group. Geneseo is already involved in a partnership agreement with Monroe Community College *(NSF #0418893)* to invite several students from underrepresented groups to do summer research at Geneseo. We plan to draw from, and build upon, this new partnership. Contacts and infrastructure for this partnership are already in place.

As the student population at Geneseo is approximately 60% female, with 50% of our mathematics majors women and 61% of the biology majors women, we have every reason to expect a large percentage of our applicants to be women. Of the eight Geneseo students currently involved in research projects related to this proposal, six are women.

We believe that the recruiting and selection methods outlined above will generate a diverse research group that will be able and productive, which is one of the major goals of the UMB program.

G   Project Management

A list of faculty participants and their specific responsibilities is listed in Section B.7. A. Macula, as Project Director, will be responsible for financial administration of the grant and communication with NSF. Individual research projects will be jointly mentored by a mathematician and a biologist. Management will follow a plan that has already been shown to be effective with the grant participants. Students will meet weekly with their mentors, following the traditional "group meeting" model used in graduate institutions. Meetings will be used to plan new experiments, discuss data and experimental problems, and to provide overview. Specific training will take place one-on-one between a student or group of students and joint mentors.

Should a change occur in personnel, the money would remain at Geneseo and we have a mechanism to make certain that the project is carried out. First, we have a team of two research mentors for each project that lessens the impact of a personnel change. Secondly, we have "backup" PIs who share our goals and are willing to carry out the curricular initiatives and research. (See the end of Section B.8.)

Although the PIs meet together regularly due to common research interests, they will have at least three big organizational meetings a year (after the summer, after the fall semester, and after the spring semester) devoted to evaluating results and determining how the project is progressing.

H   Project Evaluation and Reporting

Our goal is to train undergraduate students of biology and mathematics to be able to perform research at the interface of these two disciplines. We will assess the achievement of our goal by using a mixed model approach, including instruments for summative and formative assessment as well as several specific performance indicators. Each semester, we will survey student participants and will evaluate this feedback in our organizational meetings described at the end of Section G. We also will conduct summative assessment of our program by annually surveying our graduates.

Basic student learning will be assessed through their participation and success in the program's curriculum. (See Section D.) These courses will have interdependent learning outcomes appropriate for students progressing through the program. Evaluations of student performance will be determined using a combination of new rubrics as well as the existing State University of New York's "Basic Research Rubric." Students will be expected to present oral talks about independent research projects.

Specifically, we will gauge success of the program by monitoring the following performance indicators:

- An increase in student enrollment and success in existing cross-disciplinary courses as well as monitor enrollments in proposed curricular activities.

- An increase in the number of student presentations at local and national conferences.

- The writing, submission, and publishing of student-authored and coauthored) research papers in biomathematical journals.

- Our ability to recruit students to SUNY College at Geneseo from feeder institutions (particularly community colleges where such opportunities are not available).

- Our ability to retain and graduate students from biology and mathematics and those with dual majors from these departments.

- Student perceptions after completing the program using summative surveys of all alumni about the value of the program in assisting them in their post-graduate activities.

We have agreements with two outside reviewers to assess this program over the next five years. Our two reviewers, Tom Head (Mathematics) and Susannah Gal (Biology) from the Binghamton University, have experience with assessment and NSF programs. The assessment will involve three site visits; one in the spring of 2005 (during the first biomath symposium,) midterm in the summer of 2007, at the end in the summer of 2009. We will receive their midterm report in the fall of 2007 and their final report in the fall of 2009.

Research findings will be published in research journals. Software and simulation programs will be made publicly available. co-PI, A. Macula et al. ([59]) has already submitted *DNA Code Generator* to the BIOSPICE program. It will soon be freely available to all registered (no charge) users. We plan to continue to submit software to BIOSPICE and other biomathematical community open source programs.

Undergraduate biomathematical training findings and approaches will be published in appropriate journals (e.g., the Journal of College Science Teaching, College Mathematics Journal, American Biology Teacher) We also plan to discuss this project at regional and national biology and mathematics conferences throughout the proposed period and beyond. We plan to organize special sessions on undergraduate biomathematical training at annual and regional meeting of mathematics and biological organizations (e.g., of the Ecological Society of America, Mathematical Association of America.)

References

[1] L. Adleman, Molecular computation of solutions to combinatorial problems, Science, 266, 1021-1024, (1994.)

[2] R. Braich., N. Chelyapov, C. Johnson, P. Rothemund., L. Adleman,. Solution of a 20-Variable 3-SAT Problem on a DNA Computer. Sciencexpress, 1-15,(2002)

[3] N. Alon and J. Spencer. The Probabilistic Method. New York. Wiley (1992.)

[4] Andronescu, Algorithms for predicting the secondary structure of pairs and combinatorial sets of nucleic acid strands, Masters Thesis, University of British Columbia, (2003.)

[5] R. Axelrod, The complexity of cooperation : agent-based models of competition and collaboration, Princeton University Press (1997.)

[6] A. Barabasi and R. Albert. Emergence of scaling in random networks. Science 286: 509-512 (1999.)

[7] E. Baum. DNA sequences useful for computation, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 44, 235-242, (1999.)

[8] BenDor, A., Karp, R., Schwikowski, B., and Yakhani, Z., Universal DNA tag systems: A combinatorial scheme, J. Comp. Biol., 7, 503-519, (2000.)

[9] W. Bialek and D. Botstein, Introductory Science and Mathematics Education for 21st-Century Biologists, Science, 303, 78-790, (2004.)

[10] Bio2010: Transforming Undergraduate Education for Future Research Biologists, National Academy of Sciences. Available at http://www.nap.edu, (2003.)

[11] B. Bollobas, Random Graphs, Academic Press, New York (1985.)

[12] B. Bollobas and O. Riordan, Robustness and Vulnerability of Scale-Free Random Graphs, Internet Mathematics Vol. 1, No. 1: 91-114 (2002.)

[13] M. Boni, J. Gog,V. Andreasen, and F. Christiansen, Influenza drift and epidemic size: the race between generating and escaping immunity, Theoretical Population Biology 65: 179–191 (2004.)

[14] R. Braich., N. Chelyapov, C. Johnson, P. Rothemund., L. Adleman,. Solution of a 20-Variable 3-SAT Problem on a DNA Computer. Sciencexpress, 1-15,(2002.)

[15] S. Brenner,. Methods for sorting polynucleotides using oligonucleotide tags, U.S. Patent No. 5,604,097, (1997.)

[16] S. Brenner. et al., Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrarys, Nat. Biotechnol., 18, 630-634 (2000.)

[17] A. Brenneman and A. Condon, Strand Design for biomolecular computation, Theoretical Computer Science, 287, 39-58, (2002).

[18] H. Cai, P. White, D. Torney, A. Deshpande, Z. Wang, B. Marrone, and J. Nolan, Flow Cytometry-Based Minisequencing: A New Platform for High Throughput Single Nucleotide Polymorphism Scoring, Genomics, 66, 135-143, (2000.)

[19] J. Chen, R. Deaton, Y. Wang: A DNA-Based Memory with In Vitro Learning and Associative Recall. DNA9, 145-156 (2003)

[20] A. D'yachkov and D. Torney, On Similarity Codes, IEEE Trans. on Information Theory 46, 1558-1564, (2000.)

[21] Chinese SARS Molecular Epidemiology Consortium, Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China, Science 303: 1666-1669 (2004.)

[22] F. Chung and L. Lu, The Average Distance in a Random Graph with Given Expected Degrees, PNAS, vol. 99, no. 25, 15879–15882 (2002.)

[23] F. Comellas and M. Sampels, Deterministic small-world networks, Physica A 309 231 – 235 (2002.)

[24] R. Deaton, J Chen, H. Bi, M. Garzon, H. Rabin and D. Wood, A PCR Based Protocol for in Vitro Selection of Noncrosshybridizing Oligonucleotides, DNA Computing, 8th International Workshop on DNA-Based Computers, M. Hagiya, A. Ohuchi (editors), LNCS 2568, Springer, Berlin 196-204 (2002.)

[25] R. Deaton, J Chen, H. Bi and J. Rose, A Software Tool for Generating Noncrosshybridizing Libraries of DNA Oligonucleotides, DNA Computing, 8th International Workshop on DNA-Based Computers, M. Hagiya, A. Ohuchi (editors), LNCS 2568, Springer, Berlin 252-261 (2002.)

[26] Z. Dezso and A. Barabasi, Halting viruses in scale-free networks, Physical Review E., 65: 055103(R) (2002.)

[27] L. Dugatkin, Cooperation among animals: an evolutionary perspective, Oxford Series in Ecology and Evolution, Oxford University Press, New York (1997.)

[28] J. Dunne, R. Williams, and N. Martinez, Food-web structure and network theory: The role of connectance and size, Proc. Nat. Acad. Sci. 99 (20) 12917–12922 (2002.)

[29] A. D'yachkov, Torney, D.,P. Vilenkin, and P. White, On a Class of Codes for Insertion-Deletion Metric, 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland, (2002.)

[30] A. D'yachkov, P. Erdos, A. Macula, V. Rykov, D. Torney, C. Tung, P. Vilenkin, P. White, Exordium for DNA Codes, Journal of Combinatorial Optimization, 7, No.4, 369-380, (2003.)

[31] A. D'yachkov, D. Torney, P. Vilenkin, and P. White, Reverse-Complement Similarity Codes, IEEE Trans. on Information Theory submitted

[32] A. D'yachkov, A. Macula, W. Pogozelski, T. Renz, V. Rykov and D. Torney, A Weighted Insertion-Deletion Stacked Pair Thermodynamic Metric for DNA Codes, to appear, Proceeding DNA 10, Milan (2004.)

[33] A. Dyachkov, A. Macula, V. Rykov, New Constructions of Superimposed Codes, with A. Dyachkov and V. Rykov , IEEE Transactions of Information Theory, 46, NO. 1, 284-291, (2000.)

[34] A. Dyachkov, A. Macula, P. Vilenkin and D. Torney On families of subsets where the intersection of l subsets are not covered by the union of s others, J.Combinatorial Th. Ser. A, 99, No. 2, 195-218 (2002.)

[35] A. D'yachkov, A. Macula, P. Vilenkin, Nonadaptive group testing with error-correcting inclusion matrices, to appear, Bolyai Studies

[36] A. D'yachkov, F. Hwang, A. Macula, P.Vilenlin, C. Weng, A Construction of Pooling Designs with Some Happy Surprises, submitted , Journal of Computational Biology

[37] J. Fewell, Social insect networks, Science 301 1867-1870 (2003.)

[38] N. Friedman, Inferring cellular networks using probabilistic graphic models, Science 303: 799-805 (2004.)

[39] C. Engelhart, A. Macula, M. Sobel, Adaptive group testing for pairs complexes and graphs, to appear in Discrete Mathematics.

[40] C. Engelhart and A. Macula, Sometimes optimal one-stage balance scale searching with errors, Bull. of Inst. Combinatorics, 38, 71-83, (2003.)

[41] P. Erdos, D. Torney, and P. Sziklai, A Finite Word Poset, Elec. J. of Combinatorics, 8, (2001.)

[42] P. Flodman, A. Macula, A. Spence and D. Torney, A new data mining technique for the analysis of simulated genetic data, Proceedings of Genetic Analysis Workshop 12, Wiley-Liss, 2001; Genet Epidemiol (Genetic epidemiology.) 21 Suppl 1: S390-5 (2001.)

[43] M. Girvan and M. Newman, Community structure in social and biological networks PNAS, vol. 99, no. 12, 7821–7826 (2002.)

[44] D. Goldberg, Genetic algorithms in search, optimization, and machine learning, Addison-Wesley Publishing Co., NY (1989.)

[45] D. Gusfield, Algorithms on Strings, Trees, and Sequences, Cambridge, (1997.)

[46] G. Hartvigsen, L. Worden, and S. Levin, Global cooperation achieved through small behavioral changes among strangers, Complexity 5(3): 14-19 (2000.)

[47] G. Hartvigsen, and S. Levin, Evolution and spatial structure interact to influence plant-herbivore population and community dynamics, Proc. R. Soc. Lond. B. 264: 1677-1685 (1997.)

[48] D. He and L.Stone, Spatio-temporal synchronization of recurrent epidemics, Proc. R. Soc. Lond. B 270: 1519–1526 (2003.)

[49] H. Hollman, A relation between Levenshtein-type distances and insertion and deletion correcting capabilities of codes, IEEE Trans. on Information Theory, 39 1424-1427, (1993.)

[50] L. Kaderali, A Desphane, J. Nolan, S. White, Primer design for multiplex genotyping, Nucleic Acids Research, 31, no. 6, 1796-1802 (2003.)

[51] M. Woolhouse, R. May, G. Davies, and B. Grenfell, Modelling vaccination strategies against foot-and-mouth disease, Nature 421: 136-142 (2003.)

[52] K. Keren, R. Berman, E. Buchstab, U. Sivan, and E. Braun, DNA-Templated Carbon Nanotube Field-Effect Transistor, Science 21 November; 302: 1380-1382 (2003.)

[53] J. Kleinberg, Detecting a Network Failure Internet Mathematics Vol. 1, No. 1: 37-56 (2002.)

[54] V. Levenshtein, Efficient reconstruction of sequences from their subsequences or supersequences, Journal of Combinatorial Theory, Series A, 93, 310-332 (2001.)

[55] V. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, Soviet Phys.–Doklady, 10 707-710, (1966).

[56] V. Levenshtein, Bounds for Deletion-Insertion Correcting Codes, 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland, (2002).

[57] V. Levenshtein, Elements of Coding Theory, in the book: Discrete Mathematics and Mathematical Problems of Cybernetics, Moscow, Nauka, pp.207-305, (1974) (in Russian).

[58] A. Macula, DNA-TAT Codes, USAF Technical Report, TR-2003-57, AFRL-IF-RS http://stinet.dtic.mil/cgi-bin/fulcrum_main.pl (2003.)

[59] A. Macula, G. Page, T. Renz and V. Rykov, DNA Code Generator, to be available at https://community.biospice.org

[60] A. Macula, P. Vilenkin, Constructions of superimposed codes based on incidence structures, 2000 IEEE Proceedings of International Symposium on Information Theory, p. 3.

[61] A. Macula, D. Torney and P. Vilenkin, Two-stage group testing for complexes in the presence of errors, with , Medical applications of Discrete Mathematics, DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 55, 145-158, (2002.)

[62] A. Macula and V. Rykov, Probabilistic nonadaptive group testing in the presence of errors and inhibitors, Recent Advances in Experimental Designs and Related Topics, S. Altan and J. Singh, Nova Scientific, 73-86 (2001.)

[63] A. Macula, A combinatorial profiling model for intrusion detection and analysis, Proceedings of the 2000 IEEE Conference on Information Security and Assurance, West Point , NY 47-52 (2000.)

[64] A. Macula, V. Rykov and S. Yekhanin, Trivial two-stage group testing for complexes using almost disjunct matrices, Discrete Applied Mathematics, Volume 137, Issue 1, 27 February, Pages 97-107 (2004.)

[65] A. Macula and L. Popyack, A group testing method for finding patterns in data, to appear in Discrete Applied Mathematics.

[66] A. Macula, Trivial Two-Stage Group Testing with High Error Rates, Journal of Combinatorial Optimization, 7, No. 4, 361-368, (2003.)

[67] A. Macula, Probabilistic nonadaptive group testing in the presence of errors and DNA library screening, Annals of Combinatorics, 3, 61-69, (1999.)

[68] R.May, Uses and Abuses of Mathematics in Biology, Science, 303, 790-793, (2004.)

[69] J. Montoya and R. Sole. Small world patterns in food webs, J. Theor. Biol. 214 405-412 (2002.)

[70] M. Newman, The spread of epidemic disease on networks, arXiv:cond-mat/0205009 v1 30 Apr (2002.)

[71] P. Pevzner, Computational Molecular Biology: An Algorithmic Approach, MIT Press, (2002.)

[72] J. SantaLucia Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, Proc. Natl. Acad. Sci. USA, Vol. 95, pp 1460-1465 (1998.)

[73] J Setubal, J. Meidanis, J. Setabal, Introduction to Computational Molecular Biology, Brooks Cole (1997.)

[74] D. Stauffer and D. Chowdhury, Unified "micro"- and "macro-" evolution of eco-systems: self-organization of a dynamic network, Physica A 336: 102 – 111 (2004.)

[75] S. Roweis, E. Winfree, R. Burgoyne, N. Chelyapov, M. Goodman, P. Rothemund, L. Adleman, A Sticker Based Model for DNA Computation. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 44, 1-29, (1999).

[76] V. Rykov., A. Macula, C. Korzelius, D. Englehart, D. Torney, and P. White, DNA Sequences Constructed on the Basis of Quaternary Cyclic Codes, Proceedings of the 4th World Multiconference on Systematics, Cybernetics, and Informatics, SCI 2000/ISAS2000, Orlando, Florida, July (2000.)

[77] G. Tenengolts, Nonbinary codes correcting single deletion or insertion, IEEE Trans. Inform. Theory, 30, 766-769, (1984.)

[78] M. Waterman, Introduction to Computational Biology, Chapman-Hall, London, (1995.)

[79] D. Watts, and S. Strogatz, Collective dynamics of 'small-world' networks. Nature 393: 440-442 (1998.)

[80] E. Berlow, J. Dunne, A. Barabasi, and N. Martinez . Two degrees of separation in complex food webs Proc. Nat. Acad. Sci. 99(20) 12913–12916 (2002.)

[81] H. Yan, X. Zhang, Z. Shen, N. Seeman, A robust DNA mechanical device controlled by hybridization topology, Nature, 415(6867):62-65 (2002.)

[82] A. Zuker, B. Mathews and C. Turner, Algorithms and Thermodynamics for RNA Secondary Structure Prediction: a Practical Guide