BIOINFORMATICS, GENOMICS, AND THE CHEMICAL ENGINEER
A Perspective

VASSILY HATZIMANIKATIS,[1] DAVID COLLINS,[2] SHAWN LAWRENCE,[3] SAMUEL BROWNING, KELVIN H. LEE
Cornell University • Ithaca, NY 14853-5201

Chemical engineers have made important contributions to a variety of problems relevant to the pharmaceutical and biotechnology industries. Their broad educational background has enabled chemical engineers to work on upstream problems such as the creation and screening of large libraries of proteins for drug discovery or the design and efficient operation of bioreactors, as well as downstream problems in separations and recovery. The development of new technologies in molecular biology and genetics provides an opportunity to explore some of the most basic information about an organism (*i.e.,* the genome) with the goal of understanding how the DNA sequence of an organism relates to its ability to function. Armed with this information, the biochemical engineer can begin to design more efficient products and processes.

As with many such developments in science and engineering, these new opportunities are coupled with several challenges. Among the challenges facing modern biology are difficulties in organizing and extracting useful information from tremendous amounts of raw data, as well as the development of mathematical models that will effectively simulate experimentally observed phenomena and provide a better understanding of complex cellular processes. These issues have motivated the emerging discipline of *bioinformatics.*

Bioinformatics encompasses a wide range of approaches, techniques, and philosophies, and the chemical engineer has an opportunity to make a significant impact both by working in areas within biochemical engineering and by working in collaboration with colleagues from the life sciences, computer science, and applied mathematics.

Spearheaded by Ed Lightfoot and Sang Kim, the AIChE recently recognized the importance of this emerging field and established a Topical Conference on Bioinformatics and Genomics at its annual meeting. With financial support from Cargill and Parke-Davis, three graduate students working on problems in various aspects of bioinformatics and genomics received travel grants to attend the Bioinformatics Topical Conference held at the AIChE annual meeting in Dallas, Texas, from October 31 to November 5, 1999. The fellowships were awarded based on a statement of interest submitted by the applicants.

One of the goals of the students' participation was to provide better exposure to this important new field, to help them develop their own perspectives and ideas on how undergraduate and graduate chemical engineering education

Vassily Hatzimanikatis *is an Assistant Professor of Chemical Engineering at Northwestern and he received his Diploma in chemical engineering in 1991 from the University of Patras in Greece and his PhD in chemical engineering from the California Institute of Technology in 1996. Prior to joining the faculty at Northwestern he worked as a Metabolic Engineer at Cargill Dow LLC. His interests are in computational biotechnology, bioinformatics and functional genomics.*

David Collins *obtained a BS and MEng in chemical engineering from University College London and an MS in Chemical Engineering Practice from the Massachusetts Institute of Technology. He is currently working towards a PhD degree in chemical engineering at Massachusetts Institute of Technology. His interests are in bioinformatics and the management of and modeling of biological data.*

Shawn Lawrence *received a BS in chemical engineering from Auburn University in 1996 and is currently working on his PhD in chemical engineering at Johns Hopkins University. His research interests include using bioinformatics techniques to identify novel genes useful for making therapeutic proteins produced in non-mammalian cell lines more compatible for human use.*

Samuel Browning *obtained a BS in chemical engineering from Massachusetts Institute of Technology and is currently working toward a PhD degree in chemical engineering at Cornell University. His interests are in gene networks and control of the cell cycle.*

Kelvin H. Lee *is an Assistant Professor of Chemical Engineering at Cornell University. He received a BSE in chemical engineering from Princeton University and a MS and PhD in chemical engineering from the California Institute of Technology. His interests are in the use of mRNA and protein expression profiles in the engineering of protein secretion and in the diagnosis of central nervous system diseases.*

[1] *Cargill-Dow Polymers, 15305 Minnetonka Blvd., Minnetonka, MN 55345*
[2] *Chemical Engineering Department, Massachusetts Institute of Technology, Cambridge, MA 02139*
[3] *Chemical Engineering Department, The Johns Hopkins University, Baltimore, MD 21218*

could be enhanced to better prepare chemical engineers for these challenges. This article is one mechanism to communicate these ideas and perspectives, and we hope that it will serve as a reference for chemical engineering education development within the context of the rapidly evolving field of bioinformatics.

BIOINFORMATICS AND THE CHEMICAL ENGINEER

Bioinformatics encompasses many diverse, rapidly developing disciplines, and as a result it has been difficult to describe. One possible definition is the inference of biological principles and facts from experimentally derived data sets, which are often large and complex. The ability to draw such inferences requires many of the same skills taught in the traditional chemical engineering curriculum, including process synthesis, design, and modeling.

Chemical engineers have made significant progress in developing different modeling approaches to biochemical pathways. These models, which are often initially based on experimentally derived information, serve several important roles. They help organize disparate information into a coherent whole, encourage logical thought about which components and interactions are essential to a complex system, and make corrections to conventional wisdom.[1] Among the more valuable tools that have been successfully applied by chemical engineers to the study of biological systems are nonlinear dynamics, optimization, and first-principles modeling.

Bioinformatics encompasses a wide range of approaches, techniques, and philosophies, and the chemical engineer has an opportunity to make a significant impact both by working in areas within biochemical engineering and by working in collaboration with colleagues from the life sciences, computer science, and applied mathematics.

One area in which chemical engineers have *not* made equally significant contributions is in the application of mathematics and statistics to discovery and refinement of large sets of raw data, such as those obtained from large-scale DNA sequencing or mRNA expression profiling studies. Indeed, this area, which includes the development of techniques for visualization of data sets, clustering analysis, and other computational and statistical techniques, is an extremely active area of bioinformatics among computer scientists and physicists. An important question for the community is whether chemical engineers have an appropriate skill set to tackle these new problems.

One feature that distinguishes the chemical engineer from the molecular biologist or geneticist is the "engineering approach" to problem solving. The engineer is prepared to accept incomplete knowledge about a system, asking those questions that are necessary to complete the task and not dwelling on details that are less critical. This approach is highly appropriate for problems in biological research, where many details are unknown and many of the observations are qualitative. But, chemical engineers have to be effective in communicating their unique approach to the biological problems when interacting with the biologists. This is one of the challenges for undergraduate and graduate chemical engineering education, and it is essential in an environment where interdisciplinary collaborations are required in both academic and industrial environments.

So, how might we better train undergraduates to work at the interface of modern biology and chemical engineering science? In the near term, perhaps with little change and only minor tweaking (of course, this new field is only now developing, and these comments may only be applicable in the short term). The core training in chemical engineering is highly applicable to many problems of a biological nature. The combination of skills in mathematics and engineering that most undergraduates are exposed to can be applied to biological problems. Given what appears to be an increase in student interest in biological problems during recent years, the curriculum could be infused with more examples of biological relevance in place of more traditional examples, as well as examples of systems analysis applied to a variety of problems.

One area of the undergraduate curriculum that could be particularly enhanced is in the undergraduate laboratory experience. The traditional laboratory course has been tied to unit operations; future adaptations could include a component on the scientific method to help students understand how one selects an appropriate experimental technique to answer a specific question or how one designs an experiment to obtain high-quality data. These skills will not only help students perform their own experimental designs, but will also help them interpret data and communicate with scientists from other disciplines.

EMERGING AREAS

Many of the current research activities in bioinformatics and genomics are based on parts of the core chemical engineering curriculum. One of the sessions at the AIChE Topical Conference provided a special opportunity to learn about different approaches to bioinformatics and genomics as taken by experts outside the field of chemical engineering. Professor Ruedi Aebersold (Department of Molecular Biotechnology, University of Washington) presented the development of protein analytical technology for the study of biological pathways; Professor Gary Stormo (Department of Genetics, Washington University) discussed his computational and biophysical approaches for the identification and study of genetic regulation; and Professor Michael Savageau (Department of Microbiology and Immunology, University of Michigan) presented the application of mathematical modeling for the analysis of complex cellular phenotypes and molecular processes.

A second session at the Topical Conference provided an opportunity to learn about different problems within the realm of bioinformatics and genomics that are currently being studied by the chemical engineering community. The following represent a few of these problems and their relationship to chemical engineering.

- **John Rogers**, a chemical engineer at Parke-Davis Pharmaceutical Research, has developed a software tool that maps changes in mRNA or protein expression levels to corresponding biochemical pathways. This link provides an opportunity for a scientist or engineer to quickly move from discovery-based science to engineering opportunities. The software enables one to more quickly identify the regulatory and functional relationship between different enzymes and cellular pathways. An analysis of the regulatory control structure governing a particular biological process has been one of the hallmarks of metabolic engineering.
- **Andrea Chow**, Caliper Technologies, discussed the development of Lab-On-A-Chip devices that perform unit operations on a small scale. The development of such technology requires a detailed understanding of separations and of fluid transport.
- **Rob Davis**, University of Colorado, described the optimization of batch reactor schemes to produce RNA more efficiently.
- **Chris Floudas**, Princeton University, discussed computational methods for predicting protein folding. These methods, based on thermodynamics and optimization techniques, provide an opportunity to identify drug targets and to perform *de novo* protein design.
- **Bernhard Palsson**, University of California San Diego, described the use of thermodynamic, transport, and kinetic limitations in modeling metabolic functions. Using mass and energy balances for metabolic reactions catalyzed by proteins with predicted functions, Palsson's group has created a model for the entire known metabolism of *Escherichia coli*. This model allows the prediction of a "phenotypic phase space," a mutlidimensional space that includes all possible metabolic fluxes of an organism with a given set of metabolic pathways. This effort highlights some of the challenges that the engineer will face in the Genomic Era. The model lacks kinetic data because detailed kinetic information on the various enzymes and pathways does not exist. One of the key features of biological systems that challenges bioinformaticists is the"emergent properties" of biochemical networks. Simply, these are nonobvious properties that can arise out of a system of biochemical interactions, such as the integration of signals across multiple time-scales or the generation of distinct outputs, depending on input strength or duration. The ability to predict such emergent properties, using mathematical descriptions of biological processes, will likely come from a coupling between modeling the detailed interactions of molecules and a broader systems-level approach.
- **George Stephanopoulos**, Massachusetts Institute of Technology, noted that while systems analysis provides methods for understanding complex systems, one needs creative formulations of the problems to apply them. And one needs good experimental data. The techniques used to collect informatioin about the entire genome will be only as powerful as the tools available to analyze it[2] and to use it with some predictive power. To fuel this type of creativity in defining and solving problems, a multidisciplinary understanding of critical experimental technology and theory is critical.
- **Michael Shuler,** Cornell University, made an interesting suggestion to promote successful collaboration between chemical engineers and biologists. He noted that each member of a team should be prepared to perform more than 50% of the effort required to finish the project.

CONCLUDING REMARKS

One of the unique features of the chemical engineering field is our common "language." Our profession has been able to work on very different problems and over significantly different length scales, with most of the problems being distilled into the language of chemical engineering. Unlike other engineering disciplines where individuals may be defined by the application, chemical engineering is largely defined by the core training in thermodynamics, transport phenomena, chemical reaction kinetics, and mathematics, and is largely independent of the application.[3] The chemical engineer has an expectation that a process control expert can understand the essence of a seminar in biochemical engineering or that a person with an emphasis in polymer rheology can understand a seminar in microfluidics. Similarly, most chemical engineers expect that they can teach any course in the undergraduate core curriculum—a feature that is not true of other engineering disciplines.

A burden that is placed on the graduate student who is interested in applying chemical engineering skills to the study of problems in bioinformatics and genomics is the need to learn the "life science vocabulary" in addition to the chemical engineering vocabulary—a burden that can add significant time and course requirements to the graduate experience. While the feature of a common language in our profession is a strength that brings us together, it also makes it difficult to think about changing the core curriculum in any significant way. We should take care to maintain our language and ties across the various research topics. At the same time, we should increase the awareness of students and educators about similar emerging areas that bridge scientific disciplines by creating opportunities and forums such as the Topical Conference in Dallas, Texas.

ACKNOWLEDGMENTS

The authors would like to thank Cargill, Parke-Davis, and the AIChE and RANTC for support. SB thanks Michael Shuler; SL thanks Michael Betenbaugh and acknowledges an NSF Graduate Research Fellowship NSF DGE-9843635; DC thanks Paul Barton, Doug Lauffenburger, and Merck and Company; and KHL acknowledges NSF BES-9874938 and DuPont for support. Additionally, we thank Greg Stephanopoulos and Sang Tae Kim at Parke-Davis.

REFERENCES

1. Bailey, J.E., "Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities," *Biotechnology Prog.,* **14,** 8 (1998)
2. Hartwell, L.H., J.J. Hopfield, S. Leibler, and A.W. Murray, "From Molecular to Modular Cell Biology," *Nature,* **402,** C47 (1999)
3. These are ideas that were originated by Prof. William Olbricht, Chemical Engineering Department, Cornell University. ❐

Guiding Principles for Teaching

Continued from page 345.

to wrestle with understanding concepts.

Knowing the students personally is one of the highlights of being a teacher. I enjoy learning from them! Students have a depth that is not always evident just from classroom interactions. Moreover, knowing the students personally creates a rapport that allows us to joke with each other and to feel more at ease. When I know the students, I can also identify those students who are enthusiastic and who are positive role models, and I can encourage them to help set a positive tone in the classroom.

My department chair was enthusiastic in encouraging me to pursue different ideas in the classroom when I first began teaching. I sensed that he did not think of me as I was at that particular time, but instead saw my potential for becoming a good teacher. His attitude made a real impact on me, and I have embraced it in my own interactions with students. I respect them and keep in mind that none of us has yet arrived at our destination—that we are all still part of the process of becoming assets to our society.

I am still involved in the process of becoming a better teacher. The guiding principles detailed in this essay will continue to evolve. Learning new things and adapting them to the classroom keeps teaching fresh and exciting as I continue my journey as both a learner and a teacher.

ACKNOWLEDGMENT

I would like to thank the many colleagues and professors who have shared their teaching experiences, insights, and ideas with me. I am indebted to my colleagues and friends from Rose-Hulman Institute of Technology, particularly my mentors, William Baratuci, Jerry Caskey, and Noel Moore.

REFERENCES

Literature Cited

1. Fulghum, R., *All I Really Need to Know I Learned in Kindergarten: Uncommon Thoughts on Common Things,* Mass Market Paperback (1993)
2. Hesketh, R., and R. Ybarra (workshop leaders), "Undergradute Laboratory," ASEE Summer School for Chemical Engineering Faculty, Snowbird, UT (1997)
3. De Nevers, N., *Fluid Mechanics for Chemical Engineering,* 2nd ed., McGraw-Hill, New York, NY, p. 218 (1991)
4. "Steps Toward Becoming a Self-Directed Learner," *The Teaching Professor,* **10**(4), p. 1 (1996)
5. Felder, R.M., "Any Questions?" *Chem. Eng. Ed.,* **26**(3), p. 174 (1992)
6. "Survival Kit for New Engineering Educators," *ASEE Prism,* p. 30, October (1994)

References on Teaching

7. Davidson, C.I.l, and S.A. Ambrose, *The New Professor's Handbook: A Guide to Teaching and Research in Engineering and Science,* Anker Publishing Company, Inc., Bolton, MA (1994)
8. Lowman, J., *Mastering the Techniques of Teaching,* 2nd ed., Jossey-Basqs Inc., Publishers, San Francisco, CA (1995)

Workshops on Teaching

9. Wankat, P., "Teaching Workshop for New Faculty," ASEE Summer School for Chemical Engineering Faculty, Snowbird, UT (1997)
10. Davidson, C.I., and S.A. Ambrose, "The NSF Engineering Education Scholars Workshop: Preparing Engineering Faculty of the Future," Carnegie-Mellon University, Pittsburgh, PA (1996) ❐