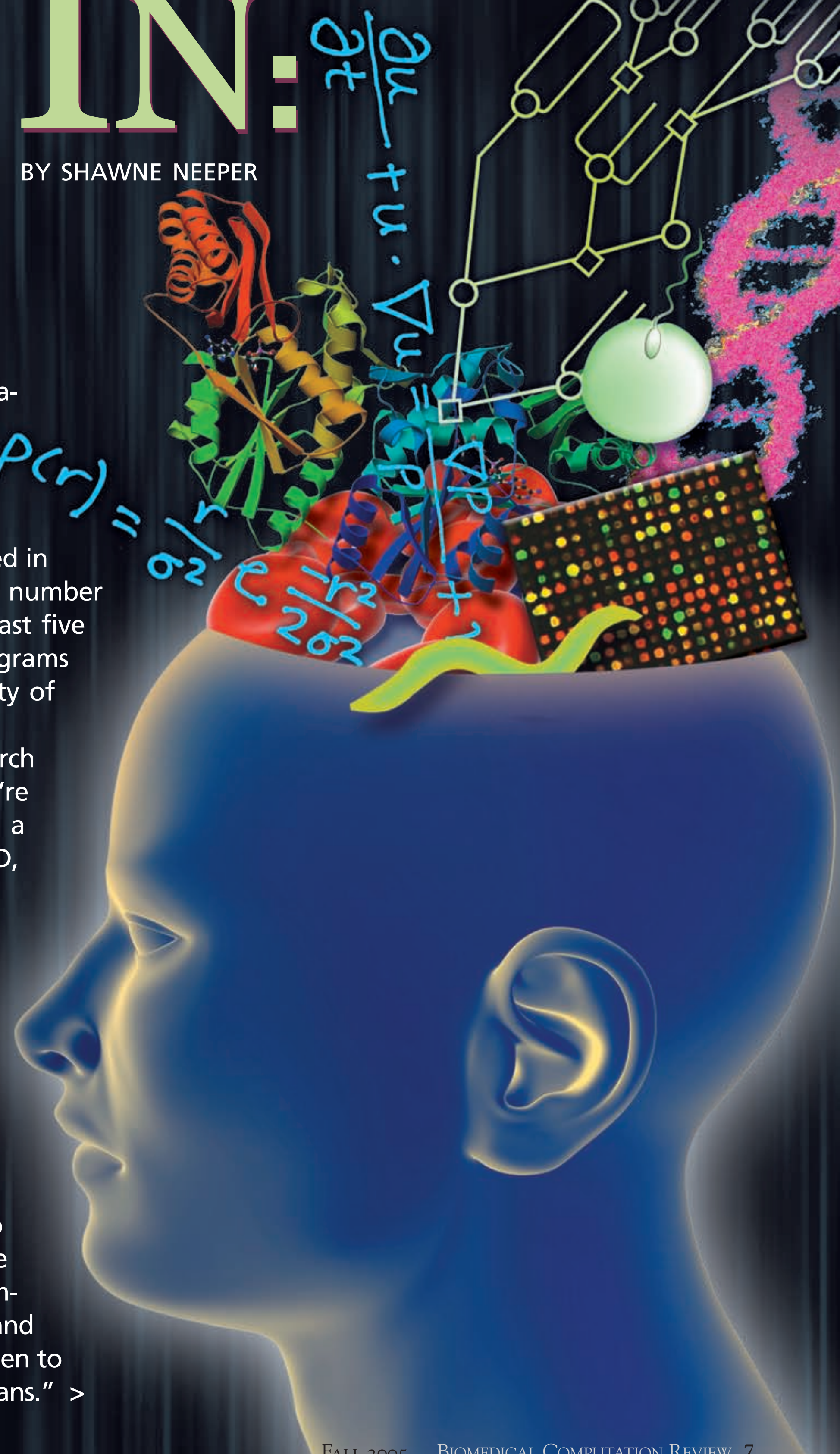
Packing It All In: Curricula for Biomedical Computing

BY SHAWNE NEEPER



The last decade saw a proliferation of training programs at the intersection of life science and computation, with more than 60 new degree and certificate programs launched in the United States alone—and a similar number worldwide. Most appeared within the last five years.[1] The number and variety of programs are growing, in step with the complexity of modern biomedical challenges.

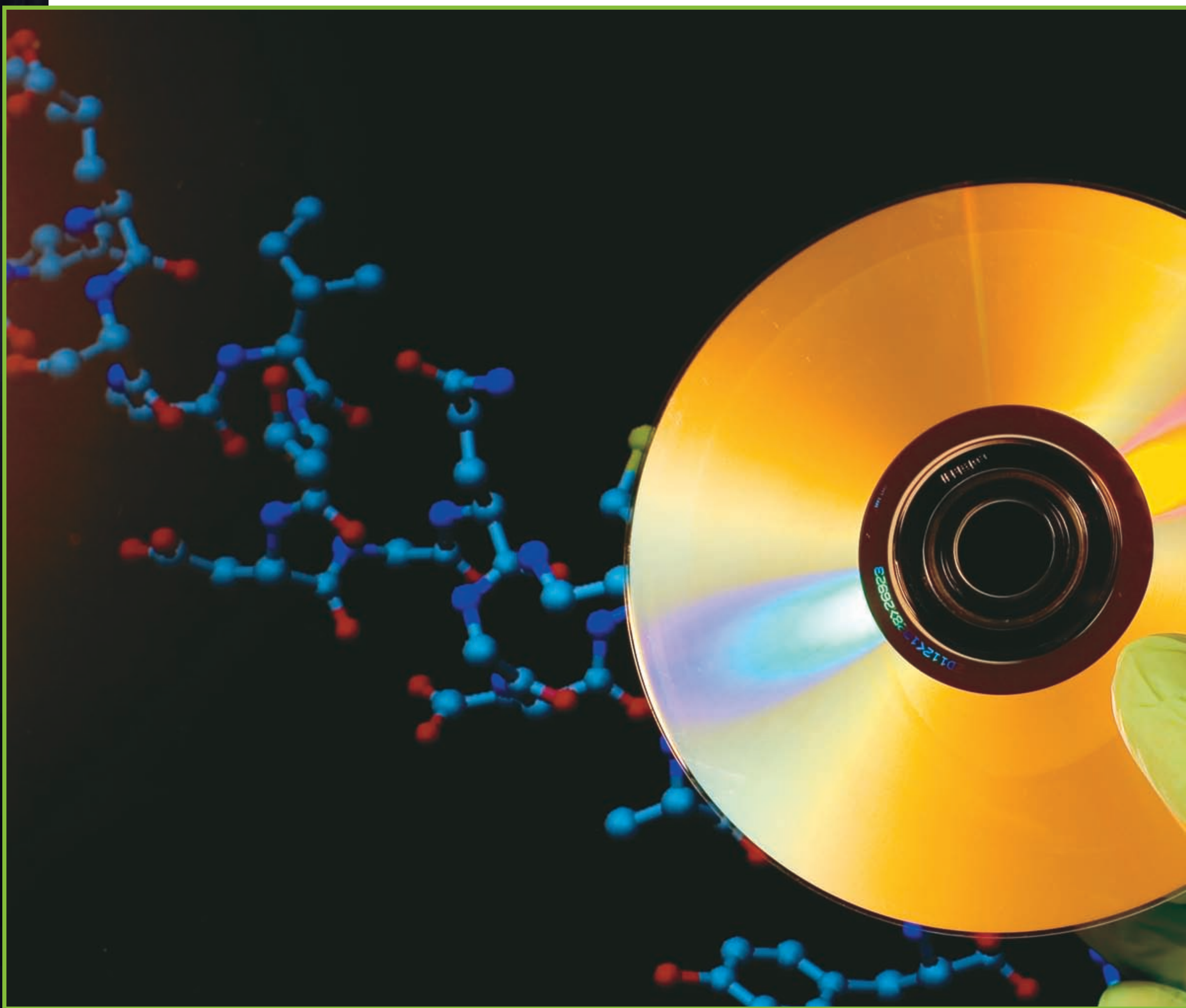
It's the evolution of biological research that's driving educational change. "We're not just sequencing DNA or looking at a gap junction," says John Wooley, PhD, Associate Vice Chancellor of Research at the University of California, San Diego and an active advocate for education in biomedical computation. "We're putting together the total cardiovascular system from the heart to the microcirculation, or the brain from neuron junctions all the way up to cognition." This work requires interdisciplinary education, says Wooley, who is also Senior Fellow of the San Diego Supercomputer Center. "It's going to be the next generation who actually accomplish these goals—students who understand both biology and computing, and can listen to experimental biologists and mathematicians." >

To train that new generation, recent reports from the National Science Foundation (NSF), the National Institutes of Health (NIH) and the Mathematical Association of America (MAA) propose changes across the educational pipeline, from K-12 through postdoctoral.[2,3,4,5] First, they suggest, introduce undergraduate coursework to make all biologists more conversant in math—not just calculus, but discrete mathematics, matrices and probability. Further, add undergraduate biology requirements in algorithm and software design. Finally, create cross-disciplinary graduate and postdoctoral programs to train future theorists and tool builders in math, computer science and life sciences.

How Much Can One Brain Do?

With so much ground to cover, educators at all levels face a tug-of-war between breadth and depth. Quantitative challenges in biomedicine draw on a daunting array of computational skills—across mathematics, statistics and computer science—as well as various strengths in physics, chemistry and engineering. And any topic in biology alone could consume a lifetime's study.

There's a spectrum of opinions about how much one brain can do. "I think depth is a critical issue," says Keith Elliston, PhD, president and CEO of Genstruct, Inc. "If you try to train in two disciplines, you end up with people who are pretty shallow in both." Genstruct needs experts with advanced degrees in molecular biology,



a disease area, or computer science, Elliston says. Genstruct trains the biologists in-house to build or apply drug and disease models. Computer scientists build the underlying software. "I certainly see providing people with some exposure and cross-training, but I think you need specialists first for that to have real value," Elliston says.

What's the best balance of depth and breadth? Genstruct's approach illustrates one end of the spectrum: computational biology as collaboration between experts in the life sciences and computation, with enough cross-training to create common ground. The opposite extreme falls just short of requiring a quadruple major in computer science, math, statistics and biology. New educational offerings span a continuum between Genstruct's position and broad interdisciplinary approaches.

© ISTOCK.COM

> Educators at all levels face a tug-of-war between breadth and depth.

Interdisciplinary training programs must evaluate which skills are essential, and determine where program boundaries should lie. As programs develop, educators are defining new fields and redefining traditional study.

Crossing Disciplines

One way to ensure depth is to train biomedical degree holders in a quantitative field, or vice versa. For example, in the 1990s, progress in genomics motivated the Alfred P. Sloan Foundation to fund postdoctoral grants that cross-trained quantitative scientists in molecular biology. Inspiration for the program came from leading molecular biologists, says Michael Teitelbaum, PhD, program director at the Sloan Foundation. "They were saying things like 'I'm computer literate and numerate, but the volume of data is expanding geometrically, and I'm not a high-powered computer scientist. I don't have the skills I'd need to deal with such large data sets.'"

At the same time, people with PhDs in computationally intense fields such as physics and mathematics faced a difficult job market. So the Sloan Foundation, joined by the Department of Energy, funded two-year postdoctoral fellowships for physicists, chemists, engineers and mathematicians to work in molecular biology laboratories. From 1996 to 2003, more than 60 awards supported cross-training at institutions across the United States.

"Our expectations of the rate of growth were too conservative," Teitelbaum says. "We didn't anticipate the almost vertical take-off that occurred in the 1990s. These people were in high demand when they finished their postdocs."

Experienced molecular biologists also need continuing education to learn how to use new bioinformatics tools as they're developed. For example, the NIH's National Center for Biotechnology Information (NCBI), which produces PubMed and such tools as the Basic Local Alignment Search Tool (BLAST) among others, teaches courses ranging from genomics search strategy to scripting of custom BLAST searches. Course attendees include roughly one-third each of NIH, academic and industry users, says David Wheeler, PhD, head of user services for the NCBI.

Cross-Training for Industry

For people who already have the depth of at least one science degree, there exists another form of cross-training in bioinformatics: the interdisciplinary science master's degree, geared toward people who are already in the workforce.

"Companies need people who can work in teams with molecular biologists, chemists, and other specialists on research and development projects that need computational molecular biology skills," Teitelbaum says.

The Sloan Foundation funds interdisciplinary science master's degree programs,[6] many of which focus on bioinformatics and computational biology. These programs train degree holders in the sciences, mathematics, engineering or a health discipline to bring complementary skills into the workplace. The two-year courses are tightly focused on applications and often include projects with industry partners.

"A significant proportion of our students are people at the prime of their careers," says Betty Cheng, PhD, associate director of Stanford University's Biomedical Informatics Training Program. "They use this program to develop new directions within the corporate setting." One such student was responsible for starting a clinical informatics program in his healthcare organization. Another is applying his new skills to create multimedia health information technology for patients with disabilities.

The Keck Graduate Institute (KGI)

launched its professional master's program four years ago. "[This] is a professional degree, on a different track than research," says Gregory Dewey, PhD, vice president for academic affairs and dean of faculty at KGI.

The program was founded on the premise that industry needs scientifically and computationally astute team members who aren't theorists or bench scientists. KGI graduates find positions in bioinformatics, product development, project management, regulatory and clinical affairs, and sales and marketing—primarily with biotech, pharmaceutical and medical device companies.

Integrated Study

Rather than teaching computation to life scientists, or vice versa, interdisciplinary programs increasingly teach both at once. They aim for breadth, but also to define new fields in bioinformatics and computational biology.

Concurrent training in life science and computation began decades ago with an emphasis on clinical informatics, and grew to embrace genomics, proteomics, and more.

Interdisciplinary programs have proliferated with the help of funding organizations like the NIH's National Library of Medicine (NLM). In 2001, the NLM launched a program that now supports 18 medical informatics research training programs. The programs offer integrated graduate degrees and promote interdisciplinary research.[7]

"We've accepted a potpourri of programs," says Milt Corn, MD, associate director, Division of Extramural Programs for NLM. "The field is so new. We do not prescribe a specific curriculum or organizational structure."

Recently, the NLM inspected all 18 of its programs and found that universities have shaped them in a variety of ways, Corn says. In some cases, programs are divided by topic area. Other universities are trying to put computation under one roof, with the idea that different topics share methodologies. "Which of these models will have ascendancy is not for me to say," Corn says. "Given how our country works, they'll all probably coexist with one another."

"We envision a change in the current approach to graduate school where most of the students take the same courses, join a lab, and work on some aspect of the advisor's project," says Pamela Silver, director of the new systems biology program at Harvard University.

Many life scientists in different fields use similar computational approaches, says Atul Butte, MD, PhD, who recently obtained his doctorate in integrative biology at the Harvard-MIT Division of Health Sciences and Technology. "So I'm more optimistic than many that there are ways to train these folks at the same time."

As an example, Butte points to machine learning. "The computer learns patterns in data, to distinguish *x* and *y*," he says. "For a medical researcher, *x* and *y* could indicate disease state; for a structural biologist, whether a molecule goes here versus there."

Similarly, genomic, systems, and medical researchers often require related methods of classification and tree building, or may share approaches to model building. Butte—who joined the medical informatics faculty at Stanford University this year—argues that these common methods can form the backbone of a core curriculum in computational biology.

Beyond Bioinformatics

Biomedical computation now reaches well beyond the boundaries of bioinformatics, as reflected in emerging graduate programs in computational and systems biology. These programs follow the lead of pre-existing neuroscience and ecology programs by teaching students to use computational methods to model complex biological problems. For example, they might develop algorithms for protein folding, blood flow through vessels, neuron action potentials, or tumor growth. Informatics might play a part in such programs, but it is not necessarily the focus.

> The MIT curriculum is designed around what faculty call the “four Ms”: Measurement, mining, modeling and manipulation.

Harvard, Yale and the Massachusetts Institute of Technology (MIT) have all created new PhD programs in the last two years. Harvard’s new Systems Biology PhD program gets underway this fall with a surprisingly flexible curriculum. It has only two required courses: a survey of faculty research, and an exploration of potential new applications of computation in biomedicine. Unlike most bioinformatics degrees, it has no course requirements in molecular biology.

“We envision a change in the current approach to graduate school where most of the students take the same courses, join a lab, and work on some aspect of the advisor’s project,” says program director Pamela Silver, PhD, professor of biological chemistry and molecular pharmacology at Harvard. “We would like to get away from that model and empower students more.”

Students will have two faculty advisors in their first year: one in physical or quantitative science, another in biology. But they are encouraged to propose inter-lab collaborations for their thesis work and may select an advisor from anywhere in the university. The program was announced in the fall of 2004. By December, about 100 applications had arrived from students of computer science, biology, chemistry, engineering, medicine, and more.

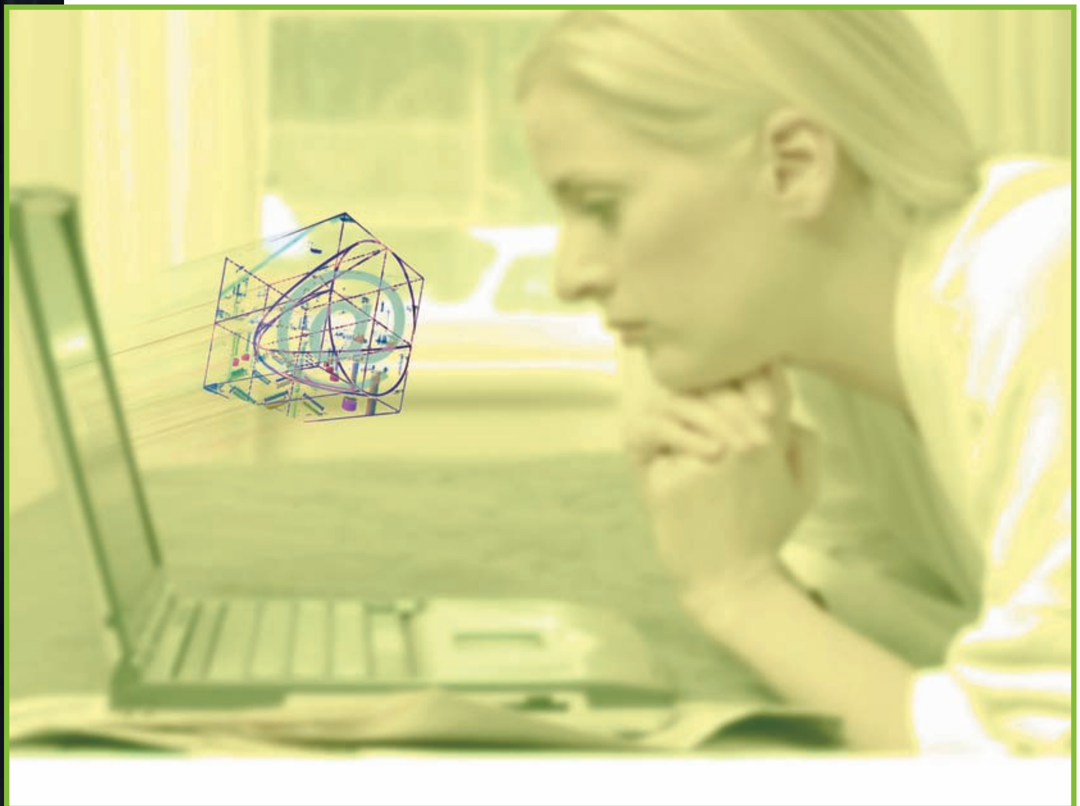
Founded in 2004, MIT’s PhD program in computational and systems biology offers similar flexibility to the Harvard program. The curriculum is designed around what faculty call the “four Ms”: Measurement, mining, modeling and manipulation. Classes stress real-world examples to explore the integration of these approaches in modern research.

“Research these days is multi-investigator and multi-site,” says Bruce Tidor, PhD, professor of bioengineering and computer science at MIT and director of the new program. “Learning how to work in that environment is crucial. Many of our students will have more than one supervisor, and thesis projects will integrate their different expertise. By learning ways of solving problems in each of their research groups, our students will be in a position to bridge disciplines.”

© PHOTOS.COM

Many of the new PhD programs overlap with research areas in existing programs at the same institutions. For example, students interested in PhD study in computational biology and bioinformatics at Columbia University can apply to a new interdepartmental program, or pursue similar research through existing PhD programs in biochemistry, molecular biophysics, biomedical informatics, biological sciences, computer science, electrical engineering or applied mathematics and applied physics. So why have both options? The difference is in the design: the new PhD programs, such as the one at Harvard, span the other departments, and more. “It gives us a degree of flexibility you don’t have in a traditional department setting,” Silver says. “It allows us to define the field as we go.”

MIT’s program favors applicants with dual backgrounds in the life sciences and a quantitative discipline, already prepared for immersion in cross-



© PHOTOS.COM

disciplinary work. And more students are now coming to graduate school with that kind of background.

Starting Younger

Some in the field believe that a more thorough integration of computation and biology should begin before graduate school. But the conflict of breadth and depth intensifies in younger students who haven't yet had time for depth. How can educators include more quantitative work in an already crowded biology curriculum?

Several universities have already launched undergraduate degrees in bioinformatics or systems biology. These programs can be rigorous—often the equivalent of a double major—and frequently must balance biology and computational content.

For example, George Washington University's departments of biology and computer science offer a dual degree in computer science and biology, or a degree in computer science with a bioinformatics specialization. The latter includes all typical premedical courses in chemistry and physics, a full computer science curriculum, and special classes in bioinformatics and computational biology. This leaves room for only two biology classes: introductory and molecular biology.

The University of California, Santa Cruz Department of Biomolecular Engineering offers a BS degree in bioinformatics with similar requirements, including two biology courses and three upper-division bioinformatics classes.

Most undergraduate degrees narrow the scope by focusing on bioinformatics. Some venture further, but it's tough to reach consensus about how and what to teach undergraduates in a broad, new field. "When you say computational biology, it's not just one thing," says Bruno Olshausen, PhD, principal investigator at the Redwood Neuroscience Institute and associate professor of neurobiology, physiology, and behavior at the University of California, Davis.

Olshausen and other faculty were asked to design an undergraduate program at UC Davis. "To some extent, a core curriculum is possible; modeling methods are similar enough," Olshausen says. However, he says, at UC Davis, genomics faculty wanted to include courses in databases, which he would not require in neurosciences. He, on the other hand, might want students to take a course in the theory of computation, which genomics faculty might not agree to. "It's going to be difficult to create one unified computational biology curriculum."

"It's going to be difficult to create one unified computational biology curriculum," says Bruno Olshausen, PhD, a professor at the University of California, Davis.

Stanford's undergraduate degree in biomedical computation addresses the issue of breadth and depth directly. Students select a "depth" track in biology or computation, with two corresponding "breadth" tracks in the other. There are four tracks: molecular and cell biology, organ and organismal biology, informatics and simulation. So, for example, a student choosing depth in organ and organismal biology would study both informatics and simulation approaches to that area.

"An element of the solution is to not make curriculum for biomedical computation simply the sum of the component fields," says Scott Delp, PhD, chairman of Stanford University's Department of Bioengineering and co-director of Simbios, a National Center for Biomedical Computation. "By definition, then they're only getting introductory material in each field. By carefully designing curricula that teach biological and physical, math and computation concepts together you can achieve a much

"By carefully designing curricula that teach biological and physical, math and computation concepts together you can achieve a much higher degree of efficiency and depth at the same time," says Scott Delp, PhD, a professor at Stanford University.

higher degree of efficiency and depth at the same time."

Interdisciplinary courses and study modules can benefit all biology students, argues a recent publication of the National Academies Press, called *Bio2010: Transforming Undergraduate Education for Future Research Biologists*.[8] The report includes a long list of computational concepts important to modern biology, for inclusion in revised undergraduate programs. This and a more ambitious suggestion to integrate undergraduate courses across sciences[9] have met with mixed reviews.[10,11,12]

Society for Mathematical Biology president Louis Gross, PhD, points out the difficulty of fitting separate courses covering those concepts into already busy curricula.[13] However, he cites successes at the University of Tennessee, which introduced essential mathematics concepts—beyond calculus and statistics—in a redesigned introductory course for biologists. The class teaches math themes through life science examples. He endorses this concepts-wise integration of quantitative skills throughout undergraduate biology curricula.[14]

Towards A Consensus

Some educators are trying to develop unified recommendations for teaching computational biomedicine. A series of workshops held at national and international meetings has built the basis for an upcoming report, "Educating Bioinformaticians for the Challenges of 21st Century Biology: A white paper for the National Science Foundation and the International Society for Computational Biology." It will be submitted to the ISCB's education committee for comment, and shared with the NSF and appropriate Department of Energy and NIH organizations.

"We hope to inspire them to further discussion," says co-author John Wooley, who is shepherding the new white paper's growth.

Inspired by similar white papers in Canada, Japan and elsewhere, the document suggests goals and curricula for degree and certificate programs in biomedical computation, and addresses strategies to better integrate biological and quantitative science education in the United States.

"To make progress in the 21st century we need the very powerful approach that bioinformatics and computational biology offer," Wooley says. "And we need students who understand both biology and computing."

The next decade will reveal how the interdisciplinary approach fares as more graduates hit the job market. A 1999 survey of 16 schools shows that many with interdisciplinary undergraduate and master's degrees found positions in software development, pharmaceutical and biotechnology industries. Most PhDs and postdocs, on the other hand, continued with academic careers. As interdisciplinary education evolves, a new generation of scientists will emerge. Only time will tell which blends of computation and biomedicine will rise to the fore.

REFERENCES

1 See http://www.iscb.org/univ_programs/program_board.php for a partial list. See also http://bioinformatics.org/faq/#study.

2 Palmer M, Arzberger P, Cohen J, Hastings A, Holt R, Morse J, Sumners D, Luthey-Schulten Z (2003). Accelerating mathematical-biological linkages: Report of a joint NSF-NIH workshop held February 12-13, 2003 at the National Institutes of Health. Bethesda, MD.

3 Hastings A, Arzberger P, Bolker B, Ives T, Johnson N, Palmer M (2002). Quantitative biology for the 21st century. Report from NSF funded workshops on quantitative environmental and integrative biology, September 7-9, 2000 and December 11-13, 2002, San Diego, California.

4 Byer B, LeBlanc M (2002). Meeting report: Incorporating genomics research into undergraduate curricula. *Cell Biology Education* 1(4).

5 The Mathematical Association of America, Curriculum Reform and the First Two Years, summary reports in biology and health/life sciences. See http://www.maa.org/cupm/crafty/summary_reports.html.

6 See http://www.sciencemasters.com/fields.php.

7 See http://www.nlm.nih.gov/ep/GrantTrainInstitute.html.

8 National Research Council of the National Academies, Division on Earth and Life Studies, Board on Life Sciences (2003). *Bio2010: Transforming Undergraduate Education for Future Research Biologists*. Report of the Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, Performed by the National Academy of Sciences with funding from the NIH and Howard Hughes Medical Institute.

9 Bialek W and Botstein D (2004). Introductory science and mathematics education for 21st-century biologists. *Science* 303(5659).

10 Diniz-Filho J (2004). Back to the future: the 1960s view of statistics and mathematics for 21st-century biologists. e-letter response to *Science*.

11 Fambrough D (2003). Review of: *Bio2010*: Read it! *Cell Biology Education* 2(2).

12 Steitz J (2003). Commentary: *Bio2010*—New challenges for biology educators. *Cell Biology Education* 2(2).

13 Gross, L (2004). Interdisciplinarity and the undergraduate biology curriculum: finding a balance. *Cell Biology Education* 3(2).

14 Gross L (2000). Education for a Biocomplex Future. *Science* 288(807).