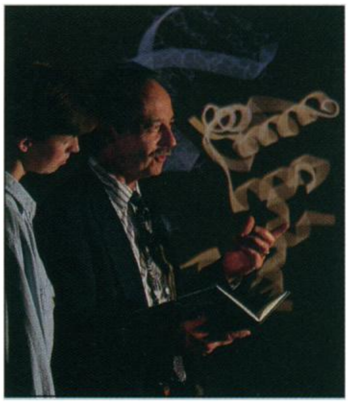
Hot Property: Biologists Who Compute

Drug companies, hoping to mine biological databases for hidden treasures, are raiding the universities for experts in "bioinformatics"

Ask anyone in the rarefied world of bioinformatics about job prospects these days and you'll hear the same thing: This field is hot. In fact, says Chris Rawlings, a British researcher with a Ph.D. in molecular genetics and a background in computer science, the top names in computational biology have become "so hot you can't touch them." A pioneer in this field, Temple Smith of Boston University, tells of a postdoc in his lab who is entertaining offers from five companies, trying to decide which he likes the most. Sean Eddy of Washington University in St. Louis, a developer of new statistical techniques for identifying genes, says: "You get a phone call from a headhunter, and when you hang up, you can hear the phone ring down the hall" as the headhunter tries to recruit colleagues.

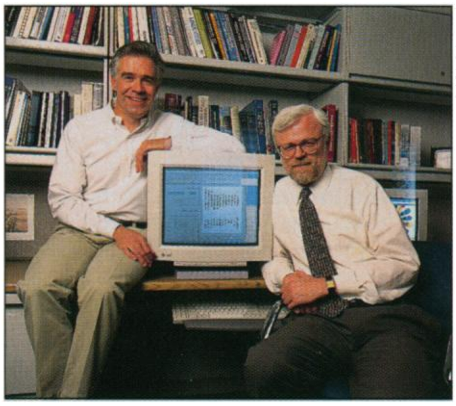
It was only about 6 months ago that pharmaceutical companies began buying up experts in number-crunching and molecular genetics. Now some of the biggest drug firms in the world—including SmithKline Beecham (SB), Merck, Johnson & Johnson, and Glaxo Wellcome—are vying with one another to hire experts in bioinformatics. At the same time, smaller biotech firms, some of them pioneers in computational biology, are finding it difficult to get the staffers they want, as they must fish for talent in the same pond as the big companies. The frantic bidding has driven salary offers sky-high. It's not uncommon, according to Smith, for a person with a master's degree and good credentials to be offered a starting salary of $65,000. Ph.D.s are getting offered $90,000 or more. And some observers worry that the privatization of expertise could deplete academic centers of the talent needed to educate the next generation (see box).


F. SWAY/BOSTON UNIVERSITY

In demand. Temple Smith says his students are getting multiple offers.

Driving this talent hunt is the explosive growth of genetic data accumulating in public data centers. Initially, it was academic researchers who were most enthusiastic about the possible uses of DNA sequence data. They found that simple organisms—yeast and bacteria, for example—share many genes with humans, suggesting that relatively easy-to-obtain animal genomes could shed light on human development and disease. In addition, it became clear that genetic data could be used to understand not only familial patterns of disease but also the day-to-day biochemical processes within cells, as genes are turned "on" and "off" (see p. 1736). Many small companies sprang up to provide computational tools for academic researchers interested in these topics, as a few large companies began to invest in private databases and customized software to help them identify genes and proteins in DNA sequence data.

Then the field exploded, just as the federal government announced that it was ready to begin funding large-scale sequencing of the human genome. As an executive at one big company says, people have come to believe that genetic data are becoming "the major driving force" in drug discovery, and that by the turn of the century, DNA data banks will be "the source of most, if not all, new drug targets." Companies that want to stay in the race have suddenly realized that, to explore this new territory, they will need employees who are adept in computer science, information technology, and genetics—a rare combination of skills.


GEORGE TATE

Industry-bound. David Searls (*right*) and James Fickett, recruited by SmithKline Beecham.

It's a heady time for those who have these skills. Many bioinformatics leaders who spoke with *Science* said that they had gone through graduate school twice—once for biology and a second time for computer science. Few set out deliberately on this arduous path. But those who followed it are now being rewarded. "I've never seen such an intense recruiting environment—it's ridiculous," says Rawlings, who accepted an offer himself from SB this year. Rawlings, 41, left Britain's Imperial Cancer Research Fund in London, where he had been working on gene mapping and cancer genetics for 15 years, joining SB's unit in Harlow, just north of London.

SmithKline's gamble

Rawlings joined a rapidly expanding team at SB. The company's big move came in 1993 when it signed a $125 million deal with Human Genome Sciences (HGS) of Rockville, Maryland, gaining access to a large database of fragmentary human DNA sequences that HGS had created. Since then, according to staffers, SB has built up a staff of 24 to help manage the flow of genetic data, analyze results, and distribute information to staff biologists.

Last fall, SB marked another critical step along this path. David Searls, a 44-year-old former computational linguist, announced that he was leaving his tenured post at the University of Pennsylvania to take a job at SB's center in Upper Merion, Pennsylvania, where he would start up a new bioinformatics research group. Recruiting this ex-linguist was a coup, says an academic colleague, noting that Searls is respected for the quality of his research, which uses methods of grammatical analysis to identify genes in DNA sequences.

Shortly after Searls's arrival, SB signed up three more stars: James Fickett, 43, of the Los Alamos National Laboratory, a computational biologist who "was almost present at the creation" of GenBank, says Searls; Randall Smith, 44, a top bioinformatics professor at Baylor College of Medicine in Houston; and Rawlings, in London. Searls says that this string of new hires signals the "kickoff" of a recruiting effort that will continue through this year and beyond. It's intended to double the size of SB's bioinformatics team to about 60.

The new group, according to Searls, will do more than offer the traditional support provided by computer-system staffs: It will initiate research projects on its own. "The

Demand Outstrips Supply

When the phone rings in Doug Brutlag's office, the odds are better than even that the person on the other end will be a headhunter, not a scientist, he says. "They all want my students," sighs Brutlag, a researcher in bioinformatics at Stanford University. He adds: "I only graduate one student a year; they're all snapped up right away into industry." Yet, even as companies are scouring the campuses for people with the skills needed to run complex analyses of genetic data, Brutlag says, universities have not recognized bioinformatics as a discipline. The result: There are few training grants to support young researchers in bioinformatics, few tenured faculty positions, and no large dedicated centers.

Like many of his peers in academe, Brutlag sees this situation as ironic. While the companies recognize the value of the special training his students get, the universities do not. And, as companies sequester talent in private labs, universities will find it more difficult to build up bioinformatics programs to instruct the next generation of researchers. "I just wish some of the students could be left to teach other students," Brutlag says. But the incentives to stay and teach often cannot compete with companies' high salary offers and commitment to the field.

Leaders in bioinformatics who spoke with *Science* say that universities have not created centers dedicated to the field, although pockets of excellence—none of them set up to attract students—can be found at the University of Pennsylvania, Stanford, Washington University in St. Louis, and a handful of other universities. Even these are having trouble hanging onto top faculty, however. The director of Penn's program until recently was David Searls, a computational linguist who developed algorithms designed to find "meaning" in DNA code. Last fall he became chief of the bioinformatics group at SmithKline Beecham, a pharmaceutical company in Philadelphia. Christopher Overton, who has succeeded Searls, says, "We're in the process of creating a center for bioinformatics," but funding remains uncertain.

Penn's program, however, is better off than some. Take Baylor College of Medicine in Houston, which has been a leader in the field. After C. Thomas Caskey, head of the genetics group there, resigned to take a top R&D post at Merck last year, the group's computer-wise biologists scattered. Victor Solovyev, an émigré Russian physicist who developed a program called "Gene Finder," went to Amgen Inc.; another bioinformatics expert, Charles Lawrence, went to Sequana; and Randall Smith, who held one of the few tenured posts in the field of bioinformatics, went to join Searls at SmithKline Beecham. A clean sweep.

Solovyev says he quit because he couldn't get a grant to create or improve Gene Finder, a project he says the National Institutes of Health rated below the funding line. Yet Solovyev notes that a competitive group of his peers at the University of California, Santa Cruz, ranked Gene Finder one of the best on a long list of similar genetics programs. Baylor is now out on the recruiting circuit along with company headhunters, looking for a bioinformatics chief to take the position vacated by Smith.

Outside industry, the most vigorous center of bioinformatics in the United States, according to many, is the National Center for Biotechnology Information (NCBI) at the National Library of Medicine, in Bethesda, Maryland. As manager of GenBank and other public databases, NCBI's primary task is to help biologists and others make use of archived sequence information. NCBI has also built up a strong research team and is trying to help support the next generation of experts in the field by funding some training grants. Staffer Mark Boguski says NCBI supports 16 doctoral or postdoctoral trainees at present. In addition, the National Science Foundation supports 11 postdoctoral fellows in a math and molecular biology program, and the Sloan Foundation together with the Department of Energy supports another five.

NCBI's counterpart in Cambridge, England, the European Bioinformatics Institute, supports a few graduate students interested in bioinformatics, as does the Sanger Center nearby. But, according to Sanger's Richard Durbin, the problem from his viewpoint is not so much inadequate funding for graduate studies as inadequate means of getting the word to students that they are wanted.

One nascent development could ease the situation, however: Pharmaceutical companies are beginning to see the value of these academic programs and are planning to support them with grants. The Burroughs Wellcome Fund, for example, intends to award 5-year, $2.5 million grants this fall to five research groups working in the general area of computational biology, and other companies are planning to jointly fund a bioinformatics effort at the Jackson Laboratories in Bar Harbor, Maine. Overton notes that such support serves the industry's best interests. Otherwise, the bioinformatics world could end up "eating its seed."

–E.M.

emphasis … will be on what I would call academic values," Searls says, but "I don't think we'll be doing much in the way of ivory tower or pure theory research." Instead, Searls will ask staffers to team up with SB's bench scientists to solve practical problems: "I would like to set up cycles where you develop computational models for some biological system, then test them at the bench, then take the results of those experiments and refine the model."

Searls envisions four major categories of research. The first, he says, is traditional gene hunting. Many programs have already been developed to help scan sequence data to find significant patterns—mainly by identifying a match between a known gene or protein and a new one. More sophisticated techniques must now be developed, Searls says, to describe the basic traits of families of genes and identify outliers and obscure relatives. Second on Searls's list is "knowledge management." The main task under this heading is to "make use of the babble of information that's out there" by integrating heterogeneous data into a coherent program and delivering it to the researcher's desk in a simple format. Developing better visual aids, for example, should make it easier to identify significant patterns in the "gray sequence data." Third, SB's team will focus on methods of mapping genes and proteins to their physical point of origin on the genome and developing methods of cross-referencing each map to others. Fourth, Searls says, is the big challenge—devising algorithms that will pull useful information about the structure and function of unknown genes from sequence data.

Forging links

Although no other company has yet gambled on bioinformatics as aggressively as SB, most are quickly building up their in-house capabilities and forging links with academic centers to tap into the top talent. For example, Merck & Co. of Whitehouse Station, New Jersey, supports a "core staff" of seven in bioinformatics, according to group director Keith Elliston, and a distributed staff of about 20 more. One of the core group's jobs is to manage the flow of information from Merck's Gene Index, a public database containing human gene fragment sequences produced by academic researchers whom Merck supports (*Science*, 28 October 1994, p. 538). Elliston, who has been at Merck since 1988, says that his group "concentrates on analysis" of genetic data. In addition, it maintains an independent, proprietary database of sequence information, makes the information accessible to Merck researchers at four sites, and tailors software for intramural use.

Another pharmaceutical giant, Glaxo Wellcome of London, hired its first bioinformatics staffer 3 years ago, but its effort expanded in 1994, after the company recruited Rainer Fuchs, a well-known leader in the field, from the European Molecular Biology Laboratory in Heidelberg, Germany. Fuchs now heads a group at Glaxo's research center in Research Triangle Park, North Carolina, with "a strong genetics component," working on computer programs that will recognize "unexpected patterns" in sequence data representing new genes and proteins. Glaxo's European staff focuses primarily on protein modeling. Fuchs estimates that about 35 people are working full-time on bioinformatics throughout the company, including the protein group in Geneva and another group in Britain that supports Glaxo's databases.

But Fuchs says Glaxo is pacing itself: "We have been very careful in trying to choose people we hire; you won't see us advertising for 20 positions. We depend more on our informal networks, and we will use those networks to identify good people and bring them in." Nor is the company interested in supporting basic informatics research in-house, although Fuchs says staffers make basic discoveries as a "side effect" of the work they do.

Pfizer Inc., in Groton, Connecticut, is taking a similar tack. Ian Williams, group director of molecular sciences, says that the company will rely on researchers at universities and government-funded labs such as GenBank to do much of the fundamental work in bioinformatics, particularly developing new data-sorting algorithms. While the company will need staffers to "develop complex relational databases," integrate information from heterogeneous sources, and develop pattern-recognition software, Williams doesn't foresee a major intramural effort in this area. At present, Williams says, there are about six bioinformatics experts on staff.

Pfizer is, however, making some strategic investments in bioinformatics outside its own walls: It is one of seven pharmaceutical giants that have invested more than $100 million in Incyte Pharmaceuticals of Palo Alto, California. This small company has developed a proprietary collection of human gene sequences—similar to HGS's—and is marketing the information as a service. Pfizer, according to Williams, is also planning to spend about "one third of a million dollars" on bioinformatics research at academic centers, and possibly more, if the initial results look good.

Among the more recent bioinformatics initiatives is one by Wyeth Ayerst of Philadelphia, which recently advertised for four bioinformatics specialists, including a director of a new "core technology group," to help develop a research strategy for drug applications. Johnson & Johnson of New Brunswick, New Jersey, has called for a project manager and postdocs to develop new information systems to analyze sequence data.

As big companies consider how best to hire or borrow the number-crunching expertise they need, the medium and small fry are competing intensely with one another for scarce and increasingly expensive talent. Phil Green of the University of Washington, Seattle, says he's heard that recruiters are looking for bioinformatics experts for a slew of companies, including Darwin Molecular, Genentech, Mercator, Millennium, and Sequana. Zeneca Pharmaceuticals, Schering-Plough, and others are also searching for talent.

No one can predict, of course, whether the big investment in intramural staffing of the kind made by SB will pay off. But one thing is certain. As federally funded projects fill the public databases with human and mouse DNA sequences over the next few years, the demand for well-trained computational biologists who can make sense of this torrent of data can only increase.

–Eliot Marshall

NASA

Goldin Drops Plan for New Institutes

NASA has scrapped an ambitious plan to create a half-dozen nonprofit science institutes after failing to win White House support for the administrative and legislative changes required to get them up and running. The decision will force NASA Administrator Dan Goldin to find other ways to revitalize research at the agency's dozen centers through stronger ties with academia.

Last May Goldin and his top science managers announced their intention to convert civil servants into employees of nonprofit institutes, operated mostly by universities, that would focus on specific disciplines (*Science*, 26 May 1995, p. 1122). The goal was to shrink the size of NASA's payroll while broadening contact between agency researchers and the larger scientific community. In a 3 April letter to Jack Gibbons, the president's science adviser, Goldin explained that the institute idea "represent[s] a positive response to the Administration's budget challenges" as well as a means to preserve and improve the agency's science.

But opposition from other government agencies has forced Goldin to fold his tent. On 7 June he ordered his managers to halt work on an astrobiology institute at Ames Research Center in Mountain View, California, and a microgravity sciences institute at Lewis Research Center in Cleveland. Officials had already decided that plans to spin off two institutes at Marshall Space Flight Center in Huntsville, Alabama, and a third from Langley Research Center in Hampton, Virginia, required further study.

The idea of private institutes was never very popular among researchers, who felt that NASA had failed to define the concept clearly. "I didn't quite understand where the institutes were going," says Claude Canizares, chair of the National Research Council's Space Studies Board and a Massachusetts Institute of Technology astrophysicist. "It was all quite confusing."

NASA

Slipping away. Institute may have continued Ames's blood-pressure work with snakes.

However, what doomed the plan was its impact on the federal government's personnel and ethical-conduct policies, as well as fiscal uncertainties and skepticism from Congress (*Science*, 17 November 1995, p. 1109). The Office of Government Ethics was loath to exempt would-be employees of the institutes from conflict-of-interest regulations that apply to workers who leave government to work for private organizations that receive federal contracts. And the Office of Personnel Management objected to NASA's request that civil servants be exempt from laws that prohibit employees from retaining health benefits and contributing to their pension fund after leaving the government. "Without [legislation], it's not going to work," says a NASA official.