fields, like John Hennessy, a pioneer in reduced–instruction-set computing (RISC) and president of Stanford University, in California, routinely visit to give lectures.

Opportunities for advancement within the PTO follow traditional management routes. Primary examiners and supervisory examiners perform almost the same functions as new examiners, but are subject to fewer reviews of their work. Primary examiners, a level that can be achieved in four years, sign their own work and have a remarkable degree of autonomy. Reaching the level of supervisory examiner takes about 10 years. A manager might find herself in Geneva, Switzerland, as the U.S. representative "on loan" to the World Trade Organization, or in the White House as an IP advisor.

Examiners love the flexible work schedule. The PTO's MaxiFlex program lets workers put in 80 hours every two weeks in any schedule that suits them. Other perks include law school tuition reimbursement after two years of service (provided the person returns to the PTO after graduating). Salaries last summer got a boost—new examiners make in the US $50 000 to $70 000 range, depending on work experience, plus overtime and bonuses based on productivity.

Legal eagles

Patent examiners who leave the PTO are often recruited by private law firms eager to employ a PTO "insider," as happened with Philip Marsh. The draw here boils down to money, potentially lots of it, as well as prestige and power. Patent attorneys in big-name law firms easily make in the mid–six figures.

Law firms specializing in patent law boast that many, if not all, of their professional staff have technical degrees. Having such in-house expertise is a necessity. In theory, anyone can file a patent application, but the intricacies of the process generally demand the services of a patent agent or patent attorney. The main difference between an agent and an attorney is a law degree, but surprisingly, a law degree is not required in this line of work. What is required is that the attorney or agent register with the PTO, indicating that he or she has a college degree in some technical discipline and has passed the PTO's registration exam.

An inventor engaging a law firm to obtain a patent may thus only meet with a patent agent, not an attorney. But this is not short-shrift treatment; again, an agent can do anything to win a patent that an attorney can. After the patent is granted, though, only an attorney can draw up a contract to assign the rights to the patent; inventors who are employees of companies typically assign their rights to their companies, which then hold the patent. Once the patent is granted, it becomes like any other property— it can be sold, protected, and shared.

And, of course, stolen. The other main function of IP law firms is to protect their clients' patents. Litigating is done by the attorneys, who sue infringers on the inventor's behalf and represent him or her in court. Some large firms also employ technical staff to do research in support of litigation.

Patent litigation can be heady stuff. Millions or even billions of dollars may be at stake, which is why the richest corporations turn gladiator in court, throwing in every resource at their disposal to defend their IP. Disagreements may involve the subtlest of subtleties—whether a gear with 14 cogs is functionally the same as one with 13, or whether one Web site's one-button checkout is the same as another's. Vast fortunes, corporate destinies, and social trends await the outcomes. ●

Next month, IEEE Spectrum *will introduce a department on intellectual property, including patents, legal issues, and practical advice on navigating the roiling waters of IP.*

Mining the Genome

Having tackled the human genome, Celera's Gene Myers is now advancing on proteins and drug development

BY ASHTON APPLEWHITE
Contributing Editor

Gene Myers' interest in computation kicked in during the early 1970s, while he was an undergraduate at the California Institute of Technology (Pasadena). "It was extremely exciting to me that you could basically write something down, feed it to a machine, and have a program that actually does things—accomplishes tasks, manipulates data, answers questions," he recalled.

Years later, having established himself as a leading researcher in computer science, Myers wrote a program that certainly did something: it assembled, in record time, a map of all the DNA in a human cell—the genome—for the life sciences company Celera Genomics (Rockville, Md.). Celera's horse race with the Human Genome Project consortium rocketed the company into the public eye. It also brought to light the extraordinary advances being made in computational biology, a field that Myers had gotten into before it even had a name.

A beautiful solution

Myers' first taste of this kind of work came through Andrzej Ehrenfeucht, his thesis advisor at the University of Colorado (Boulder). There Myers studied what were then among the hardest problems in the analysis of genetic data. He developed algorithms to determine the most thermodynamically stable configuration of RNA, which plays a vital role in determining a cell's protein structures. He liked the work because "the problems were interesting, they were new, they were solvable, and they had elegant solutions. It all comes down to your sense of aesthetics."

After receiving his Ph.D. in computer science in 1981, Myers headed for the University of Arizona. Though he'd enjoyed his graduate research, the field was still nascent, and so he focused on more traditional problems in computation.

In 1985, a university colleague of Myers' proposed setting up a center for computational biology, with the goal of analyzing the genetic sequence of *E. coli*, a widely studied bacterium. Needing a computer scientist, the professor tapped Myers. A distinguished panel from the National Institutes of Health (NIH) roundly rejected the proposal, citing its poorly constructed overall plan.

They liked Myers' work, though, declaring "the kid's not bad," and invited him to one of the first-ever gatherings of computational biologists, held in Waterville Valley, N.H., in 1985. "The buzz was just incredible," Myers recalled. "I was interacting with scientists and biologists, not just with computer guys....It was fun to be with scientists who wanted to understand how the universe worked."

By 1985, researchers studying the genomes of various organisms had sequenced a total of 5.2 million base-pairs, the chemical "letters" that spell out the genetic code. Complete genomes, with hundreds of millions of base-pairs, were on the horizon. Of necessity, biologists looked to computation for handling information about genomic sequences. To elucidate a newly discovered human gene's function, for example, computer scientists wrote algorithms to compare the gene to those from laboratory mice whose functions were already known.

The shotgun method

Back in Arizona, as Myers worked his way up the ranks to full professor, he grew interested in what geneticists call assembly: how to determine the order of the DNA sequences in the genome. Rather than proceeding through the genome in ordered sections, he looked at "shotgun" sequencing.

In this approach, several copies of a portion of DNA are each broken into millions of segments short enough to be sequenced automatically by machine. The short sequences are then reassembled, in order, using the overlapping portions as landmarks. The automation allows shotgun sequencing to proceed far faster than traditional methods. But comparing all the tiny pieces and matching up the overlaps requires massive computation.

Conventional wisdom held that shotgun sequencing wouldn't work on segments of over 30 000 base-pairs. Researchers remained skeptical even after Rob Fleischmann, Hamilton Smith, and Craig Venter in 1995 shotgun-sequenced the two million base-pairs of *H. Influenzae*, which causes bacterial meningitis.

The skeptics pointed out that the human genome is 1500 times as large and contains lots of repetitive sequences, which makes assembly harder because they are hard to distinguish from one another. Myers likens that problem to assembling "a jigsaw puzzle with large areas of blue, like a map of the world, as compared to areas with high relief where you can intuit where the pieces belong." What is more, shotgunning a whole

genome requires that the DNA be sequenced many times over before the assembly can even begin.

The trick, Myers concluded, was first to identify where the repetitive sequences lay. After that, he said, simulations showed that he "could deliver 98 percent of the genome, no problem."

Someone else was thinking along those same lines: a geneticist named Jim Weber at Marshfield Clinic in Wisconsin. He asked Myers to collaborate with him. But when the two applied to the NIH in 1996 for a $12 million pilot grant to shotgun-sequence the human genome, "everyone told us we were nuts," Myers recalled cheerfully.

Soul of a new machine

All might have been lost, but around the time Myers and Weber were being rejected, a new machine was being developed by Applied Biosystems Inc. (Foster City, Calif.) to do automatic gene sequencing [see "Gene Sequencing's Industrial Revolution," *IEEE Spectrum*, November 2000, pp. 36–42]. The ABI Prism 3700 had 10 times as much throughput as its predecessors, which radically reduced the wait time for results. "Psychologically, that was a huge sea change," Myers said.

In 1998 Celera was formed by ABI's parent company, PerkinElmer Corp., to apply the shotgun approach to the human genome, "pretty much the way Jim Weber and I had spelled it out," said Myers. Myers took a leave from Arizona to join the start-up as vice president of informatics research. His task was to build the assembler: to write the program that would take the 27 million random reads of 600 letters each that would come off the three hundred Prism 3700s, then put them together to reconstruct the genome's nearly 3 billion base-pairs.

It was a daunting task. Starting in January 1999, and over the next 15 months, Myers' team wrote half a million lines of code. The team was headed by Myers and Granger Sutton, who in turn hired "a NASA physicist, a couple of great software engineers, a mathematician from the National Bureau of Standards, and a couple of great German computer scientists," Myers said.

"We didn't sleep much," he added. "But it was really a wonderful time—we knew that this was it."

First the fly...

The scientific community learned as much when the Celera team showed up at the Genome Sequencing and Analysis Conference in Miami in September 1999 and presented their results on the *Drosophila* genome. As a pilot project, they had done three million reads of DNA fragments to assemble all 120 million base-pairs of the fruit fly's genome, "and it worked," Myers said.

Gene Myers' team at Celera Genomics wrote the program to assemble the human genome's 3 billion base-pairs.

If scaled upward to the human genome, the program used to sequence *Drosophila* would have demanded an unattainable 600 GB of memory, but Myers and his team continued to refine their algorithms. "Today we can do the human [genome] assembly in 6000 CPU hours with 24 GB of memory," he reported. Assembly of the *H. Influenzae* genome now takes less than five minutes, compared to 25 hours back in 1995.

As Celera researchers turned their sights on the human genome, they were up against some stiff competition. Formed in 1990, the publicly funded Human Genome Project had been using a much slower (but to their minds, more accurate) "hierarchical" form of sequencing. First, the genome was broken up into overlapping segments whose relative locations were known; each segment was then shotgun-sequenced.

In June 2000, Celera and its public counterpart jointly announced the completion of a "rough draft" of the human genome. But the race to complete the genome had generated animosity between the two groups. Though Myers would not comment on it, the controversy has not abated: a recent pair of papers in the *Proceedings of the National Academy of Sciences* debates Celera's whole-genome shotgun approach [see To Probe Further, p. 81].

Miracle cures to come?

Although Celera's initial business plan was to sell its genome database, it recently announced its move into the pharmaceutical business. To that end, it is now developing high-throughput mass spectrometry to identify what a specific protein is up to in a given sample of cells. Comparing protein levels in normal versus diseased tissues can help doctors to diagnose particular conditions, such as prostate cancer, which is now detected only in blood tests.

Diagnostics should come on quickly, Myers predicted. "If a protein is occurring overabundantly, maybe it's the one you want to build a small molecule—a drug—against," he said. What excites Myers is the evolution of drug development from serendipity to science, "from stumbling on penicillin to saying, 'This is how I want to affect this bacteria in order to stop it.' We're entering the age of rationally designed therapies, and that's very exciting." Computation will play a big part in this, he added, "and I'd like to see computer scientists and engineers contribute."

The shift in focus is welcomed by Myers' team, which he calls "the informatics think tank for Celera. We go off and provide novel computation techniques wherever they're needed." It also fits in with Myers' own interests, which have turned toward the inner workings of cells. "Though I still love solving computational challenges, my work is really directed toward trying to understand how cells work at the molecular level," he said. "Over the next five to 25 years, [we'll] begin to get the complete circuit diagram of the molecular machinery that operates the cell. I don't know what's going to come out of it, but it's going to be phenomenal." ●