Over a 30-year career, his lab has contributed to numerous branches of molecular biology and founded a few. Here are some of the highlights.

David Baltimore: A Short Portrait of a Long Career

by Douglas L. Smith

On May 13, it was announced to the world that Caltech's next president would be biologist and Nobel laureate David Baltimore. Here he chats with two of the people who persuaded him to take the job. From left: Kip Thorne (BS '62), Feynman Professor of Theoretical Physics and chair of the faculty search committee; Baltimore; and Gordon Moore (PhD '54), chair of the Board of Trustees.

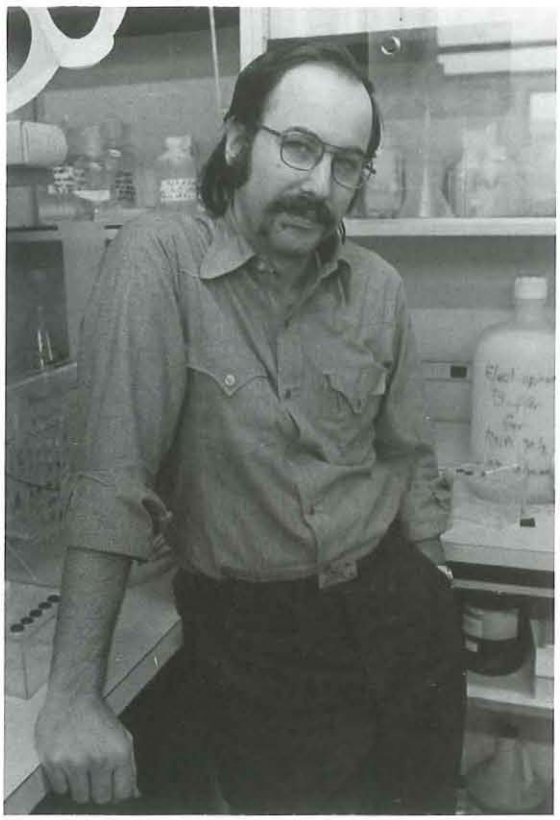
If you've been reading the papers at all, you doubtless know that David Baltimore, Cottrell Professor of Molecular Biology and Immunology at MIT, founding director of the Whitehead Institute, and former president of Rockefeller University, has agreed to become Caltech's fifth president. He should fit in admirably here—he has been a virus man all his life, and viruses are about as small as biological systems come, so Caltech's diminutive size should have a natural appeal. And like Caltech he does small science in a big way—over a 30-year career, his lab has contributed to numerous branches of molecular biology and founded a few. Here are some of the highlights.

Baltimore earned his BA in chemistry from Swarthmore College in 1960, and he finished off his PhD in biology at Rockefeller University in a remarkable three years (he actually got the sheepskin in '64). Postdoctoral positions followed at MIT and the Albert Einstein College of Medicine for a year each, leading to a three-year stint at the Salk Institute in La Jolla, California, where he continued in the burgeoning field of molecular biology, specializing in research on the polio virus. He returned to MIT in 1968 as an associate professor, becoming a full professor in 1972.

Even then, although Baltimore would spend most of the next three decades at That Other Institute of Technology, a Caltech connection had been formed. He had spent the summer after his high-school junior year at the Jackson Laboratory in Bar Harbor, Maine, where Howard Temin was the resident guru. Temin, who died in 1994, had taken his PhD at Caltech in 1960 under Renato Dulbecco, who in turn had come to Caltech in 1949 to work with Max Delbrück. At Caltech, Dulbecco developed (with Marguerite Vogt) the techniques needed to grow animal viruses in culture. He then headed south to the Salk Institute in 1962, where he would become one of Baltimore's mentors. Baltimore, Dulbecco, and Temin would share the 1975 Nobel Prize in physiology or medicine.

Temin and Baltimore independently and simultaneously (the two papers were published back-to-back in the June 27, 1970, issue of *Nature*) discovered an enzyme dubbed "reverse transcriptase" by the unfortunately anonymous *Nature* correspondent who wrote up the finding in that journal's "News and Views" section. (An enzyme is a protein molecule that expedites a specific chemical reaction by providing a pocket into which the reactants can nestle in just the right orientation for the reaction to proceed.) Temin had hypothesized the existence of reverse transcriptase—although not by that name—in 1964, but the idea was considered so far-fetched that most biologists dismissed it out of hand. Reverse transcriptase allows a molecule of ribonucleic acid (RNA) to copy itself into deoxyribonucleic acid (DNA), swimming against the current of information flow.

By the early 1960s, everyone knew that genes, which are the blueprints for making every protein an organism will ever need, are encoded in DNA molecules in the cell's nucleus. The DNA consists of two strands carrying complementary information, like a photograph and its negative, and the two strands mesh together like a zipper. At the appropriate time, the nuclear machinery unzips a portion of the DNA to expose the gene's negative strand and makes positive prints in the form of RNA, a chemically very similar molecule. The messenger RNAs then leave the nucleus and go out into the cytoplasm (the soupy gel that makes up the bulk of the cell) and say to protein-making machines called ribosomes, "Here. Make this." "This" could be a piece of cellular machinery, an intracellular regulator that turns other genes on and off, or even a signal to other cells—for instance, it might tell adjoining cells in a developing embryo that it's time to start dividing and become liver tissue. But information, it seemed, never traveled backward—there was no way to touch up a DNA negative from an RNA print.



Baltimore in the lab in the early days.

People also knew that RNA can be a repository for genetic information in its own right. A host of viruses had been discovered that have no DNA in them, but only RNA—either double-stranded like DNA, or a single positive strand. In either case, the positive RNA strand instructs a ribosome to make a protein, called RNA polymerase, that in turn makes a negative RNA strand—out in the cytoplasm, mind you, not in the nucleus—from the original positive RNA print. The RNA polymerase then makes many prints from the new negative, and the new prints fan out to all the other ribosomes in the cell, co-opting them into making new viruses. Thus, all it takes is one positive RNA print at one ribosome to launch an infection. No DNA is needed, nor is the cell's nucleus involved. In fact, many of these processes will even occur in cytoplasmic extracts from which all the nuclei (and with them, the DNA-handling machinery) have been removed.

But Temin had discovered that the Rous sarcoma virus, a single-positive-strand RNA virus that causes cancer in chickens, can't infect cells whose DNA-handling apparatus has been shut down. This led him to postulate that the viral RNA must somehow be getting translated back into DNA as a prerequisite to converting the cell into a cancer cell.

Meanwhile, back at MIT, Baltimore was trying to apply the techniques he'd developed for studying polio (another single-positive-strand RNA virus) to vesicular stomatitis virus (VSV), a line of research catalyzed by his postdoctoral fellow Alice Huang, an expert on VSV. (Huang joined his lab at the Salk Institute and returned with him to MIT, where they married in 1968. She later became a professor at Harvard Medical School, and is now dean for science at New York University.) Baltimore, Huang, and a graduate student, Martha Stampfer, rapidly discovered that VSV contained a single negative strand of RNA. How, then, could the infection get started?

This hunch was verified in short order, which got Baltimore thinking further. If a virus could carry around an enzyme to make RNA, why not one to make DNA? The two molecules are very similar, after all. Perhaps Temin was on to something.

Simply running the negative RNA strand through the ribosome won't suffice. Any bit of protein you make will come out all wrong, but the odds are the ribosome won't get very far along the negative RNA before running into a so-called stop codon. A stop codon shuts down the ribosome and ejects the completed protein and normally appears at the tail end of the RNA strand. But in a world where black is white and white is black, a stop codon is just as likely to appear in the middle of the RNA. In fact, it's a sure thing. There are no meaningless instructions in the RNA code, so the "negative" versions of stop codons are valid assembly instructions. Obviously, then, the "negative" versions of those assembly instructions are stop codons, and as soon as the ribosome comes across one—and at every step it has roughly a 5 percent chance of doing so—it turns itself off. So at best, you'll get a useless snippet of protein.

Since nobody had ever found RNA polymerase in the cytoplasm of a normal, uninfected cell, Baltimore concluded that the VSV must be bringing not only a negative RNA strand, but also a working RNA polymerase molecule into the cell in order to get the infection started. Once this RNA polymerase had used the negative strand as a template for assembling a positive copy of the virus RNA out of ingredients scavenged from the cytoplasm, the print could run through the ribosome in the usual way.

This hunch was verified in short order, which got Baltimore thinking further. If a virus could carry around an enzyme to make RNA, why not

Above: The Nobel class of '75 in Stockholm. It was a large group, as only the literature and peace prizes were unshared that year. From left: Tjalling Koopmans (economics), Leonid Kantorovich (economics), Aage Bohr (physics), Eugenio Montale (literature), Dulbecco, Vladimir Prelog (chemistry), Ben Mottelson (physics), Temin, John Cornforth (chemistry), Baltimore, and, from the Caltech class of 1939, Leo James Rainwater (physics).

Above, right: Baltimore in a less formal moment, at the MIT press conference that followed the announcement of the Nobel Prize in October.

one to make DNA? The two molecules are very similar, after all. Perhaps Temin was on to something. As Baltimore said in his Nobel acceptance speech, "Luckily, I had no experience in the field and so no axe to grind—I also had tremendous respect for Howard dating back to my high school days when he had been the guru of a summer school I attended at the Jackson Laboratory." And sure enough, Baltimore's lab found that another cancer-causing single-positive-strand RNA virus, called the Rauscher virus, contains a working molecule of what has come to be called reverse transcriptase. Rather than creating more RNA, the reverse transcriptase makes a negative strand of DNA from the positive RNA, followed by positive DNA that binds to the negative DNA in the nor-mal, two-stranded fashion—all in the host cell's cytoplasm. This viral DNA then sneaks into the nucleus and splices itself into the regular DNA, where it gets handled just like the cell's own DNA. It's now known that all cancer-causing RNA viruses get their carcinogenic genes into the host cell this way. (Temin was simultaneously making the same discovery with the Rous sarcoma virus. Dulbecco was awarded the prize for unrelated work that indicated that once a cell has become cancerous, this new state is genetically stable—a finding that dovetailed neatly with Baltimore's and Temin's discovery.) Viruses that insinuate their own genes into the host cell's genes are now known as retroviruses and include in their number HIV, the AIDS virus.

Recalls Baltimore, "After I discovered the reverse transcriptase, I worked for some time on the biochemistry of the enzyme and tried to understand how it actually carried out the process of reverse transcription. That was my first foray into working in cancer-inducing viruses, or cancer at all. I decided that if I was going to go any further, I needed a biological system to work with, and I was introduced by a lucky accident to the Abelson virus, which wasn't at all well known at that time." The Abelson virus induces leukemia in mice. Leukemia is a cancer of the blood, but instead of the cancerous cells congregating in a tumor, they circulate individually. Normal cells divide a fixed number of times—foreordained in their genes—and then quit reproducing and die. Cancer cells, with their altered genes, instead go on dividing forever. Leukemia cells are actually aberrant white blood cells, whose oversupply turns the blood milky and gives the disease its name—leukemia is Latin for "white blood."

Naomi Rosenberg, then a postdoc in the Baltimore lab, developed a method for infecting normal mouse cells with the Abelson virus and then culturing them in vitro, in glass petri dishes. This provided an infinite supply of cells to experiment on, allowing each step of the cancer-inducing process—of which the introduction of a cancer-causing gene, or oncogene, into the cell's DNA is just one facet—to be studied at the molecular level. The process by which a healthy cell turns into one of the undead is a complex molecular ballet, the choreography of which is still being charted worldwide. Up to that point there had been no easy way to study mammalian leukemia in vitro—researchers had to use chicken cells instead. But a bigger payoff awaited—one that would establish a new branch of immunology.

One of the body's chief defenses against infection is protein complexes called antibodies, which constitute about 20 percent of the free protein circulating in our blood. Antibodies ferret out alien substances by means of a pocket that recognizes and binds to an invader—for example, the pocket might fit snugly over a protein that only exists as part of a virus's protective coating. The bound antibody then summons nearby white blood cells to engulf and devour the intruder. Each antibody's pocket is tailor-made to fit one specific shape, but the immune system has to be alert for an infinite number of potential threats. Some of the menacing shapes—viral mutations

that might occur in the future, for example—don't even exist, yet the system has to be ready for them when they appear. Antibodies are secreted by white blood cells called B cells, each of which produces an antibody with one specific pocket. Bone-marrow tissue continuously cranks out generic B-cell precursors which, after a few days spent choosing the antibody they'll make for the rest of their lives, become mature B cells. So, given that a B cell's nucleus can't hold an infinite amount of DNA, how could the immune system store the potential to generate what Baltimore estimates to be in the vicinity of 100 billion possible antibody pockets? Well, while Rosenberg was perfecting the culture system, Susumu Tonegawa was discovering that snippets of DNA spontaneously rearranged themselves in nascent B-cell nuclei. (Tonegawa would win the Nobel Prize in 1987.)

As the Baltimore lab analyzed the Abelson cultures, it became apparent that Rosenberg had unwittingly infected (and thus immortalized) B-cell precursors. "When we realized that," recalls Baltimore, "it occurred to me that this might be a way of studying the events of immunodifferentiation." Finding out how a B cell decides what to do when it grows up has since turned into a growth industry. The Baltimore lab's contribution to this field perhaps culminated in the late 1980s and early '90s, when grad students David Schatz and Marjorie Oettinger performed a classic series of experiments that revealed the enzymes that actually carry out the rearrangements.

But finding those enzymes was just the beginning. Each enzyme has to appear at the right time in relation to the others, or the DNA they assemble will be a useless mishmash that won't make a functional antibody. The genes that make the enzymes are controlled by one or more activation sites that lie in stretches of the DNA that are adjacent to the gene proper. Each activation site has its own regulatory protein that recognizes and binds to it. Baltimore's lab was able to locate these sites, which then became the bait that enabled the regulators themselves to be fished out of the cellular soup. These regulators, a class of proteins called transcription factors, collectively coordinate the overall sequence of events—besides turning the genes they regulate on and off, many transcription factors bind to (and thus affect the behavior of) other transcription factors. Tracing the interplay between the transcription factors is yet another field of research the Baltimore lab has spun off.

Finding out how a B cell decides what to do when it grows up has since turned into a growth industry.

One of these transcription factors, published in 1988, is called NFκB. At the time of NFκB's discovery, the Baltimore lab found that it binds to a segment of DNA that helps to synthesize one part of the antibody complex. The researchers therefore assumed that NFκB also played a role in antibody production. Further research, however, indicated that this probably isn't the case. But at around the same time, the Baltimore lab began working on the AIDS virus. It now appears that, regardless of what NFκB's "real" job may prove to be, it also plays a large role in controlling HIV production in a class of white blood cells called T cells. Says Baltimore, "In active T cells, it might contribute 90 percent of HIV production."

In all the excitement of unraveling immunodifferentiation, the Abelson oncogene itself hasn't escaped scrutiny. The gene is related to a normal gene named *abl* after the Abelson viral gene, which was discovered first. (By convention, italics are used for gene names; the protein produced by that gene has the same name, but in Roman letters.) The *abl* gene is somehow involved in DNA repair and in the formation of the cytoskeleton, which is the protein-fiber trusswork within the cytoplasm that holds a cell in shape. (It's presumably the former function, when set awry by a few strategic mutations, that enables the viral version to cause cancer.) What *abl* does, exactly, is still unknown, but it codes for an enormous protein that weighs more than 13,000 carbon atoms. This protein appears to be a sort of Swiss army knife. It consists of several independent units, called domains, at least some of which are involved in intracellular signaling—the means by which cells coordinate such collaborative processes as tissue growth, wound healing, digestive-juice secretion, and embryonic development. The protein's signaling domains are somehow related to its DNA-repair and cytoskeleton-formation functions, but nobody has yet figured out how.

A cell communicates with another cell by secreting molecules that bind to receptor molecules on the target cell's surface. Forwarding that message to the appropriate destination within the cell—intracellular signaling—takes a chain of events that rivals the complexity of Rube Goldberg's finest machines. But instead of a bowling ball knocking over a watering can that fills a bucket that pulls a string that fires a pistol that eventually causes the bread to be toasted, the receptor molecule (for example) cuts free another molecule that seeks out yet another molecule and together they trigger yet another molecule to do something further, and eventually the cell divides or does whatever else the signal told it to do. The Abl protein's signaling domains are components of such intracellular pathways.

The first Abl signaling domain to give up its secrets to the Baltimore lab was a new twist on an old method of governing protein activity within the cell. It's called a tyrosine-specific protein kinase (catchy name!), and it was also discovered independently at the Salk Institute. A protein kinase attaches a phosphate group (a cluster of four oxygen atoms bound to one phosphorus atom) to an amino acid, in this case one called tyrosine. All proteins are made up of long strings of amino acids, which interact with one another through their electric charges, the degree to which they

Collectively, these three domains are involved in coordinating many aspects of cell division, cell differentiation, and cellular activation—just about everything a cell would "want" to do.

attract or repel water, and their stiffness or floppiness, among other things. This web of forces folds the protein up into its preferred shape. Thrusting a phosphate group (which has a negative charge, seeks out water, and is bulky) into the web distorts the balance of forces and affects the protein's shape. And as we've seen over and over again, the protein's shape confers its function—proteins are built around pockets that are designed to do something. In some cases, the phosphate group completes a missing part of the pocket, turning the protein "on." In other cases, the phosphate group obstructs or alters the pocket, turning the protein "off."

Other labs had discovered kinases that attach phosphate groups to the amino acids serine and threonine, but this was the first kinase that sought out tyrosine. The knowledge that tyrosine-rich proteins are amenable to the same control systems as serine- and threonine-rich ones has opened up new intracellular signaling pathways to explore. But the finding has much broader implications. Many classes of enzymatic reactions are controlled by the attachment and removal of phosphate groups—a discovery for which Edmond Fischer and Edwin Krebs won the Nobel Prize in 1992.

Baltimore's lab then went on to discover two more Abl domains that are very similar to the Src (pronounced "sarc") protein, which is also involved in intracellular signaling in as-yet-unknown ways. The *src* gene, as its name suggests, is also related to an oncogene, its cancerous cousin having been discovered in the Rous sarcoma virus we met earlier. (It's easy to imagine how, if the Src protein is part of a signaling cascade that tells the cell that it's time to divide, having a mutated protein stuck in the "on" position can lead to runaway cell division and cancer.)

One Src-related domain, discovered in 1986 and called Src-homology region two, or SH2, binds to the tyrosine-phosphate units created by the tyrosine-specific protein kinase the Baltimore lab had previously discovered. (The kinase itself is SH1, but people rarely call it that.) In other words, this binding event is the next step down the signaling pathway initiated by the kinase. "That is a major event in cell signaling," says Baltimore. "We spent a lot of time characterizing the nature of the interaction, but it's been largely taken over by structural biologists."

The other Src-related domain, SH3, was discovered in the early '90s and was the first example of an entirely new class of signaling interactions. SH3 binds to stretches of protein that contain large amounts of the amino acid proline. The notion of using protein-protein interactions as a signaling mechanism had been bruited about for years, but no examples of protein domains designed expressly for that purpose had been found. Again, the discovery of SH3 opened up new avenues of research, as SH3 domains have since been found far and wide.

Collectively, these three domains are involved in coordinating many aspects of cell division, cell differentiation, and cellular activation—just about everything a cell would "want" to do. Says Baltimore, "We're still trying to put these signaling elements together to understand their integration, as well as what pathways they're involved in. It's a long, complicated business." And there are vast tracts of the *abl* gene still to be explored.

With all this going on, it's amazing that Baltimore gets out at all. Yet he helped develop national guidelines for genetic research back in the 1970s, and he has been a prominent figure in the public debate over genetic engineering ever since. His work with retroviruses and reverse transcriptase led to his being invited to help plan the research assault on AIDS in the late '80s. In the early '90s, he was one of the architects of the federal Human Genome Project, which is now working to discover all of the 50,000-plus genes in human DNA. Most recently, in 1996, when the National Institutes of Health created the AIDS Vaccine Research Committee to expedite the search for a vaccine, Baltimore was tapped to lead it—a post he will retain as president of Caltech.

Baltimore sees continuing to be a public figure as part and parcel of charting Caltech's course into the 21st century. At the press conference that announced his selection, he noted that the pace of scientific advance, in fields ranging from cloning to computer science, is allowing us to do things that were impossible only a few years ago. "I look forward to working with the Caltech faculty in advising our society as it adjusts to these changing capabilities. I will also work with the leaders of government, industry, and academia to help prepare society to deal with the profound implications of modern science. The role of national leadership is an important one, and Caltech has a responsibility to play a role in the debates, as they occur, about the development of modern science and engineering." □