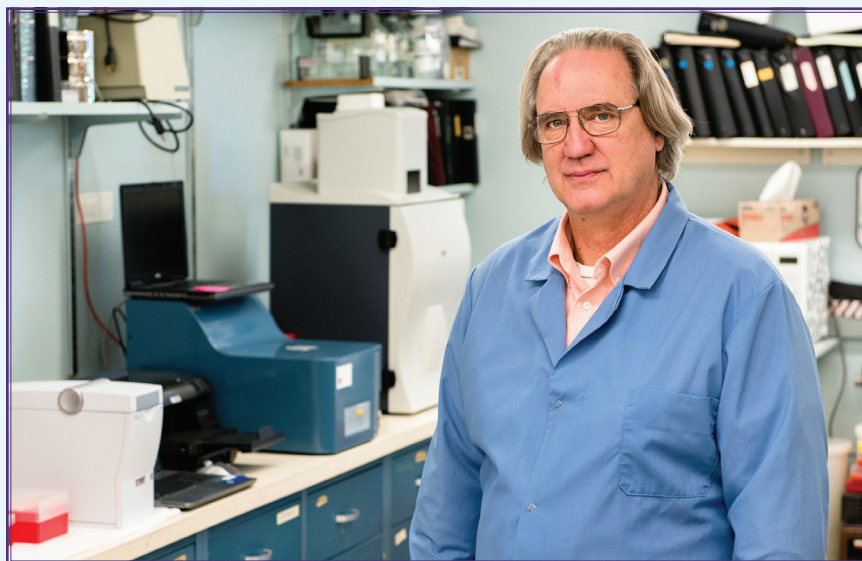


# Viral Activity

*In the last four decades, HIV has gone from being an unknown killer to the cause of a manageable chronic disease. Stephen Hughes, Ph.D., Chief of CCR's Retroviral Replication Laboratory, began his study of retroviruses before HIV was identified, but quickly made the virus the main focus of his research career. Hughes is internationally recognized for his work on two of the three essential enzymes in the HIV life cycle: reverse transcriptase (RT) and integrase (IN). His work has shed light on the emergence of drug resistance and, more recently, the nature of reservoirs of HIV that persist despite combination antiretroviral therapy. He has also used engineered host proteins that redirect HIV integration as tools for understanding eukaryotic chromatin organization.*

If asked, Stephen Hughes will tell you that the retrovirus HIV is a fascinating creature, marvelous in its complexity. "It's only 10 kilobases. You could memorize the sequence of its nucleic acids; you could have it built for you. But knowing all that it does to survive is still far beyond us," said Hughes.

Hughes committed to studying retroviruses after completing his graduate training in the laboratory of Mario Capecchi, Ph.D., who later won a Nobel Prize. He viewed retroviruses as primarily a tool for understanding how genes worked in higher eukaryotes. "It seemed to me, at the time, that retroviruses were probably masquerading as genes in their integrated state," said Hughes. He arrived as a Postdoctoral Fellow to work with the future Nobel Laureates, Harold Varmus, M.D., and Michael Bishop, M.D., at the University of California, San Francisco (UCSF), at what he describes as a magical moment. "I showed up in 1976, and, when I left three years later, the fundamental questions about how the RNA was organized and proteins were made had been answered."



(Photo: R. Frederickson)

Stephen Hughes, Ph.D.

During his years in San Francisco, men in the Castro district where he and his wife lived were just beginning to die of what was then termed GRID for "gay-related immune deficiency." By the time HIV was identified as the probable cause of AIDS, Hughes had recently arrived at the Advanced Bioscience Laboratories—Basic Research Program at NCI at Frederick, under the direction of George Vande Woude, Ph.D., who nudged Hughes in the direction of HIV.

## Reverse Transcriptase as a Drug Target

Hughes was interested in studying retroviral enzymes. Two key steps distinguish the retroviral life cycle: 1) the genome is RNA that is converted, in infected cells, into DNA through the actions of RT, and 2) the DNA is permanently embedded in the host genome through the actions of IN.

Working with a visiting scholar from Israel, Amnon Hizi, Ph.D., Hughes succeeded in using recombinant DNA in *Escherichia coli*

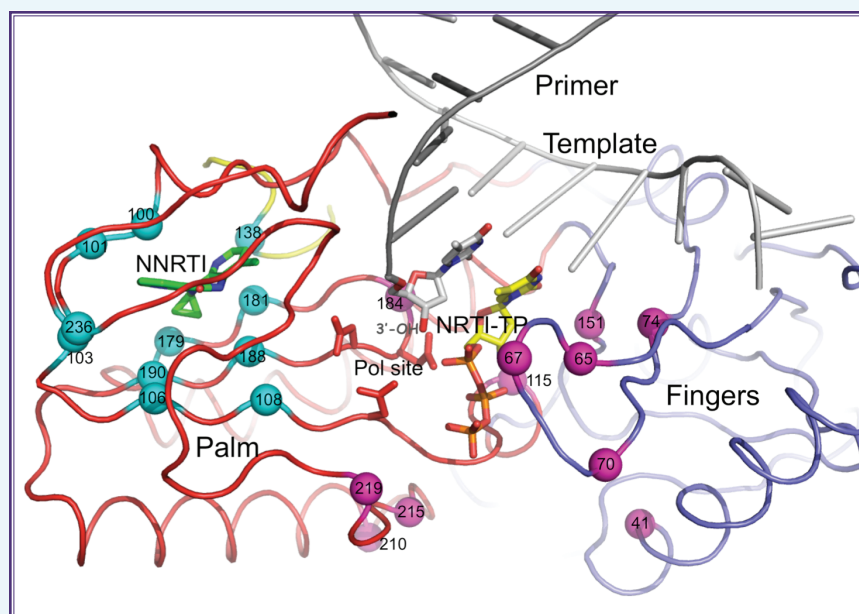
to produce RT from the murine leukemia virus (MLV) in useful quantities. “George Vande Woude came to talk with us because we were wildly excited about the amount of MLV RT we had purified,” said Hughes. “George said, ‘This is really good. I don’t mean to throw cold water on your efforts, but you should probably do this for HIV.’”

HIV RT was more challenging to express and purify, but Hizi, Hughes, and their colleagues overcame the obstacles. As in the case of MLV, however, RT was much more tractable than the two other key HIV enzymes. “Protease was toxic to *E. coli* and integrase had unfortunate physical properties, but we had an active, soluble RT,” said Hughes. Meanwhile, the nucleoside analog AZT, acting on RT, was found to be the first highly effective anti-HIV drug.

“It was obvious to retrovirologists that as soon as you began to treat HIV with drugs, you would get resistance,” said Hughes. “So we thought having large quantities of purified HIV RT would give us a tool to study resistance biochemically and, with some luck, structurally.”

It took some effort to persuade structural biologists to share this view. “When we began making milligram quantities of RT, I literally couldn’t give it away to prominent crystallographers. They all had their own proteins, which they thought were more interesting,” said Hughes.

Fortunately, he met Eddy Arnold, Ph.D., who was excited by the challenge of crystallizing RT. The Hughes and Arnold laboratories worked together for about four years, until eventually they were able to form good crystals of HIV RT that could be used for structural analysis. RT is a physically flexible protein, which resists the orderly stacking that is so important for X-ray crystallography. “We used some tricks to help stabilize the protein,” said Hughes. “We made



(Figure: K. Das and E. Arnold, Rutgers)

The structure of HIV-1 reverse transcriptase (RT). Image shows a close up of the region around the polymerase active site where mutations can cause resistance to anti-AIDs drugs. The RT backbone is shown as a wire diagram, and the p66 fingers subdomain is shown in blue and the palm is in red. In this image, the dsDNA is shown as two wires with branches to represent the bases. The incoming dNTP is shown as a wire frame model. Positions in RT where mutations give rise to resistance to nonnucleoside inhibitors (NNRTIs) are shown in light blue, sites where mutations give rise to resistance to nucleoside analogs (NRTIs) are shown in purple.

a family of monoclonal antibodies, and Eddy and his colleagues cocrystallized RT with an antibody fragment and a nucleic acid substrate to improve the structure.”

Arnold and Hughes worked together for more than 25 years on understanding the structure and function of HIV-1 RT, how drugs inhibit the enzyme, and how resistance mutations overcome the actions of different drugs. Arnold’s lab has analyzed the structure of the wild-type and mutant RT proteins, and Hughes’ lab has done the biochemistry and virology of the same mutants.

Some months after their collaborative efforts began, Hughes was surprised to see the tide turning, as other crystallographers began to reach out to him to obtain HIV RT for structural studies. It transpired that Marvin Cassman, Ph.D., National Institute of General Medical Sciences, started a new program to support structural work on HIV proteins, through which Hughes and Arnold were able to

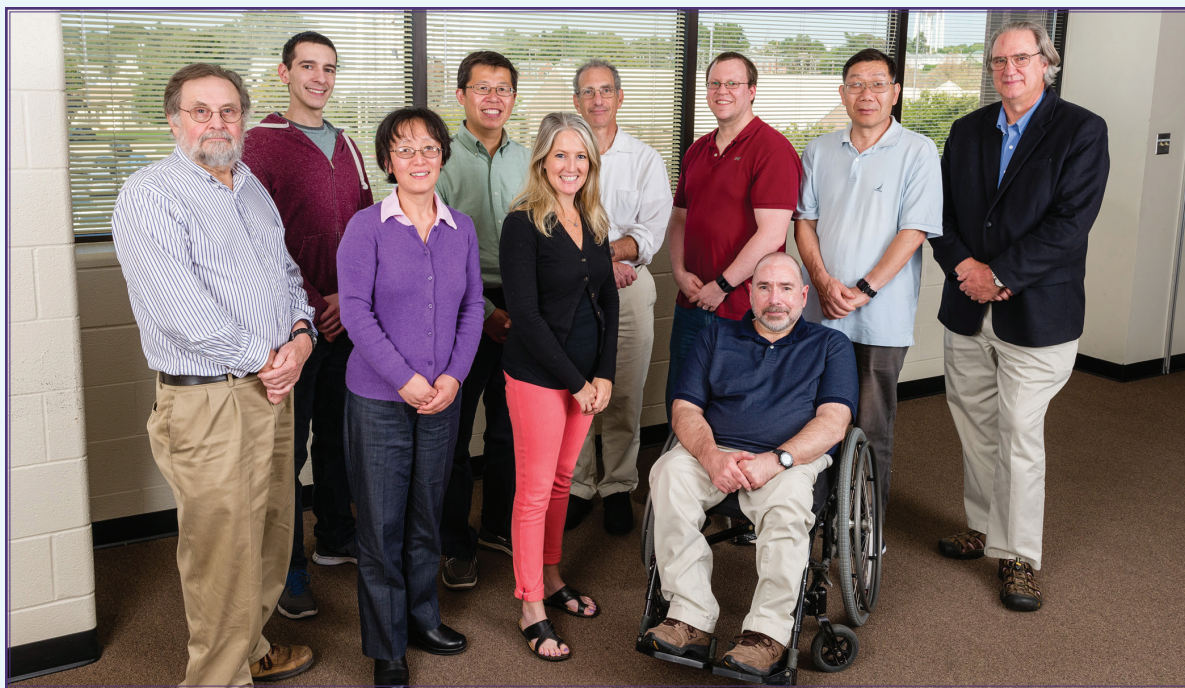
continue their ongoing research. “Marvin had the deep insight that understanding the structure of HIV proteins would be important. Several important protein structures came from this initiative. It was one of those instances where a single person changed how things were done in the field,” said Hughes.

## Integration as a Tool

“I have always had a soft spot for integration,” said Hughes.

During his postdoctoral work, Hughes solved the structure of the provirus, the viral DNA that is integrated into the host genome, but when he established his own laboratory, most of the work was focused on other problems. When the work in his laboratory shifted to HIV, technical hurdles prevented him from working on the enzyme central to integration, IN. “We did play with it a couple of times, trying to do experiments in parallel with our work on RT. It was just intractable. We set IN aside for a long time,” said Hughes.





(Photo: R. Frederickson)

The HIV integration site research team. Front row (left to right): John Coffin, Ph.D., Ling Su, M.S., Mary Kearney, Ph.D., and Shawn Hill, M.S. Back row (left to right): David Wells, M.S., Xiaolin Wu, Ph.D., Frank Maldarelli, M.D., Ph.D., Jonathan Spindler, B.S., Wei Shao, Ph.D., and Stephen Hughes, Ph.D. Not shown: John Mellors, M.D., Francesco Simonetti, M.D., and Andrea Ferris, M.S.

The HIV provirus integrates into host DNA by forming a poorly defined preintegration complex (PIC), which interacts with a chromatin-associated protein, lens epithelium-derived growth factor (LEDGF). LEDGF is a bipartite protein; one end has two sequences that interact with histone modifications and DNA, respectively, and the other end interacts with IN in the PIC. LEDGF preferentially directs HIV integration to the sequences of highly expressed genes.

“Eric Poeschla (then at the Mayo Clinic, now at the University of Colorado, Denver) did an experiment which just floored me,” said Hughes. “He showed that if he took away the nucleic acid and histone binding component of LEDGF and replaced it with something else that would also bind chromatin, the resulting fusion protein still enabled efficient HIV integration.”

Poeschla’s experiment immediately suggested to Hughes that not only would the fusion protein preserve integration efficiency, but it could also direct that integration to

different genomic sites depending on the specificity of the engineered chromatin-binding component. This integration could be important not only for gene therapy applications, where integration into the wrong piece of DNA can have disastrous effects, but also for chromatin mapping. In 2010, Hughes and his colleagues published, in *PNAS*, the finding that substituting two different chromatin-binding domains (CBDs)—the plant homeodomain finger from inhibitor of growth protein 2 (ING2) and the chromodomain of heterochromatin protein 1- $\alpha$  (HP1 $\alpha$ )—directed HIV to different integration sites according to their known specificities.

Thus, determining the sites of HIV integration could be used as a tool to map where the fusion protein binds to chromatin. Hughes collaborated with Xiaolin Wu, Ph.D., Senior Scientist at Leidos Biomedical Research, Inc., to develop the technique, which they called HIV integration targeting (HIT-seq). In 2013, in collaboration with Robert Roeder (Rockefeller), they published a paper in *Cell*, in

which they used HIT-seq to describe the effects of a common histone modification on p53-dependent transcription of active genes.

“In order to get HIT-seq to work, we had to be reasonably efficient at recovering the integration sites. It was a considerable amount of work, but we got good results using Illumina deep sequencing,” said Hughes. “So we wondered if we could take this ability back to our HIV research and study where HIV integrates in patients.”

“Thus,  
determining  
the sites of HIV  
integration  
could be used  
as a tool to map  
where the fusion  
protein binds to  
chromatin.”

## Integration and Disease

"Why can't we cure a patient with HIV?" asked Hughes. "If you can completely suppress viral replication in patients with combination antiretroviral therapy (cART) for eight to ten years, why don't all the virally infected cells die?" Many have suspected that long-lived memory T cells are a reservoir, but data from the study of HIV integration sites in patient cells, published last year in *Science*, suggests a more disturbing possibility to Hughes and his colleagues.

HIV DNA integration can occur at millions of different sites in the host DNA. Thus, if two cells have identical HIV integration sites, they were probably derived from the same originally infected cell. Hughes and his colleagues sequenced the HIV DNA integration sites in peripheral blood mononuclear cells (PBMCs) or CD4+ T cells from the blood of five patients treated with long-term cART. Of the 2,410 integration sites they identified, approximately 40 percent were found multiple times, showing that these sites came from cells that had clonally expanded after infection. In one striking example, more than 50 percent of the infected cells in a patient were from a single clone. Moreover, some of the clones of HIV-infected cells were shown to persist in patients for more than a decade.

More recently, Hughes and his colleagues have gone on to show that the virus from the expanded clone is produced at low levels in the patient and is capable of replication. "People had assumed that cells were infected and went to sleep, but suppose that's not true? Suppose there is a population of clonally expanding cells, but they do not all

behave identically, and only a small fraction are actively making virus at any one time?"

Perhaps more surprising than the presence of clones was that the data from these patients also showed there was selection for cells with integration sites in specific portions of two of the genes, *MLK2* and *BACH2*, where there were, respectively, 16 and 17 independent integrations. The sites and orientations of the integrations in *MLK2* and *BACH2* suggested that these integrations altered the expression or the protein products produced by these two genes. Meanwhile, in control experiments performed by infecting cultured cells with HIV, there was no preferential integration in one orientation in either *MLK2* or *BACH2*, nor was there preferential integration in the target regions of these genes. Thus, cells with integration sites in these two genes appeared to have gained a selective growth and survival advantage.

"Most of us were reasonably convinced we would find clones of infected cells," said Hughes. "But we weren't prepared for the fact that HIV integration could drive clonal proliferation of the cells. We are quite confident that in the case of *BACH2* and *MLK2*, integration of the provirus is a major contributor. It remains to be seen what fraction of the other integration sites are driving proliferation."

Much more work is needed to establish the importance of these cells to the course of the disease. And, if one believes they are important, the questions turn to when these cells start to expand and where they persist.

Meanwhile, Hughes is also

turning his attention to the implications of the integration work on cancer. In mice, MLV integration into the *BACH2* gene is known to cause tumors. In people, cancer is usually a multistep process that may not have had the time to develop in untreated HIV patients. However, in the last 10–15 years, better anti-HIV therapies have allowed patients to start to achieve relatively normal life spans. The higher rate of cancer in HIV-infected patients is usually attributed to the failure of the immune system to control herpes viruses. The question is whether all cancers will be attributable to that cause, and, if not, what role (if any) HIV integration sites might be playing.

Despite these looming questions, Hughes sees the progress that the field of HIV research has made over the last 30 years as a testament to human ingenuity and a matter of fortunate, if imperfect, timing.

"I think there is no question that HIV jumped from chimps to humans in West Africa around 100 years ago," concluded Hughes. "Imagine if that had happened 100 or 150 years earlier. We would have been intellectually and medically completely unprepared. As bad as it is now, it would have been much more severe. Conversely, I think if it had appeared 100 years from now, it would not have been a difficult problem to resolve. If you wanted to imagine a problem that was just beyond our intellectual grasp, and one that would make us work as hard as we could and reach as far as we might, with important consequences for millions of people, the rise of HIV is it."

"HIV DNA integration can occur at millions of different sites in the host DNA."

To learn more about Dr. Hughes' research, please visit his CCR website at <http://go.usa.gov/3JG5W>.