David Baltimore, Part 1: Introduction to Viruses: HIV and Non-equilibrium Viruses (34:06)

Hello, I am David Baltimore, President emeritus of the California Institute of Technology, and now Professor of Biology at Cal Tech. I'm going to talk today about HIV, particularly in the last two lectures, in the context of what we know about viruses in general, where HIV came from, and then go into therapeutic / vaccine approaches to HIV that are novel, and that had been forced on us by the fact that HIV has been so difficult to deal with by standard methodologies.

In the first of the talks, I am going to talk about a concept of equilibrium and non-equilibrium viruses, and in particular, describe how HIV can be considered as a non-equilibrium virus like so many other viruses, that are dangerous to us. But first, let's introduce viruses.

Viruses represent a separate kingdom of the living world. Viruses are not like bacteria, fungi or other organism that can exist and grow as independent agents in the environment. Viruses are very, very small, and so all they really have is the genetic material to drive their duplication. But to actually carry out their duplication, they have to go inside of a cell. Penetrating living cells becomes very important. So a virus is then able to divert the cell's macro-molecular synthesis toward ends which are dictated by the virus.

Viruses have one of two kinds of genetic information: either RNA or DNA. This differentiates them from everything else in the living kingdom, that all use DNA as their genetic material – to send information from one generation to the next. Viruses can use RNA. RNA and DNA are very similar to each other chemically, but are very distinct in how they are functioning in biological systems. It is interesting, and may be even a remnant of an earlier world, that RNA can be a genetic material.

Viruses can grow inside cells of all sorts – animal, plant and even bacterial cells – and bacteriophages were the first viruses that were studied by molecular biology. If we want to understand viruses, we have got to understand them in the context of the central dogma of molecular biology. That central dogma, as first elucidated by Francis Crick, says that DNA can duplicate (make an exact copy of) itself, so that it can go from cell to cell in an exact copy. The DNA in any given cell can be transcribed into RNA, and that the RNA constellation of a given cell is very important in terms of the specificity of that cell. A skin cell has a different constellation of RNA than a liver cell, or any other. And finally, RNA that encodes protein; proteins are the workhorses of viruses, as they are of cells.

In 1970, I had the good fortune to make an observation that added a piece to this central dogma, and that is reverse transcription. Reverse transcription is the ability to reverse the ordinary flow of information, so that RNA gives rise to DNA, and that's something that was thought to be particular to viruses; as we will discuss, it has much wider implications than that.

Since proteins are the workhorse of any biologic system, the messenger RNAs that encode proteins are the most important end point of a molecular system that allows for controlling cell behaviour. We can classify viruses by how they make their messenger RNA. In particular, there are DNA viruses and RNA viruses.

There are two kinds of DNA viruses: one kind has double stranded DNA – the standard kind of DNA that we find in the nucleus of all of our own cells, and the cells of other organisms in our environment. For that double-stranded DNA to make messenger RNA, the process is pretty straight forward; there's a copying mechanism called DNA-dependent RNA polymerase that copies DNA into RNA. Viruses either hijack the cell's DNA-dependent RNA polymerase, or they bring in their own.

Then there are some viruses with only one strand of DNA – and those viruses have to duplicate that strand to make a double strand, which can then be transcribed into message. These viruses have a somewhat more complicated life cycle.

Then we have four kinds of RNA viruses.

- RNA viruses that have double stranded RNA as their genome. They have a polymerase that can copy double stranded RNA into single stranded messenger RNA.
- Viruses that have only the 'plus' (sense) strand of RNA in their genome, such as the common cold virus. They get copied into the minus strand, which then becomes a template for more 'plus' mRNA.
- Viruses that have 'negative' strands of RNA as their genome. They get copied into 'plus' mRNA in a direct fashion, but because they have to be duplicated, so they have to copy back the 'plus' strand into the minus strand RNA.
- Retroviruses – that have RNA as their genetic material, but copy that into DNA (first into one strand, then into two strands) – and finally that can be transcribed into mRNA.

We can take all the viruses in the world and put them in these six categories, and that's a convenient way of thinking about them, at least for a molecular biologist.

Viruses turn up in the news all the time. HIV, which we will talk about a lot, of course has been in the news ever since its discovery in the early 1980s. On the other hand, SARS appeared once in our environment as a potentially

lethal virus. SARS came out of the animal world to cause a relatively small number of deaths before we got it under control. Smallpox is a virus that used to be rampant in our environment. We have actually gotten rid of it, and as far as we know, it only exists in two places right now – in freezers in Russia and the United States. Influenza is a virus that is continually in the news, because it keeps changing itself, comes back every year in a different form, and now we are particularly worried about whether the avian flu will become a human flu

There are some very famous incidences in which viruses just appeared out of the wild. Marburg virus appeared in Germany; it turns out that it had come from Africa, killing a significant fraction of the people it infected. Then about ten years later a extremely lethal form of the Ebola virus appeared (although there are other benign forms of Ebola). Hantaan virus appeared in the four corners of the United States in the dessert, as a virus coming out of rodents, also extremely lethal to the relatively small number of people infected. And Norwalk virus is a very different sort of agent. It turns up periodically on the news when there is a cruise ship on which people get infected with a diarrhoeal virus – that is norovirus, or Norwalk agent.

You need to understand molecular biology to understand viruses because they really are just molecular agents. But viruses were also very important in the history of molecular biology, because many of the key experiments were done using viruses. Hershey and Chase showed that DNA is the hereditary material; they were actually re-discovering something that had been found with bacteria, but it was their experiments that convinced the scientific world that DNA was the hereditary material. Luria and Delbruck showed that mutations pre-exist before evolutionary selection; Meselson and Stahl showed DNA replicated by copying each strand into a duplex; Hershey and Benzer showed that genes had fine structure, and Brenner and company finally showed that messenger RNA existed at all in experiments with viruses done at Cal Tech.

Molecular biology of animal cells has also been helped by understanding the behaviour of viruses: Splicing of nuclear RNA to generate messenger RNA was discovered in viruses; we discovered reverse transcription, as I have just described. Plant viruses were very important when RNA was first shown to be the genetic material in plant viruses, and it led to the idea (which is very current and very exciting) that perhaps the whole world we see today was once an RNA based world, that DNA came later. And finally, protection against pathogens by interfering RNA was first shown with plant viruses.

How many viruses are there? There are literally an uncountable number of them, and that's because every plant, every bacterium and every animal has its own set of viruses, and we are never going to find them all. We now recognise something like 1,500 plus species of viruses, and each species has multiple kinds, so the numbers are already in the thousands, and can clearly be more than that.

How do viruses grow? We said that the basic thing is that they grow inside of cells. They can multiply very fast; once a virus gets inside a cell, it may increase itself by a thousand fold in a few hours. In fact, bacterial viruses can increase themselves by a hundred fold in 20 minutes.

Because viruses have to grow inside of cells, they have to be passed from one host, to another host, to another host, to another host, continually; from me to you to somebody else, to somebody else. If that chain is broken, the virus has lost its ability to function in the environment. Viruses specialise themselves to take advantage of the sociology of a species. We get together in groups, and in particular our children get together in very intimate play games. Viruses have taken advantage of that, to spread among children, from children to their parents, and of course among adults when we sneeze or cough or shake hands… something we should think about changing.

If a virus is only a human virus, and we vaccinate everybody so that nobody is sensitive to the virus for just a little while (a year maybe), then the virus should be dead – it should have no place to hide. And in fact that's true. And that's why we eradicated smallpox; we virtually eradicated polio, but there are some places where it is still spreading.

How do viruses get out of cells? They develop inside cells and get out one of two ways. The virus can grow inside the cell (imagine this is the cytoplasm of the cell and nucleus of the cell) And the virus can grow in here and can burst the cell open – then it can release itself. Some viruses do that – bacterial viruses in particular. They grow inside the host cells, then cause the host cell to break, releasing the daughter viral particles. But there is a way that viruses have to get out of the host cell that does not involve killing the cell – and that is to bud off the surface of the cell. The core of the virus is inside the cell, the virus is pushing its way out of the cell, taking with it modified cell membrane that has virus-specific proteins out to the environment, ultimately releasing itself as an infectious agent.

As I have said before, viruses only make sense in the context of molecular biology. When there was no molecular biology, that is, before WWII effectively, it was impossible to understand the nature of viruses. Viruses were just something small (in fact the definition of a virus was that it would pass through a filter that held back bacteria). Viruses are very small, but their genetic information is very powerful and it can make itself a coat which can be impervious to environmental stress. Some viruses have a little more than a protective coat to them; they hold inside themselves key elements that enable a virus to grow, and we'll come back to that. But we couldn't understand any of this unless we had molecular biology.

I said viruses are very small… the question you might have is how do you see a virus? Well, you can use an electron microscope. Electron microscopes have a resolution that allows you to get down to nanometres, the size of viruses. Small viruses are about 20 nm in diameter, and larger viruses can be many times that. Biologists developed a trick to visualise individual viruses, which is how virus work was done for many years. And that is, to grow a monolayer of cells (animal or bacterial – whatever was appropriate), take a suspension of virus – say a million particles per millilitre, dilute it down to a few particles, and then put the particles on top of the monolayer. Those particles will find cells to grown in, the cells will start to release virus, infect cells around them, and ultimately, you get a plaque. You can count the number of infectious particles by the number of plaques, and if you have diluted the virus you can figure out how many virus particles you had in the original suspension, by simply multiplying back.

These are plaques of bacterial viruses growing on a bacterial lawn; these are plaques of polio viruses, growing on a lawn of human cancer cells.

I said I was going to talk about equilibrium and non-equilibrium viruses. Let me make that distinction here. Equilibrium viruses are viruses that have been with a species for a very long time. They have generally developed a *modus vivendi* with that species. They're generally not lethal or very lethal, but they spread extremely well from animal to animal, person to person, bacterium to bacterium. A very good example is the common cold virus: we all know that if we come in contact with a person who has a cold, we are likely to get a cold. Somehow that virus is going to transfer, either through a sneeze, or through physical contact. These are equilibrium viruses in the sense that they have developed an equilibrium status with their host. They're not asking too much of their host, they're not going to kill their host generally; on the other hand, the host provides a way of transmitting the virus on into generations into the future.

Non-equilibrium viruses have jumped from another species where it is an equilibrium virus, into a new host organism, to which it is not well adapted. Often these viruses can be very lethal, because they haven't been selected to be temperate; they may spread poorly but sometimes they spread well; and they represent most of the difficult problems that we have. Most of the viruses mentioned in the slide that I showed you are non-equilibrium viruses, like Marburg, Ebola and HIV. HIV has developed the ability to spread well, whereas something like Ebola virus spreads poorly, and in fact tends to make its infected people so sick so fast, that they can only transmit to very close family members, or to healthcare providers.

So, let's look at some equilibrium and non-equilibrium viruses.

Equilibrium viruses:

- Polio – almost eradicated;
- Smallpox – eradicated;
- Common cold – there are too many kinds to eradicate it, but it's a standard equilibrium virus.
- Measles, mumps, herpes – lots of them.

All very specificly human viruses.

Non-equilibrium viruses:

- Flu (birds)
- HIV (chimp)
- SARS (bats?)
- Ebola (bats)
- Hantaan (rodent).

Let me turn now to HIV. You will remember that the HIV epidemic started, or became evident to us in the 1970s. It became evident when people came down with a disease called AIDS – acute immune deficiency syndrome. What happened was we first saw people who had this very severe immunodeficiency and were dying of organisms that ordinarily should not harm them - organisms that are around us all the time, and are not that much of a problem to us. So, their immune systems somehow had totally failed, and could not handle the very simplest of challenges. But we didn't know what caused it. It took a couple of years to figure that out. You may remember that during that period of time, people thought of all sorts of ideas: they thought sperm had caused it, they thought all sorts of things – because it's sexually transmitted. But it turned out to be a virus, a virus which is sexually transmitted. That virus was then named Human Immunodeficiency Virus, or HIV.

How did we find HIV? Actually, only when reverse transcription was discovered would HIV have made any sense. We discovered reverse transcription around 1970, and HIV came around 1980, so we were lucky. We knew about reverse transcription, we knew how to find HIV by assaying for reverse transcriptase. But imagine the reverse: imagine HIV came in 1970, and we hadn't yet discovered reverse transcription. In that case, it would have been very hard to find HIV, because there just isn't very much of it around: it's hard to see it in the electron microscope, and it's hard to do anything but measure the enzymatic activity of reverse transcriptase. So we were lucky in the order in which that came.

Let me go back to that discovery of reverse transcriptase, and put it in the context of my particular life in science. In 1960, I entered the field of molecular biology, just out of college and going to graduate school. We knew about two kinds of polymerases – enzymes that can catalyse the synthesis of nucleic acids – RNA and DNA. One of them was the DNA polymerase, which we believed to be involved in DNA replication, and the other one was the DNA-dependent RNA polymerase, which was involved in gene transcription.

Then in 1962, as I began to work on polio virus, I realised that it may have an RNA-dependent RNA polymerase, and assayed for that, and found that, in polio virus infected cells, there was in fact such an enzyme. Now we know that all RNA viruses encode enzymes of that sort.

Then in 1969, we realised that there was a class of negative strand viruses – viruses that have the minus strand. We looked for polymerase for that, and in fact there was one. That was found in the virus particle, not in the infected cell so much. So that said that maybe virus particles have interesting enzymes.

Let's go back in time, 1960s: Howard Temin, just having gotten his degree from Cal Tech and going to the University of Wisconsin, speculated that a certain kind of tumour-causing cancer virus which carried RNA as its genetic material could transfer that material into DNA. He suggested that cancer-causing viruses cause permanent changes in cells. They change cells from normal growing cells to cancerous cells that grow without control. And so, they are doing something permanent to cells, but RNA generally in cells has a transient role. Messenger RNA is degraded; other types of RNA are turned over; they very rarely have a stable "information-carrying" role. He then said that maybe RNA can be turned into DNA. For ten years, he had worked on trying to prove that, then in 1970, Howard Temin and I independently discovered that, in the virus particles of RNA tumour viruses, there was a RNA-dependent DNA polymerase, which was then dubbed the "reverse transcriptase", because it reverses the standard flow of information in biological systems.

Because that discovery had lots of implications, in 1975 we shared the Nobel Prize in Physiology or Medicine for that discovery with Renato Dulbecco, who had been a mentor to both of us, and who had also done very important work on another class of viruses.

Here's reverse transcriptase. Over here is a ribbon molecule that you can't see much. Over here is a schematic model where you can see a lot. What you can see is that the RNA template comes in the top of the enzyme. Inside is the polymerization machinery that makes a copy into DNA – the red molecule – that then continues further down within the reverse transcriptase to the ribonuclease H component, which recognises that the material is a hybrid of RNA and DNA, and degrades the RNA portion of the hybrid. So, the genome that came in here, is actually degraded right away once it has been copied, and out the other end comes a pure DNA strand that then can be duplicated again by reverse transcriptase.

That was the discovery. What have been the implications since that discovery in 1970? Well, first of all, we changed the name of RNA tumour viruses. They became retroviruses, indicating that they were reversing the flow of information. Secondly, and much more importantly, it was recognised that cancer can be caused by the permanent association of new genes with a cell. That model was correct and the implication was that cancer was due to mutation of genes, which has been proven to be true in many forms of human cancers in humans and other animals. A third implication was, since we are copying RNA into DNA, maybe you could copy messenger RNAs into DNA. We very quickly worked out a way to do that; and that was one of the starts of biotechnologies, because in a sense what you are doing is capturing in a genetic form (in double stranded DNA) the information in a messenger RNA, so you've found a way to make a protein in very large amounts using artificial DNA constructs that were copied from messenger RNA. And that's what started biotechnology.

Something we (I at least) were not prepared for was the discovery that many mobile DNA, genetic elements – pieces / things that hop around in the genome – first discovered by Barbara McKlintock in plants – that many of these use reverse transcription. They copy a piece of DNA into RNA, the RNA is then copied back into DNA, the DNA hops to another place in the genome. That happens in the history of all species, and in humans it's happened so frequently, that some 45% of our genome arose in this manner. So, 45% of our genome came from reverse transcription.

Finally, relevant to this discussion, is that HIV was found to be a retrovirus, something that was discovered only ten years after the discovery of the polymerase.

So now we can put together a life cycle for HIV. HIV, as a free agent, has reverse transcriptase inside of it (it also has some other things inside of it that we'll come back to); on the outside it has little spikes, which recognise the surface of certain kinds of cells (we'll come back to which cells in a minute), and fuse with the surface of that cell, putting the inner components of the virus into the cell. There the reverse transcriptase gets activated, and it now makes a double stranded DNA copy of the RNA genetic information, that then, (because there is an integrase protein in here also that integrates with the host chromosome) goes into the nucleus and encodes messenger RNA (because it is a DNA that can be read by RNA polymerase), makes messenger RNA and that then makes new genomic RNA. The messenger RNA encodes viral proteins, the cell gets viral proteins on its surface as well as inside it, you get budding out of the sort I mentioned before and you make new viral particles.

That budding process is indicated in this slide of measles. You can see very nicely how it buds off the surface, carrying with it a coat of protein going through the membrane, genetic material inside it, and ultimately it will pinch off here, and release a free virus particle.

Now, with HIV, the virus particle has a very strange property, and that is what we call the "nucleoid" inside the particle. It is not symmetric. In almost all of the viruses it is spherically symmetric. Here it is an asymmetric structure for reasons that are not clear but represent a signature of HIV.

This is a nice diagram of the HIV particle: you can see the spikes on the outside, this is what attaches to cells; you can see the genome, which is actually two molecules of RNA; you can see the reverse transcriptase in here; you can see the integrase which is going to help the double strand DNA to go into the chromosome; and you can see another enzyme, a protease, that's important for maturing the virus.

You can't talk about HIV without talking about how awful the AIDS epidemic has been. 65 million people have been infected at this point (2005 data), 25 million people dead and 40 million living with AIDS (5 million in India and almost one million in China, as far as we know – I call those out because these are the most populous countries in the world, and they are spreading the virus).

Today, there are about 13,000 infections per day, mostly still in Africa, 5 million a year -three plus million dying every year of the viral infection – as much as the two other infections that have been the scourge of the less developed world – tuberculosis and malaria.

The effect on African countries has been dramatic and awful, in some countries seeing a reduction in average life span of as much as 20 years or more.

Why is this so lethal? Fundamentally, it is a classic, non-equilibrium virus. HIV, or a very close relative of HIV, almost certainly its parental virus, was endemic in Africa in great apes, found in chimpanzees. It got a foothold in the human population some significant length of time ago, 70 years ago, probably was spread at very low levels, maybe very inefficiently and not very well recognised because it doesn't cause a characteristic disease, that looks like a lot of other things, and then came to the United States, probably on an aeroplane, in the 1970s and started killing people in numbers that were large enough that they were recognised by physicians in the late 1970s.

Why is HIV lethal? Well, it grows in one of the key cell types of our immune system. Our immune system has lots of different cells in it, but one of them is a helper cell, that helps everything else take place. If you don't have helper cells, the immune system is pretty puny. HIV targets itself on helper cells. Infected people lose their helper T cells, over years (because it takes that long for the virus to get through the cells), and ultimately they become immunodeficient, and they die from infections by organisms that healthy people would ordinarily just get rid of.

I have taken us from a very general consideration of viruses, to the very specific virus HIV – the virus that grows in helper T cells and causes lethality in humans – the virus that we HAVE to stop. And yet we have found it very difficult to stop its transmission. Why is that true? That will be the subject of the discussion in our next two parts.