Bruce Alberts speech in Shenyang, China; September 22, 2012
Biology Past and Biology Future: Where have we been and where are we going?

Good morning! It's a pleasure to be here in Shenyang. I have often visited China, but this is my first time in this beautiful city with a great, interesting history. I want to acknowledge the many Chinese-born American scientists who have traveled here for this meeting. They represent an important tie between our two countries. I have been in China now for 12 days, emphasizing the importance of future Chinese-American friendship for the peace and prosperity of the world. And to make that possible, connections through science, and friendships between scientists, are going to be very important. Both of our nations are blessed by having so many outstanding scientists in the United States who were born in China. Of course they contribute to the teaching to our universities, and to our science; but very uniquely, they form a connection, a tight connection between China and the United States.

I am going to talk about biology. And I am going to start with a little personal history because there are many students here today. Students tend to think that successful scientists never make mistakes. Well, that's not true, as you will see. I will also emphasize what I learned from my research, which is the central role in all living things of the protein complexes that we call "protein machines." But there is an enormous amount of information that we still don't know. The complexity of life is such that we're going to need new methods and many new scientists to really come to an understanding of how cells and organisms work. In particular, I'll talk about the so-called "emergent properties" in biology. Finally, I'll address the important issue of stimulating more innovation in science and how *Science* magazine is trying to help.

First a little personal history. I was a high school student in 1953 when this picture was made of the famous discovery by Watson and Crick of the structure of DNA that solved the mystery of heredity. Watson was only 25 years old. I use this fact to emphasize the importance of encouraging young scientists who have new ideas and giving them a chance to be independent scientists.

Watson and Crick solved the mystery of heredity, but they did not know anything about the machinery that enables a DNA sequence to be copied from one generation to the next. The next major breakthrough came a few years later in 1957 when a biochemist named Arthur Kornberg discovered the enzyme DNA polymerase**,** which earned him a Nobel Prize. This famous enzyme will copy a DNA single strand sequence, one nucleotide at a time. But it will not copy DNA in its double helical conformation. Of course, inside the cell, DNA is a double helix. So this is what I tried to figure out during a very difficult and unsuccessful time as a graduate student at Harvard University. Inspired by Watson and Crick's theory for how heredity occurs, I made my own theory about how DNA polymerase might be able to start DNA synthesis on a double helix. And I did many experiments trying to see if my theory was right. Of course, it wasn't right. The many "no" answers did not surprise anyone

and they did not add to the store of scientific knowledge.

Nevertheless, my professor, after five years at Harvard, encouraged me write a thesis and then said that I was finished and should get my PhD. All I had left to do was walk into a room with five distinguished professors and talk to them for an hour and it would be done. To my knowledge, nobody had ever failed at this late stage, but to my great surprise, they failed me. That was the first of many failures I've had in my career, and all of them, if I look back, were learning experiences very important for making me a better scientist. Students can read about in a short essay I published, called " A wake-up call: How failing a PhD led to a strategy for a successful scientific career." (see *Nature* **431**, 1041; 2004); or watch my short iBioMagazine video talk at http://ibiomagazine.org/issues/august-issue/learning-from-failure.html.) Young people have to know that everybody makes mistakes, and those people who are successful learn from their mistakes. Older people can be enormously useful as mentors to younger people because they have made many mistakes and learned from them. We have thereby gained a certain amount of wisdom.

The first lesson from my failure as a graduate student is that theoretical biology is much more difficult than my generation had imagined. We were misled by Watson and Crick's great success in predicting how biology works. But most importantly, I learned that having a good strategy in scientific research is the key to success. There are so many experiments you could do, but most of them you should not do. Being a scientist is like being an artist, somebody who uses the resources that are available to do something important. And this requires designing a very thoughtful, strategic approach to what you are doing.

Following my failed thesis exam, I did more experiments, and six months later they finally awarded me a PhD. I went to Geneva, Switzerland as a post-doctoral fellow for one year. And there I suddenly discovered that DNA replication requires much more than DNA polymerase, something that I should have known all along. In Geneva, I met Richard Epstein who had carried out an extensive analysis of bacteriophage T4, publishing a beautiful genetic study of its more than 100 genes. He and his coworkers at Cal Tech had discovered and published in 1963 that at least 7 different T4 genes are needed to replicate the viral DNA.

Thus another lesson I learned the hard way is that one needs to read very widely in the scientific literature. In the middle of my PhD, this paper had been published from Cal Tech that proved that many different proteins are needed to replicate DNA in addition to DNA polymerase, but I had not read it because I had been reading only biochemical papers, not papers in genetics.

What all the other proteins did was a mystery that I would spend 20 years working on in my own laboratory. The geneticists had discovered genes that were given numbers, 32, 41, 43, 44 and so on, because they didn't know what these genes did. One, gene

43, had been shown to produce DNA polymerase; but there are at least 6 additional proteins needed for any replication of the T4 chromosome when that virus infects the E-coli bacterium. Therefore, DNA replication had to be much more complicated than anyone had imagined.

After my year in Geneva, I moved to Princeton University. I was 28 years old, and had my own small laboratory as an assistant professor, where students and I focused on purifying each of the seven bacteriophage T4 proteins identified by the mutants, using the genetics to help us. And our hope, and our dream, was to someday mix them together in a test tube with double-stranded DNA and get the DNA to replicate. The photo shows my lab at that time to show that science is fun. Which one is me? I am the one with the red shirt in the back, and the rest are students.

After many years, we discovered that the 7 purified proteins could indeed replicate DNA, when they were mixed together with double-strand DNA. And by analyzing what was happening in that purified mixture of DNA and 7 proteins, we came to recognize that DNA is replicated by a protein machine. There are two DNA polymerases at any one time at the so-called "replication fork," the place where DNA is being replicated. One, the DNA polymerase on the bottom in this diagram, is synthesizing the leading strand, and it goes continuously at hundreds of nucleotides per second, because it's moving in the same direction as the DNA that needs to be synthesized. The whole mechanism is complicated by the fact that, as pointed out by Watson and Crick, the two strands have opposite polarities: they are anti-parallel. The second DNA polymerase on the lagging strand, the one on the top, therefore has to go backwards. And to do this, it has to make DNA discontinuously as a series of short fragments. That means this that a little loop forms, breaks-up, and then reforms as the DNA fragments on the lagging strand are synthesized – a very complicated mechanism, but a beautiful one. (B. Alberts, Prokaryotic DNA replication mechanisms. *Phil. Trans. R. Soc. Lond. B*. **317**, 395-420, 1987).

So we now know that the same basic mechanism is used to replicate DNA in all organisms: from this virus, to bacteria, to yeasts, and to mammals like ourselves. However, as more complex organisms evolved, each function in T4 was carried out by more proteins, so a human cell has more than 25 different protein molecules replicating its DNA. So I was very lucky to start with a simple organism, working on a mechanism that was relatively easy to decipher.

We also know now that almost every process in the cell is carried out by a complex of 10 or more proteins. We call them "protein machines" because they work very much like the machines we all know. Of course, they are not driven by electric energy. They are driven by ordered movements, because you need ordered movements to make a machine. And those ordered movements require energy, being driven, generally, by the energy of ATP hydrolysis. And this is a wonderful, beautiful chemistry. And I am going to show you a movie about it – my favorite movie. This movie was made by

the Cold Spring Harbor Laboratory to celebrate the 50-year anniversary of the discovery of DNA.

The movie shows the DNA replication fork moving at the same speed as it does in your body. This is what a protein machine looks like – there are many different protein machines, and they work in different ways, but are similarly beautifully complicated. On the left the DNA double helix enters to be replicated; its two strands are being opened and separated by this blue turning, whirling machine. That's an ATP hydrolyzing enzyme called DNA helicase. On the right, moving continuously down to the bottom, is new DNA double helix being synthesized on the leading strand. Up at the top is the DNA polymerase molecule that's going backwards on the lagging strand. That's the loop forming and breaking-up. This is the essence of how life works. Of course, this is only DNA replication, but there are many other machines that work in terrific ways, beautiful ways; otherwise, we couldn't be talking to each other. The movie I showed is also on a CD ROM at the back of our textbook, the *Molecular Biology of the Cell*. (*5th edition.* B. Alberts, A. Johnson, J. Lewis, K. Roberts, and P. Walter. Garland Publishing, Inc., New York, 2007; see http://www.garlandscience.com/garlandscience_resources/resource_detail.jsf?landing=student&resource_id=9780815341291_CH06_QTM04).

So, some personal lessons learned:

First of all, if you want to do something important in biological research, work on a mystery. I worked on the mystery of why there were 7 proteins needed to replicate DNA. There are many more mysteries that we need to unscramble today—I'll talk about some of them. Work on a mystery because when you solve the problem, then you'll contribute in a significant way to the world's store of knowledge.

The second lesson is that there are remarkable homologies between living things; therefore use model organisms wherever possible, to try to work your way up to the understanding of humans. In the case I described, research in a set of different laboratories working with viruses and bacteria turned out to be a short-cut to understanding human DNA replication.

And finally, nearly all cell processes will be based on elegant mechanisms, too hard to predict. You have to do experiments – this is one of the hard lessons learned from my failed PhD thesis.

An important challenge for the students here, some of you, is obtaining the information needed to accurately describe the mechanism of every type of protein machine in a cell. There are hundreds of different protein machines, and each one does different things. This will require biochemistry. One needs to reconstitute each protein machine from its purified components, so that the detailed chemistry of each

machine can be deciphered in a test tube. And this means you also need chemistry.

We will also need to work out the many interactions between different protein machines. Of course, protein machines have to be able to get along with each other. An obvious case is the DNA replication fork machine that I showed you, which constantly collides with RNA polymerase molecules—again a big machine-like complex, transcribing RNA on the DNA. As cells evolved, they had to work out how to pass each other without creating problems. So, there are many wonderfully interesting problems to study.

I am now going to shift to focus on not what I learned from my own research, but what I learned from writing the textbook *Molecular Biology of the Cell*. I have been working on this textbook since 1978. The co-authors are in the midst of writing the 6th edition. The first black book took us more than 365 days to write, with all of us sitting together in a room at different times over the course of 3 or 4 years. This is a photo of the first set of authors. We were crossing Abbey Road, because the place where we work is right around the corner from the Beatles recording studio there. And the man at the back is Jim Watson with the hat. It was his idea to start this textbook, a very good idea. Of course at that point he was no longer 25 years old.

And this is what it looks like to write our textbook because every chapter is worked on by several different authors. In this case I was the author using the pen on top of the typescript and then someone else wrote over me. This is how every chapter gets written. Of course today we do it on computers. This was done before there was word processing.

We had a lot of fun with our last edition. Somebody had the idea to release it on a tour of South America. We were flown by a big Chilean Air Force plane to Antarctica, and here I am presenting the book as the official release of our 5th edition to the general who ran the Air Force base. He had set up a podium on a hill, and lucky for us it was summer in Antarctica, so it was not too cold. The general then got up at the podium and gave us a speech, and I think that he had no real idea why we were there or why he was getting a textbook in cell biology. But a young doctor at the base was very happy to have the book because he was going to be there through the long dark winter when there is nothing to do but to read books.

The authors have all learned a lot from writing our textbook. Teaching is important for doing good research; if you teach, you'll learn. And you'll think more creatively about your own research. And likewise, when you write a textbook, it makes you think quite broadly. Every time we write a new edition, we learn something new that we didn't realize before. In the last edition, we came to recognize that positive and negative feedback loops underlie nearly all chemistry. We also came to recognize the importance of so-called "scaffold proteins," which form networks that produce biochemical sub-compartments in the cell, without requiring a membrane. I am going

to talk a little bit about each of these.

When I was a student, we learned that A goes to B, goes to C, goes to D, goes to E. Of course, that's right, but in addition, there are all these feedforward and feedback loops that make this set of reactions, wherever it is, work in a cell, so that the processes can be controlled. And there is no way to understand such pathways, such networks, without mathematics. Those of you who are neuroscientists will recognize this very well. I remember many years ago there was an idea that if we knew every synapse and connection in the worm brain, the worm *C. elegans* – if we mapped all these connections – then we would understand something about how the worm thinks. Well, you know that was done in Cambridge in England. Once we had all that information, we still did not understand it, because it is just too complicated. And that's another example of the complex nature of life.

I am going to give a different example, the example that I actually worked on for a while before I left and become the full-time president of our Academy, and then had to close my laboratory. I am going to talk a bit about the actin filament network. Actin is a major protein in the cell that forms filaments that gives the cytoplasm some structure in all eukaryotic cells. It interacts with many other "actin-binding proteins" that control its activity, and this figure from our textbook shows an example of some of them. The one on the top left is a formin. It catalyzes the production of actin bundles and helps them grow in defined ways. Other proteins such as the ARP complex make actin networks; there are other proteins that break up or stabilize the filaments. Off the bottom of the slide is the famous protein myosin that causes actin filaments to move, using the energy of ATP hydrolysis.

This is just a small set. There are several hundred actin-binding proteins in human cells. The network I am showing here is a yeast cell, so it's not quite as complicated. But because of such complexity, even if we gain a complete knowledge of all of the molecules, protein machines, and molecular interactions in a cell, we will not be able to understand it. It's too complicated; it's like the worm brain I talked about.

Instead, life reflects the "emergent properties" that result from very complex networks of interactions. This is something we did not realize when I started as a student. We had no idea of it. For example, organized cell movement is one emergent property of the actin filament network. The cells are sending out projections at all times and moving around in their environment.

We will need new mathematical and computational models and tools to decipher such complex systems. I am going to give you just one example. We are very fortunate that we can use computers, because they enable us to handle huge amounts of data and complexity, and one of the things they have made possible are individual molecule based simulations — so-called "agent-based modeling". In this approach to understanding complexity, one allows tens of thousands of molecules to

simultaneously diffuse randomly and “react” in a virtual space that is in the computer. One can simulate the positions of all the molecules in cycle times of microseconds. After running the computer for a day or so, after billions of such cycles, you can see the result of having that set of molecules “react” in the computer – the so-called emergent properties of that system. If those computer properties resemble those in the cell, then one can hopefully use the modeling to go back and forth in association with actual wet chemistry experiments to gain important insights into biological mechanisms.

The reason why I am talking about this is that there is some work that is being done that I’m going to show you. This type of computational modeling cannot be done with very complicated systems because they require a large amount of quantitative data to make them worthwhile. You have to know all the interacting molecules, and you need to know their partners, and very importantly, you need to know the chemistry: the rate and equilibrium constants for forming each partner, because otherwise there are too many variables to guess at. You can do whatever you want in a computer, but you need constraints to make the computer models worthwhile. Of course, you also have to have an understanding of the actual behavior of the system, in reconstituted *in vitro* systems like the one I talked about for DNA replication, with pure proteins – and also of course know how it looks in a living cell. So for this reason, we need to start with model systems that are much simpler than the entire process.

A good model system for the actin filament network is the movement of the *Listeria* bacterium. This bacterium is very clever because when it enters the cell, it tricks the cell to move it around. It does that by making the cells think it’s a vesicle, so the system that moves vesicles inside the cells is used by the bacteria to move the bacteria around. It uses this system to move itself not only within the cell, but also directly from cell to cell, so that it can spread while avoiding immune responses. This is a useful system for working on new computational methods because there is an in vitro system developed in France in which a mixture of 5 purified proteins moves a polystyrene bead mimic of the bacterium. Just like our DNA replication system. It’s all been reconstituted in a test tube.

Secondly, work done in the United States has measured nearly all of the rate constants and affinities for the proteins in this set, so you have the constraints you need for the computer-based modeling. In this way, one can produce an agent-based computer model that actually mimics the bacterial behavior. On the left is the actual bacterium with its actin tail pushing it, and on the right is the computer model. In both systems, the actin filament tail forms to push the bacteria ahead and later dissolves behind the bacteria.

The next slide shows a movie of the way it looks in the computer, and since it mimics quite well what the bacteria does, it’s likely to be the way it looks inside of a cell as well. Here the yellow is ATP-actin in the red is ADP-actin. The tightly bound ATP

molecule is hydrolyzed to ADP after the actin polymerizes, and then the red ADP-actin filaments are depolymerized by one of the 5 proteins in the set, called cofilin.

For the many neuroscientists here, the ultimate "emergent property" – one we cannot effectively attack for quite a while until we know a lot more – is the fact of human consciousness, which of course comes from an enormous network of nerve cells in our brains. An ultimate challenge for neuroscientists is to understand how that happens. I'm only talking about much simpler properties we can tackle. Even in the case of the actin cytoskelton inside the organism, we can only hope to move to slightly more complicated systems, stepwise adding more proteins until we hope to eventually understand the entire actin filament network in the cell.

For some more details, we published a special issue of *Science* on computational biology, with review articles that give an indication of how scientists are trying to figure out how to deal with networks of reactions and the emergent properties that we need to understand. Sometimes students get the impression from reading textbooks that we understand almost everything, because we only write about what we understand in textbooks. I think the authors of our textbook would think that we probably only understand 1/20 or 1/10 of what we need to understand, about even cells, much less organisms. So there is a huge amount left to understand.

My second lesson learned from the last edition will also be familiar to neuroscientists because scaffold proteins form an important part of the biochemical network under synapses, on both sides of the synapse. In early editions of our textbooks, we wrote that cells form vesicle bounded compartments like lysosomes or peroxisomes because inside a vesicle, one can concentrate a specific set of proteins, accelerating their reactions by making them collide more rapidly. That's true. But what wasn't true, was that we thought that, "this is the only way that cells can do this." In fact we now know that cells most often speed up their biochemical reactions without requiring a membrane-bound vesicle, utilizing scaffold proteins instead.

In a simple schematic diagram from our textbook, a scaffold protein has binding sites for other macromolecules – a specific set of proteins and RNA molecules that are going to interact with each other. Unstructured regions of polypeptide chain keep these binding sites in close proximity, causing the scaffold's bound macromolecules to collide with each other (and thereby react) unusually rapidly. Scaffolds of this type are positioned in many different places in the cell to allow the cell to compartmentalize its biochemistry. Thus, for example, they give rise to the very specialized chemical regions that underlie each side of a synapse.

Here is a diagram of the simplest actual scaffold that I know about, again from our textbook. As mentioned previously, there is a protein called formin that helps to grow actin filaments. The formin is the red molecule, and the blue molecule is the actin

filament that's growing in an upward direction. Experiments that are carried out in a test tube show that these actin filaments grow at a rate that's faster than "diffusion controlled," -- that is, faster than the rate at which actin molecules would normally collide with the end of the actin filament. But formin has whiskers, which are the scaffolds that make it possible for the actin filament to grow at a much more rapid rate than otherwise possible, because they bind the little blue actin monomers and keep them close by so that they can collide very rapidly with the growing end of the actin filament. This is a very simple example of a scaffold protein used at the end of an actin filament to speed up the chemistry of life.

There are much more complicated scaffolds. We have published a review in *Science* about scaffolds, focusing on the ones that are involved in making protein kinase signaling work. These are still relatively simple scaffolds, more complex than formin, but not nearly as complicated as many. (M.C. Good, J. G. Zalatan, and W. A. Lim, *Science* 332: 680-686, 2011).

Protein scaffolds should be familiar to neuroscientists. We draw cartoons to illustrate the special regions that underlie synapses. When we make such drawings, they show a tangle of protein molecules that include scaffolds with three letter abbreviations for names, revealing our ignorance about what's really happening. My main point is that we will need much more work to really understand in purified systems how these various scaffolds proteins work, and not many people are working on these problems as of yet—very important problems in biology.

In addition, cells are constantly covalently modifying their proteins to move them to specific places in their interior, altering their sub-compartments in response to environmental changes. Thus, cells are nothing like a test tube. When I was a graduate student, we thought the cell was just a concentrated bag of randomly colliding proteins and other molecules. But in fact even in the simplest cells, the bacteria, the proteins are highly organized to create different chemical compartments in different places throughout the cell cytoplasm. And moreover, those compartments are changing all the time as the cell goes through the cell cycle and needs to respond to environmental changes. For example, attaching ubiquitin to a protein can move it to a proteasome for degradation, attaching lipids to a protein can move it to the plasma membrane, while marking it in another way can attach it to different scaffolds throughout the cell. As a result, the biochemical organization of the cell is always changing.

My conclusion is that it will probably take most of this century to gain a true understanding of how cells and organisms work. We will need much more biochemistry in purified systems to reconstitute reactions, as was the approach for deciphering the DNA replication and the Listeria bacterium movement that I showed you. We need much more of that kind of work. But also – and this is why so many physicists and engineers and computer scientists are trying to get involved in helping

us biologists – we are going to need new quantitative methods for analyzing and understanding the enormous complexity of life's chemistry, the so-called "emergent properties" that we don't yet understand. Having whole catalogues of proteins is nice, and being able to draw the many arrows between reactions that happen in the cell is important; but we still can't claim to understand them, even when we have all the details.

Another important challenge that we face is very familiar to everybody here. This is what this symposium is about. It's using our increasingly profound understanding of molecular cell biology to design intelligent strategies for improving human health. Cancer provides a great example. If we really understood what an individual tumor cell was thinking, how it was organizing itself, how it was misbehaving in a deep way, we could effectively make it kill itself. But we cannot do that today.

I'm now going to shift to the critical subject to about how we keep science healthy. There is a tendency in every endeavor for problems to arise from our social organizations. In the case important to us here, these problems can prevent doing the best science.

For science to drive innovation, it's important to recognize how new knowledge arises. I moved to the US National Academy of Sciences in 1993, where for 12 years I was president. That's a full-time job so I closed my research laboratory. But I didn't stop learning because all of life is an education. And one of the things that I came to appreciate over the 12 years, much better than I did before, is how science advances. Part of my education came from the fact that our Academy produced a series of pamphlets in the late 1990's to explain the benefits of science to politicians. Our aim was to encourage the US government to continue to strongly support fundamental research. If you look forward, you can't predict what new discoveries in fundamental research are going to do for humanity. But if you look backwards in time, one discovers that fundamental research has driven most of the innovation that has brought benefit to human beings, mostly by increasing our understanding of how the world works.

The US National Academy of Sciences eventually produced a series of 20 short 8-page pamphlets called *Beyond Discovery*; there is a website that's still makes them freely available. Called *Beyond Discovery: the path from research to human benefit*, each pamphlet takes something that everybody appreciates today, like the *Cure for HIV-AIDS*, and explains its origins. The first one we produced was for the *Global Positioning System* (GPS). Where did GPS come from? How did scientists and engineers create that invention that helps humanity? Each pamphlet contains a centerfold with a two-page timeline. The one for the Global Positioning System starts with the discovery of atomic clocks by physicists in the 1940s. That discovery earned a Nobel Prize; everybody thought it was wonderful but useless, because it enables humans to keep time to one billionth of a second. They said "well, that's a great

discovery, but who cares about keeping time to one billionth of a second?" Well, over time, that ability was combined in unexpected and unpredictable ways with other findings, until at the end of the timeline, there are 24 satellites that enable us to position exactly where we are on earth by the time it takes for their signals to arrive to where we are. And for that purpose, one wants atomic clocks to keep time to a billionth of a second because light travels so fast. Each pamphlet takes a different end product and shows how, over and over again, fundamental research has contributed to bringing great benefits to humanity.

We all know science today is moving much more rapidly than it did 30 years ago, and these pamphlets make the fundamental reason for the explosive growth of science easy to understand. New knowledge comes from combining old knowledge in new ways. If you have more old knowledge, you have many more ways of producing inventions. So, for example, the powerful methodologies that we have today for sequencing DNA, or for analyzing proteins by mass spectroscopy, all depend on earlier discoveries, just like the GPS system does. And these in turn speed the pace at which further fundamental discoveries are made in science.

The accelerating pace is seen in neuroregeneration research, as well as in every other scientific field. The next Figure presents a schematic diagram of my own field, DNA replication, when I was working on it. Our work is at the top – the paper by Barry and Alberts in 1972 was one of our important papers. Of course we only were able to do that research because we could build on other peoples' work: the work on the T4 genetics I talked about, the work by Kornberg on DNA polymerase, as well as many other previous findings.

But if I were to make a similar graph of my field today, you would see that it is moving much faster because there are more elements that a researcher can combine to make a new discovery – many more powerful methods, many more pieces of information to use in your own research.

But there is a catch! As knowledge grows, it becomes increasingly difficult to find the right combinations. There is so much you could do. How do you decide what you are going to do to make a creative new discovery? There is a wonderful quote from the famous French mathematician Henri Poincaré about the source of creativity in science. He said, "To create consists precisely in not making useless combinations and in making those which are useful, which are only a small minority. Invention is discernment, choice… Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart."

This is critical. I tell graduate students in my university, "If you can only go to one seminar a week, go to the seminar that you don't know anything about." Generally the students will go to the seminar they already know almost everything about, so they won't learn every much. They won't help themselves to be different from every other

scientist. To be innovative, you have to be different, and so you need to do something that combines domains that are far apart.
Poincaré says, “The true work of the inventor consists in choosing among these combinations so as to eliminate the useless ones,” and of course, most of them are useless. So this is why being a scientist is like being an artist. A great artist can paint a beautiful picture; but if given all of that artists materials, I can’t paint anything worthwhile. Like a great artist, a great scientist will make fascinating discoveries by combining the right things.

Today there is a major problem that inhibits new discovery in science: the channeling of research topics due to “training inertia.” I am an expert on this problem because I write a cell biology textbook every five years with other authors. Many very important areas of cell biology remain unexplored with modern methods. And yet, at the same time there are all these overcrowded experimental spaces, where many scientists are doing the same experiments. The reason for this is clear. A very successful scientist working on, for example, the Ras protein (to take a historical example) will attract many students and, in the United States, many post-doctoral fellows. Those young scientists will often go on to start their own laboratories, and when they do will continue working on something closely related to what they did before. So we get all these people working on the same questions, trying to publish two weeks before their competitors, and this is not fun. It’s not a good way to do science. It’s also not productive for the scientific enterprise, because we need many more peopling working on the unexplored spaces, like, for example, all those proteins scaffolds underneath synapses. Where are all the biochemists working on them? Well, there are hardly any!

There are many other examples I could give. How then should research be organized to stimulate innovation? My own view was first presented in a little essay in *Cell* magazine in 1985, titled “Limits to Growth: in Biology Small Science is Good Science.” This essay made a lot of people angry, and there were following editorials by scientists saying that I was wrong. But I still believe it. Because I wrote this article, I was asked to chair the *Committee on Mapping and Sequencing the Human Genome* at the U.S. National Academy of Sciences. We produced the 1988 report that led to the Human Genome Project. And because of that success I was eventually asked to become president of our Academy. So, had I not written this essay, I probably would still be doing experiments and never would have gone into science policy.

I believe that to structure institutions for maximum innovation, we should encourage our research institutes and universities to support a set of laboratories of modest size, say 9-12 people at most. (When I was chair of the Biochemistry and Biophysics Department at UCSF, the University of California, San Francisco, we had a rule that no laboratory could be bigger than 12 people). Each of these labs should be headed by an outstanding, innovative, independent investigator. Of course, a small lab cannot cover all the techniques one needs to do modern biology. Each institution therefore

needs to have clustered laboratories embedded in a cooperative culture in which techniques and equipment are freely shared. At my university, for example, graduate students who want to use a technique for the first time in my lab can go to some other lab and just ask another graduate student to help them. They don't need the professor to approve. This is what I mean by cooperative culture: people sharing what they know. Both techniques and equipment need to be freely shared.

Our reward systems are also very important. They must change to strongly encourage risk taking and originality. Now I know that this is a problem everywhere. But basically if we tell young scientists that what's important is the number of your publications, we will discourage creative science. If a scientist needs to produce a lot of publications quickly, he or she won't do anything new. If you want to do something that might be really important, there may not be any publications for a couple of years, as you develop a new system. If senior scientists had told me when I was young that I had to publish several papers every year, I could have never done productive research on the T4 bacteriophage DNA replication system.

And finally, if we want science to thrive, everything must be done to encourage a random collision of people and ideas – as I've implied in my previous discussion of Poincaré's quote. For this reason, specially organized scientific meetings provide another important way to stimulate new science. In the area of neuroscience, as in other types of science, most scientific meetings bring together people who work in the same field. But other kinds of scientific meetings can be even more productive of innovation. The U.S. National Academy of Sciences runs several of these. One was a workshop on schizophrenia that brought 5 experts together with 20 leading scientists in other areas that might be relevant to schizophrenia. In this two-day workshop, there was lots of discussion between experts and non-experts, and their job was to come up with new ideas. One needs some experts because they must educate the non-experts about what we already know and what's reasonable and not reasonable. But we found in another example, that if we had too many experts, it inhibits new ideas: the experts don't necessarily like the fact that people from other areas will have different ideas that the experts missed. So that's why one wants only a small number of experts at such workshops. The schizophrenia results were published in the *Proceedings of the National Academy of Sciences* in 1997. You can read them there. The same type of meetings can of course be held for many other areas of science.

The spaces where productive science is done must be carefully thought out. My office is located on the new Mission Bay Campus of the University of California, San Francisco in a building that looks quite ugly on the outside. It's an ugly building because it was designed from the inside out by scientists to make the science work, and the architect had to make the outside of the building fit what the scientists needed inside. Many other research laboratory buildings have been built where the architects tried to make the outside beautiful so they could win a prize, leaving the inside the building a poor fit for scientists. A good building for science has been designed to

maximize the number of random collisions between scientists. For example, my office is a small one that is clustered with 4 other offices in little area – 5 professors from 4 different departments. Every time I go into my office, I have to walk through the lounge where the students read journals and eat lunch, so I have to walk by the students 5 or 10 times a day.

All the laboratories are big open spaces with multiple groups inside so that everybody collides, and much space is devoted to shared equipment for these laboratories. Nobody owns the space, so when a professor's group gets smaller, somebody else uses those laboratory benches.

I now want to shift to a very important, even broader point. As Academy president for 12 years, the major thing I learned is that it's critically important that science, and scientists, achieve a much higher degree of influence, both throughout their own nations and the world. China is fortunate to have a top leadership now that is largely composed of engineers and scientists. When I interviewed Premier Wen Jiabao for *Science* magazine, I discovered that he thinks scientifically. President Obama is also a strong supporter of both science and scientists. Other nations are not so lucky. But it's very important that the scientific community have a strong influence on what every nation does.

In particular, we need much more of the creativity, rationality, openness, and tolerance that are inherent to science – what's been called a "scientific temper"– I like that word "scientific temper" – for both the United States and all other nations. My favorite quote is from a book, called *Science and Human Values,* written by a physicist, Jacob Bronowski, in 1956. He had flown over Hiroshima and Nagasaki in 1945, just after the bombs had been dropped, when he was in the British Army. And that got him very depressed. Is science good or bad for the world? He spent ten years thinking and writing about this question. The book that resulted is still in print. In it, he concludes that science on balance has been very good for the world. As he writes, "The society of scientists is simple because it has a directing purpose: to explore the truth. Nevertheless, it has to solve the problem of every society, which is to find a compromise between the individual and the group. It must encourage the single scientist to be independent, and the body of scientists to be tolerant. From these basic conditions, which form the prime values, there follows step by step a range of values: dissent, freedom of thought and speech, justice, honor, human dignity and self-respect. Science has humanized our values. Men have asked for freedom, justice and respect precisely as the scientific spirit has spread among them."

As you'll see soon, I spend a lot of time in the United States working on science education for children. This is a critical issue to get our societies, both scientists and non-scientists, to think like scientists do. My next slide shows the image we want for science. This is a picture taken in the front yard of the Academy in Washington DC,

where there is a giant statue of Albert Einstein. He's got a very big lap, and he looks very rumpled and friendly. And the children who come to Washington each year to see their nation's capital often end up sitting on his lap to take their class picture.

Einstein's statue presents an image of science being something that is accessible and friendly. Establishing such an image is something we all need to work on in every country, because many people are afraid of science and think it's some kind of magic. For example, China has fantastic work being done on genetically modified rice, beautiful work done by some of the world's leading scientists. If you ever are going to be able to use that to help feed both your citizens and other people in the world, your citizens must understand the nature of science and how science works, and they must be able to think scientifically about genetically modified crops.

I am going to end this talk by discussing how *Science* magazine can help to stimulate progress in both scientific research and education, which is of course integral to our mission: *Science* magazine exists to help spread science around the world. I encourage young scientists to read the whole front half of *Science* magazine, because this makes you part of the scientific community. There is News and Commentary about what's happening in many different fields of science around the world, plus discussions about data policies and scientific ethics that everybody should know about. To be an innovative and creative scientist, you need to know about approaches from many different fields so you can combine them to do something new, as I have previously emphasized. So, part of my goal as Editor-in-Chief is to try to make the exciting developments in science more accessible everywhere.

Science magazine publishes a lot on science education. This little picture of a girl looking at a plant in one of my many editorials symbolizes what we want science to be like. My grandson, who wants to be a scientist, said that in his high school in the United States, supposedly a good high school, the worst course he ever took was this first year biology course where he had to memorize all the parts of everything. For example, he had to memorize all the parts of a plant. I ask him, "Did you ever look at a plant?" He said, "No, we looked at the pictures of a plant in the textbook." This is not science.

My next slide shows what science should look like in school. This is a photo of 12-year old students in San Francisco; it is a noisy classroom, and the teacher in the back is a coach, walking around, not the source of all knowledge. The students are trying to solve problems, working together in groups of four. Of course, problems are designed to match their age group. We know how to do this. People in China are trying to introduce this kind of science teaching into Chinese schools. But it will only happen in a big way if the scientific community keeps pushing for it to happen, because it is more difficult to teach this way than to teach science as the memorization of science words and relationships.

Science magazine will have published four special issues on science education by the time that I finish as Editor-in-Chief next summer. The first one focused on what we know from research about the use of computers in schools, and the second on what we know from scientific studies about how active science learning improves learning how to read, write and communicate.

We have held two years of contests for the best free science education websites on the web, because the web has wonderful free materials on it, but you cannot find them. Each of the 24 winners of this SPORE contest has published a two-page essay in *Science* that enables a reader to quickly determine whether he or she wants to use the website. Our newest contest is for college science modules that involve students in active inquiry; each winner of this prize for the best inquiry laboratories for introductory college science courses also publishes a two-page essay in *Science*.

I am pleased to be able to announce that *Science* magazine has created a new education website that's freely available to everyone around the world (http://www.sciencemag.org/site/extra/education/). You don't have to subscribe to *Science* to use it. All of our special issues on education are available there, along with all the essays by the prize winners for both the best science education website and best college course module contests. Also on this website is a new project that I am particularly proud of, called *Science in the Classroom*, where we present *Science* research papers annotated with a glossary and other helpful tools, so that even high school students can read them with help from a teacher. If we are to spread a "scientific temper" around the world, we will need to have as many people as possible know what scientists do and how science works. One of the great ways to do this is to have them actually read at least one real scientific paper in their lives. That's our very ambitious goal.

I want to end by summarizing what I have tried to say about biology. Where has biology been? In the past 50 years, tremendous advances have been made in our understanding of the molecular basis of life, largely driven by the development of powerful new techniques. DNA sequencing, when I started, cost almost a million times more than it costs today per base pair. The analysis of proteins has similarly speeded up. When I was a graduate student, we never thought, looking forward, that we could possibly know as much about cells as we do today.

So we can see our way to the end of a remarkable descriptive phase in cell biology, since all of the molecular structures and pathways can now be deciphered in principle, although there is still a lot of work to do. We are collecting vast amounts of data and information. But we now know that the chemistry of life is incredibly complex, by far the most sophisticated chemistry known. For example, if we think of that movie of the protein machine that replicates DNA, we see that life is extremely complicated. But without such amazing chemistry, we couldn't be talking to each other today.

The bacterium that is infected by T4 bacteriophage, *Escherichia coli*, served as the model system for the early days of molecular biology. We know from its genome sequence that it is composed of a little over 4,000 different proteins. It's humbling to recognize that, even today, we have no idea what a quarter of these proteins do. Some Nobel Prizes may be hidden in figuring this out, because among these 1000 proteins one may find some new classes of chemistry, or new types of functions for proteins. That's one of the white spaces, in which there is hardly anybody working today. Innovative new methods and approaches will be needed before we can claim to "understand" even the simplest living cells. And if we don't understand even the simplest living cells, how can we pretend to understand the cells of humans?

Finally, many of the most interesting attributes of life are due to so-called "emergent properties": properties that stem from very complicated networks of chemical interactions, whose consequences cannot be deciphered from the details of a few individual parts alone. Just think of the actin network, with a hundred interacting proteins; even when we know everything about those interactions, the human mind cannot figure out what it means – what will happen inside the cell. Likewise, neuroscience is full of these very complicated networks. We have to invent new ways to figure out how we can understand them.

In conclusion, there are many wonderfully exciting challenges for young scientists. I wish I were still young; it would be great to be a young scientist. The job of us older scientists is to make it possible for you to be successful. Those of us in the United States, in Europe, and elsewhere around the world need Chinese science to be successful, because we all contribute to the world's store of knowledge.

I thank you all for listening and I very much appreciate the opportunity to be here. Thank you!