Biology Past and Biology Future: Where have we been and where are we going?

IAP, Alexandria
December 3, 2006

Bruce Alberts
Professor of Biochemistry and Biophysics
University of California, San Francisco (UCSF)
My start in science

Paul Doty,
Harvard
University:
1958 to 1965
I was a high school student, when the revolution in biology began with the Watson and Crick structure for DNA in 1953.
The Watson-Crick model for information transfer: DNA templating through complementary base pairs

A problem for the DNA replication mechanism: the two DNA strands run in opposite chemical directions
1) The **DNA polymerase** enzyme was discovered by Arthur Kornberg and earned him a Nobel Prize.

2) This protein will add a new nucleotide to the end of one DNA strand (the “primer strand”) only if that strand is paired to a complementary strand that can serve as the template (the “template strand”).
The DNA polymerase in action
How we viewed DNA replication in the 1960’s

DNA polymerase adds one nucleotide and then dissociates

A second DNA polymerase molecule adds the next nucleotide
My dismal career as a graduate student, Harvard 1961-1965

• Can we get the DNA double helix to replicate in a test tube?

• I did many experiments trying to see how this might happen using only the DNA polymerase enzyme.

• They all failed.
When I moved to Geneva Switzerland as a post-doctoral fellow in 1965, I discovered that DNA replication must require much more than the DNA polymerase
Bacteriophage T4, a large bacterial virus, had about 100 genes discovered by genetics by 1965.
A major mystery in 1965: why were there at least 7 T4 genes that were absolutely required for replication of the T4 virus?

1) These 7 T4 genes had been given numbers: 32, 41, 43, 44, 45, 61, 62.

2) One of these, the gene 43, had been shown to produce the T4 bacteriophage DNA polymerase.

3) Why are at least 6 additional proteins needed for any replication of the T4 chromosome when the virus infects the E. coli bacterium?

4) Clearly, DNA replication must involve at least 7 proteins and be much more complicated than anyone had imagined!
A strategy for solving the mystery of so many replication genes:

Try to find the mutant proteins

• Many proteins that function on DNA in the cell will have a site on their surface for binding to DNA.

• By fixing a high concentration of single stranded or double helical DNA on a solid support such as cellulose, one should be able to trap these proteins specifically and purify them.

• Many proteins in a T4 bacteriophage-infected cell bound to a DNA-cellulose column. Would one of them be missing if a mutant virus was used for the infection?
My favorite protein (for historical reasons)

Single-strand DNA binding protein (SSB)

One 32 protein molecule per every 10 nucleotides
Circular DNA single strands with and without 32 protein
Finally, in 1975 we had collected purified preparations of all of the 7 proteins. We discovered that when we mixed them all together we could now replicate a double-helical DNA template!

In contrast, DNA polymerase alone, or a mixture missing any one protein, would only make new DNA when provided with a single-stranded DNA template.

It took another 6 years for us to figure out why each of the proteins was needed in this reaction.
Result of 20 years of research at Princeton and UCSF

The folded replication fork

A protein machine
We now know that the same basic mechanism is used to replicate DNA from large viruses, like T4 bacteriophage, to mammals.

- However, as more complex organisms evolved, each function in T4 was carried out by more proteins.

- For example, bacteria use 13 protein molecules instead of 7, and humans use at least 27!
The magic of protein machines is best appreciated by a movie that shows such a machine in action.

The movie was made by Bruce Stillman at the Cold Spring Harbor Laboratory, as part of the 50 year DNA celebration there. It can be found on Cold Spring Harbor DNA Learning Center website in the section that deals with “Copying the code”.
Two lessons learned in the last 20 years

1) There are remarkable homologies between living things; therefore use model organisms wherever possible.

2) Nearly all cell processes will be:

   • driven by 10 to 20 proteins, organized as a protein machine and incorporating ordered protein movements driven by the energy of ATP hydrolysis

   • based on elegant mechanisms that are too complex to predict.
An Important Challenge for the Next Generation of Cell Biologists

1) Obtaining the information needed to accurately describe the mechanism of every type of protein machine in a cell.

- This will require the reconstitution of many hundreds of protein machines from their purified components, so that their detailed chemistry can be deciphered through reactions studied in a test tube.

- Then, we will also need to work out the many interactions between different protein machines.
My other future challenges come from 25 years of writing this very large textbook.
A cell is the fundamental unit of life
The cell as a collection of catalysts that work together to reproduce the entire collection.
Bacterial cells on the tip of a pin
Today's cells are very complex!

The reaction network of cellular metabolism published by Boehringer-Ingelheim.
A Second Important Challenge for the Next Generation of Biologists

2). Completing our understanding of one type of cell.

• It is not enough to have a catalogue of all the pieces. We must also be able to explain how these pieces all add up to make a living thing.

• For this purpose, many laboratories will need to focus on the same, “simple” cell (for example, Mycoplasma, a tiny bacterium with only 500 genes compared to the 25,000 genes of humans).
The simplest living cell known, a tiny bacterium called Mycoplasma

Dividing cell, about to produce two cells from one
Why did multicellular organisms evolve?

- Multicellularity permits cell specialization, with cells in different places having different functions for the organism as a whole.

- Consider a plant: the advantages include being able to have root cells deep underground to absorb water and leaf cells in the sun to carry out photosynthesis.
Many cells must cooperate to form a multicellular organism: but cooperation is very difficult!

• Single-celled life was all there was on the earth for about 2 billion years.

• Finally, about 1.5 billion years ago, the first cells learned how to form cooperatives and larger and larger organisms began to evolve.
A simple multicellular organism: a “cell cooperative”
More complicated
Much more complicated
A very complex multicellular organism!

Thousands of billions of cells
3). Understanding how cells make decisions in a complex multicellular organism like ourselves.

Cells constantly “talk” to each other. Then each cell integrates what it is “hearing” to control its behavior for the good of the entire assembly of cells that makes up the organism.

We need to understand this complex process of “cell thinking”.
Cell signaling

Signals coming from other cells

Decision network in one cell

Plasma membrane
A sense of how biology really works: the *Bacillus subtilis* competence network
It will probably take most of this century to gain a true understanding of how multicellular organisms work.

Why should we care?

1. Understanding life has great intellectual and spiritual value.

2. Understanding life has great practical value.
The world looks so different after learning science.

For example, trees are made of air, primarily. When they are burned, they go back to air, and in the flaming heat is released the flaming heat of the sun which was bound in to convert the air into tree. And in the ash is the small remnant of the part which did not come from air, that came from the solid earth, instead.

These things are beautiful things, and the content of science is wonderfully full of them. They are very inspiring, and they can be used to inspire others.

Richard Feynman
4). Using our increasingly profound understanding of molecular cell biology to design intelligent strategies for improving human health.

For example, rare aberrant cells give rise to the uncontrolled cell proliferation known as cancer.

Once we truly understand how cells “think”, we can make these cells commit suicide without affecting the other, normal cells of the body.
Part of the signaling system that can cause apoptosis (cell suicide)
A Fifth Important Challenge for the Next Generation of Cell Biologists

5) **Understanding how cells organize, and reorganize, their internal space.**

The cell “cytoskeleton” that organizes this space is much more than the sum of its parts.
Cells have a cytoskeleton formed from networks of protein filaments.
The properties of the cytoskeleton are due to a large set of filament-associated proteins. Some are emergent properties that cannot be understood or predicted without new methods.
For example, our cells can chase bacteria
Some of the molecular components known to drive oriented cell movement.
How can we figure out how such complex systems work?

- One new tool allows us to follow the position of each individual molecule in a living cell by making only that one type of molecule fluorescent.

- When we use this tool, we find some surprising things.
For example, cells are moved by wave-like changes in the organization of proteins that make up the cell cytoskeleton.

(Courtesy of Orion Weiner, UCSF)
We will need mathematical/computational models to decipher such complex systems.

- For example, modern computers make possible individual-molecule based simulations, in which tens of thousands of molecules are simultaneously allowed to diffuse randomly and “react” in a virtual space.

- The simulated positions of all of the molecules can be calculated in cycle steps of one microsecond or less.

- If after billions of such cycles, the emergent behaviors observed in the model closely resemble those in the cell, then one can hope to gain important insights into mechanisms.
We will need mathematical/computational models to decipher such complex systems.

These models require a lot of quantitative data:

1) An inventory of the set of interacting molecules and their molecular structures.

2) Identification of all molecular partners in the set, with determination of the rate and equilibrium constants for forming each partner -- as well as quantitative measurements of the effect of partner formation on other molecules.

3) An understanding of the behavior of this set of molecules in living cells.
One therefore needs to start with model systems that are simpler than the entire process.
One model system: the actin-catalyzed movement of *Listeria* bacteria

Actual bacterium  Computational model
Details from computational model
More complex: the fusion of the sperm and the egg nucleus catalyzed by microtubules

(Computational models courtesy of Gary O’Dell and Jon Alberts, University of Washington)
6). Deciphering the complicated pathways by which cells and organisms evolved on the Earth.

Some powerful tools: comparative genomics, biochemistry, chemistry.
A comparison of genome sequences
Where do we come from?
In summary, where has biology been?

• In the past 50 years, tremendous advances have been made in our understanding of the molecular basis of life, driven largely by the development of powerful new techniques.

• We can see our way to the end of remarkable descriptive phase in cell biology, since all of the molecular structures and pathways can now be deciphered, although this will still require a lot of hard work.
In summary, where is biology going?

• We now know that the chemistry of life is incredibly complex, by far the most sophisticated chemistry known.

• Many of the most interesting attributes of life are properties that emerge from very complicated networks of chemical interactions, whose consequences can not be deciphered from the details of each individual part alone.

• New methods and approaches will therefore be needed before we can claim to “understand” even the simplest living cells.
There are many wonderfully exciting challenges for young scientists!