Research in Progress

One Genome, Fully Deciphered

NINETEEN YEARS of concentrated labor have had their final payoff: this year Giuseppe Attardi, the Grace C. Steele Professor of Molecular Biology, and his research group have completed the functional identification of all the genes in human mitochondrial DNA (mtDNA). In the course of this work they've uncovered quite a few surprises, including the highly efficient mode in which information is packed into the genome, a mode that Attardi calls "a lesson in economy."

In most genetic systems, including the chromosomes in the nucleus of human cells, the great majority of DNA does not actually code for any gene product. There are great stretches of "nonsense" DNA in the spaces between genes, and even within most eukaryotic genes there are long noncoding regions called introns. Not so in mtDNA. Except for one short region, which anchors the mtDNA to the inner mitochondrial membrane and is important in initiating transcription, every part of the mitochondrial genome codes for some product, either a protein, or a ribosomal RNA (rRNA), or a transfer RNA (tRNA).

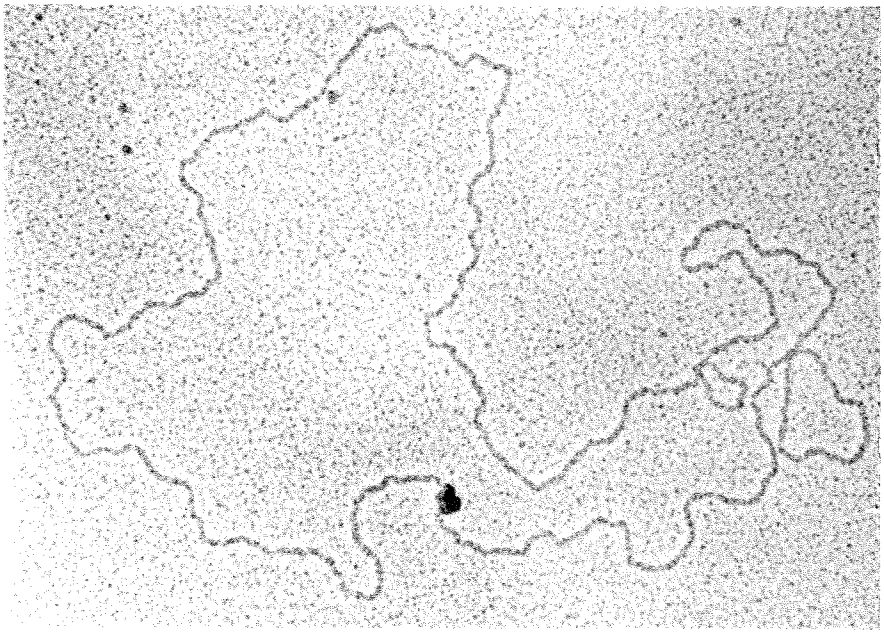
Also, in most genetic systems only one strand of the double-stranded DNA molecule in a given segment codes for gene products. In some cases genes can be found on both strands, but normally when this happens one strand is transcribed in one segment and the other strand is transcribed in the next segment. Human mtDNA is unique in that both strands are completely transcribed.

But even with this extreme degree of information compression, mitochondria (the cell's power plants) are far from autonomous. Attardi thinks that this may not always have been the case. Like many biologists, he believes that the mitochondrion was once an independent cell that long ago mounted an invasion of another cell. In a process known as endosymbiosis, the proto-mitochondrion became its host's partner, providing quantities of energy in the form of adenosine triphosphate (ATP) in return for a protected environment. Eventually, after hundreds of millions of years of evolution, most mitochondrial genes became incorporated into the host's genetic machinery. In mammalian cells this continued until the mitochondrial genome was a ghost of its former self — it contains just 16,500 base pairs in a circle 5 micrometers in circumference. Because of this limited coding capacity, the mitochondrion must import at least 95 percent of the proteins it needs to function.

Much of the content of mtDNA is devoted to coding for the mitochondrion's own protein synthesizing machinery. With some help from the nucleo-cytoplasmic compartment of the cell, mitochondria produce their own ribosomes — the protein-synthesizing factories that translate messenger RNA (mRNA). These mitochondrial ribosomes are scaled-down versions of the ribosomes used in the cytoplasm for the same purpose. And mtDNA codes for 22 different tRNAs. These are the molecules that bring the 20 different amino acids to the lengthening polypeptide chain as a protein is actually being assembled on the mRNA template.

At first, this number of tRNAs was something of a puzzle — the cytoplasm requires 32 tRNAs because in the universal genetic code a single amino acid is often coded for by two or more nucleotide triplets, and each tRNA can recognize only one or two triplets. Although it initially appeared possible that mtDNA-coded proteins might simply lack several amino acids or that the missing tRNAs were imported, Attardi and others were able to discard these two proposed solutions to the discrepancy. The actual explanation is that mitochondria use a simplified version of the genetic code, one in which (among other differences) each of several amino acids is coded for by four triplets that can be recognized by a single tRNA.

Apart from the different RNA species coded for by mtDNA, it became obvious, when the DNA sequence was determined in F. San-



This high-resolution electron micrograph shows one molecule of human mitochondrial DNA. The molecule forms a circle 5 micrometers in circumference, and it contains about 16,500 base pairs. The dark knob at bottom center represents a residue of the inner mitochondrial membrane, to which the DNA molecule is attached.

ger's laboratory in England, that the mitochondrial genome codes for 13 different proteins. It proved relatively easy to identify six of these proteins because of their amino-acid sequence homology to known proteins. Four turned out to be proteins important in the respiratory chain — a series of electron carriers that transfer to oxygen the electrons released during the final stages of the oxidation of food molecules. (These four proteins are cytochrome b and three subunits of cytochrome *c* oxidase.) The fifth and sixth proteins were found to be two subunits of ATPase, the enzyme that uses the energy derived from the oxidation of food molecules to synthesize ATP.

This left seven "unidentified reading frames" (URFs) in mtDNA — 60 percent of its protein coding capacity. Although it was possible that the URFs were nonsense sequences, Attardi and his colleagues were able to reject that hypothesis for two reasons. For one thing, these sequences were highly conserved in different mammalian species, something that never happens with nonsense DNA. For another, the researchers had clear evidence that the URFs were expressed — that is, transcribed into mRNA, which in turn is translated into protein. But, says Attardi, determining the functions of these proteins "appeared to be a terrible job. I thought that a possible shortcut would be to show that these proteins were somehow related to each other, that they belonged to the same complex."

This indeed turned out to be the case. Two of Attardi's postdocs, Anne Chomyn and Paolo Mariottini, collaborating with Russell Doolittle's group at UC San Diego, and using antibodies raised to the proteins encoded by individual URFs, were able to show that each of these antibodies under certain conditions precipitated six of the proteins produced by the URFs, indicating that they were all part of the same enzyme complex. Further investigation showed that this enzyme complex is another important component of the respiratory chain — NADH dehydrogenase. Finally, this year the remaining URF was shown to encode another subunit of this same enzyme, a finding that will appear soon in the journal *Science.*

Even though the last URF has finally been identified, Attardi claims that he has not put himself out of business. "Our effort now goes in the direction of studying how the expression of this genome is regulated. These are very intriguing problems."

And Attardi also hopes to study a number of genetic diseases that seem to result from damage to the mitochondrial genome. "There is a whole group of diseases called mitochondrial myopathies, which mostly affect the muscular system but sometimes also affect the nervous system and the heart. All these myopathies are characterized by alterations in mitochondria — both structural alterations and functional alterations. And the enzyme complex that is most frequently altered is NADH dehydrogenase."

This brings Attardi back to his original motivation for working with human cells rather than with yeast or *Neurospora,* both of which are more amenable to genetic analysis. "I'm an M.D. I'm interested in man not only as an example of a mammalian species or a higher eukaryotic cell, but also as an organism for which there are medical problems, practical problems to be solved." □ — *RF*

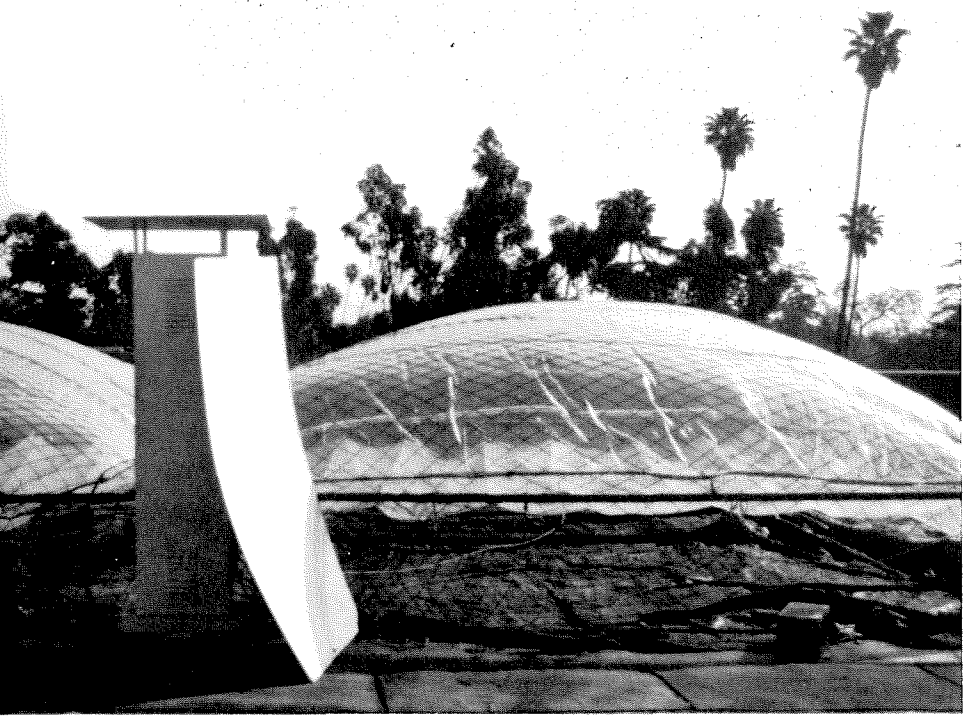
Smog — In Bags and Balances

DURING AN AVERAGE Los Angeles morning rush hour, thousands of tons of pollutants from hundreds of thousands of cars join thousands of tons of pollutants from industrial smokestacks. As the prevailing winds carry these compounds in a generally westerly direction, the sun causes them to begin reacting with each other. These chemical reactions produce ozone, nitric and sulfuric acid, and many other things as well. Some of these condense to form suspended aerosol particles, which cause the haze that we call smog.

Understanding this process is a job every bit as complex as predicting tomorrow's weather, but John Seinfeld and his colleagues are very close to achieving this goal. Seinfeld, the Louis E. Nohl Professor and professor of chemical engineering, has developed a detailed computer model of Los Angeles smog.

"You could, in principle, use this model to predict tomorrow's smog," says Seinfeld. "The only problem is the expense. You would need a supercomputer more or less dedicated to this use. The chemical mechanisms of smog involve 100 to 150 reactions. Then on top of that you've got the particle formation process, which is of comparable complexity. Our long-term goal is to have such a model on a computer that the Air Quality Management District could use."

However, "a computer model like this is only as good as the fundamental chemistry and physics in it," notes Seinfeld. In particular, the formation and composition of aerosol particles



On the roof of the Keck Laboratory is a large Teflon bag that Caltech researchers fill with hydrocarbons and nitrogen oxides to simulate early morning air. Sunlight drives the reactions that produce smog, which is analyzed by instruments housed in the structure in the center of this photograph.

are still only poorly understood. In an effort to learn more about atmospheric aerosol, Seinfeld and Richard Flagan, professor of environmental engineering science and mechanical engineering, together with graduate students Mark Cohen, Carol Jones, Gideon Sageev and Jennifer Stern, are conducting both large-scale and small-scale experiments.

To study large-scale smog phenomena, the researchers have installed a smog chamber — a room-sized Teflon bag — on the roof of the Keck Laboratory. They fill the bag at night with precisely controlled concentrations of hydrocarbons and nitrogen oxides to simulate early morning air, and then they cover the bag with a tarpaulin. The next morning they remove the tarp, and the sunlight begins driving the chemical reactions. Periodically they withdraw air through ports in the bag and measure gas concentrations and particle numbers and sizes.

One important question that remains to be answered is the extent to which aerosol particles are formed by nucleation or by condensation on pre-existing particles. Nucleation occurs when the partial pressure of a gas gets so high that aerosol particles precipitate spontaneously from the saturated air. It had been thought that nucleation played an insignificant role in atmospheric aerosol formation, since condensation will continually occur when the atmosphere is full of particles on which gases can condense. But Jennifer Stern has found that nucleation can and does occur at a significant rate. This happens when the concentration of a gas increases at such a rate that it exceeds the capacity of the condensation process.

The small-scale studies of aerosols are conducted at a very small scale indeed — that of individual aerosol particles. In a variant of the Millikan oil-drop experiment, the researchers use a device called an electrodynamic balance to levitate a single charged droplet, just five micrometers or so in diameter, between hyperbolic electrodes. Once the particle is in place, they can determine its precise size and, in a most remarkable recent accomplishment, Gideon Sageev has measured the particle's infrared spectrum and thus its composition.

Determining the particle's size is fairly straightforward — they simply turn off the charge on the electrodes and let gravity take over briefly. By measuring how long the particle takes to fall a predetermined distance, a computer can automatically calculate its size.

Sageev has recently developed an ingenious way to determine the infrared spectrum of an aerosol particle. Normally, to determine an infrared spectrum, light of various infrared wavelengths is shined through a sample of liquid or gas. When the wavelength of the light matches a particular molecular resonance, the molecules absorb the light and this absorption is inferred by detectors that see a decrease in light transmission through the sample. But with such a small particle the amount of light absorbed is infinitesimal, so measuring it is impossible.

Instead, the researchers take advantage of the fact that when the droplet absorbs some infrared energy, it heats up and a tiny bit of the liquid evaporates. Less than a single layer of atoms on the surface of the particle evaporates, but the laser light scattered by the particle changes enough to signal even this minuscule change in size. The device slowly steps through the infrared wavelengths, measuring size changes with each step, a process that takes about one hour.

The researchers plan to use the device to look at the composition of particles containing sulfates. Such particles are a common by-product of industries that burn coal. This is an important question since Carol Jones has shown that the composition of these particles changes with time, especially if they contain small quantities of catalysts such as manganese.

As the results of these experiments come in, they will add detail to Seinfeld's numerical model of smog. It's particularly appropriate that these developments have happened at Caltech, since the late Arie Haagen-Smit virtually invented smog research here. Notes Seinfeld, "Around 1950 Arie Haagen-Smit really figured out the relationship between oxides of nitrogen and ozone — a critical step in understanding the system. In a sense my research is a natural progression from what Haagen-Smit started." □ — *RF*