

A Look Into Life's Chemical Past

A computer model of gene regulation yields some evolutionary clues

By RICHARD LIPKIN

Inscribed into the twisted strands of every organism's DNA is the genetic information necessary to give it life. But without the ability to extract and translate that information, an organism could not survive: Its genetic message would remain unreadable, like an ancient tablet locked in a vault.

Biologists marvel at nature's method of chemical code breaking — how each cell manages to make from the whole DNA only the proteins it needs.

the many diseases in which cellular processes have gone awry.

Moreover, the existence of a universal code for regulating gene transcription may affect our view of life's origins. If the mechanism for decoding genes has persisted through eons of evolution as a remnant of our primordial past, then it may offer provocative insights into biological history.

His instincts as a biochemist about how one molecule ought to recognize another told Harris that a universal code must explain the actions of regulatory proteins. "Most biochemists at the time didn't know of any code," he recalls. "But I knew that there had to be some code there."

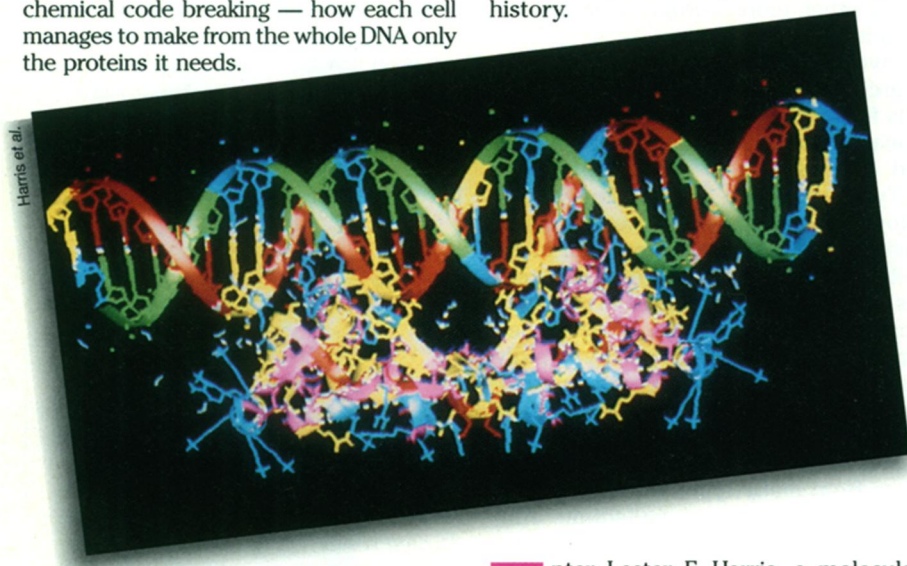
Many scientists believed that once they could see what was happening at the molecular sites where the regulatory protein and DNA bind to each other, a code directing this process would become evident. Researchers used X-ray crystallography to figure out exactly what those specific sites look like and how the various chemical parts fit together.

In fact, researchers did determine the structure of the proteins interacting with DNA. But no obvious code for protein-DNA recognition appeared.

So Harris and his colleagues decided to gamble on an entirely different approach to the problem. Convinced that a molecular recognition code must exist, they began studying the interactions between regulatory proteins and DNA on a computer. Working in conjunction with the University of Minnesota's Supercomputer Institute in Minneapolis, Harris' team used sophisticated computer models and some new search methods to compare gene sequences of a wide variety of protein binding sites.

Their gamble appears to have paid off.

They hypothesized that during the evolution of DNA and regulatory proteins, "genetic information is conserved." They observed that the pieces of the regulatory proteins that recognize DNA share common genetic information



In this simulated time series, a DNA double helix (ribbons at top) recognizes a glucocorticoid receptor protein (bottom molecule). In a fraction of a second, they meet, join, and separate (left to right).

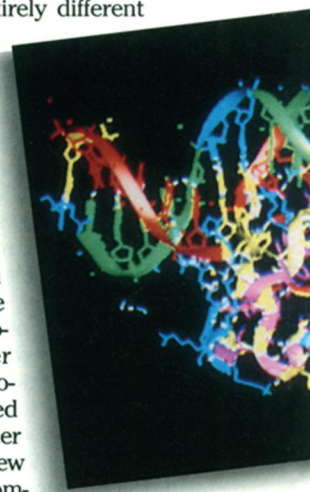
What controls the process of transcription, whereby a cell extracts from DNA the information it needs to make proteins? Do the many molecules that regulate the transcription process work in unrelated ways, or does some common chemical mechanism — a universal code — guide their action?

The answer to this question has strong implications for cancer treatment, genetic engineering, and biotechnology. The ability to influence the chemical machinery that makes proteins from genetic instructions would give scientists powerful tools to help people suffering from

Enter Lester F. Harris, a molecular biologist at Abbott Northwestern Hospital Cancer Research Laboratory in Minneapolis. A decade ago, he and his colleagues began to ponder some basic questions about transcription.

"We were studying cancer, looking at the role of certain regulatory proteins when we realized that very little was known about how regulatory proteins recognize and bind to DNA," Harris says. "We were amazed. So we decided to take a closer look at the whole process."

Researchers have long known that regulatory proteins act as shepherding molecules, making cameo appearances during transcription. But how do these proteins know where to go and what to do? How do they find just the right spot on DNA to park themselves and set in motion transcription's cascade of chemical events? If transcription fails, a distressed cell is left with a botched gene product.



with the DNA sites to which they bind.

This led them to predict what the regulatory protein's binding structure ought to look like. They also began to see common features and gene sequences that suggested to them a possible underlying code for recognition.

That code might explain how regulatory proteins could maneuver their way along DNA and trigger the transcription process.

In one case, that of a glucocorticoid receptor protein, they found a specific spiral structure — an alpha-helix — that appears to spur transcription. Almost like a key turning on an engine, this structure, upon binding to DNA, seems to cause the twisted strands to begin to bend, separate, elongate, and unwind, all events associated with initiating transcription, Harris says.

By modeling the molecular interactions with a supercomputer — then making predictions, which were confirmed in laboratory experiments — Harris's team put forth some provocative hypotheses.

"Our findings... support the idea of a stereochemical [three dimensional] basis for the origin of the genetic code," the researchers stated in issue 2 (1994) of the JOURNAL OF BIOMOLECULAR STRUCTURE & DYNAMICS.

They conclude that a few primitive molecular structures may have determined the manner in which genetic information formed. Because amino acid interactions with DNA depend largely on their chemical structures, they have probably

DNA sequences later used to transmit genetic information.

The binding sites of regulatory proteins and DNA provide compelling examples of the conservation of genetic information, Harris says. Their simple, modular forms suggest that they are "primordial remnants" of life's chemical past. On primordial Earth, peptides and nucleotides could have acted as templates for one another in a process of "autocatalysis," or self-directed molecular synthesis, he says.

This process not only would have allowed basic biological molecules to flourish, but also would have preserved primitive chemical machinery. This notion fits a larger view, shared by other molecular biologists, of how modern biological processes may have emerged.

"I find Harris' work fascinating," says Lawrence B. Hendry, a biochemist at the Medical College of Georgia in Augusta. "The idea that there is a code for the interaction of proteins and nucleic acids makes a lot of sense, as does the notion of a stereochemical relationship between proteins, amino acids, and nucleic acid bases.

"His basic concept is probably correct in that it seems to correspond to what people are finding in their laboratories. While these ideas are still somewhat controversial, people have been discussing them in the [scientific] literature, and there's data to support them."

Hendry adds that if Harris' results are fully confirmed, "they could have important implications.... Looking way into the future, I think that potentially important medical consequences could come out of this research."

John H. Miller, a computational chemist at the Department

"Nearly every investigator in human genome research is excited about the possibility of understanding gene expression at the molecular level," Miller says. "It's a fundamental question in biology. One aim of medical technology is to turn on or off specific genes responsible for diseases. So finding some general principles for gene regulation is very important. Right now, the real bottleneck in genome research is interpreting vast amounts of genetic data so they can be put to practical use. Harris' work could help."

Using computers to model biological processes has its dicey side. Not all models pan out, some don't accurately represent reality, and others just confuse people. Those stumbling blocks have left some researchers skeptical about Harris' hypothesis.

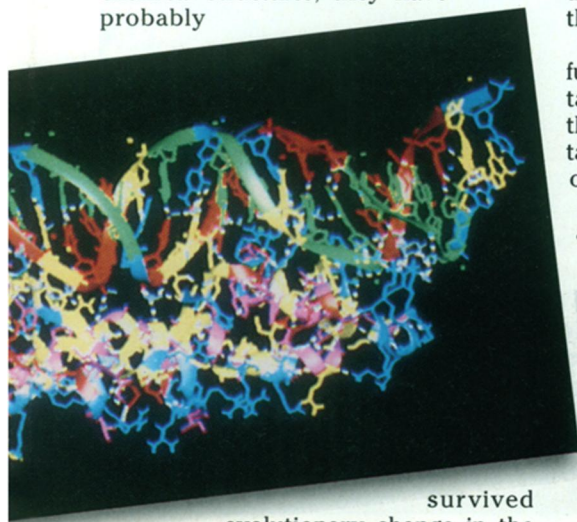
Ultimately, models must prove their worth in the real world of flesh and blood. The truth, the skeptics point out, always comes out in the more traditional "wet labs."

"It's been tough to convince skeptics about the validity of what Harris has done, because his computational methods are so advanced that few people can understand them," says Matthew Witten, a computational medicine researcher at the University of Michigan in Ann Arbor. "It's taken a while for people to realize the significance of what he's doing. But I think his work is very leading edge."

In coming years, Witten believes, gene researchers will see increasing interplay between wet labs and the newer computational labs in what has come to be called molecular medicine.

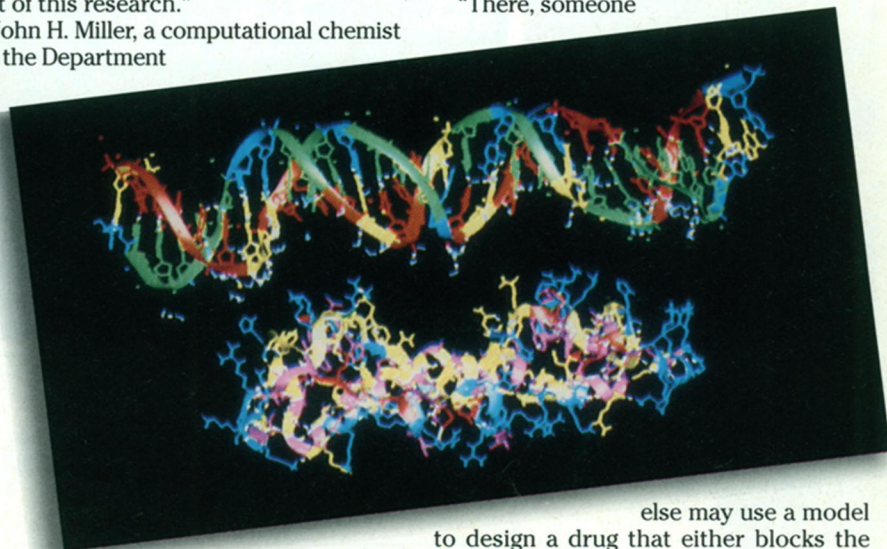
"Someone in a wet lab will find a gene, sequence it, and ship the data over to someone in a computational lab.

"There, someone



survived evolutionary change in the genome. As such, they would have been conserved through time.

The roots of our present system for passing genetic information from one generation to the next, they contend, may have evolved earlier than previously suspected. In the prebiological world, as a few molecules became predominant, they would have influenced other molecules with which they interacted. Harris calls this process "template-dependent evolution" because pieces of proteins, or peptides, could have served as molds for



of Energy's Pacific Northwest Laboratory in Richland, Wash., agrees. "Harris' ideas about the conservation of genetic information are very novel, though it will take a while to ferret out all of the generalities."

else may use a model to design a drug that either blocks the gene or the gene product. Then they'll take the drug back to the wet lab for testing.

"This back-and-forth process," says Witten, "is what molecular medicine is all about." □