
July 25, 2006

Scientists Say They've Found a Code Beyond Genetics in DNA

By [NICHOLAS WADE](#)

Researchers believe they have found a second code in [DNA](#) in addition to the genetic code.

The genetic code specifies all the proteins that a cell makes. The second code, superimposed on the first, sets the placement of the nucleosomes, miniature protein spools around which the DNA is looped. The spools both protect and control access to the DNA itself.

The discovery, if confirmed, could open new insights into the higher order control of the genes, like the critical but still mysterious process by which each type of human cell is allowed to activate the genes it needs but cannot access the genes used by other types of cell.

The new code is described in the current issue of Nature by Eran Segal of the Weizmann Institute in Israel and Jonathan Widom of [Northwestern University](#) in Illinois and their colleagues.

There are about 30 million nucleosomes in each human cell. So many are needed because the DNA strand wraps around each one only 1.65 times, in a twist containing 147 of its units, and the DNA molecule in a single chromosome can be up to 225 million units in length.

Biologists have suspected for years that some positions on the DNA, notably those where it bends most easily, might be more favorable for nucleosomes than others, but no overall pattern was apparent. Drs. Segal and Widom analyzed the sequence at some 200 sites in the yeast genome where nucleosomes are known to bind, and discovered that there is indeed a hidden pattern.

Knowing the pattern, they were able to predict the placement of about 50 percent of the nucleosomes in other organisms.

The pattern is a combination of sequences that makes it easier for the DNA to bend itself and wrap tightly around a nucleosome. But the pattern requires only some of the sequences to be present in each nucleosome binding site, so it is not obvious. The looseness of its requirements is presumably the reason it does not conflict with the genetic code, which also has a little bit of redundancy or wiggle room built into it.

Having the sequence of units in DNA determine the placement of nucleosomes would explain a puzzling feature of transcription factors, the proteins that activate genes. The transcription factors recognize short sequences of DNA, about six to eight units in length, which lie just in front of the gene to be transcribed.

But these short sequences occur so often in the DNA that the transcription factors, it seemed, must often bind to the wrong ones. Dr. Segal, a computational biologist, believes that the wrong sites are in fact inaccessible because they lie in the part of the DNA wrapped around a nucleosome. The transcription factors can only see sites in the naked DNA that lies between two nucleosomes.

The nucleosomes frequently move around, letting the DNA float free when a gene has to be transcribed. Given this constant flux, Dr. Segal said he was surprised they could predict as many as half of the preferred nucleosome positions. But having broken the code, “We think that for the first time we have a real quantitative handle” on exploring how the nucleosomes and other proteins interact to control the DNA, he said.

The other 50 percent of the positions may be determined by competition between the nucleosomes and other proteins, Dr. Segal suggested.

Several experts said the new result was plausible because it generalized the longstanding idea that DNA is more bendable at certain sequences, which should therefore favor nucleosome positioning.

“I think it’s really interesting,” said Bradley Bernstein, a biologist at [Massachusetts General Hospital](#).

Jerry Workman of the Stowers Institute in Kansas City said the detection of the nucleosome code was “a profound insight if true,” because it would explain many aspects of how the DNA is controlled.

The nucleosome is made up of proteins known as histones, which are among the most highly conserved in evolution, meaning that they change very little from one species to another. A histone of peas and cows differs in just 2 of its 102 amino acid units. The conservation is usually attributed to the precise fit required between the histones and the DNA wound around them. But another reason, Dr. Segal suggested, could be that any change would interfere with the nucleosomes’ ability to find their assigned positions on the DNA.

In the genetic code, sets of three DNA units specify various kinds of amino acid, the units of proteins. A curious feature of the code is that it is redundant, meaning that a given amino acid can be defined by any of several different triplets. Biologists have long speculated that the redundancy may have been designed so as to coexist with some other kind of code, and this, Dr. Segal said, could be the nucleosome code.