Do Notches Mark DNA?

A very short DNA sequence has an unusual structure and occurs in control regions of bacterial and eukaryotic DNA

Certain proteins, it is clear, "recognize" DNA sequences. But what is it that they recognize and how do they do it? A group of researchers at the University of Pennsylvania proposes that DNA in regulatory regions may have notches, or "molecular detents," like the notches on the balance beam of a scale. When proteins slide along DNA searching for the regions where they are to bind, they may be stopped by these detents. The investigators find that a very short DNA sequence seems to form notches in the DNA of both bacteria and higher organisms.

The University of Pennsylvania investigators, Ponzy Lu, Susannie Cheung, and Kim Arndt, came upon their molecular detent while studying one of the oldest and most difficult questions in molecular biology. Ever since the earliest days of molecular biology, researchers have been asking what is it about certain sequences of DNA that makes them regulatory regions. Although they know the genetic code that associates the sequence of DNA bases with corresponding sequences of protein amino acids, they have no idea what sort of code signals regulatory proteins to bind to certain segments of DNA. But they do have reason to suspect that regulatory regions of DNA have more to them than just their sequences. The consensus among molecular biologists is that these regions have unusual structures that may be dictated by their sequences but that may also be determined by the structures of sequences that surround them.

Lu has spent much of his scientific career asking why a particular regulatory sequence—the *lac* operator of the bacterium *Escherichia coli*—is recognized by a particular protein—the *lac* repressor. One of his methods is to use nuclear magnetic resonance, or NMR, to examine the structure of the *lac* operator and to see how it changes in mutants whose *lac* operators can no longer bind repressor. What Lu does is to perturb the DNA and then look at the resonances of hydrogen-bonded base pairs of DNA, the so-called imino protons. "In principle, this part of the NMR spectrum is very far to the left and you see no resonances of proteins. It is easy to see DNA and it is a nice place to look at DNA sequences," Lu explains.

The standard procedure for looking at an NMR spectrum is to heat a DNA segment and to assume that it starts to melt from either end. With that assumption, investigators can make sense of the spectral data they get. And when they tested the assumption by looking at the NMR spectra and melting of symmetric, synthetic pieces of DNA, the pieces did indeed melt from the ends inward.

But another technique is to irradiate specific protons of the DNA and observe perturbations of nearby protons, a method that takes advantage of an effect called nuclear Overhauser enhancement (NOE). The application of this technique to nucleic acids was first suggested by Alfred Redfield of Brandeis University in 1981 and has been used extensively ever since, particularly by Brian Reid of the University of Washington. The advantage of NOE, says Lu, is that "you get the imino proton sequence with no prejudice, you don't have to assume the DNA melts from the end."

So Lu, Cheung, and Arndt decided to try the NOE method, in combination with a measurement of the imino protein exchange rate, on the *lac* operator. They immediately saw that, contrary to their expectations, something in the middle melted first. "We were confused," Lu recalls. But he remembered that there was some rationale for their finding because David Kerns of the University of California in San Diego had reported that A-T sequences and G-C sequences melt independently. From the sequence of the *lac* operator, this should mean that melting should start in the middle as well as at the ends.

Putting their NMR findings together with the biological information they had on the *lac* operator sequence, Lu and his colleagues realized that the sequence of interest in the middle of the operator is GTG/CAC. Not only does this sequence seem to have a peculiar property, as indicated by the NMR data, but it seems to be crucial for the binding of *lac* repressor protein, since when it is altered genetically the protein does not bind at all.

Next, Lu, Cheung, and Arndt asked whether this short sequence occurs in other key regulatory regions of DNA. First they looked in bacterial DNA at places where proteins recognize DNA sequences. They found GTG/CAC sequences everywhere. The sequences appear, for example, in the two operator sequences of the *gal* operon of *E. coli*. It appears in three places where the cyclic AMP binding protein interacts with the *lac, gal*, and *ara* operons. It is in the six operators of the virus lambda.

The sequence also is in control regions of eukaryotic DNA. It shows up in areas where Z-DNA forms, the strange structure that seems to be associated with active genes. It is in the long terminal repeat regions of some retroviruses and in the DNA coding for the joint in immunoglobulins. It is in the enhancer sequences of animal viruses—the sequences that are needed to turn on genes. And, Lu finds, the sequence in the enhancer of the tumor virus SV40 also melts early.

But since the GTG/CAC region is so short, could these appearances of the sequence simply have occurred by chance? Lu thinks not. He did a computer search of the European Molecular Biology Laboratory DNA sequence bank and also examined published frequencies of dinucleotides to ascertain that GTG/CAC occurs no more often than would be expected if it occurred at random, which is 1/64 of the time. But in the regulatory regions that Lu, Cheung, and Arndt examined, it occurs at five times that frequency. Moreover, in many cases there are two of these sequences symmetrically arranged, a finding that, Lu says, "further reduces the possibility of a random quirk."

Not every regulatory site has a GTG/CAC sequence, but Lu and his associates find other short sequences, such as GAG/CTC, seem to occur unusually often. He suspects that these sequences too will have unusual structures when examined by NMR.

Lu concludes that local alterations in DNA structure, such as that provided by GTG/CAC, may be just enough to tell proteins that are sliding down DNA to slow down and take a second look. He does not think that this is a sequence used for specific binding, but he believes that it is structures like this that may tell the proteins where to bind. "This is a testable hypothesis," he remarks.

—GINA KOLATA

Additional reading

1. P. Lu, S. Cheung, K. Arndt, *J. Biomolec. Struct. Dynamics* **1**, 509 (1983).
2. S. Cheung, K. Arndt, P. Lu, *Proc. Natl. Acad. Sci. U.S.A.*, in press.