

together with another plasmid in which a yeast *UAS_G* is placed upstream of a CAT reporter gene, induction of GAL4 results in substantial activation of the reporter. Janice Fischer, working in Tom Maniatis's laboratory at Harvard, extended these experiments. She developed lines of *Drosophila* in which the synthetic GAL4 genes were placed under the control of a tissue-specific *Drosophila* promoter; the reporter gene was *lacZ*. Since the developing embryos are transparent, *lacZ* activity can easily be demonstrated by staining for β -galactosidase. The clear results were that β -galactosidase was synthesized, and at high levels, *only* in those parts of the embryo where the promoter is known to function. Preliminary experiments show that analogous constructions also work in tobacco cells. Perhaps most importantly for further analysis of the detailed aspects of activation interactions, purified GAL4 hybrids, synthesized in bacteria, work quite efficiently when added to *in vitro* transcription systems.

Having regulatory sites located far apart—as with *UAS_G* and the *GALI* promoter—amplifies the number of components that can be brought together to form a genetic control mechanism. A particularly instructive system for elucidating the molecular mechanisms by which proteins sepa-

rated by hundreds of base pairs can still interact has been the arabinose operon of *Escherichia coli*. This operon is principally controlled by the AraC protein, which both positively and negatively regulates transcription of the *araBAD* gene-cluster. In the absence of arabinose, the protein dimer binds at two sites (*araO₂*, and *araI*) that are separated by 210 base pairs, and is thought to form a loop between them by tetramerizing the two proteins. This contact prevents AraC bound at the *araI* site from assuming an inducing conformation and turning on the genes for arabinose catabolism. In the presence of arabinose, the loop opens and the protein is able to contact RNA polymerase and activate transcription.

Schleif has used plasmid constructs containing the relevant sites, and the *araBAD* promoter fused to easily assayed enzymes like galactokinase, to ascertain the topological and distance constraints to DNA looping. If looping occurs as postulated, insertion of base pairs between the binding sites—which would rotate AraC to the opposite face of the helix—should inhibit repression as long as the binding energy of the tetramer was not sufficient to counter-rotate the DNA. This prediction was confirmed experimentally using a series of insertions, each two base pairs longer than its prede-

cessor: Shifting the relative orientation of the binding sites by half a turn, alternately inhibited and allowed *araBAD* repression.

By varying the length of the inserted DNA, Schleif was also able to show that the binding sites can be separated by as many as 400 base pairs without significantly affecting tetramer formation. He noted that, in general, the maximum distance between two looping sites was a measure of the strength of the interaction between the bound proteins. When deletions were made between the two sites, the minimal distance that still allowed repression was one helical turn. Since this introduces a considerable bend in the DNA, Schleif suggested that the protein itself may constitute a significant part of the loop.

Protein-protein regulatory interactions can also be modified by conformational changes induced as a result of DNA binding itself, as shown by the existence of mutants in the *araI* site formally equivalent to λ *pc* mutations. These mutants bind AraC essentially as well as the wild-type, but abolish the ability of the bound protein to act as a repressor. Thus contacts in the helix that are *not* used in protein binding can act as allosteric effectors to influence the ultimate regulatory functions the bound protein may perform. —Harvey Bialy

COMMENTARY

VISIONS OF THE FUTURE, ACCORDING TO HOOD

MIAMI—Does technology drive science or vice-versa? Whichever, it is perfectly clear that without the microchemical instrumentation now available—protein and DNA sequencers and synthesizers—scientists would not be embarking on their most enormous journey yet—sequencing the human genome. That Cal Tech's (Pasadena, CA) Leroy E. Hood is a major force in this effort goes without saying.

Scientists at Cal Tech's microchemical facility continue to refine existing technology while simultaneously creating new technology that makes it obsolete. For instance, it is now possible to microsequence individual protein fragments directly from two-dimensional (2-D) gels. Together with colleagues at the Jet Propulsion Lab, Hood's group is developing new image analysis techniques that will allow researchers to compare a 2-D gel with *any other* 2-D gel ever run. This makes possible subtractive analysis—a snapshot of a cell's complete physiological state. "We can identify proteins in quantitative state changes," Hood explains. "We believe that this will give

us enormous access to organisms' phenotypes."

In the next year or two, Hood hopes to improve the capacity for sequencing by another three orders of magnitude—with a combination of new chemistries, new instrumentation, and new techniques for separation and analysis. Increased capacity to sequence proteins and clone their genes demands faster DNA sequencing methods. Current instruments can sequence about 1500 nucleotides per day; Hood claims that next year's model will be able to approach 100,000 bases in 24 hours.

How does Hood propose to extract useful information from this exploding data base? One approach, which Cal Tech scientists have embarked on with TRW (Cleveland, OH), is to use fast data finder chips—which were developed to analyze satellite data gathered over the Soviet Union. These chips can extract units of information from massive electronic backgrounds. According to Hood, this is exactly the kind of fast reading we need to analyze DNA. The basic idea is to feed the entire 15 million nucleo-

tides now in data bases such as Genbank through a series of chips programmed to extract data that fit search criteria. That informational filtrate passes to another chip—programmed with other conditions. And so on, until, "At the end, you have extracted from that enormous data base just that information containing the particular patterns of interest." In fact, his group has already used this approach to extract all the sequences that are consistent with DNA rearrangements of immunoglobulin T-cell receptor genes—in minutes.

With all these technologies in place, it will become possible, for the first time, to compare a total genotype with a total phenotype. With subtractive analysis, predicts Hood, "We can identify the spots that change qualitatively and quantitatively, microsequence them, enter these sequences into the human genome data bank, identify the genes they correspond to, make homology searches to see if they resemble genes whose functions are already known, and then begin to study directly the functions of those genes." —Jennifer Van Brunt