Dimers Direct Development

A common molecular mechanism mediated by the "helix-loop-helix" proteins has unified a remarkable array of different areas in developmental biology

NOT LONG AGO, RESEARCHERS FISHING cancer genes from mammalian viruses might have thought they had little in common with geneticists busily counting bristles on a fruit fly wing. But recently workers in these disparate fields—and others—have found their research paths merging in an extraordinary way, leading them toward what could be a unifying molecular theme underlying much of animal development.

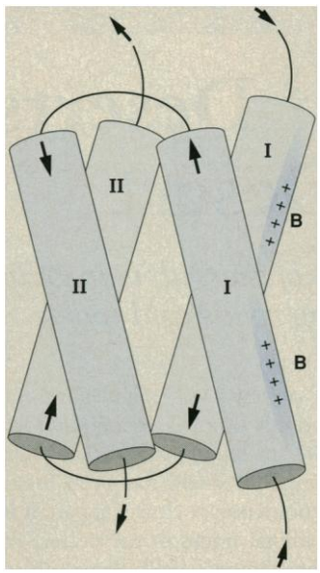
What they have found—in most cases quite unexpectedly—is that they are working on the same family of proteins. And those proteins have a striking trait in common: they link up in mixed pairs that then bind to DNA, guiding the development of muscle, nerve, and sexual characteristics (and no doubt other features) in animals ranging from flies to mammals.

The connectedness of these systems might have been overlooked had it not been for something the proteins had in common—a bit of similarity to the cancer-causing gene *myc*. But although *myc* gave a powerful initial impetus to the field, the mechanism of *myc*'s action remained poorly understood. Now the story has come full circle—with the discovery that *myc* too has a protein partner that helps it bind to DNA (see article by Blackwood and Eisenman on page 1211 of this issue of *Science*). And that mechanism in turn provides an intriguing—but as yet unexplored—link between the proteins that guide cells toward differentiation and those—like *myc*—that tell them to keep dividing.

It's a tale of "fields pulling each other along," says Robert Eisenman, whose research team at the Fred Hutchinson Cancer Research Center found *myc*'s protein partner. Thomas Cline, who studies sex determination in flies at the University of California, Berkeley, adds: "It's such a beautiful example showing that if work is well done in several fields...it will come together in a way that will be synergistically informative.

The seeds of this synergism were sown in the 1970s, when *myc* was first identified as the cancer-causing gene in several tumor viruses. Like other oncogenes, *myc* is also found in normal cells, where it seems to have a role in cell division, although no one is sure exactly what its mechanism is.

While virologists were isolating oncogenes such as *myc*, researchers in the laboratory of Antonio Garcia-Bellido at the Universidad Autonoma de Madrid were studying an apparently unrelated cluster of genes in the fruit fly *Drosophila*. Known as the *achaete-scute* complex, these genes appeared to be required for development of the sensory nervous system. In flies having mutations in *achaete*, *scute*, or one of the other genes in the complex, bristlelike sensory organs called chaetae fail to develop.



C. Murre and D. Baltimore

Two to tango. *The helix-loop-helix segments of two proteins (one protein shown in front of the other) may join to create a stretch of amino acids (+) that bind DNA.*

As the Spanish group studied *achaete-scute*, Harold Weintraub of the Hutchinson Cancer Research Center in Seattle was taking a different approach to another subject in development: muscle differentiation. Weintraub and co-workers used cultured cells to search for proteins that could activate muscle-specific genes. Sure enough, they found a protein (they called it MyoD) that converts a variety of mammalian cells into musclelike cells. MyoD turned out to be part of a small family of proteins that seem to play overlapping roles in muscle-cell differentiation—much as the four genes of the *achaete-scute* complex do in *Drosophila* nervous system development.

Then came the first hint of correspondence between these diverse systems. While Weintraub's group was cloning the *myoD* gene, Carlos Cabrera and Juan Modolell, also at the Universidad Autonoma de Madrid, cloned the *achaete-scute* genes. In 1987 both Cabrera and Weintraub published the same startling observation: the proteins of both families share the same similarity with the *myc* protein. In each case, they found a smattering of a dozen or so identical amino acids, spaced over a 50–amino acid stretch in the tail of the protein. "We practically fell off our chairs," says Cabrera, now at the Marie Curie Institute in Oxted, England. The similarities between the three proteins suggested that all of these proteins may work via a similar mechanism, though no one yet knew what that mechanism might be.

Soon another developmental system was drawn into the story. Lily and Yuh Nung Jan of the University of California, San Francisco were studying a mutation in *Drosophila* that prevents development of sensory neurons. Tom Cline, then at Princeton, had been studying the genes that lead to sex determination, including a gene, called *daughterless*, that is required for female development. To their surprise the Jans found that their *Drosophila* neuronal mutation also mapped to that same gene. Cline and the Jans thus found themselves simultaneously sequencing *daughterless*—based on two apparently unrelated roles in development. And, in what was beginning to be a recurring theme, *daughterless* turned out to be homologous to *myc*.

The sequence similarity between these genes—each of which clearly played a central role in a different type of development—was intriguing, but what it meant was quite unclear. It took the entry of a fourth group from yet another field—immunology—to come up with a clever idea about what the common region might do.

In 1988 Cornelis Murre, in David Baltimore's lab at the Whitehead Institute, was looking for proteins that activate the immunoglobulin genes of the antibody-producing white blood cells called B lymphocytes. He found two related proteins, which he called E-12 and E-47—and found they

belonged to the by now rapidly expanding club of *myc*-like proteins.

Murre and his co-worker Patrick McCaw then focused their attention on the region that all the proteins had in common. They found that its amino acid sequence made the protein capable of forming a pair of helices. Each helix seemed to have a stretch of hydrophobic amino acids lined up along one side—significant because such amino acids can play a role in linking proteins.

That structure rang a bell in the Baltimore lab because not long before researchers in Steven McKnight's lab at the Carnegie Institution had proposed that pairs of protein helices rich in the amino acid leucine can form a "zipper" that binds some proteins together. Murre and McCaw proposed that their sequence, which they dubbed a "helix-loop-helix"(HLH) motif, also provides means by which proteins can join to form a dimer (a pair of proteins) that then goes on to regulate gene expression.

According to Baltimore this was a key insight, because it linked the protein sequence data with function in a new way. "There had been so much [sequence information] in the literature, but none of it fell together until they realized that the sequence had this dimerization potential."

Indeed, this structural clue paved the way for quick advances. The Baltimore group divided HLH proteins into three groups. Class A contains *daughterless*, E-12, and E-47; each was found in all tissues tested. Into class B went the *achaete-scute* and MyoD proteins, which were expressed only in certain tissues. In mix-and-match experiments, they found that class A proteins join with class B proteins, and that the dimers can bind specific DNA sequences. (Because it bound to neither A nor B proteins, *myc* was put in a class by itself.)

The recognition that it may take two different proteins linked in a dimer to bind DNA was crucial, says Cabrera: "We had been trying for a year and a half to get DNA binding [with the *achaete-scute* proteins] because we were convinced they were transcription factors. But working with one at a time, we didn't get any results." After Murre's finding, Cabrera's group realized it takes two to tango—and that *daughterless* was the missing partner.

But that was by no means the end of the story, partly because other results suggested that—in some cells—there is a system opposing the action of the joined A and B type proteins. Those findings came simultaneously from the labs of Modolell, Weintraub, and James Posakony at the University of California, San Diego. Posakony and Modolell had both cloned a gene called *extramachrochaetae* (*emc*), discovered in Garcia-Bellido's laboratory in the early '80s. Flies with a mutation in *emc* have too many chaetae, apparently because, in the absence of normal *emc* protein, the *achaete-scute* genes are active in places where they normally would not be. Normal *emc* protein, therefore, seemed to be opposing the action of *achaete-scute* proteins.

DEVELOPMENTAL DIMERS

	NEUROGENESIS (DROSOPHILA)	MYOGENESIS (MAMMALS)
POSITIVE REGULATOR	ACHAETE-SCUTE	MYOD FAMILY GENES (INCLUDING MYOGENIN, MYF-5, MRF-4)
NEGATIVE REGULATOR	EMC	ID
UBIQUITOUS PROTEIN	DAUGHTERLESS	E-12, E-47

At the same time, Weintraub's lab, during a screen for more HLH genes, had also come up with a negative regulator. They called their gene *Id* (for inhibitor of DNA binding), because when introduced into cells, it blocked the action of MyoD. Once the three groups sequenced the genes, the reason for these proteins' inhibitory role became crystal clear. Both had the HLH sequence needed for protein-protein pairing—but they lacked an adjoining region that Weintraub's group had shown to be required for binding to DNA. "We talked to each other and got very excited about the parallel result," says Weintraub. "The two studies complement each other quite nicely."

Indeed, while the *Drosophila* groups had solid genetic evidence that *emc* plays a real negative regulatory role in the animal, they lacked the hard biochemical evidence for how *emc* works. Weintraub's group, working in cultured cells, could not vouch for *Id*'s role in animals, but they had the evidence of its biochemical function.

With the revelations about *Id* and *emc*, the outlines of the full picture appeared. In cells that make all three protein types, positive regulators (such as MyoD or the *achaete-scute* genes) compete with negative regulators (*Id* or *emc*) for binding to a "ubiquitous" protein that appears in many cells (*daughterless*, E-47, E-12). The winner of the competition determines the cells' fate. Positive regulators seem to turn on key developmental genes, and so when they win, the cell heads down a specific developmental pathway—becoming, say, a muscle cell. If the negative regulators win, the cell does not choose that particular fate.

This arrangement seems ideal for laying down patterns in development, Cabrera says. It gives biological systems a way to create an intricate pattern using only broad strokes. The pattern of expression of any one gene needn't be highly resolved, but the overlap of positive and negative regulators produces protein-protein interactions that yield a refined final pattern—defining muscles, nerve cells, or other features.

Competition between proteins is also an appealing means for determining sex. In *Drosophila* sex is decided by the ratio of X chromosomes to autosomes (nonsex chromosomes). If one of the HLH competitors were coded on the X-chromosome, it would be twice as abundant in females (which have two X chromosomes) as in males (which have only one). If that twofold boost were enough to beat out a negative competitor, the protein could prevail specifically in females.

In fact, results from several different laboratories, including that of Lucas Sanchez of the Centro de Investigaciones Biologicas in Madrid, suggest that this may be how sex is determined in *Drosophila*, with *scute* (which is located on the X chromosome) serving as the positive regulator and *daughterless* as the ubiquitous protein. The story is not finished, notes Cline, until a negative regulator is found, but HLH fans like Posakony are betting on *emc* as a likely candidate.

"It's really incredible that [HLH proteins] play such a vital role in three completely different developmental processes," says Posakony. "It certainly gives you a sense of the potential versatility of a system like this."

And now that remarkable versatility has been extended to *myc* itself—the system that got the ball rolling in the first place. Until now, *myc* wouldn't pair with any known HLH protein. But now Eisenman's group has found *myc*'s partner—a protein they call *max*. When linked to *max*, *myc* binds well and specifically to DNA, a finding that will help researchers close in on *myc*'s true function, in normal cells and in cancer.

The concentration of *myc* protein seems to be key to *myc*'s action, in both normal and cancer cells. Competition for pairing—the dominant theme of the HLH proteins—could be what makes *myc* levels so crucial, says Eisenman. Having found that *max* proteins will bind to each other, Eisenman favors a scheme in which *max* is bound to *max*, "until *myc* comes knocking at the door," to compete for *max*, forming a *myc-max* dimer with special—and still unknown—gene-regulating potential.

"The breakthrough in *myc* is the fact that the partner has been found," says Posakony. "The underlying theme in this whole story is dimerization."

■ MARCIA BARINAGA