Gene Splicers Contemplate the Rat Brain

Genes active in the brain may carry a tag in noncoding regions, and a potential new neurotransmitter has been identified

The brain is being probed with a newfound audacity by collaborating cell and molecular biologists at the Research Institute of Scripps Clinic in La Jolla, California, and the nearby Salk Institute, whose skills allow them to pluck genes from complex mixtures with considerable selectivity. This pick-and-choose approach, applied with an admitted initial component of "brute force," is leading these researchers to identify apparently brain-specific genes, including what could represent a wholly new family of neuropeptide transmitter molecules.

Also, and perhaps more important, this approach has led them to speculate that genes used in the central nervous system carry an identification tag: a specific sequence of 82 nucleotides. This identifier, or ID sequence, is located not at the beginning of the gene, where most other control signals are found, but within the noncoding regions (introns) that interrupt the protein-coding segments of eukaryote genes.

Not only does this discovery challenge the notion of introns being "junk DNA," it also raises the possibility that these noncoding sequences serve a key function for eukaryotes during development and differentiation. Though Floyd Bloom of the Salk Institute confined his discussion at a recent meeting in Boston* to the ID's occurrence in the rat brain, his collaborator J. Gregor Sutcliffe at Scripps says that a similar phenomenon may occur in other tissues, where different ID-like sequences have been found. The additional members of this research team include Thomas M. Shinnick, Joel M. Gottesfeld, and Richard A. Lerner at Scripps and Robert J. Milner at the Salk Institute.

The California team found that DNA clones containing an 82-nucleotide, noncoding segment hybridized to an unusual RNA molecule found in the cytoplasm of brain cells, Sutcliffe says. That molecule is long-lived, contains 160 nucleotides, and is present in large quantities in the brain but not the liver or kidney; it now is known to consist of the 82-nucleotide ID segment and a polyadenylate tail. Early on, the California researchers observed that this small molecular weight RNA molecule would hybridize to many clones of DNA copies of RNA [called complementary DNA (cDNA)] from the brain, even when that cDNA clearly represented much larger genetic messages.

The abundance of this small RNA—there are about 2000 copies of it present per brain cell—sparked curiosity. Trying to understand it at first was much like "the blind men trying to describe the elephant," Sutcliffe says. But eventually nucleotide sequence comparisons led him and his colleagues to notice that the 82-nucleotide sequence kept popping up elsewhere in brain transcripts. Now the 160-nucleotide RNA is considered peripheral to their interests and merely the molecule that "provided a way into the problem" of ID sequences. "We don't know if it's functional or why it's long-lived. We believe it's the polyadenylated version of the primary transcript of the ID sequence," Sutcliffe says.

> A virtually blind approach has turned up an altogether new neurotransmitter.

The ID sequence has been found—with an identical match for 62 of its 82 nucleotides and close to that in the remaining 20—in all 11 of the "brain-active" genes studied so far, according to Sutcliffe. He and his colleagues estimate there are 30,000 genes active in the rat brain.

On a grander but also considerably cruder scale, more than half but probably fewer than all of the unprocessed RNA transcripts in the rat brain contain ID sequences, Sutcliffe says, referring to a measure that is based on hybridization experiments rather than on nucleotide sequencing.

There are often several copies of the ID sequence in any one gene, and there also appear to be copies in pseudogenes, which are inactive but preserved genetic regions resembling genes. It seems likely that, like transposable elements, ID sequences can jump around the genome. Thus, there are many more ID sequences in the genome than there are genes believed to be active in the brain—perhaps five times as many, or 150,000. The California researchers even speculate that ID sequences might move "anywhere where their effect is not lethal."

These estimates for how widely and selectively the ID sequences are distributed within the rat genome are based on several different experiments and also, partly, on a comparison of other published gene sequences with the 82-nucleotide ID sequence. One basis for that estimate derives from a series of 191 cDNA clones obtained from rat brain messenger RNA (mRNA) constructed by the California group. Five of these clones, including four whose sequences have been determined, contain an ID sequence, Sutcliffe says. The gene corresponding to another clone, obtained specifically from the brain and designated 1B236, has been studied more extensively (see below). This gene contains an ID sequence in two of its four introns.

A second basis for those estimates comes from studying five other brain-specific DNA clones, obtained from genomic DNA rather than from DNA copies of messengers. Each contains an ID sequence, a finding that Sutcliffe considers extremely unlikely to be due to chance.

Third, and certainly the weakest link in this chain of evidence so far, is that ID sequences have been identified in genes that do not fit so neatly into the hypothetical pattern. For example, the same "brain" ID sequence is found within the published gene sequences for growth hormone, prolactin, a pseudogene for tubulin, and for a small nuclear RNA gene. Both growth hormone and prolactin are made in the pituitary gland, and Sutcliffe argues that their ID sequences may have alighted there by chance and are likely to be inactive. The appearance in a pseudogene of an ID sequence is rationalized differently—by arguing that, if such segments are mobile like transposable elements, they may sometimes alight in odd places.

The explanation for how ID sequences work is far from complete. It seems likely that the ID sequence must be transcribed into RNA before the gene containing it can be switched on. This activation might involve a change in chromatin structure, Sutcliffe says.

Unexpectedly, the enzyme RNA polymerase III is believed to transcribe ID-containing sequences in the brain (with the help of a protein factor) as a prelude

*Molecular Biology Now & Tomorrow—Thirty Years of DNA, Boston, 19 to 21 September 1983.

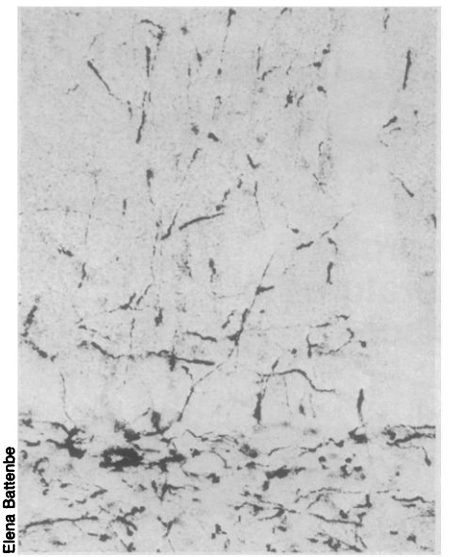
to activating the whole ID-containing gene. Typically, another enzyme, RNA polymerase II, is solely responsible for transcribing protein-encoding genes into RNA messages. "The simple notion, based on experience with bacteria, is that genes will be controlled at the 5′ end of a message," Sutcliffe says. "Here, we seem to have another element—the ID sequence—that is thousands of nucleotides away from the RNA polymerase II initiation site and one that seems to be transcribed by RNA polymerase III." Experiments with the inhibitor α-amanitin, which specifically blocks polymerase II, clearly show that polymerase III is involved in ID transcription. "We need to connect that to the transcription of genetic messages by polymerase II, but we have no idea of a biophysical mechanism," Sutcliffe says.

The brain ID sequence is thought to be a "positive and general neural signal," one that is set because of a cell's developmental history that establishes it will reside in the brain. A similar mechanism may be at work in other tissues, Sutcliffe says, noting that he and his colleagues find ID-like sequences arising elsewhere in the rat. These molecules "are similarly sized but of a different nucleotide sequence" from the ones in the rat brain.

The ID sequence cannot be telling the whole story of differentiation, however, in that a cell's position within the brain also is important in determining that cell's ultimate role. Thus, the California researchers are arguing that the ID sequence acts as a general marker for the genes that are used in the brain. A corollary to this argument is that other factors besides ID must operate during differentiation to fine tune the process. With such studies it becomes possible to define the brain's "hardware," Sutcliffe says. "Someone still has to figure out the program."

Among the brain-unique genes containing an ID sequence and belonging to the growing collection of brain-specific "hardware," referred to by Sutcliffe, is one designated 1B236, the catalog number for its place in the rat-brain gene library developed by the California team. Though the protein and peptides encoded by 1B236 have been characterized fairly extensively, their true function is not yet proved, Milner says. Nonetheless, the prospect is tantalizingly open that, the first time it has been tried, a virtually blind recombinant DNA approach has turned up an altogether new neurotransmitter molecule.

The clonal library of the brain-active genes was constructed by making cDNA from messenger molecules obtained



Elena Battenbe

Rat brain pathways

Neuronal fibers stained by antibodies to the brain-specific protein 1B236 in the CA3 region of rat hippocampus.

from the brains of adult rats. Only those mRNA molecules that were present in brains, but not in either kidney or liver cells, were considered. The 1B236 clone was selected from that library, before its promising features were recognized, partly on the basis of its size and the completeness of the clone, and partly because it is one of "the rarest we can routinely detect" among active genes from rat brain cells, Sutcliffe explains. Also, the cloned version of the molecule contains a large open reading frame, another early indication to the researchers that it codes for a protein.

The inferred amino acid sequence of 1B236 shares significant features with other active neuropeptides. Specifically, it has along its length several pairs of basic amino acids that in other neuropeptides serve as markers for enzymes that carve large precursor molecules into small, active, diffusible peptides (*Science*, 22 April 1983, p. 395; 14 May 1982, p. 720). The pairs of basic amino acids in 1B236 potentially define three complete peptides and a fragment, according to Milner and his colleagues.

Taking the inferred amino acid sequences of those three peptides and gambling on their biological importance, the California team synthesized each of the three peptides and made antibodies to them. Each of the antibodies was expected to recognize only one of three of the peptides, which are nonoverlapping, and the precursor protein from which they are derived. The first use of the antibodies was for doing fairly detailed neuroanatomy to answer the fundamental question: Are the peptides found in the brain?

The answer is more than just affirmative. The peptide-recognizing antibodies react selectively at various sites in the brain, describing a pattern throughout, Milner says. The antibody experiments "strongly suggest that an extensive system of neurons with a reproducible pattern of circuitry relationships contain this protein either as a full-length species or in processed pieces," he and his colleagues have reported. Although the patterns suggest there is "more than one neural pathway," Milner says, the patterns are "very specific in each of the regions. . . . They are fairly broadly but discretely organized." The most intense staining occurs in the neocortex, particularly in the somatosensory cortex, and in the hippocampus and the cerebellum, he adds.

The target of those antibodies turns out to be a membrane-bound, carbohydrate-containing protein having an apparent molecular weight of 100,000. Its distribution in the brain "corresponds pretty closely" to that determined by the antibody-staining pattern, he adds.

Milner notes that although the putative precursor protein is abundant compared to the peptides presumably derived from it, this does not rule out their biological importance. By comparison, the precursor of enkephalin is predominant in adrenal cells, but the smaller peptides derived from that precursor are biologically active. Thus, it remains an "open question" whether the 100,000-dalton protein is a precursor that gives rise specifically to those smaller peptides.

Yet another piece of evidence is consistent with a neurotransmitter role for some or all of the 1B236 peptides. When they are squirted onto appropriate brain cells, they can alter the firing rate, or action potential, of those cells. However, proving that this is a specific effect is not so easy, Milner says.

Milner and his collaborators are not worried by the absence so far of an "established" biological function for the 1B236 peptides, especially as their discovery was made in such a novel manner. "This is only the beginning of a long story," he says. "The initial protein chemistry using recombinant DNA is a lot easier to do compared to the traditional approach. I'm quite happy to do protein chemistry on something interesting and novel that is possibly functional in the brain."—**JEFFREY L. FOX**

Additional Reading

1. J. G. Sutcliffe, R. J. Milner, F. E. Bloom, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4942 (1982).
2. J. G. Sutcliffe *et al.*, *Cell* **33**, 671 (1983).
3. R. J. Milner and J. G. Sutcliffe, *Nucleic Acid Res.* **11**, 5497 (1983).