On the Origin of Introns

Conflicting data are causing a revision of ideas on the concept of exon shuffling and its correlation with structural units in proteins

One of the great puzzles of modern molecular biology is the origin of introns. Are they ancient features of genes, which in some cases have been selectively or randomly lost (completely so in prokaryotes)? Or do they represent relatively recent insertions, in the manner of transposable elements, into previously intact coding sequences?

Four years ago Walter Gilbert, then of Harvard University, suggested that the mosaic structure of eukaryotic genes was the basis of exon shuffling, the recombination of different coding regions through crossing-over. Colin Blake, of Oxford University, England, refined the idea by arguing that exons might code for discrete, stable regions of proteins. Exon shuffling might therefore assemble new proteins through the novel reassortment of stable substructures.

These proposals are consistent with the idea that the mosaic nature of eukaryotic genes is an ancient feature. And they carry the corollary prediction that introns would always be found between regions that code for definable structural units in proteins. Although gene sequencing has proceeded apace in recent years, information on protein structure has been less forthcoming. In those cases where it has been possible to test the exon-shuffling notion in some way, the results have fallen fairly sharply into one of two categories. Either they strongly support the proposed correlation between exons and protein substructures, such as with the immunoglobulins and globin; or they equally strongly appear to contradict it, such as with actin and myosin. "It is all very confusing," comments Blake.

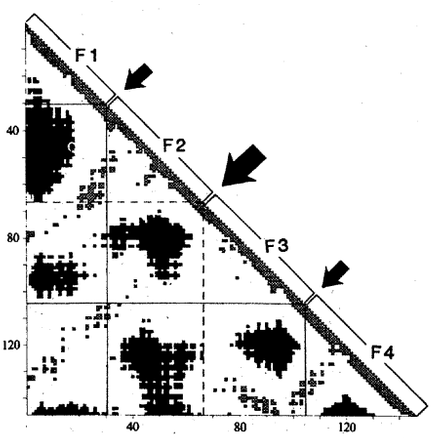
Immunoglobulins represent the most striking support for exon shuffling, both in structural and functional terms. Not only do different exons encode discrete domains of the protein, but the generation of antibody diversity also involves the somatic shuffling of exons that code for antigen-binding regions.

For globin the picture is a little less clear, but is still persuasive. The fact that the basic structure of globin genes—three exons separated by two introns—is preserved over great phylogenetic distance argues for stable rather than mobile introns. Unlike immunoglobulins, however, globin cannot readily be divided into discrete structural domains. Nevertheless, when Mitiko Gō, of Kyushu University, Japan, constructed a distance map of alpha carbons for globin in June last year, she found that, although the protein is not constructed from discrete domains, it can be divided into compact regions she calls modules.

Gō, who recently gave a talk at the National Institutes of Health in Bethesda, describes protein modules as smaller than domains, containing between 20 and 50 amino acids as compared with 100 or more. And whereas two domains have only limited contact with each other, neighboring modules may have extensive areas of contact. A single domain may be made up of several modules.

Initially Gō's distance map for globin was puzzling, because it revealed four modules rather than three as would be expected if they were to correlate directly with coding regions. As shown in the diagram, the two known introns fall precisely at the junction between modules F1/F2 and F3/F4. Why was the central region of the protein split into two modules, F2 and F3, if no intron divided them? Gō speculated that an intron in an ancestral gene had been precisely excised from that junction. Her prediction was borne out almost immediately with the publication of the structure of leghemoglobin, thought to represent a primitive form of globin, which showed a third intron in the required position.

In her NIH talk, Gō described distance maps, not yet published, for hen egg white lysozyme and cytochrome c. She divides lysozyme into five modules and says that the known introns fall at the junctions between modules 1/2, 3/4, and 4/5. She suggests that an intron once separated modules 2 and 3, but that it has been recently excised.



Distance map of β hemoglobin

Distances between alpha carbons greater than 27 angstroms in black; between 9 and 27 angstroms in grey; less than 9 angstroms in white. Lines through white regions show module boundaries. Small arrows show position of introns. Large arrow indicates position of excised intron. [Nature ***291***, *90 (1981)*]

The distance map for cytochrome c shows four modules, says Gō, which would imply a maximum of three introns. The rat gene for cytochrome c has just one intron at codon 56, and this corresponds closely with the junction between modules 2 and 3. The "missing introns," says Gō, have been excised.

Precise excision of introns has of course been demonstrated in nature, with the recent discovery of processed genes. In this case it appears that a gene is transcribed into RNA, its introns are spliced out, and the processed message returns to the genome via reverse transcription, possibly with a retrovirus as intermediary. It is easy to see how all introns in a gene might be lost, but less so for just one of, for example, two. The loss of a single intron from a pair has apparently occurred, however, in the case of rat insulin genes. Rats have two insulin genes, one of which has two introns whereas a second has just one of the pair.

Gō speculates that at some early stage in the history of all genes, modules would have been separated by introns. Exigencies of space and energetics served to remove introns from prokaryotes, she says, whereas removal in eukaryotes is much less complete. The evolutionary benefits of a continued facility for exon shuffling presumably favors their retention. "I wouldn't expect to see introns between all modules in modern genes," explains Gō. "But where there are introns I would expect them to be between modules."

The cases in support of legitimate exon shuffling look strong, but there are examples that simply do not rest easy with it. For instance, John Karn, of the Medical Research Council's Laboratory of Molecular Biology in Cambridge, England, shows that the mysoin genes in nematodes present several problems. There is a family of four myosin genes in this worm, each member of which has about seven introns. Not only are the positions of the introns not uniform be-

0036-8075/82/0903-0921$01.00/0

Pulsed Neutron Sources Okayed

A Department of Energy (DOE) panel has reaffirmed that the Los Alamos National Laboratory should become the lead institution for U.S. pulsed neutron scattering research in the late 1980's. But, in a turnabout of previous policy, the panel also endorsed the continued operation of a pulsed source at the Argonne National Laboratory until Los Alamos is in full swing in 1986.*

There are two main sources of neutrons for neutron scattering, a technique that gives spectroscopic and structural information about a wide range of materials, from crystalline solids to biological molecules. Nuclear reactors, which emit a steady stream of thermal neutrons, are the mainstay of DOE's neutron scattering program. Accelerator-based sources of pulsed neutrons are expected to become increasingly important, however.

Los Alamos already has a pulsed neutron facility. Part of the beam from the laboratory's 800-million-electron-volt proton linear accelerator LAMPF is diverted to a tungsten target. Protons from the beam strip away neutrons from the nucleus. To enhance the usefulness of the setup, Los Alamos has just begun the construction of a $19-million storage ring that will collect protons from LAMPF and shape them so that the resulting neutron pulses are shorter and more intense. Most of the Los Alamos facility is being paid for by DOE's Office of Military Application, which is interested in the use of neutrons for weapons studies. This was part of the reason why an earlier panel (headed by William Brinkman of Bell Laboratories) recommended to DOE's Office of Basic Energy Sciences, which supports neutron scattering research, that Los Alamos rather than Argonne become the center for pulsed neutron studies (*Science*, 16 January 1981, p. 259). The new panel (also headed by Brinkman) reaffirmed this recommendation saying, "optimal use should be made of DOE resources to ensure a natural, gradual shift to and buildup of the Los Alamos facility by the 1986–1987 time frame," provided that its experimental hall is upgraded considerably.

The original Brinkman panel concluded that without a substantial increase in funding, DOE should not support two pulsed neutron sources, and recommended that Argonne's source, then just about to begin running, be terminated. DOE was unwilling to take such a drastic step and told Argonne to continue full operation through fiscal 1983 but begin winding down in fiscal 1984 and cease running in fiscal 1985.

Argonne has made the most of its opportunity. Although limited finances prevent running for more than 6 months per year, the reliability of the source is such that a neutron beam is available 90 percent of the scheduled time, and the intensity of the beam is close to the promised level. Some 70 experiments by researchers from 25 universities, industrial laboratories, and DOE laboratories were performed in the first 6 months of running. Argonne is establishing a large user community, with benefits both to U.S. neutron scattering and to the laboratory's security as a valued national facility. Louis Ianniello of DOE acknowledges that "we were reaching an irreversible decision point on what to do about Argonne's pulsed neutron source," and that this uncertainty was one reason for setting up the panel.

Although Argonne, which wants to expand to 8 months of running, could face severe financial restraints in fiscal 1983, the "big decision is for fiscal 1984 and beyond," says Ianniello. The Brinkman panel found that "Without question, [Argonne's pulsed neutron source] has demonstrated its value in a variety of experiments and will be effectively used for research in condensed matter physics, materials science, and molecular biology for the next few years if funding is available." It specifically called for funds to keep Argonne's source going through fiscal 1984.

In the end, however, it remains DOE's policy to shift pulsed neutron scattering research to Los Alamos, where the intensity of the neutron beam will be considerably higher than at Argonne. The life of Argonne's source may be prolonged, but not beyond 1986.—ARTHUR L. ROBINSON

*"Report of the Pulsed Neutron Research Review Committee" (unpublished). Panel members were: J. D. Axe, Brookhaven National Laboratory; R. J. Birgeneau, Massachusetts Institute of Technology; W. F. Brinkman (chairman), Bell Laboratories; H. A. Mook, Oak Ridge National Laboratory; and J. J. Rush, National Bureau of Standards.

tween different myosin genes, but in some of the genes introns fall in the middle of repeated structural units, units that code for a 28 amino acid repeat that constitutes the rod part of the protein.

These as yet unpublished data on myosin contrast markedly with those for collagen. Like myosin, collagen is composed of a multiple amino acid repeat, three residues in this case. But unlike myosin, these repeats as represented in the gene are never interrupted by introns. Collagen supports the modified exon-shuffling notion while myosin apparently does not.

A second problem example comes from the recently published data on human α_1-antitrypsin gene.* Savio Woo and his colleagues at Baylor College of Medicine, Houston, show that this gene has important homologies with chicken ovalbumin. And yet the number and position of the introns in the two genes is different. Antitrypsin has three introns while ovalbumin has seven, with possibilities for overlap limited to one pair.

One interpretation of these data is that the ancestral gene for antitrypsin and ovalbumin had as many as ten introns, seven of which were lost en route to one descendant gene, while a different three were shed en route to the other. Woo and his colleagues consider this unlikely and suggest that "some of the introns could be vestiges of transposable elements that had been inserted into preexisting exons of the two genes after their divergence from an ancestral gene."

This second interpretation—that of mobile introns—has also been invoked on a number of occasions to explain the disparity in number and position of introns in actin genes throughout the animal and plant kingdoms. It is certainly possible to argue that introns can insert and excise themselves, in the manner of transposable element, and those that happen to fall between regions coding for protein modules will be preserved because of the benefits of recombining exons. In other words, exon shuffling could be a secondary phenomenon following from propitious intron insertion. The one great drawback in the argument for intron mobility is the total lack of evidence of sequence structure in introns that is so characteristic of transposable elements.

The idea of exon shuffling and all that it implied was powerfully seductive, so much so that, as one commentator said, "it led people to enthusiastic over interpretation." The contrary examples provide a cautionary and provocative force.
—ROGER LEWIN

*M. Leicht *et al.*, *Nature (London)* **297**, 655 (1982).

0036-8075/82/0903-0922$01.00/0