MEETING REPORT

# GENETIC ENGINEERING IN THE PRECAMBRIAN

NEW BRUNSWICK, N.J.—Even though bacteria have a much longer evolutionary history than plant or animal cells (when measured by the number of cell divisions), they are likely derived from an ancestor with a eucaryote-like genome. This heterodox view was elaborated here by Walter Gilbert (Biogen, Cambridge, MA) at the second *Bio/Technology*-Waksman Institute Symposium—Biotechnology: Current Aspects.

Since 1977 scientists have known that the genes of eucaryotic cells are not functionally continuous. In many cases they consist of relatively short (20–300 bp) coding regions (exons) interrupted by much longer (200–10,000 bp) noncoding regions (introns). Procaryotic genes are continuous. Gilbert originally hypothesized that this eucaryotic exon-intron arrangement reflects a basic, modular design principle in evolution.

The present version of Gilbert's argument runs along the following lines: Since introns vary greatly in size, they must have a sequence-independent role in the genome. Gilbert suggested this role is to facilitate recombination between exons, thus placing the exons within the same transcription unit. Different splicing patterns to remove intron RNA from the final message could then be tried through evolution without disrupting the integrity of the original genes. The probability of recombination between exons increases with the square of the intervening sequence's length. Thus recombinations between introns are orders of magnitude more likely than the precise events required to link the 5′ end of one exon to the 3′ end of another. One consequence of this view is that "exon shuffling" should produce classes of proteins which share some of the same exons, Gilbert noted. Indeed this is precisely what is seen in a number of multigene families. In the case of immunoglobins, this conservation is striking. All the globin genes have three exons separated by two introns located at constant positions relative to the coding sequences. Further, Gilbert reported the recent finding that three otherwise unrelated genes—those coding for the epidermal growth factor (EGF) precursor, Factor IX, and the low-density lipoprotein (LDL) receptor—all share a common exon domain.

These facts about intron-exon relationships lead to speculation on whether the earliest genes contained introns. In order to approach this question, Gilbert has been investigating the structure of triose phosphate isomerase (TIM) genes from divergent species. TIM is a ubiquitous catabolic enzyme that must have arisen very early in the evolution of cells. It has also been extensively studied by protein chemists, and its three-dimensional structure is known from crystallography. Two questions can be unambiguously answered by examining this gene. Does it contain introns? If it does, do their positions reflect properties of the protein structure?

Gilbert reported that the TIM gene from chicken consists of seven exons separated by six introns, and that the exons closely correspond to structural domains of the enzyme. That is to say, the exons do in fact represent modular units which are closely packed in the three-dimensional structure of the protein. He finds it easier to imagine a noninterrupted bacterial gene arising by the loss of introns than to imagine the converse—the precise insertions into a continuous ancestral gene necessary to generate the eucaryotic version.

To emphasize this point, Gilbert cited very recent studies carried out in his laboratory at Biogen by Mark Marchionni. Marchionni has cloned the TIM gene from corn. Comparison of the structures of the maize and chick genes reveals striking structural similarities (see illustration above). Three of the five maize introns correspond exactly to intron positions in the chick gene, and a fourth (chick intron 5 in the diagram) is shifted by an integral number of codons. Since animal and plant cells are thought to have diverged approximately one billion years ago, the interrupted version of TIM is at least that old. In evolutionary terms, this leaves only a short period prior to the divergence in which the introns would have to be inserted into an original, noninterrupted sequence to produce the ancestor of both the plant and animal genes. Gilbert therefore argues that the ancestral TIM contained introns, which were progressively lost as bacteria evolved. **—Harvey Bialy**

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

***Barrel Building From Exons.*** **Depicted is a schematic representation of triose phosphate isomerase (TIM) in which secondary structural elements are shaded according to the positions of introns interrupting the coding sequences. The lower portion is a linear representation showing how the introns divide the protein. Maize introns are given above the line, chicken positions below the line. Beta sheets are shown as arrows, and helices as rectangles.**