which greatly exceeds the escape velocity from Jupiter. Sodium is present only at the 1% trace level; however, it provides a compelling model for atomic processes that can be inferred to exist also for the dominant species oxygen and sulfur which, although more abundant by a factor of 100 than sodium, are not as readily detectable. Observations of the neutral sodium cloud thus give information on temporal changes in the Io torus as described by Mendillo *et al.* One of the more significant conclusions that can be drawn from the study is the source rate for ions that are created out of the magnetonebula (primarily by solar photoionization) and form a population of hot ions in the outer magnetosphere after acceleration to local corotation speed by the planet's magnetic field (*2*). This process is thought to be a principal source of the hot heavy ion (S^+ and O^+) plasma that the SWICS and HI-SCALE instruments on Ulysses were able to measure directly. In all it would seem that Ulysses was very successful in returning a valuable data set that illuminates the workings of Jupiter's magnetosphere in great detail.

REFERENCES

1. Special issue on the Ulysses encounter with Jupiter, *Science* **257**, 1503–1557 (1992).
2. M. Kane, B. H. Mauk, E. P. Keath, S. M. Krimigis, *Geophys. Res. Lett.* **19**, 1435 (1992).

Exons as Microgenes?

H. Martin Seidel,* David L. Pompliano,† Jeremy R. Knowles‡

Most eukaryotic genes consist of coding sequences (exons) interrupted by noncoding sequences (introns). After transcription into RNA, the introns are removed by splicing to generate the mature messenger RNA that carries a continuous coding sequence. The splice points are marked by consensus sequences that act as signals for the splicing process. Although much is known about the role of these consensus sequences, there has been little speculation about their evolutionary origin (*1*). Some recent findings now encourage musing on these matters.

The enzyme phosphoenolpyruvate mutase is an unusual enzyme that is responsible for catalyzing the formation of the carbon-phosphorus bonds of phosphonate metabolites. The few organisms that are known to make phosphonates are spread across the evolutionary spectrum, and this biosynthetic activity is believed to be an ancient one (*2*). In the sequence of the messenger RNA for the mutase from the protozoan *Tetrahymena*, there are two in-frame amber codons (*3*). In most organisms, amber (TAG) is one of the three "stop" signals in the genetic code, but *Tetrahymena* reads amber (and ochre) codons as glutamine at these loci (*4*).

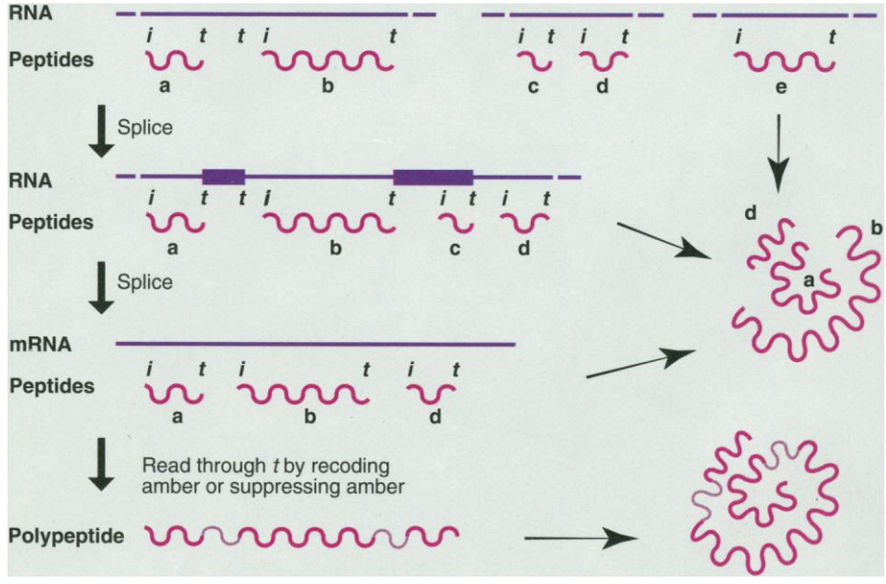
In the mutase gene (sequenced from the genome of *Tetrahymena*) there are three introns, two of which start precisely after the amber codons (TAG) and one that begins after a lysine codon (AAG). Moreover, two of the introns also terminate with amber. The base sequences around the three pairs of junctions are

```
   exon        intron        exon
--ACTTAGgtactt......aaatagGTTTTG--
--TTCTAGgtaagc......caatagGTCGTC--
--ATGAAGgtaaag......ataaagGAATGG--
```

In all, 12 of the 17 known *Tetrahymena* introns end with TAG, 4 with AAG, and 1 with CAG. These junctions all conform to the global consensus sequences for exon-intron junctions (of ..AG|gta..) and intron-exon junctions of (..ag|G..) found in all eukaryotes (*5*). These consensus splice-junction sequences do not occur in the non–protein-encoding ribosomal RNA and transfer RNA genes, which have quite different splicing mechanisms.

What is the origin of the splice-site consensus sequences in protein-encoding genes? The trivial explanation is that the sequence consensus is simply a historical accident and carries no information about its evolutionary origin. Given the existence of ancient RNA splicing mechanisms, however, the evolution of a splicing consensus sequence peculiar to protein-encoding genes seems unlikely to have been accidental and more likely reflects an early functional role of these sequences.

The location of the introns in the *Tetrahymena* mutase gene both explains the origin of the 3′-termini of eukaryotic exons and introns and suggests the possibility that exons were once "microgenes," originally terminating with amber and encoding relatively short oligopeptides that assembled spontaneously into active protein. The argument runs as follows: from various segments of primordial RNA, initiation of protein synthesis (at *i*) and termination (at *t*) would produce a library of oligopeptides (**a, b, c**, and so forth), some combinations of which would spontaneously combine to form multichain protein assemblies having catalytic activity (see figure). Such assemblies, in which protein fragments come together to generate a catalytically active unit, are well known (*6*). Splicing of these microgenes by using their common amber termini (*t*) as a recognition element for splicing (*7*) could then bring the appropriate microgenes closer together, thereby improving the chances for the linked inheritance of all the fragments that contribute to a particular catalytic activity. This notion is



The authors are in the Departments of Chemistry and Biochemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138.

*Present address: Ligand Pharm Inc., 9393 Town Center Drive, San Diego, CA 92121.
†Present address: Merck, Sharp & Dohme Research Laboratories, Department of Cancer Research, West Point, PA 19486.
‡To whom correspondence should be addressed.

also consistent with the view (8) that today's genes evolved by the shuffling of exons, a process facilitated by the presence of intervening segments of genetic material. Excision of the introns (heavy bars in the figure) then leads to the coding sequence.

Read-through of the termination codons, rather than translational termination at the end of each microgene, would then (for the microgenes that were in frame) produce a continuous polypeptide stretch that would fold into a protein of greater thermal stability. Read-through would, of course, introduce extra peptide sequences [encoded by the intervening RNA between one termination (*t*) and the next initiation (*i*)] (see figure). This suggestion is attractive because exon-intron junctions map predominantly at the protein surface (*9*, *10*) where the charges at each end of the peptide fragments could have more stably resided. Surface locations can also best accommodate the problem of how to connect the ends of exon products (the original amino and carboxyl termini of the microgene products) that, in the first-formed assembly, happen not to be close to one another (see figure). If the juxtaposition of a pair of microgenes were to result in the exons being out of frame, the intervening sequence would be retained by selection, leaving later intron excision from the RNA transcript as the route to a functional protein. Subsequent migration of the splice sites would then refine and trim the early structure, and result in the distribution of intron phase classes observed today (*11*).

There are only two ways that read-through can occur: either the termination codon *t* must be changed to a coding triplet or the meaning of the terminator must be suppressed. The consensus sequence at the end of eukaryotic exons (... AG| ...) may simply be a consequence of the first way of reading through. If the first base in the terminating amber triplet TAG was changed to C, A, or G, translation into protein could continue from the spliced message. In *Tetrahymena*, we see the second solution to the read-through problem. The terminal codon TAG was retained, but its meaning was changed from amber to sense [in this case, to glutamine (*4*)]. Given the prior existence of the three "universal" stop codons, the ease of this type of suppression has been persuasively argued (*12*). We therefore suggest that exon microgenes originally terminated in TAG and that this feature was used to recognize the ends for splicing. Pressure for read-through then led directly to the two ways of reading through outlined above, while the need to maintain a splicing signal kept ... AG| ... at the 3′ ends of both exons and introns. The existence today of TAG termini in some exons and many introns from *Tetrahymena* would thus be the consequence of a preference for ... TAG| ... for splicing recognition (in addition to the absence of any need in *Tetrahymena* to select against T as the first base of the last triplet), while most eukaryotes select more simply for ... AG|.. splice sites (and respond to pressure against a T in front of in-frame splice site AG termini).

Many of today's introns may simply be failed exons: they were once microgenes (then, as now, ending in ...AG|...) originally encoding polypeptides that failed to contribute to any functional protein assembly. Only the useful microgenes have become what we now call exons. In this view, there is no imperative that exons have to encode polypeptide domains that fold independently (*13*), for rudimentary proteins could as well be assembled by the stable complementation of exon products that alone have no three-dimensional structure (*6*, *14*). Although many questions are not addressed by the figure (such as why amber codons should have dominated the splicing process, whether all the steps occurred in the RNA world, and whether splice sites were originally all in frame), our proposal provides a link between what is known of the structure of eukaryotic genes and of the proteins that they encode.

REFERENCES AND NOTES

1. P. Senapathy, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1129 (1988).
2. T. Hori, M. Horiguchi, A. Hayashi, *Biochemistry of Natural C-P Compounds* (Maruzen, Tokyo, 1984).
3. H. M. Seidel, D. L. Pompliano, J. R. Knowles, *Biochemistry* **31**, 2598 (1992).
4. D. W. Martindale, *J. Protozool.* **36**, 29 (1989); S. Osawa, A. Muto, T. H. Jukes, T. Ohama, *Proc. R. Soc. Lond. B* **241**, 19 (1990).
5. M. B. Shapiro and P. Senapathy, *Nucleic Acids Res.* **15**, 7155 (1987).
6. C. B. Anfinsen, P. Cuatrecasas, H. Taniuchi, in *The Enzymes, ed. 3* **4**, 177 (1971); F. M. Richards and H. W. Wyckoff, in *ibid.*, p. 647; I. Saint Girons *et al.*, *J. Biol. Chem.* **262**, 622 (1987).
7. T. R. Cech, *Annu. Rev. Biochem.* **59**, 543 (1990).
8. W. Gilbert, *ibid.* **271**, 501 (1978); *Science* **228**, 832 (1985).
9. C. S. Craik, S. Sprang, R. Fletterick, W. J. Rutter, *Nature* **299**, 180 (1982).
10. Y. Naito, C. K. Riggs, T. L. Vandergon, A. F. Riggs, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6672 (1991).
11. P. A. Sharp, *Cell* **23**, 643 (1981).
12. N. Hanyu, S. Nishimura, Y. Kuchino, *EMBO J.* **5**, 1307 (1986).
13. M. Go, *Nature* **291**, 90 (1981); _____ and M. Nosaka, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 915 (1987).
14. L. E. Orgel, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 9 (1987); C. H. Li and T. A. Bewley, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1476 (1976); N. G. Galakatos and C. T. Walsh, *Biochemistry* **26**, 8475 (1987).

15. We thank R. L. Dorit, W. Gilbert, and M. Meselson for clarifying discussions.