Implications of RNA·RNA Splicing in Evolution of Eukaryotic Cells

James E. Darnell, Jr.

Summary. The differences in the biochemistry of messenger RNA formation in eukaryotes compared to prokaryotes are so profound as to suggest that sequential prokaryotic to eukaryotic cell evolution seems unlikely. The recently discovered noncontiguous sequences in eukaryotic DNA that encode messenger RNA may reflect an ancient, rather than a new, distribution of information in DNA and that eukaryotes evolved independently of prokaryotes.

For some years evidence has been accumulating that messenger RNA (mRNA) formation in eukaryotic cells is substantially different and more biochemically complex than in bacteria (*1–4*). At the 5′ terminus most eukaryotic mRNA's from yeast to man contain a modified methylated structure called a "cap" (*2*). In addition, there is an average of one 6-methyladenylic acid per 500 to 1000 bases in the mRNA of higher cells. A segment of polyadenylic acid is added after transcription to the 3′ end of possibly all, certainly most, eukaryotic mRNA's except histone mRNA (*1, 4, 5*). Finally, the most startling posttranscriptional event in mRNA formation is suggested from discoveries in the past year. It appears that primary nuclear RNA transcripts that have been known for many years to be larger than mRNA [(*6, 7*), see (*1, 3*) for review] not only may be cleaved, but selected pieces may be "spliced" back together to make the mRNA molecule. This very unexpected conclusion came first from work on adenovirus mRNA's. Late in adenovirus type 2 (Ad2) infection a series of at least 13 individual mRNA molecules (*8, 9*) were found to contain sequences from noncontiguous sites on the adenovirus genome (*10, 11*). That each of these "mosaic" mRNA molecules comes about by RNA·RNA "splicing" or "ligation" was inferred from studies on the synthesis of Ad2 specific RNA in the nucleus of the infected cells (*12, 13*). The only primary RNA transcript that can be detected from the regions of the spliced late mRNA's embraces all the spliced regions as well as all the major mRNA regions. Moreover, it appears that approximately 20 percent of each long primary transcript is conserved in the formation of one of five separate 3′ coterminal groups of mRNA molecules, suggesting that each long transcript produces one mRNA molecule (*9, 14*). Breakage and reunion of pieces of the long primary transcript would seem essential to form the "spliced" mRNA. Transcribing a folded template in the "correct" places or recombining the DNA prior to transcription could theoretically produce shorter primary transcripts containing only the mRNA sequences. However, such short primary transcripts are not detected (*13*).

In addition to the initial reports of presumed splicing in late Ad2 mRNA formation, a large amount of evidence has accumulated that many other virus mRNA molecules (*14, 15*) as well as hemoglobin mRNA (*16, 17*), immunoglobulin mRNA (*18–20*), ovalbumin mRNA (*21–23*), and other cellular mRNA's (*24, 25*) are probably derived from larger primary transcripts and are composed of sequences that are not contiguous in the DNA from which the mRNA is transcribed. Splicing of RNA in cell-free extracts has been reported in two cases. Yeast transfer RNA (tRNA) precursors contain a 15-base intervening sequence that is not present in mature tRNA (*26, 27*). Extracts of yeast cells have been shown capable of removing these extra nucleotides and rejoining the remaining tRNA segments (*28*). A second case of in vitro RNA·RNA splicing involves formation of an mRNA (*29*). Nuclei from adenovirus-infected cells labeled for 10 minutes early in infection contain a single labeled polyadenylate [poly(A)]-terminated high-molecular-weight RNA complementary to one of the several regions of early transcription. This nuclear RNA species contains the sequences of one of the early Ad2 mRNA's as well as intervening sequences that do not exist in the mRNA. Incubation of the labeled nuclei together with cell extracts results in the conversion of about half of this labeled precursor RNA into poly(A)-terminated RNA of the same size and "spliced" sequence composition as authentic mRNA. Thus, it seems highly likely that RNA·RNA splicing is truly the mechanism for bringing "mosaic" mRNA's together. Perhaps in every case where a larger nuclear RNA containing mRNA sequences can be demonstrated, splicing will be involved in producing the final mRNA product. None of the posttranscriptional events in mRNA formation—capping, posttranscriptional poly(A) addition, or splicing—has been reported in bacteria.

During the past few years as the differences in the mechanics of gene expression between prokaryotes and eukaryotes have been revealed, it has seemed more and more difficult to envision a stepwise cellular evolution connecting prokaryotes and eukaryotes. This article explores the idea that the complex of biochemical reactions that result in mRNA formation is the chief evolutionary basis that sets eukaryotes apart from prokaryotes. Further, the key evolutionary step is the ability of eukaryotes to utilize noncontiguous information in DNA.

Acceptance of the hypothesis that, in eukaryotes, "spliced" mRNA molecules are frequently formed from noncontiguous sequences raises several interrelated questions: Why, when, and how in evolution did the divided genes arise? What function is served today by having genes remain divided?

When Did Divided Genes Arise?

Two very different views might be presented for the origin of the divided genes. (i) The first genes that arose were composed of contiguous sequences and functioning cells employing such genes

The author is a professor of molecular cell biology at the Rockefeller University, New York 10021.

0036-8075/78/1222-1257$01.00/0

evolved. As eukaryotic cells evolved from a prokaryotic cell precursor the need for more sophisticated controls was satisfied by first the development of RNA·RNA splicing ability with subsequent insertions into the structural genes, possibly frequently into regions where transcription of genes is initiated. These insertions might have been similar to the bacterial insertion sites that are known in several instances to affect transcriptional regulation (*30*). An attractive feature of this proposal is that the segments of DNA in or around inserted elements might be possibly transposed to change the resulting expression of the involved genes as occurs, for example, with *Salmonella* serotypes (*31, 32*). Such "transposons" have been suggested as the mechanism underlying "controlling elements" in *Zea mays* (maize) (*30*). The union of the variable and constant regions of immunoglobulin genes is also seen as an example of DNA rearrangement to promote proper transcription and ultimate gene expression (*33*). (ii) A contrasting view for the origin of divided genes is that many (perhaps most) genes of eukaryotic cells are composed of DNA that was never contiguous and therefore has not "become divided." This view denies an intermediate role in evolution to a cell organized like present-day prokaryotes. According to this view, the separated DNA segments that represent a eukaryotic gene were "recruited" (albeit by chance) very early in evolutionary time within the same transcriptional unit and became established as a functioning gene because RNA·RNA ligation allowed the use of all the protein coding information within the primary RNA transcript. In this scheme, where RNA·RNA splicing is held as basic to the origin and design of the eukaryotic genome, it might be the unusual eukaryotic gene in which splicing at the RNA level is not necessary for expression; DNA rearrangement of the type that occurs in immunoglobulin production would be unusual but, when it does occur, developmental consequences would be profound (*33*).

Proving the order of evolutionary development of present-day cell functions is clearly an impossible task. However, there are several reasons that might seem to favor the view that eukaryotic development was not the end of a sequence passing through cells designed like present-day prokaryotes.

Both prokaryote and eukaryotes existed at least 1.0 to 1.5 $\times$ 10^9 years ago (*34, 35*), and studies to date provide no evidence of sequential prokaryotic to eukaryotic evolution (*35–37*). For example, although the genetic code is universal (or nearly so) and the machinery for protein synthesis is quite similar in prokaryotes and eukaryotes, the tRNA molecules for specific amino acids—even including initiator tRNA (*37, 38*), and ribosomal RNA's (rRNA) (*35, 36, 39*)—bear little resemblance even between lower eukaryotes and prokaryotes while there is considerable sequence overlap between various eukaryotic tRNA and rRNA molecules. Furthermore, even in yeasts, which are among the least complex eukaryotic organisms, some tRNA's are formed from a precursor tRNA by the removal of about 15 to 20 nucleotides from the middle of the tRNA sequence (*26–28*), with subsequent RNA·RNA splicing. Likewise, there is little evidence of any overlap between prokaryotes and eukaryotes of primary amino acid sequences even for similar proteins although, it must be admitted, there has been very little work done on which such comparisons can be made (*37, 40, 41*). Thus, there is at present no evidence of a "core" or residue of prokaryotic genes that are still present within a now expanded set of eukaryotic genes.

Other points might also seem to favor the separate origin of eukaryotes rather than insertions into prokaryotic genomes to make eukaryotes. With the exception of histone mRNA, virtually every individual mRNA from vertebrates or from DNA viruses infecting vertebrates that has been examined so far derives from noncontiguous regions of the DNA. In addition almost all mRNA's in cultured cells appear to derive from nuclear RNA precursor molecules longer than the mRNA itself (*25*). Thus, almost every gene would have had to have one or more inserts (as many as seven in the case of ovalbumin) (*21, 23*). Bacterial insertion sequences are characterized by movement in and out of the genome (*30*). In contrast, the intervening sequences in some of the vertebrate structural genes that have been analyzed would seem to be stable. For example, both the DNA in specialized cells that make rabbit hemoglobin (*16*) and chick ovalbumin (*21*), as well as the DNA of a variety of other nonspecific tissues not making hemoglobin or ovalbumin, contained the same array of noncontiguous DNA sequences complementary to each of the mRNA's. In mice, an apparently duplicated β-globin gene exists and both β major and β minor genes have very similar configurations for the protein coding and intervening sequences (*42*). Therefore, there is no experimental support in these cases for DNA rearrangement for the purpose of transcription or any other purpose. Even in the case of immunoglobulin formation, where rearrangement does occur during the differentiation of the lymphocyte, the variable (V) and constant (C) regions of the light-chain gene remain 1250 nucleotides apart in a myeloma cell producing the light chain (*18, 19*).

Strong evidence that the arrangement of noncontiguous segments of present-day genes occurred long ago and evolved to become fixed and stable comes from a comparison of the arrangement of sequences for the β chains of the hemoglobins in two different species. Both the length and position of the intervening sequences between the two segments of DNA represented in the mRNA is about the same in rabbit and mouse DNA (*16, 17*). The major intervening sequences are estimated to be 550 to 600 nucleotides in both cases and occur after sequences encoding amino acid 104 in the mouse globin gene and somewhere between amino acids 100 and 120 in the rabbit gene. Thus, the event that led to the formation of this "split" gene occurred 5 $\times$ 10^7 to 10^8 years ago, and the DNA appears to have been quite stable since then. If a similar, common configuration is found to be the rule for the "pieces" of a variety of homologous genes in a variety of species, then it would seem even more likely that the parts of the "split" genes came together a very long time ago and remained relatively unaltered. It should be added that no changes in comparative gene structure between organisms would be most surprising, but the point is that lack of frequent changes within the developmental cycle of an organism speaks against a regular programmed "use" of intervening sequences.

A final argument that might favor the possibility of independent evolution for the transcriptional units in present-day prokaryotes and eukaryotes has been recognized for many years. Even in lower eukaryotes where biosynthetic pathways exist that are equivalent to those in bacteria (synthesis of a complex amino acid such as histidine, for example), the genes for the related enzymes of a pathway are not linked in "operon" fashion but are scattered on different chromosomes (*43, 44*). If eukaryotes developed from prokaryotes, it is hard to see an evolutionary advantage to eukaryotes in jettisoning such an efficient mechanism for the coordinate regulation of a whole metabolic pathway.

Why Did Divided Genes Aid Evolution?

The development of RNA·RNA splicing from primary transcripts of divided

genes during mRNA formation could have facilitated evolution in at least two times and in two ways: (i) very early, when a functioning, integrated genome that was capable of directing the regulated growth of a cell was first evolved or (ii) during cellular evolution when wider cellular capacities, eventually leading to metazoan life, were developed (*19*).

"Recruitment" of DNA segments into transcription units early in evolution. After a biological energy-generating system had evolved and an apparatus capable of programming and carrying out polypeptide synthesis had developed [(the "progenote" state in the terminology of Woese and Fox (*35*)], the primary thrust of "precellular" evolution was, presumably, improvement in, diversification of, and regulation of polypeptide synthesis and function, and the construction of a genome responsive to regulation that harbored information for these polypeptides. When a polypeptide that could perform a specific function first arose, association with other randomly evolved peptides might have resulted in enhancement, extension, or regulation of that function. For example, a polypeptide with a catalytic site for phosphatase activity might have encountered a complementing peptide that would cause the phosphatase activity to function best at alkaline or acid *p*H, or aerobically or anaerobically, providing for restriction or extension of conditions under which the activity was expressed. Or, a complementing peptide might have caused the phosphatase to localize in a particular position within the evolving cell structure, allowing for the development of site-specific phosphatase activity.

The possible structural residues of such functional "domains," the idea for which originated with studies on immunoglobulins (*45*), have been discussed for a number of proteins (*46, 47*). The "pocket" in which heme is bound in cytochrome b5 and hemoglobin has structural and some distant amino acid similarity. The nicotinamide adenine dinucleotide (NAD)-binding sites (*46*) in lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase are similar (*47*); nucleotide-binding proteins (*48*) and Ca^{2+}-binding proteins also have structural similarities (*49*). The general argument has been made (*40, 48, 50*) that these overall structural similarities in proteins with similar binding capacities may represent a more reliable phylogenetic guide than amino acid sequence. Changes in sequences without change in shape might obscure the similar origins of present-day structural similarities. If "precellular" evolution proceeded by the random development of such complementing peptides that might serve as function-directing "domains" within proteins, a molecular event that brought the information for these "domains" together into one genetic unit whose expression could be coordinately regulated would have been a powerful aid to the evolution of a functioning genome.

The incorporation during evolution of information for such complementing "domains" into one chromosomal gene in a genome organized like present-day bacteria could be achieved only by a tedious series of point mutations or by precise recombination of the DNA specifying the two interacting polypeptides, such that a continuous mRNA representing both polypeptides would be synthesized. The evolution of a functioning "divided" gene containing information for both peptides could have occurred only if RNA·RNA splicing already existed. Given the existence of RNA·RNA splicing, a random DNA recombinational event that joined the information for two independently evolved complementing polypeptides ("domains") would not have been required to be precise but only to bring the two DNA's into the same transcriptional unit. Even occasionally successful RNA·RNA ligation would result in the formation of a functional mRNA. It is recognized that eventually a precise mechanism for RNA·RNA ligation of the newly recruited RNA segments that encoded amino acid sequence would have been required. But, prior to this stage of evolution, if some of the RNA·RNA ligations were successful, than a marginally competent mRNA could have been formed from divided genes as a very early evolutionary step. Gradually in evolution the process of making the mRNA more precisely could have been perfected. Possibly all the modifications of primary transcripts such as methylation and polyadenylation are later evolutionary adaptations that ensure proper processing.

More efficient recombination. The second contribution to cellular evolution that splicing might have made is to facilitate the production of new genes through recombination. As has been pointed out by others (*19, 51*), recombinations of functional parts (domains) of a split gene would allow cells to readily try out various combinations because of recombination of the pieces into one transcription unit. An occasional correct splicing event would satisfy the demands of selective pressure until the new gene was established. Moreover, mutations that promoted variable splicing patterns would have increased variety without necessarily destroying the original product (*51*). Theorists have puzzled over the speed and diversity of evolution through conventional molecular mechansims proceeding from prokaryotes to eukaryotes (*37, 52*); perhaps the answer lies in a separate eukaryotic evolution featuring the "en bloc" recombination of functions between transcription units whose primary transcripts contain several regions that can be ligated and may represent "domains" of proteins.

Maintaining the Presence of Intervening Sequences Today

One of the persistent puzzles, especially to molecular geneticists who have studied the behavior of "excess" DNA segments in bacteria, is why the intervening sequences that are not expressed in protein but are nevertheless quite long are maintained in the eukaryotic genome. First, it should be pointed out that what is an intervening sequence for one mRNA may be included at another time in another mRNA. This is the situation for a great many of the virus mRNA's (*8–11*). Moreover, at this stage in our thinking about intervening sequences it should be recognized that all the interruptions may not exist for the same reason. For instance, those at the 5′ ends in particular might have more to do with regulating transcription or translation (or both) of the eventual mRNA products than do those within the coding regions of a gene that can only produce one protein. Thus, we may come to recognize classes of intervening sequences or, rather, classes of retained regions within an mRNA when more structural and sequence data become available. Nevertheless, it does seem likely that some intervening sequences that interrupt protein coding regions, for example, in ovalbumin and hemoglobin genes, may never be used as part of an mRNA; the question then arises why these regions, which can be thousands of nucleotides long, have not been deleted by the pressures of evolution.

In *Escherichia coli* even sequences with a function, identical tRNA genes for example, can be deleted in the absence of evolutionary pressures for retention (*53*) but the intervening sequences in hemoglobulin, for example, have been maintained for at least 10^8 years. Perhaps many eukaryotic cells simply do not have an equivalent facility for deleting "excess" sequences so that nonfunctional "intervening" sequences would be maintained for a longer time in

eukaryotes. Doolittle (*54*) suggests that the capacity to "streamline" a genome by discarding DNA sequences that are not demanded by environmental stress is a late development in evolution. Present-day bacteria, which so readily excise sequences, would be viewed as evolutionarily highly advanced from original cells. Furthermore, present-day eukaryotes, such as yeasts, that are highly adapted for rapid growth and have perhaps fewer and smaller intervening sequences (they at least have no large heterogeneous RNA molecules) also have acquired the capacity to reassort DNA segments and discard "extra" segments more efficiently than the cells of metazoans. However, during its early evolutionary development, a genome that was built by recruitment of ligatable regions might have been successful only if its system for deleting unnecessary DNA was inefficient.

Conclusion

The discovery of "mosaic" mRNA's in eukaryotic cells (*10, 11, 15-23, 26, 27*) and the conclusion that posttranscriptional RNA·RNA ligation is responsible for the biogenesis of these mRNA's, has profound implications in the study of both evolutionary and regulatory biology. If it is true that sequential prokaryotic to eukaryotic evolution did not occur, then it seems not only possible but logical that the basic rules of genome organization might also differ between present-day prokaryotes and eukaryotes. If the molecular basis of eukaryotic gene regulation is to be explained in relation to developmental biology or cancer biology or endocrinology or many other topics, it is at least possible that we cannot rely on bacterial models but must again solve the molecular control mechanisms of eukaryotic genes.

References and Notes

1. J. E. Darnell, W. Jelinek, G. Molloy, *Science* **181**, 1215 (1973).
2. A. J. Shatkin, *Cell* **9**, 645 (1976).
3. J. E. Darnell, *Prog. Nucleic Acid Res. Mol. Biol.* **22**, 327 (1978).
4. G. Brawerman, *ibid.* **17**, 118 (1975).
5. Y. Kaufman, C. Milcarek, H. Berissi, S. Penman, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4801 (1977); R. E. Gray and A. R. Cashmore, *J. Mol. Biol.* **108**, 595 (1976).
6. K. Scherrer and L. Marcaud, *J. Cell. Physiol.* **72** (Suppl. 1), 181 (1968); G. Georgiev and V. L. Mantieva, *Biochim. Biophys. Acta* **61**, 153 (1962).
7. K. Scherrer and J. E. Darnell, *Biochem. Biophys. Res. Commun.* **7**, 486 (1962); J. E. Darnell, *Bacteriol. Rev.* **32**, 262 (1968).
8. J. Flint, *Cell* **10**, 153 (1977); L. T. Chow, J. M. Roberts, J. B. Lewis, T. R. Broker, *ibid* **11**, 819 (1977).
9. J. R. Nevins and J. E. Darnell, *J. Virol.* **25**, 811 (1978).
10. S. N. Berget, C. Moore, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171 (1977).
11. L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts, *Cell* **12**, 1 (1977); D. F. Klessig, *ibid.*, p. 9; R. E. Gelinas and R. J. Roberts, *ibid.* **11**, 533 (1977).
12. J. Weber, W. Jelinek, J. E. Darnell, *ibid.* **10**, 612 (1977); S. Goldberg, J. Weber, J. E. Darnell, *ibid.*, p. 617.
13. R. Evans, N. W. Fraser, E. Ziff, J. Weber, M. Wilson, J. E. Darnell, *ibid.* **12**, 133 (1977); S. Goldberg, J. R. Nevins, J. E. Darnell, *J. Virol.* **25**, 806 (1978).
14. J. R. Nevins and J. E. Darnell, *Cell* **15**, 1477 (1978).
15. G. R. Kitchingman, S. P. Lai, H. Westphal, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4392 (1977); A. Berk and P. A. Sharp, *Cell* **14**, 659 (1978); Y. Aloni, O. Dhar, M. Laub, M. Horowitz, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3686 (1977); M-T. Hsu and J. Ford, *ibid.*, p. 4982.; E. Rothenberg, D. J. Donoghue, D. Baltimore, *Cell* **13**, 453 (1978); A. Berk and P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1274 (1978); L. V. Crawford, C. N. Colec, A. E. Smith, E. Paucha, P. Tegtmeyer, K. Rundell, P. Berg, *ibid.*, p. 117.
16. A. J. Jeffreys and R. A. Flavell, *Cell* **12**, 1097 (1977).
17. S. Tilghman, D. C. Tiemier, J. G. Seidman, B. M. Peterlin, J. Sullivan, J. V. Maizel, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 125 (1978).
18. C. Brack and J. Tonegawa, *ibid.* **74**, 5652 (1977).
19. S. Tonegawa, A. M. Maxam, R. Tizard, O. Bernhard, W. Gilbert, *ibid.* **75**, 1485 (1978).
20. M. Gilmore-Hebert and R. Wall, *ibid.*, p. 342.
21. R. Breathnach, J. L. Mandel, P. Chambon, *Nature (London)* **270**, 314 (1977).
22. R. Weinstock, R. Sweet, M. Weiss, H. Cedar, R. Axel, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1299 (1978).
23. E. C. Lai, S. L. C. Woo, A. Dugaiczyk, J. F. Catterall, B. W. O'Malley, *ibid.*, in press.
24. R. Giorno and W. Sauerbier, *ibid.* **75**, 4374 (1978).
25. S. Goldberg, H. Schwartz, J. E. Darnell, *ibid.* **74**, 4520 (1977).
26. H. M. Goodman, M. V. Olson, B. D. Hall, *ibid.* p. 5453.
27. P. Valenzuela, A. Venegas, F. Weinberg, R. Bishop, W. J. Rutter, *ibid.* **75**, 190 (1978).
28. G. Knapp, J. S. Beckmann, P. F. Johnson, S. A. Fuhrman, J. Abelson, *Cell* **14**, 221 (1978).
29. J-M. Blanchard, J. Weber, W. Jelinek, J. E. Darnell, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
30. P. Nevers and H. Saedler, *Nature (London)* **268**, 109 (1977).
31. T. Ino, T. Oguichi, T. Hirand, *J. Gen. Microbiol.* **89**, 265 (1975).
32. J. Zeig, M. Silverman, M. Silmen, M. Simon, *Science* **196**, 170 (1977).
33. L. Hood, H. V. Huang, W. J. Dreyer, *Supramol. Struct.*, in press.
34. R. M. Schwartz and M. O. Dayhoff, *Science* **199**, 395 (1977); J. W. Schopf and D. Z. Oehler, *ibid.* **193**, 47 (1976).
35. C. R. Woese and G. E. Fox, *J. Mol. Evol.* **10**, 1 (1977).
36. ———, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977).
37. M. O. Dayoff, W. C. Barker, J. McLaughlin, *Origins Life* **5**, 311 (1974).
38. M. Simsek, V. L. RajBhandary, M. Boisnard, G. Petrissant, *Nature (London)* **247**, 518 (1974).
39. D. D. Brown, C. S. Weber, J. A. Sinclair, *Carnegie Inst. Washington Yearb.* (1967), pp. 8–17.
40. M. O. Dayhoff, *Atlas of Protein Sequences and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, Suppl. 2.
41. G. D. Fasman, Ed., *Handbook of Biochemistry and Molecular Biology*, vol. 1, *Proteins* (CRC Press, Cleveland, ed. 3, 1975).
42. D. C. Tiemeier, S. M. Tilghman, F. I. Polsky, J. G. Seidman, A. Leder, M. H. Edgell, P. Leder, *Cell* **14**, 237 (1978).
43. G. Dorn, in *Handbook of Biochemistry and Molecular Biology*, G. D. Fasman, Ed. (CRC Press, Cleveland, 1976), p. 792.
44. G. Pontecorvo, *Prov. R. Soc. Biol. London Ser. B* **158**, 1 (1973).
45. B. A. Cunningham, P. D. Gottlieb, M. N. Pflumm, G. M. Edelman, in *Progress in Immunology*, B. Amos, Ed. (Academic Press, New York, 1971), pp. 3–24.
46. M. G. Rossman and P. Argos, *J. Mol. Biol.* **105**, 75 (1976).
47. ———, *J. Biol. Chem.* **250**, 7525 (1975).
48. M. G. Rossman, D. Moras, K. W. Olsen, *Nature (London)* **250**, 194 (1974).
49. R. H. Kretsinger and C. E. Nuckolds, *J. Biol. Chem.* **248**, 3313 (1973).
50. G. E. Schulz, *J. Mol. Evol.* **9**, 339 (1977).
51. W. Gilbert, *Nature (London)* **271**, 501 (1978).
52. T. Dobzhansky *et al.*, *Evolution* (Freeman, San Francisco, 1977).
53. W. R. Folk and P. Berg, *J. Mol. Biol.* **58**, 595 (1971); P. W. J. Rigby, B. D. Burleigh, Jr., B. S. Hartley, *Nature (London)* **251**, 200 (1974).
54. While this manuscript was being prepared, a note appeared in *Nature (London)* [**272**, 581 (1978)] by W. F. Doolittle which proposes the same general premise as this article—that splicing and "genes in pieces" is representative of an early state of cell evolution. Doolittle then goes on to advocate the idea that bacteria become "streamlined" and, in so doing, lost the flexibility afforded by splicing and that eukaryotes successfully specialized by retaining this very early arrangement of genetic material. During the review of this manuscript a second note [C. C. F. Blake, *Nature (London)* **273**, 267 (1978)] appeared welcoming the proposal (*19, 51*) that the disparate portions of a primary transcript that are conserved as mRNA might represent "domains" of folded proteins.

55. Supported by grants from NIH (CA 16006-05) and the American Cancer Society (VC-295H). I thank Eva Derman for a number of helpful discussions of ideas contained in the manuscript.