Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1271–1275, March 1986
Biochemistry

Speculations on the early course of evolution

(RNA processing/cell lineages)

J. E. DARNELL* AND W. F. DOOLITTLE†

*Molecular Cell Biology Department, Rockefeller University, New York, NY 10021; and †Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

Contributed by J. E. Darnell, Jr., October 23, 1985

ABSTRACT **The proposal that RNA preceded DNA in evolution is more than 15 years old. In light of recent studies on RNA processing (including protein-free reactions), present knowledge about eukaryotic gene structure, and studies comparing ribosomal RNA sequences, we propose a train of events for precellular and early cellular evolution.**

When RNA–RNA splicing in the manufacture of adenovirus mRNA was discovered in 1977 (1, 2) and soon thereafter introns were found in dozens of cell genes, widespread speculation on the evolutionary "meaning" of noncontiguous protein coding was stimulated. Two rather different streams of thought were generated. (*i*) Noncontiguous coding in exons may have promoted more vigorous DNA recombination, thus leading to rapid eukaryotic evolution. The implication of this proposal is that noncontiguous coding came after contiguous coding (3, 4). (*ii*) RNA processing in precellular time may have been important in allowing the earliest genes to be assembled and to function (5, 6).

Noncontiguous Coding and New Gene Function. Many authors (3–11) called attention to the fact that introns frequently divide DNA into regions that encode functional domains or subdomains (7, 8) in proteins. The evolutionary importance of functional domains in proteins was already known (12–14); and conservation of physical and functional structure, not necessarily amino acid sequence, in the dehydrogenases was the outcome of long-standing evolutionary demands for binding to nucleotide cofactors. Gilbert (3, 4) saw the arrangement of exons as units encoding protein domains as one that could, once operational, promote a greater genetic variation and therefore possibly assist evolution of higher cells. He predicted the existence of "exon shuffling"—the appearance of the same functional domain in proteins of different overall function.

Several striking cases of exon shuffling of recognized domains are now known. The key carbohydrate-metabolizing enzymes glyceraldehyde-3-phosphate dehydrogenase (15), phosphoglycerate kinase (16), alcohol dehydrogenase (17), and pyruvate kinase (18) all have binding sites for the nucleotides NAD or ATP. The nucleotide binding domains are encoded in similar exons in these genes, whereas the exons encoding other parts of these enzymes have independent evolutionary pedigrees. Gilbert (19) reported that the gene for triose phosphate isomerase, both in animal and in plant cells, has an exon structure that separates the coding region for the nucleotide binding domain of that enzyme. The similar structural domains of the dehydrogenases evident by DNA sequence analysis are those that were recognized by Rossman and coworkers (13, 14) from crystallographic analyses. The divisions of these presumably ancient genes into exons that correspond to known protein domains in a number of different vertebrate species and in plants strongly suggests an early existence both of introns and of shuffling of introns between genes. At least some of the genes for these enzymes are not divided in yeasts (20–22) and presumably in bacteria; we will argue later that this is a result of intron loss in rapidly growing organisms.

Another case of apparent sharing of exons is evident in the gene encoding the receptor for low density lipoprotein (LDL) (23, 24). This gene has 18 exons, 8 encoding a sequence of about 400 amino acids that resembles a portion of the primary translation product of epidermal growth factor (EGF), in whose gene 8 exons also encode the homologous region. The LDL receptor also shares a 40 amino acid sequence with C9, a complement factor. In the LDL receptor, this 40 amino acid unit is repeated seven times and four of these units are encoded in separate exons, as is the single C9 receptor domain. R. Doolittle (25) pointed out that by protein sequence analysis other proteins (e.g., coagulation factors IX and X) possess a segment of similar amino acid sequence and, at least in factor IX, it is encoded by a separate exon (26). Another case of partial homology, with the hormone-coding portion of the EGF gene, is known within the *Drosophila* Notch gene (27) whose protein product(s) regulate differentiation of epidermal-like tissues in fruit flies (M. Young, personal communication). There seems little doubt that exon duplication and exon shuffling have occurred in evolution, bearing out Gilbert's prediction.

Origin of Noncontiguous Coding. In addition to the proposed exon shuffling, a second line of thought was stimulated in 1978 by the discovery of introns in genes. We and Darryl Reanney (5, 6, 28) recognized the likelihood that exons might represent protein domains but were concerned not with how intron–exon arrangements, once present, would affect the course of evolution, but with the problem of how the introns and the protein-coding domains got there in the first place. We were especially taken with the possibility that the widespread presence of exons in eukaryotic cells revealed something about the very earliest precellular evolution. We (29–31) and Reanney (32) proposed that the earliest cells contained introns because the first coding nucleic acid was likely to have some stretches with and some without meaningful information (protein domains). The use of RNA–RNA splicing early in evolution would thus have been a great boon, if not a prerequisite, to early genomic evolution. We used several general arguments to develop these suggestions further. (*i*) Introns are ancient, as exemplified by their constant presence through very long times in evolution in, for example, the globin genes (33), related myoglobin genes (34), and even leghemoglobin genes (35) of plants. (*ii*) Introns seemed unlikely to have entered preexisting genes, all coded in a contiguous fashion, in the absence of an already functioning splicing mechanism to ensure their accurate removal. Wholesale entrance of introns into contiguous genes implies anticipatory evolution of an as yet unneeded biochemical capacity. (*iii*) The "aim" of the majority of the entering introns in a set of already contiguous functioning genes would have been extraordinary: entrance of introns at the DNA level would have divided genes at sites that often mark the boundaries of protein domains. For all these reasons, we took

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

as very unlikely the entrance late in evolution of introns into organisms with already functioning contiguous genes.

Introns in the Earliest RNA? At the time of the earliest discussions of the evolutionary meaning of splicing, it was already clear that RNA processing was very widespread in all types of RNA in eukaryotic cells (6) and in tRNA and mRNA in bacteria. So removal of unused sequences from RNA transcripts and ligation of useful sequences were no longer unrealistic possibilities for early cells (30, 31). Reanney (28, 32) called attention to the idea discussed in the late 1960s by Crick (36), Orgel (37), and Woese (38) that RNA preceded DNA in early evolution. He suggested that since what introns in present-day genomes really signified was the necessity for RNA–RNA processing, why not assume the introns existed in the early RNA before DNA?

A corollary to all these speculations, of course, is that the eukaryotic nuclear genome, especially those of multicellular organisms, which are filled with introns, may have more resemblance to the genome of the earliest cells than do the genomes of present-day bacteria (5, 6). The idea that the eukaryotic nuclear genome did not arise directly from prokaryotes was put forward by Reanney before splicing was discovered (39) but did not capture wide attention at that time.

Cell Lineages from Ribosomal RNA Comparisons: A New Way to Look at the Prokaryote–Eukaryote Dichotomy. That eukaryotic genomes are, at least in some features, more primitive than bacterial genomes ran counter to prevailing concepts about the nature of the relationship between prokaryotes and eukaryotes. It seemed only common sense, before eukaryotic gene structure was understood, to view prokaryotes as more primitive than eukaryotes because prokaryotic cells seem structurally less complex and they have smaller genomes. Life was then viewed as progressing from some ill-defined precellular precursor, to bacteria that were fundamentally like modern bacteria, then on to simple eukaryotes perhaps like modern yeasts, and finally to complex multicellular eukaryotes like ourselves, with increasingly complex adaptations of the hereditary apparatus accumulating along the way.

This "common sense" view was unsupported by data and is now strongly ruled against by the extensive work of Woese and his collaborators (40–42). Their ribosomal RNA sequence comparisons show clearly that eubacterial, archaebacterial, and eukaryotic (nuclear-coded) ribosomal RNAs are sufficiently similar within groups and dissimilar between groups to define three separate genomic lineages, none of which is earliest in any phylogenetic sense. These rRNA data and comparable but less extensive data on tRNA, ribosomal proteins, RNA polymerases, and other components of the transcription and translation machineries (43) suggest that eubacteria, archaebacteria, and the nuclear components of eukaryotes diverged from a common ancestor at an early stage in cellular evolution (40–42, 44). Woese and Fox (42) have called this common ancestor a "progenote." One prominent idea about the lineage of eukaryotic cells, the endosymbiotic theory (45), is strongly confirmed by the work of Woese and his colleagues. Mitochondria in plants and animals and chloroplasts in plants encode rRNAs very similar to eubacterial rRNAs. Chloroplasts are thought to descend from cyanobacteria, and mitochondria from purple sulfur bacteria (40–42, 46). But the origin of cytoplasmic organelles tells us nothing about the lineage of the nucleus, which according to rRNA sequence differences came, as did eubacteria and archaebacteria, from the progenote.

Advances in RNA Chemistry and Predictions about the Earliest Genomes. If the progenote gave birth to several cell lineages, it presumably contained DNA. What clues are there, if any, of how the progenote came to be? Recently discovered nonenzymatic RNA reactions have returned our attention to the early participation of RNA, including the possibility of original RNA genomes with at least limited function. It would be well first to repeat that after the discovery of mRNA and the solution of the genetic code in the early 1960s, Crick, Orgel, and Woese (all writing about precellular evolution) made similar speculations about early genetic systems (36–38). Translation, perhaps the most important molecular genetic event, whose origin in evolution is still most shrouded in mystery, is based on three separate roles for RNA (encoding information, mRNA; translation of the information, tRNA; and facilitation of the translation, rRNA). Because no genetic system is conceivable without a coding and translating function, all three writers suggested that RNA arose before DNA. Let us now outline some reactions of which RNA is capable that provide support for and extend the proposal that RNA may have arisen first.

Synthesis of Oligo- and Polynucleotides. Orgel and his coworkers have studied protein-free oligonucleotide synthesis from activated nucleotide precursors (47, 48). Template-directed, protein-free RNA synthesis producing 3′,5′-linked chains >30 nucleotides long occurs; Watson–Crick base-pairing directs nucleotide selection in such protein-free RNA chain elongations. Although no claim is made that the reaction under study did occur on primitive earth, the results offer impressive testimony that protein-free RNA synthesis is at least possible. Protein-free oligopeptide formation is possible but much more difficult, and no real success in oligodeoxynucleotide synthesis has been described in protein-free reactions.

RNA-Directed Site-Specific Cleavage. Cleavage of bacterial RNAs, was documented in cells in the late 1960s and early 1970s, and enzymatic activities that would carry out these reactions were described (49, 50). Several cases of RNA-directed, protein-free cleavages of bacterial and of eukaryotic RNAs have now been discovered. (*i*) Many tRNAs in bacteria must be processed from primary transcripts by cleavages at both the 5′ and the 3′ ends. Altman and his colleagues (50) purified RNase P, the *Escherichia coli* enzyme capable of carrying out this cleavage. The active enzyme contains a 375-nucleotide RNA chain termed M1. When pre-tRNA is mixed with pure M1, correct site-specific cleavage of the precursor occurs without any protein (51).

(*ii*) In bacteriophage T4, the protein-coding sequence for thymidine kinase is noncontiguous, requiring the excision of an intron and religation of the RNA primary transcript to achieve a functional mRNA (52). At least the first step of this process, excision of the intron, can occur without protein. Another case of T4 mRNA cleavage without protein has also been reported (53).

(*iii*) Perhaps the most dramatic RNA-directed reaction known is the splicing out of an intron and ligation of the exons in *Tetrahymena pyriformis* rRNA. Cech *et al.* (54, 55) found that, with guanosine as a cofactor, a 408-nucleotide intron is correctly removed from the primary transcript and the two ends ligated without the presence of protein. Protein-free RNA splicing in the mitochondrial rRNAs and mRNAs of *Neurospora crassa* (56) and yeast (57) have also been reported. The sequences of all the spontaneously excised introns are very similar and, in fact, resemble many other introns found in mitochondrial mRNAs (58). Finally, Zaug and Cech (59) have shown that the spontaneous ligation reaction that joins the exons and also joins the ends of the excised intron to form a circle can also join two introns from different primary transcripts, a case of so-called "trans-splicing." Konarska *et al.* (60) and Solnick (61) also have described trans-splicing, in protein-containing extracts of HeLa cells. Both Sharp (62) and Zaug and Cech (59) point out that such a reaction provides a basis by which long chains of RNA might be built up; in the case of the *Tetrahymena* reaction, we see this "synthetic" event occurring without

protein. In addition, Sharp (62) points to the biochemical similarities in the RNA-mediated splicing reaction discovered by Cech and the reaction carried out by eukaryotic nuclei to splice mRNAs. In both cases the 2′ and 3′ hydroxyls of a guanosine moeity may be active in effecting attack at the splice site, and in both cases the phosphate that unites the two exons preexists in the RNA chain and can be thought of as participating in a transesterification.

A Possible Pathway of Precellular and Early Cellular Evolution. Thus, we see that protein-free RNA synthesis, site-specific cleavage, and splicing could all have been available for use in assembly of an early, RNA-dominated genetic system. Fig. 1 traces possible events of early RNA genome organization leading eventually to cell evolution.

RNA chemistry is assumed to be the dominant successful polymer chemistry in prebiotic times (28, 30–32, 50–57), and fairly sophisticated RNA molecules are envisioned as early participants in evolution. A great, if not the greatest, stumbling block in these considerations is met early and is the problem addressed by Crick (36), Orgel (37), Woese (38), and

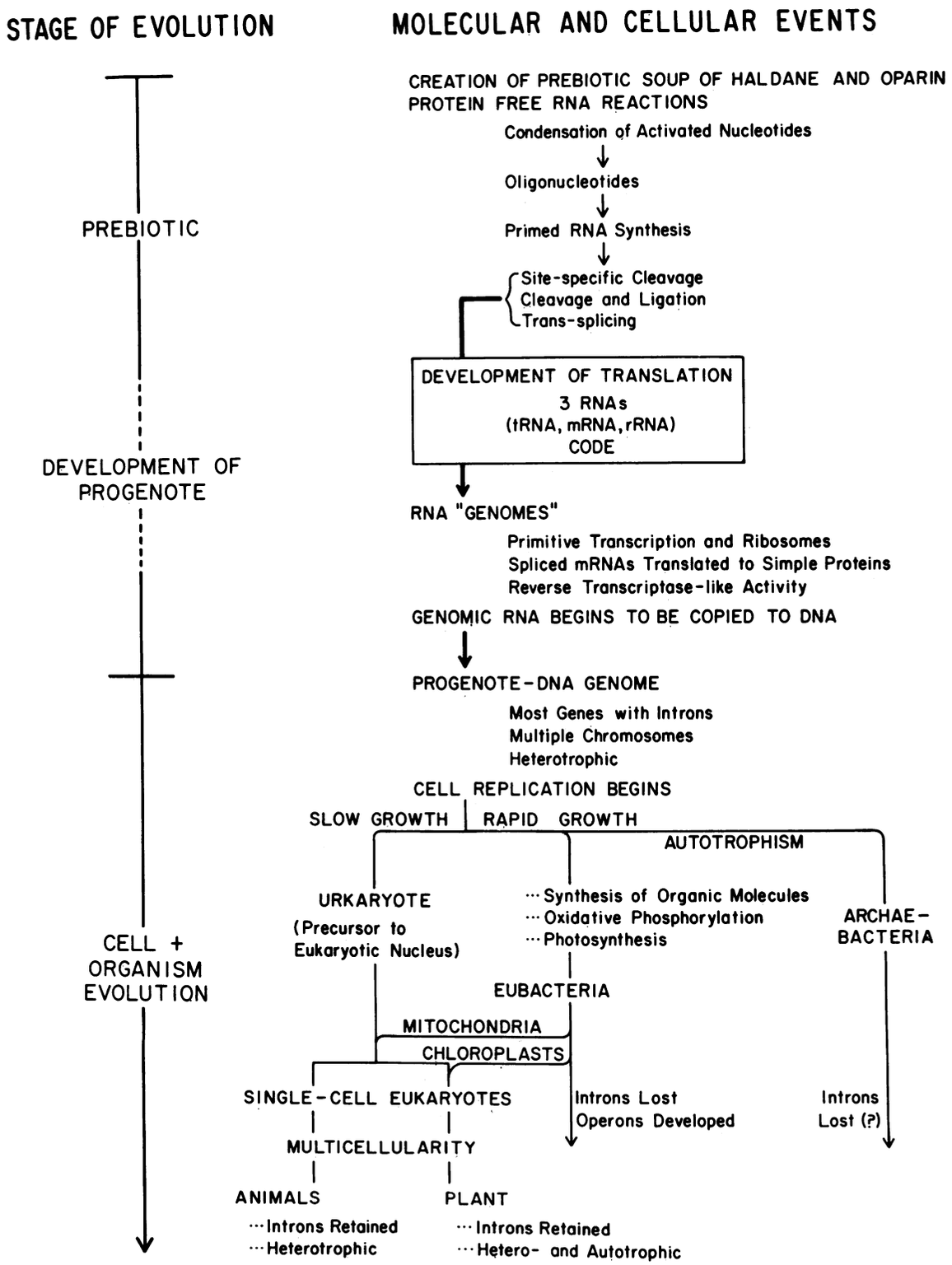

FIG. 1. Some possible events in evolution, from the prebiotic "soup" of Haldane and Oparin (63, 64) through multicellular development.

others. Even given RNA molecules with which oligopeptide–oligonucleotide interactions might occur, how did primitive translation arise? Crick *et al.* (65) have described a model that would allow an oligonucleotide (heptamer) to function as a tRNA, and Eigen and Schuster (66) have speculated on which interactions were most likely for the heptamer in the earliest prebiotic times. Although these discussions present plausible solutions to the problem of evolution of translation, no experimental breakthroughs have been reported that illuminate this crucial area.

Given a solution to the translation problem, the earliest coding system is conceived of as RNA-directed and filled with introns. It seems reasonable that genomic RNA(s) existed as well as primitive transcription and translation. Whether for reasons of stable information storage by DNA compared to the chemically less stable RNA or for other reasons having to do with successful gene replication, DNA as the genetic material was likely before true replicating cells arose. How did DNA evolve?

Here we also have recourse to recent experimental results. Reverse transcriptase (RNA-dependent DNA polymerase) activity as a necessary step in the life cycle of retroviruses was discovered 15 years ago (67, 68), and these viruses were known then to be widespread in vertebrates. In recent years, many segments in vertebrate genomes, such as *Alu*-type sequences (69), pseudogenes for the U series of small RNAs (70), and so-called processed genes (71, 72) [stretches of inert, promoterless DNA that are obvious partial or full copies of mRNAs, even including their posttranscriptionally added poly(A)], have been shown likely to move within the genome through a reverse transcriptase-like activity that copies transcribed RNA to DNA (70, 73, 74). Also, in *Drosophila* cells (75), in cauliflower mosaic virus (76), and in yeast (77), reverse transcriptase-like activity is known and may be the major cause of movement of mobile DNA elements within the nuclear DNA for perhaps all eukaryotes. Moreover, the putative protein has amino acid homology in all these cases. Thus, the reverse transcriptase-like enzyme may be very ancient, possibly the original mechanism by which DNA was copied from RNA near the end of precellular time. If the earliest RNA-based genetic system already had functioning transcription units filled with introns, then only the copying of the genomic RNA, and not the promoterless mRNA, would have resulted in transfer to DNA of usable genetic information. Introns that existed in the RNA genome would have been preserved in the transfer to the DNA of the progenote.

If the genome of the progenote were replete with introns, and both prokaryotic and eukaryotic genomes descend from it, why do only the latter retain introns in protein-coding chromosomal genes? We have argued previously (5, 6, 29, 30) that this reflects selection against the retention of introns in fast-growing cells because their replication, transcription, and excision involve bioenergetic expense. Selection against the retention of introns should be weaker in organisms for which cell replication represents a smaller fraction of total energy expenditure; that is, complex, highly differentiated, multicellular animals and plants. In neither case would the possession of introns *per se* be an adaptation (although certain specific intron sequences may have been coopted into regulatory functions). This is not to say, however, that introns have not had an important, perhaps even crucial effect of the sort Gilbert described (3, 4) in the evolution of complex eukaryotes. They might thus best be seen as "exaptations," in the terminology of Gould and Vrba (78).

Another feature of eukaryotes, including yeast, is mentioned in Fig. 1. For example, the DNA of *Saccharomyces cervisiae* is only 3–5 times the size of the *E. coli* genome but is contained in 17 chromosomes, all of which have multiple origins of replication (79). As suggested before (30), this state of affairs could reflect the organization of the progenote. In precellular time, those transcription units that had replication origins would have been the most populous and the most likely to be included in the first successful collection of functioning transcription units. If such a cell survived without streamlining its genome and making efficient, rapid growth a priority, perhaps a disjointed genome also is a vestige of an early state of relative disorganization. At any rate, it is interesting that all eukaryotic DNAs seem to have replication origins every 10^4–10^5 base pairs (79, 80) and multiple chromosomes. This of course contrasts with the much faster growing prokaryotes, which have very few introns, one chromosome, and one tightly controlled origin of replication (81).

Also in Fig. 1, an implicit parallel is drawn for which no obvious reason exists: the eukaryotes that did not become fast growers solved their environmental challenges by forming multicellular masses that were the forerunners of plants and animals. Together with this "community-mindedness" and slower growth, this cell lineage is pictured as having retained introns more faithfully during evolution. Perhaps this relative sluggishness in discarding unused DNA allowed the necessary genome expansion and consequent variation that allowed the evolution of genetic programs for development. It is at least true that *E. coli* loses unnecessary sequences promptly (82, 83) and that mice retain apparently useless, newly introduced sequences for 15 generations or more (84, 85).

While the scheme in Fig. 1 can be defended as possible or even plausible, it is simply one assembly of possibilities. It does embrace three important findings of the last decade: all cell lineages do not "begin" with prokaryotes, RNA chemistry that occurs in the laboratory today is a reasonable guide to what could have happened in precellular times, and introns are ancient in the genomes of eukaryotes and no wholesale creation of intron-containing genes has ever been posited. Therefore, a model in which the precursor of the eukaryotic nuclear genome is traced back to the earliest times in evolution has reasonable logical force.

Note Added in Proof. Zaug and Cech (86) have reported that *Tetrahymena* intervening-sequence RNA can function as an RNA synthetic enzyme to add single residues through the transesterification reaction, and Peebles and coworkers (87) have shown that a yeast mitochondrial intron can be removed by self-splicing to yield a branched, "lariat" structure, like the introns in genes transcribed by RNA polymerase II. These additional protein-free reactions greatly strengthen the suggestions in Fig. 1 that sophisticated protein-free RNA chemistry initiated the formation of biopolymers and that introns were present in the earliest RNA.

We thank P. Model and M. Young for helpful comments. J.E.D. is supported by grants from the National Institutes of Health and the American Cancer Society, and W.F.D. by the Canadian Research Council and National Sciences and Engineering Research Council. W.F.D. thanks the T. S. Guggenheim Foundation for sabbatical fellowship support.

1. Berget, S. M., Moore, C. & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3171–3175.
2. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. T. (1977) *Cell* **12**, 1–8.
3. Gilbert, W. (1978) *Nature (London)* **271**, 501.
4. Gilbert, W. (1979) in *Eukaryotic Gene Regulation: ICN-UCLA Symposia on Molecular and Cellular Biology*, eds. Axel, R., Maniatis, T. & Fox, C. F. (Academic, New York), Vol. 14, pp. 1–12.
5. Doolittle, W. F. (1978) *Nature (London)* **272**, 581–582.
6. Darnell, J. E. (1978) *Science* **202**, 1257–1260.
7. Blake, C. C. F. (1978) *Nature* **273**, 267.
8. Blake, C. C. F. (1983) *Nature (London)* **306**, 535–537.
9. Craik, C. S., Buchman, C. S. & Beychok, S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1384–1388.
10. Go, M. (1981) *Nature (London)* **291**, 90–92.
11. Tonegawa, S., Maxam, A. M., Tizard, R., Bernhard, O. &

Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1485–1489.
12. Cunningham, B. A., Gottlieb, P. D., Pflumm, M. N. & Edelman, G. M. (1971) in *Progress in Immunology*, ed. Amos, B. (Academic, New York), pp. 3–24.
13. Rossman, M. G., Moras, D. & Olsen, K. W. (1974) *Nature (London)* **250**, 194–199.
14. Rossman, M. G. & Argos, P. (1981) *Annu. Rev. Biochem.* **50**, 497–532.
15. Stone, E. M., Rothblum, K. N. & Schwartz, R. J. (1985) *Nature (London)* **313**, 498–500.
16. Michelson, A. M., Blake, C. C. F., Evans, S. T. & Orkin, S. H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6965–6969.
17. Branden, C.-I., Eklund, H., Cambillan, C. & Pryor, A. J. (1984) *EMBO J.* **3**, 1307–1310.
18. Lonberg, N. & Gilbert, W. (1985) *Cell* **40**, 81–90.
19. Gilbert, W. (1985) *Science* **228**, 823–824.
20. Hitzeman, R. A., Hagie, F. F., Hayflick, J. C., Chen, C. Y., Seeburg, P. H. & Derynek, R. (1982) *Nucleic Acids Res.* **10**, 7791–7808.
21. Russell, E. W., Smith, M., Williamson, V. M. & Young, E. T. (1983) *J. Biol. Chem.* **258**, 2674–2682.
22. Holland, J. P. & Holland, M. J. (1980) *J. Biol. Chem.* **255**, 2596–2605.
23. Sudhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* **228**, 815–822.
24. Sudhof, T. C., Russell, D. W., Goldstein, J. L., Brown, M. S., Sanchez-Pescador, R., Bell, G. & Greene, I. (1985) *Science* **228**, 839–895.
25. Doolittle, R. F. (1985) *Sci. Am.* **253**, 88–99.
26. Anson, D. S., Chou, K. H., Rees, D. J. G., Gianelli, P., Gould, K., Huddleston, J. S. & Brownlee, G. G. (1984) *EMBO J.* **3**, 1053–1060.
27. Bargiello, T. & Young, M. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2142–2146.
28. Reanney, D. C. (1979) *Nature (London)* **227**, 597–600.
29. Doolittle, W. F. (1981) in *Evolution Today: Proceedings of the Second International Congress of Systematic and Evolutionary Biology*, eds. Scudder, G. G. E. & Reveal, J. J. (Hunt Inst. for Botanical Documentation, Carnegie-Mellon Univ., Pittsburgh), pp. 197–206.
30. Darnell, J. E., Jr. (1981) in *Evolution Today: Proceedings of the Second International Congress of Systematic and Evolutionary Biology*, eds. Scudder, G. G. E. & Reveal, J. J. (Hunt Inst. for Botanical Documentation, Carnegie-Mellon Univ., Pittsburgh), pp. 207–213.
31. Darnell, J. E. (1983) *Sci. Am.* **249**(4), 90–100.
32. Reanney, D. C. (1981) in *Evolution Today: Proceedings of the Second International Congress of Systematic and Evolutionary Biology*, eds. Scudder, G. G. E. & Reveal, J. J. (Hunt Inst. for Botanical Documentation, Carnegie-Mellon Univ., Pittsburgh), pp. 214–234.
33. Stamatoyannopoulos, G. & Nienhuis, A. W., eds. (1981) *Organization and Expression of Globin Genes* (Liss, New York).
34. Blanchetot, A., Wilson, V., Wood, D. & Jeffreys, A. J. (1982) *Nature (London)* **301**, 732–734.
35. Jensen, E. O., Paludan, K., Hyldig-Nielsen, J. J., Jorgensen, P. & Marcker, K. A. (1981) *Nature (London)* **291**, 677–679.
36. Crick, F. (1968) *J. Mol. Biol.* **38**, 367–379.
37. Orgel, L. (1968) *J. Mol. Biol.* **38**, 381–393.
38. Woese, C. R. (1967) *The Origins of the Genetic Code* (Harper & Row, New York).
39. Reanney, D. C. (1974) *J. Theor. Biol.* **48**, 243–251.
40. Woese, C. R. (1981) *Sci. Am.* **244**, 98–125.
41. Woese, C. R. (1983) in *Evolution from Molecules to Man*, ed. Bendall, D. S. (Cambridge Univ. Press, Cambridge, MA), pp. 209–233.
42. Woese, C. R. & Fox, G. E. (1977) *J. Mol. Evol.* **9**, 369–371.
43. Doolittle, W. F. (1985) *Trends Genet.*, in press.
44. Woese, C. R., Gutell, R. R., Gupta, R. & Noller, H. F. (1983) *Microbiol. Rev.* **47**, 621–669.
45. Margulis, L. (1981) *Symbiosis in Cell Evolution* (Freeman, New York).
46. Young, D., Oyaizu, Y., Oyaiza, H., Olsen, G. J. & Woese, C. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4443–4447.
47. van Roode, J. H. G. & Orgel, L. E. (1980) *J. Mol. Biol.* **144**, 579–585.
48. Inoue, T. & Orgel, L. E. (1983) *Science* **219**, 859–862.
49. Nikolaev, N., Silengo, L. & Schlessinger, D. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3361–3365.
50. Stark, B. C., Kole, R., Bowman, E. T. & Altman, S. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3717–3721.
51. Guerrier-Takada, C., Dardiner, K., Marsh, T., Pace, N. & Altman, S. (1983) *Cell* **35**, 849–857.
52. Belfort, M., Pedersen-Lane, J., West, D., Ehrenman, K., Maley, G., Chu, F. & Maley, F. (1985) *Cell* **41**, 375–382.
53. Watson, N., Gurewitz, M., Ford, J. & Spirion, D. (1984) *J. Mol. Biol.* **172**, 301–323.
54. Cech, T. R. (1983) *Cell* **34**, 713–716.
55. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gotschling, D. E. & Cech, T. R. (1982) *Cell* **31**, 147–157.
56. Burke, J. M. & Rajbhandary, U. L. (1982) *Cell* **31**, 509–520.
57. van der Horst, G. & Tabak, H. F. (1985) *Cell* **40**, 759–766.
58. Waring, R. B., Scazzocchio, C., Brown, T. A. & Davis, R. W. (1983) *J. Mol. Biol.* **167**, 595–605.
59. Zaug, A. J. & Cech, T. R. (1985) *Science* **229**, 1060–1064.
60. Konarska, M. M., Padgett, R. A. & Sharp, P. A. (1985) *Cell* **42**, 165–171.
61. Solnick, D. (1985) *Cell* **42**, 157–164.
62. Sharp, P. A. (1985) *Cell* **42**, 397–400.
63. Haldane, J. B. S. (1929) *Rationalist Annals*, p. 2; reprinted in Smith, J. M., ed. (1985) *On Being the Right Size and Other Essays* (Oxford Univ. Press, Oxford), pp. 101–112.
64. Oparin, A. I. (1938) *The Origin of Life* (Macmillan, New York).
65. Crick, F. H. C., Brenner, S., Klug, A. & Piecznik, G. (1976) *Origins Life* **7**, 389–397.
66. Eigen, M. & Schuster, P. (1979) *The Hypercycle* (Springer, Berlin).
67. Baltimore, D. (1970) *Nature (London)* **226**, 1209–1211.
68. Temin, H. & Mizutani, S. (1970) *Nature (London)* **226**, 1211–1213.
69. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51**, 813–844.
70. van Arsdell, S. W., Dennison, R. A., Bernstein, L. B., Weiner, A. M., Manger, T. & Gesteland, R. F. (1981) *Cell* **26**, 11–17.
71. Wilde, C. D., Crowther, C. E., Croipe, T. E., Lee, M. G-S. & Cowan, N. J. (1982) *Nature (London)* **297**, 83–84.
72. Sharp, P. A. (1983) *Nature (London)* **301**, 471–472.
73. Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26**, 141–142.
74. Baltimore, D. (1985) *Cell* **40**, 481–482.
75. Flavell, A. (1984) *Nature (London)* **302**, 119–124.
76. Pfeiffer, P. & Hohn, T. (1983) *Cell* **33**, 781–789.
77. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) *Cell* **40**, 491–500.
78. Gould, S. J. & Vrba, E. S. (1982) *Paleobiology* **8**, 4–15.
79. Newton, C. S. & Burke, W. (1981) in *Mechanistic Studies of DNA Replication and Genetic Recombination*: ICN-UCLA Congress on Molecular and Cell Biology, ed. Alberts, B. M. (Academic, New York), Vol. 19, pp. 399–410.
80. Hand, R. (1978) *Cell* **15**, 317–325.
81. Kornberg, A. (1980) *DNA Replication* (Freeman, New York).
82. Folk, W. R. & Berg, P. (1971) *J. Mol. Biol.* **58**, 595–610.
83. Rigby, P. W. J., Burleigh, B. D. & Hartley, B. S. (1974) *Nature (London)* **251**, 200–204.
84. Jaennisch, R. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1260–1264.
85. Johler, J., Timpi, R. & Jaenisch, R. (1984) *Cell* **38**, 597–607.
86. Zaug, A. J. & Cech, T. R. (1986) *Science* **231**, 470–475.
87. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. & Cheng, H.-L. (1986) *Cell*, in press.