REVIEW ARTICLE

The precursor to animal cell messenger RNA

Charles COUTELLE

*Central Institute for Molecular Biology, Academy of Sciences of the German Democratic Republic,
DDR-1115 Berlin-Buch, German Democratic Republic*

The idea that the primary products of transcription are different from the mature polysomal mRNA emerged in the early 1960s, due in part to the observation that the majority of the radioactivity after pulse-labelling of RNA in mammalian cell cultures remained in the nucleus and was never transferred to the cytoplasm (Harris, 1962; Sibatani *et al.*, 1962; Georgiev *et al.*, 1963). On extraction, this nuclear RNA showed a heterogeneous, 'bell-shaped' sedimentation profile, including molecules apparently much larger than cytoplasmic polysomal mRNA. In spite of this size difference, the nuclear RNA is very similar to mRNA in base composition, which clearly distinguished it from ribosomal RNA and its precursors (Scherrer *et al.*, 1963).

These results led to several proposed models, among them the hypothesis that the primary products of transcription are large precursors of mRNA which are processed while still within the nucleus. Most of each nuclear RNA sequence is degraded, and only the mRNA portion is transferred to the cytoplasm where it becomes active in directing protein synthesis (Scherrer & Marcaud, 1968; Georgiev *et al.*, 1972; Darnell *et al.*, 1973). Scherrer & Marcaud (1968) emphasized, in their cascade hypothesis, that processing of RNA from nucleus to cytoplasm could include specific regulatory steps of importance in the fine control of gene expression.

Samarina *et al.* (1968) demonstrated that hnRNA, as it was called, is complexed with specific nuclear proteins. They named these ribonucleoprotein particles informers, implying that they are transport forms of processing intermediates of newly synthesized mRNA. Although these RNPs have been quite extensively investigated (Williamson, 1973; Pedersen, 1974) their involvement and function in the processing mechanism is still unclear.

Specific gene transcripts

For the rather long time of about 10 years, investigations were limited to studies of total,

Abbreviations used: hnRNA, heterogeneous nuclear RNA; RNP, ribonucleoprotein; cDNA, complementary DNA; pre-mRNA, precursor mRNA.

unfractionated hnRNA. Only after the introduction of techniques allowing the isolation of specific eukaryotic mRNAs, and their transcription into highly radioactive cDNA molecules, did the study of individual pre-mRNA molecules become possible. Imaizumi *et al.* (1973) were the first to show the presence of globin sequences in RNA fractions sedimenting much faster than duck globin mRNA in a sucrose density gradient, using a cDNA prepared from mRNA as a sequence-specific probe.

The main criticisms of these results arose from the possibility that hybridization was to aggregates of smaller RNA, perhaps the size of mature mRNA, rather than to genuinely large precursor molecules. Although several authors, including ourselves, have found sequences hybridizing to globin cDNA of a size up to that of 28S ribosomal RNA even under strictly denaturing sedimentation conditions (Spohr *et al.*, 1976; Strair *et al.*, 1977; Hunger *et al.*, 1979), most investigations only detect sequences with a maximum size of 16S (MacNaughton *et al.*, 1974; Knochel & Grundmann, 1978; Ross & Knecht, 1978). This is also the size of the major β -globin pre-mRNA species isolated by affinity chromatography on cDNA columns (Curtis *et al.*, 1977a; Kwan *et al.*, 1977; Ross, 1978). Only Bastos & Aviv (1977) claim the isolation of a larger (27S) globin mRNA precursor.

Despite differences concerning the exact size, these results clearly established the long-disputed existence of pre-mRNA molecules. Since the poly(A) sequence is added to the 3'-terminus of the pre-mRNA fairly early after transcription, it was generally assumed that the large precursor molecule was processed primarily by trimming from the 5' end. These views were radically changed by the exciting discoveries of intervening sequences in structural genes, first for adenovirus and SV40 virus (Tooze, 1980). Both viruses were shown to possess a genomic leader sequence which is transcribed to give an RNA molecule which does not code for a protein sequence, but becomes attached to different protein-coding mRNAs transcribed from widely separated parts of the virus DNA. One of several suggested mechanisms to explain this attachment was to postulate a long primary RNA transcript

from which sequences between the leader sequence and the different mRNA sequences are looped out and excised, while leader and mRNA sequences are spliced together. This mechanism has proved to be correct.

Recombinant DNA technology *in vitro* gives the possibility of obtaining specific and highly radioactive molecular probes for the identification and analysis of single eukaryotic genes and their transcription products by Southern transfer-hybridization, cloning and nucleic acid sequencing. The surprising observation of unexpected restriction sites (Jeffreys & Flavell, 1977) and R-loop structures (Tilghman *et al.*, 1978) within the globin structural genes was soon interpreted, in the light of the tumour virus data, as caused by intervening non-coding DNA sequences interrupting the coding sequences of the globin genes. The logical consequence of this gene structure is that transcription involves the primary synthesis of a large pre-mRNA molecule containing both intervening sequence transcripts as well as sequences represented in the final, polysomal mRNA. It may seem ironic that the most important information now available about the structure of pre-mRNA has come from the analysis of individual gene sequences at the DNA rather than the transcript level.

Pre-mRNA structure

At present, the eukaryotic genes coding for proteins that have been most fully investigated are the various globin genes from different species, the chicken genes coding for the oviduct proteins ovalbumin, conalbumin and ovomucoid, the insulin genes of several species including man, the immunoglobulin genes and the histone genes. The knowledge obtained in the investigation and comparison of these genes, together with studies on the primary transcription products themselves, has given us detailed knowledge of the structure of several pre-mRNA species, although not necessarily about the whole length of the primary transcript. The general pattern emerging from these data is summarized as follows (Fig. 1):

(1) The pre-mRNA molecules are several times larger than the size of the mature mRNA. Each contains not only the coding and non-coding mRNA sequences, but also intervening sequences and possibly also sequences extending to both the 5' and the 3' ends of the mature mRNA sequence. Due to their short half-life and low concentration, the exact size and structure of the primary transcripts are still not absolutely clear in most cases.

(2) The non-translated 3' sequence following the

Fig. 1. Structure of an eukaryote gene and its transcription and processing products

coding sequence in all eukaryotic mRNAs studied to date contains a pentanucleotide AAUAA approx. 10–20 nucleotides before the poly(A) starting site (Proudfoot & Brownlee, 1976). The corresponding DNA sequences may be termination signals for transcription, causing RNA synthesis to finish at the poly(A) addition site. Alternatively, an as yet unknown transcription termination signal could be situated further downstream to the poly(A) attachment site, giving a large non-polyadenylated pre-mRNA molecule that must be trimmed at its 3' end before poly(A) can be added. Such a mechanism is known for adenovirus-2 late mRNA processing (Fraser *et al.*, 1979). A readthrough of the poly(A) addition site has also been observed recently in β -globin gene transcription in dimethylsulphoxide-induced Friend cells (Hofer & Darnell, 1981).

(3) Convincing evidence points towards the 5' end of the mature mRNA being identical with the starting point of transcription. In several eukaryotic genes (Gannon *et al.*, 1979) a region of homology of approx. 12 nucleotides has been found immediately adjacent to the 5' end of the mature mRNA. The prevalent sequence

CTCA
GAGT

immediately preceding mature mRNA sequences is now generally thought to be the capping signal. At least for the 16S globin pre-mRNA, the cap structure is identical with that of the mature mRNA (Curtis *et al.*, 1977b), and the 5' termini of both molecules map at the same position in the genomic DNA (Weaver & Weissmann, 1979). Similar investigations have led to the same conclusion for ovalbumin gene transcription (Roop *et al.*, 1980).

Therefore, either this region of homology adjacent to the 5' cap site is a transcription initiation signal, with transcription starting immediately after the

CTC
GAG

sequence and the first nucleotide of the primary transcript being methylated and capped, or this region serves in a larger transcript as a signal for 5' trimming and capping. The former possibility is the more likely, since a characteristic sequence

TATAAA
ATATTT

very much resembling the promoter site for *Escherichia coli* RNA polymerase (Pribnow, 1975), has been found in all animal genes investigated so far. This signal, known as the TATA box, is about 23 nucleotides upstream from the cap site. It has been shown to play a role in the initiation of transcription for several genes in transcription systems *in vitro* (Wickens & Laskey, 1981). Even

so, the remote possibility exists that the actual starting point of transcription is upstream, rather than downstream, from the promoter, as shown for the 5S ribosomal RNA genes by Engelke *et al.* (1980). Such a mechanism could explain the observation of Kinniburgh & Ross (1979), who concluded that there could be as many as 250 nucleotides to the 5' end of the cap site in pre-mRNA by comparison of the electrophoretic mobility of the largest detectable globin pre-mRNA species with the nucleotide length as determined by sequencing (Konkel *et al.*, 1978).

In the 5' untranslated region of several mature eukaryotic mRNAs, a possible stable secondary structure forming a base-paired stem and loop can be suggested theoretically. This has been proposed as a region of possible base pairing between mRNA and 18S rRNA, functioning as a site for the initiation of translation (Hagenbüchle *et al.*, 1978; Rosenberg & Paterson, 1979) corresponding to the specific base pairing of the 16S rRNA with the 5' end of prokaryotic mRNAs. It is interesting to note that, for several genes containing an intervening sequence in the gene region corresponding to the 5' untranslated part of mature mRNA, the excision and splicing mechanism is necessary to bring together the mRNA sequences that could form the stem and loop structure mentioned above.

Optional splicing

It is now known that, in some cases, the use of different splice acceptor sites generates more than one mRNA. For instance, the mRNA for mouse α -amylase is transcribed in liver and salivary gland from the same gene. The liver mRNA contains a 5' non-coding region which is separated in the pre-mRNA from the coding region by a 4500 base intervening sequence. However, in the salivary gland, transcription starts approximately 2800 bases further upstream with a non-coding sequence, which in this pre-mRNA is separated from the coding sequence by an intervening sequence approx. 7500 bases in length. In order to form the two tissue-specific α -amylase mRNAs, a differential splicing of the two pre-mRNAs occurs, by which the two different mRNA non-coding sequences are ligated to the common coding sequence. The intervening sequence of the larger, salivary gland pre-mRNA (which contains, of course, the non-coding sequence preserved in liver α -amylase mRNA) is excised in this way (Young *et al.*, 1981).

The mRNAs for secreted and membrane-bound μ -immunoglobulins are also transcribed from the same gene. They differ in their 3' regions, which contain the mRNA sequences coding for a terminal polypeptide region which is specific for membrane binding or for secretion. These sequences are added

to the main mRNA by differential splicing. In the case of the membrane-specific μ -immunoglobulin mRNA a readthrough of the secretory μ -mRNA poly(A) addition site has to occur in order to transcribe the membrane-specific sequences situated further to the 3' side. Differential splicing then eliminates the intervening sequence containing the terminal sequences specific for μ -immunoglobulin secretion.

These mechanisms could explain the remaining ambiguities concerning the specificity of pre-mRNA initiation and termination (Hofer & Darnell, 1981), if we assume that a certain degree of 'leakage' towards the 'wrong option' might occur during the process of transcription. Molecules with incorrect sites might form faulty secondary and tertiary structures in a particular cellular environment, and would therefore be degraded more rapidly than 'correct' mRNAs, as has been demonstrated for modified 5S RNA precursors (Stahl *et al.*, 1979).

Intron structure

There seem to be no general rules concerning the sizes and numbers of introns in different genes. The genes of the β -globin gene complex, analysed from different species, all contain at least two intervening sequences at corresponding positions, but of various sizes. Within a species, the size of each intron is more constant for either family of globin genes (Leder *et al.*, 1978).

The human insulin gene also contains two introns, the larger of which interrupts the coding sequence at exactly the same position as in the rat insulin gene II. The insert itself differs considerably in size and sequence between the two species, and is missing completely in the rat insulin gene I (Lomedico *et al.*, 1979). A smaller intron present in all insulin genes investigated so far is located in the region coding for the 5' nontranslated portion of the mRNAs, and is displaced by three bases in the human genes as compared with the rat genes (Bell *et al.*, 1980). The ovalbumin gene is split into exons by seven intervening sequences (Dugaiczky *et al.*, 1978), and the chicken conalbumin gene by 16 introns into 17 exons, each of approx. 60–200 base pairs (Cochet *et al.*, 1979).

Although intervening sequences differ considerably in length and sequence even for the same gene in different species, there is remarkable conservation of the regions flanking the coding sequences. As pointed out by Chambon's group (Breathnach *et al.*, 1978) the prototype sequence

UCAG GUA----UNCAG G...
exon intron exon

is characteristic of all mRNA intron/exon junctions investigated to date. Konkelt *et al.* (1978) have

emphasized that this type of sequence repetition allows splicing to occur in different reading frames without altering the correct amino acid sequence. Extensive computer analyses have been made in search of regions of complementarity which would allow recognition of a splice site through base pairing of either mRNA or intron sequences. No convincing evidence for such a mechanism has been found. Hamer & Leder (1979) have also argued that Watson-Crick base pairing cannot play the main role in RNA splicing, as correct splicing did not occur when a part of the globin gene including the large intervening sequence was inserted into SV40 in the reverse direction. Intramolecular base pairing should be independent of the orientation of the inserted fragment.

Obviously there is no straightforward recognition mechanism by intramolecular base pairing. It has been suggested (Lerner & Steitz, 1979) that the small nuclear RNA species might be able to base pair with the splice boundary sequences and thereby direct the correct cutting and joining of the RNA. Studies on the yeast mitochondrial gene coding for cytochrome *b* have led Halbreich *et al.* (1980) to suggest that intron sequences remote from the splice junction may also function in the specificity of processing. The conservation of the splice junctions and the ability of cells to carry out correct splicing of pre-mRNA molecules from other species (Hamer & Leder, 1979) points to a rather universal recognition mechanism, perhaps involving a limited number of enzymes with quite broad specificity. Knapp *et al.* (1979) and Peebles *et al.* (1979) have studied the splicing of yeast pre-tRNA in a system *in vitro* and found evidence that the cutting and joining reactions are independent of each other. Both reactions are rather unusual: the cutting produces tRNA half molecules with 5'-OH and 3'-phosphate termini, whereas the ATP-dependent joining reaction ligates these tRNA halves to give mature tRNA.

In this case, how is the correct sequence of the mature mRNA ensured during the joining of several exons? Two facts come to mind when considering this problem. First, in all genes investigated, the order of the exon sequences from 5' to 3' end is exactly the same as that found in the mature mRNA. Secondly, in the nucleus the pre-mRNA is complexed with nuclear proteins, which have a rather complex size profile (Maundrell & Scherrer, 1979). The arrangement of these proteins on the mRNA sequence, together with the hybridization of the mRNA exons and introns to low molecular weight nuclear RNAs, could serve to ensure that the specificity of splicing is correct (Maxwell *et al.*, 1981).

It seems from several investigations on the globin and ovalbumin pre-mRNAs that the excision of the introns and splicing of the mRNA exon sequences is not performed simultaneously, but takes place in

discrete and sequential steps. At least two independent, and probably sequential, cleavage and joining reactions are necessary to splice out the large insert in mouse globin pre-mRNA (Kinniburgh & Ross, 1979). A primary transcript approx. 1800 nucleotides long is first spliced to give a 1030 nucleotide long RNA, which in turn gives rise to two different 900 nucleotide RNAs before the mature 790 nucleotide mRNA transcript is obtained. In chicken, precursor RNAs 4–5 times the size of the mature ovomucoid and ovalbumin mRNAs have been identified, and six discrete intermediates between the primary transcript and the mature mRNA have been found (Nordstrom *et al.*, 1978; Roop *et al.*, 1978).

Although the sites of splicing can be determined with accuracy, and the intermediates themselves isolated in pure form, the exact size of the primary transcript is still not clear. It is necessary to determine the initiation and termination sites of transcription unequivocally, and even then it is possible that processing occurs too rapidly to allow isolation of a primary transcript which is processed as it is synthesized. However, if the size of precursor claimed by Scherrer and others exists (Spohr *et al.*, 1974, 1976), this predicts a length of about 12000 nucleotides, longer than the size of a single globin cistron including the intergene regions (Bernards *et al.*, 1979). Readthrough of the poly(A) addition site of the δ -globin gene resulting in a δ - β -globin pre-mRNA could account for such a large molecule, but the lack of any evidence for polycistronic mRNA in eukaryotes to date makes this explanation as a normal event quite unlikely.

Several theories have been proposed to explain the sequence evolution, and the evolutionary significance, of introns (Crick, 1979). Hamer & Leder (1979) have demonstrated quite impressively that at least one intron, from any source, is a necessary prerequisite for correct processing and stabilization of eukaryotic mRNA from a gene which normally contains introns. On the other hand, interferon genes contain no introns, and yet are expressed normally by lymphocytes and fibroblasts. The regulation of splicing plays a crucial role in gene expression, as shown by the fact that a disturbance in splicing can have fatal results in humans. Comi *et al.* (1977) and Maquat *et al.* (1980) have described cases of thalassaemia in which nuclear globin RNA exists but its processing to functional mRNA is disturbed, leading to the accumulation of significantly less mRNA than for normal persons.

In spite of the enormous amount of new information concerning the structure of eukaryotic genes and their transcription products, our knowledge about the functional aspects of the regulation of transcription and post-transcriptional processing is still rather poor. We are certainly now going through a period of accumulation and systematization of

detailed structural facts. When correlated with functional studies of nucleic acid-protein interactions, and especially in conjunction with selective sequence alteration by site-directed mutagenesis, these structural facts will help to clarify some of the mysteries of gene function.

A major part of this review was written at the Department of Biochemistry, St Mary's Hospital Medical School, London, U.K., during an exchange fellowship of the Royal Society and the Academy of Sciences of the G.D.R. The author is indebted to Professor Robert Williamson for stimulating discussions and critical reading and editing of the manuscript.

References

- Bastos, R. N. & Aviv, H. (1977) *Cell* **11**, 641–650
- Bell, G. J., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer, E. & Goodman, H. M. (1980) *Nature (London)* **284**, 26–32
- Bernards, R., Little, P. F. R., Annison, G., Williamson, R. & Flavell, R. A. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4827–4831
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4853–4857
- Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. & Chambon, P. (1979) *Nature (London)* **282**, 567–574
- Comi, P., Giglioni, B., Barbarano, L., Ottolenghi, S., Williamson, R., Novokova, M. & Masera, G. (1977) *Eur. J. Biochem.* **79**, 617–622
- Crick, F. (1979) *Science* **204**, 264–271
- Curtis, P. J., Mantei, N., Van den Berg, J. & Weissmann, C. (1977a) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3184–3188
- Curtis, P. J., Mantei, N. & Weissmann, C. (1977b) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 971–984
- Darnell, J. E., Jelinek, W. & Molloy, G. (1973) *Science* **181**, 1215–1221
- Dugaiczyk, A., Woo, S. L. C., Lai, E. C., Mace, M. L., McReynolds, L. & O'Malley, B. W. (1978) *Nature (London)* **274**, 328–333
- Engelke, D. R., Ng, S. Y., Shastry, B. S. & Roeder, R. G. (1980) *Cell* **19**, 717–728
- Fraser, N. W., Nevins, J. R., Ziff, E. & Darnell, J. E. (1979) *J. Mol. Biol.* **129**, 643–666
- Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J. P., Benoist, C., Cochet, G., Breathnach, R., Royal, A., Garapin, A., Cami, B. & Chambon, P. (1979) *Nature (London)* **278**, 428–437
- Georgiev, G. P., Samarina, O. P., Lerman, M. I., Smirnov, M. N. & Severtzov, A. N. (1963) *Nature (London)* **200**, 1291–1294
- Georgiev, G. P., Ryskov, A. P., Coutelle, C., Mantieva, V. L. & Avakyan, E. R. (1972) *Biochim. Biophys. Acta* **259**, 259–283
- Hagenbüchle, O., Santer, M., Steitz, J. & Hans, R. J. (1978) *Cell* **13**, 551–568
- Halbreich, A., Pajot, P., Foucher, G., Grandchamp, C. & Slonimski, P. (1980) *Cell* **19**, 321–329

- Hamer, D. & Leder, P. (1979) *Cell* **17**, 737–747
- Harris, H. (1962) *Biochem. J.* **84**, 60–61
- Hofer, E. & Darnell, J. E. (1981) *Cell* **23**, 585–593
- Hunger, H. D., Grutzmann, H., Speer, A., Hanke, R., Hahn, V. & Coutelle, C. (1979) *Acta Biol. Med. Germ.* **38**, 1555–1563
- Imaizumi, T., Diggelmann, H. & Scherrer, K. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1122–1126
- Jeffreys, A. J. & Flavell, R. A. (1977) *Cell* **12**, 1097–1108
- Kinniburgh, A. J. & Ross, J. (1979) *Cell* **17**, 915–921
- Knapp, G., Ogden, R. C., Peebles, C. L. & Abelson, J. (1979) *Cell* **18**, 37–45
- Knochel, W. & Grundmann, U. (1978) *Biochim. Biophys. Acta* **517**, 99–108
- Konkel, D., Tilghman, S. M. & Leder, P. (1978) *Cell* **15**, 1125–1132
- Kwan, S. P., Wood, T. G. & Lingrel, J. B. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 178–182
- Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M. & Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6187–6191
- Lerner, M. R. & Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5495–5499
- Lomedico, P., Rosenthal, N., Efstradiatis, A., Gilbert, W., Kolodnes, R. & Tizard, R. (1979) *Cell* **18**, 545–558
- MacNaughton, M., Freeman, K. B. & Bishop, J. O. (1974) *Cell* **1**, 117–125
- Maquat, L. E., Kinniburgh, A. J., Beach, L. R., Honig, G. R., Lazeron, J., Ershler, W. B. & Ross, J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4287–4291
- Maundrell, K. & Scherrer, K. (1979) *Eur. J. Biochem.* **99**, 225–238
- Maxwell, E. S., Maundrell, K., Pavion-Dutilleul, F. & Scherrer, K. (1981) *Eur. J. Biochem.* **113**, 233–247
- Nordstrom, J. L., Roop, D. R., Tsai, M. J. & O'Malley, B. M. (1978) *Nature (London)* **278**, 328–331
- Pedersen, T. (1974) *J. Mol. Biol.* **83**, 163–183
- Peebles, C. L., Ogden, R. C., Knapp, G. & Abelson, J. (1979) *Cell* **18**, 27–35
- Pribnow, D. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 784–788
- Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214
- Roop, D. R., Nordstrom, J. L., Tsai, S. Y., Tsai, M. J. & O'Malley, B. M. (1978) *Cell* **15**, 671–685
- Roop, D. R., Tsai, M.-J. & O'Malley, B. W. (1980) *Cell* **19**, 63–68
- Rosenberg, G. & Paterson, B. G. (1979) *Nature (London)* **279**, 696–701
- Ross, J. (1978) *J. Mol. Biol.* **119**, 21–35
- Ross, J. & Knecht, D. A. (1978) *J. Mol. Biol.* **119**, 1–20
- Samarina, O. P., Lukanidin, E. M., Molnar, J. & Georgiev, G. P. (1968) *J. Mol. Biol.* **33**, 251–263
- Scherrer, K., Latham, H. & Darnell, J. E. (1963) *Proc. Natl. Acad. Sci. U.S.A.* **48**, 240–248
- Scherrer, K. & Marcaud, L. (1968) *J. Cell Physiol.* **72**, 181–212
- Sibatani, A., De Kloet, S. R., Allfrey, V. G. & Mirsky, A. E. (1962) *Proc. Natl. Acad. Sci. U.S.A.* **48**, 471–477
- Spohr, G., Imaizumi, T. & Scherrer, K. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 5009–5013
- Spohr, G., Dettori, G. & Manzari, V. (1976) *Cell* **8**, 505–512
- Stahl, D. A., Walker, T. A., Meyhack, B. & Pace, N. R. (1979) *Cell* **18**, 1133–1143
- Strair, R. K., Skoultchi, A. J. & Shafritz, D. A. (1977) *Cell* **12**, 133–141
- Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Matija Peterlin, B., Sullivan, M., Maizel, J. V. & Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 725–729
- Tooze, J. (1980) *The Molecular Biology of the Tumour Viruses*, vol. 2, Cold Spring Harbor, New York
- Weaver, R. F. & Weissmann, C. (1979) *Nucleic Acids Res.* **7**, 1175–1193
- Wickens, M. P. & Laskey, R. A. (1981) in *Genetic Engineering 1* (Williamson, R., ed.), pp. 103–167, Academic Press, London
- Williamson, R. (1973) *FEBS Lett.* **37**, 1–6
- Young, R. A., Hagenbüchle, O. & Schibler, U. (1981) *Cell* **23**, 451–458