Proc. Natl. Acad. Sci. USA
Vol. 78, No. 7, pp. 4471–4474, July 1981
Genetics

Novel models for RNA splicing that involve a small nuclear RNA

(mechanism of splicing/U2 RNA/U1 RNA/intron/exon)

YASUMI OHSHIMA†, MICHIYASU ITOH†, NORIHIRO OKADA†, AND TAKASHI MIYATA‡

†Institute of Biological Sciences, University of Tsukuba, Sakura-mura, Ibaragi-ken 305, Japan; and ‡Department of Biology, Faculty of Science, Kyushu University, Hakozaki, Fukuoka-City 812, Japan

Communicated by Motoo Kimura, March 31, 1981

ABSTRACT **Nucleotide sequences of mammalian small nuclear RNAs (snRNAs) have been analyzed with a computer program for complementarity with sequences around a splice junction of various eukaryotic mRNA precursors (pre-mRNAs). A region in U2 RNA or some other snRNAs can form base pairs with both exons surrounding an intron of certain pre-mRNAs and, thereby, can align the two junctions leading to correct splicing of the pre-mRNA. These findings suggest that a snRNA such as U2 can be involved in splicing certain pre-mRNAs by pairing with exons, which we call an "exon model" for splicing, as compared with the model involving U1 RNA presented by Lerner *et al.* [Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) *Nature (London)* 283, 220–224]. We constructed a secondary structure model of U1 RNA and studied the capacity of base pairing with pre-mRNAs on the basis of both primary and secondary structures of U1 RNA. We present an alternative model for splicing that involves U1 RNA, which assumes base pairing of noncontiguous regions of U1 RNA with an intron of a pre-mRNA. Pairing of an snRNA with exons could explain correct matching of the two junctions that bound one and the same intron, which is not explained by pairing with consensus sequences at the ends of an intron as proposed by Lerner *et al.* Pairing of an intron with U1 RNA and pairing of the surrounding exons with another snRNA such as U2 RNA could take place at the same time to insure specificity of splicing.**

It has been known for over a decade that eukaryotic nuclei contain a group of small stable RNAs (1). They are called small nuclear RNAs (snRNAs), and many results have accumulated, waiting for elucidation of their function (1, 2). Recently Lerner *et al.* (3) found a sequence in mammalian U1 snRNA that is complementary to consensus sequences at the intron–exon junction of eukaryotic precursor RNAs (pre-mRNAs) or heterogeneous nuclear RNAs. They presented a model for alignment of the two junctions for correct splicing, in which a region of U1 RNA can form base pairs with both ends of an intron of a pre-mRNA (3). This model seems to be very attractive and has drawn much attention, but it waits for experimental verification. Even if the model turns out to be true, the function of other snRNAs is still obscure. To surmise a function of other snRNAs and to look for other mechanisms of splicing involving U1 RNA, we studied complementarity or homology between the nucleotide sequence of a snRNA and that of a vertebrate gene or pre-mRNA. Our results suggest that U2 snRNA and U1 RNA are involved in splicing. We present novel models for splicing involving these snRNAs. Other snRNAs also might be involved in splicing, according to our results.

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.

METHODS

Computer Program. The program used for searching homology or complementarity between two RNA or DNA sequences was developed in Kyushu University and is written in FORTRAN (unpublished data). Its principal features are as follows. (*i*) The sequence of a snRNA, for example, is printed on the abscissa as the DNA sequence. (*ii*) The sequence of a gene or that around a splice junction of a gene is printed on the ordinate to search for homology. For a search of complementarity, the sequence is transcribed into the complementary one and printed also on the ordinate. (*iii*) The program is directed to print a mark (*) at the intersection of the abscissa and the ordinate if the base on the ordinate is homologous with that on the abscissa. (*iv*) If contiguous four or more bases share homology, the number of bases of contiguous homology is printed instead of the mark (*), resulting in a diagonal array of numbers that allows easy recognition of block homology or complementarity. ACOS 800 system 2 computer in the University of Tsukuba was mainly used for the results presented in this report.

Calculation of Free Energy for Base Pairing. Free energy for base pairing between two RNA sequences was calculated according to Tinoco *et al.* (4).

Construction of a Secondary Structure Model of U1 RNA. The sequence of U1 RNA in the 5′–3′ direction was written on the abscissa, and its complementary sequence in 3′–5′ direction was written on the ordinate. Intramolecular complementary regions were printed out as described. Several secondary structure models were constructed, allowing as many pairings as possible. The models were compared for stability or free energy and for the consistency with the reported distribution of nuclease-sensitive regions (5, 6). Thus, we finally chose the two models described in the text.

RESULTS

Sequence Complementarity Between U2 RNA and Exons (Exon Model). To surmise functions of snRNAs other than U1 RNA, we tried to look for homology or complementarity between a snRNA and each of various genes in vertebrates whose sequences are known. The first interesting result we obtained was the complementarity between a region in rat U2 snRNA (7) and both exon sequences just outside of an intron in mouse β-globin major gene (8). This resulted in a model shown in Fig. 1. In Fig. 1, the nucleotide sequence 126–146 of U2 RNA (7) can pair with both exon regions surrounding the first intron of mouse β-globin gene transcript or pre-mRNA. The nucleotide sequence of a snRNA seems to be severely conserved in mammals (3, 6), and the sequence of mouse U2 RNA must be quite

Abbreviations: snRNA, small nuclear RNA; pre-mRNA, mRNA precursor.

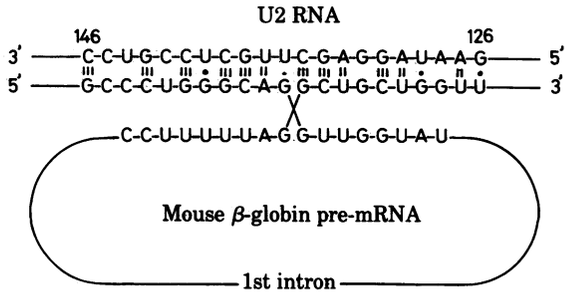

FIG. 1. Base pairing between U2 snRNA (7) and both exons surrounding the first intron of the mouse β-globin (major) pre-mRNA (8).

similar to or identical with that in rat. The same region in U2 RNA can pair also with exons surrounding the second intron of the mouse β-globin pre-mRNA (Fig. 2). We then restricted our attention to sequences around a splice junction and looked for complementarity with a sequence in U2 RNA and other snRNAs. Fig. 2 shows various cases in which significant base pairing between both exons surrounding an intron and a region in U2 RNA is possible. In all cases shown in Fig. 2, the same region in U2 RNA is involved in predicted base pairing. Fig. 2 also shows the free energy of base pairing, neglecting the effect of an intron as a bulge loop (4). In the most stable case, mouse α-globin second intron, the energy is −28.2 kcal, and a very extensive base pairing can occur. We call these models "exon models" because they involve base pairing with exons in contrast to the model by Lerner *et al.* (to be called an "intron model"). Exon models can also be built for other snRNAs—4.5S RNA (15), U1 RNA, or RNA-bound 4.5S RNA (16)—with certain pre-mRNAs (Fig. 3). An efficient exon model has already been proposed for VA RNA-I with hexon pre-mRNA of adenovirus (20) and for 4.8S (U6) nuclear RNA with mouse IgGλI pre-mRNA (21).

Secondary Structure Model of U1 RNA. The sequence of U1 RNA has been reported for rat Novikoff hepatoma cells by Reddy *et al.* (5). Recently the sequences of U1 RNA from rat brain, man, and chicken have been reported by Branlant *et al.* (6). The sequence of rat U1 RNA is different in several regions from that published previously. We have included U1 RNA in the search for complementarity with splice junctions in various pre-mRNAs by initially using the sequence of Reddy *et al.* (5) and recently using that of Branlant *et al.* (6). We confirmed complementarity of the 5'-end region of U1 RNA (nucleotides 1–22) with the ends of an intron of an individual pre-mRNA as proposed by Lerner *et al.* (3). We also noticed that a region in U1 RNA near the 3' end (nucleotides 131–144 of the revised sequence) has complementarity with the 3' end of an intron, which complementarity, with many pre-mRNAs, is better than or as good as that with nucleotides 12–22 proposed by Lerner *et al.* If this second complementary region in U1 RNA (nucleotides 131–144) occupies adjacent spacial position to the 5'-end region, it could join in base pairing with an intron of many pre-mRNAs in a more stable manner than could the region 12–22. This possibility led us to construct a secondary structure model and to examine its probability on that basis. The secondary structure model based on the revised sequence is shown in Fig. 4*a*. Fig. 4*b* shows an alternative model of approximately equal stability, which is different only at the two terminal regions. Both models are quite similar in base-pairing characteristics and are consistent with the distribution of the sites sensitive to partial digestion with RNases (5) or S1 nuclease (6). There are only two base changes in human U1 RNA and seven base changes in chicken U1 RNA from rat U1 RNA (6), which seem to exert little effect on the secondary structure shown in Fig. 4. The severe conservation of the U1 RNA sequence through the whole molecule may indicate that not only the primary structure near the 5' end but also the whole structure of U1 RNA is important for its function.

Base-Pairing Capacity of U1 RNA and a Novel Intron Model. In the models shown in Fig. 4, nucleotides 1–12 at the 5' end of U1 RNA do not form strong intramolecular base pairing. Therefore, this region can probably be involved in base pairing with an intron of a pre-mRNA as predicted by Lerner *et al.* (3), although unknown modified nucleotides at positions 6 and 7 (U*) might not form base pairs. Nucleotides 13–22 form more stable base pairs with nucleotides 120–123 and 43–48

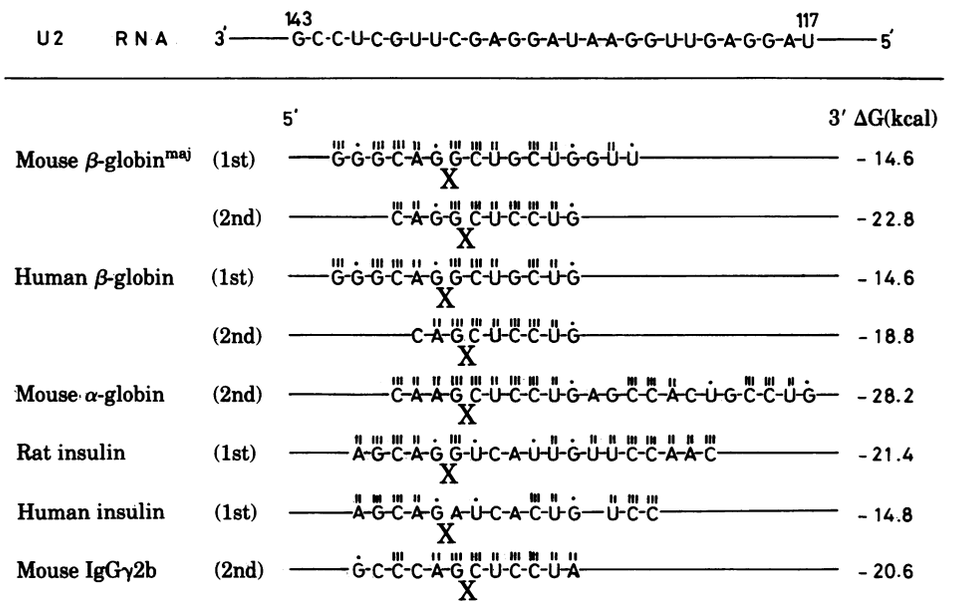

FIG. 2. Base pairing between U2 RNA and exons surrounding an intron of various pre-mRNAs. The intron to be excised is specified in the parentheses. The exon sequences of pre-mRNAs shown are taken from the corresponding gene sequences: mouse β-globin (major) (8), human β-globin (9), mouse α-globin (10), rat or human insulin (11, 12), and mouse IgGγ2b chain (13, 14).

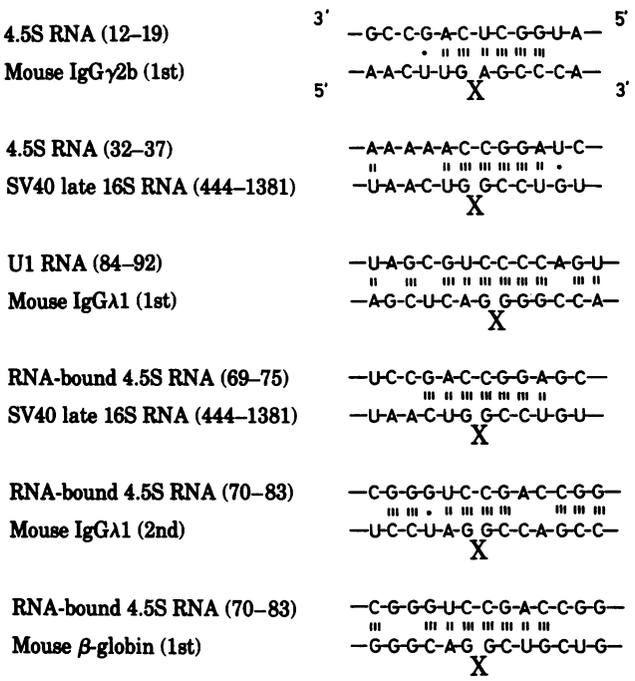

FIG. 3. Exon models with various snRNAs. In the parentheses after a snRNA species, the nucleotides involved in base pairing are shown, and in the parentheses after a pre-mRNA, the intron to be excised is specified. The exon sequences of the pre-mRNAs are taken from the following references: mouse IgGγ2b (13, 14), SV40 late 16S RNA (17, 18), mouse IgGγ1 (19), and mouse β-globin (8).

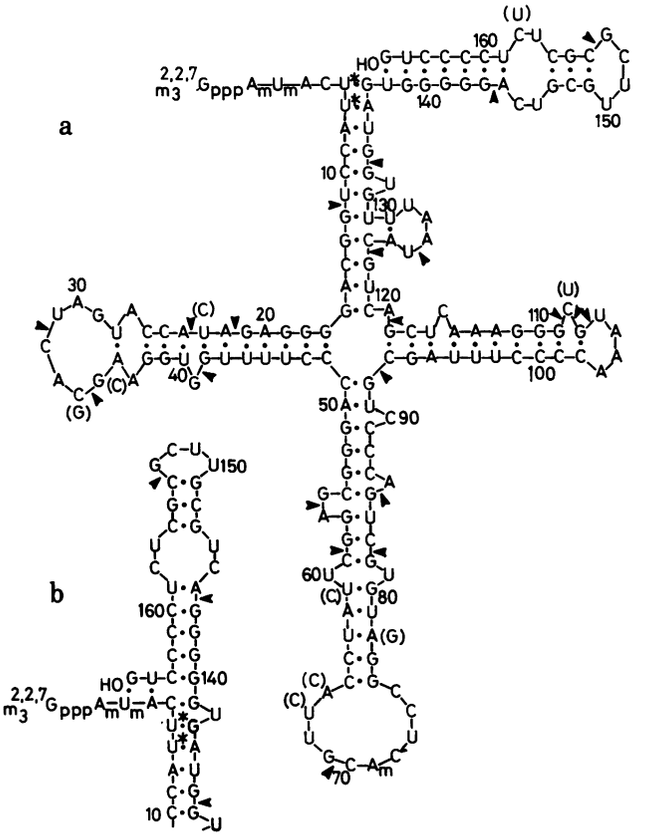

FIG. 4. Secondary structure models of rat U1 RNA. The models are constructed as described, with *b* showing only the 5′- and 3′-end regions because the other regions have the same structure as in *a*. Nucleotide changes observed in chicken or human U1 RNA (6) are shown in the parentheses. Arrows in the figure indicate the sites sensitive to partial digestion with RNases (5) or S1 nuclease (6).

(Fig. 4), and this region may have some difficulty in pairing with the 3′ end of an intron, unless the intermolecular pairing is strong enough. Concerning nucleotides 131–144, which are complementary to the 3′ end of many introns, nucleotides after 137 probably cannot be involved in intermolecular pairing because this region is very resistant to ribonucleases (6). However, nucleotides 131–137 are free to form intermolecular pairing when the 5′-end region dissociates from this region for pairing with an intron of a pre-mRNA (Fig. 4). Fig. 5 shows such a partially dissociated structure of U1 RNA pairing with an intron in various pre-mRNAs as examples. If nucleotides 138–144 are free, more extensive pairing with the intron is possible in some cases, but this probably will not be the case. The models in Fig. 5 may be called intron models involving noncontiguous regions of U1 RNA. We calculated the free-energy of predicted base pairing between U1 RNA and each of about 30 introns and compared the free energies with those predicted by the model of Lerner *et al.* (3). With about half of the introns, including those in Fig. 5, our present model gives more stable base pairing than that by Lerner *et al.* In the rest of the introns, their model seems better. We suggest that, at least in some cases, base pairing between U1 RNA and an intron of the type proposed here (Fig. 5) will take place.

DISCUSSION

We found an extensive complementarity between a region in U2 RNA and both exons adjacent to an intron of certain pre-mRNAs. We also found similar complementarity between a region in other snRNAs and exons of a certain pre-mRNA. These findings led us to propose a model for splicing that involves a snRNA such as U2 RNA, in which both exons are aligned by base pairing with the RNA to excise the intron loop (Figs. 1–3). This model, designated exon model, is different from the model proposed by Lerner *et al.* (3), in which both ends of an intron are aligned by pairing (designated intron model).

Possible significant features of the exon model are discussed. First, an exon model can be more effective than an intron model for joining of exons because exons remain aligned after the intron is excised.

Second, sequences at both ends of an intron roughly agree with the consensus for most of the introns and, therefore, U1 RNA, which has a complementary region to the "consensus sequence," can play a universal role in splicing. However, U1 RNA alone cannot choose the correct pair of junctions (donor- and acceptor-sites) bounding the same intron just because the ends of an intron agree with the consensus (24). Usually a pre-mRNA in eukaryotes has two or more introns, and so U1 RNA can align the donor site of an intron and the acceptor site of another intron in the same precursor molecule as efficiently as in the correct alignment. On the other hand, because exons surrounding an intron are generally different from each other, alignment of exons by pairing with a snRNA such as U2 RNA, could explain correct matching of the donor site and the acceptor site.

Third, most of the introns studied can pair with U1 RNA either in the way proposed by Lerner *et al.* (3) or in the way proposed here (Fig. 5). Therefore, as for an intron where an exon model with another snRNA molecule such as U2 RNA is valid, base pairing of the intron with U1 RNA and that of adjacent exons with the second snRNA can probably take place at the same time. It may be a very favorable situation for correct splicing to occur. Thus, an exon model and an intron model do not necessarily exclude each other but can work in cooperation for precise and biologically significant splicing.

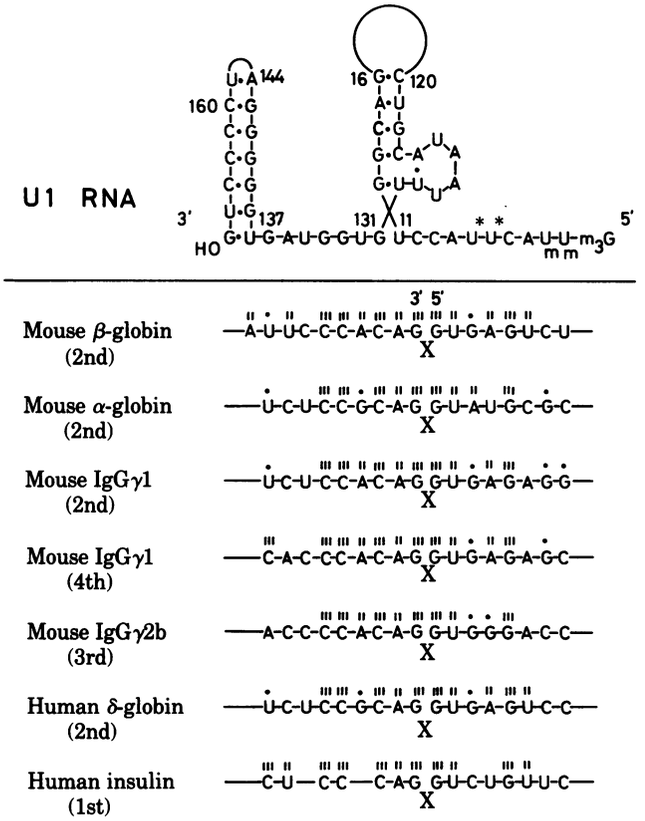

FIG. 5. Models for base pairing between U1 RNA and an intron in various pre-mRNAs. Intron sequences are shown below the line: only the 5′ end (on the right side) and the 3′ end (on the left side) are drawn, but both are, in fact, contiguous, making an intron loop as shown in Fig. 1. The intron sequences are taken from the following references: mouse α-globin (10), mouse IgGγ1 (21, 22), mouse IgGγ2b (13, 14), human δ-globin (23), and human insulin (12).

Fourth, exon sequences are gene-specific, and so only certain class of exons or pre-mRNAs can be involved in pairing with a region in an snRNA. Therefore, a snRNA such as U2 RNA could regulate splicing in a gene-specific or group-specific manner by pairing with a certain class of pre-mRNAs.

Our present models and the model by Lerner *et al.* (3) are the models for one-step splicing of an intron. On the other hand, there are some evidences for multi-step excision or splicing of a single intron (25–27). However, all of these results do not seem to exclude the existence of one-step excision or splicing. Probably one-step splicing for an intron is occurring as well, even in these cases. It is likely that most of the splicing for an intron takes place as a single step with small introns and with the introns for which an efficient model for one-step splicing is valid.

In conclusion, it is suggested that U2 RNA and U1 RNA take part in splicing. 4.5S RNA, RNA-bound 4.5S RNA, and U6 RNA also might be involved in splicing. We believe that one or a combination of these snRNAs are involved in splicing of a specific pre-mRNA by pairing with either an intron, exons, or both.

We would like to thank Prof. Yoshio Watanabe and other members of the Institute of Biology for constant hospitality and encouragement.

1. Weinberg, R. (1973) *Annu. Rev. Biochem.* **42**, 343–354.
2. Zieve, G. & Penman, S. (1976) *Cell* **8**, 19–31.
3. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) *Nature (London)* **283**, 220–224.
4. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246**, 40–41.
5. Reddy, R., Ro-Choi, T. S., Henning, D. & Busch, H. (1974) *J. Biol. Chem.* **249**, 6486–6494.
6. Branlant, C., Krol, A., Ebel, J-P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. & Jeanteur, P. (1980) *Nucleic Acids Res.* **8**, 4143–4154.
7. Shibata, H., Ro-Choi, T. S., Reddy, R., Choi, Y. C., Henning, D. & Busch, H. (1975) *J. Biol. Chem.* **250**, 3909–3920.
8. Konkel, D. A., Tilghman, S. M. & Leder, P. (1978) *Cell* **15**, 1125–1132.
9. Smithies, O., Blechl, A. E., Denniston-Thompson, K., Newell, N., Richards, J. E., Slightom, J. L., Tucker, P. W. & Blattner, F. R. (1978) *Science* **202**, 1284–1287.
10. Nishioka, Y. & Leder, P. (1979) *Cell* **18**, 875–882.
11. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizard, R. (1979) *Cell* **18**, 545–558.
12. Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tisher, E. & Goodman, H. M. (1980) *Nature (London)* **284**, 26–32.
13. Tucker, P. W., Marcu, K. B., Newell, N., Richards, J. & Blattner, F. R. (1979) *Science* **206**, 1303–1306.
14. Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M. & Honjo, T. (1980) *Nature (London)* **283**, 786–789.
15. Ro-Choi, T. S., Reddy, R., Henning, D., Takano, T., Taylor, C. W. & Busch, H. (1972) *J. Biol. Chem.* **247**, 3205–3222.
16. Harada, F. & Kato, N. (1980) *Nucleic Acids Res.* **8**, 1273–1285.
17. Ghosh, P. K., Reddy, V. R., Swiscoe, J., Lebowitz, P. & Weissman, S. M. (1978) *J. Mol. Biol.* **126**, 813–846.
18. Reddy, V. B., Thimmapaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. & Weissman, S. M. (1978) *Science* **200**, 494–502.
19. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* **15**, 1133–1144.
20. Murray, V. & Holliday, R. (1979) *FEBS Lett.* **106**, 5–7.
21. Harada, F., Kato, N. & Nishimura, S. (1980) *Biochem. Biophys. Res. Commun.* **95**, 1332–1340.
22. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
23. Spritz, R. A., DeRiel, J. K., Forget, B. G. & Weissman, S. (1980) *Cell* **21**, 639–646.
24. Lewin, B. (1980) *Cell* **22**, 324–326.
25. Kinniburgh, A. & Ross, J. (1979) *Cell* **17**, 915–921.
26. Reymond, C., Agliano, A. & Spohr, G. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4683–4687.
27. Avvedimento, V. E., Vogeli, G., Yamada, Y., Maizel, J. V., Jr., Pastan, I. & deCrombrugghe, B. (1980) *Cell* **21**, 689–696.