Proc. Natl. Acad. Sci. USA
Vol. 75, No. 11, pp. 5627–5630, November 1978
Genetics

Evidence for translocation of DNA sequences during sea urchin embryogenesis

(inverted repeat sequences/hairpins/base pair mismatches/methylation/cell differentiation)

DAVID G. DICKINSON AND ROBERT F. BAKER*

Molecular Biology Division, University of Southern California, Los Angeles, California 90007

Communicated by Martin D. Kamen, August 21, 1978

ABSTRACT **Hairpin-like DNA was prepared *in vitro* from the family of sequences that are inverted relative to each other and, as pairs, are relatively homologous and adjacent on the sea urchin genome. The majority of these hairpins are shown to have base pair mismatch positions distributed along their stems. Comparison of the hairpins derived from the DNA of morula, blastula, and gastrula stage embryos shows that during embryogenesis there are changes in the average number and position of S1 nuclease-sensitive base pair mismatch sites on the majority of the hairpin stems. Our data indicate that during early embryogenesis there are sequence changes *in vivo* within the majority of the adjacent inverted repeat sequences of the sea urchin genome. We have also found that there is higher specificity for the occurrence of sequence-change events within that fraction of the inverted repeat sequences that are methylated *in vivo*.**

Evidence is presented here that suggests that there are translocations of inverted repeat DNA sequences within the sea urchin genome during early embryogenesis. The method used for testing for translocations is sensitive to sequence changes within those pairs of contiguous or nearly contiguous (adjacent) sequences that are inverted relative to each other: when isolated DNA is denatured and allowed to renature, intrastrand hydrogen bonding of adjacent homologous inverted sequences on the same DNA strand will cause pairs of these sequences to fold back on themselves and form hairpin-like structures (1–4). We have found that the majority of the hairpins formed *in vitro* from the sea urchin genome have a small number of base pair mismatch positions distributed along their stems; the average distance between those mismatch sites that can be cut by *Aspergillus oryzae* S1 nuclease has been measured. Most importantly, we have found that additional mismatch positions are placed into the majority of the derived hairpin stems as embryogenesis proceeds, indicating sequence changes *in vivo*.

Biologically, these sequence changes during embryogenesis could have resulted from (*i*) hypermutation within inverted repeat regions or (*ii*) translocations resulting in exchanges of part or all of either the left or right sequence of the majority of adjacent pairs of inverted homologous sequences. The latter mechanism is highly favored over hypermutation as an explanation of the data. We have also found that there is specificity within the genome for sequence changes within adjacent inverted repeat sequences that are methylated at cytosine positions.

MATERIALS AND METHODS

Isolation of Sea Urchin Embryo DNA. *Strongylocentrotus purpuratus* embryos were cultured and their DNA was isolated as described (5). DNA was labeled *in vivo* by addition of L-[*methyl*-^{3}H]methionine (10.7 Ci/mmol, 1.0 μCi/ml) or $H_3[^{32}P]PO_4$ (<0.1 μCi/ml) to the sea water media in which the embryos were grown. Labeling with either isotope was continuous from fertilization until termination of development. Control studies have shown that the results of the experiments reported here are not affected by any time-course changes in the specific activity of the radiolabels during development; labeling during only the last two cell divisions prior to termination of development gave the same experimental results as did continuous labeling from the time of fertilization until termination.

Preparation of Hairpin DNA. Isolated DNA from embryos was sheared to an average size of approximately 3500 base pairs, diluted to <1 μg/ml, denatured in alkali, and allowed to renature at pH 6.8 in 0.15 M sodium phosphate for less than 1 min at room temperature; the double-stranded portion of this DNA was removed by a batch process using hydroxyapatite [all procedures were essentially as described (6)]. The renatured DNA was twice more denatured and renatured in the same manner at the resulting successively lower concentrations, each time the renatured duplex DNA being collected on hydroxyapatite.

Nuclease Treatment of Hairpin DNA. *Escherichia coli* exonuclease VII was kindly provided by J. Chase and C. Richardson and used under the conditions specified (7, 8). *Aspergillus oryzae* S1 nuclease (9) was isolated according to the method of Sutton (10). S1 nuclease treatment of DNA was in 0.2 M NaCl/0.03 M Na acetate pH 4.7/1 mM $ZnSO_4$ as described (11).

The sample of S1 nuclease used in these experiments was shown not to break duplex ϕ80 or λ bacteriophage DNA under conditions sufficient for complete hydrolysis of an equivalent amount of single-stranded DNA (data not shown). Neither the turnarounds nor the base pair mismatch positions along the stems of the hairpins are cut by exonuclease VII. This has been shown by control experiments utilizing native morula DNA having gaps (5) and by the ability to denature and refold hairpin DNA after incubation with exonuclease VII.

Electrophoresis of DNA. DNAs were electrophoresed on agarose or acrylamide gels as indicated in the figure legends. Tris/EDTA/borate buffer (12) was used as running buffer in both types of electrophoresis. Appropriate molecular weight marker DNAs were mixed with the radiolabeled DNAs prior to electrophoresis. Control experiments showed that the marker DNAs did not influence the distribution profiles of radioactive DNAs displayed on the gels. Alignment of patterns from separate gels was critical to the interpretation of the experiments; this alignment was effected by staining the gels for 20 min with 1 μg of ethidium bromide per ml of water and then visualizing the positions of molecular weight marker DNAs by fluorescence under UV light. The gels were sliced and assayed for radioactivity, and equivalent molecular weight marker positions for the different gels were put in register on the abscissa of the

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.

* To whom reprint requests should be addressed.

graphs. By slicing each gel at a position determined by a marker and then continuing to slice the gel in each direction from the initial slice position, the three radioactivity profiles shown on Figs. 1 and 2 were put into almost exact register with molecular weight of DNA over the long section of each gel that was of most interest in comparisons of length distributions of radiolabeled DNAs.

RESULTS

Hairpin DNA was prepared from isolated sea urchin DNA by three rounds of opening and refolding the hairpins. The derived hairpin DNA used in these experiments need not be composed entirely of hairpin-forming sequences in order for the assay for translocation events to be valid. As indicated in the *Discussion*, the presence of other sequences (e.g., bimolecularly reassociating highly repetitive sequences or fast reassociating interstrand crosslinked DNA) as a contaminating fraction of the derived hairpin DNAs would not interfere with the assay. However, the hairpin DNA used in the assays did display the properties expected of folded pairs of sequences that are nearly homologous, inverted relative to each other, and closely adjacent on the same strand of DNA: after three rounds of preparative renaturation, the yield of *A. oryzae* S1 nuclease-resistant DNA or *E. coli* exonuclease VII-resistant DNA was essentially constant when the DNA concentration during renaturation was varied over a 10-fold range (the data for this concentration-independent renaturation are similar to those shown in table 1 of ref. 6). The mean (± range) yield of exonuclease VII-resistant DNA *or* S1 nuclease-resistant DNA was 3.0 ± 0.11% of bulk morula DNA, 3.1 ± 0.13% of blastula DNA, and 2.9 ± 0.10% of gastrula DNA, based upon three preparations of DNA from each of the three stages of development.

After the hairpins were trimmed of their single-stranded tails by incubation with *E. coli* exonuclease VII (7, 8), the distribution of lengths of the resulting hairpin population was measured by electrophoresis on agarose gels. The three gel patterns shown in Fig. 1 have been normalized on the ordinate of the graph at 1020 base pairs and have been put into register on the abscissa of the graph with molecular weight markers run on the same gels (as described in *Materials and Methods*) to allow comparison of the length distribution of the morula, blastula, and gastrula hairpins. Some of the hairpins could, perhaps, contain hairpins within hairpins; therefore, the DNA pieces displayed on these gels may not be migrating during electrophoresis as linear duplex pieces. However, because exonuclease VII is both a 3′ and a 5′ exonuclease (7, 8), the DNA sizes displayed on gels should reflect the total stem length, the small amount of turnaround sequence, and any nonhomologous sections on each hairpin. In any case, the purpose of this experiment was to compare the length distributions of exonuclease VII-treated foldback DNAs resulting from the bulk DNAs isolated from embryos at various stages of development. The population of sizes of hairpin DNA did not change measurably during early embryogenesis. This experiment also suggests that preferential loss or amplification of portions of the genome containing adjacent inverted repeat sequences does not occur either *in vivo* during embryogenesis or during isolation of DNA from the different stage embryos.

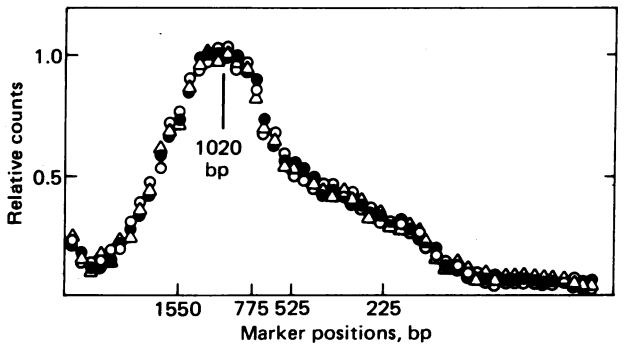

FIG. 1. Distribution of sizes of sea urchin embryo hairpin [^{32}P]-DNA after trimming with *E. coli* exonuclease VII. Subsequent to this enzyme treatment to the limit of digestion, the hairpin DNA was mixed with unlabeled simian virus 40 DNA that had been separately restricted with *Hae* III endonuclease in order to provide internal molecular weight markers (13, 14) on the gels. After electrophoresis of the DNA on 2.4% agarose, the gels were stained with ethidium bromide to identify, by UV-induced fluorescence, the marker positions; the gels were then sliced and assayed for radioactivity. The corresponding gel slices for each of the three stages of development were critically aligned according to the marker DNA positions shown and the slice position was plotted on the abscissa of the graph; the position having the maximum counts (and containing DNA 1020 base pairs in length) for each developmental stage was plotted on the ordinate of the graph as having 1.0 relative counts: bp, base pairs. ○, morula; ●, blastula; △, gastrula.

When hairpin DNA was incubated with *A. oryzae* S1 nuclease, the average length of the surviving double-stranded pieces was reduced to less than 10% of the length of the same population of hairpins whose single-stranded tails had been trimmed with exonuclease VII (compare Figs. 1 and 2). The reduction, caused by S1 nuclease, in the apparent length of hairpin stems indicates that these hairpins have an average of 12–18 S1 sites along their stems, depending upon the stage of embryogenesis. S1 nuclease is a single strand-specific endonuclease that will cleave double-stranded DNA at positions that have some minimum number of base pairs that are not homologous. We do not know the minimum number of mismatched base pairs necessary in order for S1 nuclease to cut the stem of a hairpin at a particular site; however, the size pattern of S1 pieces seen on gels does not depend upon the degree of S1 digestion over a range of 5 to 100% of the enzyme activity required to digest an equivalent amount of single-stranded DNA (data not shown; ref. 6). Even so, we have used the same reaction conditions to generate the S1 pieces from the hairpins of the different stages of development.

In comparing the isolated DNAs from different stages of sea urchin embryogenesis, the lengths of S1 pieces derived from hairpins should remain the same if the sequences within the inverted repeats *in vivo* have not changed during the course of embryogenesis. However, we found that the size patterns of S1 pieces on gels systematically changes with development; as development progressed, the average length of the S1 pieces became shorter (Fig. 2).

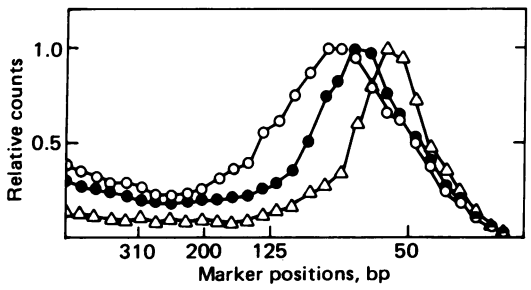

FIG. 2. Distribution of sizes of sea urchin embryo hairpin [^{32}P]-DNA after treatment with *A. oryzae* S1 nuclease to the limit of digestion. The S1-treated DNAs were separately mixed with unlabeled simian virus 40 DNA that had been previously restricted with *Hae* III and *Eco*RII endonucleases (13, 14); the mixtures were electrophoresed independently on 7% acrylamide gels. The internal marker positions were identified, the gels were sliced and assayed, and the data were plotted as in Fig. 1. ○, morula; ●, blastula; △, gastrula.

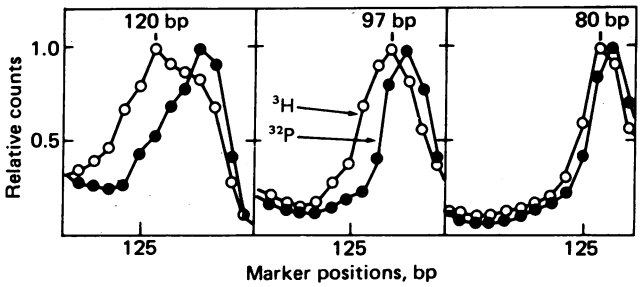

FIG. 3. Distribution of sizes of sea urchin embryo hairpin [^{3}H]-methyl- and ^{32}P-labeled DNA after treatment with *A. oryzae* S1 nuclease to the limit of digestion. The S1-treated DNAs were separately mixed with unlabeled simian virus 40 DNA that had been previously restricted with *Hae* III and *Eco*RII endonucleases (13, 14); the mixtures were electrophoresed independently on 7% acrylamide gels. The internal marker positions were identified, the gels were sliced and assayed, and the data were plotted as in Fig. 1. (*Left*) Morula; (*Center*) blastula; (*Right*) gastrula. ○, ^{3}H; ●, ^{32}P.

We also examined the distribution of sizes of the pieces resulting from S1 nuclease digestion of hairpin DNA that was methylated *in vivo* during embryogenesis. Hairpin DNA, double-labeled *in vivo* with [^{3}H]methionine and ^{32}P, was isolated from embryo DNA. [^{3}H]Methionine has been shown to label sea urchin DNA during embryogenesis by adding [^{3}H]-methyl groups to those cytosine bases that are normally methylated after DNA replication (15, 16). The distribution of sizes as well as the average size of the methylated S1 pieces changed during early embryogenesis and does so in a manner different from that of the general population of S1 pieces derived from the ^{32}P-labeled hairpins (Fig. 3).

DISCUSSION

The distribution of lengths of the family of hairpins derived from inverted repeat DNA sequences remains constant during early development (Fig. 1) and yet the number and position of S1 nuclease-sensitive base pair mismatch sites along the stems of the majority of the hairpins change during the course of embryogenesis (Figs. 2 and 3). There are three explanations that could account for these observations: (*i*) There are translocations during early embryogenesis of sequences within at least those inverted repeat sequences that can be folded *in vitro* to make hairpin DNA. (*ii*) Inverted repeat sequences undergo a high rate of accumulative mutation whereas informational gene sequences must be spared this fate. Any mechanism generating hypermutation would necessarily be highly specific for *methylated* adjacent inverted repeat sequences in order to satisfy the data shown in Fig. 3. Considering the calculation at the end of the *Discussion*, a hypermutation mechanism with or without this specificity seems very unlikely. (*iii*) There is preferential loss or gain of certain species of hairpins or other S1 nuclease-sensitive pieces of DNA that are isolated along with the hairpins. This condition might have resulted from stage-dependent loss of some components of the genome during isolation of the DNA, amplification or deletion *in vivo* of a portion of the inverted repeat DNA sequences, an increase in the rate of breakage of DNA strands (and, therefore, a decrease in the average length of hairpin stems and S1 pieces) as a result, for example, of greater ^{32}P incorporation or radio decay events in the DNA samples isolated from progressively older embryos, or an increase in the number of interstrand crosslinks in the duplex DNA as embryogenesis proceeds; isolated hairpin DNA from subsequent stages might contain an increasing proportion of fast-renaturing crosslinked DNA strands that behave like hairpin DNA during isolation.

However, this third explanation, including all of the contributing factors listed, could not explain the patterns seen on the gels shown in Fig. 1, 2, or 3. First, Fig. 1 shows that the spectrum of lengths of exonuclease VII-trimmed hairpin DNA does not change in DNA samples from the different stages of embryogenesis. This strongly suggests that there is no change in the spectrum of lengths of adjacent inverted repeat sequences *in vivo* or in the components within isolated DNAs in the different stages of embryogenesis. Second, the proportion of the bulk DNA that can be converted into exonuclease VII-trimmed hairpin DNA remains relatively constant (about 3% of the starting DNA by our measurement) in morula, blastula, and gastrula DNAs. Third, and by far the most important control, an internal standard shows that, if *any* of the previously listed subpoints under the third explanation are in fact occurring, they could not change the primary conclusion that there are sequence changes within inverted repeats during development. This control is the simultaneous double labeling (with [^{3}H]-methionine and ^{32}P, *in vivo*) of those sequences that become hairpin DNA *in vitro*. The ^{32}P should uniformly label all of the hairpins at any stage of development; as evident in Fig. 3, the ^{3}H nonrandomly labels the hairpins and does so (15, 16) at positions of methylation. For each of the three stages of embryogenesis, the ^{3}H and ^{32}P separate differently on the electrophoretic gels carrying the double-labeled S1 pieces derived from the hairpins. Because, at any stage, ^{3}H is labeling a subset of the total S1 pieces labeled by ^{32}P, a shift during development in the size pattern of ^{3}H-labeled S1 pieces *relative* to ^{32}P-labeled S1 pieces must mean that there are sequence changes *in vivo* in the methylated inverted repeat sequences that do not occur in the nonmethylated inverted repeat sequences.

The presence of base pair mismatches along the lengths of the DNA pieces isolated as hairpin stems argues strongly that these DNA pieces are indeed hairpins and arose *in vitro* by fold-back of adjacent inverted repeat sequences on the same DNA strand rather than by renaturation of interstrand crosslinked DNA. If the putative hairpin DNA were actually formed by pairs of strands held in register by crosslinks (introduced into duplex DNA *in vivo* or *in vitro*), then S1-sensitive mismatch positions would not be expected along the lengths of the majority of the renatured duplex pieces. Any mismatches arising *in vivo* along crosslinked DNA would normally have been repaired. If some of the population of isolated hairpin DNA is actually interstrand crosslinked DNA, it could not interfere with the analyses of the shift during embryogenesis in the position or number of mismatches along the stems of the derived hairpins. This follows because there should be no S1-sensitive mismatch sites in crosslinked DNA held in register during renaturation *in vitro*. This latter point is shown particularly for the methylated hairpin DNA; if the increase during embryogenesis in the number of S1 sites were the result of an increase in the number of crosslinks in the genome, the data shown in Fig. 3 would also require that preferential and increasing methylation occur at or near new crosslink sites that were S1 sensitive, even though the sites should not contain base pair mismatches. Thus, it seems unlikely that the data could be explained by an increase in the number of crosslinks as embryogenesis progressed. A more trivial argument is that the isolated hairpin DNA contained a component resulting from fast renaturation of crosslinked or highly repetitive DNA and that the proportion of this component in the isolated hairpin DNA increased with further embryogenesis. If this were true the distribution of sizes of exonuclease VII pieces shown in Fig. 1 should shift with embryogenesis. This was not seen.

The presence of S1-sensitive base pair mismatch sites in hairpin DNA preparations also requires that at least the relevant portion of the DNA pieces being compared on gels (that is, those

pieces which *are* sensitive to S1 nuclease) could not be fast-renaturing highly repetitive (e.g., satellite) DNA. If this latter type of DNA were present as a component of the hairpin DNA preparations, it seems unlikely that this component would have S1-sensitive base pair mismatch positions that are nonrandomly spaced (on the order of 60–80 base pairs apart) or developmental stage-specific.

The results shown in Figs. 2 and 3 might also be explained if genomic DNA becomes modified as embryogenesis progresses and S1 nuclease has preference for this modified DNA. However, to our knowledge, S1 nuclease has not been shown to have preference for modified DNAs (e.g., methylated DNA). As described in the *Results*, S1 nuclease digestions of morula or gastrula hairpin DNA over a wide range of total nuclease activity per unit amount of DNA did not alter the distribution of sizes of S1 pieces seen on gel profiles. This suggests that, if embryo DNA is modified during development, any preference that S1 might have for the modified DNA is not a matter of efficiency of the enzyme for its substrate over the range tested. A more compelling argument against S1 being selective for some modified form of DNA is that replacement *in vivo* of thymidines in DNA by 5-BrdUrd (17–19) does not allow the shifts during development in the sizes of S1 pieces shown in Figs. 2 and 3 even though [^{3}H]methionine labeling of DNA proceeds at the same rate in control embryos as in embryos grown in BrdUrd (unpublished data). This suggests that, unless BrdUrd prevents the hypothetical S1-sensitive, developmentally programmed modification of DNA, S1 nuclease has no special preference for this modified DNA. The corollary is that culturing embryos in medium containing BrdUrd inhibits translocations of DNA sequences.

It is interesting that S1 treatment of hairpin stems does not break them into random length pieces. Figs. 2 and 3 show that the dominant size of S1 pieces is 60–80 base pairs. This suggests that base pair mismatch positions tend to repeat at this length along hairpin stems which may have an average length of about 1000 base pairs, as is suggested by the data shown in Fig. 1. Because the mismatch positions would not be present *in vivo* and are created *in vitro* when the inverted repeats are folded to form hairpins, one might hypothesize that the longer inverted sequences may have evolved by clustering of shorter length sequences at various positions throughout the genome.

Taken together, the best explanation for the results shown in Figs. 1–3 would seem to be that, during embryogenesis, there are translocations or reorganizations of DNA sequences *at least* within inverted repeat sequences. In principle these events could include sequences from any part of the genome. These present data do not allow us to specify the terminal positions of sequences undergoing translocation. However, Fig. 1 shows that the lengths of the stems are constant in hairpins resulting from inverted repeats that are undergoing translocation events. This consideration allows the speculation that the positions of the centers of the 2-fold axes of symmetry do not shift within the genome during embryogenesis. Excision of a sequence from within an inverted repeat should be followed by integration of another sequence as long as or longer than the sequence that was removed. Therefore, it is reasonable that an individual translocating sequence could extend in one direction beyond an inverted repeat sequence region.

A minor portion of the adjacent inverted repeat sequences is methylated *in vivo* and undergoes a high rate of sequence change compared with the general population of adjacent inverted repeat sequences (Fig. 3). It is not possible to determine from this present data whether methylation of specific DNA sequences is obligatory to DNA sequence translocation.

During the period of development between morula and gastrula stages, a minimum of 10^4 translocation events per haploid genome per cell division would need to have taken place to account for the observed shifts in sizes of S1 pieces. This calculation is made by considering (*i*) the average length of the exonuclease VII-trimmed hairpin DNA, (*ii*) the proportion of the sea urchin genome that will form hairpin DNA, (*iii*) the number of cells per embryo at the morula and gastrula stages, (*iv*) the average number of S1 sites per derived hairpin at gastrula stage compared with morula stage, and (*v*) the amount of DNA per haploid genome.

We thank Drs. J. Chase and C. Richardson for a generous supply of *E. coli* exonuclease VII. This study was supported by Grant HD04015 from the National Institutes of Health.

1. Graham, D. E., Neufeld, B. R., Davidson, E. H. & Britten, R. J. (1974) *Cell* **1**, 127–137.
2. Schmid, C. W., Manning, J. E. & Davidson, N. (1975) *Cell* **5**, 159–172.
3. Wilson, D. A. & Thomas, C. A., Jr. (1974) *J. Mol. Biol.* **84**, 115–144.
4. Cech, T. R. & Hearst, J. E. (1975) *Cell* **5**, 429–446.
5. Case, S. T. & Baker, R. F. (1975) *J. Mol. Biol.* **98**, 69–92.
6. Baker, R. F. & Thomas, C. A., Jr. (1977) in *DNA Insertion Elements, Plasmids, and Episomes*, eds. Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 463–469.
7. Chase, J. W. & Richardson, C. C. (1974) *J. Biol. Chem.* **249**, 4545–4552.
8. Chase, J. W., & Richardson, C. C. (1974) *J. Biol. Chem.* **249**, 4553–4561.
9. Ando, T. (1966) *Biochim. Biophys. Acta* **114**, 158–168.
10. Sutton, W. (1971) *Biochim. Biophys. Acta* **240**, 522–531.
11. Case, S. T. & Baker, R. F. (1975) *Anal. Biochem.* **64**, 477–484.
12. Peacock, A. C. & Dingman, C. W. (1968) *Biochemistry* **7**, 668–674.
13. Subramanian, K. N., Pan, J., Zain, S. & Weissman, S. M. (1974) *Nucleic Acids Res.* **1**, 727–752.
14. Yang, R. C.-A., Van de Voorde, A. & Fiers, W. (1976) *Eur. J. Biochem.* **61**, 101–117.
15. Grippo, P., Iaccarino, M., Parisi, E. & Scarano, E. (1968) *J. Mol. Biol.* **36**, 195–208.
16. Grippo, P., Parisi, E., Carestia, C. & Scarano, E. (1970) *Biochemistry* **9**, 2605–2609.
17. Kotzin, B. L. & Baker, R. F. (1972) *J. Cell. Biol.* **55**, 75–81.
18. Baker, R. F. & Case, S. T. (1974) *Nature (London)* **249**, 350–352.
19. Case, S. T. & Baker, R. F. (1975) *Nature (London)* **253**, 64–66.