Proceedings of the National Academy of Sciences
Vol. 67, No. 2, pp. 943–950, October 1970

In Vitro Biosynthesis of Pseudouridine at the Polynucleotide Level by an Enzyme Extract from *Escherichia coli*

Lee Johnson and Dieter Söll

DEPARTMENT OF MOLECULAR BIOPHYSICS AND BIOCHEMISTRY, YALE UNIVERSITY, NEW HAVEN, CONNECTICUT 06520

Communicated by Kenneth B. Wiberg, July 27, 1970

Abstract. DNA from *Mycoplasma sp.* Kid which was enriched for tRNA genes (containing about 10% tDNA) was transcribed by *E. coli* RNA polymerase. The RNA transcription product labeled with [^{14}C]uridine was formed in good yield (70-fold net synthesis). After incubation of this [^{14}C]uridine-labeled RNA with *E. coli* extracts, nucleotide analyses revealed that [^{14}C]pseudouridine was formed. The experiments support the idea that the conversion of uridine to pseudouridine takes place at the macromolecular level. Furthermore, the conversion was shown to be specific for a uridine residue in tRNA-like material since neither [^{14}C]polyuridylic acid nor the [^{14}C]uridine-labeled RNA transcribed from λ DNA served as substrate for the pseudouridine-forming enzyme(s).

---

tRNA is unique among cellular RNAs in its content of a large number of modified nucleosides.[1] Studies on the biosynthesis of such modified nucleosides have been limited by the lack of a proper substrate for the modifying enzymes. Two useful substrates can be envisaged: tRNA that is lacking a specific nucleoside, or completely unmodified tRNA. The former substrate is found in tRNA of *Mycoplasma sp.* Kid,[2] which lacks many modified nucleosides and has been used for the isolation and characterization of the *E. coli* enzyme that forms isopentenyladenosine.[3] The latter substrate can be produced by transcribing DNA sequences that contain the information for tRNA (tDNA) with DNA-dependent RNA polymerase.

To date, there are three useful sources of tDNA: (*i*) chemically synthesized tRNA genes;[4] (*ii*) the DNA of $\phi 80 Su_3^+$, a transducing phage carrying a tRNA gene,[5] and (*iii*) double-stranded DNA from *Mycoplasma sp.* Kid which has been enriched for the sequences corresponding to tRNA.[6] The enriched Kid DNA appeared to be more advantageous for this purpose than $\phi 80 Su_3^+$ DNA, since its content of tDNA is much higher (10% compared to 0.4%).

Pseudouridine is the most common modified nucleoside in tRNA. It may be crucial to tRNA function since it is found in all tRNAs active in protein synthesis. Recently it has been shown[7] that a tRNA which is not active in protein synthesis lacks pseudouridine. The mechanism of pseudouridine biosynthesis in RNA is not known. From earlier work[8–11] comes indirect evidence

that the mechanism involves the modification of a uridine residue in the polynucleotide chain. However, the isolation of an enzyme which catalyzes the synthesis of pseudouridine[12,13] has prompted speculation that pseudouridine in RNA arose through incorporation of monomers.[14]

In this paper we report the transcription of tDNA from *Mycoplasma sp.* Kid and the use of this RNA (Kid RNA) as a substrate for detecting the pseudouridine-forming enzyme(s) in *E. coli* extracts. The studies indicate that the *in vitro* reaction results in the formation of pseudouridine contained in Kid RNA, and that the conversion occurs at the macromolecular level.

Materials and Methods. [2-^{14}C]Uridine triphosphate (50.9 mCi/mmol) or [U-^{14}C]uridine triphosphate (diluted to 25 mCi/mmol) was obtained from New England Nuclear Corporation. 2-[^{14}C]Polyuridylic acid (mol wt >50,000, 0.843 mCi/mmol P) was obtained from Miles Laboratories, Inc. Bacterial alkaline phosphatase and T2 RNase were products of Worthington Biochemical Corp. T2 RNase was also obtained from Calbiochem. Omnifluor was a product of New England Nuclear Corp. Microcrystalline cellulose for thin-layer chromatography was purchased from Sigma Chemical Co. λ-phage DNA was a gift of Dr. Charles Radding. Kid tDNA was a gift of Dr. John Ryan. *E. coli* RNA polymerase was prepared according to Chamberlin and Berg.[15] Kodak rapid processing x-ray films were used for autoradiography.

Determination of radioactivity: Paper disks or paper chromatograms (after being cut into 2-cm strips) were counted in a Packard Tri-Carb liquid scintillation counter (model 3320). The scintillation medium consisted of Omnifluor dissolved in toluene (4 g/liter). The counting efficiency for ^{14}C was 65% and that for ^{32}P was near 100%. Radioactive spots on thin-layer chromatograms were quantitated by scraping areas of cellulose containing labeled material into the counting vials and counting the samples in the liquid scintillation counter.

Paper and thin-layer chromatography: Descending paper chromatography was performed at room temperature using Whatman no. 1 paper. Ascending thin-layer chromatography was carried out at room temperature on 20 × 20 cm glass plates coated with 0.25 mm microcrystalline cellulose. Solvent systems used were: solvent I, isobutyric acid–0.5 M ammonium hydroxide (5:3, v/v); solvent II, isopropyl alcohol–concentrated hydrochloric acid–water 70:15:15; solvent III, isopropyl alcohol–concentrated ammonium hydroxide–water 7:1:2.

tDNA: The preparation of *Mycoplasma sp.* Kid DNA enriched in genes corresponding to tRNA and rRNA was described previously.[6,16] About 10% of the double-stranded DNA (which was brought by sonication to a molecular mass of approximately 200,000 daltons) was tDNA.

[^{14}C]Uridine-labeled RNA (Kid RNA and λRNA): Labeled RNA was obtained by the reaction of *E. coli* DNA-dependent RNA polymerase (containing σ-factor) with either Kid tDNA or λ DNA as template under standard conditions.[15] The reaction mixture contained per ml: 40 μmol of Tris·HCl (pH 7.9), 4 μmol of $MgCl_2$, 1 μmol of $MnCl_2$, 12 μmol of β-mercaptoethanol, 0.5 μmol each of ATP, CTP, and GTP, 0.25 μmol (6.25–12.5 μCi) of [^{14}C]UTP, 0.08 A_{260} unit* of Kid tDNA or λ DNA, and 650 units of RNA polymerase. Incubation was at 37°C for 2 hr. The reaction mixture was extracted with phenol and the RNA solution was dialyzed first against 0.2 M NaCl–0.05 M sodium acetate (pH 5.3)–0.01 M $MgCl_2$ and then against water. The resultant ^{14}C-labeled RNAs were designated Kid RNA and λ RNA respectively.

Sedimentation of RNA: Zone sedimentation was as described by Burgi and Hershey.[17] A linear concentration gradient of 20 to 80% deuterium oxide (v/v) in 0.01 M sodium acetate (pH 5)–0.005 M $MgCl_2$ was used. The sample of RNA in 0.1 ml of aqueous buffer was layered on a 5.2 ml gradient and centrifuged in the SW 65 rotor of a Spinco L2-65B at 4°C and 65,000 rpm for 6 hr. The polyallomer tube was punctured

and 12-drop fractions were collected on filter disks. The disks were dried and counted for radioactivity.

S-100 preparation: The preparation of a 100,000 × *g* supernatant fraction from *E. coli* Q 13 and from *Mycoplasma sp.* Kid was as described previously.[18] The final preparations (freed from tRNA by DEAE-cellulose chromatography) contained 10% glycerol and were stored frozen in liquid nitrogen.

Pseudouridine formation: The reaction mixtures contained, per ml: 0.5–1 × 10^6 cpm of [^{14}C]uridine-labeled RNA; 50 μmol of Tris·HCl, 5 μmol of $MgCl_2$, about 1 mg of S-100 protein, and other compounds as specified in Table 1. Incubations were

TABLE 1. *Pseudouridine formation in vitro.**

Kid RNA incubated with	[^{14}C]ψp isolated	
	%†	cpm‡
...	0.14§	..
E. coli S-100	0.26	96
E. coli S-100 (heated)	0.14	68
E. coli S-100 + UTP¶	1.1	112
Mycoplasma S-100	0.38	114

* 1.0 μmol/ml GTP and ATP and 10^5 cpm of Kid RNA were used in a 0.15 ml reaction.
† Amount of radioactivity in spot compared to total amount of radioactivity in U, Up, and pppUp.
‡ Amount of radioactivity recovered from thin-layer chromatography plate.
§ Average value of several control experiments.
¶ UTP concentration was 0.3 μmol/ml. No GTP or ATP was included.

carried out at 37°C for 30 min, after which the mixture was extracted with phenol and dialyzed against 0.2 M NaCl–0.05 M sodium acetate (pH 5.3)–0.01 M $MgCl_2$ and then against glass-distilled water. This procedure removed unincorporated nucleoside triphosphates and other low molecular weight components. The dialysate was then evaporated, digested with T2 RNase, and chromatographed as described below.

Identification of nucleotides and nucleosides: [^{14}C]uridine-labeled RNA was digested with T2 RNase and the resulting nucleotide mixture was analyzed by two-dimensional thin-layer chromatography in solvents I and II and subsequent autoradiography of the plate. The radioactive materials were eluted with 0.15 M ammonium hydroxide from the cellulose scrapings and then characterized by chromatography with authentic markers. Spot 3 (Fig. 4) was tentatively identified as pppUp on the basis of its chromatographic behavior. Bacterial alkaline phosphatase degraded it to uridine.

Results. Transcription of Kid tDNA: DNA from *Mycoplasma sp.* Kid was transcribed with *E. coli* RNA polymerase in the presence of [^{14}C]UTP as the only radioactive nucleoside triphosphate. The kinetics of RNA synthesis are shown in Fig. 1. It can be calculated that the total net synthesis of RNA was approximately 70-fold over input DNA. The amount of RNA synthesis varied with the age of the DNA preparation, but at least 30-fold net synthesis was achieved in each case.

In order to study the size of the transcription product, we subjected the isolated [^{14}C]Kid RNA to gradient centrifugation together with *E. coli* [^{32}P]-tRNA as a marker. The Kid RNA sediments as a broad band somewhat faster than the sharp peak of *E. coli* tRNA (Fig. 2). The heterogeneity of the transcription product indicated by the sedimentation profile is not surprising, since the tDNA is also heterogeneous in size (average molecular weight of 200,000).

***In vitro* pseudouridine formation:** The assay scheme used to detect pseudouridine formation is outlined in Fig. 3. The isolated Kid RNA was incubated with an S-100 supernatant enzyme preparation from *E. coli*. After incubation

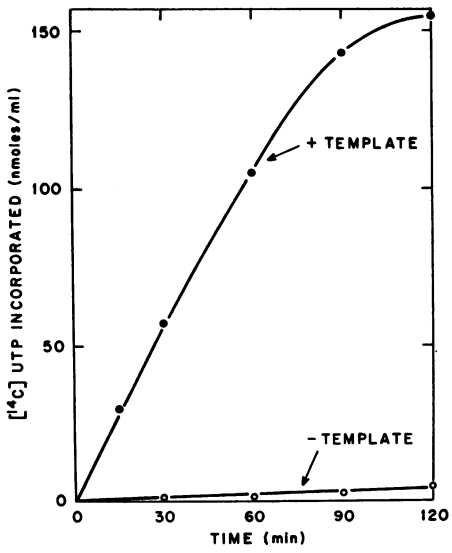

FIG. 1. Kinetics of formation of Kid RNA.

the mixture was extracted with phenol and dialyzed extensively to remove all mono- and small oligonucleotides. Density gradient centrifugation of this material revealed that it had a sedimentation profile similar to that of nonincubated Kid RNA (data not shown).

The presence of pseudouridine in the nucleotide mixture resulting from a T2 RNase digestion of incubated Kid RNA was detected after two-dimensional thin-layer chromatography and autoradiography of the plates. As a control, the nucleotide composition of nonincubated Kid RNA was analyzed. A typical result is shown in Fig. 4.

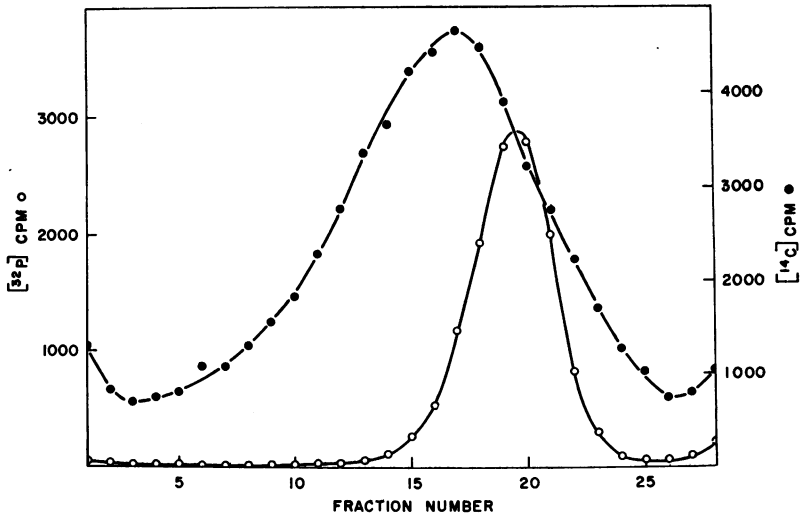

FIG. 2. Sedimentation of [14C]Kid RNA and *E. coli* [32P]tRNA.

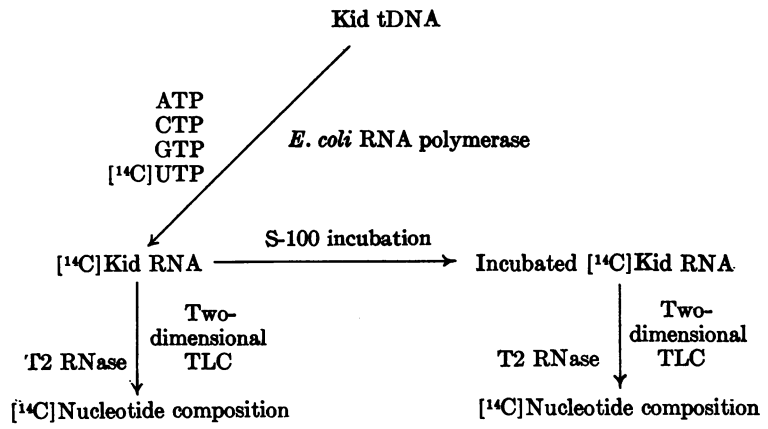

FIG. 3. Assay of formation of pseudouridine *in vitro*. TLC, thin-layer chromatography.

The major spot (1) corresponds to Up. Spot 2 is U, which originates from the digestion of an RNA molecule containing uridine at the 3′ terminus. Spot 3 is probably pppUp, or a similar uridine polyphosphate which would originate from the 5′ end of an RNA. The amounts of terminal nucleotides compared to Up are consistent with an average chain length for the Kid RNA of 80–120 nucleotides. Spot 5 is ψp. The radioactive material chromatographed with a marker of ψp in solvents I, II, and III. After treatment with bacterial alkaline phosphatase and chromatography in solvents I and II it showed the same R_f values as pseudouridine. Spot 4 is most probably dUp, but rigorous chromatographic characterization has not yet been completed. Occasionally some unidentified material (Spot 6) was observed. Small amounts of radioactivity remaining at the origin (O) indicated incomplete digestion.

Quantitation of the radioactive materials on the plate (Table 1) showed that Kid RNA incubated with S-100 fractions contained 2–8 times the amount of ψp as the control RNA. The conversion of uridine to pseudouridine seems to be an enzymatic reaction since the boiled cell extract does not stimulate the conversion. Extracts from *E. coli* and *Mycoplasma sp.* Kid seem to be equally effective in catalyzing pseudouridine formation. The addition of cold nucleoside triphosphates, especially UTP, does not inhibit the reaction. The presence of low amounts of ψp found in the control reactions was attributed to contamination of

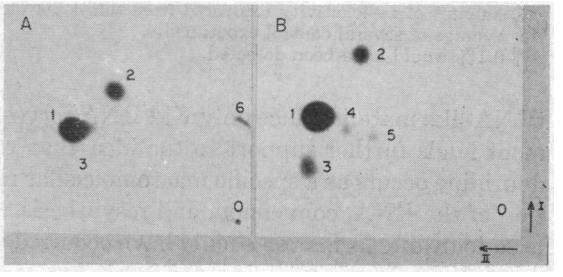

FIG. 4. Autoradiogram of a two-dimensional separation of a T2 digest of [14C]uridine-labeled Kid RNA. *A*, Nonincubated Kid RNA. *B*, Incubated Kid RNA.

the [^{14}C]UTP sample with small amounts (about 0.1%) of [^{14}C]ψTP which would have been incorporated randomly into RNA during transcription.[19]

The cofactor or energy requirements of the pseudouridine-forming enzyme(s) are not known. Nucleoside triphosphates were routinely included in the reactions except where indicated. If other cofactors are needed, the minute quantities required for the pseudouridine formation in these assays could have been supplied by some agent present in the S-100 preparation.

The results suggest that pseudouridine is formed at the macromolecular level. An alternative explanation of the observed results could be that the Kid RNA is degraded during incubation with the *E. coli* supernatant enzymes to nucleotides which are transformed to a pseudouridine derivative and, after enzymatic phosphorylation, polymerized to a polynucleotide. This possibility was tested by incubating both [^{14}C]UTP and [^{14}C]pU with supernatant enzymes under the same conditions as for the Kid RNA incubations except that unlabeled UTP was omitted from the mixture. After dephosphorylation with bacterial alkaline phosphatase and one-dimensional paper chromatography in solvent II no pseudouridine was detected. Furthermore, the observed RNA polymerase and polynucleotide phosphorylase activities contained in the amount of S-100 used are not sufficient to synthesize a polynucleotide of the size of Kid RNA. These results make it unlikely that pseudouridine is formed at the nucleoside or nucleotide level. A definite answer can be obtained from an experiment in which pseudouridine formation in Kid RNA is examined in the presence of unlabeled uracil, uridine, and uridine mono-, di-, and triphosphates.

Some studies to test the specificity of the pseudouridine-forming enzyme(s) were also made. [^{14}C]poly U and λ RNA were incubated with the *E. coli* supernatant enzymes. Neither RNA should contain tRNA.[20] The results shown in Table 2 indicate that the pseudouridine-forming enzyme(s) are specific for

TABLE 2. *Substrate specificity of pseudouridine-forming enzyme(s).**

Source of RNA (cpm/0.25 ml)	Incubation with *E. coli* S-100	[^{14}C]ψp isolated	
		%†	cpm‡
Kid RNA (2×10^5)	−	0.14§	...
	+	1.40	885
λ RNA (4×10^5)	−	0.16	123
	+	0.27	447
Poly U (10^6)	−	none detected¶	
	+	none detected¶	

* For details see *Materials and Methods*. The reaction mixture also contained 0.1 μmol/ml of all four unlabeled nucleoside triphosphates.
† Amount of radioactivity in spot compared to total amount of radioactivity in U, Up and pppUp.
‡ Amount of radioactivity recovered from thin-layer chromatography plate.
§ Average of several control experiments.
¶ 0.1% would have been detected.

tRNA-like material since only Kid RNA serves well as a substrate. This experiment lends further support to the idea that the conversion of uridine to pseudouridine occurs as a specific macromolecular rearrangement and not by degradation of the RNA, conversion, and resynthesis. If the latter scheme were correct pseudouridine formation should have occurred with [^{14}C]poly U.

Discussion. The present study has utilized the unmodified RNA transcription product of Kid tDNA to investigate the formation of pseudouridine in tRNA. The results have shown that pseudouridine is formed *in vitro* by a simple incubation of the unmodified RNA with *E. coli* cell extract. Our results present strong evidence that the pseudouridine is formed at the polynucleotide level and that pseudouridine is derived from uridine contained in RNA. Whether the mechanism involves the cleavage of an *N*-glycoside bond in uridine, rotation of a uracil residue, and formation of a *C*-glycosidic linkage remains to be seen. The possibility that pseudouridine is inserted after excision of specific nucleoside residues from the RNA seems less likely in light of our results. The formation of pseudouridine appears to be specific for tRNA-like material since in neither polyuridylic acid, nor the transcription product of λ DNA, was significant conversion of uridine to pseudouridine found. The actual macromolecular substrate in this reaction and the sequence of events are not yet known. However, it appears that this modification occurs in the absence of other modified nucleosides and that the RNA need not assume its mature size and conformation to be modified. Preliminary experiments showed that incubated Kid RNA could not be aminoacylated.

The formation of mature tRNA may be considered to proceed as follows:

$$\text{tRNA genes} \xrightarrow[\text{transcription}]{} \text{tRNA precursor} \xrightarrow[\text{maturation}]{} \text{mature tRNA}$$

The availability of the unmodified RNA transcription product permits the investigation of the biosynthesis of the other modified nucleosides and of functional tRNA. At present little is known about the sequence of events or the enzymes involved in these reactions. For instance, it is of crucial importance to know whether the precursor is longer than a mature tRNA molecule. The precursor could be brought to the proper length by an enzyme after transcription of tDNA into a large RNA chain. Such a mechanism appears to exist in HeLa cells, where a large precursor tRNA, still incompletely modified, is found.[21] Another possibility would be to control the size of tRNA through the transcription process, which would require a termination signal for RNA polymerase, possibly mediated by termination factors.[22] It would also be interesting to clarify whether the C-C-A nucleotide sequence of the amino acid acceptor end of tRNA is coded for by tDNA, or if the tRNA-CMP-AMP-pyrophosphorylase is required to add these nucleotides.[23] Studies to answer these questions can be attempted with Kid RNA.

We are greatly indebted to Dr. J. Ryan for his help in the early experiments of tDNA transcription. This study was supported by grants from the National Institutes of Health (GM 15401), The National Science Foundation (GB 7269 and GB 19085) and from the American Cancer Society (E 590).

* One absorbance unit, A_{260}, is defined as the amount of material per ml of solution that produces an absorbance of 1 in a 1-cm light path at 260 nm.

[1] Hall, R. H., The Modified Nucleosides in Nucleic Acids, Columbia University Press, New York (1970).
[2] Hayashi, H., H. Fisher, and D. Söll, *Biochemistry*, **8**, 3680 (1969).
[3] Bartz, J., L. Kline, and D. Söll, *Biochem. Biophys. Res. Commun.*, in press.

[4] Agarwal, K. L., H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. RajBhandary, J. H. Van de Sande, V. Sgaramella, H. Weber, and T. Yamada, *Nature*, **227**, 27 (1970).

[5] Smith, J. D., J. N. Abelson, B. F. C. Clark, H. M. Goodman, and S. Brenner, *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 479 (1966).

[6] Ryan, J. L., and H. J. Morowitz, *Prod. Nat. Acad. Sci. USA*, **63**, 1282 (1969).

[7] Stewart, T. S., R. J. Roberts, and J. L. Strominger, submitted for publication.

[8] Dubin, D. T., and A. Gunlap, *Biochim. Biophys. Acta*, **134**, 106 (1967).

[9] Kusama, K., D. M. Prescott, L. O. Fröholm, and W. E. Cohn, *J. Biol. Chem.*, **241**, 4086 (1966).

[10] Weiss, S. B., and J. Legault-Demare, *Science*, **149**, 429 (1965).

[11] Ginsberg, T., and F. F. Davis, *J. Biol. Chem.*, **243**, 6300 (1968).

[12] Suzuki, T., and R. M. Hochster, *Can. J. Microbiol.*, **10**, 867 (1964).

[13] Heinrikson, R. L., and E. Goldwasser, *J. Biol. Chem.*, **239**, 1177 (1964).

[14] Goldwasser, E., and R. L. Heinrikson, *Progr. Nucl. Acid Res. Mol. Biol.*, **5**, 412 (1966).

[15] Chamberlin, M., and P. Berg, *Proc. Nat. Acad. Sci. USA*, **48**, 81 (1962).

[16] Ryan, J., Ph.D. Thesis, Yale University (1969).

[17] Burgi, E., and A. D. Hershey, *Biophys. J.*, **3**, 309 (1963).

[18] Söll, D., J. D. Cherayil, and R. M. Bock, *J. Mol. Biol.*, **29**, 97 (1967).

[19] Slapikoff, S., and P. Berg, *Biochemistry*, **6**, 3654 (1967).

[20] λ DNA does not contain tRNA genes, but λ-specific 4S RNA molecules have been found in phage-infected *E. coli* cells (M. Pearson and D. Hogness, personal communication). Should such RNA molecules serve as substrates for the pseudouridine-forming enzymes then the low amount of pseudouridine found in the experiment with λ RNA may be explained.

[21] Mowshowitz, D. B., *J. Mol. Biol.*, **50**, 143 (1970).

[22] Roberts, J. W., *Nature*, **224**, 1168 (1969).

[23] Daniel, V., S. Sarid, and U. Z. Littauer, *Science*, **167**, 1682 (1970).