Enzymatic Synthesis of Deoxyribonucleic Acids with Repeating Sequences. A New Repeating Trinucleotide Deoxyribonucleic Acid, $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$†

A. Richard Morgan,* Marion B. Coulter, Wayne F. Flintoff, and Verner H. Paetkau

ABSTRACT: Synthetic DNAs with repeating nucleotide sequences have been valuable as models for more complex DNAs. The enzymatic copying of synthetic DNAs poses several problems including (a) production of covalent links between complementary strands (clc-DNA) and (b) contamination of the product with other polymers, such as $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ and $dG_n \cdot dC_n$. This paper details (i) the optimum conditions for the enzymatic copying of synthetic DNAs, (ii) a simple fluorescence assay for the extent of synthesis and the presence of clc-DNA, (iii) a quick isolation procedure, (iv) conditions necessary for producing DNAs without clc strands, (v) a very sensitive assay for estimating contaminating $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$, and (vi) the synthesis of a new polymer $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ and its physical properties.

Synthetic DNAs with repeating nucleotide sequences were originally chemically synthesized and enzymatically copied with DNA polymerase I of *Escherichia coli* in the laboratory of Dr. H. G. Khorana (Byrd *et al.*, 1965; Wells *et al.*, 1967). They have been used in many types of studies both biological and physicochemical, for example, the elucidation of the genetic code (Khorana *et al.*, 1966) and the specificity of three-stranded helices (Morgan and Wells, 1968). Difficulties had arisen in synthesizing DNAs with high GC content (R. D. Wells, unpublished data). Under conditions successfully used for copying synthetic polymers with DNA polymerase, the expected product $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ was not formed from the mixture of oligomers $d(C\text{-}C\text{-}T)_4$ and $d(G\text{-}G\text{-}A)_3$. By altering the conditions of synthesis it is now possible to make $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ and also to improve considerably on the net-fold replication of other synthetic DNAs. We routinely follow the synthesis of DNA by a very sensitive fluorescence assay which depends on the intercalation of ethidium bromide between the bases in duplex structures specifically (Le Pecq and Paoletti, 1966, 1967; Morgan and Paetkau, 1972). This assay is not only sensitive, but an immediate reading is obtained allowing calculation of the amount of DNA synthesized. If degradative nuclease action is beginning to overtake the synthetic reaction, isolation of DNA can be initiated at once. The fluorescence assay also allows a determination of the amount of clc[1]-DNA present, and since $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ is a frequent contaminant of synthetic reactions (Burd and Wells, 1970) it can usually be estimated by the fluorescence assay. An alternative assay for $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$, which will detect at little as 0.01% contamination, depends on transcription using ATP and UTP as the only rNTP substrates, at relatively high KCl concentrations (0.15–0.2 M). We have developed a simple gel filtration procedure for isolating DNAs in any desired buffer within 10–20 min of terminating the reaction.

Materials and Methods.

Chemicals. $d(C\text{-}C\text{-}T)_4$ and $d(G\text{-}G\text{-}A)_3$ were chemically synthesized in Dr. H. G. Khorana's laboratory (Narang *et al.*, 1967). The other DNA templates with repeating sequences have already been described (Wells *et al.*, 1967). Nucleases were from Worthington Biochemical Corp., ethidium bromide from Sigma Chemical Co., Agarose 5m from Bio-Rad Labs, CsCl was from Pierce Chemical Co. and Schwarz/Mann, and Cs_2SO_4 from Schwarz/Mann. Radioactive (^{14}C and ^{3}H) nucleoside and deoxynucleoside triphosphate were obtained from Schwarz/Mann, unlabeled nucleotides from P-L Biochemicals and Raylo Co., and [α^{32}-P]dATP and -dCTP from International Chemical and Nuclear Corp. *E. coli* B cells, grown to $^{3}/_{4}$ log phase in minimal medium, were purchased frozen from the Grain Processing Co., Muscatine, Iowa.

DNA polymerase I from *E. coli* was prepared by the method of Jovin *et al.* (1969), and RNA polymerase by a modification (Paetkau and Coy, 1972) of the procedure of Chamberlin and Berg (1962). A by-product of RNA polymerase preparation was fraction DIII. Material extracted from the protamine pellet with 0.01 M Tris-HCl (pH 8.0)–0.1 M $MgCl_2$–0.1 mM EDTA–10 mM β-mercaptoethanol was fractionated with ammonium sulfate and the fraction precipitating between 45 and 65% saturation redissolved in 20 mM potassium phosphate (pH 7.4) as DIII (protein concentration 1.8 mg/ml). DIII was added to prevent the accumulation of clc strands during the DNA polymerase reaction. Fraction IV DNA polymerase (Richardson *et al.*, 1964) was also effective in preventing clc sequences (Paetkau, 1969) and was used in some experiments instead of DIII. The factor in DIII blocking clc accumulation is a heat-stable protein other than endonuclease I (Coulter *et al.*, 1974). DIII was heated (97° for 10 min) to remove most of the nuclease activity, and tRNA was added to the DNA polymerase reactions to block residual endonuclease I activity, as indicated in the experiments. Pancreatic DNase was added in some experiments to stimulate synthesis, as indicated.

Assays. The DNA polymerase reaction mixtures contained: 50 mM potassium phosphate (pH 7.5), 7–15 mM $MgCl_2$ depending on the triphosphate concentration, deoxynucleoside triphosphate 0.8–8 mM total in a ratio corresponding to the base ratio of the template, 20 μg/ml of DNA polymerase I

† From the Department of Biochemistry, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada. *Received February 20, 1973.* This work was supported by the Medical Research Council of Canada, Grants MA3758 and MA3764.

[1] Abbreviation used is: clc-DNA, DNA with covalently linked, complementary sequences.

fraction 7 (Jovin *et al.*, 1969), either fraction 4 or DIII, and the DNA template, as indicated.

The RNA polymerase reaction mixtures contained: 40 mM Tris-HCl (pH 8), 5 mM $MgCl_2$, 0.15 M KCl, 0.2 mM dithiothreitol, 0.5 mM of each required rNTP, 0.15 mg/ml of RNA polymerase, and the template concentration as indicated. [^{14}C]UTP was 1580 cpm/nmol and RNA synthesis was measured by acid insolubility as previously described (Nishimura *et al.*, 1964). The synthesis of DNA was followed either by the incorporation of radioactive deoxynucleoside triphosphate into acid-insoluble material (Nishimura *et al.*, 1964) or by the enhanced fluorescence of ethidium bromide. The standard ethidium bromide solution for assaying DNA contained 2 mM Tris-HCl (pH 8), 0.2 mM EDTA, and 0.5 μg/ml of ethidium bromide. The excitation wavelength was 525 nm and the emission was 600 nm, using the Turner spectrofluorometer, Model 430; 2 ml of standard ethidium bromide solution was added to 1-cm^2 cuvets (Spectrosil) as a blank, and a sample of DNA (up to 20 μl volume) was added to a second cuvet with 2 ml of standard ethidium bromide solution. Routine precautions included calibrating the instrument from time to time with a fluorescent standard, as the intensity of the xenon lamp sometimes fluctuated. This fluorescence assay responded linearly to DNA concentrations up to 1 μg/ml (Morgan and Paetkau, 1972). None of the components of the DNA polymerase reaction mixture interfered with the fluorescence assay under our conditions. All measurements were at 25°.

CsCl Density Gradient Centrifugation. Analytical CsCl density gradient centrifugation was carried out on the Beckman Model E ultracentrifuge with uv optics. Centrifugaton was at 48,000 rpm for at least 20 hr at 25°. The density of the solution was determined by the refractive index on an Abbe refractometer. The density of the DNA after equilibrium had been reached was calculated either by the isoconcentration method (Vinograd, 1963) or by relating the density to that of a known market DNA. The values of β were taken from Chervenka (1969).

The strands of d(T-C-C)$_n$·d(G-G-A)$_n$ were separated in a preparative alkaline CsCl density gradient; 6.9 OD_{260} units of d(T-C-C)$_n$·d(G-G-A)$_n$, 0.2 ml of 1 M NaOH, and 0.1 ml of 0.1 M EDTA were made up to 5.11 ml with water and 7.8 g of CsCl was added. The density was 1.805 g/ml. The solution was overlaid with Dow-Corning 556 fluid in polyallomer tubes, which were centrifuged at 35,000 rpm in the Ti 50 fixed angle rotor at 4° for 64 hr using the Beckman L2 65B ultracentrifuge. The gradient was pumped out into 16 fractions through a needle lowered into the tube, using a peristaltic pump to obtain a uniform flow rate. Fractions 1–8 contained d(G-G-A)$_n$ and fractions 9–16 d(T-C-C)$_n$. The DNAs were dialyzed against 5 mM Tris (pH 8) and 0.05 mM EDTA. The ratio of the OD at 260–280 was 2.15 for d(G-G-A)$_n$ and 1.28 for d(T-C-C)$_n$. The polymers were further characterized by RNA polymerase assays.

Molecular Weights. Polymers were examined in the Spinco Model E ultracentrifuge, using the method and equations of Studier (1965).

Nearest Neighbor Analyses of d(T-C-C)$_n$·d(G-G-A)$_n$. The standard polymerase assay conditions were used with one of the deoxyribonucleoside triphosphates labeled with ^{32}P in the α-phosphate. Reactions (50-μl scale) were terminated by adding a tenfold excess of EDTA over Mg^{2+} and sodium dodecyl sarcosinate to 0.5%. Polymers were dialyzed free of triphosphates in a collodion bag (Sartorius) by repeatedly concentrating them with suction and adding water and unlabeled dNTPs to chase out any remaining radioactivity not in DNA. Finally the DNA was taken up in 0.25 ml of 10 mM glycine (pH 10) and 2 mM $CaCl_2$, containing 0.2 μg/ml of micrococcal nuclease. After overnight digestion at 37° most of the radioactivity was acid soluble. The solution was then made 100 mM in NH_4OAc (pH 5.9) and spleen phosphodiesterase added to 50 μg/ml. After further incubation at 37° for 2 hr, about 0.5 OD_{260} unit of each of the four 3′-dNMPs were added to the solution. The solution was evaporated to dryness, taken up in the minimum volume of water (~50 μl), and electrophoresed on Whatman 3 mm paper at 30 V/cm for 3.5 hr in 0.1 M citrate (pH 3.5). The electropherograms were analyzed for radioactivity either by cutting them into 1-cm strips or else by running them through a radiochromatogram scanner.

Nucleotide Analyses. The samples of DNA were degraded to the mononucleotide level with micrococcal nuclease and spleen phosphodiesterase as for the nearest neighbor analyses. When pancreatic nuclease and snake venom phosphodiesterase were used a contaminating enzymatic activity converted the nucleotides to other unidentified compounds. The mixture of nucleotides was separated and analyzed by high-pressure liquid chromatography using a Varian Aerograph LCS-1000 chromatograph equipped with a 3-m capillary column packed with Varian PA-38 anion exchange resin, essentially according to Brown (1970). The amounts of each nucleotide were calculated by comparing the peak areas, as determined by planimetry, with those of known amounts of authentic nucleotides. Approximately 1–2 nmol of total nucleotide was used per analysis.

Results

The Effect of the dNTP Concentration on DNA Synthesis. In order to obtain the maximum amount of DNA replication in a small volume it seemed that a high triphosphate level might be advantageous. Figure 1 shows the effect of varying the dNTP concentrations on the synthesis of d(T-G)$_n$·d(C-A)$_n$. When each triphosphate is 0.2 mM, maximum synthesis occurs by about 90 min, and thereafter nuclease activity predominates. At a tenfold higher concentration, however, there is still active synthesis at 10 hr, and the degradative action appears to be inhibited. This inhibition may be due to the buildup of pyrophosphate concentration which would compete for OH^- in the degradative reaction (Beyersmann and Schramm, 1968). At the higher triphosphate levels the level of Mg^{2+} must be increased. A good general rule is to use 5–7 mM excess Mg^{2+} above the total triphosphate concentration. Thus at 2 mM in each dNTP (total 8 mM), a Mg^{2+} concentration of 15 mM is used. If 5 mM Mg^{2+} is used instead there is a complete lack of DNA synthesis, suggesting that triphosphates not bound to Mg^{2+} are potent inhibitors of DNA polymerase. The net-fold synthesis of DNA can be as high as 100-fold in 5 hr, depending on the particular template and the initial DNA concentration.

Correlation of Ethidium Bromide Fluorescence and Acid Insolubility Assays for DNA. It has been shown that ethidium bromide fluorescence is specifically enhanced by duplex structures, and the sensitivity of the assay suggested it might be a suitable alternative to the acid insolubility or the hypochromicity assay. The latter assay is not feasible at the high triphosphate concentrations we use in any case. The disadvantage of the acid insolubility assay is the length of time taken from sampling to obtaining a value for the amount of DNA synthesized. The ethidium bromide fluorescence assay

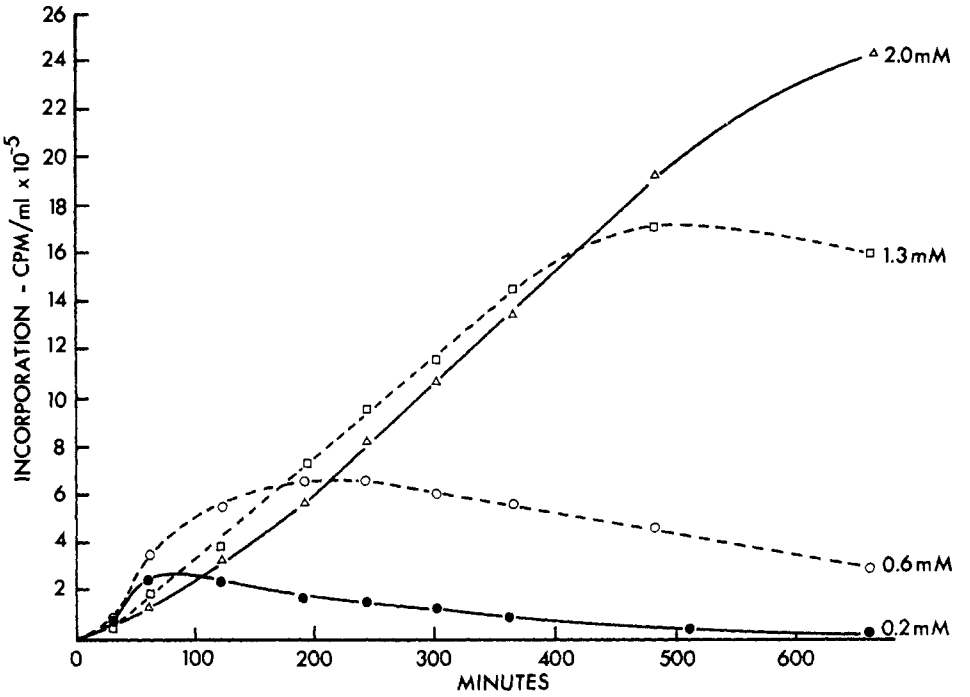

FIGURE 1: The effect of dNTP concentration on $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$ synthesis. The standard DNA polymerase conditions were used. Heated DIII (10 min, 97°), 25 μg/ml, 0.5 OD_{260} of unfractionated tRNA, 50 ng/ml of DNase I, and 0.206 OD_{260} unit/ml of $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$ were present. Each dNTP was present at the concentration indicated on the graph, with TTP ^{3}H-labeled (2700 cpm/nmol) and dCTP ^{14}C-labeled (2000 cpm/nmol); 10-μl samples were assayed by acid insolubility.

provides an immediate value for the DNA synthesized and also gives further information on whether clc-DNA or $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ is being produced. In principle the ethidium bromide assay coupled with the acid insolubility assay can be used to determine if an appreciable amount of single-stranded DNA is present, since ethidium bromide measures only duplex structures, but acid insolubility measures total DNA. As can be seen from Figure 2, the synthesis of $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$ under several concentrations of triphosphates gave identical results when followed by acid insolubility or ethidium bromide fluorescence. This indicates that both strands are being synthesized in identical amounts, and this was confirmed by equal incorporation of ^{3}H and ^{14}C into the two strands. The input template is of course also measured in the ethidium bromide assay, but can be subtracted out. While radioactive incorporation can be made extremely sensitive with substrates of very high specific activity, in fact it is the size of the aliquots taken which will usually limit the repro-

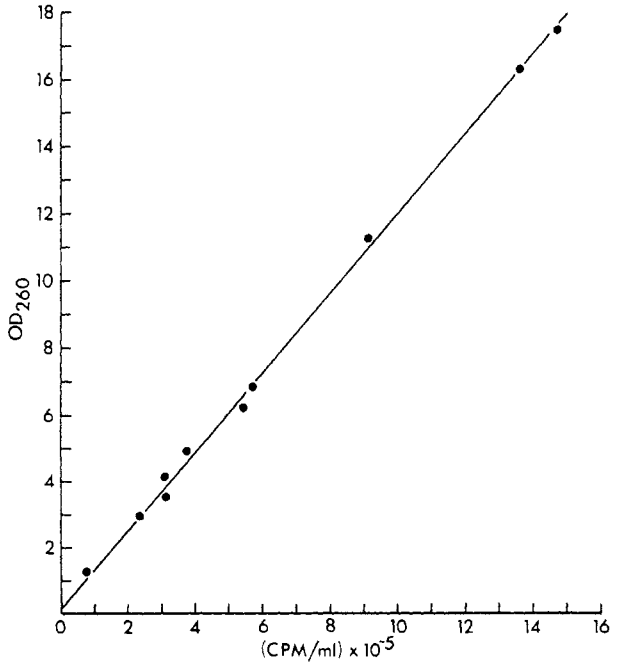

FIGURE 2: Correlation of acid insolubility with the ethidium bromide fluorescence assay. The reaction mixtures and the samples were the same ones as used in Figure 1. The optical density of DNA synthesized was determined by the fluorescence of the test sample, compared to that of a known concentration of $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$.

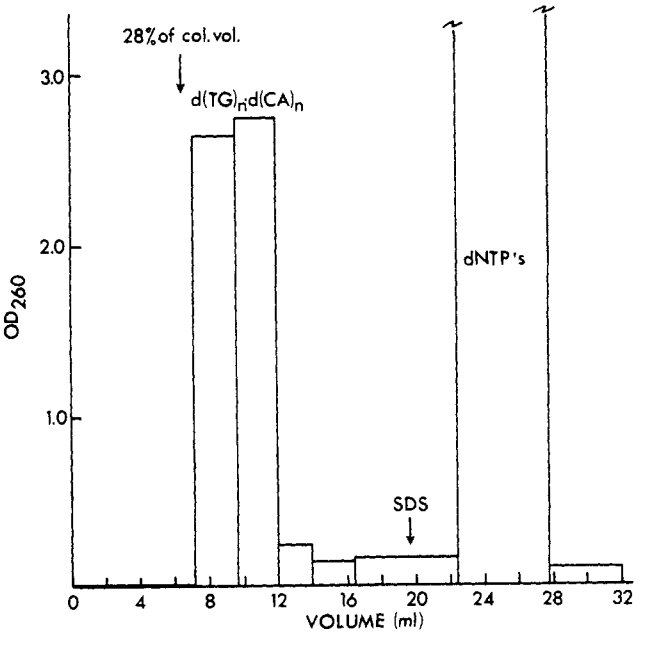

FIGURE 3: Agarose 5m gel filtration of $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$ (see text).

TABLE I: Nearest Neighbor Analysis of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$.

α-^{32}P-Labeled Triphosphate	Radioactivity in 3′-Deoxymononucleotides							
	dCp		dAp		dGp		dTp	
	cpm	%	cpm	%	cpm	%	cpm	%
dTTP	35,600	95.5	1,310	3.5	290	0.8	160	0.4
dATP	0	0	8	0	5,602	99	52	0.9
dCTP	1,369	50	0	0	0	0	1372	50
dGTP	0	0	52,950	50	52,590	50	0	0

ducibility. Since in the later stages of synthesis 5-μl samples already contain far more DNA than is required for the ethidium bromide assay it is evident that little advantage is obtained from the potentially higher sensitivity of using radioactive substrates.

Isolation of DNAs by Gel Filtration. Agarose gels provide a simple and quick procedure to isolate DNA free of protein and small molecules. Using a modification of an earlier report (Paetkau and Khorana, 1971), the reaction was stopped by the addition of excess EDTA (tenfold over Mg^{2+}) and the solution made 0.05% in either sodium dodecyl sulfate or sodium dodecyl sarcosinate (Sarkosyl). Sarkosyl is often advantageous because potassium dodecyl sulfate is relatively insoluble. The solution was heated at 50° for a few minutes to ensure complete disruption of any DNA–enzyme complex. We routinely used Agarose 5m or 15m (50–100 mesh or finer), 0.8 cm^2 × 31 cm, as shown in Figure 3 for the isolation of $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$. The eluting buffer in Figure 3 was 0.05 M triethylammonium bicarbonate (pH 8), which is readily removed by evaporation with ethanol. We routinely use 10 mM Tris-HCl (pH 8)–0.1 mM EDTA, since the DNAs are stored in this buffer.

Synthesis of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$. After some preliminary variations of parameters the following procedure was adopted for the large scale synthesis of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ from oligomers. Since polypurines are known to form secondary structures, the oligomers $d(C\text{-}C\text{-}T)_4$ and $d(G\text{-}G\text{-}A)_3$ (each 2.5 OD_{260} units/ml) were heated to 100° for 2 min and cooled in ice in the absence of any added salts, before the rest of the reaction components were added. The first 30 min of the reaction were at 22° (to prevent melting of the oligomers) and the temperature was then raised to 37° and fraction IV of the DNA polymerase preparation was added to a final concentration of 21 μg/ml. The reaction mixture (0.5 ml) contained 0.5 OD_{260} unit of each oligomer. After 6 hr at 37° the reaction was terminated by addition of EDTA to 50 mM and sodium dodecyl sulfate to 25 mM. After heating at 50° for 5 min the DNA was isolated by Agarose 5m gel filtration as above, 5.5 OD_{260} units of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ appearing as a well-separated peak at the calculated void volume. This sample of DNA was then replicated on a small scale (50 μl) with [^{14}C]dCTP and [^{3}H]TTP at the same time as $d(TG)_n \cdot d(CA)_n$ and $d(TTC)_n \cdot d(GAA)_n$. The ratio of $^{14}C/^3H$ incorporated into DNA was constant for each DNA during 9 hr of synthesis and the ratios were consistent with the base composition of the DNA. In Figure 4 the kinetics of synthesis for $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ are shown. The reaction was scaled up to 1 ml and after 9.5 hr at 37° the DNA was isolated as before; 0.5 OD_{260} unit of DNA was used as template with a final yield of 17.5 OD_{260} units (35-fold synthesis). The DNA was checked for $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ contamination by transcription and contained about 0.75% (see last section).

Characterization of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$. The most definitive criterion for the structure of repeating polymers is nearest neighbor analysis. This polymer was also characterized by transcriptional studies (Morgan, 1970), its buoyant density, the separation of the complementary strands in alkaline CsCl density gradients, the melting temperature, and nucleotide analyses.

The nearest neighbor analysis of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ is summarized in Table I. α-^{32}P-labeled TTP, dATP, dGTP, and dCTP were each used in turn to label the DNA. Occasionally the nearest neighbor analysis was not that expected theoretically for $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$, and this was due to a minor amount of contaminating $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ arising *de novo* in the reaction. A phenol extraction usually removed the $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ (Laskowski, 1972) although it is irreproducible as a routine procedure in our hands. In Table I the radioactivity found in dAp and dTp after labeling with [α-^{32}P]dTTP and [α-^{32}P]dATP, respectively, is almost certainly due to traces of $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$. In the [$\alpha$-^{32}P]dGTP experiment about 0.5% of the radioactivity moved in the region of Tp on the electropherogram but did not coincide with it. It is most probably the nucleoside 3′,5′-diphosphate from the 5′-end of the DNA which would not be further degraded by micrococcal nuclease and spleen phosphodiesterase.

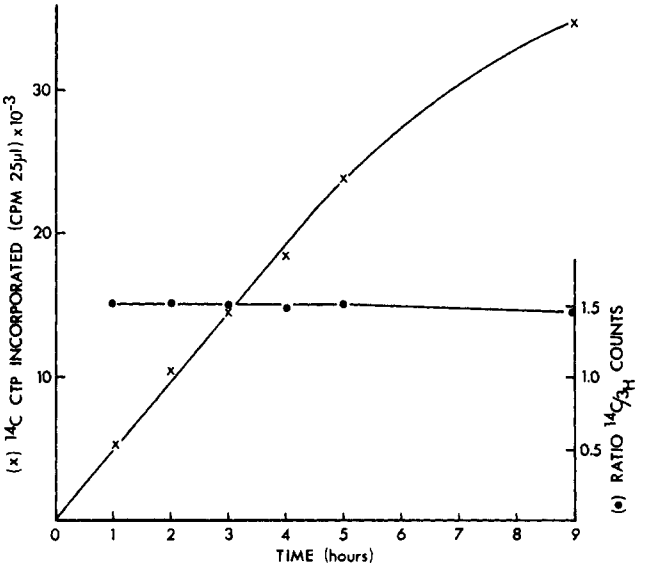

FIGURE 4: The kinetics of synthesis of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$. The standard conditions for DNA polymerase were used with the following concentrations/ml: total dNTPs 8 mM, with dCTP ^{14}C labeled (5000 cpm/nmol) and [^{3}H]dTTP (916 cpm/nmol), nucleotide on paper, $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ 0.55 OD_{260} unit, DNA polymerase fraction VII 80 μg, fraction IV 96 μg and 15 mM $MgCl_2$. Acid-insoluble [^{14}C]dCTP (with restricted channels, no overlap of ^{3}H) (×); ratio of $^{14}C/^3H$ counts (in restricted channels) (●).

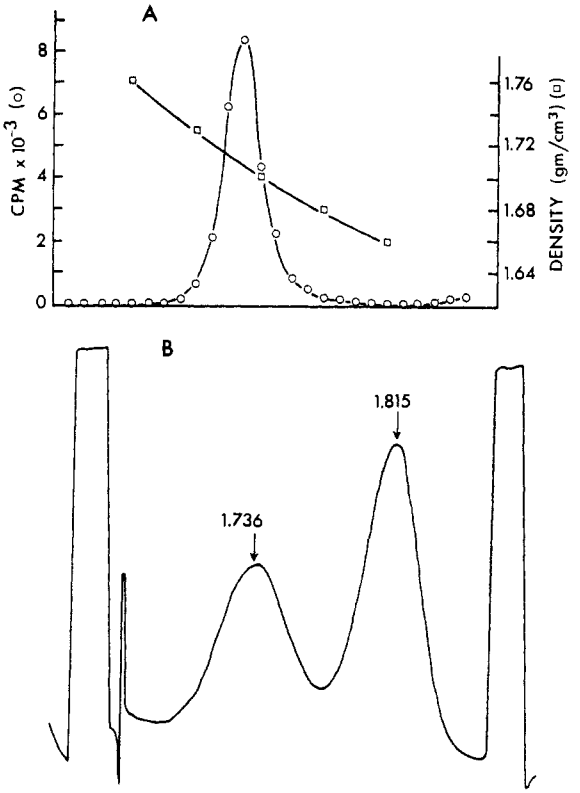

FIGURE 5: Buoyant density gradient centrifugation of d(T-C-C)$_n$·d(G-G-A)$_n$. (A) A neutral preparative CsCl density gradient. The ^{3}H d(T-C-C)$_n$·d(G-G-A)$_n$ was made up to 7 ml in CsCl solution ρ = 1.70 and centrifuged for 48 hr at 40,000 rpm, 20° in a Ti 50 rotor. The fractions were analyzed for density and ^{3}H (present in TMP residues of the DNA). (B) An analytical bouyant density of d(T-C-C)$_n$·d(G-G-A)$_n$ in alkaline CsCl. The CsCl solution had a density of 1.777 and was 0.02 M in NaOH. The sample was centrifuged at 48,000 rpm for 19 hr at 25°. The densities of the peaks were calculated by the isoconcentration method (Vinograd, 1963).

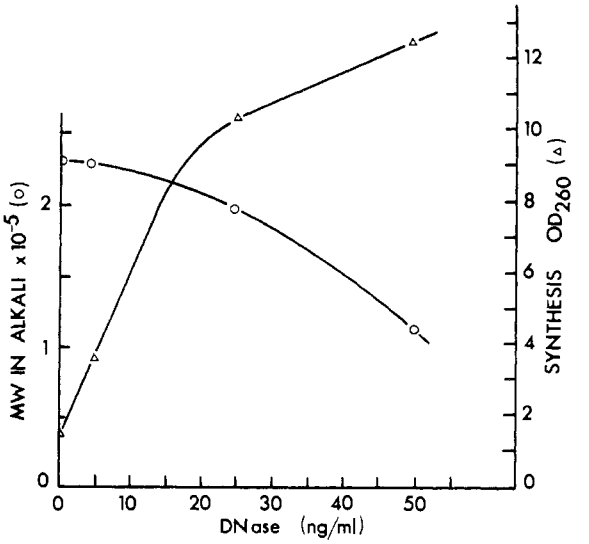

FIGURE 6: The effect of pancreatic DNase on the synthesis and molecular weight of d(T-G)$_n$·d(C-A)$_n$. The reaction conditions for synthesis were those of Figure 1 but with varying DNase I concentrations. Incubation was for 6 hr. The final total dNTP concentration was 4 mM, and $MgCl_2$ 12 mM. Molecular weight in alkali, (○); synthesis as monitored by ethidium bromide fluorescence (△).

It is approximately the amount that would be expected on the basis of the DNA molecular weight. It was not analyzed further since it would not affect the overall nearest neighbor results, which shows that the DNA consists of two strands, one a repeating trinucleotide sequence d(T-C-C)$_n$ and the other the complementary sequence d(G-G-A)$_n$.

Further characterization by buoyant densities in neutral and alkaline CsCl were consistent with a duplex structure for the DNA. Figure 5A shows a preparative banding of ^{3}H-labeled d(T-C-C)$_n$·d(G-G-A)$_n$ in a CsCl gradient. It was prepared in the presence of DIII, tRNA, and DNase I (see next section). There was no evidence for d(A-T)$_n$·d(A-T)$_n$ which would have contained [^{3}H]TMP, and banded at a density of 1.67–1.68. An analytical banding of d(T-C-C)$_n$·d(G-G-A)$_n$ in neutral CsCl gave a density of 1.720 by the isoconcentration method or *vs.* d(A-T)$_n$·d(A-T)$_n$ (ρ 1.672). This is precisely what one would expect on the basis of 67% GC content (Sueoka *et al.*, 1959) but as shown by Wells and Larson (1972) the density of synthetic polymers varies unpredictably. While the density of some synthetic d(pyrimidine)$_n$·d(purine)$_n$ DNAs falls on the line obtained by plotting GC content *vs.* buoyant density for natural DNAs (*e.g.*, d(T-T-C)$_n$·d(G-A-A)$_n$) most do not. Therefore better evidence for the duplex nature of the DNA was obtained by an alkaline CsCl equilibrium centrifugation of the material from the preparative gradient. Figure 5B shows two well-resolved bands were obtained. On the basis of T and G content, we have assigned the band of density 1.736 to d(TCC)$_n$, and that of density 1.815 to d(G-G-A)$_n$. When the areas under the bands were determined by planimetry the ratio of the denser to the lighter band was 1:1.34. The calculated ratio of the expected strands d(G-G-A)$_n$ to d(T-C-C)$_n$ at pH 12 and 265 nm is 1:1.38, confirming that the two strands are present in equimolar ratio. The strands were then separated in a preparative alkaline CsCl gradient (see Methods) for further characterization. The d(T-C-C)$_n$ and d(G-G-A)$_n$ were found not to enhance the fluorescence of ethidium bromide indicating the absence of any duplex DNA. The d(T-C-C)$_n$ strand readily acted as template for transcription with the synthesis of r(G-G-A)$_n$ as proved by nearest neighbor analysis and the ratio of G and A incorporated (Morgan, 1970). The d(G-G-A)$_n$ strand was an extremely poor template for r(U-C-C)$_n$ synthesis, as it is in the duplex DNA as well. This again is consistent with the generalization that all d(purine)$_n$ DNAs are very poor templates for RNA polymerase (Morgan, 1970). The molecular weights of the single strands and duplex were determined in case there should be any difference leading to pronounced asymmetry of transcription, and they were found to be 211,000 and 415,000, respectively. The nucleotide analysis (see Methods) of the isolated d(T-C-C)$_n$ strand showed a molar ratio of dTp to dCp of 1:2.01. For the d(G-G-A)$_n$ strand the nucleotides were not completely resolved but assuming a Gaussian curve the ratio of dAp to dGp was 1:2.2.

Final characterization of d(T-C-C)$_n$·d(G-G-A)$_n$ as a duplex was a melting analysis in 0.1 M NaCl–1 mM sodium phosphate (pH 7.4) and 0.01 M NaCl–0.1 mM sodium phosphate (pH 7.4), conditions identical with those used by Wells *et al.* (1970) for synthetic polymers. A sharp transition was obtained with a T_m of 71° in 0.01 N NaCl and 89.5° in 0.1 M NaCl. The transition occurred within 1–2° with 30% hyperchromicity. There was no evidence for any other sharp transition. A slight decrease in optical density was observed as the temperature was raised up to the T_m possibly due to a slight unwinding of the DNA (Wang, 1969) which was also seen for some polymers by Wells *et al.* (1970). The difference in T_m for a tenfold rise in NaCl concentration is 18.5° which is the same as that found for natural DNAs. For repeating

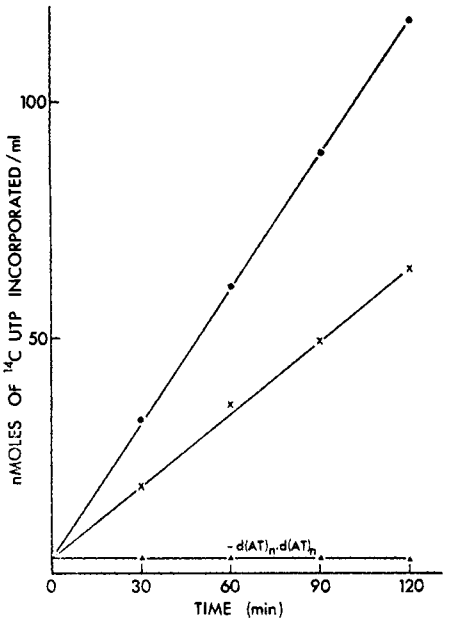

FIGURE 7: Kinetics of r(A-U)$_n$ synthesis. The standard RNA polymerase assay conditions were used with 0.15 M KCl. The reaction mixture contained per ml: ^{14}C labeled 0.5 mM UTP (1578 cpm/nmol), 0.47 mM ATP, 0.2 mg of RNA polymerase, and d(A-T)$_n$·d(A-T)$_n$, 0.018 or 0.036 OD_{260} unit. The blank contained all components except DNA.

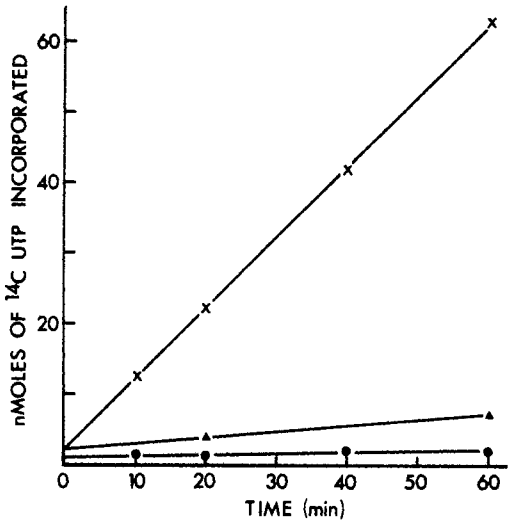

FIGURE 8: Kinetics of r(A-U)$_n$ synthesis to determine contaminating d(A-T)$_n$·d(A-T)$_n$ in d(T-G)$_n$·d(C-A)$_n$ and d(T-C)$_n$·d(G-A)$_n$. The assay conditions were as for Figure 7 with the template concentrations per ml as follows: d(T-C)$_n$·d(G-A)$_n$, 0.47 OD_{260} (●); d(T-C)$_n$·d(G-A)$_n$, 0.47 OD_{260} + d(A-T)$_n$·d(A-T)$_n$, 0.032 OD_{260} (×); d(T-G)$_n$·d(C-A)$_n$, 0.42 OD_{260} (▲).

trinucleotide DNAs the difference is 12 ± 3° (Wells *et al.*, 1970) so that again the unpredictability of synthetic polymers is illustrated. If the T_m's at 0.1 M NaCl are plotted *vs.* % GC content for the known synthetic d(purine)$_n$·d(pyrimidine)$_n$ DNAs (dA$_n$·dT$_n$, d(T-C-C)$_n$·d(G-A-A)$_n$, d(T-C)$_n$·d(G-A)$_n$, d(T-C-C)$_n$·d(G-G-A)$_n$, and dC$_n$·dG$_n$) a straight line can be drawn in which only d(T-T-C)$_n$·d(G-A-A)$_n$ falls off and below the line.

Other Factors Affecting DNA Replication. To obtain DNAs with high molecular weights it was necessary to inhibit the endonuclease I in DIII with tRNA. However, DNA synthesis was then so sluggish that pancreatic DNase was added to stimulate the reaction. DIII is necessary since it contains the factor(s) necessary to prevent clc-DNA forming (see Methods). Endoclease I will not suffice (Coulter *et al.*, 1974). Figure 6 shows the effect of the level of DNase on the molecular weight of d(T-G)$_n$·d(C-A)$_n$, and the amount of synthesis; 25 ng/ml of pancreatic DNase is a good compromise for a reasonable synthetic rate without an appreciable reduction in molecular weight.

Assay for d(A-T)$_n$·d(A-T)$_n$ Contaminating a DNA Preparation. We have recently shown (Morgan, 1970) that the transcription of most synthetic DNAs was strongly inhibited by KCl at concentrations greater than 0.15 M. However, transcription of d(A-T)$_n$·d(A-T)$_n$ to r(AU)$_n$ is stimulated by 0.20 M KCl and proceeds linearly until most of the triphosphates are utilized (Figure 7). With ATP and UTP alone, the possible RNA products are rA$_n$, rU$_n$, and r(AU)$_n$. Since only UTP is labeled, rA$_n$ synthesis will not register in the assay. Poly(U) synthesis with dA$_n$·dT$_n$ as template is almost negligible at 0.2 M KCl (Morgan, 1970) and in any case, ceases after a few minutes. The molar ratio of UTP incorporated to dT$_n$ is less than half, perhaps because of triplex formation (Murray and Morgan, 1973). Transcription of d(AT)$_n$·d(AT)$_n$, on the other hand, proceeds at about onefold/min. Thus, after 2 hr of d(AT)$_n$·d(AT)$_n$ transcription there is about 100 times as much r(AU)$_n$ as template supplied under saturating RNA polymerase conditions. An additional advantage of the high KCl concentration is that the binding of RNA polymerase to other DNAs is minimized so that very high levels of the enzyme are not required for saturation. In Figure 7 it is seen that the incorporation of [^{14}C]UTP into r(AU)$_n$ is linearly related to d(AT)$_n$·d(AT)$_n$ concentration. In Figure 8, assays for contamination of d(T-C)$_n$·d(G-A)$_n$ and d(T-G)$_n$·d(C-A)$_n$ by d(A-T)$_n$·d(A-T)$_n$ are shown. d(T-C)$_n$·d(G-A)$_n$ is not contaminated at all since no [^{14}C]UTP incorporation is observed in the presence of ATP. However d(T-G)$_n$·d(C-A)$_n$ is apparently contaminated with d(A-T)$_n$·d(A-T)$_n$ to the extent of 0.6%. When a small, known, amount of d(A-T)$_n$·d(A-T)$_n$ is added to either template (as shown in the figure for d(T-C)$_n$·d(G-A)$_n$) the expected increase in [^{14}C]UTP incorporation takes place, confirming that other DNA templates do not interfere with the transcription of d(A-T)$_n$·d(A-T)$_n$ under these conditions. UTP is also usually labeled because of the possibility of some poly(A) polymerase activity in RNA polymerase preparations (Chamberlin and Berg, 1964). The dependence of UTP incorporation on the presence of ATP can be shown although it is advisable to purify the UTP as it sometimes contains trace amounts of ATP. We have tested various natural DNAs (T7, *E. coli*, and calf thymus DNAs) by this assay for d(A-T)$_n$·d(A-T)$_n$ and there is some indication that they all contain such sequences, the amount increasing in the order T7 < *E. coli* < calf thymus DNA. However, nearest neighbor analyses of the product are desirable to confirm this conclusion.

Discussion

Understanding the complex enzymology and physical properties of DNA and RNA *in vivo* depends to a large extent on the study of simple model systems. The preparation of DNAs with repeating nucleotide sequences was accomplished only when defined oligonucleotides could be chemically synthesized (Wells *et al.*, 1967). Incubating complementary oligonucleotides of repeating sequences with DNA polymerase and the appropriate dNTPs produces DNAs of much longer chain length but still containing the strictly repeating

sequence present in the oligomer templates. The mechanism of this reaction is still unclear. Once such a DNA is obtained it can be further replicated, obviating the necessity of going back to chemically synthesized materials. In principle such DNAs could be replicated indefinitely and it is desirable to build up large quantities of these DNAs.

Over the past few years it has become evident that a rigorous characterization of the product of replication is always required. There are three major problems which can arise out of repeated replication.

(1) The replicative act is not a perfect one and *in vitro* mutations may be expected to arise. Kornberg's group has shown that in $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ replication (Trautner *et al.*, 1962), the level of misincorporation is extremely low. However, this polymer is exceptional in many of its properties, and the fact that it is the frequent product of unprimed DNA synthesis suggests the enzyme may have an innate tendency to form such a strictly alternating sequence. We have not sufficient data yet to judge whether the nearest neighbor analyses indicating incorrect transfers are due to mutations arising within the polymer or due to small amounts of other DNAs being formed (see Burd and Wells, 1970). It is instructive to note that transcriptional "mutations" (mispairing) are easily induced with $d(\text{pyrimidine})_n \cdot d(\text{purine})_n$ DNAs (Paetkau *et al.*, 1972). G can be made to base pair with T, such that rG_n is made from $d(T\text{-}C)_n \cdot d(G\text{-}A)_n$ if ATP is omitted. However, G cannot be made to pair with T in the polymer $d(T\text{-}G)_n \cdot d(C\text{-}A)_n$ to form $r(G\text{-}C)_n$. This stresses the importance of the nearest neighbor in transcriptional errors.

(2) The synthesis of unwanted DNAs, apparently unprimed, during the course of the replication of a synthetic DNA is a particular hazard. Burd and Wells have shown that depending on the pH and other parameters, $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$, $dG_n \cdot dC_n$, $d(G\text{-}C)_n \cdot d(G\text{-}C)_n$, and $d(I\text{-}C)_n \cdot d(I\text{-}C)_n$ can all be formed *de novo* by the *E. coli* and *Micrococcus luteus* DNA polymerases (Burd and Wells, 1970). Our use of high dNTP concentrations produces high levels of DNA, inhibits nucleases, and has the added benefit of depressing $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ synthesis. However, some recent transcription studies and nearest neighbor analyses indicate that in the presence of DIII, $(pyr)_n \cdot (pur)_n$ DNA, in addition to the above polymers, may contaminate the product of the replication of repeating polymers.

(3) DIII contains a protein factor which either prevents clc-DNA forming or removes such structures as they are formed. However, DIII is still a rather impure fraction containing some nucleotidic material. The preliminary indication that $(pyr)_n \cdot (pur)_n$ DNAs may form as a result of DIII suggests it may be enriched for $(pyr)_n \cdot (pur)_n$ DNA.

The newly synthesized DNA $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ has all the characteristics and peculiarities of $(pyr)_n \cdot (pur)_n$ DNAs. That the polymer $d((T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ has indeed been synthesized is evident then from many lines of investigation: the original oligonucleotide templates used, the nearest neighbor and nucleotide analyses, and physicochemical studies. In addition to the evidence cited in Results, we have already shown (Paetkau *et al.*, 1972) that after depurination of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ isotopically labeled in the pyrimidine strand $d(T\text{-}C\text{-}C)_n$ could be isolated in 85% yield and with unchanged molecular weight. This confirms that there are no covalent links between the purine and pyrimidine strands. The buoyant density is that expected for a DNA with 67% GC content as is the T_m if only $(pyr)_n \cdot (pur)_n$ DNAs are considered. The strands of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ can be readily separated in alkaline density gradients as expected from the difference in the content of ionizable bases in the two strands. It also will dismutate into a triplex and $d(G\text{-}G\text{-}A)_n$ at pH 6, typical of all $(pyr)_n \cdot (pur)_n$ DNAs containing C (unpublished data). Transcription studies (Morgan, 1970) have shown that the strands are read asymmetrically by RNA polymerase. In the presence of GTP aolne, $r(G)_n$ is synthesized (Paetkau *et al.*, 1972). $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}C\text{-}A)_n$ (67% GC content) completes a series of $(pyr)_n \cdot (pur)_n$ DNAs ranging from $dA_n \cdot dT_n$ (0% GC) to $dG_n \cdot dC_n$ (100% GC) with $d(T\text{-}C)_n \cdot d(G\text{-}A)_n$ (50% GC) and $d\text{-}(T\text{-}T\text{-}C)_n \cdot d(G\text{-}A\text{-}A)_n$ 33% GC) having intermediate GC content. They all show the same phenomenon of forming triple stranded structures.

A more detailed and thorough investigation of the properties of $d(T\text{-}C\text{-}C)_n \cdot d(G\text{-}G\text{-}A)_n$ will be reported in a future publication of the unique properties of $(pyr)_n \cdot (pur)_n$ DNAs, especially with regard to dismutation of the duplex. Care was then taken in this work to avoid conditions of low pH which might lead to dismutations, but for this reason spectra will be published later when the dynamics of the systems are understood.

The assay for $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ has proved extremely sensitive, readily detecting as little as 0.01% contaminating a synthetic DNA. Since $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ is a major component of crab satellite DNA (for a review see Laskowski (1972)) it suggests that it may have a functional role to play if not an informational one. We have preliminary evidence of $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ regions in T7, *E. coli*, and calf thymus DNAs. However, the complexity of natural DNAs make it desirable to isolate the putative $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ regions, in case the $r(A\text{-}U)_n$ synthesis is a primed reaction and does not require a template. $r(A\text{-}U)_n$ itself is a very poor template for RNA polymerase.

Acknowledgments

We thank Mr. M. Burrington for help with the nearest neighbor analyses and Dr. H. G. Khorana for the gift of $d(C\text{-}C\text{-}T)_4$ and $d(G\text{-}G\text{-}A)_3$. Dr. Larry Brox kindly performed the nucleotide analyses.

References

Beyersmann, D., and Schramm, G. (1968), *Biochim. Biophys. Acta 159*, 64.

Brown, P. R. (1970), *J. Chromatogr. 52*, 257.

Burd, J. F., and Wells, R. D. (1970), *J. Mol. Biol. 53*, 435.

Byrd, C., Ohtsuka, E., Moon, M. W., and Khorana, H. G. (1965), *Proc. Nat. Acad. Sci. U. S. 53*, 79.

Chamberlin, M., and Berg, P. (1962), *Proc. Nat. Acad. Sci. U. S. 48*, 81.

Chamberlin, M., and Berg, P. (1964), *J. Mol. Biol. 8*, 708.

Chervenka, C. H. (1969), A Manual of Methods for the Analytical Ultracentrifuge, Palo Alto, Calif., Spinco Division, Beckman Instruments, Inc.

Coulter, M., Flintoff, W., Paetkau, V., Rulleyblank, D., and Morgan, A. R. (1974), *Biochemistry 13*, 1603.

Jovin, T. M., Englund, P. T., and Bertsch, L. L. (1969), *J. Biol. Chem. 244*, 2996.

Khorana, H. G., Buchi, H., Ghosh, H., Gupta, N., Jacob, T. M., Kossel, H., Morgan, R., Narang, S. A., Ohtsuka, E., and Wells, R. D. (1966), *Cold Spring Harbor Symp. Quant. Biol. 31*, 39.

Laskowski, Sr. M. (1972), *Progr. Nucl. Acid Res. Mol. Biol. 12*, 161.

Le Pecq, J.-B., and Paoletti, C. (1966), *Anal. Biochem. 17*, 100.

Le Pecq, J.-B., and Paoletti, C. (1967), *J. Mol. Biol. 27*, 87.
Morgan, A. R. (1970), *J. Mol. Biol. 52*, 441.
Morgan, A. R., and Paetkau, V. (1972), *Can. J. Biochem. 50*, 210.
Morgan, A. R., and Wells, R. D. (1968), *J. Mol. Biol. 37*, 63.
Murray, N. L., and Morgan, A. R. (1973), *Can. J. Biochem. 51*, 436.
Narang, S. A., Jacob, T. M., and Khorana, H. G. (1967), *J. Amer. Chem. Soc. 89*, 2167.
Nishimura, S., Jacob, T. M., and Khorana, H. G. (1964), *Proc. Nat. Acad. Sci. U. S. 82*, 1494.
Paetkau, V. (1969), *Nature (London) 224*, 370.
Paetkau, V., Coulter, M. B., Flintoff, W. F.. and Morgan, A. R. (1972), *J. Mol. Biol. 71*, 293.
Paetkau, V., and Coy, G. (1972), *Can. J. Biochem. 50*, 142.
Paetkau, V. H., and Khorana, H. G. (1971), *Biochemistry 10*, 1511.
Richardson, C. C., Schildkraut, C. L., Aposhian, H. V., and Kronberg, A. (1964), *J. Biol. Chem. 239*, 222.
Studier, F. W. (1965), *J. Mol. Biol. 11*, 373.
Sueoka, N., Marmur, J., and Doty, P. (1959), *Nature (London) 183*, 1429.
Trautner, T. A., Swartz, M. N., and Kornberg, A. (1962), *Proc. Nat. Acad. Sci. U. S. 48*, 449.
Vinograd, J. (1963), *Methods Enzymol. 6*, 854.
Wang, J. C. (1969), *J. Mol. Biol. 43*, 25.
Wells, R. D., Buchi, H., Kossel, H., Ohtsuka, E., and Khorana, H. G. (1967), *J. Mol. Biol. 27*, 265.
Wells, R. D., and Larson, J. E. (1972), *J. Biol. Chem. 247*, 3405.

In Vitro Synthesis and Detection of Deoxyribonucleic Acids with Covalently Linked Complementary Sequences†

Marion Coulter, Wayne Flintoff,* Verner Paetkau, David Pulleyblank, and A. Richard Morgan

ABSTRACT: An activity has been demonstrated which blocks the accumulation of DNA with covalently linked, complementary (clc) sequences during the copying of defined DNAs with DNA polymerase I of *Escherichia coli*. The factor is necessary for the synthesis of both d(T-G)$_n$·d(C-A)$_n$ and d(T-T-G)$_n$·d(C-A-A)$_n$ free of clc sequences. A modification of the ethidium fluorescence method for measuring clc sequences is described. This facilitates measurement of clc structures in natural polynucleotides without interference by nonspecific, intrastrand associations. The defined polymer d(T-T-G)$_n$·d(C-A-A)$_n$ has been prepared with separable strands. Defined DNAs of the type d(purine)$_n$·d(pyrimidine)$_n$ do not give rise to clc-DNA, even in the absence of the above factor, in our experience. Since the complementary strands of d(T-C)$_n$·d(G-A)$_n$ have the same buoyant densities in alkaline CsCl, an independent method was developed to separate them physically, and thereby prove that the polymer was of the non-clc type. In this method, the polymer was treated with a water-soluble carbodiimide under mild conditions, to derivatize T and G residues. The strands were separated by neutral CsCl equilibrium centrifugation, and dederivatized by mild alkali to regenerate the original strands.

Naturally occurring DNAs generally consist of complementary strands which may be topologically, but not covalently, linked. Several exceptions exist: newly synthesized phage T7 DNA (Barzilai and Thomas, 1970) and λ-DNA (Ihler and Kawai, 1971) appear to contain covalent linkers between complementary strands, at least under some conditions. Similarly, electron micrographs of replicating colicin E1 DNA (Fuke and Inselburg, 1972) can be interpreted to indicate covalent continuity between the complementary daughter strands at the growing fork. During the *in vitro* copying of natural (Schildkraut *et al.*, 1964) and defined (Paetkau, 1969; Harwood and Wells, 1970) DNAs by DNA polymerases, covalent linkers appear between complementary sequences (clc sequences[1]). Clc sequences also arise when DNA becomes cross-linked by agents such as mitomycin C. A technique for detecting such structures has been developed (Morgan and Paetkau, 1972), based on the enhanced fluorescence of ethidium bromide when it is bound to bihelical DNA (LePecq and Paoletti, 1967). In the present work, this technique has been modified to prevent the nonspecific, relatively unstable, bihelical structures occurring in heat-denatured, naturally occurring DNAs. Several other procedures are described which permit either separation of the complementary strands of synthetic DNAs of defined sequence, or detection of clc structures when these exist. Conditions for enzymatically synthesizing defined polymers free of clc sequences are described.

Experimental Section

Materials. Chemicals. CsCl was obtained from Pierce Chemical Co. or Schwarz/Mann and *N*-cyclohexyl-*N'*-β-(4-methylmorpholium)ethylcarbodiimide *p*-toluenesulfonate (hereafter referred to as the carbodiimide) from Aldrich Chemical Co. Technical grade Cs_2SO_4 was obtained from American Potash and Chemical Corporation and purified by recrystallization from hot water. BrdUTP was prepared by a modification (D. E. Pulleyblank and A. R. Morgan, unpublished) of the procedure of Inman and Baldwin (1964). Other materials were indicated earlier (Morgan *et al.*, 1974).

Enzymes. Endonuclease I was purified from *E. coli* by the method of De Waard and Lehman (1966). A unit of endonuclease I is defined as causing a hyperchromicity of 0.001

† From the Department of Biochemistry, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada. *Received November 27, 1973.* Supported by The Medical Research Council of Canada.

[1] Abbreviations used are: clc-DNA, DNA with covalently linked, complementary sequences; BrdUTP, deoxybromuridine 5'-triphosphate.