Biochemistry

Volume 16, Number 9 May 3, 1977

Studies on Gene Control Regions. 1. Chemical Synthesis of Lactose Operator Deoxyribonucleic Acid Segments†

D. V. Goeddel, D. G. Yansura, and M. H. Caruthers*

ABSTRACT: The chemical synthesis of lactose operator DNA segments is described. The 31-base-paired duplex contains the DNA recognized by *lac* repressor protein and twofold rotationally symmetric base pairs on either side of the tight binding region. The synthesis includes the deoxyoligonucleotides d(T-G-T-G-G), d(A-A-T-T-G-T-G-A-G), d(C-G-G-A-T-A-A-C-A-A-T-T), d(T-C-A-C-A), d(T-G-T-G-A-A-A-T-T-G-T), d(T-A-T-C-C-G-C-T-C-A-C), and d(A-A-T-T-C-C-A-C-A). These deoxyoligonucleotides were characterized by two-dimensional sequencing techniques, paper chromatography, and thin-layer chromatography.

The central aim of our work is to contribute toward understanding in precise chemical terms how genes are regulated. One aspect of this problem is to explore how various regulatory proteins interact with DNA that serves a control function in operons and chromosomes. Studies of this interaction must examine protein and nucleic acid structure that leads to tight binding complexes and also must investigate how various proteins interact cooperatively to enhance or block transcription.

One choice for such studies on the molecular basis of gene regulation is the lactose (*lac*) operon in *Escherichia coli*. This operon has been analyzed in detail by both genetic and biochemical techniques (Beckwith and Zipser, 1970). The control region of the *lac* operon contains elements of positive and negative control (Figure 1). The operator, o, is the DNA binding site for the repressor protein (Gilbert and Müller-Hill, 1967). In the absence of the specific inducer allolactose (Jobe and Bourgeois, 1972), the *lac* repressor–operator complex blocks transcription. At high levels of cAMP,[1] the catabolite gene activator protein (CAP)–cAMP complex activates transcription of the *lac* operon by DNA-dependent RNA polymerase (de Crombrugghe et al., 1971; Eron and Block, 1971). The promoter, p, is the binding site for CAP and DNA-dependent RNA polymerase (Silverstone et al., 1970; Beckwith et al., 1972). These control proteins therefore interact at specific sites on DNA located in the regulatory region of the *lac* operon. We would like to understand the precise chemical features of this recognition process. More specifically, the present work is directed toward understanding how *lac* repressor protein interacts with *lac* operator DNA. Our approach is to chemically prepare operator DNA with various sequence modifications and then study how these changes affect the operator–repressor recognition process. This paper outlines the chemical synthesis of *lac* operator DNA.

The *lac* operator DNA we have chosen to synthesize is shown in Figure 2. The rationale for this plan is as follows. In the central part of this figure is the operator region sequence of lactose operon DNA (Gilbert et al., 1973; Dickson et al., 1975). The region which binds repressor tightly is sequence 3–26 (Gilbert and Maxam, 1973). The heavy lines above and below mark base pairs that are related by a twofold rotational axis of symmetry. Sequence changes which are found in different classes of operator constitutive mutants (Gilbert et al., 1974) are shown at the appropriate sites below this operator DNA. These mutants were isolated either by Gilbert (1973) or by Sadler and Smith (Smith and Sadler, 1971; Sadler and Smith, 1971). Also shown in this figure for various mutant classes are the *P* values (the percent constitutivity or the ratio of basal to the induced β-galactosidase level) and the in vitro lifetimes of the repressor–DNA complex (Jobe et al., 1974). For wild-type *E. coli*, the *P* value is 0.1 and the half-life for the repressor–DNA complex is 75 min. Using these data we constructed the synthetic plan shown in the top part of this figure. The brackets enclose chemically synthesized compounds. Using T4 ligase (Khorana et al., 1972, 1976) these compounds

† From the Department of Chemistry, University of Colorado, Boulder, Colorado 80309. *Received November 18, 1976.* This research was supported by grants from the National Institutes of Health (GM 21120 and GM 21644), a Cottrell Research Grant from Research Corporation, a Biomedical Sciences Support Grant from the National Institutes of Health to the University of Colorado, an Institutional Research Grant from the American Cancer Society to the University of Colorado, a National Science Foundation equipment grant (BMS 75-14541), and the University of Colorado. M.H.C. was supported by a Career Development Award from the National Institutes of Health (1 K04 GM 00076).

[1] Abbreviations used are: cAMP, cylic adenosine 3′,5′-monophosphate; TPS, triisopropylbenzenesulfonyl chloride; CE, cyanoethyl; DEAE, diethylaminoethyl; TEAB, triethylammonium bicarbonate; MeOTr, monomethoxytrityl; ib, isobutyryl; an, *p*-anisoyl; bz, benzoyl; MS, mesitylenesulfonyl chloride; CAP, catabolite gene activator protein; DCC, dicyclohexylcarbodiimide; TLC, thin-layer chromatography. The one-letter symbols for nucleotides and the symbols for polynucleotides are according to the IUPAC-IUB Commission on Biochemical Nomenclature Recommendations [(1970) *Biochemistry 9*, 4022].

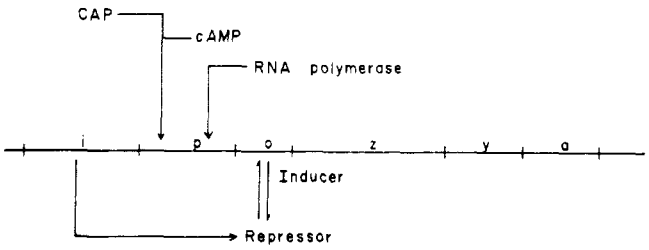

FIGURE 1: The *lac* operon. Symbols i, z, y, and a represent structural genes for the proteins *lac* repressor, β-galactosidase, permease, and transacetylase, respectively. Symbols o and p represent regions of the operon which serve a control function.

have been joined in various combinations to form either the complete or many different partial operator DNA regions (Yansura et al., 1977).

This synthetic plan was designed so that research could be conducted on how repressor protein interacts with operator DNA. Structural areas of particular interest are inner symmetric regions (base pairs 6–11, 21–26) and the less symmetric central region (base pairs 12–20). Our synthetic plan was designed so that selected base pairs within these regions could be readily modified. The chemical syntheses were carried out using procedures developed by Khorana and his associates (Khorana, 1968). Therefore syntheses begin at the 5′ end of a chemical fragment and proceed to the 3′ end. As a consequence, a series of intermediate deoxyoligonucleotides was prepared, all of which have the same 5′ end. Many of our investigations will involve changing the sequence of operator DNA. If this is done chemically, then changes are much easier near the 3′ end of deoxyoligonucleotides because intermediate deoxyoligonucleotides can be used and extended with altered sequence. Our synthetic plan was designed with these concepts in mind. Sequences in a highly symmetric region (base pairs 21–26) and a less symmetric region containing many operator constitutive mutants (base pairs 10–14) can be readily changed since 3′ ends of deoxyoligonucleotides are located within these regions. Already we have inserted base pair changes at site 13 (Loder and Caruthers, unpublished experiments). We have included a major portion of the outer symmetry region (Dickson et al., 1975) in our synthesis (base pairs 1–4, 28–31). Our reasons for doing this are twofold. First, although evidence is lacking at this time, these regions may be important for the operator–repressor interaction. Also, including these additional sequences in the synthesis permits considerable maneuverability for modification of operator DNA because more 3′ ends are available for synthetic alterations. The present paper reports the chemical synthesis of *lac* operator DNA sequences. Brief reports on this work have appeared during the past 2 years (Caruthers et al., 1974, 1975). Recently, the synthesis of the central 21 base pairs of *lac* operator DNA has been reported (Itakura et al., 1975).

Materials and Methods

Materials. Nucleotides and nucleosides were purchased from Papierwerke Waldhof-Aschaffenburg. The protected mononucleotides dpbzA, dpanC, dpibG and their cyanoethylphosphoryl, 3′-*O*-acetyl, and 3′-*O*-isobutyryl derivatives were prepared as described elsewhere (Khorana et al., 1961; Büchi and Khorana, 1972). The protected nucleosides, d(MeOTr)T and d(MeOTr)anC, were prepared as published previously (Schaller et al., 1963). Di-, tri-, and other oligonucleotide blocks containing a 5′-phosphate were prepared as previously described (Weber and Khorana, 1972). Blocks containing *N*-isobutyryldeoxyguanosine were reisobutyrylated before use. Bacterial alkaline phosphatase and snake venom phosphodiesterase were purchased from Worthington Chemical Co. T4 polynucleotide kinase was prepared as described elsewhere (Panet et al., 1973). Materials such as cellulose acetate strips (Schleicher and Schuell), Cellulosepulver MN300HR, and MN300 DEAE-cellulose for thin-layer chromatography (Brinkman), crude yeast RNA (Gallard-Schlesinger), precoated silica gel plates (Eastman-6064), and Eastman RP-royal blue X-OMAT x-ray film were purchased from commercial sources. TPS and MS were obtained commercially (Aldrich Chemical Co.) and recrystallized from anhydrous pentane immediately before use. Pyridine (reagent grade, J. T. Baker Co.) was purified as described previously (Kössel et al., 1967). Diisopropylethylamine (Aldrich Chemical Co.) was distilled from toluenesulfonyl chloride before use. Triethylamine and hydracrylonitrile from Eastman were used without further purification. Dicyclohexylcarbodiimide (DCC) was purchased from Aldrich Chemical Co.

Methods. DEAE-cellulose column chromatography, paper chromatography, paper electrophoresis, and detection of the monomethoxytrityl group were as described previously (Weber and Khorana, 1972). Phosphorylation of deoxyoligonucleotide 5′-OH groups using [γ-^{32}P]ATP and polynucleotide kinase was performed according to previous procedures (Sgaramella and Khorana, 1972; van de Sande et al., 1972). Two-dimensional analysis of synthesized deoxyoligonucleotides was as described by published procedures (Ling, 1972a,b; Sanger et al., 1973; Galibert et al., 1974; Jay et al., 1974). Homomix V (Jay et al., 1974), 20 × 20 cm glass plates, and dry cellulose acetate paper were used for these analyses.

The following method was generally used for condensation reactions. In all syntheses, the coupling was between a free 5′-phosphate and 3′-OH on mono- or oligonucleotides. All other reactive groups were masked with protecting groups. Both reaction components were dissolved in anhydrous pyridine and the mixture rendered anhydrous by repeated evaporation of added dry pyridine in a vacuum. Each pyridine addition was in the atmosphere of a drybox containing phosphorus pentoxide. During the last pyridine evaporation, TPS or MS was added and the solution concentrated to the minimal volume necessary for complete solubilization of the reaction components. The reaction mixtures were kept at room temperature 2.5–3.0 h (MS) or 5 h (TPS) with exclusion of moisture. They were then cooled in ice–water baths and diisopropylethylamine as a 1 M solution in pyridine added (2 mmol/mmol sulfonyl chloride plus 1 mmol/mmol phosphoryl dissociation). Water was added to a final concentration of 20–30% and solutions left at room temperature for 2 days.

Cyanoethyl, 3′-*O*-acetyl, and 3′-*O*-isobutyryl groups were removed by saponification. The reaction solutions were diluted to 50% water. After cooling to 0 °C, enough 2 M sodium hydroxide was added to give final solutions 1 M in alkali after accounting for all anions present. Occasionally ethanol had to be added to obtain homogeneous solutions. After 5 min in ice–water baths, solutions were neutralized with excess pyridinium Dowex 50 ion-exchange resin. The mixtures were poured into columns and resin washed thoroughly with 20% aqueous pyridine. Filtrates and washings were pooled and fractionated on DEAE-cellulose columns.

Fractionation of protected deoxyoligonucleotides on DEAE-cellulose was by a combination of salt and alcohol gradients. The volatile salt was triethylammonium bicarbonate. The general types of gradients used were outlined previously (Khorana et al., 1976). Column effluents were monitored by

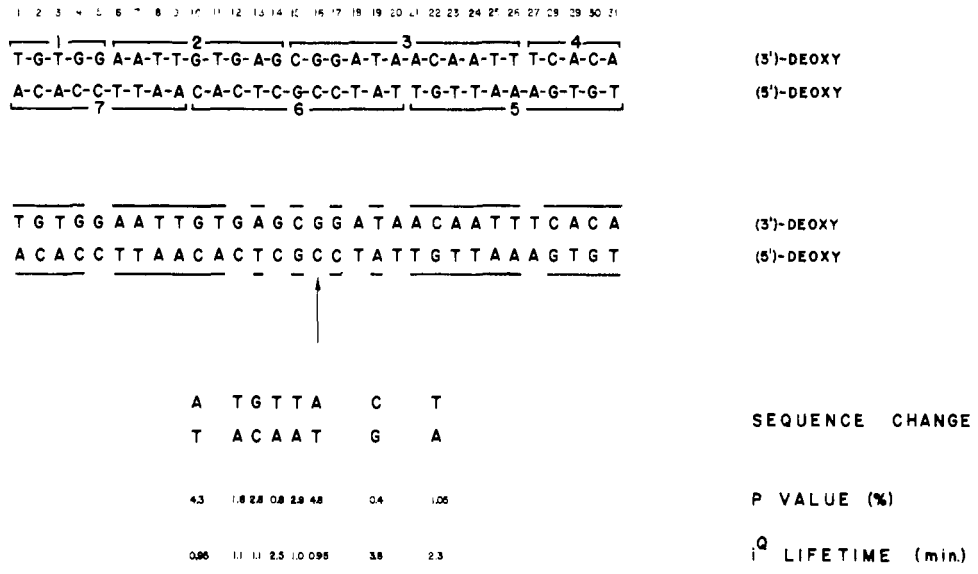

FIGURE 2: The plan for synthesis of lactose operator DNA. The duplex shown includes the DNA that binds repressor tightly. The deoxyribooligonucleotide segments are shown within the brackets, the serial number of the segment being inserted into the brackets.

TABLE I: Summary of Reaction Conditions, Recoveries, and Yields for Segments 1, 2, and 5.

Expt No.	Starting Chain (3'-OH component) (A)	Amount (mmol)	Oligonucleotide Block (5'-phosphate end) (B)	Amount (mmol) (B/A)	MS[a] or TPS (equiv/P)	Reaction Time (h)	Starting Chain Recovered (%)	Block[b] Recovered (%)	Product	Yield (%)
	d(MeOTr)T	5.80	dpibG-T(ib)	2.9 (0.5)	1.0	5		32	d(MeOTr)T-ibG-T	34
	d(MeOTr)T-ibG-T	0.78	dpibG-ibG(ib)	1.3 (1.7)	1.0	5	26	32	d(MeOTr)T-ibG-T-ibG-ibG	28
	d(MeOTr)T-ibG-T	0.40	dpibG-bzA(ib)	1.8 (4.4)	1.0	5	22	57	d(MeOTr)T-ibG-T-ibG-bzA	21
1	dCEpbzA-bzA	3.5	dpT(Ac)	14 (4.0)	0.7†	2.8	9		dpbzA-bzA-T	45
2	dCEpbzA-bzA	3.0	dpT(Ac)	9.5 (3.2)	0.7†	2.3	8		dpbzA-bzA-T	48
3	dCEpbzA-bzA	2.0	dpT(Ac)	10.8 (5.4)	1.0	5			dpbzA-bzA-T	36
1	dCEpbzA-bzA-T	2.6	dpT(Ac)	12.5 (4.8)	0.7†	3			dpbzA-bzA-T-T	39
2	dCEpbzA-bzA-T	0.4	dpT(Ac)	2.5 (5.6)	1.0	5			dpbzA-bzA-T-T	23
	dCEpbzA-bzA-T-T	0.8	dpibG-T(ib)	1.6 (2.0)	1.0	5	12		dpbzA-bzA-T-T-ibG-T	28
	d(MeOTr)T-ibG-T-ibG-bzA	0.08	dpbzA-bzA-T-T-ibG-T(ib)	0.11 (1.25)	1.0	5	61	44	d(MeOTr)T-ibG-T-ibG-bzA-bzA-bzA-T-T-ibG-T	19
	dCEpibG-bzA	1.4	dpibG(ib)	7.0 (5.0)	1.0	5			dpibG-bzA-ibG	9
	dCEpbzA-bzA-T-T-ibG-T	0.04	dpibG-bzA-ibG(ib)	0.04 (1.0)	1.0	5	21	39	dpbzA-bzA-T-T-ibG-T-ibG-bzA-ibG	12

[a] TPS unless indicated with †, equiv per phosphate group dissociation in the two components. [b] Percent of the original amount for further condensation without additional purification.

absorbance and purity checked by thin-layer and paper chromatography before and after removal of protecting groups. *N*-Acyl protecting groups were removed by treating aliquots of column fractions with concentrated ammonium hydroxide in a sealed tube at 50 °C for 2–3 h. Solutions were diluted to 50% pyridine and evaporated under vacuum. The pH was controlled by addition of dilute TEAB buffer. Monomethoxytrityl groups were removed from N-deprotected compounds by room temperature treatment with 80% acetic acid for 1 h. After monitoring the purity of column effluents, the desired protected deoxyoligonucleotides were pooled, diluted to 50% with pyridine, concentrated in vacuo at room temperature, and collected as precipitates from anhydrous ether.

Isolation of deoxyoligonucleotides free of protecting groups was by gradient elution on columns of DEAE-cellulose packed in 7 M urea. Sodium chloride was the eluting salt. Urea and salt were removed by passage through a column of Bio-Gel P-2 (Bio-Rad). Samples were stored frozen at −80 °C in water adjusted to pH 7 with ammonium hydroxide.

Results

The plan outlined in Figure 1 required the synthesis of seven deoxyoligonucleotides ranging in size from pentanucleotide to dodecanucleotide. These deoxyoligonucleotides are indicated by brackets with the segment numbers inserted into the brackets. Because of the twofold rotational symmetry in this DNA, several deoxyoligonucleotides such as d(A-A-T-T) and d(A-A-T-T-G-T) can be used more than once. Wherever possible, the syntheses were designed so that multiple use of such blocks was possible.

TABLE II: Summary of Reaction Conditions, Recoveries, and Yields for Segments 4 and 7.

Starting Chain (3′-OH component) (A)	Amount (mmol)	Oligonucleotide Block (5′-phosphate end) (B)	Amount (mmol) (B/A)	MS[a] or TPS (equiv/P)	Reaction Time (h)	Starting Chain Recovered (%)	Block[b] Recovered (%)	Product	Yield (%)
dCEpbzA-anC	1.4	dpbzA(Ac)	6.3 (4.3)	0.7†	2.75	31		dpbzA-anC-bzA	27
dCEpT-anC	0.2	dpbzA-anC-bzA-(Ac)	0.1 (0.5)	1.0	5			dpT-anC-bzA-anC-bzA	31
dCEpbzA-bzA-T-T[c]	0.2	dpanC-anC(Ac)	0.5 (2.5)	1.0	5			dpbzA-bzA-T-T-anC-anC	35
dCEpbzA-bzA-T-T-anC-anC	0.07	dpbzA-anC-bzA-(Ac)	0.3 (3.9)	1.0	5		39	dpbzA-bzA-T-T-anC-anC-bzA-znC-bzA	22

[a] TPS unless indicated with †, equiv per phosphate group dissociation in the two components. [b] Percent of the original amount for further condensations without additional purification. [c] Synthesis summarized in Table I.

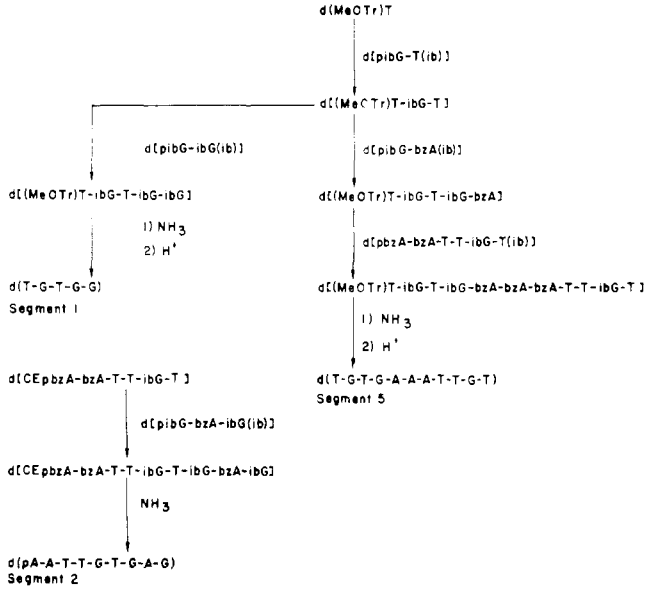

FIGURE 3: Steps in the synthesis of segments 1, 2, and 5.

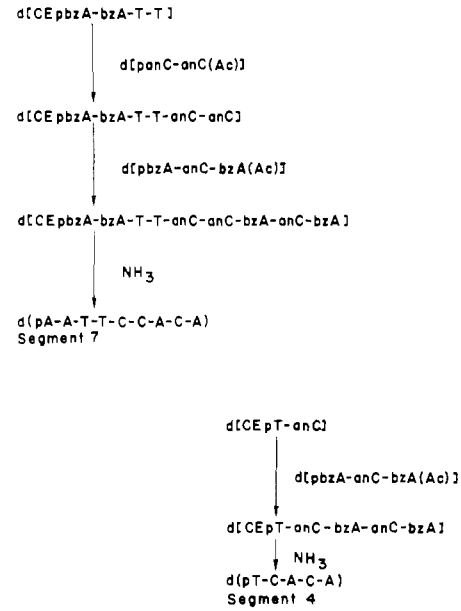

FIGURE 4: Steps in the synthesis of segments 4 and 7.

Segments 1, 2, and 5 were synthesized using the steps shown in Figure 3. A summation of the synthetic steps is shown in Table I. For the synthesis of segments 1 and 5, the first step was condensation of the nucleoside, d(MeOTr)T, with the dinucleotide d[pibG-T(ib)]. The condensing agent was TPS and the product was isolated by a partition method developed previously (Caruthers and Khorana, 1972). To complete the synthesis of segment 1, the trinucleotide was condensed with the dinucleotide d[pibG-ibG(ib)] and the product purified by anion-exchange chromatography. The synthesis of segment 5 utilized the same trinucleotide but the dinucleotide was d[pibG-bzA(ib)]. Again the product pentanucleotide was isolated by anion-exchange chromatography. As the final step in this synthesis, d[pbzA-bzA-T-T-ibG-T(ib)] was condensed with the pentanucleotide. The product in 19% yield was isolated by anion-exchange chromatography. As can be seen in Table I, the ratio of pentanucleotide to hexanucleotide in the condensation step was only 1.25 and TPS was used for the condensation. Less degradation of these rather large deoxyoligonucleotide blocks occurs when the ratio is close to equivalence and TPS is used as the activating agent. Approximately 61 and 44% of the methoxytrityl segment and hexanucleotide block, respectively, were recovered. Since these compounds have been used for many different condensations, a high recovery was desirable. The hexanucleotide d[pbzA-bzA-T-T-ibG-T] was synthesized in a stepwise fashion using the β-cyanoethyl group to protect the 5′-phosphate (Weber and Khorana, 1972). Finally, segment 2 was prepared using equivalent molar amounts of d[CEpbzA-bzA-T-T-ibG-T] and d[pibG-bzA-ibG(ib)]. The isolated yield of homogeneous nonanucleotide was 12%.

For the synthesis of segments 4 and 7, the trinucleotide d[pbzA-anC-bzA] was used twice. The tetranucleotide d[pbzA-bzA-T-T] used in segment 7 was recovered from the synthesis of d[pbzA-bzA-T-T-ibG-T]. The multiple use of these blocks reduced significantly the chemical synthesis outlined in Figure 4 and summarized in Table II. Segment 7 was prepared by condensation of d[CEpbzA-bzA-T-T] with d[panC-anC(Ac)]. The resulting hexanucleotide was obtained pure in 35% yield. Further condensation with d[pbzA-anC-bzA(Ac)] gave the nonanucleotide in 22% yield. In a similar fashion, segment 4 was prepared by condensation of dCEpT-anC] with d[pbzA-anC-bzA(Ac)]. The pentanucleotide in fully protected form was isolated by anion-exchange chromatography before removing the β-cyanoethyl group. The product has a net negative charge of 5 whereas the pyrophosphate of d[pbzA-anC-bzA(Ac)] has six negative charges.

TABLE III: Summary of Reaction Conditions, Recoveries, and Yields for Segment 3.

Expt No.	Starting Chain (3′-OH component) (A)	Amount (mmol)	Oligonucleotide Block (5′-phosphate end) (B)	Amount (mmol) (B/A)	MS[a] or TPS (equiv/P)	Reaction Time (h)	Starting Chain Recovered (%)	Block[b] Recovered (%)	Product	Yield (%)
1	d(MeOTr)anC	2.8	dpibG-ibG(ib)	1.4 (0.5)	1.0	5		43	d(MeOTr)anC-ibG-ibG	32
2	d(MeOTr)anC		dpibG-ibG(ib)					38	d(MeOTr)anC-ibG-ibG	48
	d(MeOTr)anC-ibG-ibG		dpbzA-T(Ac)				13	66	d(MeOTr)anC-ibG-ibG-bzA-T	17
	dCEpbzA-bzA	2.0	dpanC(Ac)	10 (5)	0.7†	2.75			dpbzA-bzA-anC	32
	d(MeOTr)anC-ibG-ibG-bzA-T	0.2	dpbzA-bzA-anC-(Ac)	0.6 (3.2)	1.0	5	28	42	d(MeOTr)anC-ibG-ibG-bzA-T-bzA-bzA-anC	44
	d(MeOTr)anC-ibG-ibG-bzA-T-bzA-bzA-anC	0.05	dpbzA-bzA-T-T(Ac)[c]	0.1 (1.9)	5	34	20		d(MeOTr)anC-ibG-ibG-bzA-T-bzA-bzA-anC-bzA-bzA-T-T	25

[a] TPS unless indicated with †, equiv per phosphate group dissociation in the two components. [b] Percent of the original amount for further condensations without additional purification. [c] Synthesis summarized in Table I.

Separation was therefore possible. Attempted separation after removal of the β-cyanoethyl group would have been unsuccessful since both components then have a net negative charge of six. Furthermore, TPS which generates very little pyrophosphate was used during the condensation step. The pentanucleotide was isolated pure in 31% yield.

The synthesis of d(C-G-G-A-T-A-A-C-A-A-T-T) was performed using the steps shown in Figure 5. Table III contains a summation of the experimental data. The first step was condensation of d(MeOTr)anC with d[pibG-ibG(ib)]. The product was isolated by the partition method in 32% yield. The trinucleotide was next condensed with d[pbzA-T-Ac)] to form the pentanucleotide which was purified by anion-exchange chromatography. The next step was condensation of the pentanucleotide with d[pbzA-bzA-anC(Ac)]. The yield was 44% and the recovery of starting material was satisfactory. The final step was condensation of the octanucleotide with d[pbzA-bzA-T-T(Ac)]. Again the yield of dodecanucleotide was satisfactory (25%) and a significant amount of octanucleotide (34%) was recovered.

Synthesis of the undecanucleotide (segment 6) which represents nucleotide sequence 10–20 of the bottom DNA strand was performed by the steps shown in Figure 6 and is summarized in Table IV. The initial condensation was d(MeOTr)T with d[pbzA-T(Ac)]. The product in 62% yield was isolated by the partition procedure. The trinucleotide was condensed with d[panC-anC(Ac)] to form the pentanucleotide and the product (40%) isolated by anion-exchange chromatography. The pentanucleotide was condensed with d[pibG-anC(ib)] to give the heptanucleotide in 25% yield. Finally the undecanucleotide was formed by condensation of the heptanucleotide and d[pT-anC-bzA-anC(Ac)]. Again equivalent molar amounts of the two components were used and the yield was 12% for the undecanucleotide.

Amounts of deoxyoligonucleotides available in protected form are shown in Table V. Further analysis by removal of protecting groups followed by chromatography on DEAE-cellulose in 7 M urea indicated that segments 1, 2, 4, and 7 were homogeneous. No additional peaks or shoulders on the main peak were observed. However, segments 3, 5, and 6 fractionated by this technique into a main band which in all cases was

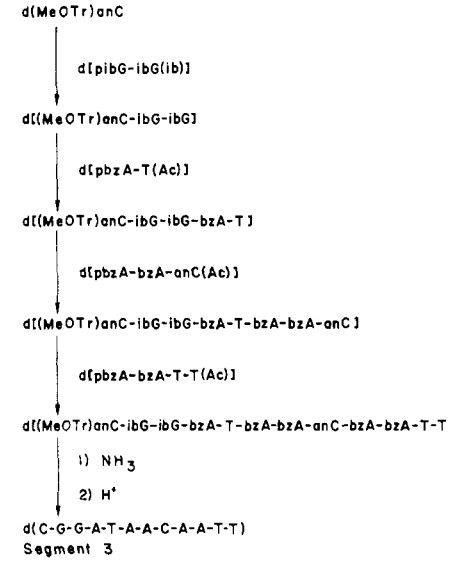

FIGURE 5: Steps in the synthesis of segment 3.

found to be the product and at least one other peak containing unidentified components. These purified segments were analyzed by the two-dimensional sequencing technique. Segment 5 was analyzed by the same technique after passage through Sephadex G-75. The results of these analyses are shown in Figures 7 and 8. All compounds studied displayed a major set of spots which confirmed the nucleotide sequence of the synthetic products. When analyzed carefully by prolonged exposure to x-ray film, all samples were found to contain minor unidentified spots. This was especially obvious with samples of d(A-A-T-T-C-C-A-C-A) and d(T-A-T-C-C-G-C-T-C-A-C). Segment 5 was impure. No technique has yet been found satisfactory for purifying segment 5. The sequencing data indicated that approximately half the sample contained a modified nucleotide at site 27 (see Figure 2). This pentanucleotide had a net negative charge that was greater than the correct sequence (dTGTGA). Possibly adenine had been deaminated during the chemical synthesis to form hypoxanthine. For the

TABLE IV: Summary of Reaction Conditions, Recoveries, and Yields for Segment 6.

Starting Chain (3′-OH component) (A)	Amount (mmol)	Oligonucleotide Block (5′-phosphate end) (B)	Amount (mmol) (B/A)	MS[a] or TPS (equiv/P)	Reaction Time (h)	Starting Chain Recovered (%)	Block[b] Recovered (%)	Product	Yield (%)
d(MeOTr)T	4.0	dpbzA-T(Ac)	2.5 (0.6)	1.0	5		21	d(MeOTr)T-bzA-T	62
d(MeOTr)T-bzA-T	1.6	dpanC-anC(Ac)	2.0 (1.25)	1.0	5		24	d(MeOTr)T-bzA-T-anC-anC	40
d(MeOTr)T-bzA-T-anC-anC	0.55	dpibG-anC(ib)	0.55 (1.0)	1.0	5	13		d(MeOTr)T-bzA-T-anC-anC-ibG-anC	25
dCEpT-anC	0.5	dpbzA-anC(Ac)	0.25 (0.5)	1.0	5			dpT-anC-bzA-anC	35
d(MeOTr)T-bzA-T-anC-anC-ibG-anC	0.09	dpT-anC-bzA-anC(Ac)	0.09 (1.0)	1.0	5	30	30	d(MeOTr)T-bzA-T-anC-anC-ibG-anC-T-anC-bzA-anC	12

[a] TPS unless indicated with †, equiv per phosphate group dissociation in the two components. [b] Percent of the original amount for further condensation without additional purification.

TABLE V: Summary of Yields for Segments 1–7.

Deoxyoligonucleotide	Segment No.	Amount (μmol)
d[(MeOTr)T-ibG-T-ibG-ibG]	1	220
d[pbzA-bzA-T-T-ibG-T-ibG-bzA-ibG]	2	4.8
d[(MeOTr)anC-ibG-ibG-bzA-T-bzA-bzA-anC-bzA-bzA-T-T]	3	13.2
d[pT-anC-bzA-anC-bzA]	4	30.9
d[(MeOTr)T-ibG-T-ibG-bzA-bzA-bzA-T-T-ibG-T]	5	16
d[(MeOTr)T-bzA-T-anC-anC-ibG-anC-T-anC-bzA-anC]	6	10.8
d[pbzA-bzA-T-T-anC-anC-bzA-anC-bzA]	7	15.4

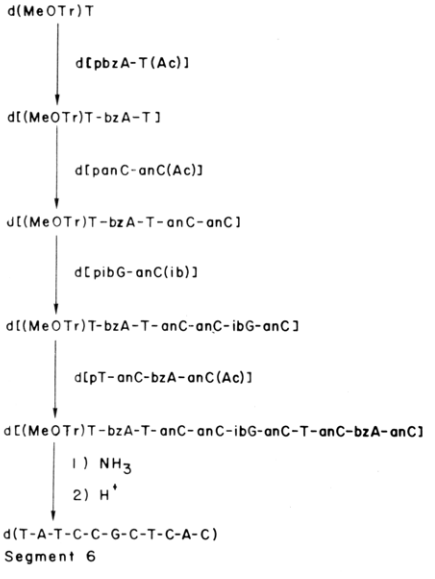

FIGURE 6: Steps in the synthesis of segment 6.

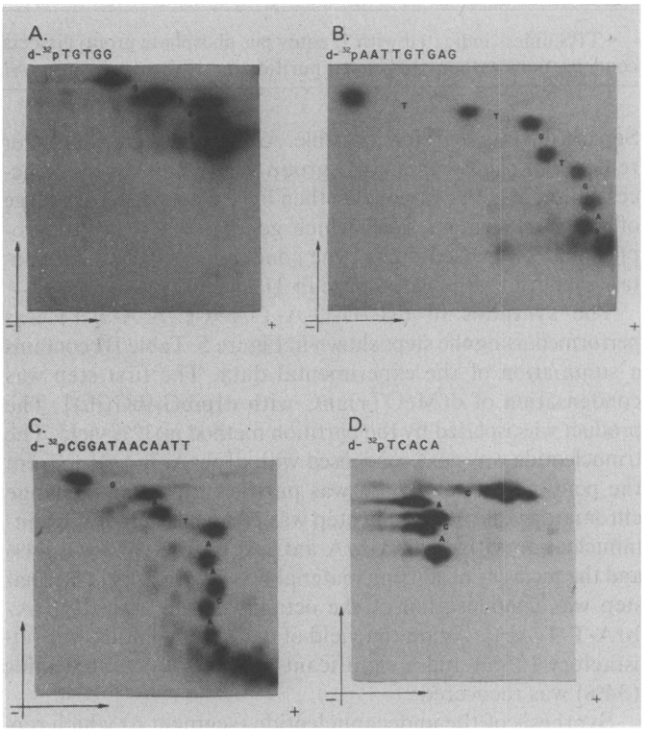

FIGURE 7: Two-dimensional analysis of segments dTGTGG(A), AATTGTGAG(B), dCGGATAACAATT(C), and dTCACA(D). Electrophoresis was along the longitudinal axis and homochromatography the verticle axis. Nucleotide losses are recorded between the appropriate spots.

other segments, the mobility shifts are consistent with those calculated using an empirical formula previously developed (Jay et al., 1974). Some of the mobility shifts for di- and trinucleotides were difficult to interpret using this formula. In segments 2 and 7, for example, the degradation of dpA-A to dpA is not consistent with the calculated shift. The calculated result would indicate loss of dpG or dpA. This problem was solved by extracting the dinucleotide spot from the TLC plate followed by cochromatography with a mixture of the four possible dinucleotides (dpA-A, dpA-C, dpA-G, and dpA-T). All the radioactivity comigrated in two solvent systems (solvents C and D, Weber and Khorana, 1972) with the spot corresponding to d(pA-A) and d(pA-C). Because of the mobility shift, the dinucleotide could not be d(pA-C) and therefore had to be d(pA-A) which is consistent with the expected synthesis. The two-dimensional pattern shown for segment 3 begins with the dinucleotide d(pC-G). The loss of dpG to give dpC was demonstrated by running the electrophoresis in the first dimension for a much shorter time before the homochromatography step (data not shown).

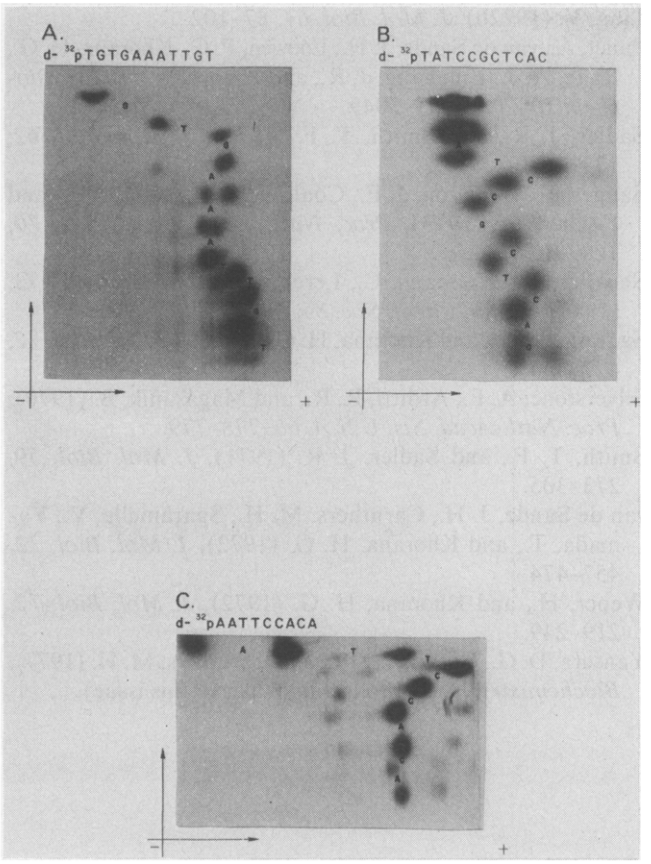

FIGURE 8: Two-dimensional analysis of segments dTGTGAAATTGT(A), dTATCCGCTCAC(B), and dAATTCCACA(C).

Discussion

The synthesis of *lac* operator DNA was designed so that studies could be carried out on how *lac* repressor protein interacts with this DNA. Therefore, intermediates of various syntheses will undoubtedly prove to be as valuable as the final products. These intermediates will be used in future work to alter the operator sequence as we learn more about this system. Much of the work reported here utilized TPS rather than MS as the condensing agent. TPS appeared to cause less deoxyoligonucleotide degradation during the syntheses. As a result less complex fractionation patterns were obtained, indicating fewer side products. For most synthetic steps, a very limited excess of the oligonucleotide block was used over the 3′-hydroxyl component. For example, the synthesis of segment 3 (Table III) utilized about a twofold excess of d[pbzA-bzA-anC(Ac)] and an equivalent amount of d[pbzA-bzA-T-T(Ac)]. Yet, the yields of octa- and dodecanucleotides were 44 and 25%, respectively. In addition, recoveries of the oligonucleotide block and the starting 3′-hydroxyl component were quite high (Table III). These results compared very favorably with previous syntheses where much larger excesses were used (Khorana et al., 1972, 1976). For all the work reported here, oligonucleotides which contained the 5′-phosphate component and guanine nucleotide were reisobutyrylated immediately before use. Considerable evidence has indicated that partial but cumulative deisobutyrylation of the guanine nucleotide occurs during synthetic work (Jay et al., 1976). Therefore by this step, we could at least ensure that the 5′-phosphate containing oligonucleotide was completely isobutyrylated. Because of the acid lability of the methoxytrityl ether, the 3′-hydroxyl component cannot be treated with isobutyric anhydride. For many steps, the isolated yields of protected oligonucleotides are not high. The yields reported here represent only the column fractions of highest purity which were used for the next step in the sythesis. Impure fractions containing mostly product are not included in the reported yields.

Deoxyoligonucleotides isolated in completely deprotected form appear essentially homogeneous when analyzed by the two-dimensional sequencing technique. Segment 5 was an exception since it appeared to be about 50% pure. However, most of the work presently being pursued does not utilize this segment. The base paired duplex 1–26 has been prepared with d(A-A-T-T-G-T) (Yansura et al., 1977). Furthermore by enzymatic joining with T4 ligase and repair synthesis with DNA polymerase I, segment 5 can be utilized to prepare the correct 31-base-paired duplex. DNA polymerase I does not appear to use the modified segment 5 as a template for repair synthesis (unpublished experiments of D. Yansura and D. Goeddel). For future work the effect of trace impurities, present in all chemical syntheses, on interpretation of results with *lac* repressor must be critically evaluated. Previous results indicated that these impurities may be removed during the ligase catalyzed joining steps (Caruthers et al., 1972). This possibility is being actively investigated in our laboratory.

Acknowledgments

The technical assistance of Mrs. Patricia Sullivan is gratefully acknowledged.

References

Beckwith, J., Grodzicker, T., and Arditti, R. R. (1972), *J. Mol. Biol. 69,* 155–159.

Beckwith, J. R., and Zipser, D. (1970), The Lactose Operon, Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Büchi, H., and Khorana, H. G. (1972), *J. Mol. Biol. 72,* 251–288.

Caruthers, M. H., Goeddel, D. V., and Yansura, D. G. (1974), Abstracts of the Southwest Regional Meeting of the American Chemical Society, No. 120.

Caruthers, M. H., Goeddel, D. V., and Yansura, D. G. (1975), *Fed. Proc. Abstr. 34,* No. 1891.

Caruthers, M. H., and Khorana, H. G. (1972), *J. Mol. Biol. 72,* 407–427.

Caruthers, M. H., Kleppe, K., van de Sande, J. H., Sgaramella, V., Agarwal, K. L., Büchi, H., Gupta, N. K., Kumar, A., Ohtsuka, E., RajBhandary, U. L., Terao, T., Weber, H., Yamada, T., and Khorana, H. G. (1972), *J. Mol. Biol. 72,* 475–492.

de Crombrugghe, B., Chen, B., Anderson, W., Nissley, P., Gottesman, M., Pastan, I., and Perlman, R. (1971), *Nature (London), New Biol. 231,* 139–142.

Dickson, R., Abelson, J., Barnes, W., and Reznikoff, W. (1975), *Science 187,* 27–35.

Eron, L., and Block, R. (1971), *Proc. Natl. Acad. Sci. U.S.A. 68,* 1828–1832.

Galibert, F., Ziff, E., and Sedat, J. (1974), *J. Mol. Biol. 87,* 377–407.

Gilbert, W., Gralla, J., Majors, J., and Maxam, A. (1974), in Proceedings of the Symposium on Protein–Ligand Interaction, de Gruyter, Berlin, pp 193–210.

Gilbert, W., Maizels, N., and Maxam, A. (1973), *Cold Spring Harbor Symposium Quant. Biol. 38,* 845–855.

Gilbert, W., and Maxam, A. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 3581–3584.

Gilbert, W., and Müller-Hill, B. (1967), *Proc. Natl. Acad. Sci. U.S.A. 58,* 2415-2491.
Itakura, K., Katagiri, N., Narang, S. A., Bahl, C. P., Marians, K. J., and Wu, R. (1975), *J. Biol. Chem. 250,* 4592-4600.
Jay, E., Bambara, R., Padmanabhan, R., and Wu, R. (1974), *Nucleic Acids Res. 1,* 331-353, and preceding papers in this series.
Jay, E., Cashion, P., Fridkin, M., Ramamoorthy, B., Agarwal, K., Caruthers, M. H., and Khorana, H. G. (1976), *J. Biol. Chem. 251,* 609-623.
Jobe, A., and Bourgeois, S. (1972), *J. Mol. Biol. 69,* 397-408.
Jobe, A., Sadler, J. R., and Bourgeois, S. (1974), *J. Mol. Biol. 85,* 231-248.
Khorana, H. G. (1968), *Pure Appl. Chem. 17,* 349-381.
Khorana, H. G., Turner, A. F., and Vizsolyi, J. P. (1961), *J. Am. Chem. Soc. 83,* 686-698.
Khorana, H. G., et al. (1972), *J. Mol. Biol. 72,* 209-217.
Khorana, H. G., et al. (1976), *J. Biol. Chem. 251,* 565-570.
Kössel, H. R., Büchi, H., and Khorana, H. G. (1967), *J. Am. Chem. Soc. 89,* 2185-2194.
Ling, V. (1972a), *Proc. Natl. Acad. Sci. U.S.A. 69,* 742-746.
Ling, V. (1972b), *J. Mol. Biol. 64,* 87-102.
Panet, A., van de Sande, J. H., Loewen, P. C., Khorana, H. G., Raae, A. J., Lillehaug, J. R., and Kleppe, K. (1973), *Biochemistry 12,* 5045-5049.
Sadler, J. R., and Smith, T. F. (1971), *J. Mol. Biol. 62,* 139-169.
Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H., and Fischer, O. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 1209-1213.
Schaller, H., Weimann, C., Lerch, B., and Khorana, H. G. (1963), *J. Am. Chem. Soc. 85,* 3821-3827.
Sgaramella, V., and Khorana, H. G. (1972), *J. Mol. Biol. 72,* 427-444.
Silverstone, A. E., Arditti, R. R., and Magasanik, B. (1970), *Proc. Natl. Acad. Sci. U.S.A. 66,* 773-779.
Smith, T. F., and Sadler, J. R. (1971), *J. Mol. Biol. 59,* 273-305.
van de Sande, J. H., Caruthers, M. H., Sgaramella, V., Yamada, T., and Khorana, H. G. (1972), *J. Mol. Biol. 72,* 457-474.
Weber, H., and Khorana, H. G. (1972), *J. Mol. Biol. 72,* 219-249.
Yansura, D. G., Goeddel, D. V., and Caruthers, M. H. (1977), *Biochemistry 16* (the following paper in this issue).

Studies on Gene Control Regions. 2. Enzymatic Joining of Chemically Synthesized Lactose Operator Deoxyribonucleic Acid Segments†

D. G. Yansura, D. V. Goeddel, and M. H. Caruthers*

ABSTRACT: The T4 polynucleotide ligase catalyzed joining of six chemically synthesized deoxypolynucleotides corresponding to lactose operator DNA has been investigated. Joining was studied using various combinations of segments. Joining reactions involving multiple sites and the formation of duplex operator DNA were complete in a few hours. Joining reactions involving a single site and the formation of only one strand of operator DNA required several days and repeated annealing in order to go to completion. These studies have permitted the synthesis on a preparative scale (several nanomoles) of operator duplexes and operator single strands.

For investigations on how *lac* operator interacts with *lac* repressor, the DNA that binds repressor tightly has been chemically synthesized (Goeddel et al., 1977). The plan involved the chemical synthesis of seven deoxyoligonucleotides ranging in size from a pentanucleotide to a dodecanucleotide. The present paper describes investigations on the enzymatic joining of these segments or partial segments in various combinations to form polydeoxyoligonucleotides corresponding to *lac* operator DNA.

The plan for synthesis of *lac* operator DNA[1] is shown in Figure 1. The rationale for this particular plan is outlined in the previous paper (Goeddel et al., 1977). The purpose of the work presented in this paper was to investigate possible pathways for preparing operator DNA. Previous work had indicated that in general each ligase joining reaction must be examined individually (Khorana et al., 1972, 1976; Harvey et al., 1975) in order to maximize yield. For reasons which are not yet well defined, some ligations go better than others. Furthermore one set of reaction conditions is usually not satisfactory for all ligase joining reactions. Therefore of major

† From the Department of Chemistry, University of Colorado, Boulder, Colorado 80309. *Received November 18, 1976.* This research was supported by grants from the National Institutes of Health (GM 21120 and GM 21644), a Cottrell Research Grant from Research Corporation, a Biomedical Sciences Support Grant from the National Institutes of Health to the University of Colorado, an Institutional Research Grant from the American Cancer Society to the University of Colorado, a National Science Foundation equipment grant (BMS 75-14541), and the University of Colorado. M.H.C. was supported by a Career Development Award from the National Institutes of Health (1 K04 GM 00076).

[1] Abbreviations used: DEAE, diethylaminoethyl; Tris, tris(hydroxymethyl)aminomethane; hyphenated numbers within brackets represent chemically synthesized deoxyoligonucleotides as defined in Figure 1 that have been enzymatically joined. The numbers left to right within brackets refer to the DNA sequence 5′ to 3′. The actual sequence of each deoxyoligonucleotide defined by the numbers is given in Figure 1. The symbol [/] is used to indicate that the segments within brackets form a base-paired duplex.