TOTAL SYNTHESIS OF THE GENE FOR AN ALANINE TRANSFER RIBONUCLEIC ACID FROM YEAST

H. G. KHORANA*

The Institute for Enzyme Research, University of Wisconsin, Madison, Wisconsin

ABSTRACT

The methodology developed for the synthesis of bihelical DNA consists of the following three steps. (1) Chemical synthesis of deoxyribopolynucleotide segments containing 8 to 12 nucleotide units. These should represent the entire two strands of the intended DNA and those belonging to the complementary strands should have an overlap of 4 to 5 nucleotides. (2) The phosphorylation of the 5'-hydroxyl group with ATP carrying a suitable label in the γ-phosphoryl group using the T4 polynucleotide kinase. (3) The head-to-tail joining of the appropriate segments when they are aligned to form bihelical complexes using the T4 polynucleotide ligase.

Using this methodology total synthesis of yeast alanine transfer RNA structural gene has been achieved.

Methods have been developed in recent years for the chemical synthesis of deoxyribopolynucleotides of defined nucleotide sequence[1,2]. However, in practical terms, there is a severe limit on the length of the polynucleotide chains which can be assembled unambiguously by purely chemical methods. On the other hand, for biological studies of the nucleic acids, it is often the high molecular weight nucleic acids which are the most useful. Therefore, it is desirable or even necessary to couple methods of organic chemistry, which alone can afford oligonucleotides of predetermined sequence, with other concepts or methods in order to prepare nucleic acids of defined nucleotide sequences. In earlier work reported from this laboratory, it was possible to prepare double-stranded DNA-like polymers of known repeating sequences by using short synthetic deoxypolynucleotides as templates for the DNA polymerases. The availability of the resulting polymers permitted extensive studies of the cell-free protein synthesis and of the genetic code[1,2]. Clearly, a major aim of future synthetic work in the field of nucleic acids must be the development of methods for the synthesis of bihelical DNAs with specific nucleotide sequences. Towards this goal, the central concept which we have wanted to exploit is the inherent ability of polynucleotide chains to form ordered bihelical complexes by virtue of base-pairing. Thus, it was hoped to join relatively short chemically synthesized deoxyribopolynucleotides while

* Present address: Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

these were held together in properly aligned bihelical complexes. As a specific first objective, the total synthesis of the gene, i.e. the double-stranded DNA, specifying a transfer RNA was undertaken. It is the purpose of the present article to report briefly on the total synthesis of the gene corresponding to the major yeast alanine transfer RNA (*Figure 1*)[3]. This tRNA

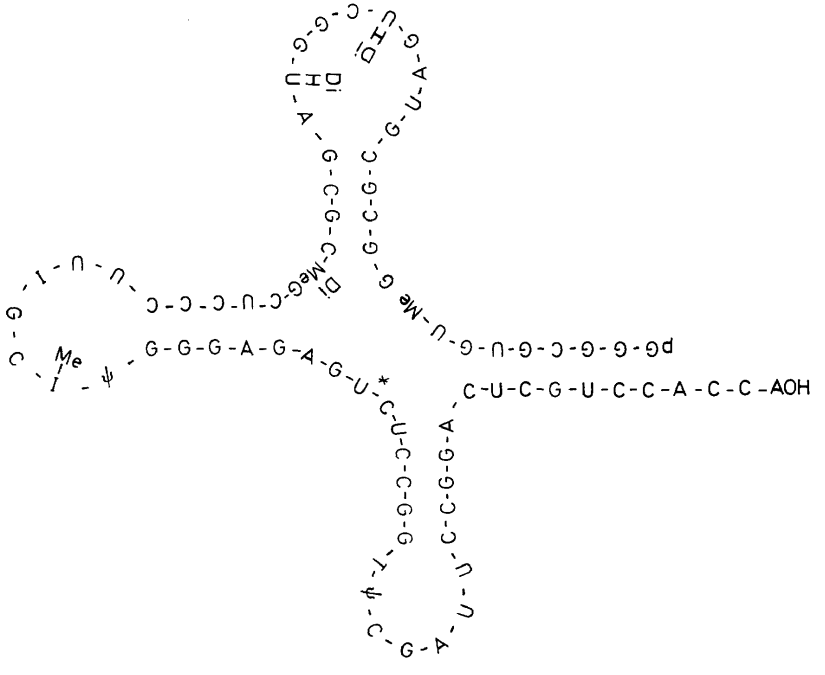

Figure 1. Structure of the major yeast alanine transfer RNA as deduced by Holley and his co-workers[3]. The structure is shown in the familiar cloverleaf model for the secondary structure of the tRNA. The abbreviations U, C, A and G stand, respectively, for the nucleotides of uracil, cytosine, adenine and guanine. ψ stands for pseudouridine, T for ribothymidine, I for inosine, Me-I for methylinosine and DiH-U for 5,6-dihydrouridine.

was the only one whose sequence was known at the time (1965) that the present work was undertaken.

The decision to synthesize the gene for a transfer RNA followed from a variety of considerations. The first clear argument in its favour was that the sequence of the deoxynucleotides in the gene can be derived directly from the sequence of the transfer RNA. Additional considerations were that the general functions of the tRNAs are clearly established. These molecules have to be recognized by a rather large number of components of the protein-synthesizing machinery such as by the aminoacyl-tRNA synthetases, by the pyrophosphorylase which repairs the CCA end, by the ribosomes and by several protein factors involved in protein chain initiation, elongation and

termination, and finally by messenger RNA. Furthermore, the tRNAs are indeed a unique class of molecules possessing attributes of both nucleic acids and of proteins. The total area of the structure/function relationships in these molecules is an open field, despite intensive current research effort. It is clear that chemical synthesis could, in principle, offer a definitive approach of wide scope. Different parts of the tRNA structure could be systematically modified at the gene level. For example, one could have deletions in different parts, one could take an anti-codon loop from a tRNA specific for one amino acid and replace it with the anti-codon loop from another tRNA.

EARLY WORK AND TOTAL SYNTHETIC PLAN

In chemical work carried out between 1965 and 1967, chemical syntheses of the two icosa-deoxyribonucleotides shown in *Figure 2* were accomplished[1]. The two icosanucleotides together span nucleotides 21–50 of the tRNA gene, belong to the complementary strands and overlap through halves of their lengths in the required antiparallel manner. It should be pointed out that in deriving the DNA sequence from the primary sequence of the tRNA, the assumption has been made that the rare bases present in the tRNA arise by subsequent modification of the 'nascent' tRNA which contains only the four standard bases (A, C, U and G) and is the product of transcription of the gene. For example, inosine is formed by deamination of adenosine and so comes from an A–T base-pair in DNA. While there was only limited experimental support for this assumption at the time the present work was started, there is now extensive and strong evidence for the validity of this important assumption.

Following the discovery of the polynucleotide ligases in 1967, studies were carried out with these enzymes to determine the minimum length of the deoxyribo-oligonucleotide chains which these enzymes require to bring about the joining reaction. The chain lengths necessary turned out to be small[1,4,5]. Thus, in the system consisting of the two icosanucleotides and the appropriate oligonucleotides (*Figure 3*), oligonucleotides as short as the tetranucleotide (T–C–T–C) could indeed be joined to the appropriate icosanucleotide under suitable conditions of temperature, magnesium ions and enzyme concentration. These results appeared to promise a remarkably simple strategy for the synthesis of bihelical DNA which would comprise the following three steps: (1) Chemical synthesis of deoxypolynucleotide segments of chain length in the range of 8 to 12 units only with free 3′- and 5′-hydroxyl end groups; the segments would represent the entire two strands of the intended DNA and those belonging to the complementary strands would have an overlap of four to five nucleotides; (2) the phosphorylation of the 5′-hydroxyl group with ATP carrying a suitable label in the γ-phosphoryl group using the T4-polynucleotide kinase and (3) the head-to-tail joining of the appropriate segments when they are aligned to form bihelical complexes using the T4-polynucleotide ligase.

Being already in possession of the two icosanucleotides and the corresponding shorter segments, which had served as intermediates in chemical

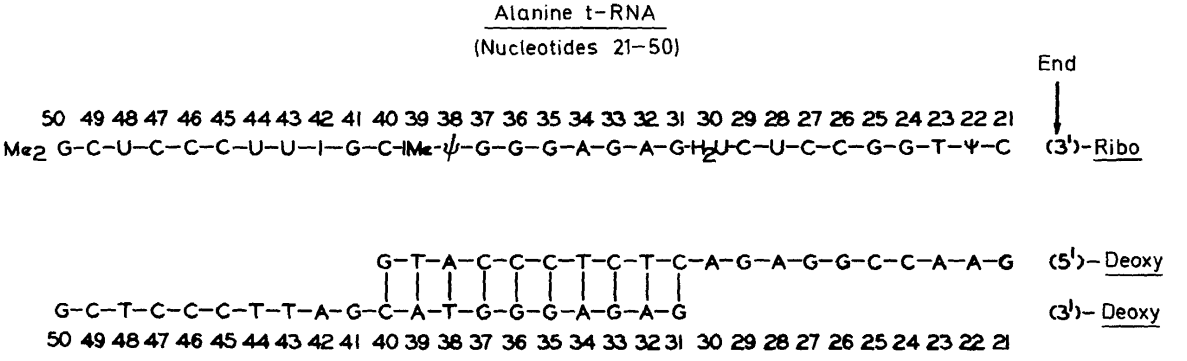

Figure 2. Deoxyribo-icosanucleotide segments corresponding to nucleotide sequences 21–50 of the gene for yeast ala-tRNA. The two icosanucleotides are complementary to each other through halves of their length in the required antiparallel manner.

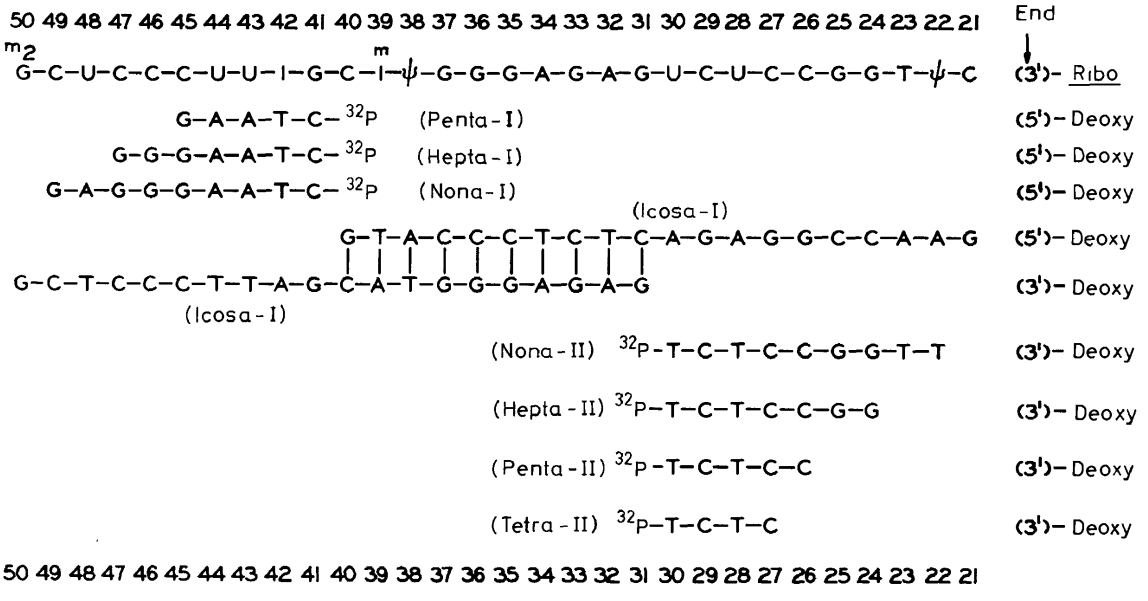

Figure 3. Chemically synthesized deoxyribopolynucleotides corresponding to sequences 21–50 of yeast ala-tRNA. The icosanucleotide-I (Icosa-I) represents a sequence complementary to nucleotides 21–40 of the tRNA and has polarity opposite to that of the tRNA; the nona-, hepta- and penta-nucleotides (Nona-I, Hepta-I, etc.) similarly contain sequences complementary to nucleotides 41–49 or less, and again have polarity opposite to that of the tRNA. The deoxyribopolynucleotides (Icosa-II, Nona-II, Hepta-II, Penta-II, and Tetra-II) are segments, complements of the complement, and therefore contain the same sequences and polarity as the tRNA itself. ^{32}P represents the 5'-phosphate end group wherever shown.

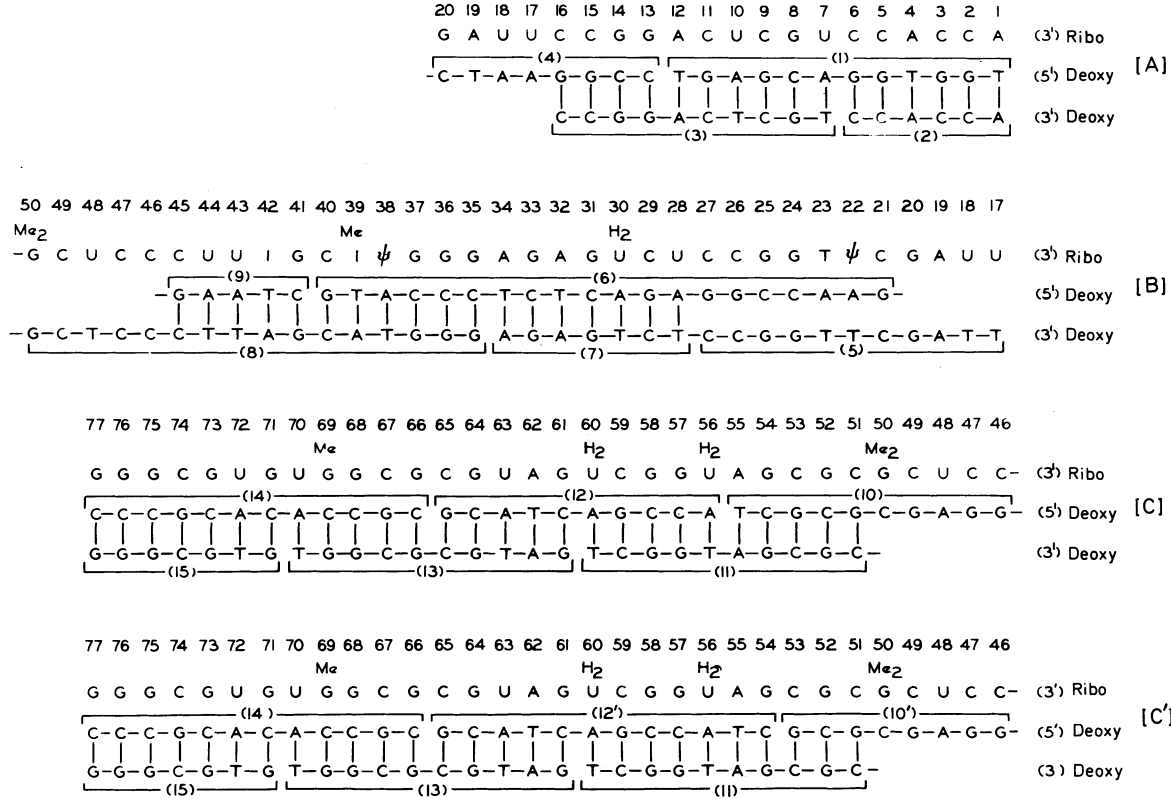

Figure 4. Total plan for the synthesis of a yeast alanine tRNA gene. The chemically synthesized segments are enclosed in brackets, the serial number of the segment being shown within the brackets. A total of 17 segments (including 10′ and 12′ varying in chain length from penta- to icosa-nucleotides were synthesized.

synthesis, the plan shown in *Figure 4* was formulated for the total synthesis of the transfer RNA gene. Thus the gene was divided into three parts shown as A, B and C (or C') and each part was to consist of several chemically synthesized segments. The segments are indicated in *Figure 4* by brackets, the serial number being inserted into the bracket. Additional considerations which led to the derivation of this plan were as follows. Starting with the B part (bottom strand) and proceeding to the right, segment 8 consisted of the hexadecanucleotide rather than the icosanucleotide shown in *Figures 2* and *3*. The former was preferred because (1) the heptanucleotide (segment 7) could be prepared readily and (2) the starting tetranucleotide block (C–C–G–G) in the chemical synthesis of segment 5 could be used, in addition, for the synthesis of both segments 3 and 4 (part A of the gene). This left only the dodecanucleotide (segment 1) and the hexanucleotide, C–C–A–C–C–A (segment 2) to complete the part A of the gene. (However, as seen below, this plan for the A part of the gene created an interesting and instructive difficulty in the enzymatic work.)

Segment 6 (part B, top strand) consisted of the icosanucleotide already shown in *Figures 2* and *3*. Proceeding to the left of segment 6, it is seen that this region predominates in purine nucleosides which cause greater difficulty in chemical synthesis. Therefore, it was decided to use a pentanucleotide as segment 9 and then to go on to divide the remainder of the gene (part C) arbitrarily into decanucleotides (segments 10, 11, 12 and 13) and, finally, provide the terminal sequences in the form of a dodecanucleotide (segment 14) and a heptanucleotide (segment 15). (The latter again predominates in the guanine nucleotide.) Because of the high G–C content in part C, problems due to intramolecular secondary structure and possible self-aggregation were anticipated and, in particular, the self-complementarity of segment 10 was obvious. Thus, as seen in *Figure 5* (a), this segment can form a tight

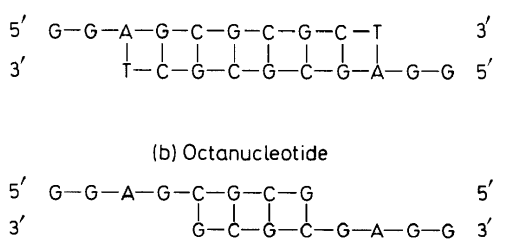

Figure 5. Self-complementarity of segments 10 and 10′. For numbering of the segments see *Figure 4*.

duplexed structure by itself. (Absorbance/temperature profiles recently have borne this out, the T_m in 5 mM Mg^{2+} ion + 0.01 M tris-HCl buffer being 62°.) To guard against the possibility of failure in the joining of segment 10 to B or C or both simultaneously, a modified scheme (C′ of *Figure 4*) for this part of the gene was also adopted. In this, the length of the segment 10′ is reduced to eight nucleotides [probable secondary structure as shown in *Figure 5* (b)]

P.A.C.—25/1—E

while the adjoining segment 12′ is increased to a dodecanucleotide. In the work described below, no difficulty has in fact been encountered in realizing synthesis according to the plan C. While the total synthesis using the scheme C′ has so far not been completed, preliminary experiments indicate that this altered scheme would also succeed.

ASPECTS OF EXPERIMENTAL METHODOLOGY

Methods for the chemical synthesis of the deoxyribopolynucleotide segments have been reviewed previously[1]. In particular, the synthesis of the icosanucleotide, G–A–A–C–C–G–G–A–G–A–C–T–C–T–C–C–C–A–T–G, was discussed in some detail[1]. Time will not permit any discussion of these methods here. However, it should be emphasized that the chemical synthesis of all these segments easily represents the greater part of the total effort in the present project. In passing, it may be noted that the synthesis of every segment (shown within brackets in *Figure 4*) required careful planning with respect to the choice of the protected blocks and it should be emphasized that all the segments were rigorously purified, firstly at the fully protected stage, and secondly, after removal of the protecting groups. The final step in purification was invariably chromatography on DEAE-cellulose columns in the presence of 7 M urea. Desalting was effected on Bio-Gel columns. The pure segments have been stored in aqueous buffered solutions in the frozen state at − 20° to − 120° C, there being no evidence of any detectable breakdown.

The chemically synthesized segments were phosphorylated using the T4-polynucleotide kinase as described previously[4]. Several segments appeared to possess secondary structures which inhibited the 5′-phosphorylation. Therefore, the reaction mixtures were routinely heated prior to phosphorylation. The kinase reactions were judged to be quantitative when a further cycle of heating, cooling to 37° and further incubation with fresh kinase gave no increase. The phosphorylated segments were separated from excess of the γ-labelled ATP by passing the reaction mixtures through Sephadex columns. The specific activity of the γ-labelled ATP used for phosphorylation of the oligonucleotides was used as a measure of the molar proportions of the phosphorylated segments used in the joining reactions.

The reactions for enzymatic joining were set up under different conditions of Mg^{2+} ion concentration and temperature on 10–50 μl scale. The conditions used for preannealing the components varied, some typical examples being given in the legends to figures. The kinetics of joining were followed by measuring resistance to bacterial alkaline phosphatase. The reactions were analysed by chromatography on DEAE-cellulose paper strips[4, 5]. An example is shown in *Figure 6*. In the control (bottom part), from which the ligase was omitted, all the label was sensitive to the phosphatase and moved as inorganic phosphate. After the ligase reaction, most of the label stayed at the origin as part of a polynucleotide.

The joined products were separated from the starting materials on Agarose or Sephadex columns and in all cases subjected to careful analysis, usually consisting of the following steps: (1) resistance to phosphatase; (2) degradation to 3′-nucleotides using the micrococcal and the spleen phosphodiesterase (*Figure 7*) and (3) hydrolysis to 5′-nucleotides using pancreatic DNAase

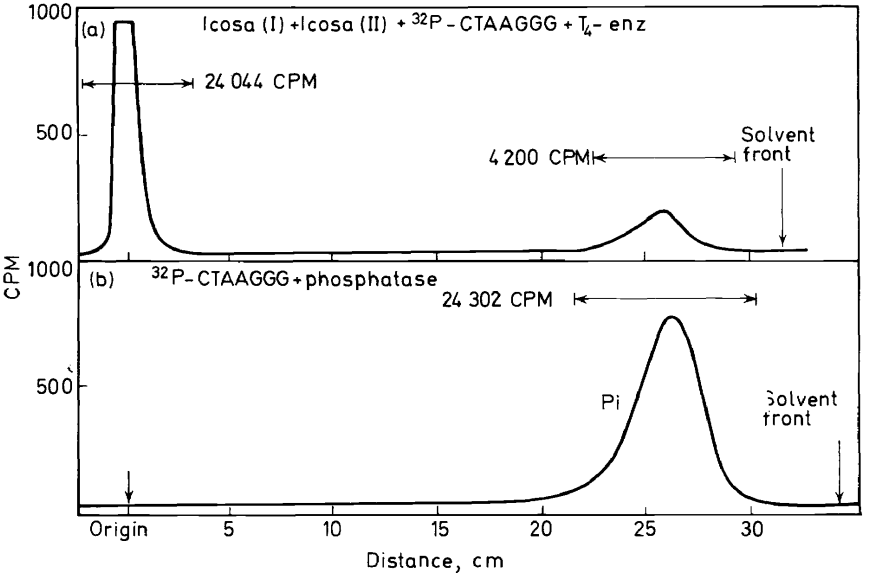

Figure 6. DEAE-cellulose paper-strip assay of the reaction catalysed by the DNA-joining enzymes. The polynucleotides and ^{32}P-labelled oligonucleotides were as shown in *Figure 3*. The top (a) shows the joining of Icosa-I to ^{32}P-CTAAGGG in the presence of Icosa-II and the T4-induced ligase; (b) shows the control in which only the ^{32}P-CTAAGGG is exposed to the phosphatase without incubation with the ligase. For further details see refs 4 and 5.

and venom phosphodiesterase. In a few cases repair with DNA polymerase from *E. coli* was also used to characterize the product further[6]. The mononucleotides were separated by solvent systems described previously[6] and also by electrophoresis at pH 3.5.

Characterization of the joined products was aided by manipulation of the specific activity of γ-^{32}P-ATP used in phosphorylation reactions. Thus, radioactivity at the internal linkage points in A, B and C was much weaker than that used at the termini which were to connect A to B and B to C. Further, in multiple joinings (e.g. in C), stoichiometry of the joining of different segments was aided by the concomitant use of ^{32}P and ^{33}P isotopes. Stoichiometric joining of A to B and B to C was similarly confirmed by using *hot*^{32}P at the 5′-termini in B while the appropriate 5′-termini in A and C contained *hot* ^{33}P.

THE SYNTHESIS OF B

Of the three parts (*Figure 4*), the synthesis of this part was relatively straightforward. Segments 5, 7 and 9 were phosphorylated at the 5′-termini using γ-^{32}P-ATP of the same specific activity. The synthesis was carried out in two ways. In the first, segments 6 and 8 and the 5′-^{32}P-segments, 7 and 9,

Nearest neighbour analysis of DNA ligase product

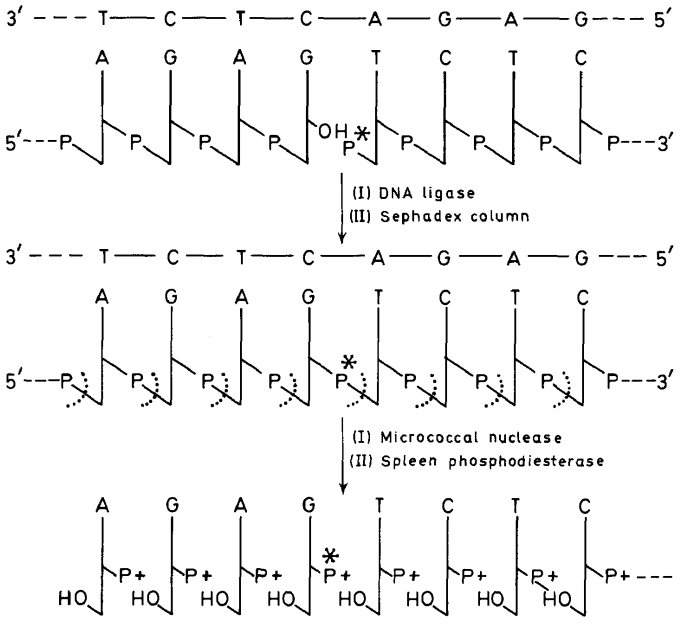

Figure 7. Illustration of the method for 'nearest-neighbour analysis' by degradation of the ^{32}P-labelled joined products to 3'-nucleotides. In the joining reaction, one starts with the defined oligonucleotides, one of which carries the 5'-^{32}P label and the second of which contains the 3'-hydroxyl-bearing nucleoside. In the example shown, 5'-^{32}P is present on T-nucleotide and 3'-OH is present in the guanosine nucleoside. After the enzyme-catalysed phosphodiester bond formation, degradation to 3'-nucleotides as shown should give d-Gp as the labelled deoxynucleotide.

were reacted and the product was isolated as shown in *Figure 8*. The characterization of the joined product (Peak 1 of *Figure 8*) by degradation to 5'- and 3'-nucleotides is shown in *Table 1*. Thus as expected, the radioactivity was equally distributed between the 5'-nucleotides, d-pA and d-pC and no radioactivity was present in the other two 5'-mononucleotides. On degradation to 3'-nucleotides, all of the radioactivity was found in d-Gp and this result was again to be expected.

Table 1. Characterization of the product of peak 1, *Figure 8*

Radioactivity in 5'-nucleotides	(CPM)	Radioactivity in 3'-nucleotides	(CPM)
d-pA	5124	d-Ap	6
d-pG	14	d-Gp	9078
d-pT	27	d-Tp	47
d-pC	5107	d-Cp	8

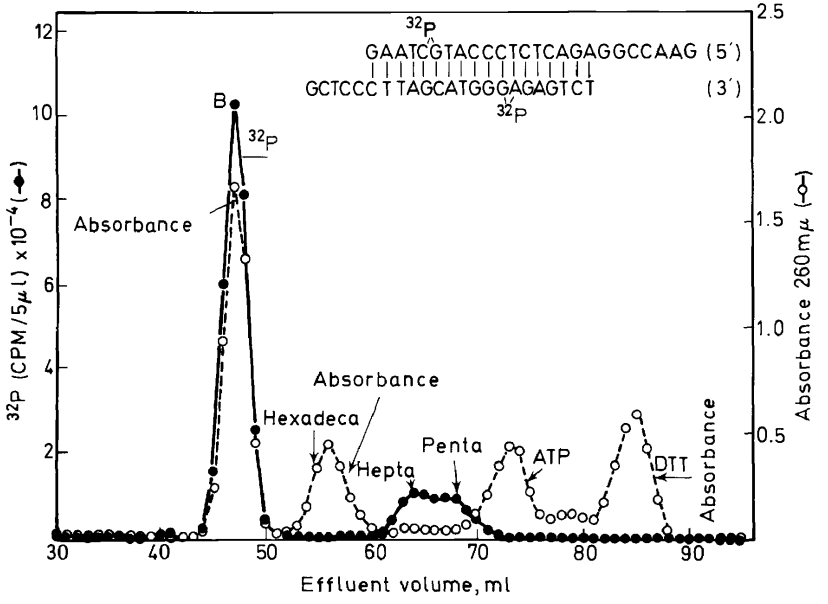

Figure 8. Separation of the ligase-catalysed joined product corresponding to the B (without segment 5) part of the gene. The separation was carried out on an Agarose (0·5 M) column (1 × 104 cm) at 4° using 0·1 M triethylammonium bicarbonate as the eluant.

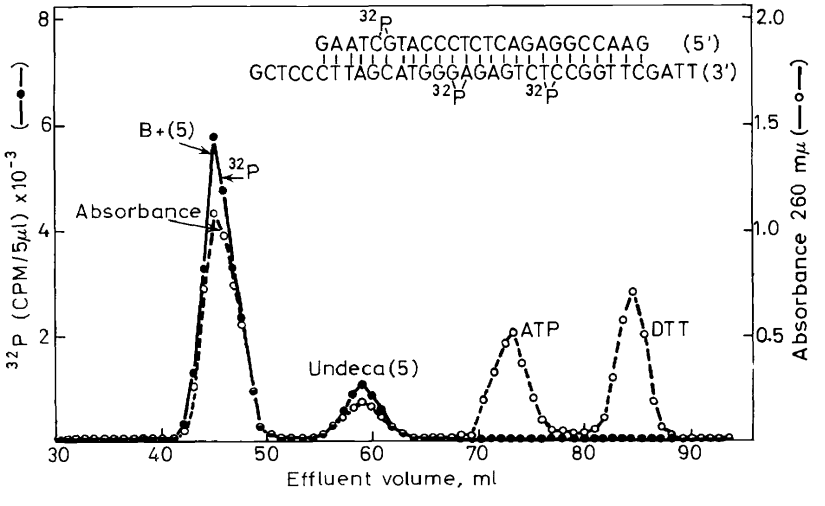

Figure 9. The joining of segment 5 (^{32}P-undecanucleotide) to the product of peak 1, *Figure 8*. The product was purified by using an Agarose (0·5 M) column (1 × 104 cm) at 4° using 0·1 M triethylammonium bicarbonate as the eluant.

In the second step, the ^{32}P-phosphorylated undecanucleotide (segment 5) was joined to the product obtained in *Figure 8* and B was purified as shown in *Figure 9*. Analysis of the pooled peak (excluding the skewed right portion

of the first peak in *Figure 9*) showed the product to contain largely B. However, contamination by B lacking the segment 5 seems very likely as indicated by the rather low yield obtained subsequently in the experiment of *Figure 21*.

The above results indicated that the quantitative joining of segment 5 required more specialized conditions and an alternative preparation of 'B' is shown in *Figures 10* and *11*. Here, after the joining of segments 7 and 9 to, respectively, segments 8 and 6, the undecanucleotide was added and the incubation temperature was increased to 25° for this reaction. Kinetics of the reaction are shown in *Figure 10*. The product was isolated as shown in

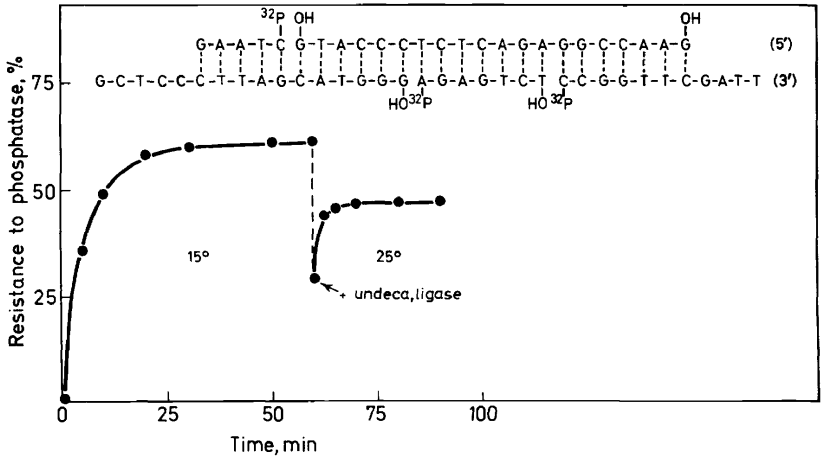

Figure 10. Alternative enzymatic synthesis of B. This experiment is illustrated with experimental details as follows. The reaction mixture (0·84 ml) contained 1.8 mμmoles of the icosanucleotide (segment 6), 2.5 mμmoles of the hexadecanucleotide (segment 8), 2.5 mμmoles of 5′-^{32}P-labelled heptanucleotide (segment 7), and 2.5 mμmoles of ^{32}P-labelled pentanucleotide (segment 9) in 15 mM tris-Cl, pH 7.6, 10 mM DTT, 66 μM ATP, and 2.5 mM Na_2HPO_4. The mixture was heated to 60° for 3 min, then $MgCl_2$ was added to give 10 mM concentration. After 15 min at 15° the ligase (250 units) was added. After 1 h at 15°, when the reaction had reached a plateau, the mixture was heated to 100°C for 2 min and pre-heated ^{32}P-labelled undecanucleotide (segment 5) (3.5 mμmoles) was added. A 20 min pre-incubation at 25° was again allowed and the joining re-started by the addition of more ligase (150 units). Further additions of ligase during either the first or the second step made no difference in the extent of the reaction.

Figure 11. Completely satisfactory analyses were obtained for this product (*Table 2*). Thus, as expected, enzymic degradation to 5′-nucleotides gave radioactive d-pC and d-pA in the ratio of two to one. Degradation to 3′-nucleotides gave d-Gp and d-Tp in the ratio of two to one, no radioactivity being present in any of the other nucleotides.

In various subsequent experiments on the joining of B to A and C, the two G units at the 5′-termini of B were phosphorylated with γ-^{32}P or ^{33}P labelled ATP of much higher specific activity than that used at the internal positions.

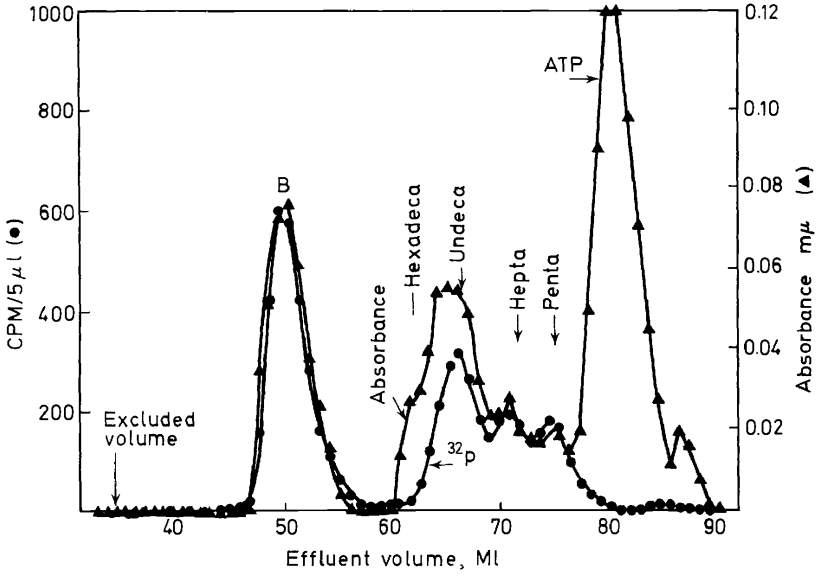

Figure 11. Purification of B by gel filtration. The reaction mixture (legend to *Figure 10*) was concentrated in the presence of 20 mM EDTA to about 0.25 ml and then applied on top of an Agarose-0.5 M column (100 cm × 1 cm diameter) pre-equilibrated at 5° with 0.1 M triethylammonium bicarbonate. Fractions of 0.90 ml were collected every 20 min and analysed for radioactivity using 5 μl aliquots and for absorbance at 260 mμ. The identification of the different peaks is shown in the figure.

Table 2. Characterization of B (Material of first peak, *Figure 11*)
(Numerals in parentheses following counts represent experimental molar ratios)

Phosphatase assay				
	Resistant	Sensitive		
^{32}P-CPM	1113	0		
5'-Nucleotide analysis				
I. B as such				
	d-pA	d-pG	d-pT	d-pC
^{32}P (counts/5 min)	747(1)	0	0	1457(1.9)
II. After phosphorylation with γ-^{32}P-ATP				
	d-pA	d-pG	d-pT	d-pC
^{32}P (CPM)	240(1)	428(1.8)	10	465(1.9)
3'-Nucleotide analysis				
I. B as such				
	d-Ap	d-Gp	d-Tp	d-Cp
^{32}P (CPM)	0	1230(2.1)	595(1)	0
II. B 'repaired' with DNA polymerase (α-^{32}P-dGTP, the other triphosphates being cold)				
	d-Ap	d-Gp	d-Tp	d-Cp
^{32}P (counts/5 min)	6998(1)	12658(1.8)	363	343

An example of phosphorylation of B using γ-^{32}P-ATP of very high specific activity followed by isolation of the product is given in *Figure 12.*

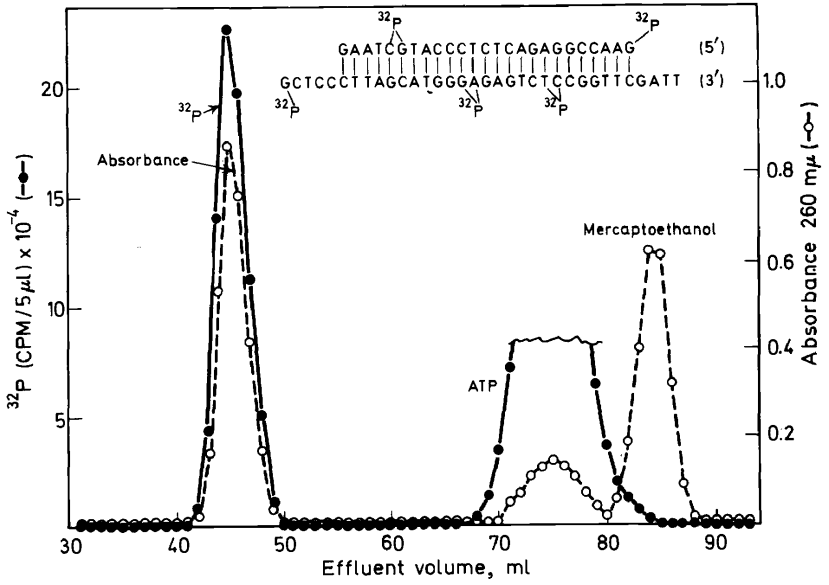

Figure 12. Phosphorylation of B (product of *Figure 9*) with γ-^{32}P-ATP of very high specific activity. After the reaction with polynucleotide kinase, the product was isolated essentially as described in legend to *Figure 9.*

THE SYNTHESIS OF A

In the initial experiments the synthesis of A was attempted by using 5'-^{32}P-labelled segments 2, 3 and 4 together with unphosphorylated segment 1 (top part of *Figure 13*). While the joining of segment 2 to segment 3 went to completion, no joining of segment 4 occurred. Instead, the product formed was the dimer shown in the bottom part of *Figure 13.* This was clearly due to the self-complementary nature of the protruding C–C–G–G single-stranded end in the duplex comprising segments 1, 2 and 3. To circumvent this, the synthesis of A was carried out in two steps. In step 1, segment 1 was joined to ^{32}P-labelled C–C–G–G–A–A–T–C (segment 4) in the presence of segment 3, the latter containing *unphosphorylated* 5'-end. This favoured kinetic rather than the thermodynamic control. The kinetics of the synthesis of the icosanucleotide (segments 1 + 4) are shown in *Figure 14.* It is interesting that in the initial stages, a product was formed (dashed line in *Figure 14*) which was identified as AMP-*O*-^{32}P-CCGGAATC. Ultimately, the formation of the required icosanucleotide was essentially complete as based on the amount of the segment 1. The icosanucleotide was isolated free from the unused segment 4 and the complementary segment 3 by passage through a column of Sephadex G-50 at room temperature with 0·05 M triethyl-

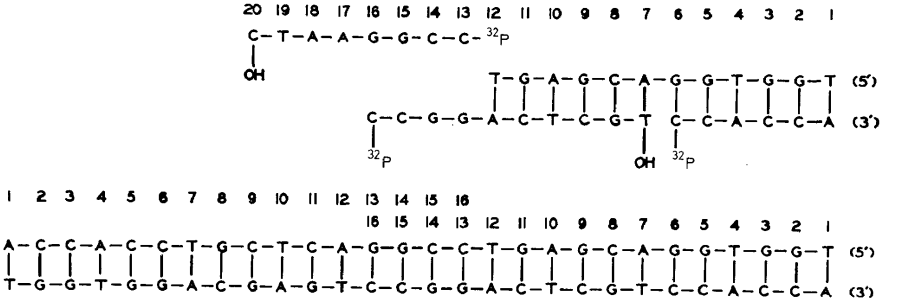

Figure 13. Attempted synthesis of the A part of the gene. The four segments used are shown in the top part of the figure, while the product (a dimer derived from segments 1, 2 and 3) is shown in the bottom part of the figure.

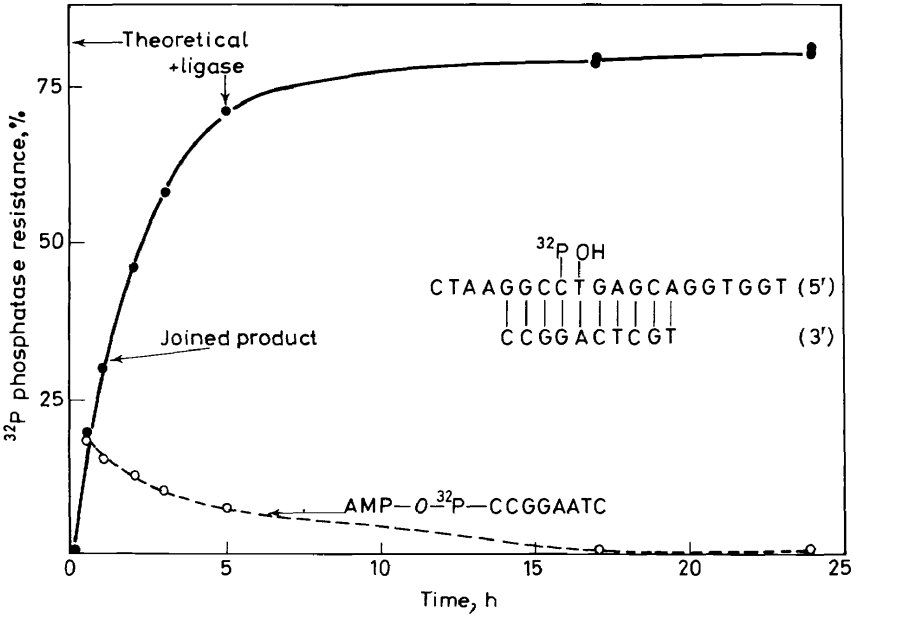

Figure 14. Enzymatic joining of segment 1 to ^{32}P-labelled segment 4 in the presence of segment 3 (see also the insert in the figure). The reaction mixture (2.4 ml) contained 20.4 mμmoles of segment 1, 23.6 mμmoles of segment 3 and 24.2 mμmoles of ^{32}P-labelled segment 4 in 15 mM tris-Cl buffer, pH 7.6, 10 mM $MgCl_2$, 10 mM DTT, 26.6 μM ATP, and 2.5 mM Na_2HPO_4. Before addition of the ligase (2000 units), the mixture was heated at 60° C for 3 min, cooled to 15°, pre-incubated at that temperature for 15 min. The joining was monitored by the development of ^{32}P-resistance to phosphatase. The assay was performed by DEAE-paper chromatography in 0.32 M ammonium formate – 7 M urea. This system allowed the detection of the fast moving phosphatase resistant radioactivity due to AMP-*O*-^{32}P-CCGGAATC. The required icosanucleotide stayed at the origin.

ammonium bicarbonate as the eluant, conditions under which the duplexed structure was unstable (*Figure 15*).

The synthesis of A was completed in a second step by using the above icosanucleotide, together with segments 2 and 3 which were labelled with

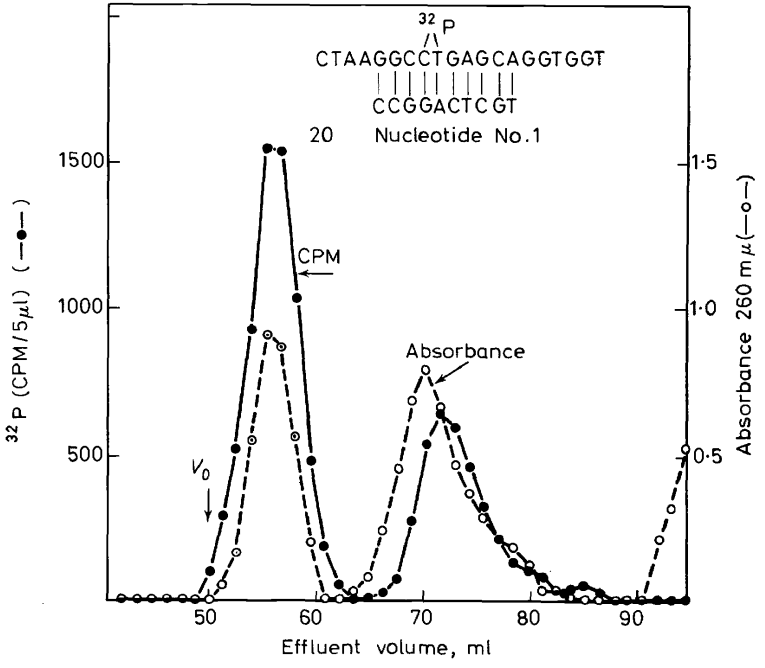

Figure 15. Purification of the icosanucleotide resulting from the joining of segments 1 and 4 in the presence of segment 3. The reaction was carried out as in *Figure 14*. The separation was on a column of Sephadex G-50 at room temperature. The eluant was 0.05 M triethylammonium bicarbonate.

^{32}P or ^{33}P at the 5′-end. In the 'large scale' preparation shown in *Figure 16*, the decanucleotide (segment 3) was phosphorylated with γ-^{33}P-ATP of very high specific activity. The same batch of ATP was also used for phosphorylation of the 5′-hydroxyls of segments 10 and 12 (see below). The preparative isolation of A was carried out as shown in *Figure 16*.

THE SYNTHESIS OF C

The synthesis of this part proved to be a problem of yet a different kind. Numerous attempts to perform joining reactions using three component systems (e.g. phosphorylated segment 11 + segments 12 and 13) in various combinations either failed or showed moderate success. It is our current interpretation that this failure was due to the high G–C content in the components belonging to this part and that under all of the earlier conditions

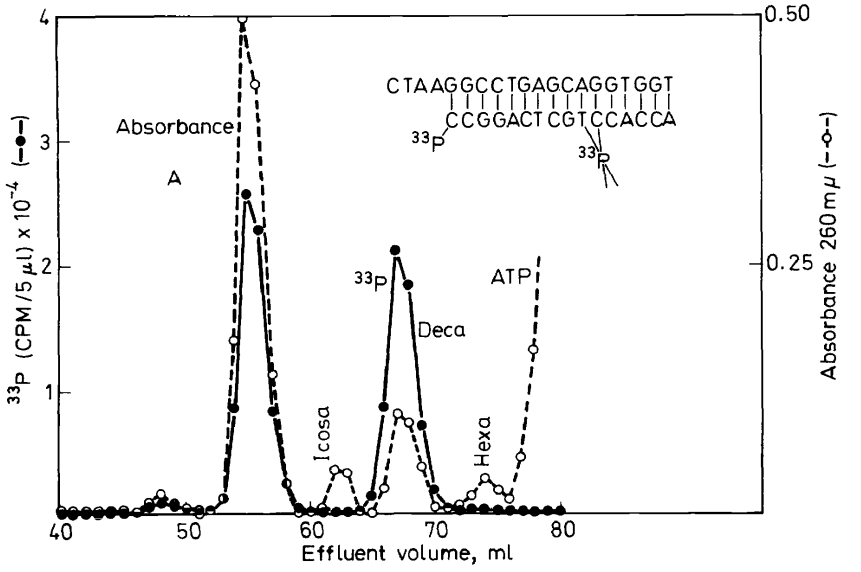

Figure 16. The synthesis and purification of A. The initial reaction mixture (1.2 ml) contained the icosanucleotide (6.5 mμmoles) prepared by separation as in *Figure 15*, weakly ^{33}P-labelled segment 2 (9.5 mμmoles), $MgCl_2$ (18 mM) and tris-Cl buffer (pH 7.7; 27 mM). This was heated to 85° and then slow-cooled (1½ h) to 10° C. To this reaction mixture were now added *hot* ^{33}P-labelled segment 3 (9.5 mμmoles), DTT (10 mM) and ATP (64 μM) and the volume was made up to 2 ml. After pre-incubation of the mixture at 10°C for 1 h ligase (200 units) was added and the joining reaction allowed to proceed for 2½ h at 10° C. The reaction mixture was then heated to 85°, again slow-cooled to 10°C and a further addition of ligase (80 units) was made. After a further incubation for 10 h at 10° C, the reaction was terminated by the addition of EDTA (2.5-fold excess over the Mg^{2+} ion concentration). The mixture was applied to the top of an Agarose-0.5 M column (90 cm × 1.2 cm diameter), and elution carried out with 0.1 M triethylammonium bicarbonate at 4° C. Fractions of about 1 ml were collected every 20 min. The identification of the different peaks is shown in the figure.

tried, the formation of only a few G–C base-pairs stabilizes imperfectly matched duplexed structures and that it takes very prolonged periods (several weeks) for the components to align to form a perfect duplex. The procedure which was effective in the synthesis of C is as follows: 5′-^{32}P-labelled segment 14 was annealed to segment 15 (heating above 70° and slow cooling to 5° C, a temperature well below the T_m of the bihelical complex which in 5 mM Mg^{2+} + 0.01 M tris-Cl buffer is 38° C). Judiciously labelled components 13, 12 and 11 were then added and the total mixture was kept at 10° for 5 hours. Subsequent addition of ligase brought about rapid concerted joining of the five components. The kinetics of the joining are shown in the insert in *Figure 17*, the extent of reaction being determined by the development of phosphatase resistant ^{32}P as well as ^{33}P-radioactivity. In this connection, the pattern of labelling of the different components in C should be noted (insert in *Figure 17*). Two decanucleotides contain T at the 5′-end and the joining of these will occur to the 3′-hydroxyl group of a G nucleoside in both cases. To aid quantitation, one of the decanucleotides was phosphorylated using ^{33}P-ATP and the second using ^{32}P-ATP. The other two com-

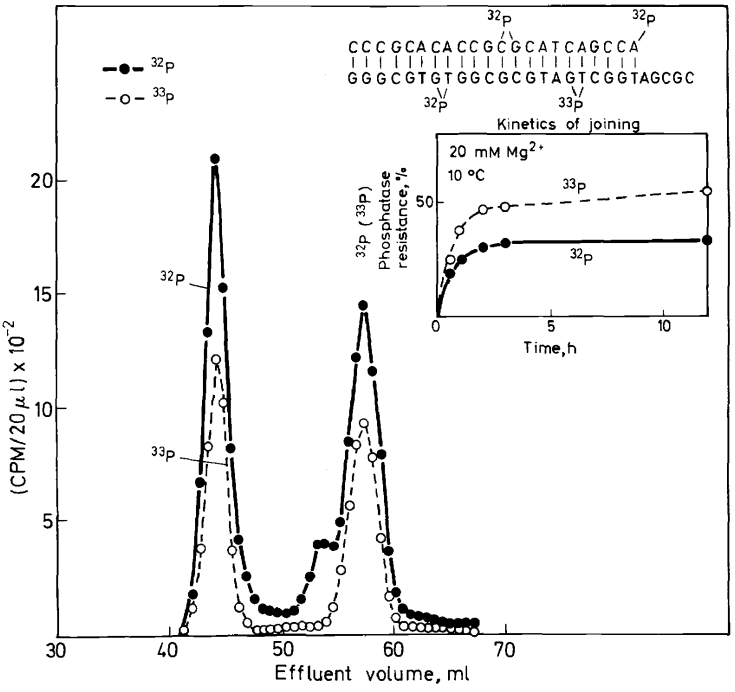

Figure 17. The synthesis and purification of C (small scale). The labelling pattern is shown in the insert in the figure as are the kinetics of joining. The incubation was at 20 mM Mg^{2+} ion concentration at 10° C. The detailed protocol for the experiment was similar to that described in the legend to *Figure 18*. The separation was on an Agarose-0.5 M column.

ponents in the reaction shown in *Figure 17* were also phosphorylated with ^{32}P-ATP. In a subsequent large-scale preparation of C (*Figure 18*), the labelling pattern was different. The extent of reaction shown in *Figure 17* was 60 to 70 per cent and the product was purified by gel filtration through an Agarose column. Analyses carried out by degradation to 3′- and 5′-nucleotides gave results as expected for C comprising the five components.

In a larger scale (such that ultra-violet absorbance can be measured) preparation shown in *Figure 18*, the labelling pattern was as shown in the insert in the figure. Thus two components carried ^{33}P (relatively weak), one component carried weak ^{32}P while the component with A at the 5′-end was labelled with ^{33}P of very high specific activity. (The same batch of ^{33}P-ATP was used as had been used in the preparation of A as shown in *Figure 16*.) The reaction was carried out as described in detail in the legend and the product was purified again by using an Agarose column. A shoulder was evident at the right of the first peak which contained the required product. The peak was pooled carefully so as to minimize contamination by the side product in the shoulder and the product (C) was analysed. The results given in *Table 3* are all as would be expected. It should also be noted that in the above preparations of C, segment 10 was deliberately excluded.

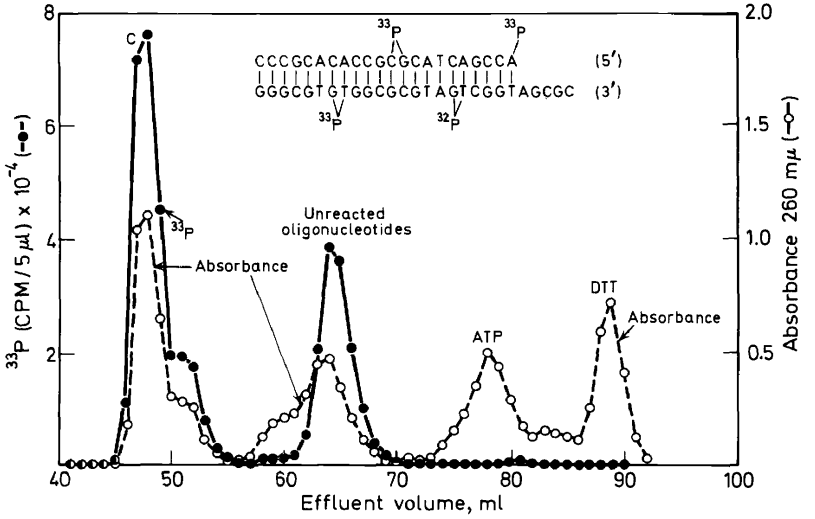

Figure 18. The synthesis and purification of C. Four of the five components of the system contained 5′-labelled phosphate groups as shown in the insert in the figure, the **^{33}P** group (segment 12) being much hotter than those used at the internal linkage points. The heptanucleotide (segment 15) was heated at 100° C for 3 min before using. The dodecanucleotide (segment 14, 18.2 mμmoles) and the heptanucleotide (18.2 mμmoles) were mixed in 5 mM $MgCl_2$ and 27 mM tris-HCl buffer (pH 7.7), the total volume being 1 ml. The mixture was heated to 70° C and then slow-cooled to 5° C to form the bihelical complex. The three decanucleotides (segments 11–13; 18–19 mμmoles of each) were then added and the total mixture (volume 2.1 ml) was pre-incubated for 5 h at 10° C. The final concentrations of other components were as follows: $MgCl_2$, 20 mM; tris-HCl (pH 7.6), 20 mM; DTT, 10 mM and ATP, 64 μM. Ligase (1 600 units) was then added and the joining was allowed to proceed at 10° C until the reaction had reached a plateau (4 h). The reaction was terminated by the addition of EDTA and the product (C) was isolated by gel filtration through an Agarose (0.5 M) column (104 × 1 cm). The elution was carried out at 4° using 0.1 M triethylammonium bicarbonate as the eluant. The identification of the various peaks is shown in the figure. C was isolated by careful pooling of the fractions belonging to the symmetrical part of the first peak.

Table 3. Characterization of C (Material of peak 1, *Figure 18*)
(Numerals in parentheses after the counts represent experimental molar ratios)

	5′-Nucleotide analysis in CPM			
	d-pA	d-pG	d-pT	d-pC
^{32}P	—	25	1084(1.2)	8
^{33}P	40900(1.0)	11	683(1.0)	722(1.1)
	5′-Nucleotide analysis after phosphatase in CPM			
	d-pA	d-pG	d-pT	d-pC
^{32}P	—	—	1088(1.1)	20
^{33}P	—	10	782(1.0)	778(1.0)
	3′-Nucleotide analysis in CPM			
	d-Ap	d-Gp	d-Tp	d-Cp
^{32}P	—	1019(1.0)	—	—
^{33}P	35	1438(1.7)	30	12

THE JOINING OF A TO B

The joining of A to B entails the very weak tetranucleotide overlap comprising the nucleotide sequence 17–20. Therefore, the joining reaction was carried out at rather high Mg^{2+} ion concentration and low temperature (5°). In the experiment shown in *Figure 19*, A contained weak ^{32}P-label while the

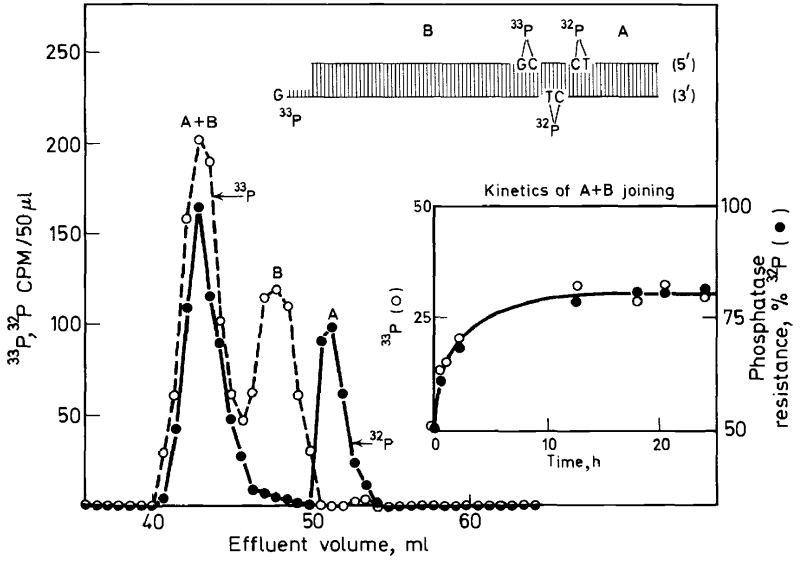

Figure 19. The synthesis and purification of A + B. The reaction mixture (25 µl) contained 27.5 µµmoles of each of the separately pre-annealed A and B in 15 mM tris-Cl (pH 7.6), 18 mM $MgCl_2$, 10 mM DTT and 66 µM ATP. After seven units of the ligase had been added, the mixture was incubated at 5°. The extent of joining was measured by the development of phosphatase-resistant labels (both ^{32}P and ^{33}P). ^{33}P was expected to increase from zero to a maximum of 50 per cent resistance and ^{32}P from 50 to 100 per cent resistance. The kinetics of joining are shown in the insert in the figure. After a total reaction time of 20 h, EDTA (30 mM) was added and the mixture was applied to the top of the same Agarose column used above in the experiment of *Figure 6*. Fractions of 0.75 ml were collected every 20 min and appropriate aliquots were examined for the ^{32}P and ^{33}P labels. The various peaks obtained are identified in the figure. Characterization of the pooled A + B peak was as shown in *Table 4*.

sample of B used was phosphorylated at its 5'-G ends using γ-^{33}P-ATP (the internal linkages in B contained insignificant ^{32}P-label). The linkages which contain significant radioactive isotope are shown in the structure inserted in *Figure 19*. Inserted also are the kinetics of the joining reaction as measured by the development of phosphatase-resistant ^{32}P and ^{33}P radioactivity. Both assays show that the joining occurs to the extent of about 60 per cent. The product (A + B) was isolated as shown in *Figure 19*. As expected, the first peak corresponding to the joined product contained both ^{32}P and ^{33}P labels in the expected ratio. Unreacted B containing only ^{33}P was eluted next and unreacted A containing only ^{32}P was eluted last. Full characterization of A + B isolated from the first peak is given in *Table 4*. All of the data are exactly as were to be expected for the product.

Table 4. Characterization of A + B (peak 1 of *Figure 19*)
(Numerals in parentheses after the counts represent experimental molar ratios)

	Phosphatase assay	
	Resistant	Sensitive
^{32}P-CPM	1240	0
^{33}P-CPM	1267	1301

3′-Nucleotide analysis after phosphatase

	d-Ap	d-Gp	d-Tp	d-Cp
^{32}P-CPM	0	26	2375(2.1)	0
^{33}P-CPM	14	70	20	2338(1)

THE JOINING OF B TO C

The preparation of B used in this experiment carried very weak ^{32}P-label in the internal linkages while the two 5′-G ends were labelled with *hot* ^{32}P. The sample of C used contained very weak ^{32}P and ^{33}P labels in the internal linkages while the terminal A carried *hot* ^{33}P. Segment 10 also was labelled with the same *hot* ^{33}P. B and C were annealed separately, mixed and segment 10 was added. Preincubation was now performed at 10° for 3 h. The kinetics of joining are shown in the insert to *Figure 20.* The reaction extent was in the range of 60 to 70 per cent. The product (B + C) was isolated as shown in the same figure. Characterization of the product is in *Table 5.* Thus, on

Table 5. Characterization of B + C (Material of peak 1, *Figure 20*)
(Numerals in parentheses after the counts represent experimental molar ratios)

Phosphatase assay in CPM

	Resistant	Sensitive
^{32}P	2890	3090
^{33}P	2755	260

5′-Nucleotide analysis in CPM

	d-pA	d-pG	d-pT	d-pC
^{32}P	38	2206(1.8)	31	0
^{33}P	710(1.0)	565(0.8)	0	0

5′-Nucleotide analysis after phosphatase in CPM

	d-pA	d-pG	d-pT	d-pC	Pi
^{32}P	49	1953(1.1)	21	0	1960(1.1)
^{33}P	1130(1.0)	935(0·9)	0	0	114

3′-Nucleotide analysis after phosphatase in CPM

	d-Ap	d-Gp	d-Tp	d-Cp
^{32}P	19	55	40	800(1.0)
^{33}P	0	567(1.1)	432(0.8)	0

degradation of the product to 5′-nucleotides, ^{32}P should be only in d-pG (the molar ratios of radioactive nucleotides calculated from the specific activities are shown in parentheses after the observed radioactivities); ^{33}P radioactivity should be in d-pA and d-pG in a molar ratio of 1 :1. On degradation to 5′-nucleotides *after* phosphatase treatment, the ^{32}P radioactivity

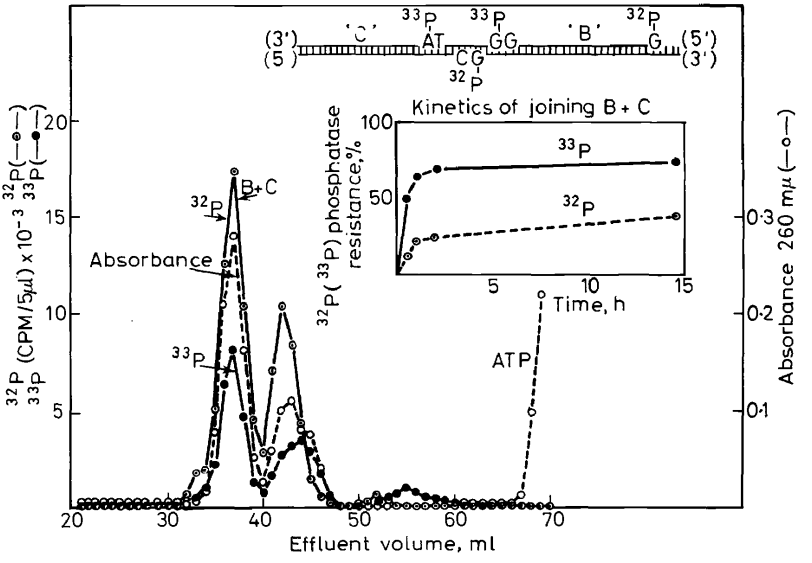

Figure 20. The synthesis and purification of B + C. As shown in the insert, C contained hot ^{33}P only on the 5'-nucleotide deoxyadenosine, B contained hot ^{32}P on both 5'-deoxyguanosines, and the decanucleotide (10) contained hot ^{33}P on the 5'-deoxyguanosine. All other radioactively labelled internucleotide bonds were relatively cold, characterization having been accomplished much earlier. Equimolar amounts (2 mμmoles) of B (0.3 ml) and of C (0.075 ml) were separately annealed at 5 mM $MgCl_2$, 20 mM tris-HCl buffer (pH 7.6) and 37° C. B and the segment 10 (2.1 mμmoles in 0.23 ml) which had been heated in a boiling water bath were then added to C. The total reaction solution (0.85 ml) was pre-incubated for 3 h at 6 mM $MgCl_2$, 23 mM tris-HCl buffer (pH 7.6), 10 mM DTT, and 10° C. Ligase (200 units) was then added and the joining was carried out for 2.5 h at 65 μM ATP. Then 70 units ligase were added and the reaction continued an additional 12 h. The reaction was terminated by addition of EDTA and the product fractionated at 4° C on a column (104 × 1 cm) of Agarose, 0.5 M (200–400 mesh). The eluant was 0.1 M triethylammoniun bicarbonate. The first peak was the product B + C. The second peak was a mixture of B and C.

should now be equally distributed between inorganic phosphate and d-pG while the ^{33}P label should be resistant to the phosphatase and again be equal between d-pA and d-pG. The results of phosphatase assay given separately in *Table 5* provide further confirmation. Degradation to 3'-nucleotides after a phosphatase treatment should show ^{32}P only in d-Cp while ^{33}P should be equally distributed between d-Gp and d-Tp. The data given in *Table 5* are all essentially consistent with these predictions.

TOTAL SYNTHESIS BY JOINING PREFORMED B + C TO A

The preformed B + C as described above was brought into reaction with A derived from the experiment of *Figure 16*. The labelling pattern at the linkage points is shown in the abbreviated structure of the gene inserted in *Figure 21*. The kinetics of the joining reaction are shown in the insert in the same figure. This reaction was carried out on a 'large' scale such that

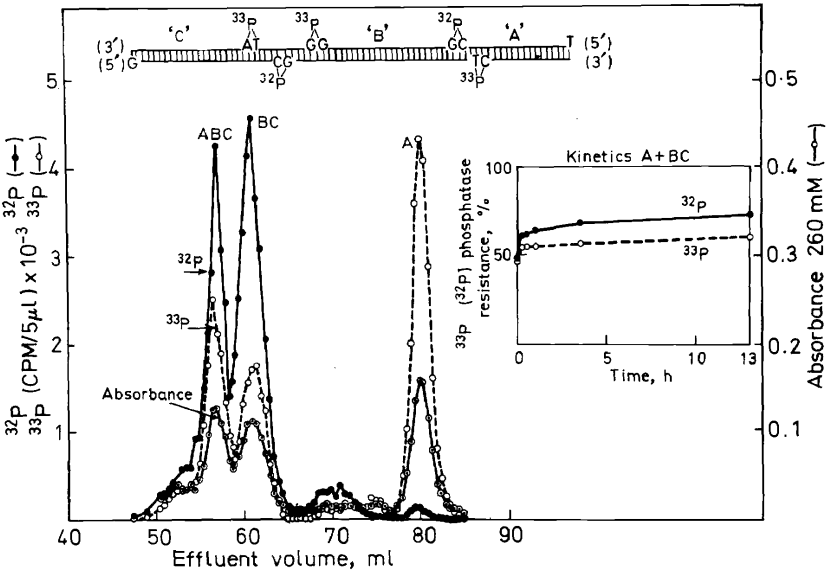

Figure 21. The synthesis and purification of A + B + C by the joining of A to preformed B + C. The insert indicates that A contains a hot ^{33}P only on the 5′-nucleotide deoxycytidine. The joined product B + C contains a *hot* ^{32}P only on one of the 5′-nucleotide deoxyguanosines. Internally, as shown in the insert, B + C contains two hot ^{33}P labels and one hot ^{32}P label. All other radioactively labelled internucleotide bonds are relatively cold and have previously been characterized. B + C (0.98 mμmole in 0.1 ml) and A (1.91 mμmole in 0.1 ml) were separately annealed at 5 mM $MgCl_2$ and 20 mM tris-HCl buffer (pH 7.6). B + C was annealed at 37° and A at 85° C. Both were slowly cooled to 5° C. The reaction mixture (0.29 ml) was then prepared and contained B + C, A and the following components: 21.5 mM $MgCl_2$, 22 mM tris-HCl buffer (pH 7.6), 10 mM DTT, and 64 μM ATP. After pre-incubation for 22 h at 5° C, ligase (160 units) was added and the joining allowed to proceed for 3.5 h. Then additional ligase (80 units) was introduced and the reaction allowed to proceed for 9.5 h. The reaction was terminated by addition of EDTA and the product fractionated at 4° C on a column (146 × 1.1 cm) of Agarose 0.5 M (200–400 mesh). The eluant was 0.1 M triethylammonium bicarbonate. The first peak was the product ABC. Re-chromatography of ABC has given a single peak. The second peak was B + C and the third corresponded to A.

ultra-violet absorbance of the different peaks could also be measured. The extent of the reaction (above 40 per cent) was much lower than expected but all the properties (the ultra-violet absorbance to ^{32}P to ^{33}P ratios) of the first peak were as expected for the total gene. Full characterization is described in *Table 6*. Thus, the phosphatase assay should show all ^{32}P and ^{33}P to be resistant and this was essentially found. Degradation to 5′-nucleotides should give ^{32}P only in d-pG while ^{33}P should be in d-pA, d-pG and d-pC in a ratio of 1 :1 :1 and this was again actually found. Analysis of the radioactivities on degradation to 3′-nucleotides was all consistent with expectation. As a further check that the product contained all the three parts (A, B and C) of the gene, the isolated product was treated with polynucleotide kinase using γ-^{32}P-ATP. The results in *Table 5* now showed ^{32}P

Table 6. Characterization of BC + A (Material of peak 1, *Figure 21*)
(Numerals in parentheses after the counts represent experimental molar ratios)

	Phosphatase assay in CPM			
	Resistant	Sensitive		
^{32}P	1752	125		
^{33}P	1845	35		
	5′-Nucleotide analysis in CPM			
	d-pA	d-pG	d-pC	d-pT
^{32}P	60	5760(2.1)	38	153
^{33}P	2205(1.1)	2045(1.0)	2155(1.1)	0
	5′-Nucleotide analysis after phosphorylation in CPM			
	d-pA	d-pG	d-pC	d-pT
^{32}P	66	3700(3.2)	102	1163(1.0)
^{33}P	1060(1.1)	930(1.0)	920(1.0)	0
	3′-Nucleotide analysis in CPM			
	d-Ap	d-Gp	d-Cp	d-Tp
^{32}P	0	205	6480(2.1)	194
^{33}P	0	2185(1.0)	0	4590(2.1)

in d-pT in addition to d-pG and the molar ratio was approximately 1:3 for d-pT to d-pG and this was exactly as was to be expected. (No ^{32}P radioactivity was found in the two other 5′-nucleotides.) The ^{33}P ratio in 5′-nucleotides remained as was found before phosphorylation with the kinase.

TOTAL SYNTHESIS BY JOINING PREFORMED A + B TO C

A second synthesis of the gene is illustrated in *Figure 22.* The preformed A + B (derived from experiment of *Figure 19*) was brought into reaction

Table 7. Characterization of (A + B) + C (Peak 1 of *Figure 22*)
(Numerals in parentheses after the counts represent experimental molar ratios)

	Phosphatase assay			
	Resistant	Sensitive		
^{32}P-CPM	739	27		
^{33}P-CPM	45	0		
	3′-Nucleotide analysis			
	d-Ap	d-Gp	d-Tp	d-Cp
^{32}P-CPM	0	25	27	0
^{33}P-CPM	10	41	1137(1)	294(2·0)
	5′-Nucleotide analysis after 5′-phosphorylation			
	d-pA	d-pG	d-pT	d-pC
^{32}P-CPM	0	737(1)	742(1)	18
^{33}P-CPM	364(1.0)	68	11	0
	5′-Nucleotide analysis after phosphatase and 5′- phosphorylation			
	d-pA	d-pG	d-pT	d-pC
^{32}P-CPM	8	513(1.0)	482(0.9)	32
^{33}P-CPM	179(0.9)	37	6	2

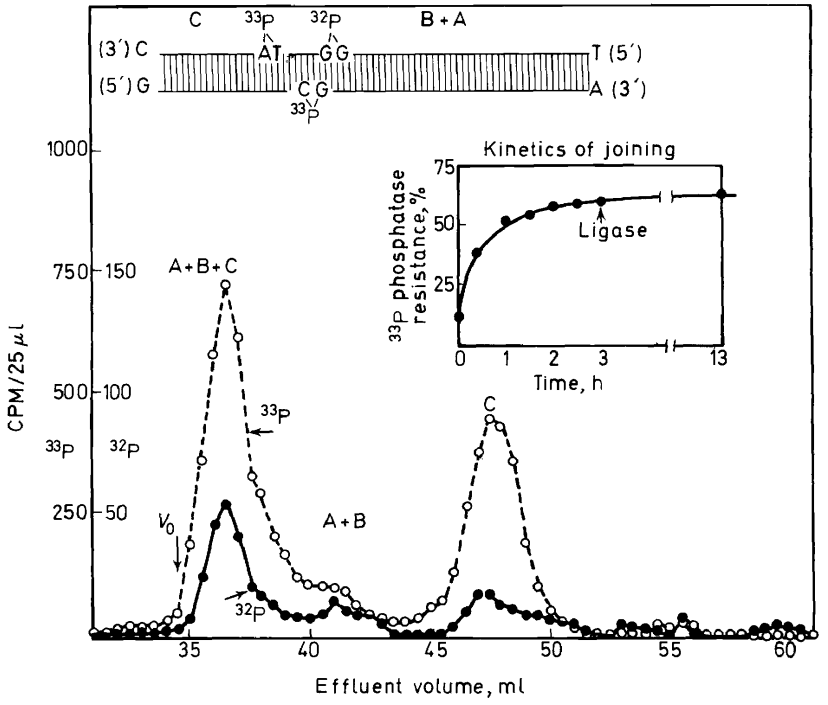

Figure 22. The synthesis and purification of A + B + C by the joining of preformed A + B with C in the presence of segment 10. The reaction mixture (35 μl) contained the preannealed A–B (from *Figure 19*) (32.5 μμmoles), C (38 μμmoles) and preheated weakly ^{32}P-labelled segment 10 in tris-HCl, pH 7.6 buffer. The $MgCl_2$ concentration was 5 mM. After annealing of the mixed components by slow-cooling (from 37° to 10° in about 4 h), seven units of the ligase were added. The kinetics of the joining reaction are shown in the insert, the assay measuring development of ^{33}P resistance to the phosphatase. For purification, the mixture after addition of EDTA (final concentration, 10 mM) was fractionated on an Agarose—0.5 M columns (115 cm × 1.1 cm diameter). Elution was at 4° by using 0.1 M triethylammonium bicarbonate. Aliquots were tested for ^{32}P and ^{33}P content. The elution pattern and identification of the peaks are shown in the figure. The total gene (A + B + C) was eluted at the void volume. The first peak pooled free from the shoulder at the right end was analysed (*Table 7*).

with C (derived from experiment of *Figure 18*) in the presence of the segment 10 containing weakly ^{32}P-labelled 5′-G end. The labelling pattern which is shown in the structure inserted in *Figure 22* shows the linkages where there was significant radioactivity, pronouncedly strong label being only at one site (the 5′-A unit in C). The kinetics of the joining reaction are also shown in the insert in *Figure 22*, the reaction having proceeded to about 65 per cent. The profile obtained on separation of the products is shown in the figure. The first peak (excepting the inflection at the right) corresponded to the total gene. Next came a small amount of unused A + B and finally the excess of C. Full characterization is reported in *Table 7* and all the data are nicely consistent with theoretical expectations. One additional experiment carried out with this preparation is also shown in *Table 7*. To check if any nick had been

made inadvertently during the total operations, the synthetic product was treated with chromatographically pure alkaline phosphatase at 70°. The product, after inactivation of the phosphatase, was phosphorylated with γ-^{32}P-ATP. Analysis carried out after degradation to 5′-nucleotides again provided additional confirmation of the continuous 77-unit long bihelical structure. We believe that these data together with those presented in the preceding tables go far in firmly establishing the identity and homogeneity of the totally synthetic gene.

CONCLUDING REMARKS

While the methods of synthesis and of characterization described above leave little doubt about the identity of the synthetic DNA, clearly a great deal more remains to be done to study the properties of the product. When this is done, the first urgent job is the replication of the synthetic gene for only then can one ensure adequate future supplies for the many studies that suggest themselves. Unpublished experiments by Drs Kjell Kleppe, Eiko Ohtsuka and Ian Molineux have given encouraging results on the use of DNA polymerase for replication of the gene in the presence of suitable primers. Transcription of the appropriate strand to produce the tRNA will be the next step and, because the gene lacks the natural start and stop signals for the action of the DNA-dependent RNA polymerase, the initiation and termination of transcription will have to be artificially controlled. Here again, unpublished results recently obtained by Drs T. Terao and R. Kleppe on the problem are promising.

For many of the studies on structure/function relationship in the tRNA molecules using well-characterized components of the protein synthesizing machinery, it would be desirable to have an *Escherichia coli* tRNA gene. Some two years ago it was decided to embark on the total synthesis of the gene for tyrosine suppressor tRNA. The primary sequence and the segmentation of the corresponding DNA duplex for purposes of chemical synthesis are shown in *Figure 23*. Chemical syntheses of all the segments are already complete (unpublished work of Drs Hans van de Sande, Marvin Caruthers, Ashok Kumar, K. Minamoto and Nina Sidorova) and the enzymatic joining is making progress (unpublished work of Drs Peter Besmer and Robert Miller).

In addition to the studies on the structure/function relationship in the tRNAs which have already been mentioned, a systematic study of the 'maturation' of the tRNA molecule could be initiated. What we hope to have first is a tRNA molecule which contains only the four common nucleosides. Step-by-step enzymic modification would then be studied concurrently with the effects of such modifications on the functions of the tRNA.

The principles used in the present work are such that they allow 'welding' of bihelical DNAs to one another. Hopefully, these principles will permit studies of the punctuation marks on DNA by addition of appropriate deoxypolynucleotide sequences at either end of the synthetic gene. The same principle could be used eventually to add the synthetic gene to other genomes such as those of the transducing phages. While all these possibilities belong

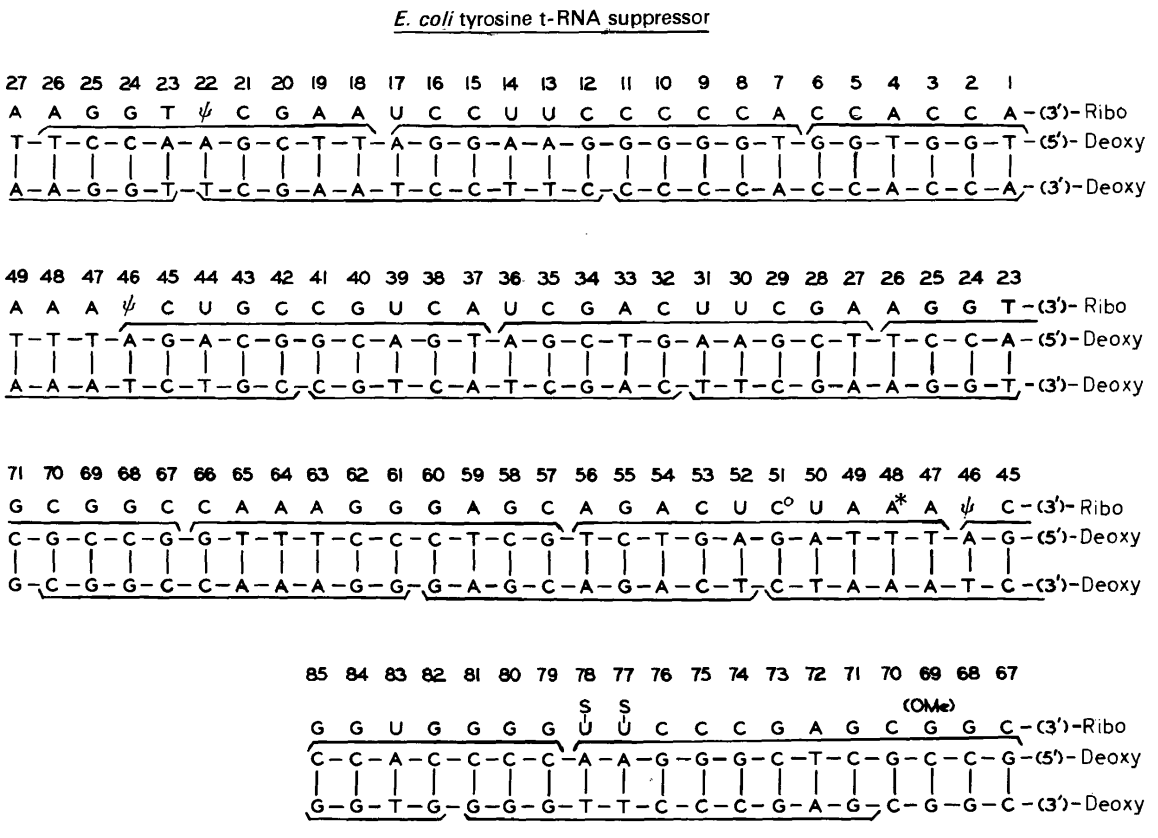

Figure 23. The primary sequence of *E. coli* tyrosine suppressor tRNA and the chemically synthesized segments representing the corresponding DNA duplex.

to the future, the present results nevertheless would appear to give an encouraging start.

ACKNOWLEDGEMENTS

In my research experience so far the work reported in this lecture has been the most demanding of sustained effort, determination and critical analysis. It is therefore my great pleasure to pay tribute to my colleagues who made the successful execution of this task possible. The joining work, which has formed the bulk of the present lecture, has been carried out by Drs V. Sgaramella, Hans van de Sande, Kjell Kleppe, Marvin Caruthers and T. Yamada. Many of the segments belonging to parts A, B and C were synthesized by Drs A. Kumar, E. Ohtsuka, M. Caruthers, K. Agarwal and H. van de Sande. Finally, all the early work on the ligase and consequent planning of the total synthesis was made possible by the painstaking and pioneering efforts of Drs Hans Weber and Henry Büchi who were responsible for the syntheses of the two icosanucleotides; this was at a time when there was little assurance that these would be of any use and the prospect of the total synthesis was very remote indeed.

The work has been generously supported by the National Cancer Institute of the National Institutes of Health (Grant No. CA05178), The National Science Foundation, Washington (Grant No. GB-7484X), The Life Insurance Medical Research Fund (Grant No. 65–44) and the Graduate School of the University of Wisconsin.

REFERENCES

[1] H. G. Khorana, *Pure Appl. Chem.* **17**, 349 (1968).
[2] H. G. Khorana, *Proceedings of the Seventh International Congress of Biochemistry*, Tokyo, 19–25 August, 1967, p 17 (1968).
[3] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick and A. Zamir, *Science*, **147**, 1462 (1965).
[4] N. K. Gupta, E. Ohtsuka, H. Weber, S. H. Chang and H. G. Khorana, *Proc. Nat. Acad. Sci., Wash.*, **60**, 285 (1968).
[5] N. K. Gupta, E. Ohtsuka, V. Sgaramella, H. Büchi, A. Kumar, H. Weber and H. G. Khorana, *Proc. Nat. Acad. Sci., Wash.*, **60**, 1338 (1968).
[6] N. K. Gupta and H. G. Khorana, *Proc. Nat. Acad. Sci., Wash.*, **61**, 215 (1968).