Synthesis of DNA's and their biological roles*

SARAN A. NARANG

Division of Biological Sciences, National Research Council of Canada, Ottawa, Canada

(*Delivered on* 20 *December* 1973)

The study of gene expression and the mechanism of its regulation in the living organism is one of the most fascinating frontier of molecular biology field. The hypothesis that the genes in all living organisms contain all the information required for the cell to reproduce is now more than 50 years old. In chemical terms, the basic idea of this hypothesis has been that the deoxyribonucleic acid (DNA) in all living organisms directs the synthesis of all the protein required by that organism. This control is exerted through the intermediacy of ribonucleic acids (messenger or mRNA, transfer or tRNA), which transcribe and transmit the information originally present in DNA (Figure 1). Although most of the predictions of the *central dogma* have been verified using biological approaches, the most successful and direct confirmation has been obtained by the use of synthetic DNA molecules during the mid-sixties.

During seventies, we have entered a most difficult era of biology to understand that how a biological activity of DNA expression is regulated. For example, when *E. coli* bacterium is grown on sugar lactose as its sole carbon source, the bacterium makes two proteins which are essential specifically for the metabolism of lactose. These are (1) *galactoside permease* which is found in the bacterium membrane and directs the transport and accumulation of lactose within the cell and (2) *β-galactosidase*, an enzyme inside the cell which catalyzes the hydrolysis of lactose into its two components monosaccharides, glucose and galactose. But when *E. coli* is grown on glucose as its carbon source, these two proteins are present in very low amounts of the order of 10 or so molecules per cell. However, in the presence of lactose, the rate of synthesis of these proteins is increased by as much as 1000 fold. The genes which determine the ability of *E. coli* to make and regulate the synthesis of these two proteins are clustered in a small region of *E. coli* chromosome. The structure of these proteins are determined by the *Z* gene (*β-galactosidase*) and the *Y* gene (*permease*). The transcription of these genes is initiated at a single site the *lac* promoter gene (*p*). A group of these genes thus transcribed into a single polygenic messenger RNA species is known as operon. In the absence of an inducer (which in this case is lactose sugar) of the *lac* operon, the transcription of these genes is prevented

* Coochbehar Professorship lecture, 1962.
** NR CC No. 13713

by the *lac* repressor peotein. This repressor acts to stop transcription by binding to the operator gene (*o*) DNA which lies between the promoter and structural genes. The inducer allows the transcription of the operon by binding to the repressor and causing its release from the operator gene[2] (Figure 2). The most

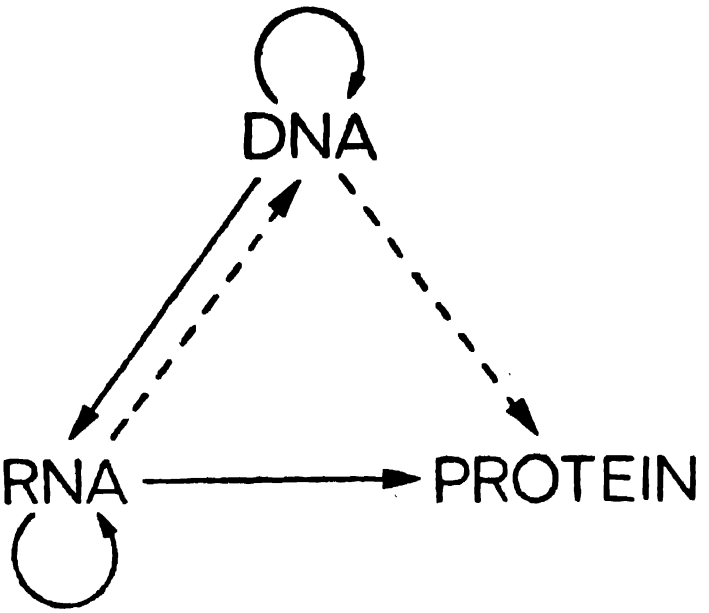


Figure 1. Modern central dogma.

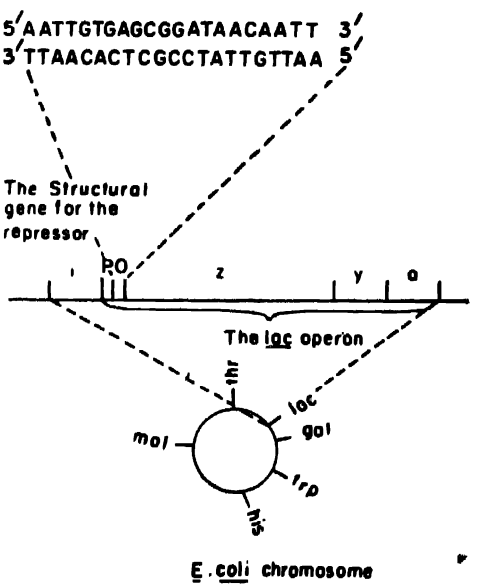


Figure 2. Lac-operon of *E. coli* chromosome.

fascinating part of this story is to understand that how a protein molecule can recognize a very small portion of a DNA molecule in a chromosome and stop the transcription process ? This question presents a biochemist with a very difficult paradox, because a molecule can *recognize* a message only in terms of geometry *i.e.*, the shape or configuration of the molecule bearing the message. It seems that nature has developed two evolutionary codes; (i) *genetic code* which signifies that how a linear sequence of four bases in a long chain of DNA is transferred into a very specific sequence of twenty amino acids in a protein; (ii) *recognition code* which relates that how a protein molecule of defined sequence can recognize and binds to a specific sequence of a DNA molecule. For many such investigations supplies of DNA molecules would be essential, but their pure form may not be derived from degradative procedures. Thus there have been attempts at a rational synthesis of such molecules not only because of the challenges of their chemical complexity but also for their usefulness in studying some cardinal biological processes.

Being an organic chemist, it is not possible theoretically to conceive any chemical approach for the total synthesis of DNA molecules. Fortunately, we are in the better position than protein chemists because of the availability of certain enzymes which have been successfully utilized for the synthesis of DNA macromolecules.

A. Genetic Code

1. *Synthesis of DNA containing the repeating base sequence*

The syntheses of these molecules were achieved by two steps (i) chemical synthesis of repeating polynucleotides up to the size nona- and decanucleotides; (ii) multification-amplication of these synthetic complementary chains by DNA polymerase enzyme into longer chains in the range 1–2 million molecule weight.

a. *Chemical Synthesis of Polynucleotide containing the repeating sequence by the Stepwise Condensation Method*

The prevalent approach used in the stepwise synthesis of specific deoxyribopolynucleotides has involved the successive addition of mono-, di-, tri-, or tetranucleotide units to the 3′-hydroxyl end of a 5′-*o*-trityl protected mono- or oligonucleotide[3], as illustrated in figure 3. As the chain enlarges, an increasing excess of the incoming 3′-*o*-acetyl protected unit 1 or 2 must be used to maintain satisfactory yields with respect to the *growing* chain. The condensing agents in current use are mesitylenesulfonyl chloride MS, trisopropylbenzenesulfonyl chloride TPS, or dicyclohexylcarbodiimide DCC. In most of the synthetic work MS or TPS reagents offer the particular advantages that trialkylammonium salts can be used to aid solubilization of the nucleotidic components in the reaction medium. Moreover, the presence of a trace of amine is a strong inhibitor

of the DCC reaction. After each condensation step, the reaction mixture was submitted to a mild alkali treatment to remove the 3′-*o*-acetyl group and the desired product was obtained by anion-exchange chromatography on a DEAE-Cellulose column using volatile triethylammonium bicarbonate *p*H 7.5 buffer.

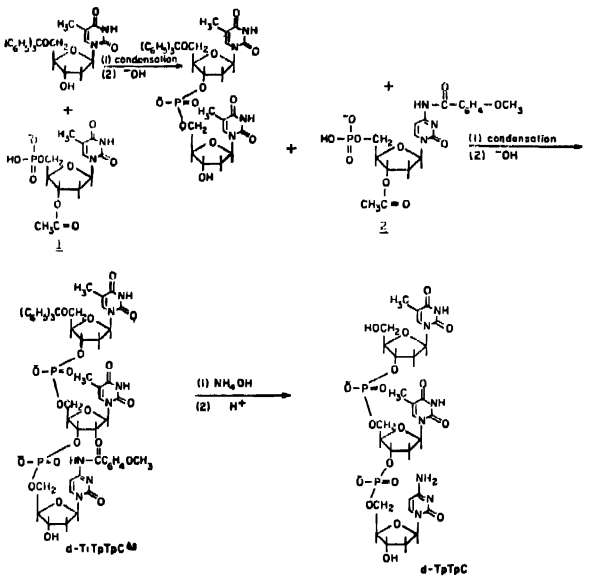


Fig. 3. A representative stepwise condensation to prepare oligonucleotide.

b. *Polymerization Method*

Homopolymers

These compounds have been prepared by the polymerization of mononucleotides containing a free hydroxyl group. Initial studies, involving the reaction of thymidine 5′-phosphate with DCC, produced a large number of polymeric products[4]. As a result, the major problem was the development of satisfactory chromatographic methods for separation and characterization of the desired linear polynucleotides, which were isolated mainly by chromatography on columns of cellulose anion exchanger (ECTEOLA) and diethylaminoethyl (DEAE) cellulose. Usually three homologous series of polynucleotides are obtained. The first are the linear polynucleotides represented by the general structure 3, while the second series of compounds contain the cyclic oligonucleotides represented by 4 (figure 4). They arise by an intramolecular phosphodiester bond formation between the 5′-phosphomonoester group and the 3′-hydroxyl group and third containing pyridine moiety at 5′-end and monophosphate at the 3′-end. The formation of the cyclopolynucleotide compounds can be reduced by adding some 25% of 3′-*O*-acetyl-thymidine 5′-phosphate to the unprotected nucleotide at

the start of polymerization. The 3'-OAc nucleotides form the terminating unit of the greater portion of the resulting polynucleotide chains thus blocking the cyclization reaction and the acetyl group can subsequently be removed by mild alkali treatment. Recently a thin-layer chromatographic technique[5] has been employed for the easy separation and characterization of various products of polymerization of thymidine 5'-phosphate (Figure 5). It has also been observed that when polymerization was carried out with one molar equivalent of mesitylene-sulfonyl chloride, the two-dimensional TLC pattern showed a much simpler and extensively polymerized products containing less side-products.

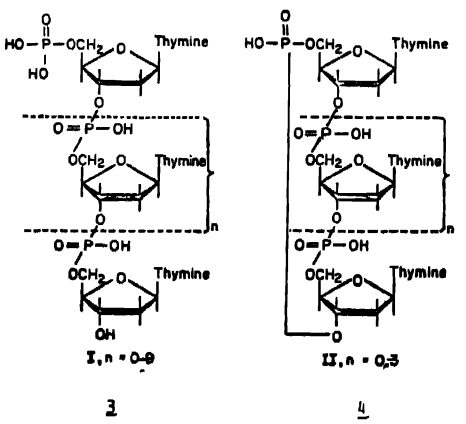


Fig. 4. Major components (linear and cyclic oligonucleotides) of a polymerization of pT.

Block Polymers

Similarly, the synthesis of polynucleotides containing repeating di-, tri- and tetranucleotides were carried by using the above procedure[6]. There is no doubt that polymerization methods available at present are inefficient and produce a plethora of contaminating side-products which require extensive chromatography. The ready formation of these side-products illustrates very well the reactivity of the phosphate group and the multitude of sites available for reaction in these molecules. Nevertheless, this method does provide a rapid means of obtaining modest quantities of short-chain deoxyribopolynucleotides containing repeating sequences.

Using these two chemical approaches, the syntheses of specific deoxyribopolynucleotides listed in table 1 were accomplished during mid-sixties. A general point about the list in this table is that the synthetic polynucleotides comprise sets which are complementary in the antiparallel Watson-Crick base pairing sense. This requirement of synthesis was vital due to the failure of DNA-polymerase to use a single-stranded deoxyribopolynucleotide as a template.

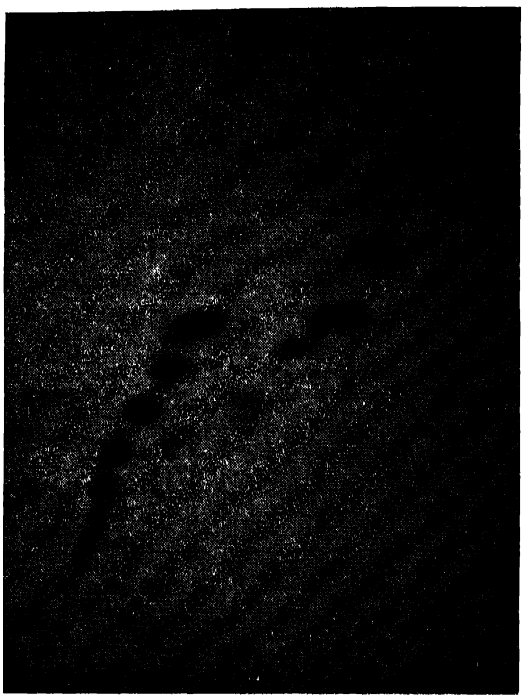


Fig. 5. Two-dimensional tlc fractionation of chemically polymerized thymidine 5′-phosphate on silica-gel tlc plate. Solvent I was used in the first dimension and solvent II in the second dimension.

Synthesis of DNA-like Polymers containing the repeating sequences

An enzyme has been isolated from *E. coli*[7] which catalyzes the synthesis of high-molecular-weight DNA in the presence of all four of the deoxynucleoside 5′-triphosphates, *d*-ATP, *d*-CTP, *d*-GTP, and TTP, plus a DNA template. No synthesis of DNA takes place in the absence of this template, which may consist of short-chain synthetic oligonucleotides or longer chains isolated from natural sources. For example, it has been demonstrated that short-chain synthetic deoxyribopolynucleotides containing alternating deoxyadenylate and thymidylate units serves as a template, in the presence of thymidine 5′-triphosphate and deoxyadenosine 5′-triphosphates, to induce the synthesis of a high-molecular-weight DNA-like polymer containing again deoxyadenylate and thymidylate units in alternating sequences (see Table 2, Reaction 1). Table 2 lists the types of reactions so far elicited from DNA-polymerase. These reactions are also very fast; polymers much larger than the original template can be produced in several hours.

The most direct method for showing the chemical structure of the polymer is by the so-called nearest neighbour analysis technique[8], in which one α-P^{32}

labeled deoxyribonucleoside triphosphate is incorporated into the chain at a time. Degradation of the synthetic polymer to deoxyribonucleoside 3′-phosphates using enzymatic methods gives results which show complete accord with theoretical expectations from the sequence of the original template. Typically, the results from such a polymerization using templates with complementary repeating trinucleotide sequences[9] is shown in table 3. Within experimental error, excellent agreement is observed concerning the incorporation of the various 5′-triphosphate monomers.

Table 1. Synthetic deoxyribopolynucleotides with repeating sequences

Repeating Trinucleotide Sequences

$\left\{\begin{matrix} d(TTC)_4 \\ d(AAG)_4 \end{matrix}\right\}$ $\left\{\begin{matrix} d(CCT)_{3-5} \\ d(CCA)_{3-5} \end{matrix}\right\}$ $\left\{\begin{matrix} d(TAC)_{4-6} \\ d(TAG)_{4-6} \end{matrix}\right\}$

$\left\{\begin{matrix} d(TTG)_{4-6} \\ d(CAA)_{4-6} \end{matrix}\right\}$ $\left\{\begin{matrix} d(CGA)_{3-5} \\ d(CGT)_{3-5} \end{matrix}\right\}$ $\left\{\begin{matrix} d(ATC)_{3-5} \\ d(ATG)_{3-5} \end{matrix}\right\}$

$\left\{\begin{matrix} d(GGA)_{3-5} \\ d(GGT)_{3-5} \end{matrix}\right\}$

Repeating Tetranucleotide Sequences

$\left\{\begin{matrix} d(TTTC)_3 \\ d(AAAG)_{3-4} \end{matrix}\right\}$ $\left\{\begin{matrix} d(TATC)_3 \\ d(TAGA)_2 \end{matrix}\right\}$ $\left\{\begin{matrix} d(TTAC)_4 \\ d(TAAG)_2 \end{matrix}\right\}$

Table 2. DNA-polymerase catalyzed reactions

(1) d-$(AT)_3$ + d-ATP + TTP → Poly d-AT

(2) T_{11} + d-A_7 + d-ATP + TTP → Poly d-A : T

(3) T_{11} + d-A_7 + d-ATP → Poly d-A

(4) d-$(TG)_6$ + d-$(AC)_6$ + $\left\{\begin{matrix} TTP \\ dATP \\ dCTP \\ dGTP \end{matrix}\right\}$ → Poly d-TG : CA*

(5) d-$(TTC)_4$ + d-$(AAG)_3$ + $\left\{\begin{matrix} TTP \\ dATP \\ dCTP \\ dGTP \end{matrix}\right\}$ Poly d-TTC : GAA*

(6) d-$(TATC)_3$ + d-$(TAGA)_2$ + $\left\{\begin{matrix} TTP \\ dATP \\ dCTP \\ dGTP \end{matrix}\right\}$ → Poly d-TATC : GATA*

* All of the DNA-like polymers are written so that the colon separates the two complementary strands. The complementary sequences in the individual strands are written so that antiparallel base-pairing is evident.

An additional property of these synthetic DNA-like polymers is their ability to reseed the synthesis by the DNA polymerase of more of the same product. The importance of this finding can hardly be overstressed. In essence, it means that once the specific sequences have been put together by well-defined and unambiguous chemical synthesis. DNA-polymerase will ensure their permanent availability. Thus the well-known, dramatic feature of DNA-structure, its ability to guide its own replication, can be exploited at the molecular level. The total number of DNA-like polymers prepared so far is listed in table 4.

Table 3. Nearest neighbour frequency analysis of poly d-TTC : GAA templates : $d(TTC)_4+d(AAG)_3$

α-P^{32}-labelled triphosphate	Radioactivity in deoxynucleoside 3′-phosphates							
	dAp		dGp		dCp		dTp	
	count/min	%	count/min	%	count/min	%	count/min	%
dATP	12,836	50.0	12,851	50.0	0	0	0	0
dGTP	13,684	100.0	0	0	0	0	0	0
dCTP	0	0	0	0	0	0	9.623	100.0
dTTP	0	0	0	0	12,860	50.6	12,565	49.4

Table 4. New DNA-like polymers with repeating sequences

Repeating Dinucleotide Sequences	Repeating Trinucleotide Sequences	Repeating Tetranucleotide Sequences
	Poly d-TTC : GAA	
Poly d-TC : AG	Poly d-TTG : CAA	Poly d-TTAC : GTAA
Poly d-TG : AC	Poly d-TAC : GTA	Poly d-TATC : GATA
	Poly d-ATC : GAT	

The complete characteristics of the DNA-polymerase-catalyzed reactions can be summarized as follows : (a) chemically synthesized segments corresponding to both strands are required for reaction to proceed : (b) minimal size of the two complementary segments used as primers varies between 8 and 12 nucleotide units : (c) synthesis is extensive : (d) products are high molecular weight and are double stranded with sharp melting transitions : (e) nearest-neighbour analysis invariably indicates that the individual strands contain the appropriate repeating sequences : (f) high-molecular-weight products can be reutilized as primers for more synthesis.

An electron micrograph of poly *d*-TG : AC showed the average size to be 0.5μ. This is indicative of a molecular weight in the range of one million.

Single-stranded RNA with repeating Di- and Trinucleotide sequences

A DNA-dependent RNA-polymerase[10] is present in bacteria and other tissues. It is similar in action to DNA-polymerase and requires a double-stranded segment of DNA as a primer plus all four ribonucleoside 5'-triphosphates to be effective. The polyribonucleotides thus synthesized are single-stranded and of high molecular weight and their composition is determined by the composition of the DNA template. If synthetic polydeoxyribonucleotides with a known base sequence are used as templates, then a polyribonucleotide having a complementary defined base sequence is obtained. Typical results obtained with various synthetic polydeoxyribonucleotide primers for DNA-dependent RNA-polymerase are given in table 5.

Table 5. DNA-dependent RNA-polymerase reactions

poly d-TG : AC—	CTP+ATP RNA Polymerase GTP+UTP	Poly CA Poly GU
poly d-TAC : GTA—	UTP+ATP+GTP RNA Polymerase UTP+ATP+CTP	Poly GUA Poly UAC
poly d-TATC : GATA—	UTP+ATP+GTP RNA Polymerase UTP+ATP+CTP	Poly GAUA Poly UAUC

Such results emphasize the fact that only by providing in the reaction mixture a set of ribonucleotide 5'-phosphates which are appropriate for copying the template, the required single-stranded polyribonucleotide can be obtained. In every case the single-stranded ribopolynucleotides thus prepared have been shown to contain the expected repeating sequences by the technique of nearest-neighbour analysis. A template of only 9 nucleotides (3 triplets) has been shown to be sufficient for synthesizing a complementary polyribonucleotide with more than 150 nucleotide units.

Cell-free Amino Acid incorporation and the Codon assignments

Using poly AAG as a messenger, three homopeptides, *i.e.*, polylysine, polyarginine and polyglutamine were obtained whereas the poly AG yielded poly arg-glu as a copeptide. These observations offered a direct proof for the three letter and the non-overlapping nature of the genetic code. The individual assignment was confirmed by tri-nucleotide-stimulated binding of amino acyl SRNA to ribosomes[11]. The results are given in Tables 6 and 7.

Table 6. Amino acids incorporations stimulated by messengers containing repeating nucleotide sequences

(Systems, *E. coli* B)

Messenger	Amino acids incorporated	Messenger	Amino acids incorporated
Repeating Dinucleotides		Poly GUA	val, ser
Poly UC	ser-leu	Poly UAC	tyr, thr, leu
Poly AG	arg-glu	Poly AUC	ileu, ser, his
Poly UG	val-cys	Poly GAU	met, asp
Poly AC	thr-his		
		Repeating Tetranucleotides	
Repeating Trinucleotides		Poly UAUC	tyr, leu, ileu, ser
Poly UUC	phe, ser, leu	Poly GAUA	none
Poly AAG	lys, glu, arg	Poly UUAC	leu, thr, tyr
Poly UUG	cys, leu, val	Poly GUAA	none
Poly CAA	gln, thr, asn* ?		

* The expected incorporation of asparagine has not been realized so far because of the presence of a powerful enzyme which deaminates asparagine in the amino acid incorporating system (cf. Schwartz 1965).

B. Recognition Code

The recognition of *lac* operator by *lac* repressor is a good example of a specific and tight interaction between protein and DNA molecules. Its dissociation constant *in vivo* is 10^{-11} *M*. There are 10 repressor molecules per cell (concentration of $10^{-8}M$) yet this very small number is sufficient to keep the operator gene 99.99% blocked. Until recently, nothing was known about the structure of operator gene and the chemistry of ineraction between the operator and the repressor. The experiments so far shows that the double-standard DNA is required. It implies that interaction involves either the outside of the native helix or requires the participation of both strands. The next question is that how big is the operator gene ? To the extent that this region is a stretch along the DNA to which a protein molecule the size of one submit of the *lac* repressor (38,000 mol wt) can bind, it can be only between one and two turns of the DNA molecule (35 to 70A, 10 to 20 bases). Since repressors bind to one site on the *E. coli* DNA with very high affinity it means that the region to which they bind must be unique in the chromosome (3×10^6 base pairs) and thus it must have at least the specificity of 12 bases size. Therefore it is not unreasonable to predict that the size of the operator gene should be between 12 to 21 base pairs. The primary objective in this problem is to elucidate the base sequence of the operator gene. After a couple of years hard work Gilbert and his coworkers have been

able to determine its nucleotide. This was achieved by sonicating the *lac* repressor bound *E. coli* DNA followed by the isolation of the complex. The free DNA component was liberated by adding lactose as inducer. The nucleotide sequence was determined (personal communication with Professor Ray Wu of Cornell University) by its transcription with DNA-dependent RNA polymerase in the presence of low concentration of triphosphates and GpA as primer. The complete sequence is shown in figure 2. Recently the sequence of about 50 amino acids at the N-terminus of the *lac* repressor has been determined. It is believed this part of the sequence is essential for the direct binding to *lac* operato DNA.

Table 7. Codon assignments from polypeptide synthesis and/or stimulation of amino acyl-sRNA binding to ribosomes

1st	2nd				3rd
	U	C	A	G	
U	*Phe*	*Ser*	*Tyr*	*Cys*	U
	Phe	Ser	*Tyr*	Cys	C
	Leu	*Ser*	*Ochre*	?	A
	Leu	Ser	*Amber*	Trp	G
C	*Leu*	Pro	*His*	Arg	U
	Leu	*Pro*	*His*	Arg	C
	Leu	Pro	*Gln*	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ilou	*Thr*	Asn	*Ser*	U
	Ileu	Thr	*Asn*	Ser	C
	Ileu	*Thr*	*Lys*	*Arg*	A
	Met	Thr	*Lys*	Arg	G
G	*Val*	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	*Glu*	Gly	A
	Val	Ala	*Glu*	Gly	G

The assignments not underlined are on the basis of binding experiments only. The assignments singly underlined are on the basis of copolypeptide and/or homopolypeptide syntheses, and gave essentially no binding. The assignments doubly underlined are derived from both polypeptide synthesis and binding experiments.

In order to study the chemistry of interaction between *lac* repressor protein and *lac* operator gene, it is important to have a reasonable amount of these materials in hand. With the recent advancement in the field of bacteriophage, it has been possible to isolate pure *lac* repressor protein in large amounts, but the amount of operator gene available from *E. coli* is very discouraging. With this objective in view we have recently undertaken a total synthesis of *lac*-operator gene. For this purpose we have investigated two approaches in our laboratory which are outlined below.

2. *Chemical synthesis of defined sequence*

(i) *Diester approach using the benzoylated DEAE-cellulose or sephadex column.* Recently we have been able to develop a new diester synthetic approach[12] which offers an easy procedure for the separation of the growing chains of intermediate oligonucleotides from the incoming nucleotidic units. The basic principle of our new approach is that the attachment of an aromatic protecting group to the 5′-phosphate of a nucleotidic unit markedly increases its binding to benzoylated DEAE-sephadex. On condensation of 5 with a second compound, a 3′-*O*-acetyl N-protected nucleoside 5′-phosphate 6 containing non-aromatic protecting groups, the aromatic group will be found only in the unreacted starting material 5 and the product 7 (figure 6). Thus, when the reaction mixture is passed through a benzoylated DEAE-Sephadex column 5 and 7 will be the only components of the reaction mixture retained by the BD-DEAE-Sephadex due to the affinity between their aromatic rings. All the other components which lack the aromatic group such as unused 6, pyrophosphate of 6 (a serious side-product) will be eluted quickly. The components 5 and 7 can then be easily eluted by washing the column with stronger buffer solutions containing ethyl alcohol (see figure 7). The desired product is then separated from the starting material by either gel[13] filtration (figure 8) or preparative tlc[14].

In order to develop this approach, it was considered essential to search for suitable phosphate protecting groups containing a benzene ring. Various substituted phenols were investigated as potential protecting groups. Although aryl esters of mononucleoside 5′-phosphate could easily be prepared by the usual dicyclohexylcarbodiimide reaction, more drastic conditions (2N sodium hydroxide, 50–100° 15–16 min) were generally required for the complete hydrolysis. This treatment resulted in the formation of small amounts of deamination and other side-products. However, 3,5-dinitrophenol, 2-nitro-4-chlorophenol, and pentachlorophenol were found to be promising, since these groups only required the action of 2N sodium hydroxide (100°, 15 min) for the complete hydrolysis of their nucleotidic esters.

Next, various aromatic hydracrylamide derivatives, prepared by condensing β-propiolactone with the appropriate aromatic amines, were studied. Their

esters, known to cleave under mild alkali condition via a β elimination mechanism, were prepared by reaction with mononucleoside 5'-phosphate in the presence of dicyclohexylcarbodiimide or triisopropylbenzenesulfonyl chloride. As expected, treatment with 2N sodium hydroxide at room temperature smoothly hydrolyzed the phenylhydracrylamide ester in 45 min, *p*-methoxyphenylhydracrylamide ester in 90 min, and the benzylhydracrylamide ester in 8 hr.

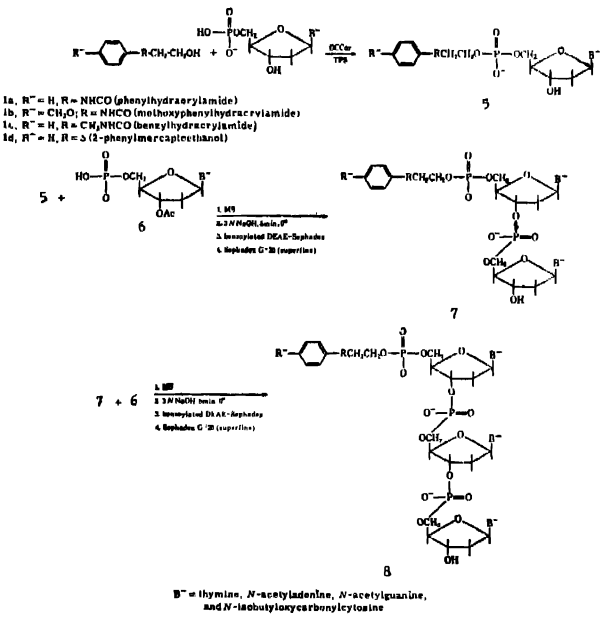


Fig. 6. A new approach for the synthesis of oligonucleotide containing terminal monophosphomonoester group.

Finally, we investigated 2-phenylmercaptoethanol[15], a new phosphate protecting group, which has many novel features. It can easily be introduced under the usual conditions and the corresponding sulphide ester is very stable to conditions commonly encountered in oligonucleotide synthesis, such as (a) 2N sodium hydroxide at room temperature for 4 hrs.; (b) concentrated ammonium hydroxide at 50° for 2 hr.; (c) aqueous pyridine at room temperature for 1 week; (d) mesitylenesulfonyl chloride in anhydrous pyridine at room temperature for 3 hr.; (e) aqueous hydrochloric acid (*p*H 2) at room temperature for 3 days. However, when *activated* this group is readily removed. Thus, the 2-phenylmercaptoethyl group was removed by treating the protected nucleotide in aqueous solution or in 0.1*M* triethylammonium bicarbonate buffer with sodium metaperiodate (approximately 5*M* excess—as a 0.1*M* aqueous solution) at room temperature for 1 hr. It was apparently oxidized to the corresponding sulfoxide derivative. After decomposition of the excess periodate with ethylene glycol the solution

was concentrated to dryness *in vacuo*. The free nucleotide was then liberated by treatment with 2N sodium hydroxide at room temperature for 30 min. It was also surprising to find that the sulfide esters bind more strongly to Sephadex than their corresponding oxidized derivatives sulfoxide. This novel property also offers a potentially new separation technique. Using this approach the synthesis of two nonanucleotides bearing the 5′-terminal sequences of *lac*-operator gene have been achieved.

(ii) *Modified tri-ester approach.* The tri-ester method has recently been examined in various laboratories in an attempt to overcome some of the difficulties inherent in the "diester approach", such as low yields with increasing chain length, instability of the internucleotidic phosphodiester bond and cumbersome purification procedures. The essential feature of the *tri-ester approach* involves phosphorylation of the terminal 3′-hydroxyl group of a 5′-protected mononucleoside with a simple monophosphate ester (for example, β cyanoethyl phosphate, phenyl phosphate or β', β', β', trichloroethyl phosphate) and triisopropylbenzenesulfonylchloride (TPS) followed by subsequent condensation with the primary 5′-hydroxyl group of an incoming 3′-suitably protected nucleoside and additional TPS. Since each internucleotidic bond is present in the desired product as a tri-ester function, purification at each stage by more conventional techniques such as organic solvent extraction followed by a simple chromatography on silica-gel columns becomes possible. However, we have observed that the initial phosphorylation reaction of the 3′-hydroxyl group of the starting 5′-protected mononucleoside is incomplete, thus the addition of the incoming 3′-protected mononucleoside leads to a very complicated reaction mixture. These mixtures cannot be completely separated using ordinary silica-gel chromatography, hence the advantage of large scale synthesis and high yields is somewhat nullified. However, this difficulty can be ovecome by the following simple modification which involves initial phosphorylation of the 5′-protected nucleoside with a suitably substituted phosphate in the presence of TPS followed by isolation of the phosphorylated product by silica-gel column chromatography in 80–85% yield. The phosphodiester group was further protected with cyanoethanol and TPS to obtain in quantitative yield.

Using this approach the synthesis of a pentadecadeoxypolynucleotide has been achieved[15]. We have now accomplished the synthesis of two deoxypolynucleotide fragments up to twelve and fifteen bases containing the specific sequence of the *lac*-operator gene. The main strategy of the present synthetic approach (figure 7) is to condense a 5′-o-tritylated deoxynucleotide containing 3′-monophosphate ester of *p*-chlorophenyl 8 with a second component 9 containing free 5′-hydroxyl of a deoxynucleotide with fully protective 3′-phosphate with *p*-chlorophenyl as well as trichloroethyl protecting groups in the presence of triisopropylbenzene-sulfonyl chloride. After the usual work, the desired product

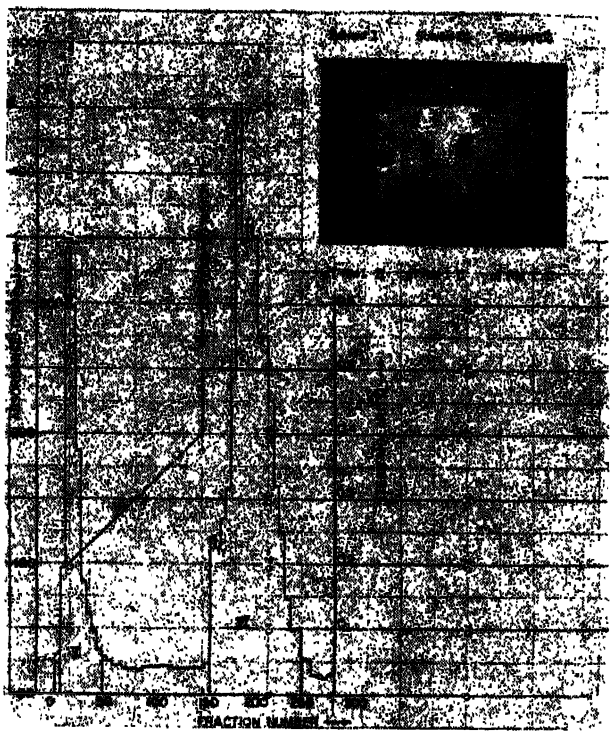
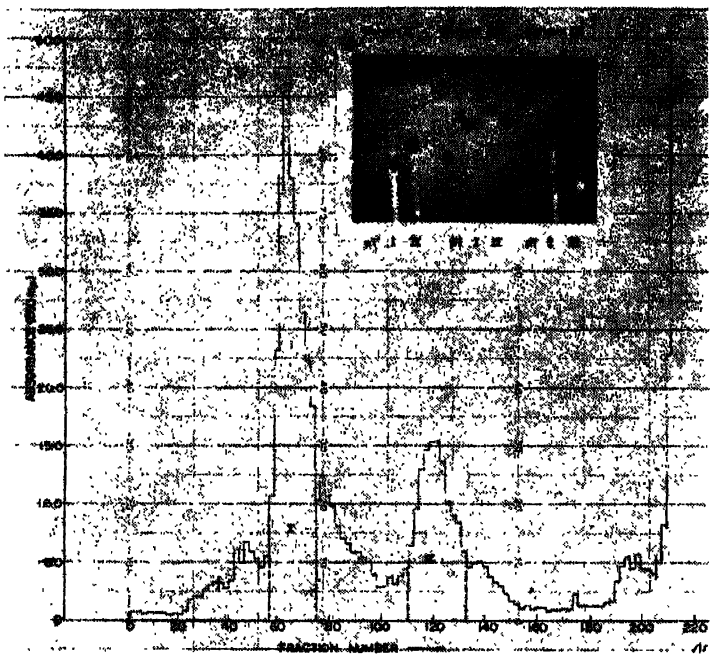


Fig. 7. Column chromatography on a benzoylated DEAE-sephadex column.

which is fully protected desired component was quite easily separated from the reaction mixture simply by column chromatography on silica-gel based upon the following rationale : Component 8 will strong bind to silica-gel column due to its one negative charge whereas the separation between the product 10 (containing tritylyl group) and the second component 9 containing hydroxyl group is quite easily achieved by eluting the column with an appropriate mixture of chloroform; methanol (methanol ratio varies from 1–4%). After the isolation of the desired product, the trichloroethyl protecting group was removed by treatment with Zn/Acetic acid in pyridine and the product was further used for the next condensation. The removal of the various protecting groups were carried out in the following order : (i) Zn/Cu couple treatment to remove trichloroethyl group; (ii) base treatment to remove N-acyl and *p*-chlorophenyl group, and finally (iii) 80% acetic acid treatment to remove trityl group.

It is noteworthy to mention that the oligonucleotides containing guanine bases were obtained in unsatisfactory yields (20–25%). It seems that the chloride ion liberated during the reaction caused the formation of side-products. We have recently developed a new condensing reagent mesitylene sulfonyl triazolide 12 which gives consistent yields.



Gel-filtration on sephadex G-25 (SF) for the isolation of trinucleotide.

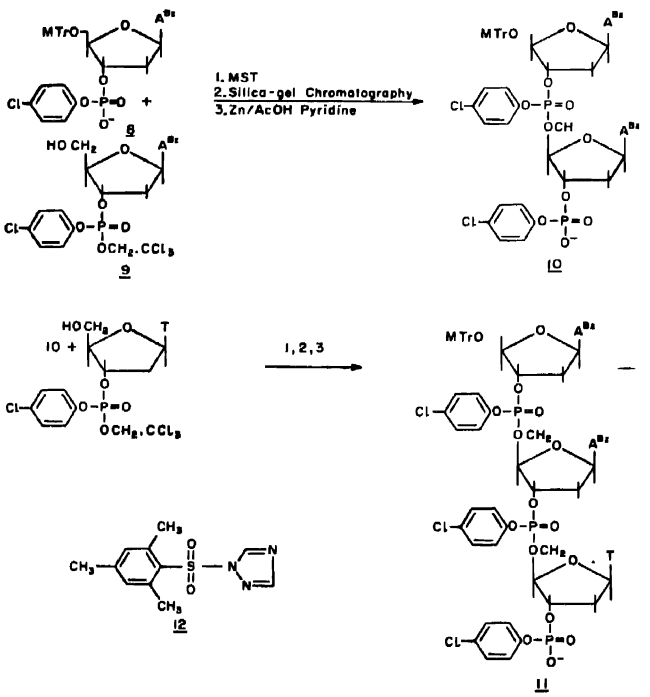
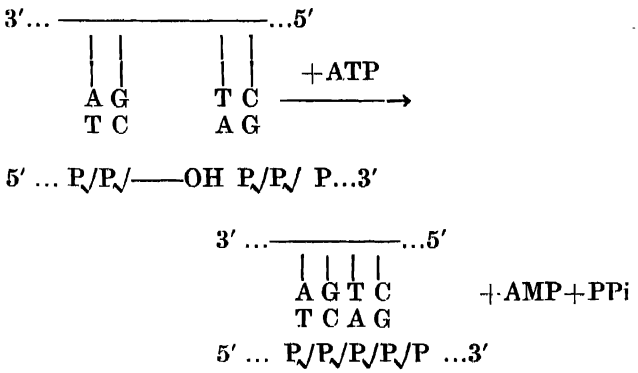


Fig. 9. Modified triester synthetic approach.

Enzymatic synthesis of the longer polynucleotides

I. *Polynucleotide ligase*

Recently, enzymes[17] which catalyze the covalent joining of breaks in a single strand of bihelical DNA have been identified and purified from *E. coli*, phage-T_4 and phage-T_7 infected *E. coli*. Given a DNA substrate containing single-stranded breaks, as shown below, the ligase enzyme accomplished the repair by the esterification of an internally located 3′-hydroxyl group with the adjacent 5′-phosphomonoester. In the reaction ATP is cleaved to AMP and PPi.



Studies on *E. coli* and T_4-induced ligases have revealed that they have many properties in common. Both enzymes are specific for the same DNA substrates and produce the same DNA end product. Furthermore, the reactions catalyzed by both enzymes are mediated by an enzyme-adenylate (enzyme AMP) complex. However, a distinguishing feature of the two enzymes is that the adenylate moiety for each of the two complexes is derived from different cofactors. The cofactor for *E. coli* enzyme is DPN, whereas that for the T_4-induced enzyme has been shown to be ATP.

Much has been deduced concerning the overall mechanism of the reaction. In both cases, the first step in the overall reaction consists of the transfer of an adenylate group from the cofactor to the enzyme to form a covalently linked enzyme-AMP intermediate. Once enzyme-AMP is formed by a reaction of the enzyme with DPN, there is a further transfer of the AMP to the 5-phosphoryl terminus of a DNA chain to generate a new pyrophosphate bond linking the AMP and DNA. In the final step, the DNA phosphate of the pyrophosphate is presumably attacked by the 3′-hydroxyl group of the neighbouring DNA chain, displacing the activating AMP and forming the phosphodiester bond. However, specific characteristic made this enzyme of singular importance to the problem of the total synthesis of biologically specific high-molecular-weight DNA. Initial studies showed that a short deoxyribopolynucleotide carrying a 5′-phosphate group at one end and a 3′-hydroxyl group at the opposite end could be joined

end-to-end in the presence of long complementary deoxyribopolynucleotides. These experiments also showed that chains as short as hexa- or heptanucleotides could be joined. Using this approach Khorana and his workers[18] have achieved the synthesis of alanni t-RNA gene.

With this successful application of the ligase enzyme in joining synthetic polynucleotides, it was considered essential to study the fidelity of the T_4-infected *E. coli* ligase enzyme. Recently it has been discovered that this enzyme can also catalyze the covalent joining of interrupted deoxyribooligonucleotide strands with one mispaired base at the 3'-hydroxyl terminus of a bihelix($P^{32}T_{11}C+$ Poly dA)[19]. Similar types of mismatched joining as well as end-to-end dimerizations have subsequently been reported to occur during the synthesis of alanine t-RNA gene[20]. A recent report showed that terminal crosslinking in DNA strands by an enzyme system consists of the T_4-ligase and another activity considered likely to be an exonuclease[21].

II. *Extension of the synthetic primer with* E. coli *DNA-polymerase enzyme*

Recently Wu and his coworkers[22] have reported an elegant approach which involves the extension of the chemically synthesized primer with labelled deoxynucleotides by *E. coli* DNA-polymerase I in the presence of an appropriate DNA-strand under the repair condition. This approach also facilitate the sequence analysis as the chain length of the prime grows in size. This approach is in full investigation (in collaboration with Professor R. Wu) to achieve the total synthesis of the *lac*-operator gene and also to carry out its sequence analysis.

Once the synthesis of *lac*-operator gene is achieved and its sequence studies completed, physico-chemical tools such as X-ray crystallography, n.m.r. and esr will be extensively employed to elucidate the mechanism of the recognition code between the specific protein and DNA molecule.

References

1. H. G. Khorana, H. Büchi, H. Ghosh, N. Gupta, T. M. Jacob, H. Kössel, R. Morgan, S. A. Narang, E. Ohtsuka & R. D. Wells, *Cold Spring Harb. Symp.* **31**, 39 (1966).
2. F. Jacob & J. Monod, *J. Mol. Biol.* **3**, 318 (1961).
3. S. A. Narang & H. G. Khorana, *J. Amer. Chem. Soc.* **87**, 2981 (1965); E. Ohtsuka & H. G. Khorana, *J. Amer. Chem. Soc.* **89**, 2195 (1967); T. M. Jacob & H. G. Khorana, *J. Amer. Chem. Soc.* **87**, 368 (1965).
4. G. M. Tener, H. G. Khorana, R. Markham & E. H. Pol, *J. Amer. Chem. Soc.* **80**, 6223 (1958); A. F. Turner & H. G. Khorana, *J. Amer. Chem. Soc.* **81**, 4651 (1959); H. G. Khorana & J. P. Vizsolyi, *J. Amer. Chem. Soc.* **83**, 675 (1961).
5. J. J. Michniewicz, C. P. Bahl, K. Itakura, N. Katagiri & S. A. Narang, *J. Chrom.* **85**, 159 (1973).

6. E. Ohtsuka, M. W. Moon & H. G. Khorana, *J. Amer. Chem. Soc.* **87**, 2956 (1965); S. A. Narang, T. M. Jacob & H. G. Khorana, *J. Amer. Chem. Soc.* **89** 2167 (1967); J. M. Jacob, S. A. Narang & H. G. Khorana, *J. Amer. Chem. Soc.* **89**, 2177 (1967).
7. A. Kornberg, I. R. Lehman, M. J. Bessman, & E. S. Simms, *Biochim. Biophys. Acta* **21**, 197 (1956); I. R. Lehman, M. J. Bessman, E. S. Simms & A. Kornberg, *J. Biol. Chem.* **233**, 163 (1958); C. C. Richardson, C. L. Schildkraut, H. V. Aposhian & A. Kornberg, *J. Biol. Chem.* **239**, 222 (1964).
8. H. Schachmann, J. Adler, C. Radding, I. Lehman & A. Kornberg, *J. Biol. Chem.* **235**, 3242 (1960)
9. R. D. Wells, T. M. Jacob, S. A. Narang & H. G. Khorana, *J. Mol. Biol.* **27**, 237 (1967).
10. J. Hurwitz, J. August, J. Davidson & W. E. Cohen, *Progress in Nucleic Acid Research*, Academic Press, New York, London, Vol. I, **59** (1963); J. Hurwitz, *Methods in Enzymology*, Academic Press, New York, London, Vol. VI, **23** (1963); M. Chamberlin & P. Berg, *Proc. Nat. Acad. Sci. U.S.A.* **48**, 81 (1962); E. Fuchs, W. Zillig, P. H. Hofschneider & A. Preuss, *J. Mol. Biol.* **10**, 546 (1964); M. Chamberlin & P. Berg, *J. Mol. Biol.* **8**, 297 (1964).
11. M. Nirenberg, T. Caskey, R. Marshall, R. Brimacombe, D. Kellog, B. Doctor, D. Hatfield,J. Levin, F. Rothman, S. Pestka, M. Wilcox & F. Anderson, *Cold Spring Harbor Symp.* **31**, 11 (1966).
12. S. A. Narang, O. S. Bhanot, J. Goodchild, R. H. Wightman & S. K. Dheer, *J. Amer Chem. Soc.* **94**, 6183 (1972).
13. S. A. Narang, J. J. Michniewicz & S. K. Dhorr, *J. Amer. Chem. Soc.* **91**, 936 (1969).
14. S. A. Narang & J. J. Michniewicz, *Anal. Biochem.* **49**, 379 (1972).
15. R. H. Wightman, S. A. Narang & K. Itakura, *Canad. J. Chem.* **50**, 456 (1972).
16 K. Itakura, C. P. Bahl, K. Katagiri, J. J. Michniewicz, R. H. Wightman & S. A. Narang, *Canad. J. Chem.* **51**, 3626 (1973).
17. S. B. Zimmerman, J. W. Little, O. K. Oshinsky & M. Gellert, *Proc. Natl. Acad. Sci. U.S.* **57**, 1841 (1967); B. Olivera, & I. R. Lehman, *Proc. Natl. Acad. Sci. U.S.* **57**, 1426 (1967); M. L. Gefter, A. Becker & J. Hurwitz, *Proc. Natl. Acad. Sci. U.S.* **58** 240 (1967); B. Weiss & C. C. Richardson, *Proc. Natl. Acad. Sci. U.S.* **57**, 1021 (1967); N. R. Cozzarelli, N. E. Melechen, T. M. Jovin & A. Kornberg, *Biochem. Biophys. Res. Commun.* **28**, 578 (1967).
18. K. L. Agarwal, H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhandary, J. H. Van de Sande, V. Sgaramella, H. Weber & T. Yamada, *Nature* **227**, 27 (1970).
19. C. M. Tsiapalis & S. A. Narang, *Biochem. Biophys. Res. Commun.* **39**, 631 (1970).
20. V. Sgaramella, J. H. Van de Sande & H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.* **67**, 1468 (1970).
21. B. Weiss, *Proc. Natl. Acad. Sci. U.S.* **65**, 652 (1970).
22. R. Padmanabhan, R. Padmanabhan, R. Wu, *Biochem. Biophys. Res. Comm.* **48**, 1295 (1972).

STATEMENT ABOUT OWNERSHIP AND OTHER PARTICULARS ABOUT NEWSPAPER INDIAN JOURNAL OF PHYSICS AND PROCEEDINGS OF THE INDIAN ASSOCIATION FOR THE CULTIVATION OF SCIENCE, VOL. 48, 1974.

FORM IV

(*See* Rule 8)

1.	Place of Publication	Indian Association for the Cultivation of Science, 2 & 3, Raja Subodh Mallick Road, Jadavpur, Calcutta-32
2.	Periodicity of its publication	Monthly
3.	Printer's Name Nationality Address	Sri Prokash Chandra Chakrobertty Indian Eka Press, 204/1, B. T. Road, Calcutta-35
4.	Publisher's Name Nationality Address	Sri Samarendranath Sen Indian Registrar, Indian Association for the Cultivation of Science, 2 & 3, Raja Subodh Mallick Road, Jadavpur, Calcutta-32.
5.	Editor's Name Nationality Address	As below

1. Dr. D. Basu
Indian,
Director,
Indian Association for the Cultivation of Science
Jadavpur, Calcutta-32.

2. Prof. A. Bose
Indian,
Department of Magnetism,
Indian Association for the Cultivation of Science,
Jadavpur, Calcutta-32.

3. Prof. S. D. Chatterjee
Indian,
91 Ballygunge Place,
Calcutta-19.

4. Prof. K. L. Chopra
Indian,
Sr. Professor of Physics
Indian Institute of Technology, Hauz Khas,
New Delhi-29.

5. Prof. Krishnaji, Indian,
Pro-Vice Chancellor
Senate House
Allahabad University
Allahabad, U. P.

6. Dr. D. N. Kundu
Indian,
Director, Saha Institute of Nuclear Physics,
92, Acharya Prafulla Chandra Road,
Calcutta-9.

7. Dr. A. P. Mitra
Indian,
National Physical Laboratory,
Hillside Road,
New Delhi-12

8. Dr. B. D. Nag Choudhuri
Indian,
Vice-Chancellor
Jawaharlal Nehru University
New Mehrauli Road
New Delhi-57.

9. Dr. T. Pradhan
Indian,
Saha Institute of Nuclear Physics
92, Acharya Prafulla Chandra Road,
Calcutta-9.

10. Prof. G. N. Ramachandran
Indian,
Institute Professor of Biophysics
Indian Institute of Science
Molecular Biophysics Unit
Bangalore-560012

11. Dr. R. Ramanna
Indian,
Director, Bhabha Atomic Research Centre
Trombay, Bombay-85.

12. Prof. U. R. Rao
Indian,
Director
Indian Scientific Satellite Project
A 3-6 Peenya Industrial Estate
Bangalore-560022

13. Prof. S. C. Sirkar
Indian,
122/1B, South Sinthi Road,
Calcutta-30.

14. Prof. A. R. Verma
Indian,
Director
National Physical Laboratory,
Hillside Road,
New Delhi-12.

15. Prof. A. K. Barua
(*Hony. Secretary*)
Indian,
Dept. of General Physics & X-Rays.
Indian Association for the Cultivation of Science,
Jadavpur, Calcutta-32.

6. Name and address of individuals who own the newspaper and partners or shareholders holding more than one per cent of the total capital

Indian Association for the Cultivation of Science.
2 & 3, Raja Subodh Mallick Road,
Jadavpur, Calcutta-32.

I, Sri Samarendranath Sen, hereby declare that the particulars given above are true to the best of my knowledge and belief.

Sd/- SAMARENDRANATH SEN

Dated : 18-2-75.

Signature of Publisher