Tetrahedron Vol. 39, No 1, pp. 3 to 22, 1983
Printed in Great Britain.
0040-4020/83/010003-20$10.00/0
Pergamon Press Ltd

TETRAHEDRON REPORT NUMBER 140

DNA SYNTHESIS

SARAN A. NARANG
Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada, K1A 0R6

(*Received in U.S.A.* 23 *February* 1982)

CONTENTS

INTRODUCTION 3
Rationale for synthetic genes
Synthetic strategy
CHEMICAL SYNTHESIS 5
Historical background
(i) Phosphodiester approach
(ii) Phosphotriester approach
(iii) Modified phosphotriester approach
(iv) Phosphite-triester approach
(v) Polymer support synthesis
Protecting groups
(i) Primary hydroxyl
(ii) Secondary hydroxyls
(iii) Primary amino
(iv) Phosphate
New condensation and phosphorylating reagents
Triazole as phosphate activator
General method of oligonucleotide synthesis by the modified phosphotriester approach
Final deblocking steps
Nature of side products
Sequence analysis of fully protected oligomers by mass spectrometry
Sequence analysis of unprotected oligomers containing $3' \rightarrow 5'$ phosphodiester linkages
(i) Mobility-shift procedure
(ii) Base-specific chemical cleavage
(iii) Dideoxy method
ENZYMATIC SYNTHESIS 14
Synthesis of double-stranded DNA
Enzymatic repair of partial duplex
Cloning of synthetic genes
(i) Specific method involving synthetic cohesive ends
(ii) General method involving linkers
(iii) Retrieval of gene from the cloning vehicle
BIOLOGICAL ROLE OF SYNTHETIC DEOXYRIBOOLIGONUCLEOTIDES 19
Probe for gene
Protein-DNA interaction
Tools in molecular cloning of DNA
Site-specific mutagenesis
Inhibition of *Rous Sarcoma* virus replication and cell transformation by a specific oligonucleotide
Molecular structure of synthetic DNA fragments
Uptake of synthetic deoxyribooligonucleotide sequence in *Haemophilus* cells
CONCLUDING REMARKS 20
Acknowledgement
REFERENCES 20

INTRODUCTION

The golden age of DNA began with the first important evidence that DNA, an organic phosphate macromolecule, is the genetic material in all the living organisms: the discovery reported in 1944. The DNA prepared from one strain of pneumococcus could "transform" another strain.[1] A second remarkable event was the discovery in 1953 of the complementary double-stranded (duplex) structure of DNA and with its recognition of how the molecule can be replicated.[2] These discoveries led to the realization

that DNA has two major and discrete functions. One is to carry the genetic information that brings about a specific phenotype of the cell. DNA is transcribed into RNA and the RNA is translated into amino acid language of proteins. The other major function of DNA is its own replication.

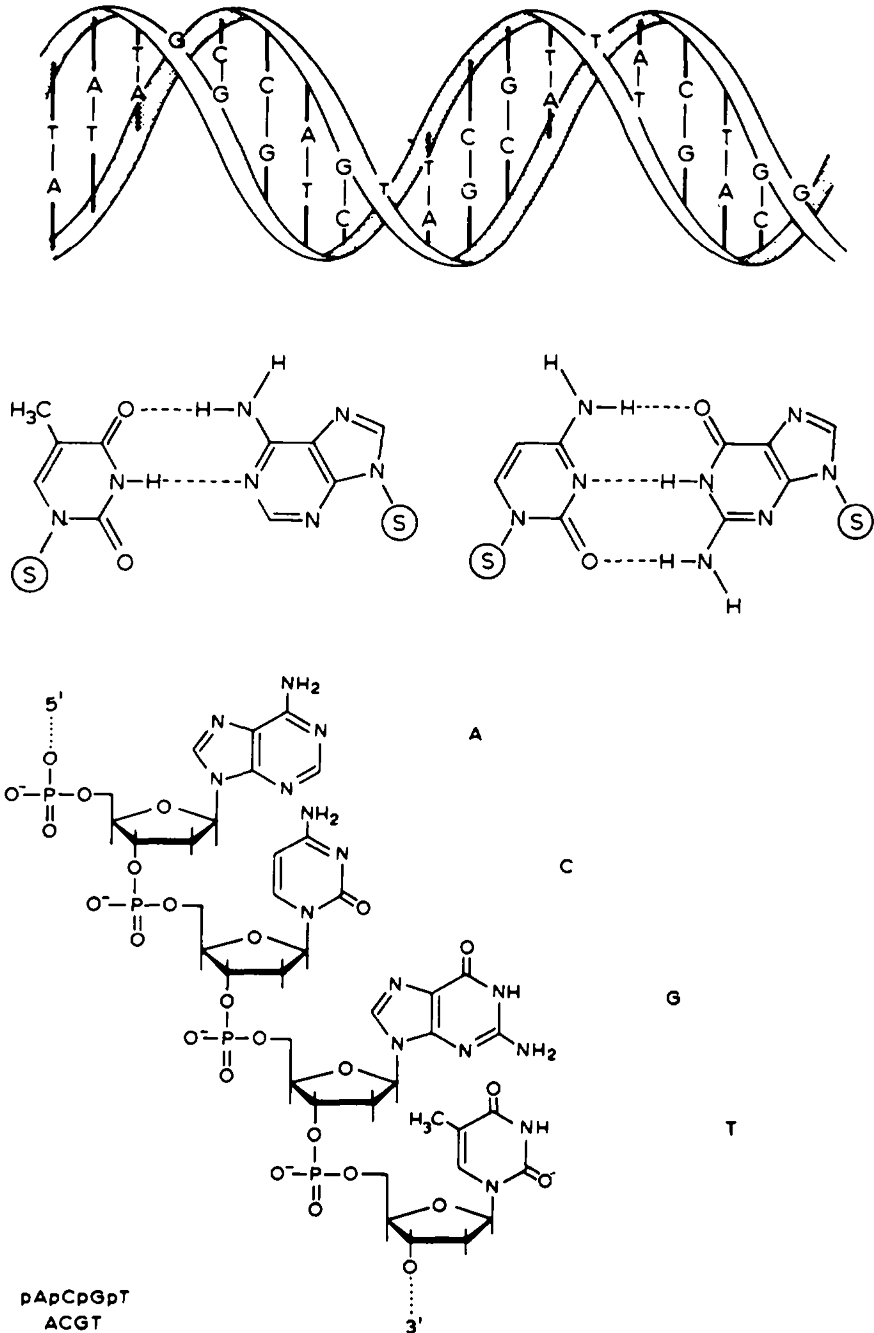

Fig. 1. Watson-Crick structure for DNA.

These discoveries brought a radical change in view-point towards DNA. Genetics and the biology of DNA became a branch of chemistry. Despite its complexity, like any other natural product, the DNA molecule was modified, dissected, analyzed and finally synthesized in the test tube. With these advancements, the beginning has been made to understand the organization, control of replication and gene expression well enough to explain how the cells develop, differentiate and die. At the molecular level, the sequence of heterocyclic bases contain information of DNA functions in terms of genetic code, recognition sites for restriction-endonuclease enzymes, control signals and invert repeats for transposal elements, etc.

Rationale for synthetic genes

Tailor-made DNA is quite valuable to understand the molecular biology of DNA. The elucidation of a genetic code using synthetic DNA containing repeating base sequences,[3] synthetic regulatory signals,[4] linkers[5] and primers,[6] has contributed significantly in understanding the biological role of different sequences in genes.

Among the several possible ways in which a gene can be obtained, total synthesis offers the following advantages: (i) it gives directly the exact desired sequence, the coding sequence and the non-coding sequences can be designed at will for prokaryotic expression, restriction enzyme sites can be removed or built in, introns deleted; (ii) it bypasses the often difficult step of isolating the relevant mRNA or genomic DNA; (iii) it simplifies the modification of the gene and its product protein by lengthening or shortening the coding regions, or by changing specific codons and the corresponding amino acids.

Until recently, the labor and time required for chemical synthesis of gene fragments were such as to make the enterprise quite unattractive, in spite of advantages cited above. Improved synthetic methods and the development of a DNA-recombinant technology during the past few years have now made the complete synthesis of a gene a practical proposition.

Synthetic strategy

The chemical synthesis of polynucleotides beyond 40 units long with any of the present methodologies is not possible for the time being. Decreasing yields dictate a practical compromise for chemical synthesis of oligonucleotides to about 20 units. A strategy has been developed for the synthesis of longer chains. Short synthetic oligonucleotides (usually 10–20 bases in length) are then joined with the use of DNA ligase enzyme in the presence of a complementary strand or template. The sequence of such a gene is preceded and terminated in specific endonuclease-restriction sites sequence. The complete synthetic sequence is ligated to a plasmid and cloned in *Escherichia coli*. The cloned DNA is selected and analyzed for the presence of a correct sequence. The importance of this method of DNA synthesis can be hardly overstressed to synthetic chemists; once the specific sequences have been assembled by a well-defined chemico-enzymatic method and cloned in *E. coli*. The bacterium containing chimeric DNA ensures their permanent availability, an expected but nevertheless dramatic feature of DNA structure, namely, the ability to guide its own replication.

CHEMICAL SYNTHESIS

From a chemical point of view, oligonucleotides are very sensitive molecules possessing diverse functionality, uncommom solubility and difficult purification problems in the conventional chemical sense. Even more difficult is their structure analysis by the standard physico-chemical methods. In spite of these obstacles, their synthesis can be a fascinating challenge to organic chemists because these are the "master molecules of living organisms."

The fundamental objective in oligonucleotide synthesis is the formation of an ester linkage between an activated phosphoric acid function of one nucleotide with the hydroxyl group of another nucleoside/nucleotide, thus ultimately forming the natural phosphodiester bridge between the 5′-OH of one nucleoside unit and the 3′-OH of the next. To achieve these ends, the nucleic acid chemist must devise (i) selective blocking/deblocking procedures for a primary, and a secondary OH (or two in ribonucleosides), a primary amino group and often two of the three dissociable groups of phosphate. In addition he must be cognizant of the variable labilities of certain purines and pyrimidines and their glycosidic bonds plus the variable reactivities of mono-, di- and trisubstituted phosphate. One must surmount some frustrating stereochemical and neighboring group effects especially noticeable as the oligonucleotide chain grows. All protecting groups must be removed without causing chain cleavages and with minimal formation of side products. The required product must be separated from these almost identical impurities and methods devised to prove unambiguously the structure.

Historical background

The pioneering research from Todd's laboratory[7] in the fifties was followed by Khorana's phosphodiester approach which dominated the field of oligonucleotide synthesis for almost twenty years beginning in the late 1950's. This group was mainly responsible in developing important techniques such as protecting groups, phosphorylation procedures, condensation reaction and enzymatic characterization which played a major role in further development of modern chemistry. In spite of the fact that the phosphodiester approach was laborious and time consuming, his group did succeed in accomplishing the

total synthesis of a biologically active tyrosine t-RNA gene[8] as a climax. By mid-1960's, other organic chemists such as Letsinger,[9] Reese[10] and Eckstein[11] reintroduced the phosphotriester approach in which the third phosphate dissociation was masked during synthesis. It is not before 1973, when the phosphotriester approach was further modified by Narang[12] and Cramer[13] independently and established its practicality in achieving the total synthesis of a first biologically active genetic element.[4] This method has now become the more practical and simpler approach as evident from a recent international symposium series on synthesis.[14]

(i) *Phosphodiester approach.* The traditional method of oligonucleotide synthesis, developed and championed by Khorana *et al.*, is the so-called phosphodiester approach.[15] As shown in Fig. 2, the

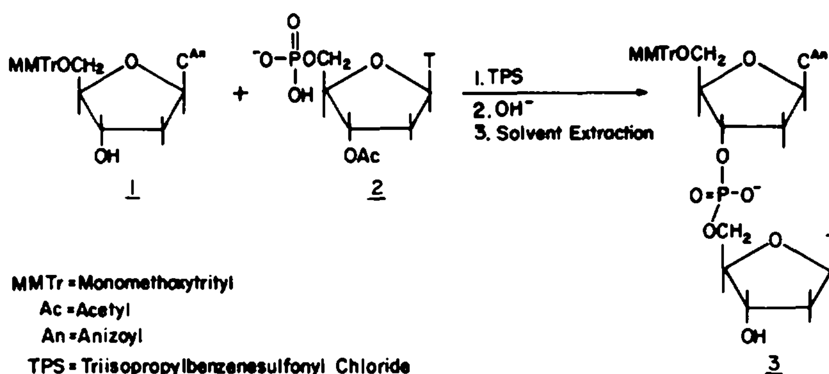

Fig. 2. Phosphodiester approach.

5'-phosphate of a commercially available nucleotide **2**, after suitable protection of other function groups was condensed with the 3'-hydroxyl of another protected nucleoside or nucleotide **1** using dicyclohexylcarbodiimide or arylsulfonyl chloride (MS or TPS) and the chain could then be elongated by further step-wise block condensation. Long reaction times, drastically dereasing yields as the chain length grew unless large excesses of one component of the condensation was provided, and time-consuming purification procedures via anion exchange column chromatography made this approach a tedious one. Recent improvements by using solvent extraction procedures[16] or hplc chromatographic techniques[17] have helped somewhat to alleviate this problem. Little wonder a computer optimized synthetic path[18] predicted 20 man-years to synthesize a gene of approximately 100–150 units by this approach.

(ii) *Phosphotriester approach.* A possible solution to some of these problems (water solubility, ion exchange chromatography and lower yields) inherent in the phosphodiester approach would accrue if the third dissociation of phosphate was masked, thus creating a neutral organic molecule amenable to the more standard manipulations of organic chemistry. Actually, this approach was first introduced by Michelson and Todd[7] in 1955 as outlined in Fig. 3.

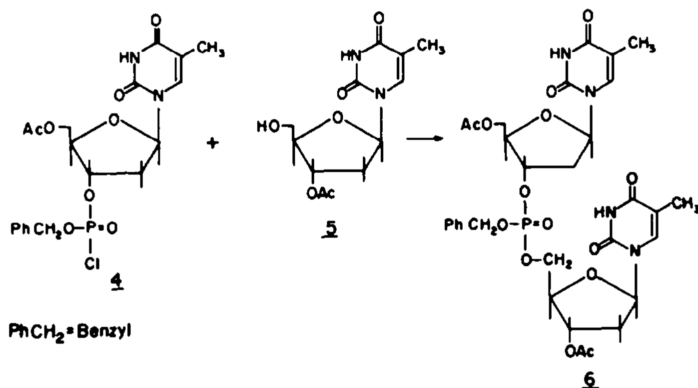

Fig. 3. Phosphotriester approach (Todd 1955).

The basic principle of the triester method is to mask each internucleotidic phosphodiester function by a suitable protecting group during the course of building a defined sequence. After completing synthesis,

all the protecting groups are removed at the final step to give a deoxyoligonucleotide containing each internucleotidic 3′→5′ phosphodiester linkage. The main advantages of this method include the opportunity for large-scale (50-75 g) synthesis, significantly shorter time periods especially in the purification steps and high yields using almost stoichiometric amounts of reactants. This is probably due to the absence of any *endo*-P-O$^-$ groups in the oligonucleotide chain thus avoiding chain scission and pyrophosphate formation.

The triester method as originally reported[9-11] involved the phosphorylation of the 3′-OH group of a 5′-protected nucleoside **7** with a substituted phosphate **8** followed by subsequent condensation with a primary 5′-OH of a second nucleoside **10** (Fig. 4). This is essentially a one-pot procedure. However, it

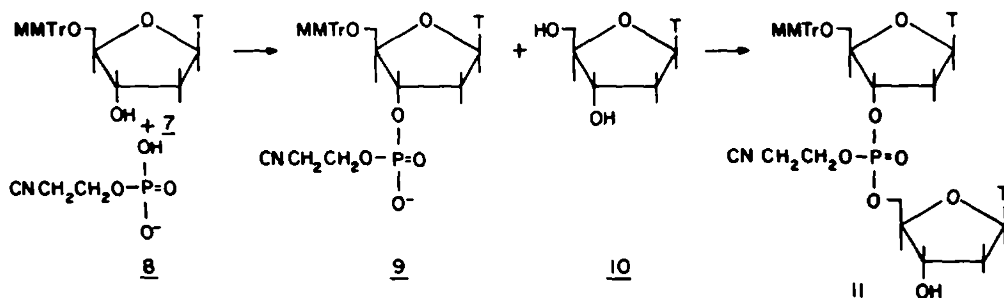

Fig. 4. Phosphotriester approach (Letsinger 1967).

was observed that owing to incomplete phosphorylation in the first state, subsequent coupling with 5′-protected nucleoside led to a complicated reaction mixture including the so-called 3′-3′ and 5′-5′ coupled products. Since these mixtures could not be completely resolved on conventional silica gel columns, the advantages of large-scale synthesis and high yield of product were somewhat nullified.[19]

(iii) *Modified phosphotriester approach.* To overcome this difficulty the "one-pot" synthetic approach which includes phosphorylation and coupling steps was modified to a "two-step" sequential procedure.[12,13] The basic feature of this approach is to start the synthesis of an oligonucleotide from a totally protected mononucleotide containing a fully masked 3′-phosphotriester group **12** (Fig. 5). Since

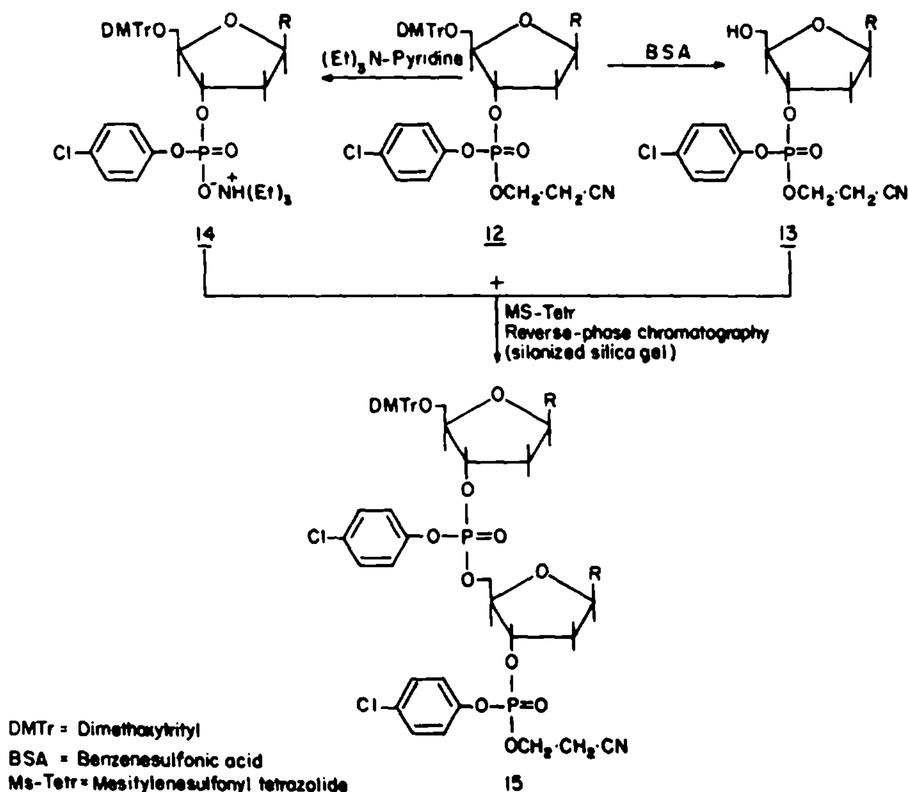

Fig. 5. Modified phosphotriester approach.

the resulting intermediate oligonucleotides contain a fully masked 3'-phosphate group **15**, the necessity for a phosphorylation step at each condensation stage is elimated thus simplifying the approach.

Such a starting material **12** was prepared by treating 5'-dimethoxytrityl N-acyl deoxymononucleoside with *p*-chlorophenyl phosphoryl ditriazolides[20] **16** (Fig. 6) followed by cyano-ethylation reaction. The

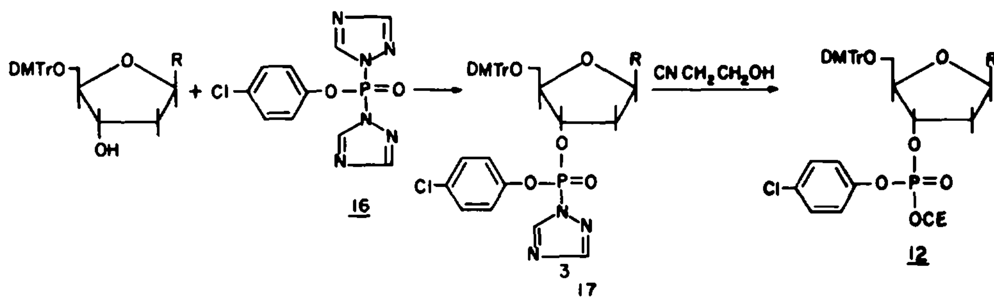

Fig. 6. Preparation of fully protected mononucleotide.

fully protected monomer is easily purified and permits chain elongation from either end. At each coupling step the *β*-cyanoethyl group was removed selectively on treatment with triethylamine-anhydrous pyridine[21,22] or diisopropylamine-anhydrous pyridine.[23] Under these conditions all the other base labile (N-acyl and *p*-chlorophenyl) protecting groups were intact. Recently the above starting material without the *β*-cyanoethyl group as a barium[24] or triethyl-ammonium[25] salt has also been directly used in the coupling reaction. In practice, longer-sized chains are built by block additions starting from the dinucleotide unit. After each cycle the desired fully protected product can be purified on a short column, medium and high pressure techniques, reversed phase technique on RP-2[26] and a simple but very effective chromatography on deactivated silica gel with an aqueous solvent system such as acetone–water–ethyl acetate.[27]

(iv) *Phosphite-triester approach*

By taking advantage of the extreme reactivity of phosphite reagents, Letsinger[28] introduced a phosphite triester approach to decrease appreciably the phosphorylation/condensation time period in building longer chains. Thus two units can be joined together in a matter of minutes rather than hours and the internucleotidic trivalent P-bond oxidized easily to the required phosphotriester by aqueous iodine (Fig. 7). This approach requires that a phosphorus be introduced at each cycle and that limits single base addition. The phosphite reagent could be dichloro-[29] or its ditetrazole[30]-derivatives.

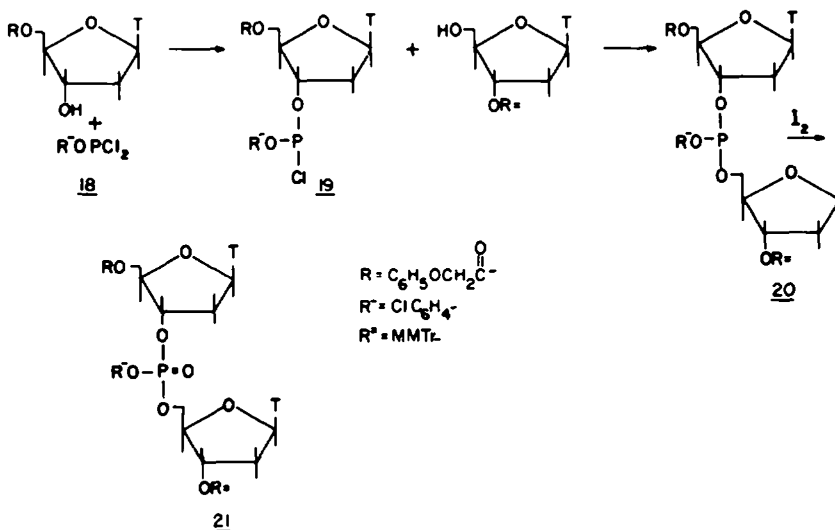

Fig. 7. Phosphite-triester approach.

(v) *Polymer support synthesis*

In analogy with peptide synthesis much effort has been extended to perfect a polymer supported oligonucleotide synthesis. Using a diester approach, two major obstacles were incompatibility of the polymers with the necessary solvents and, perhaps more importantly, yields in the phosphorylation/condensation stage of ~80% or less.[31] This of course means the creation of increasing amounts of impurities as the chain grows. Some of these problems have been overcome by applying the phosphotriester[9,32,33] (Fig. 8) or phosphite[34,30] triester methodologies on new polymers and by blocking off any unreacted chains after each cyle. Recently faster synthesis of oligonucleotide up to 31 units[35] has been accomplished by block condensation on polymer support. The preparation of these blocks were however carried out using solution phosphotriester method. Although the yields of each coupling reaction are not quantitative, the final compound could still be isolated by hplc or preparative gel electrophoresis.

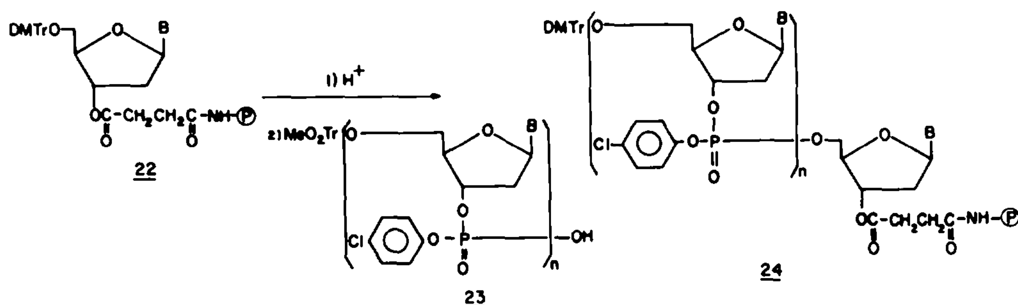

Fig. 8. Phosphotriester-polymer support synthesis.

Protecting groups

(i) *Primary hydroxyl.* The mono- and dimethoxytrityl continue to be the most popular acid labile protecting groups for the 5′-OH. Aromatic sulfonic acids,[36] trifluoroacetic acid[32] and/or zinc bromide[30] have been shown to be milder and more selective deblocking agents than aqueous acetic acid. The pixyl group[37] has been suggested as a useful variant with a lability similar to the dimethoxytrityl group.

Several examples of base labile groups have been reported especially for use in oligoribonucleotide work where an acid labile group has been used for protection of the 2′-OH. These are usually bulky ester groups such as the trityloxyacetyl or substituted phenoxyacetyl.[38] The laevulinyl group, selectively removed by hydrazine, has also been used at this position.[39]

(ii) *Secondary hydroxyls.* Protection of the 3′-OH in oligonucleotide synthesis must in general be complementary to the groups on the 5′-OH and phosphate. The most common base labile groups are acetyl[15] or benzoyl ester[36] while the more popular acid sensitive groups are the tetrahydropyranyl ethers or variations thereof.[40] Silyl ethers and laevulinyl esters have also been used in the deoxy series.

(iii) *Primary amino.* Although evidence has been presented that protection of the exocyclic amino group is not necessary during the coupling reactions for adenosine and guanosine (but definitely for cytosine), use of N-protected monomeric units is almost universally observed. The main reasons have to do with increased solubility but one should be aware of complications in deblocking of other functionalities. Amino protection is evidently not necessary with the phosphite coupling method when used at low temperature.[29] Although a certain number of more sophisticated N-protecting groups have been reported, e.g. tBOC, dimethyl-aminoacetals, by far the most common amino protecting groups are anisoyl for cytosine, benzoyl for adenosine and isobutyryl for guanosine.[8] These represent the best compromises between stability and ease of removal for each of the heterocycles. Removal is usually accomplished by ammonolysis.

(iv) *Phosphate.* With the increasing popularity of the triester method has come a revitalized interest in developing new blocking/deblocking procedures for one or more of the phosphate dissociations. Although its lability can cause complications in extended syntheses, the β-cyanoethyl group is still extensively used[12] especially since the discovery that it can be removed by organic amines[21–23] rather than NaOH aq. Other groups removed via β-elimation mechanism include the tri-chloroethyl ester with a variety of new milder conditions for its removal involving variations of the zinc catalysed method,[41,42] hydrogenation[43] and fluoride ion.[19,44] The original idea of "activation" of the stable β-thioethyl ester group before removal[45] has proven popular in a number of recent variations.[46–48]

Probably the most commonly employed groups requiring "direct" displacement are the substituted

(*o*-Cl, *p*-Cl, *p*-NO_2, etc.) phenyl esters.[10,19,49] Deblocking with ammonium hydroxide or fluoride[19] ion can lead to varying and unwanted amounts of internucleotide bond cleavage.[50] Oximate[51] or thiophenoxide[52,53] nucleophiles have been proposed as solutions to this problem. Groups requiring prior activation before direct displacement include substituted anilidates with isoamyl nitrite[54] and certain amidates by protonation.[55] The 8-hydroxyquinolyl group,[56] selectively removable with Cu ions has been used as a phosphate protecting group in oligonucleotide synthesis. Various alkyl esters,[52] methyl or benzyl, have also been proposed. It is the hope here to break a C–O, rather than a P–O, bond and thereby avoid competing internucleotide bond breakage.

New condensation and phosphorylating reagents

The development by Khorana *et al.* of dicyclohexylcarbodiimide (DCC), mesitylenesulfonyl chloride (MS) and triisopropylbenzenesulfonyl chloride (TPS) as reasonably effective condensing reagents has played a significant role in the synthesis of polynucleotides by the diester method. In the case of the triester synthetic approach, these reagents were found to be unsatisfactory as TPS caused estensive sulfonation whereas DCC does not activate the phosphodiester function; therefore a search for new condensing reagents was initiated by our group in 1974. We speculated on overcoming this problem by using another arylsulfonyl derivative with a better or less innocuous leaving group. A report was made of *p*-toluenesulfonyl-imidazole[57] as a condensing reagent but its rate of coupling was very slow.

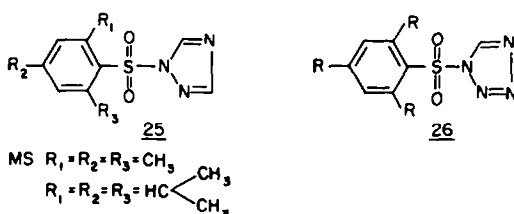

Fig. 9. New coupling reagents.

However, N-acetyltriazole has been reported to be hydrolyzed six times faster at room temperature than N-acetyl imidazole.[58] This observation prompted us to investigate the reactivity of various arylsulfonyl triazoles[59] **25** and tetrazoles[60] **26** derivatives. The various triazoles and tetrazole derivatives of arylsulfonyl were prepared in almost quantitative yields by condensing the corresponding arylsulfonyl chloride in the presence of triethylamine. The rate of coupling reaction was higher, cleaner and a faster reaction mixture, with a minimum of sulfonation reaction, as compared with previously used arylsulfonyl chloride.

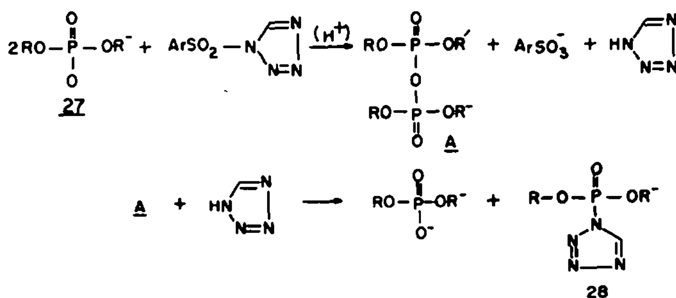

Fig. 10. Mechanism of phosphotriester linkage formation.

Knorre[61] *et al.* studied the mechanism and postulated the formation of phosphodiester tetrazolide **28** as reactive intermediate which participates in the phosphorylation step with alcohol. It seems reasonable to suggest that tetrazole liberated in the first stage reacts with diester component giving rist to the reactive intermediate. The latter should be a powerful phosphorylating reagent due to strong electronegativity of the tetrazole moiety.

Triazole as phosphate activator

Phosphorylation is one of the most crucial steps in the synthesis of polynucleotides by the triester approach. We developed *bis*(triazolyl) *p*-chlorophenylphosphate[20] **16** as a powerful phosphorylating reagent which is apparently a doubly activated reagent with triazole but will give only a diesterified product; thus the formation of $3' \rightarrow 3'$ and $5' \rightarrow 5'$ dimer formation is avoided.

General method of oligonucleotide synthesis by the modified phosphotriester approach

From a practical point of view, the availability of all sixteen dinucleotide blocks enabled a rapid build-up of any deoxyribonucleotide sequence. The basic strategy of a synthetic plan is to extend the chain from 3'- towards 5'-ends by block condensation in the presence of arylsulfonyl tetrazole[36] as a coupling reagent (Fig. 11). The reaction is generally over within 30 min; after work-up a pure fully protected product was isolated by a short column medium pressure chromatography on RP-2 adsorbent or a simple but very effective chromatography on deactivated silica gel with aqueous solvent systems such as acetone–water–ethyl acetate. The homogeneity and the determination of the molecular weight of fully protected ologomer has been determined by the Californium-252 plasma desorption mass spectrometric technique[63] (*vide infra*).

Final deblocking steps

The final and complete removal of all protecting groups from the fully protected oligomer to yield a compound containing natural $3' \rightarrow 5'$ phosphodiester linkages in the most critical step in the success of the triester approach which has introduced one additional complication. Because of the complementary nature of the various protecting groups, several sequential steps must often be employed. Not only most conditions for removal of one group be compatible with other functions: for example, (i) acid for detrilylation can cause depurination, (ii) Zn catalyzed conditions for removing trichloroethyl group can cause problems with cytidine base, (iii) condition to remove the triester phosphate blocking group can cause chain cleavages. The order of deblocking, i.e. removal of the 5'-trityl group followed by removal of base labile or in the reverse order, is also a concern which can cause rearrangement of the nucleotide sequence. The best compromise to date appears to be very specific two-step ammonia treatment[27] or oximate ion[51] followed by acid treatment.

The final compounds are then isolated by any of the fast selective procedures available. These include tlc on polyethyleneimine cellulose (PEI) plates; slab preparative gel electrophoresis and hplc on Permaphase AAX or Partisil 10Sax column.

Nature of side products

During the course of synthesis, the formation of various side products has been observed when a large excess of condensing reagents, arylsulfonyl tetrazole or 3'-nitro-1, 2, 4-triazole derivatives in the presence of large excess phosphodiester component were used for a prolonged period of time.[64] The nature of these side reactions is the incorporation of tetrazole moiety in uracil, thymidine and guanosine bases which is apparently reversed during the deblocking of oligonucleotides with *syn*-4-nitrobenzaldoximate ion. Similar side products were also formed when *p*-chlorophenyl phosphoryl ditetrazolide[23] was used as a phosphorylating reagent. In practice, it is advisable to use a slight excess of the coupling reagent for 30 min in order to avoid any modification of the heterocyclic bases.

Sequence analysis of fully protected oligomers by mass spectrometry

A mass spectrometric method for determing the sequence and molecular weight of protected oligonucleotides containing phosphotriester groups has been developed using Californium-252 plasma desorption techniques.[63] It has been demonstrated that ^{252}Cf PDMS can produce positive and negative molecular ions and fragment ions of large organic molecules. The ionization process and subsequent desorption of the ions depends upon the interaction of nuclear fission fragments from Californium-252 in thin solid films of these compounds. The fast chemical reaction that occurs in the solid film upon excitation by fission fragments leads to the formation of ions that are desorbed by ensuing shock waves. The negative ion mass spectra are characterized by a nested set of fragment ions extending from the 3'- or 5'-terminal nucleotides to the opposite terminal nucleotide, thereby identifiying the sequence.[65]

Sequence analysis of unprotected oligomers containing $3' \rightarrow 5'$ phosphodiester linkages

The development of rapid and unambiguous methods of sequence determination of DNA and RNA is

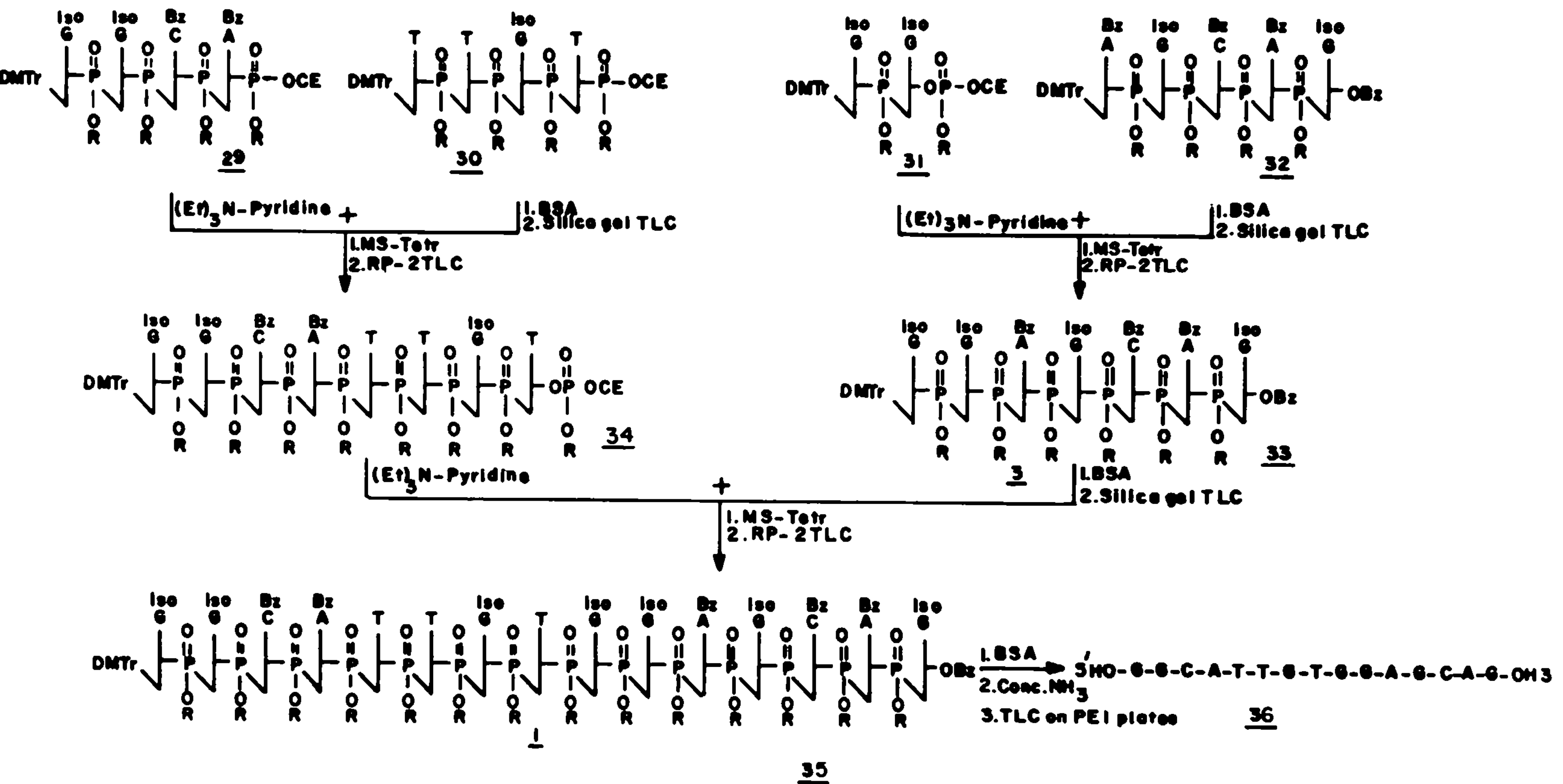

Fig. 11. General approach of deoxyribooligonucleotide synthesis by the modified phosphotriester approach.

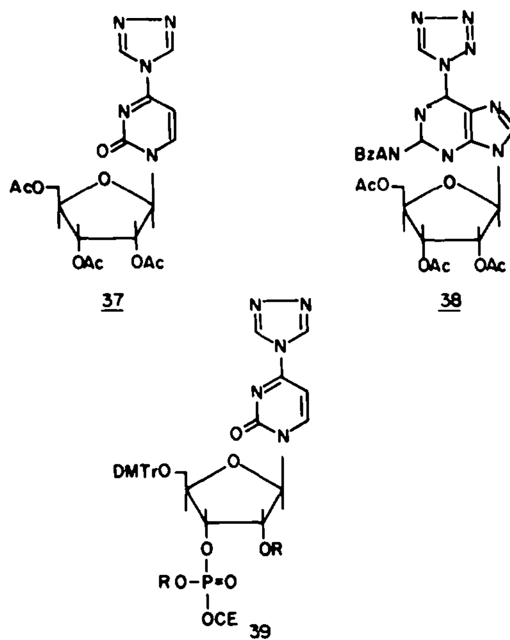

Fig. 12. Formation of the side products.

one of the great breakthroughs in molecular biology of the seventies. The techniques are so simple that DNA sequencing is now used as an indirect method of peptide sequencing. Obviously these procedures are of critical importance for synthetic oligomers as well as natural material. The three most important methods are outlined below.

(i) *Mobility-shift procedure.* The sequence determination of synthetic oligomers up to 20 units long can be carried out by the mobility-shift method (Wandering spot) developed by Sanger[66] and Wu.[67] It involves labelling of the 5′-end of a synthetic compound with T_4-polynucleotide kinase and [γ-^{32}P]-ATP. The labelled oligomer is partially digested by snake phosphodiesterase to produce labelled sequential degradation products down to mononucleotides. These sequential partial degradation products can be fractionated two dimensionally, first by cellulose acetate electrophoresis at pH 3.5, then by homochromatographic techniques on DEAE-cellulose. The sequence of an oligomer is determined by the characteristic mobility shifts of the labelled degradation products after exposing an audioradiogram (Fig. 13).

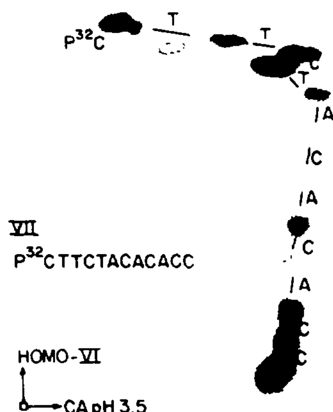

Fig. 13. Fingerprint patterns by mobility-shift method.

(ii) *Base-specific chemical cleavage.* Maxam and Gilbert[68] developed a base-specific chemical cleavage method in which a terminal labelled DNA is partially cleaved at each of the four bases in four different reactions. The products were fractionated according to their size on a slab gel, and then reads

the sequence from an autoradiogram by simply noting which base-specific agent cleaved at each successive nucleotide along the strand, the so-called "Ladder technique". Each chemical scheme, which cleaves at one or two of the four bases, involves three consecutive steps: modification of a base, removal of the modified base from its sugar and DNA strand scission at that sugar. The technique can sequence chains up to five hundred or more units in length in one run (see as example Fig. 14).

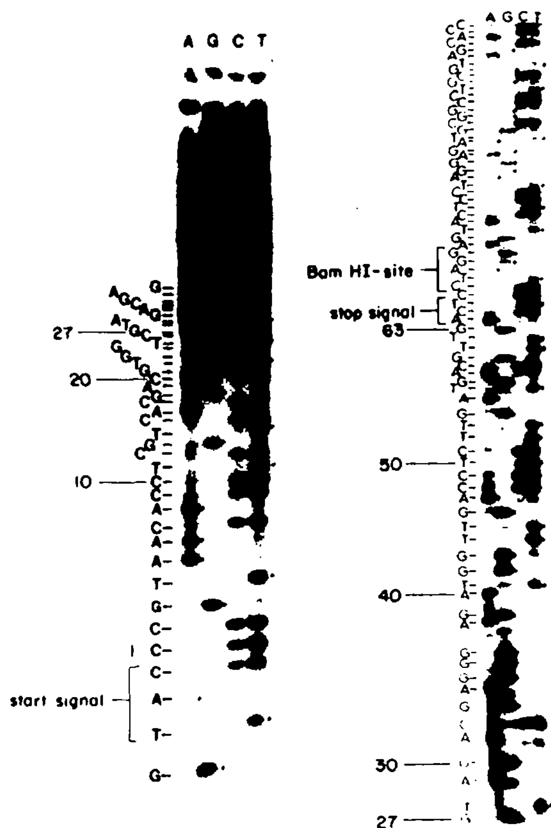

Fig. 14. Gel pattern of cloned insulin chain A by Maxam and Gilbert sequence method.

(iii) *Dideoxy method.* This method makes use of the ability of DNA polymerases to synthesize accurately a complementary radioactive copy of a single-stranded DNA template using DNA fragments as primers.[69] The synthesis is carried out in the presence of the four deoxyribonucleotide triphosphates, one or more of which is α-^{32}P-labelled, and in turn with each dideoxy or arabinose nucleoside triphosphate in separate incubations. There is, therefore, in each reaction a base-specific partial incorporation of a termination analog onto the 3′-ends of the extending DNA strands throughout the sequence. Partial fractionation by gel electrophoresis of the size range of terminated labelled transcripts from each reaction, each with the common 5′-end of the primer, allows a sequence to be deduced from the gel electrophoretic fingerprint.

ENZYMATIC SYNTHESIS

Synthesis of double-stranded DNA

Once the short segments comprising the DNA sequence of both the strands are synthesized, the next task is aligning them in proper order and linking them covalently. A strategy has been developed mainly by Khorana[70] and his group whereby short synthetic oligonucleotides are joined by DNA ligase[71] enzyme, in the presence of a complementary strand or template (Fig. 15A). This enzyme obtained from both uninfected and T_4-phage infected *Escherichia coli* catalyzes the covalent joining of the two molecules of deoxyribooligonucleotides or DNA between the 5′-phosphate from one molecule to the 3′-OH of another to form a phosphodiester bond in the presence of a third molecule which acts as a

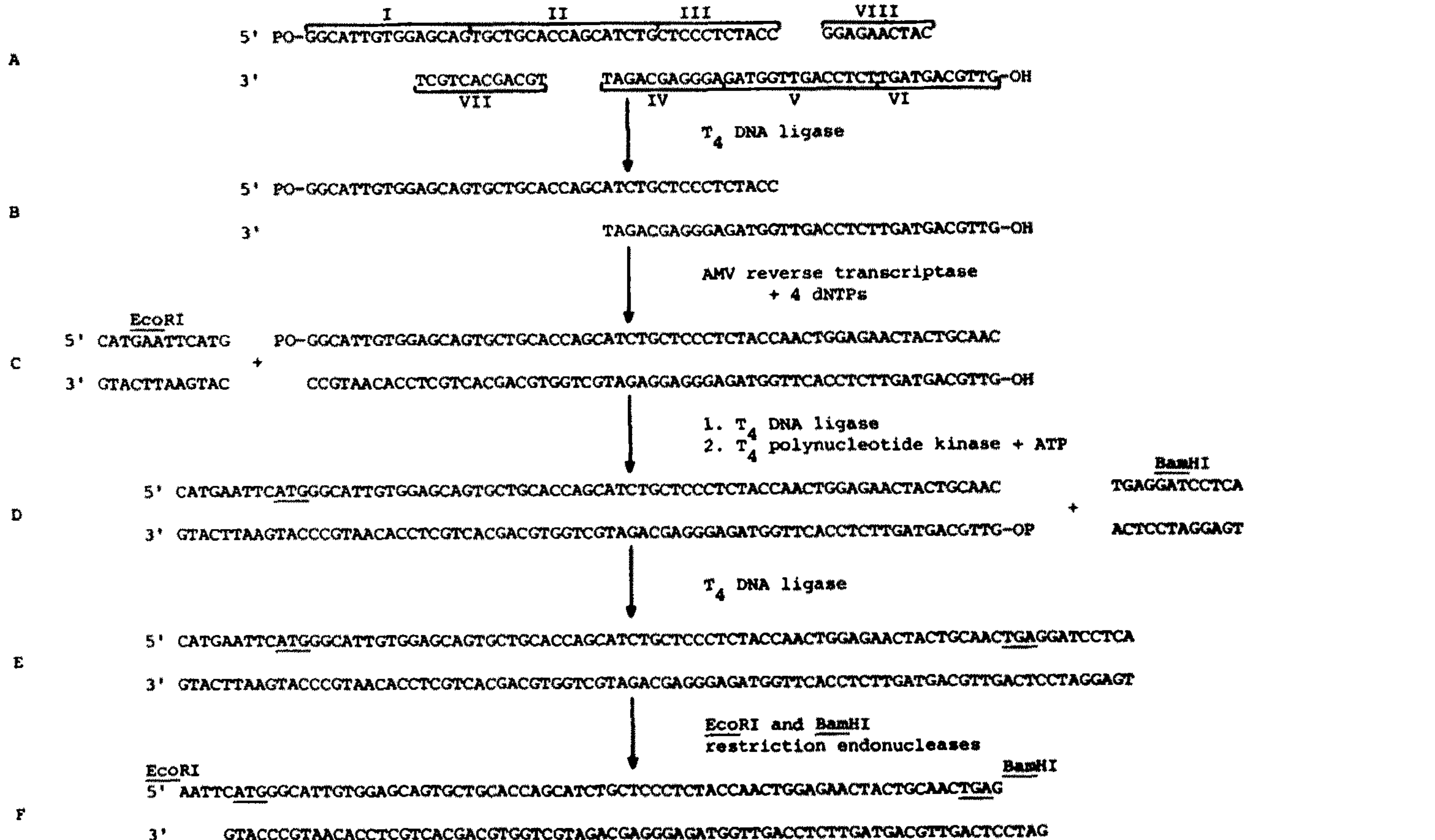

Fig. 15. A, DNA-ligase; B, DNA-synthesis by repairs; C, blunt-end ligation.

complementary strand near the junction of linkage. A minimum of four overlapping nucleotide pairs are required on each side of the junction to provide sufficient template interaction for the joining of deoxyribooligonucleotides with DNA ligase enzyme. This enzyme can also join fragments containing mismatched base pair[72] and therefore it is important to sequence the product of each enzymatic joining reaction. This method has been used to construct a double-stranded DNA molecule that includes the *E. coli* precursor tyrosine tRNA gene[8] together with its promoter, somatostatin,[73] human insulin,[74] human proinsulin,[75] leucocyte interferon[76] and partial sequence of human growth hormone.[62]

Enzymatic repair of partial duplex

An alternative to this approach is the synthesis of a partial strand (upper and lower) by DNA ligase reaction. The resulting single strands are purified by gel electrophoresis under denaturing conditions. Since the 3′-end of the upper strand is complementary to the 3′-end of the lower strand by an overlapping stretch of nucleotide, a partial duplex is formed. With the single strand region as template the remainder of the duplex was enzymatically synthesized with AMV reverse transcriptase and the four deoxynucleoside triphosphates to yield a complete duplex DNA[77] (Fig. 15B). Using this approach a 63 bases long duplex for human insulin gene A was synthesized and confirmed by its DNA sequence method.[68]

Cloning of synthetic genes

During the last several years, useful techniques have been developed for the *in vitro* joining of the duplex DNA segments to vehicle DNA molecules capable of independent replication in a host cell. The cloning vehicle may be a plasmid DNA, a phase λ DNA or a SV40 DNA which are cleaved asymmetrically at specific sequences by a certain restriction enzyme.[78] After joining the DNA segment to the cloning vehicle, the resulting hybrid DNA (known as chimeras) can be used to transform a suitable cell.[79] The hybrid DNA can then be selected from among the transform cells known as clones, and its expression in terms of DNA replication, transcription or translation studied.

Cloning of a DNA molecule requires three essential steps. First the DNA sequence to be cloned must be prepared with specific unpaired or "sticky" sequences at each end recognized by a particular restriction enzyme. Next, the DNA fragment thus obtained is joined to the cloning vehicle DNA by means of DNA ligase. This "match-up" is directed by the complementary cohesive ends present at the termini of both the DNA fragment and the cloning vehicle. Finally, the hybrid DNA is introduced into a functional living cell and a suitable genetic method is used for selecting the clone containing the specific hybrid DNA (Fig. 16).

(i) *Specific method involving synthetic cohesive ends.* A DNA segment for cloning can be chemically synthesized to include a protruding single-stranded sequence that corresponds to the recognition sequence of a restriction endonuclease. For example, a *lac* operator DNA duplex has been chemically synthesized to include at each end a protruding 5′d(pA–A–T–T) sequence corresponding to part of the recognition sequence of *EcoRI* restriction endonuclease (Fig. 17). A molecule of circular pMB9 plasmid DNA is cut once by an *EcoRI* restriction endonuclease to produce a linear *p*MB9 DNA with a protruding 5′d(pA–A–T–T) sequence at each end. New Watson–Crick H-bonds between the protruding sequences direct the orientation of the two molecules which are joined covalently using T_4-DNA ligase to produce circular hybrid *lac-p*MB9 DNA. This hybrid is capable of transforming competent *E. coli* cells and expressing its biological activity *in vivo.*[80]

(ii) *General method involving linkers.* A more general method has been developed[82] in which chemically synthesized linker oligonucleotide is used to create cohesive ends at the termini of blunt ended DNA molecule. Thus, any double-stranded DNA molecules can be cloned. The principle behind the method is shown in Fig. 18. In the first step, a blunt end DNA molecule is joined end-to-end to a synthetic linker using the blunt end ligation activity of DNA ligase. The resulting molecule, with linker molecules added to each end of the original, is digested by a suitable restriction enzyme, in this case to create the *Hind*III endonuclease cohesive ends, 5′-(pA–G–C–T). Alternatively, a ready-made *Bam*I adaptor[83] can also be used. As no digestion with *Bam*I endonuclease is required to create the cohesive end, this adaptor can be used for cloning DNA molecules that have internal *Bam*I sites. In the final step, this molecule is joined to a DNA vehicle (cut by the same endonuclease) to produce a hybrid DNA.

(iii) *Retrieval of gene from the cloning vehicle.* A general method has been developed to recover an intact gene from the cloning vehicle using synthetic retrieval adaptor[77] as shown in Fig. 19.

The eight extra nucleotides at the 5′-end (left-hand side) of the insulin A-chain gene (which include

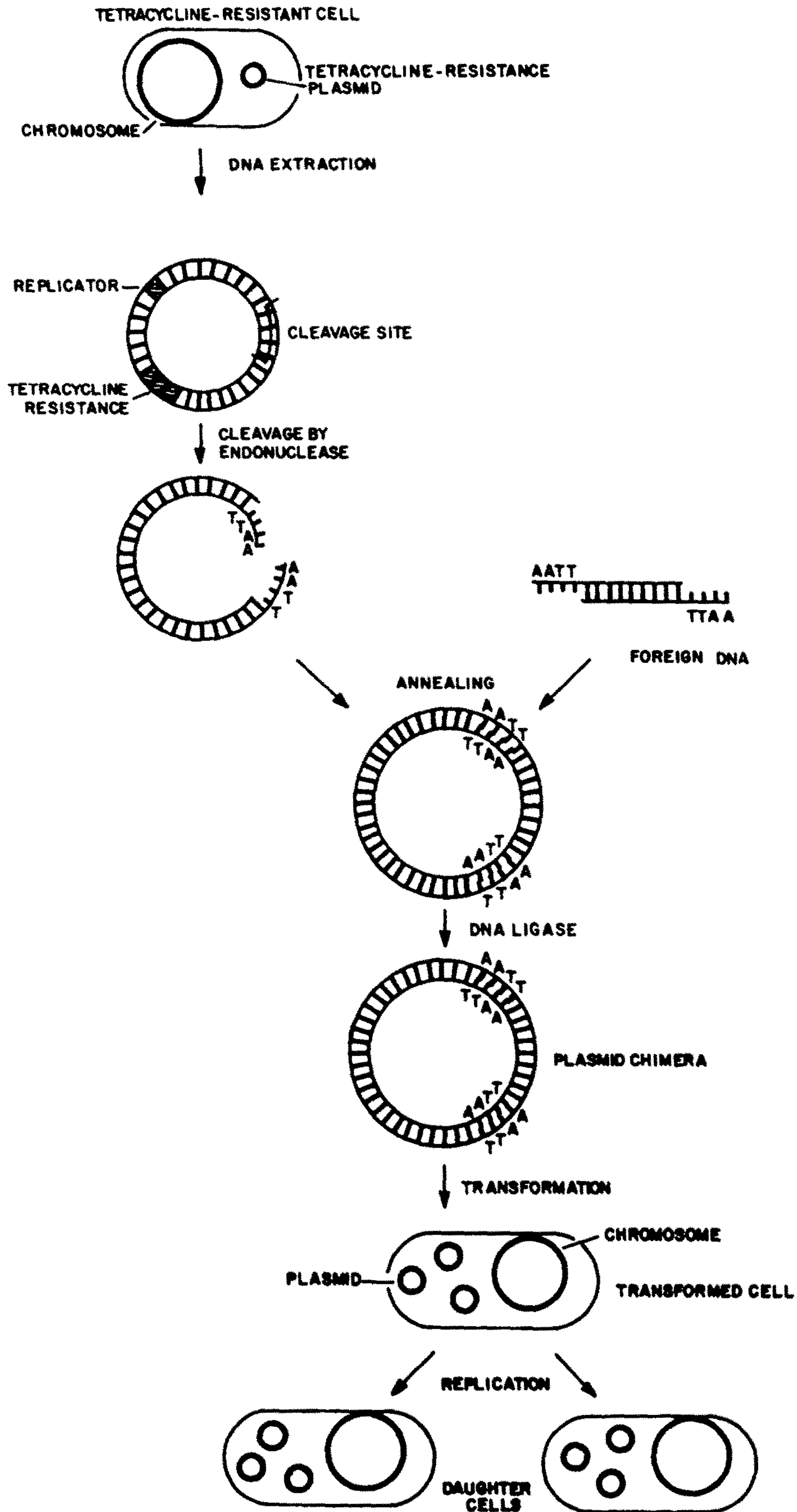

Fig. 16. Recombinant DNA method.

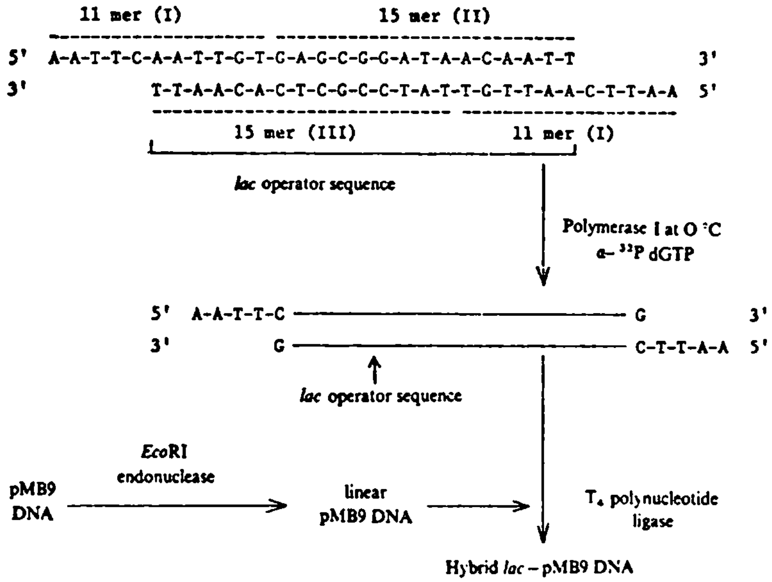

Fig. 17. Cloning of *lac*-operator DNA.

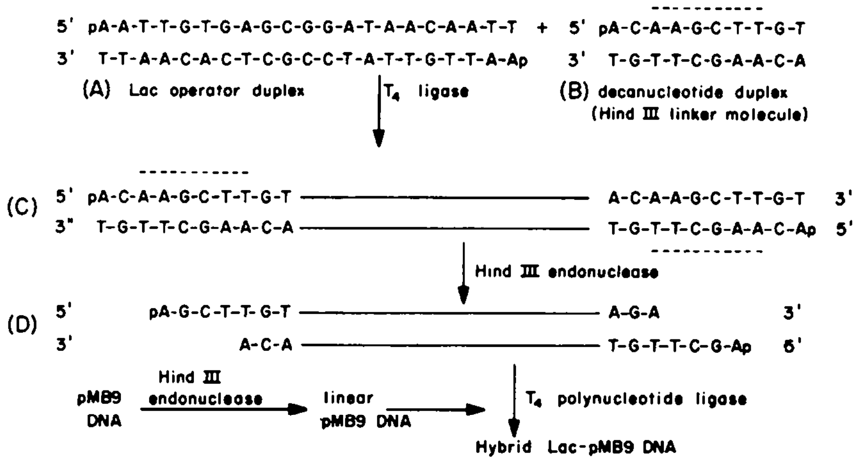

Fig. 18. Cloning of *lac*-operator DNA by blunt-end ligation.

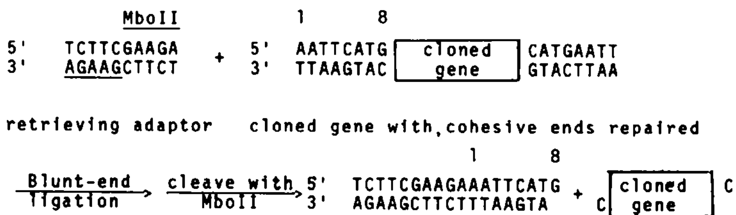

Fig. 19. Retrieval of a gene from a cloning vehicle.

the start signal) have been removed by using a *Mbo*II trimming adaptor. The specific DNA in a cloning vehicle is cut by a specific restriction enzyme and then repaired in the presence of AMV reverse transcriptase. It is next blunt end ligated with a decanucleotide *Mbo*II adaptor. Cleavage of DNA with *Mbo*II restriction enzyme (which cleaves DNA eight nucleotides away from recognition sequence GAAGA) followed by *E. Coli* DNA polymerase I (in the presence of [α-^{32}P] dCTP to insure that the end of DNA is flushed), resulted in the removal of the eight extra base pairs from the A-chain gene. This strategy was used in the recovery of insulin A-chain[77] to be used in the synthesis of proinsulin gene.[75]

BIOLOGICAL ROLE OF SYNTHETIC DEOXYRIBOOLIGONUCLEOTIDES

In the field of molecular biology, the availability of a well-defined sequence of oligomers offers a novel approach in understanding various biological processes at the molecular level. For example, synthesis of all the 64 possible triribonucleotides and DNA-like polymers containing repeating base sequences were successfully used in the elucidation of the genetic code.[3] In recent years, synthetic oligomers have been widely used as exemplified below.

Probe for gene

Synthetic oligomers can be used as specific probes for studies of various aspects of gene structure and function. Synthetic fragments corresponding to portions of several genes have been synthesized, e.g. endolysin,[84] lysozyme,[85] yeast iso-1-cytochrome C,[81,86] gastrin,[87] insulin,[88] and interferons.[89] These oligomers can be used as primers for determining the primary DNA sequences by the primer extension method and also as hybridization probes for the isolation and cloning of single-copy genes in mammalian cells.

Protein–DNA interaction

One of the most basic problems in molecular biology is to understand the mechanism of protein-nucleic acid interactions, i.e. how do the specific amino acids of a protein molecule interact with or recognize a specific nucleotide sequence of the DNA? The sequences of a number of interesting segments of DNA such as operator, promoter, and restriction endonuclease recognition sites have been elucidated in recent years. For studying the point-to-point interaction of nucleic acids with proteins large amounts of the DNA fragments are required. DNA fragments may be isolated following suitable restriction enzyme digestion, incorporation into bacterial plasmids, and amplification. Alternatively, they can be chemically synthesized.[4]

Tools in molecular cloning of DNA

During the last several years, useful techniques have been developed for the *in vitro* joining of DNA fragments to vehicle DNA molecules capable of independent replication. The resulting hybrid DNA can be used to transform a suitable cell. Synthetic linkers[5] or adaptors 8–16 bases long containing various restriction site sequences have been synthesized and successfully used in cloning experiments or in changing a restriction in a plasmid DNA (*vide supra*).

Site-specific mutagenesis

A synthetic oligodeoxynucleotide mismatched at a single nucleotide to a specific complementary site on wild-type circular ϕX174 DNA can be used to produce a defined point mutation. After *in vitro* incorporation into a closed circular duplex DNA by elongation with DNA polymerase and ligation followed by transfection of *E. coli*[90–92] can generate any specific mutation at any given site.

Inhibition of Rous Sarcoma virus replication and cell transformation by a specific oligonucleotide

The synthetic tridecamer which is complementary to 13 bases of the 3′- and 5′-reiterated terminal sequences of *Rous Sarcoma* virus 35S RNA[93] was added to chick embryo fibroblast tissue cultures infected with *Rous Sarcoma* virus. Inhibition of virus production resulted. This is because the synthetic oligomer hybridizes with the terminal reiterated sequences at the 3′- and 5′-ends of the 35S RNA and interferes with one or more steps involved in viral production and cell transformation.

Molecular structure of synthetic DNA fragments

Various self-complement synthetic deoxyoligomers have been crystallized as a left-hand double

helical molecule with Watson–Crick base pairs and an antiparallel organization of the sugar phosphate chains.[94,95] It differs significantly from right-hand B-DNA.

Uptake of synthetic deoxyribooligonucleotide sequence in Haemophilus cells

Only certain DNA fragments are taken up efficiently by competent *Haemophilus* cells, which implies that efficient uptake requires the presence of a specific nucleotide sequence on the incoming DNA. Such a fragment has been identified[96] as 11-mer base pairs in common 5'-AAGTGCGGTCA-3' and its synthesis has been accomplished and has been shown to be biologically active in transporting DNA in the *Haemophilus* cell.[97]

CONCLUDING REMARKS

Progress during the past decade has resulted in the maturation of the triester methodology of chemical synthesis, sophistication in recombinant DNA techniques and rapid DNA-sequence methods. It seems safe to predict that during the eighties the large-scale and economical synthesis of important proteins, e.g. insulin, human growth hormones and interferons, etc, directed by synthetic genes will become a reality. It should even be possible to program microorganisms to make proteins that do not occur naturally in any organism. This approach of synthesis could provide a wealth of new molecules when more is known about the relations between the architecture of proteins and their biological properties. It is also predicted that most of the gene synthetic step will be automated and our knowledge of gene regulation and expression will increase. Ultimately this branch of science may find its best application in solving basic problems such as health, food and energy so essential to mankind.

Acknowledgements—My special thanks are due to Prof. Ray Wu of Cornell University, Ithaca, for the fruitful collaboration in developong the area of DNA synthesis. I am also indebted to my research associates, Drs. C. P. Bahl, R. H. Wightman, K. Itakura, N. Katagiri, T. Hozumi, A. K. Sood, J. Stawinski, R. Brousseau, H. M. Hsiung, W. Sung, F. George, J. Michniewicz, K. J. Marians, R. Sacrpulla for their dedication to this project (NRCC.20748).

REFERENCES

[1]O. T. Avery, C. M. MacLeod and M. McCarty, *J. Exp. Med.* **79**, 137 (1944).
[2]J. D. Watson and F. H. C. Crick, *Nature* **171**, 737 (1953).
[3]H. G. Khorana, H. Büchi, H. Ghosh, N. Gupta, T. M. Jacob, H. Kössel, R. A. Morgan, S. A. Narang, E. Ohtsuka and R. D. Wells, *Cold Spring Harb. Symp. Quant. Biol.* **31**, 39 (1966).
[4]C. P. Bahl, R. Wu, K. Itakura and S. A. Narang, *Proc. Nat. Acad. Sci. USA* **73**, 91 (1976).
[5]C. P. Bahl, K. J. Marians, R. Wu, J. Stawinski and S. A. Narang, *Gene* **1**, 76 (1976).
[6]J. M. Szostak, J. E. Stiles, C. P. Bahl and R. Wu, *Nature* **265**, 61 (1977).
[7]A. M. Michelson and A. R. Todd, *J. Chem. Soc.* 2632 (1955).
[8]H.G. Khorana, *Science* **203**, 614 (1979).
[9]R. L. Letsinger and V. Mahaderan, *J. Am. Chem. Soc.* **87**, 3526 (1965).
[10]C. B. Reese and R. Saffhill, *Chem. Comm.* 767 (1968).
[11]F. Eckstein and I. Risk, Chem. Ber. **102**, 2362 (1969).
[12]K. Itakura, C. P. Bahl, N. Katagiri, J. Michniewicz, R. H. Wightman and S. A. Narang, *Can. J. Chem.* **51**, 3469 (1973).
[13]J. C. Catlin and F. Cramer, *J. Org. Chem.* **38**, 245 (1973).
[14]*Nucleic Acid Synthesis*, Proceedings of an International Symposium at Hamburg, Germany, 3–9 May (1980) published by Nucleic Acids Research, Special Publication No. 7 (September 1980).
[15]P. T. Gilham and H. G. Khorana, *J. Am. Chem. Soc.* **80**, 6212 (1958).
[16]K. L. Agarwal, Y. A. Berlin, H. J. Fritz, M. J. Gait, D. G. Kleid, R. G. Lees, K. E. Norris, B. Ramamoorthy and H. G. Khorana, *Ibid.* **98**, 1065 (1976).
[17]R. A. Jones, H. J. Fritz and H. G. Khorana, *Biochem.* **17**, 1268 (1978).
[18]G. J. Powers, R. L. Jones, G. A. Randall, M. H. Caruthers, J. H. van de Sande and H. G. Khorana, *J. Am. Chem. Soc.* **97**, 875 (1975).
[19]K. Itakura, N. Katagiri, C. P. Bahl, R. H. Wightman and S. A. Narang, *Ibid.* **97**, 7327 (1975).
[20]N. Katagiri, K. Itakura and S. A. Narang, *Ibid.* **97**, 7332 (1975).
[21]A. K. Sood and S. A. Narang, *Nucl. Acids Res.* **4**, 2527 (1977).
[22]R. W. Adamiak, M. Z. Barciszewska, E. Biala, K. Grzeskowiak, R. Kierzek, A. Kraczewski, W. T. Markiewicz and M. Wiewiorowski, *Ibid.* **3**, 3397 (1976).
[23]W. L. Sung and S. A. Narang, *Can. J. Chem.* **60**, 111 (1982).
[24]G. R. Gough, C. K. Singleton, H. L. Weith and P. T. Gilham, *Nucl. Acids Res.* **6**, 155 (1979).
[25]S. S. Jones, R. Raynes, C. B. Reese, A. Ubasawa and M. Ubasawa, *Tetrahedron* **36**, 3075 (1980).
[26]H. Hansen, R. Brousseau, J. Michniewicz and S. A. Narang, *Nucl. Acids Res.* **6**, 1371 (1979).
[27]H. Hansen, W. Sung, R. Brousseau, R. Wu and S. A. Narang, *Ibid.* **8**, 5753 (1980).
[28]R. L. Letsinger and W. B. Lursford, *J. Am. Chem. Soc.* **98**, 3655 (1976).
[29]K. K. Ogilvie, N. Theriault and K. L. Sadana, *Ibid.* **98**, 7741 (1977).

[30]M. D. Mateucci and M. H. Caruthers, *Ibid.* **103**, 3186 (1981).
[31]H. Hayatsu and H. G. Khorana, *Ibid.* **88**, 3182 (1966); F. Cramer, R. Helbig, H. Hettler, K. H. Scheit and H. Seliger, *Angew. Chem. Int. Ed. Engl.* **5**, 601 (1966) and M. J. Gait and R. C. Sheppard, *J. Am. Chem. Soc.* **98**, 8514 (1976).
[32]M. J. Gait, M. Singh, R. C. Sheppard, M. D. Edge, A. R. Greene, G. R. Healthcliffe, T. C. Atkinson, C. R. Newton and A. F. Markham, *Nucl. Acids Res.* **8**, 1081 (1980).
[33]K. Miyoshi and K. Itakura, *Tetrahedron Letters* **28**, 2449 (1979).
[34]K. K. Ogilvie and M. J. Nemer, *Tetrahedron Letters* **29**, 4159 (1980).
[35]P. Dembek, Ken-ichi Miyoshi and K. Itakura, *J. Am. Chem. Soc.* **103**, 706 (1981).
[36]J. Stawinski, T. Hozumi, S. A. Narang, C. P. Bahl and R. Wu, *Nucl. Acids Res.* **4**, 353 (1977).
[37]J. B. Chattopadhya and C. B. Reese, *Chem. Comm.* 639 (1978).
[38]R. W. Adamiak, R. Arantzen and C. B. Reese, *Tetrahedron Letters* 1431 (1977).
[39]J. H. van Boom and P. M. J. Burgers, *React. Trav. Chim. Pays-Bas*, **97**, 73 (1976).
[40]J. H. van Boom, P. H. van Deusen, J. Meewose and C. B. Reese, *Chem. Comm.* 766 (1972).
[41]E. Ohtsuka, H. Tsuji, T. Miyake and M. Ikehara, *Chem. Pharm. Bull.* Japan **25**, 2844 (1977).
[42]J. H. van Boom and P. M. J. Burgers, *React. Trav. Chim. Pays-Bas* **97**, 73 (1978).
[43]R. W. Adamiak, E. Biala, M. Grzeskowiak, R. Kierzek, A. Kraszewski, W. T. Markiewicz, J. Okupniak, J. Stawinski and M. Wiewiorowski, *Nucl. Acids Res.* **5**, 1889 (1978).
[44]K. K. Ogilvie and S. L. Beaucage, *Ibid.* **7**, 805 (1979).
[45]R. H. Wightman, S. A. Narang and K. Itakura, *Can. J. Chem.* **50**, 456 (1972).
[46]K. L. Agarwal, M. Fridkin, E. Jay and H. G. Khorana, *J. Am. Chem. Soc.* **95**, 2020 (1973).
[47]J. Smrt, Coll. *Czech. Chem. Comm.* **39**, 972 (1974).
[48]M. J. Gait and R. C. Sheppard, *Nucl. Acids Res.* **4**, 1139 (1977).
[49]W. S. Zielinski and Z. Lesnikowski, *Snythesis* 185 (1976).
[50]J. F. M. De Rooij, G. Wille-Hazeleger, P. M. J. Burgers and J. H. van Boom, *Nucl. Acids Res.* **6**, 4443 (1978).
[51]C. B. Reese and L. Yan, *Tetrahedron Letters* 4443 (1978).
[52]G. W. Daub and E. E. van Tammlen, *J. Am. Chem. Soc.* **99**, 3526 (1977).
[53]C. B. Reese and Y. T. Yan Kui, *Chem. Comm.* 802 (1977).
[54]E. Ohtsuka, T. Tanaka and M. Ikehara, *Nucl. Acids Res.* **7**, 1283.
[55]J. H. van Boom, P. M. J. Burgers, R. Crea, W. C. M. Luyten, A. B. J. Vink and C. B. Reese, *Tetrahedron* **31**, 3953 (1975).
[56]H. Takaku, M. Sato and T. Hata, *Chem. Comm.* 190 (1977).
[57]Yu. A. Berlin, O. G. Chakhmakhcheva, V. A. Efimov, M. N. Kolosov and Y. G. Kovobko, *Tetrahedron Letters*, 1353 (1973).
[58]H. A. Staab, *Angew. Chem.* Int. Ed. Engl. **1**, 350 (1962).
[59]N. Katagiri, K. Itakura and S. A. Narang, *Chem. Comm.* 326 (1974).
[60]J. Stawinski, T. Hozumi and S. A. Narang, *Can. J. Chem.* **54**, 670 (1976).
[61]D. G. Knorre and V. F. Zarytova, *Phosphorus Chemistry Directed Towards Biology* (Edited W. J. Stec), p. 13. Pergamon Press, Oxford (1980).
[62]D. V. Goeddel, H. L. Hayneker, R. Arentzen, K. Itakura, D. G. Yansura, M. J. Ross, G. Miozzari, R. Crea and P. H. Seeburg, *Nature* **281**, 544 (1979).
[63]C. J. McNeal, S. A. Narang, R. D. Macfarlane, H. M. Hsiung and R. Brousseau, *Proc. Nat. Acad. Sci. U.S.A.* **77**, 735 (1980).
[64]C. B. Reese and A. Ubasawa, *Nucl. Acids Res.* Symp. Ser. No. **7**, 5 (1980).
[65]C. J. McNeal and R. D. Macfarlane, *J. Am. Chem. Soc.* **103**, 1609 (1981).
[66]F. Sanger, J. E. Donelson, A. R. Coulson, H. Kössel and D. Fischer, *Proc. Nat. Acad. Sci. U.S.A.* **70**, 1209 (1973).
[67]E. Jay, A. Bambara, R. Padmanabhan and R. Wu, *Nucl. Acids Res.* **1**, 331 (1974).
[68]A. M. Maxam and W. Gilbert, *Proc. Nat. Acad. Sci. U.S.A.* **74**, 560 (1977).
[69]F. Sanger and A. R. Coulson, *J. Mol. Biol.* **94**, 441 (1975).
[70]H. G. Khorana, K. L. Aggarwal, P. Besmer, H. Büchi, M. H. Caruthers, P. J. Cashion, M. Fridkin, E. Jay, K. Kleppe, R. Kleppe, A. Kumar, P. C. Loewen, R. C. Miller, K. Minamoto, A. Panet, U. L. RajBhandary, B. Ramamoorthy, T. Sekiya, T. Takeya and J. H. Van de Sande, *J. Biol. Chem.* **251**, 565 (1976).
[71]M. Gellert, *Proc. Nat. Acad. Sci. U.S.A.* **57**, 148 (1967); B. Weiss and C. C. Richardson, *Ibid.* **57**, 1201 (1967); B. M. Olivera and I. R. Lehman, *Ibid.* **57**, 1426 (1967).
[72]C. M. Tsiapalis and S. A. Narang, *Biochem. Biophys. Res. Comm.* **39**, 631 (1970).
[73]K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heynekar, F. Bolivar and H. W. Boyer, *Science* **198**, 1056 (1977).
[74]D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heynekar, D. C. Yansura, R. Crea, T. Hirose, A. Krazewski, K. Itakura and A. D. Riggs, *Proc. Nat. Acad. Sci. U.S.A.* **76**, 106 (1979).
[75]R. Brousseau, R. Scarpulla, W. Sung, H. M. Hsiung, S. A. Narang and R. Wu, *Gene* **17**, 279 (1982).
[76]M. D. Edge, A. R. Greene, G. R. Heathcliffe, P. A. Meacock, W. Schuch, D. B. Scanlon, T. C. Atkinson, C. R. Newton and A. F. Markham, *Nature* **292**, 756 (1981).
[77]R. C. Scarpulla, S. A. Narang and R. Wu, *Anal. Biochem.* **121**, 356 (1982).
[78]R. J. Roberts, C. R. C. *Crit. Rev. Biochem.* **4**, 123 (1976).
[79]S. N. Cohen, C. Y. Chang, H. W. Boyer and R. B. Helling, *Proc. Nat. Acad. Sci. U.S.A.* **70**, 3240 (1973).
[80]K. J. Marians, R. Wu, J. Stawinski, T. Hozumi and S. A. Narang, *Nature* **263**, 744 (1976).
[81]S. Gillam, R. Rottman, P. Jahnke and M. Smith, *Proc. Nat. Acad. Sci. U.S.A.* **74**, 96 (1977).
[82]H. L. Heyneker, J. Shine, H. M. Goodman, H. W. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S. Lin and A. D. Riggs, *Nature* **263**, 748 (1976).
[83]C. P. Bahl, R. Wu, R. Brousseau, A. K. Sood, H. M. Hsiung and S. A. Narang, *Biochem. Biophys. Res. Comm.* **81**, 695 (1978).
[84]R. Wu, R. Tu and R. Padmanabhan, *Ibid.* **55**, 1092 (1973).
[85]S. A. Narang, K. Itakura, C. P. Bahl and Y. Y. Wigfield, *Ibid.* **49**, 445 (1972).
[86]J. W. Szostak, J. E. Stiles, C. P. Bahl and R. Wu, *Nature* **265**, 61 (1977).
[87]B. E. Noyes, M. Mevarech, R. Stein and K. L. Agarwal, *Proc. Nat. Acad. Sci. U.S.A.* **76**, 5036 (1979).
[88]S. J. Chan, B. F. Noyes, K. A. Agarwal and D. F. Steiner, *Ibid.* **76**, 5036 (1979).
[89]M. Houghton, A. G. Stewart, S. M. Doel, J. S. Emtage, M. A. Eaton, J. C. Smith, T. P. Patel, H. M. Lewis, A. G. Porter, A. J. Birch, T. Cartwright and N. H. Carey, *Nucl. Acids Rees.* **8**, 1813 (1980).
[90]A. Razin, T. Hirose, K. Itakura and A. D. Riggs, *Proc. Nat. Acad. Sci. U.S.A.* **75**, 4268 (1978).
[91]S. Gillam and M. Smith, *Gene* **8**, 99 (1979).

[92]O. S. Bhanot, S. A. Khan and R. W. Chambers, *J. Biol. Chem.* **254**, 12684 (1979).
[93]P. C. Zemecnik and M. L. Stephenson, *Proc. Nat. Acad. Sci. U.S.A.* **75**, 280 (1978).
[94]A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature* **282**, 680 (1979).
[95]H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Nat. Acad. Sci. U.S.A.* **78**, 2179 (1981).
[96]D. B. Danner, R. A. Deich, K. L. Sisco and H. O. Smith, *Gene* **11**, 311 (1980).
[97]D. B. Danner, H. O. Smith and S. A. Narang, *Proc. Nat. Acad. Sci. U.S.A.* **79**, 2393 (1982).