CHEMICAL SYNTHESIS OF DNA CONTAINING MODIFIED BASES BY POST-SYNTHETIC SUBSTITUTION AND APPLICATION OF MODIFIED DNA TO THE STUDY OF PROTEIN-DNA INTERACTION

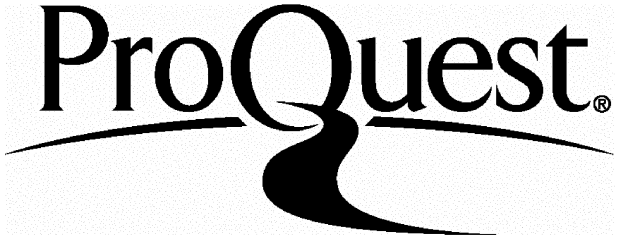
by

QINGUO ZHENG

A thesis submitted for the degree of Doctor of Philosophy in the University of London

Department of Biochemistry and Molecular Biology
University College London

May 1995



To my wife and my daughter

ABSTRACT

Because of the limitations of most methods for modified oligonucleotide synthesis, an alternative strategy (post-synthetic substitution strategy) was developed for synthesis of DNA with modification on the 4-position of thymine and the 6-position of guanine. This was achieved by incorporation of versatile nucleotides into DNA which contain a leaving group which is sufficiently stable to withstand the conditions of DNA synthesis but can be substituted by nucleophiles, after synthesis, to produce a series of oligomers each containing a different modified base.

4-Triazolothymidine phosphoramidite was prepared and inserted into the dodecanucleotide AGCGAAXTCGCT (X standing for 4-triazolothymine). By treating the oligomer with different nucleophiles the parent oligomer containing thymine (T) and other five oligomers were prepared each containing a different modified base: O^4-methylthymine (T^{OMe}), O^4-ethylthymine (T^{OEt}), 5-methylcytosine (T^{NH2}), N^4-dimethylamino-5- methylcytosine (T^{DH}), or 4-thiothymine (T^{S}).

As a further development of the above method, a fully deprotected and purified versatile oligonucleotide containing 4-phenylthiothymine was prepared with the advantage of making oligonucleotides containing very labile modified thymines. Oligomers containing 5-methyl-N^4, N^4-ethanocytosine or 4-azidothymine were prepared by treating the versatile oligomer with ethyleneimine or sodium azide (NaN_3), respectively.

A post-synthetic approach was also developed for the preparation of oligonucleotides containing guanine modified at the 6-position. The versatile guanine monomer N^2-phenylacetyl-6-(2, 4-dinitrophenyl)thio-2'-deoxyguanosine-3'-phosphoramidite was prepared and incorporated into an oligomer. The oligomer containing this versatile guanine ($G^{S\emptyset}$) was converted into oligomers containing 6-thioguanine (G^{S}), 2, 6-diaminopurine (G^{NH2}), 2-amino-6-methylaminopurine (G^{NMe}), O^6-methylguanine (G^{OMe}), or guanine (G) by treatment with appropriate reagents after synthesis.

The interactions between λ-phage Cro repressor and its operator DNA were investigated by a photochemical cross-linking approach using 21 mer oligonucleotides containing the sequence of the OR3 operator of λ-phage in which one of guanine or thymine was substituted with 6-thioguanine or 4-thiothymine. Upon irradiation with long-wave UV light (~360 nm) the oligonucleotides containing these thiobases cross-linked to the Cro protein with different efficiency, depending on the position of the substitution. The cross-linking results are consistent with the Cro-DNA interaction model and also support the suggestion that cross-linking can occur only on those amino acid residues at the interface of the DNA-protein complexes. However an attempt to identify the amino acid residues covalently linked to DNA was unsuccessful, mainly due to the instability of cross-linked Cro-DNA complexes.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Drs Peter Swann and Yao-Zhong Xu, for their tuition, guidance, and support throughout the course of this project. I am also grateful to the Cancer Research Campaign for their financial support.

I would like to thank Drs Tim Waters and Hwee Boom Tan for their assistance and useful discussions, Mr Raymond Mace for proof reading this manuscript, and all my colleagues for creating a stimulating environment in our group.

Thanks are also due to Dr Brian Coles for the peptide sequencing, Mrs Jill Maxwell for running ^{1}H and ^{31}P NMR spectra, and Dr Alex Drake for measuring the CD spectra.

Finally, I am greatly indebted to my wife and my daughter for their faith and moral support. Without their support it would have been much harder to have completed this course.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	adenine
BSA	bovine serum albumin
8-BrdA	8-bromo-2'-deoxyadenosine
5-BrdU	5-bromo-2'-deoxyuridine
C	cytosine
CD	circular dichroism
CPG	controlled-pore-glass
DMT (Cl)	4, 4'-dimethoxytriphenylmethyl (chloride)
DNA	deoxyribonucleic acid
dA	2'-deoxyadenosine
DBU	1, 8-diazabicyclo (5, 4, 0) undec-7-ene
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
DTT	dithiothreitol
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast protein liquid chromatography
G	guanine
G^{NH2}	2, 6-diaminopurine
G^{NMe}	2-amino-6-methylaminopurine
G^{OMe}	O^6-methylguanine
G^{S}	6-thioguanine
$G^{S\phi}$	6-(2, 4-dinitrophenyl)thioguanine
HPLC	high performance liquid chromatography
HPTLC	high perfomance thin layer chromatography
5-IdU	5-iodo-2'-deoxyuridine
IPTG	isopropylthiogalactoside

Ms-Cl	mesitylenesulfonyl chloride
oxime	E-2-nitrobenzaldoxime
PAC	phenoxyacetyl
PAGE	polyacryamide gel electrophoresis
Px-Cl	9-chloro-9-phenylxanthene
SAM	S-adenosyl-L-methionine
SDS	sodium dodecyl sulphate
T	thymidine or thymine
T^{DH}	4-(2, 2-dimethyl)hydrazino-5-methylpyrimid-2-one [i.e. N^4-dimethylamino-5-methylcytosine]
T^{e}	5-methyl-N^4, N^4-ethynocytosine
TEMED	N, N, N', N'-tetramethylenediamine
TLC	thin layer chromatography
Tm	melting temperature
TMG	N^1, N^1, N^3, N^3-tetramethylguanidine
T^{N3}	4-azidothymine
T^{NH2}	5-methylcytosine
T^{OEt}	O^4-ethylthymine
T^{OMe}	O^4-methylthymine
TPS-Cl	2, 4, 6-triisopropylbenzenesulfonyl chloride
T^{SH}	4-thiothymine
T^{SPh}	4-phenylthiothymine
T^{Tri}	4-(1, 2, 4-triazolyl)thymine
UV	ultraviolet

CHAPTER 1

INTRODUCTION

1.1. THE NEED FOR MODIFIED OLIGONUCLEOTIDES[1]

Over the last fifteen years, the development of efficient and simple methods for DNA synthesis has led to numerous applications of oligonucleotides. For example, the oligonucleotides have been widely used for cloning and synthesizing genes (Groger *et al.*, 1988), as primers for sequencing DNA and various PCR applications (Arnhelm and Levenson, 1990), as potential therapeutic drugs (Uhlmann and Peyman, 1990; Marshall and Caruthers, 1993), site-directed mutagenesis of genes (Leatherbarrow and Fersht, 1986), examination of nucleic acid-protein interactions (Harrison and Aggarwal, 1990), and for studies on structures of nucleic acids (Kennard and Hunter, 1989). During these studies, people discovered that if synthetic DNA was modified the potential applications could be widened, and even more significantly, some research could be done which otherwise would be impossible to carry out. For example, the therapeutic use of synthetic, unmodified oligonucleotides faces the following problems (Englisch and Gauss, 1991): (a) oligonucleotides do not easily pass through the mainly lipophilic cell membrane since they carry one negative charge per phosphate group; (b) many nucleases rapidly cleave oligonucleotides; (c) the stability of complexes formed between the oligonucleotides and their complementary target is not very high under physiological conditions. However, these problems can be overcome, or at least partially overcome, by use of modified oligonucleotides (Marshall and Caruthers, 1993). Inevitably such discoveries have stimulated the development of the chemistry for synthesizing modified DNA.

1.2. THE STRUCTURE OF DNA

The primary structure of DNA has each nucleoside joined by a phosphodiester

---

[1]Oligonucleotide, nucleotide, and nucleoside are often used to refer to both ribo- and deoxyribo- derivatives. The abbreviations oligonucleotide, nucleotide, and nucleoside used in this thesis refer only to deoxyribo- derivatives.

from its 5'-hydroxyl group to the 3'-hydroxyl group of one neighbour and by a second phosphodiester from its 3'-hydroxyl group to the 5'-hydroxyl group of its other neighbour and so on (Figure 1.1). Unlike RNA which contains a number of modified nucleosides, the only nucleosides in DNA are 2'-deoxyadenosine (dA), 2'-deoxyguanosine (dG), 2'-deoxycytidine (dC), and thymidine (T).

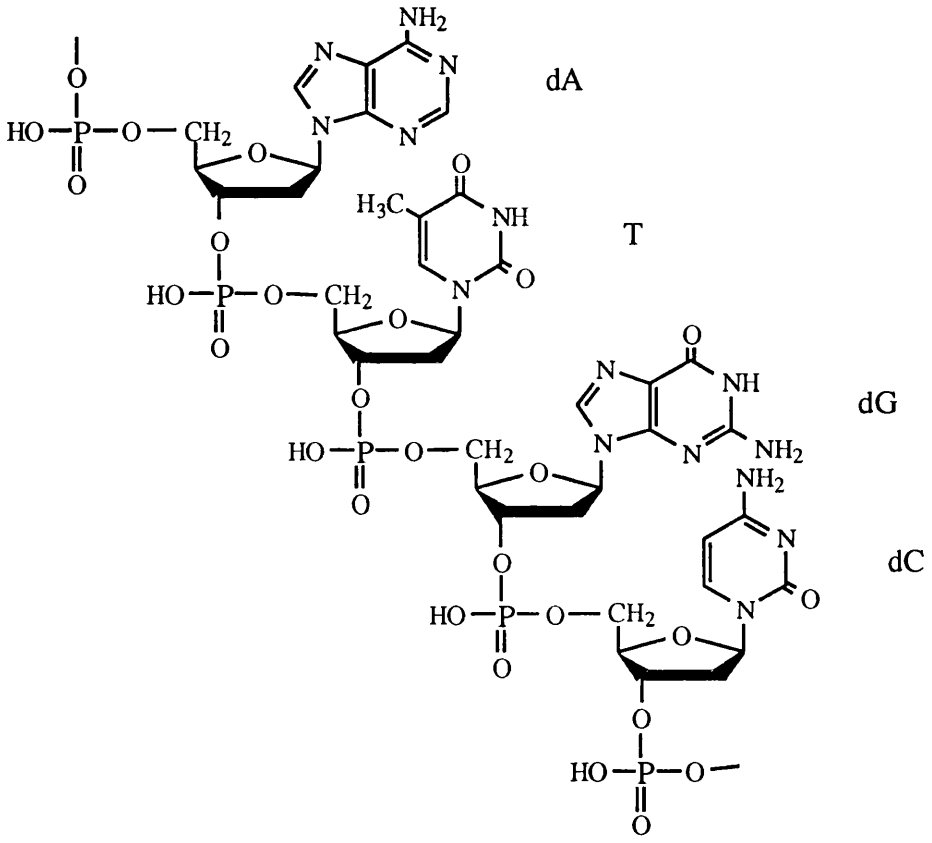
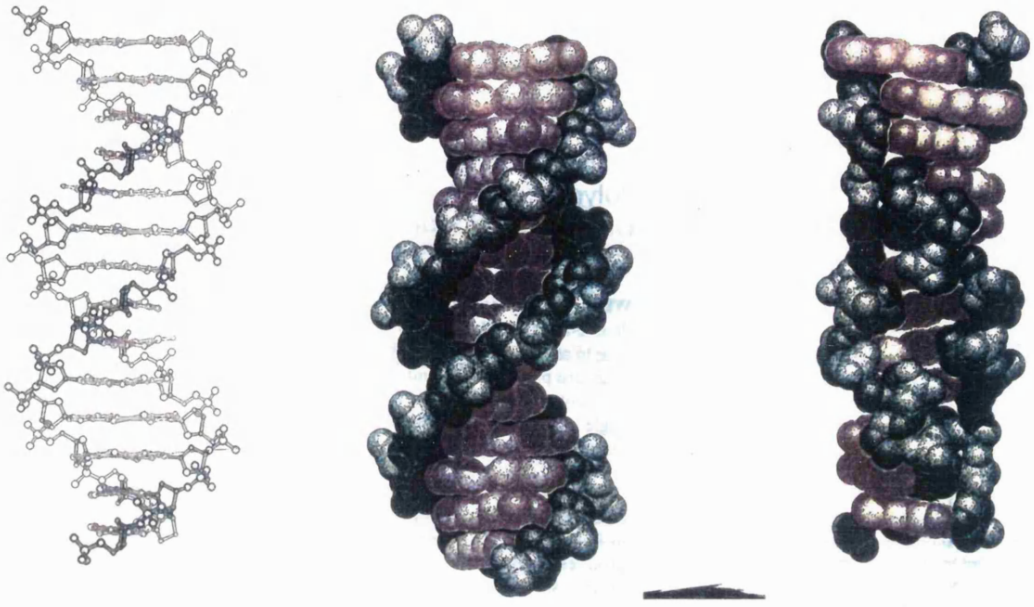


Figure 1.1. The primary structure of DNA showing the four common bases.

The secondary structure of uniform DNA, first proposed by Watson and Crick (1953) based on X-ray diffraction patterns of DNA fibres is shown in Figure 1.2a-b. The essential features of their model are: (a) two polynucleotide chains run in opposite directions and coil around a common axis to form right-handed double helix which are held mainly by hydrogen bonds between G:C and A:T base pairs (Figure 1.3); (b) purine and pyrimidine bases are on the inside of the helix, whereas the phosphate and deoxyribose unit are on the outside of the helix; (c) the diameter of the helix is 20 Å, the

distance of adjacent bases is 3.4 Å along the helix axis and related by a rotation of 36 degree, hence, the helical structure repeats after ten residues on each chain.



(a) (b) (c)

Figure 1.2. The secondary structure of B-DNA in (a) ball and stick representation; (b) space filling representation; (c) structure of Z-DNA in space filling representation [Taken from Saenger (1984)].

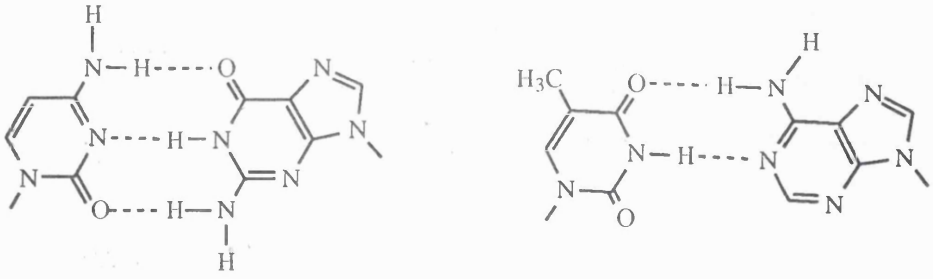


Figure 1.3 Watson-Crick base-pairing for C:G (left) and T:A (right).

When the structure of crystals of short stretches of DNA were solved by X-ray crystallography it became clear that the structure of DNA was far more variable than had first been thought (Dickerson and Drew, 1981). An extreme example was a

completely new form of DNA that was a left-handed helix (Wang *et al.*, 1979; Wang *et al.*, 1981) now known as Z-DNA (Figure 1.2c). Even oligonucleotides that had overall B-DNA structure showed considerable deviations from classical uniform B-DNA at the local level as reviewed recently by Calladine and Drew (1992).

1.3. CHEMICAL SYNTHESIS OF OLIGONUCLEOTIDES

1.3.1. Brief History of DNA Synthesis

The first successful synthesis of a dinucleotide containing a 3'-5' internucleotide linkage identical to natural DNA was achieved by Michelson and Todd (1955). The dinucleotide phosphate was prepared by coupling 5'-O-acetylthymidine-3'-(benzylphosphorochloridate) (1) with 3'-O-acetylthymidine (2) in the presence of 2,6-lutidine to give the fully protected dinucleotide (3). Subsequent removal of the protecting groups gave the dimer TpT (4) (Figure 1.4)

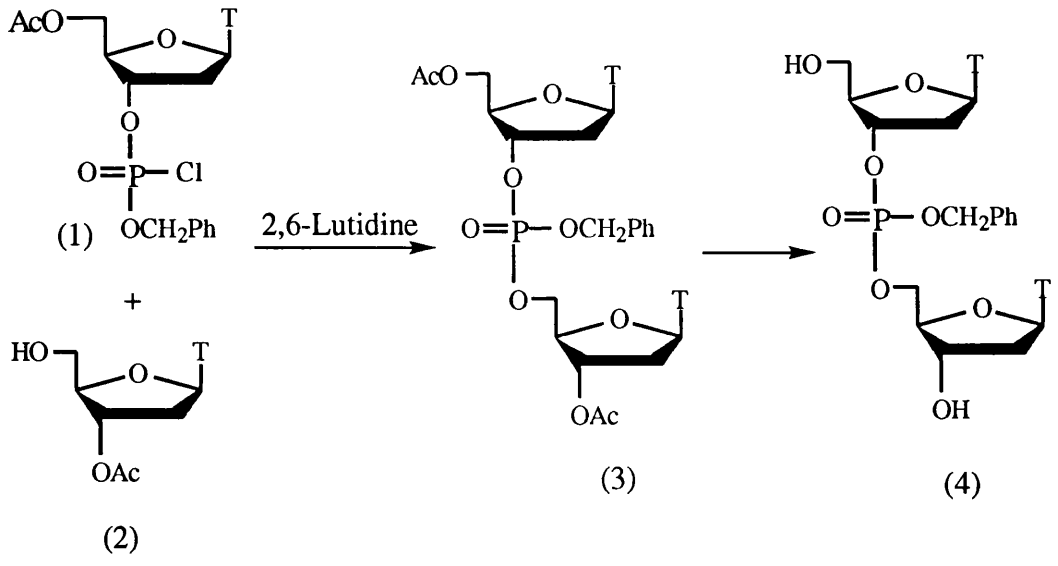


Figure 1.4. The first chemical synthesis of dinucleotide.

Since the overall yield of this method was low because of various side reactions, this approach was not useful in the synthesis of larger oligonucleotides. In the late fifties and early sixties, Khorana and his colleagues (Gilham and Khorana,

1958) successfully established the phosphodiester approach for oligonucleotide synthesis which became the routine for the next two decades. One major disadvantage relating to this approach was the ionic nature of the starting material and the various condensation products which had to be separated by tedious and time consuming ion-exchange chromatography.

To alleviate this inconvenience, the phosphotriester approach introduced by Michelson and Todd (1955) (see Figure 1.4), in which the three phosphate bonds were masked during synthesis, was reinvestigated by Letsinger and Ogilvie (1967; 1969). Through the years, this method was constantly improved by the efforts of many organic chemists and gradually it replaced the phosphodiester chemistry for the synthesis of oligonucleotides Using this method, the first biologically active genetic element, the lac operator DNA, was synthesized and cloned (Bahl *et al.*, 1976). Subsequently, genes encoding other chemically important proteins such as insulin (Goeddel *et al.*, 1979) and interferon (Edge *et al.*, 1981) were synthesized via the same technique.

The use of an organic polymer as a support for the synthesis of oligonucleotides was first investigated by Letsinger and Mahadevan (1966). Subsequently, Koster (1972) introduced inorganic carriers such as silica gel as an alternative to swellable organic supports for the preparation of synthetic oligonucleotides. Although these approaches met with only moderate success, they laid the foundation for the subsequent development of the solid-phase methodology for the synthesis of oligonucleotides.

In the mid-seventies, Letsinger and his coworkers revolutionized the chemical synthesis of oligonucleotides through the development of the phosphite triester methodology using phosphodichloridites to link nucleosides (Letsinger and Lunsford, 1976). The combination of this efficient approach with the resurgence of silica gel as a solid support in the early eighties intensified research activities to automatic oligonucleotide synthesis (Alvarado-Urbina *et al.*, 1981; Matteucci and Caruthers, 1981), but, since nucleoside chlorophosphite intermediates were unstable to moisture and air, synthesis had to be carried out in an inert gas atmosphere at -78°C,

consequently these intermediates were not suitable for automation. To alleviate these problems, Beaucage and Caruthers (1981) developed the deoxyribonucleoside phosphoramidites as a new class of intermediates for the synthesis of oligonucleotides. These intermediates were isolated as stable powders and could be stored for prolonged periods of time. With the further modification and improvement through the years, this so called phosphoramidite approach has been the predominant method for solid-phase synthesis of oligonucleotides and the basis for most commercial automated DNA synthesizers.

1.3.2. Solid-phase Supports for Oligonucleotide Synthesis

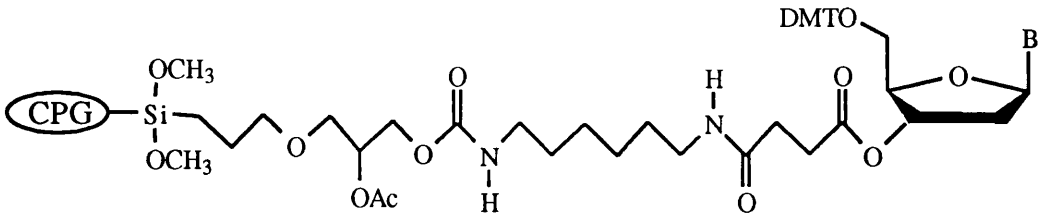
Oligonucleotides, both modified and unmodified, can be made in solution or by solid phase synthesis. DNA synthesis in solution begins with the coupling of one appropriately protected monomer unit with another and separation of products and unreacted starting materials on a column of silica gel. Removal of a protecting group at one or other end gives a dimer block, which is coupled with another block to give a tetramer, and so on, to give larger protected fragments. Each coupling step requires a chromatographic purification and, although a skilled person can do this fairly quickly, adequate resolution of long chains is difficult and the method is heavily labour intensive and time-consuming. However, when large quantities of oligonucleotides are needed (hundreds of milligrams), solution-phase chemistry may be advantageous because reactions are mixed in equimolar amounts to maximize yields and to reduce costs.

Compared with the methods of DNA synthesis in solution, synthesis of oligonucleotides by the solid-phase method has the following great advantages: (a) because an oligonucleotide chain is attached to a solid support, all excess reagents are washed away thus laborious and time-consuming purification steps after each monomer addition are avoided; consequently, the oligonucleotide can be synthesized rapidly; (b) all chemical reactions during an elongation cycle can be pushed to completion by using a large excess of chemicals relative to the support bound oligonucleotide; (c) the process is amenable to automation.

The principle of solid phase synthesis was developed by Merrifield (1962; 1963; 1965) at the Rockefeller Institute. This simple technique was initially applied to the synthesis of polypeptides. Because of this significant development he was awarded the Nobel Prize for chemistry in 1984. Soon after the solid phase method had been shown to be valuable for peptide synthesis, the technique was applied to the synthesis of oligonucleotides (Letsinger and Mahadevan, 1966). Since then numerous polymer supports have been tested for oligonucleotide synthesis, from polystyrene to polytetrafluoroethylene and silica gel. Today the solid-phase synthesis of oligonucleotide, especially on small-scale synthesis, is mainly carried out on specially defined glass beads, so-called controlled pore glass (CPG) support since this material does not swell or contract in various solvents and also possesses mechanical properties that offer distinct advantages as a polymer support for synthesis. More importantly, it gives faster and more efficient coupling than other polymer supports.

The most commonly used CPG supports have pores of 500 Å and 1000 Å. The pore size is of importance during the synthesis of relatively long oligonucleotides. An abrupt termination of chain propagation was observed when the synthesis of oligomers longer than 100 bases was attempted on CPG with a pore size of 500 Å (Efcavitch *et al.*, 1987). This termination was caused by the steric crowding of growing oligomer chains which, presumably, reduced the diffusion of the reagents through the matrix. By contrast, CPG support with a pore size of 1000 Å was satisfactory for the synthesis of large oligonucleotides, and it is now thought that CPG support with a pore size of 1000 Å should be used whenever the oligomers are longer than 50 bases.

Numerous functionalized CPG supports have been reported, among which the "long-chain alkylamine CPG support (LCAA-CPG) (5) is the one which has been most widely and successfully used in DNA synthesis and is commercially available from many companies.



(5)

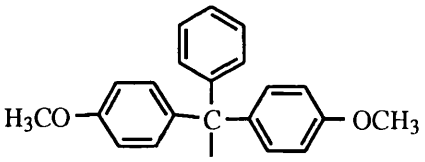
1.3.3. The Protection of the 5'-Hydroxyl Function of Nucleotides

As mentioned above (1.2), an oligonucleotide is a single-stranded chain in which nucleosides are linked by phosphodiester bridges between the 3'-hydroxyl group of one nucleoside and the 5'-hydroxyl group of another. The specific and sequential formation of this 3'-5' phosphodiester linkage is the key step in the synthesis of oligonucleotides. Since a nucleoside contains two hydroxyl groups, one must be chemically protected while the other is phosphorylated (or phosphitylated). For a number of reasons it is preferable for nucleoside to be phosphorylated (or phosphitylated) at the 3'-position. Hence there is a need for temporary protection of the 5'-hydroxyl group.

Protection of the 5'-hydroxyl group with 4, 4'-dimethoxyltrityl (DMT) (6), which was introduced by Khorana et al. (Smith *et al.*, 1962), has become almost standard in today's oligonucleotide synthesis. This group is introduced on to the 5'-position of the N-protected nucleosides by reaction with 4, 4'-dimethoxyltrityl chloride in the presence of a mildly basic catalyst such as pyridine or 4-dimethylaminopyridine. The reaction is regioselective for the primary 5'-hydroxyl group compared to the secondary 3'-hydroxyl function partly because of the bulk of the DMT group. This selective incorporation is one of the main reasons for the popularity of the DMT group for the protection of the 5'-hydroxyl function. The group is removed by treatment with acids such as dichloroacetic or trichloroacetic acid in a non-aqueous solvent.

Besides protecting the 5'-hydroxyl from unwanted reactions, this protecting group also serves other functions. First, since this group is a large lipophilic moiety, it

confers a solubility in organic solvents the unprotected nucleosides would not have, thus making these intermediates much easier to purify by chromatography. Second, since the DMT+ produced on acid-catalyzed removal of the group is intensely coloured the efficiency of the chain elongation step can be evaluated by simply measuring the released carbonium ions by spectrophotometer. Third, since the lipophilic 5'-O-trityl on the terminal base is not removed by the basic conditions used to deprotect synthetic oligomers the desired sequence, which has the 5'-O-trityl protected, can be separated by HPLC easily from the failure sequences generated during the synthesis.



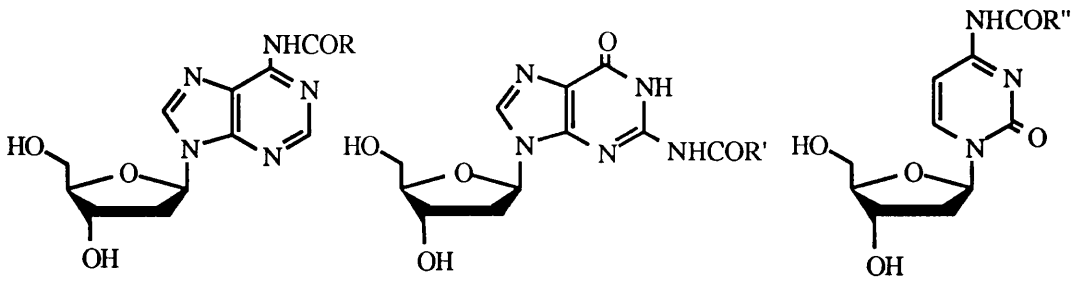
(6)

1.3.4. Protection of Exocyclic Amino Groups of Heterocyclic Bases

Before introduction of the DMT group the exocyclic amino functions of adenine, cytosine and guanine nucleosides must be protected to prevent these from being tritylated as well. Furthermore, protection is also necessary to avoid side reactions with these exocyclic amino groups during the phosphorylation and chain elongation reactions.

The conventional protecting groups for the exocyclic amino groups of nucleosides were first developed nearly three decades ago by Khorana and co-workers for use in oligonucleotide synthesis by the phosphodiester approach (Lohrmann *et al.*, 1966; Khorana, 1968). The benzoyl group was used to protect both adenine (7) and cytosine (8) and isobutyryl group to protect guanine (9). The acyl protecting groups chosen remained stable for long periods under mildly basic or acidic conditions during the synthesis, but could be removed by treatment with concentrated ammonia at the end of the synthesis. Despite the substantial changes since then in the speed and in the type of chemistry used in the synthesis of oligonucleotides, these acyl protecting groups are

still widely used today.



(7) R = Ph (9) R' = $(CH_3)_2CH$ (8) R" = Ph

(10) R = $PhOCH_2$ (11) R' = $PhOCH_2$ (12) R" = $(CH_3)_2CH$

(13) R = t-$BuPhOCH_2$ (14) R' = t-$BuPhOCH_2$ (15) R" = t-$BuPhOCH_2$

Since the above protecting groups require the use of concentrated aqueous ammonia at 55°C for at least 5 h for complete removal, a significant amount of research has been directed to find more labile base protecting groups. A very useful development has been the introduction of the phenoxyacetyl group for the protection of adenine (10) and guanine (11), and isobutyryl for cytosine (12) residues (Schulhof *et al.*, 1987a). They can be removed completely in less than four hours in concentrated ammonia at room temperature. The use of the phenoxyacetyl group not only greatly reduces the duration of ammonia treatment at the end of assembly but also facilitates the preparation of oligonucleotides carrying modified bases sensitive to alkaline conditions. Furthermore, the rate of depurination of N^6-phenoxyacetyldeoxyadenosine in 80% acetic acid is 20% slower than that observed with N^6-benzoyldeoxyadenosine under the same condition (Schulhof *et al.*, 1987b). The nucleoside phosphoramidites protected with phenoxyacetyl on adenine and guanine and isobutyryl on cytidine have been commercially available since 1990 (PAC amidites, Pharmacia). More recently, phosphoramidite monomers protected with t-butylphenoxyacetyl on the exocyclic amino functions of nucleosides (13-15) (Expedite monomer from Millipore) have been launched (Sinha *et al.*, 1993). These monomers can be completely deprotected with

ammonia at room temperature in two hours.

1.3.5. Methods for Oligonucleotide Synthesis

Several methods have been developed for the synthesis of oligonucleotides, mainly differentiated by the way in which the phosphate ester bond is formed. They are termed as the phosphodiester, phosphotriester, phosphoramidite, and H-phosphonate. Since the phosphoramidite procedure gives better overall yields, and is the only chemistry reliable beyond 50 residues, and is the method most widely used today for synthesis of oligonucleotides, it will be discussed in detail. However, an understanding of other methods is also important.

1.3.5.i. Phosphodiester Method

Even though the first chemical synthesis of a dinucleoside phosphate was by a phosphotriester methodology (Michelson and Todd, 1955), the method most widely used in the early stage of development of DNA synthesis was the phosphodiester approach which was developed, and investigated mainly by Khorana's team (Gilham and Khorana, 1958; Khorana, 1968; Agarwal *et al.*, 1972). This method involves the coupling of a nucleoside with a free 3'-hydroxyl group and a 5'-hydroxyl group protected by the acid-sensitive trityl group (16) with a nucleoside 5'-phosphate having a 3'-hydroxyl function blocked usually with an acetyl group (17). The condensation occurs in the presence of a condensing agent such as mesitylenesulfonyl chloride (Ms-Cl) (18a) or 2, 4, 6-triisopropylbenzenesulfonyl chloride (TPS-Cl) (18b). Extension of the chain involves removal of the 3'-protecting group from (19) and coupling with another suitably protected nucleoside 5'-phosphate (Figure 1.5.).

Although the phosphodiester approach was used successfully to synthesize many oligonucleotides, the method has a number of fundamental faults: (a) since phosphodiester functions are nucleophilic and therefore open to attack in subsequent phosphorylation step, the yield is usually low; (b) salts of phosphodiesters are normally soluble only in water or the more polar organic solvents, and so cannot normally be

purified by the simple techniques of organic chemistry such as adsorption chromatography on silica gel or alumina. To purify intermediates an anion-exchange chromatography such as DEAE-Sephadex or DEAE-Cellulose has to be used which limits the scale of preparative reactions; (c) the whole process of purification is tedious and time-consuming, synthesis of an oligonucleotide of 10-15 residues takes a specialist organic chemist several month to complete.

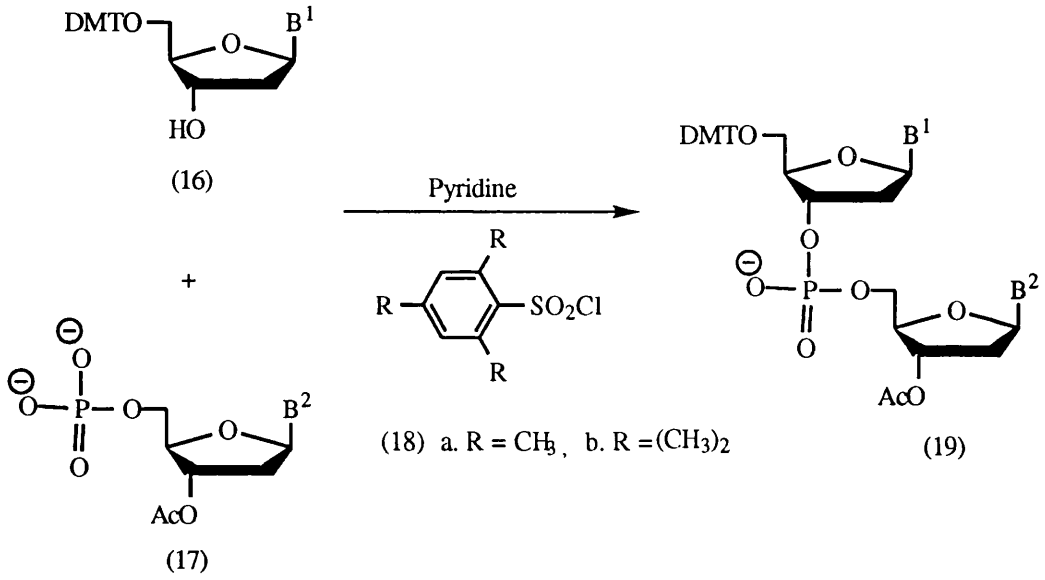


Figure 1.5. Chemistry of phosphodiester method for DNA synthesis.

Clearly, the way to overcome these problems would be to devise a method in which the internucleotide linkages are protected. Thus the phosphotriester method of oligonucleotide synthesis first introduced by Michelson and Todd (1955) was reinvestigated and improved, and became a mature method for synthesis of oligonucleotides.

1.3.5.ii. Phosphotriester Method

Although this chemistry was first applied in the solution-phase, it was particularly successful when applied to solid-phase synthesis (Sproat and Gait, 1984). In the solid-phase phosphotriester method, a 5'-O-DMT-(N-acetylated)-nucleoside-

3'-(arylphosphate) (20) is coupled to a nucleoside (21) attached through its 3'-position to a solid support. The coupling agent is mesitylenesulphonyl 3-nitro-1, 2, 4-triazolide (22). It activates the nucleoside-3'-phosphodiester and allows reaction with the 5'-hydroxyl group of the nucleoside on the support. To extend the chain, the DMT group is removed by treatment with acid to liberate the 5'-hydroxyl group ready for further coupling (Figure 1.6). Since the product (23) is a phosphotriester and thus is protected from further reaction with phosphorylating agents the yield is much better than with the phosphodiester method. However, some side reactions are important. In particular there is competitive sulphonylation of the 5'-hydroxyl group by the coupling agent which reduces the efficiency of the coupling.

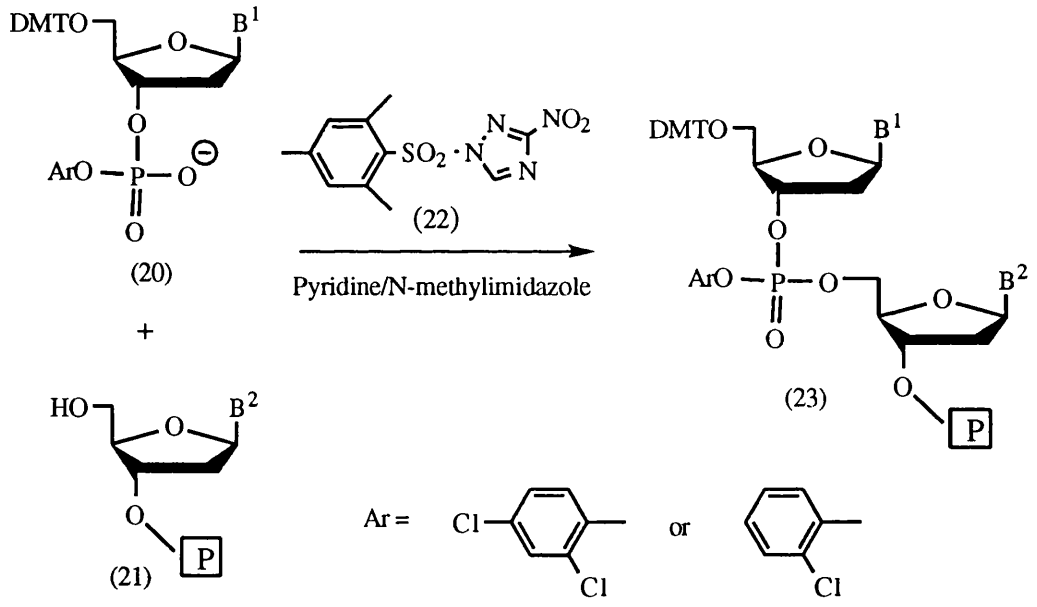


Figure 1.6. Chemistry of phosphotriester method for DNA synthesis.

The choice of the protecting group for the internucleotide linkage is the most important consideration for the phosphotriester method. This protecting group must remain intact during the whole synthesis and be stable under the conditions required to unblock the hydroxy functions but then be removable under conditions where no internucleotide cleavage or modification of the desired product occurs. The most favourable protecting groups are mono- or dichlorophenyl derivatives, since they

almost completely fulfil the requirements above and are readily cleaved by oximate reagents (Reese and Zard, 1981).

1.3.5.iii. H-phosphonate Method

Although this chemistry was introduced by Todd and his co-workers to prepare diribonucleotide phosphate almost four decades ago (Hall *et al.*, 1957), it was successfully applied in oligonucleotide synthesis only recently (Garegg *et al.*, 1985; Froehler *et al.*, 1986; Andrus *et al.*, 1988). A protected nucleoside H-phosphonate (24) is converted into a dinucleoside hydrogen phosphonate (25) by reaction with the 5'-OH of a nucleoside (26) in the presence of a condensing agent such as pivaloyl chloride (Froehler, 1986 #79) or, preferably, adamantoyl chloride (Andrus *et al.*, 1988). Oxidation of all the phosphorus centres is carried out simultaneously at the end of the synthesis since the internucleoside H-phosphonate bond is stable enough to withstand the conditions of synthesis and so the chain can be extended without prior oxidation (Figure 1.7).

One advantage of this chemistry is that oxidation is subject to general base catalysis and this allows nucleophiles other than water to be used at the "oxidation" step to give a range of oligonucleotide analogues. Another advantage is that no phosphate protecting group is needed since the internucleoside H-phosphonate diester link is relatively inert to the conditions of coupling.

The H-phosphonate chemistry is not without disadvantages. At the beginning of each coupling cycle, after activation, two H-phosphonates can rapidly dimerize to form a symmetrical phosphite anhydride, and if this reacts with a 5'-OH group from a support-linked nucleoside, a branched trinucleotide derivative is formed. It is impossible to eliminate this side reaction completely and thus the yield with this method is lower than that obtained with phosphoramidite chemistry

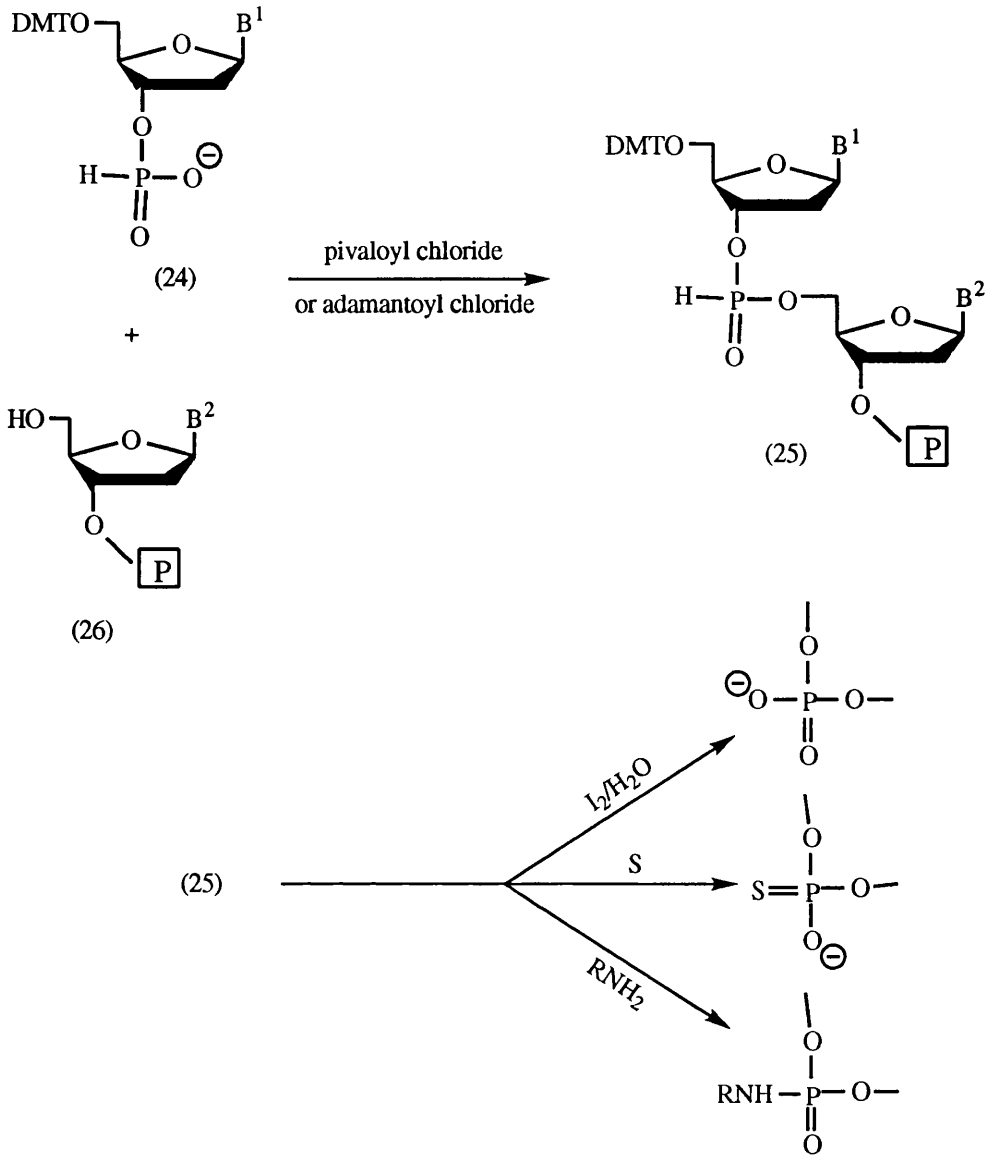


Figure 1.7. Chemistry of H-phosphate method.

1.3.5.iv. Phosphoramidite Method

The basis of the phosphoramidite approach (earlier known as the phosphite triester method) was initially developed by Letsinger et al. (1975; 1976). Essentially, a 5'-O-protected nucleoside (27) was reacted with 2, 2, 2-trichloroethyl phosphorodichloridite. The nucleoside-3'-O-phosphoro-chloridite (28) was formed within 5 min at -78°C, and subsequent addition of a 3'-protected nucleoside (29) resulted in the rapid formation of the dinucleoside phosphite triester which was then oxidized to the corresponding phosphate triester (30) by aqueous iodine (Figure 1.8).

Under optimal conditions, (30) was prepared within 1 h in 82% isolated yield.

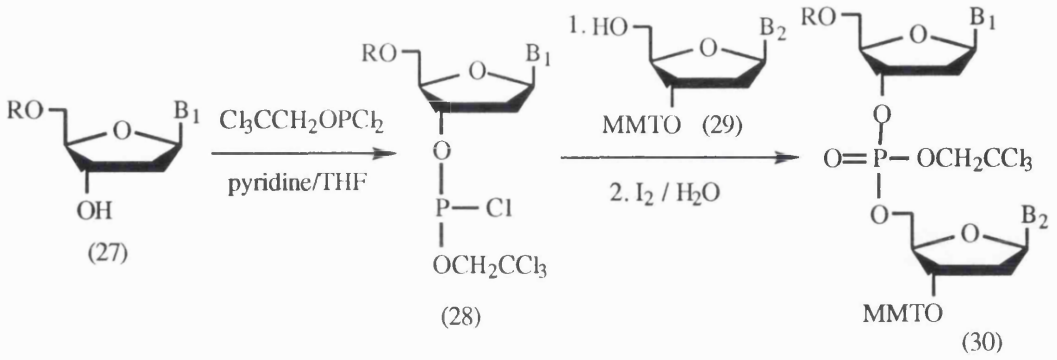


Figure 1.8. Chemistry of phosphite triester method.

At the beginning of the eighties, Matteucci and Caruthers (1980; 1981) adapted this chemistry to solid-phase oligonucleotide synthesis. In their approach a deoxyribonucleoside methoxychlorophosphite (31) or a methoxymonotetrazolphosphite (32) was coupled with a nucleoside (33) covalently linked to a silica support. The dinucleoside phosphate triester (34) was generated in a yield greater than 95% (Figure 1.9). A decanucleotide was prepared and isolated in a 30% yield.

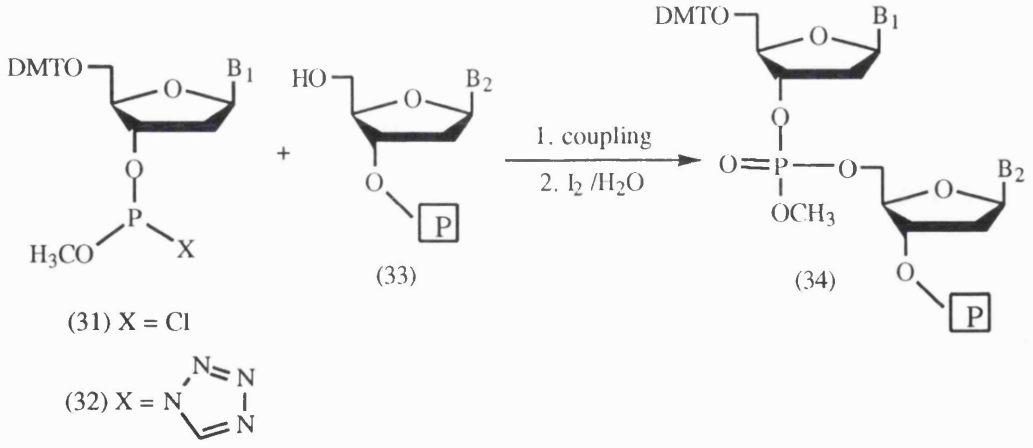
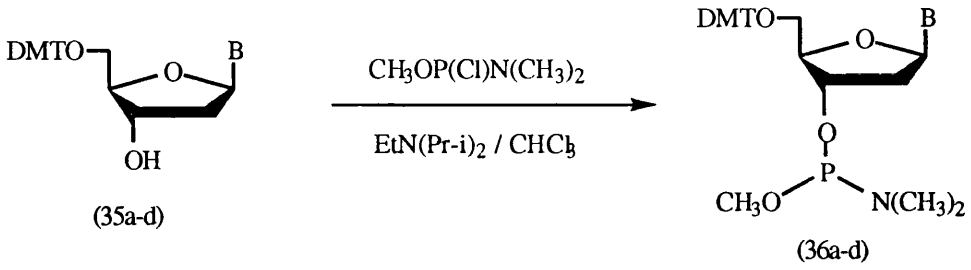


Figure 1.9. Solid-phase synthesis of DNA by phosphite triester method.

Although the phosphite triester was subsequently improved, the application of this method via nucleoside chlorophosphites or the corresponding tetrazolides to the

solid-phase synthesis of oligonucleotides was still generally problematic. The preparation of these reagents from reactive bifunctional phosphitylating agents had to be performed at low temperature in the absence of moisture and under inert atmosphere. In addition to being contaminated by variable amounts of undesired (3'-3')-dinucleoside phosphite triester, the nucleoside chlorophosphites and /or corresponding tetrazolides were sensitive to hydrolysis and, hence, difficult to handle. Although these synthons enabled the rapid formation of a large number of DNA segments on silica support, their relative instability prevents their reliable use in automated system.



a, B = T; b, B = C^{Bz}; c, B = A^{Bz}; d, B = $G^{i\text{-}Bu}$

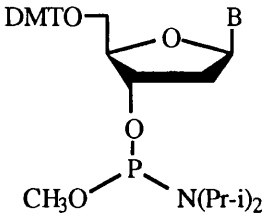
Figure 1.10. Synthesis of methyl N, N-dimethylaminophosphoramidites of nucleosides.

To satisfy the requirements for automation, the phosphitylated nucleoside monomers must be stable to hydrolysis and air oxidation under normal laboratory conditions. In addition to being easily isolated and stable upon storage, these monomers must be readily activated to reactive species to permit rapid and efficient formation of internucleotidic linkages. A breakthrough was achieved by Beaucage and Caruthers (1981) who developed nucleoside phosphoramidites as a new class of intermediates for the oligonucleotide synthesis. Essentially, the approach consisted of the reaction of 5'-DMT nucleosides (35a-d) with methyl N, N-dimethylchloro-phosphoramidite in the presence of N, N-diisopropylethylamine in dry chloroform at 20°C. After work-up, the resulting phosphoramidites (36a-d) were isolated by precipitation in cold hexane as dry powders in 90-94% overall yields (Figure 1.10).

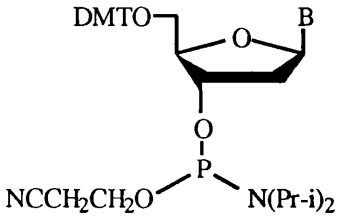
These were stored under an inert atmosphere at 20°C for at least a month without significant decomposition.

With these monomers and 1H-tetrazole as an activator, various dimers were synthesized by solid-phase in yields from 93-100% (Beaucage and Caruthers, 1981). This approach was then applied to the synthesis of much larger oligonucleotides (up to 45 bases) with automated system (Josepheson *et al.*, 1984). Coupling yields ranged from 85-100%.

In spite of the usefulness of these monomers (36 a-d) in the solid-phase synthesis of oligonucleotides, their application in automated system was unreliable because their stability in anhydrous acetonitrile varied from hours to weeks depending on their purity. To overcome this problem, McBride and Caruthers (1983) and Adams et al (1983) investigated a series of related methyl N, N-dialkylaminophosphoramidites. From these investigations it appeared that methyl N, N-diisopropylamino-phosphoramidites of the four common base-protected nucleosides (37 a-d) were, regarding their stability and reactivity, the most useful intermediates for the synthesis of oligonucleotide by the phosphoramidite approach, and hence were used extensively in the automated synthesis of oligonucleotides afterwards.



(37a-d)



(38a-d)

a, B = T; b, B = C^{Bz}; c, B = A^{Bz}; d, B = $G^{i\text{-}Bu}$

Using the methyl group as the protecting group is not without problems. The removal of this group from a synthetic oligomer requires thiolates which are unpleasant reagents. Particularly this methyl group is an active methylating agent and reacts with thymidines in the synthetic oligomer (Gao, 1985). For these reasons the

phosphoramidites (37a-d) have been superceded by ß-cyanoethyl N, N-diisopropylaminophosphoramidites (38 a-d) which were introduced by Sinha et al (1983; 1984).

The cyanoethyl group is stable during DNA synthesis, but can be removed by a ß-elimination under the same basic conditions required for the deprotection and cleavage of oligonucleotides from the solid support. The ß-cyanoethyl N, N-diisopropylaminophosphoramidites have been by far the most widely used monomers for the automated solid-phase synthesis of oligonucleotides and are commercially available from many companies.

The chemistry of automated solid-phase phosphoramidite technique for oligonucleotide synthesis using the phosphoramidites (38a-d) is illustrated in Figure 1.11. The first step of the cycle is the removal of DMT group from the 5'-hydroxyl of the nucleoside (39) covalently attached the support . The coupling reaction is catalyzed by tetrazole (40), which protonates the N, N-diisopropylphosphoramidite (41), and converts the diisopropylamino moiety into a good leaving group. The protonated amino group is displaced by the 5'-hydroxyl group of the support-bound nucleoside (42) and the dimer (43) is formed. After a capping step (see below) the dimer is oxidized with aqueous iodine to convert the phosphite triester into a more stable phosphate triester (44). The average coupling efficiency is about 98.5% and sequences in excess of 100 bases can be prepared. However, in each cycle around 1.5% of the oligonucleotide chains on the glass beads fail to react with the activated monomer and if this situation were ignored, a complex mixture of truncated sequences would accumulate, the majority of which would be only one nucleotide shorter than the correct product. These impurities would obviously have similar properties to the desired oligonucleotide and purification would be very difficult. Hence the capping step is performed to terminate the extension of these unwanted oligomers (45).

Because of its simplicity and efficiency, automatic synthesis of oligonucleotides with solid-phase phosphoramidite chemistry is now the predominant method for preparing oligonucleotides.

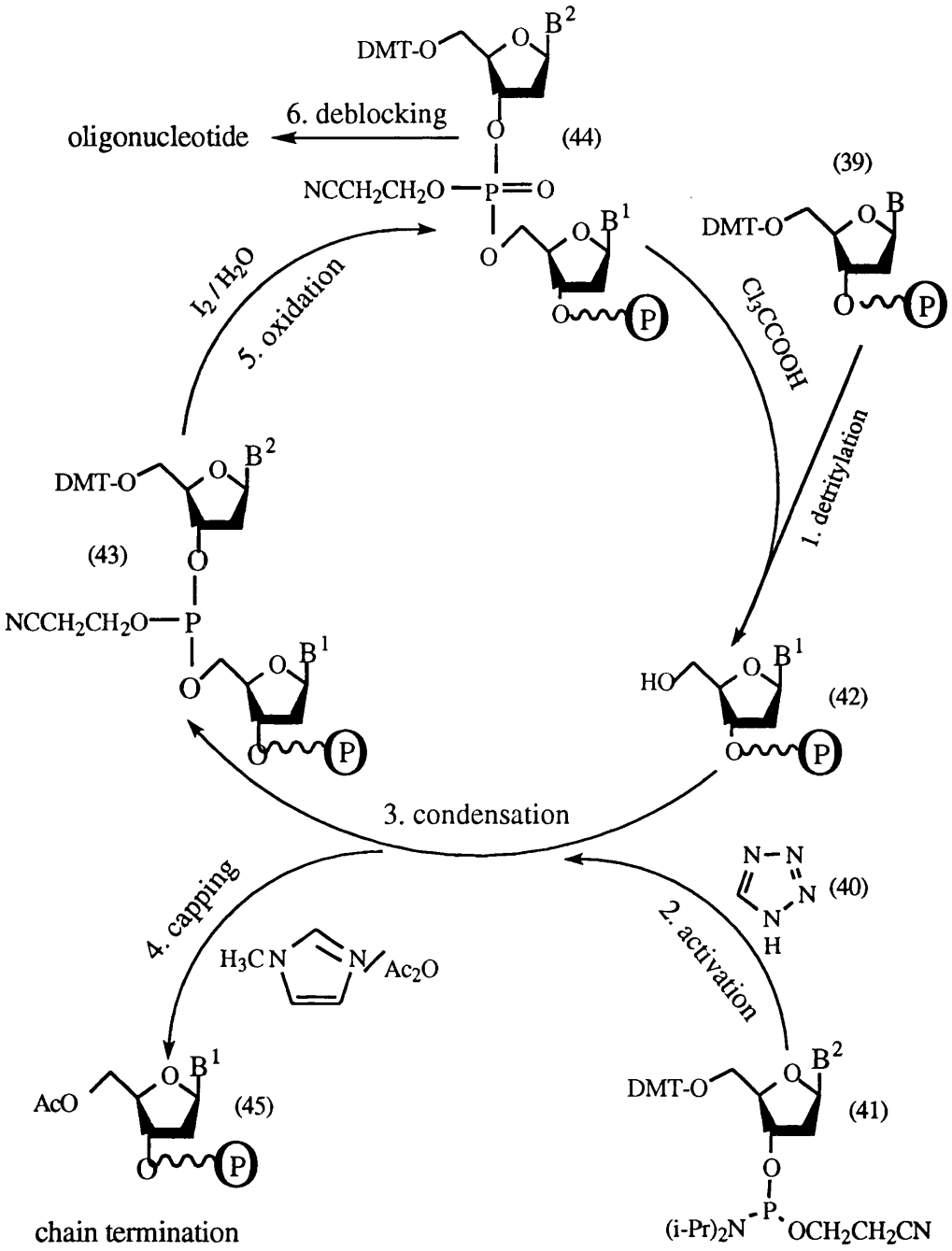


Figure 1.11. Chemistry of automated solid-phase phosphoramidite approach for DNA synthesis (Taken from Engels and Uhlmann, 1989).

With the aid of an efficient automated solid-phase method, synthesis of unmodified DNA has become routine but an experienced chemist is still needed to design and synthesize modified oligonucleotides. The next section will discuss the synthesis of modified oligonucleotides and their applications.

1.4 CHEMICAL SYNTHESIS OF OLIGONUCLEOTIDES CONTAINING MODIFIED BASES AND THEIR APPLICATIONS

An inspection of DNA suggests several potential sites for the introduction of modification. These include the sugar, base and phosphate backbone of these molecules and the 3'- and 5'- ends (Figure 1.12). Because of the importance of these modified oligonucleotides for studies related to carcinogenesis, mutation, DNA repair and for investigation of protein-DNA interaction, and for their use as probes, inhibitors and cytotoxic drugs (for reviews see: Basu and Essigmann, 1988; Swann, 1990; Englisch and Gauss, 1991; Marshall and Caruthers, 1993), many methods have been recently developed for synthesizing different modified oligomers (Eckstein, 1991). The particular interest of our lab in base-modified oligonucleotides has meant that this project focuses on developing methods for synthesizing this sort of modified oligonucleotides and investigating their applications. Hence, the following discussion will focus on oligonucleotides containing modified bases, especially, those widely used for study of DNA-protein interactions, which is a part of this project, and those used for investigation of DNA damage by alkylating agents, which is the main interest of our lab.

1.4.1. Oligonucleotides for Study of DNA-Protein Interactions

DNA sequence recognition by enzymes and binding proteins plays fundamental roles in many biological processes. Many proteins that interact with double strand B-DNA show a very high specificity for a particular target sequence. Good examples are repressor proteins and transcription factors which usually bind to sequences 12- 20 bases in length, and restriction endonucleases which cut DNA at short recognition sequences, usually 4-8 base pair in length. The mechanisms by which high selectivity is achieved are not very well understood. The selectivity is achieved in part by these proteins making direct contacts to exposed portions of bases in the major or minor groove. It is possible to probe these contacts by using base-modified oligonucleotides by several different approaches. Usually one of the most difficult parts of these

approaches is the preparation of suitable base analogues and their incorporation into oligonucleotides.

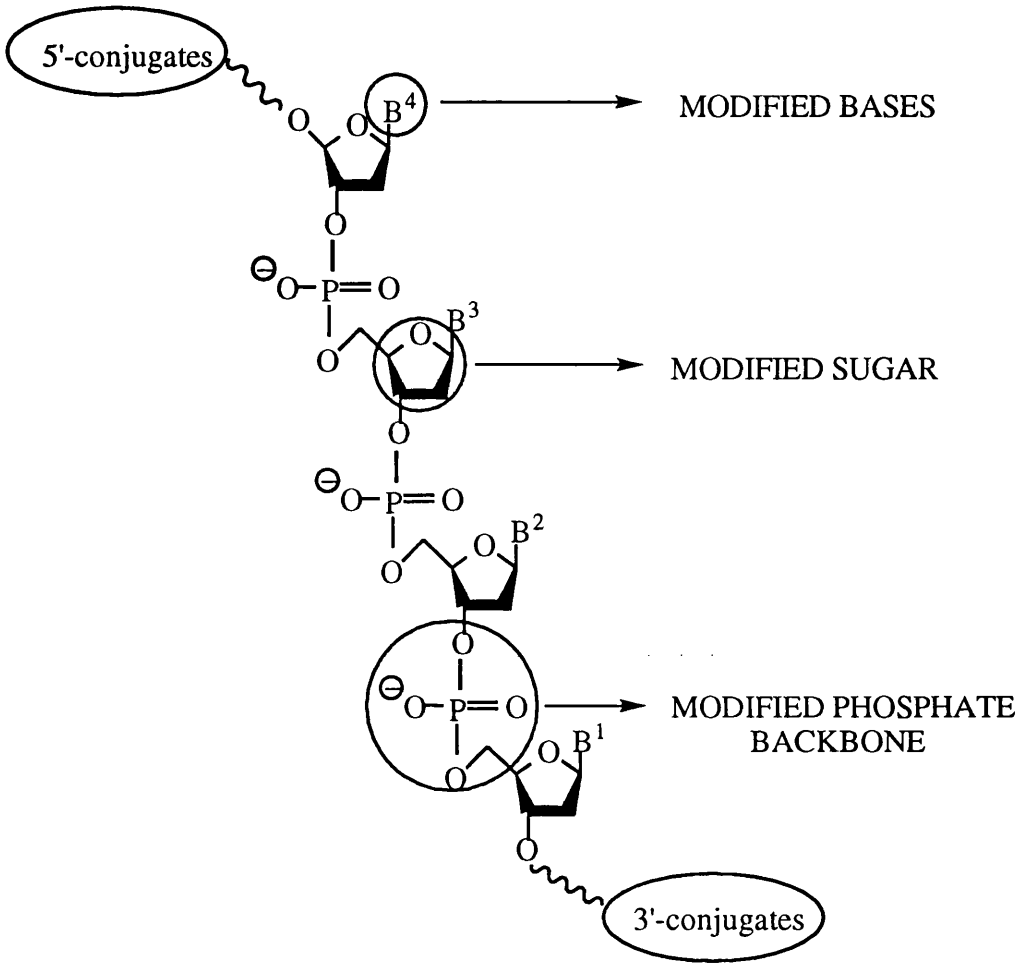


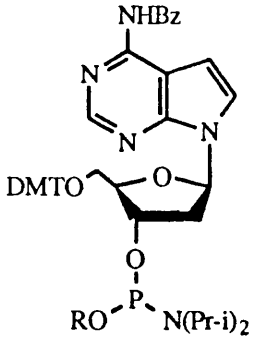
Figure 1.12. Potential sites for the introduction of modification in DNA.

1.4.1.i. Base Analogue Approach

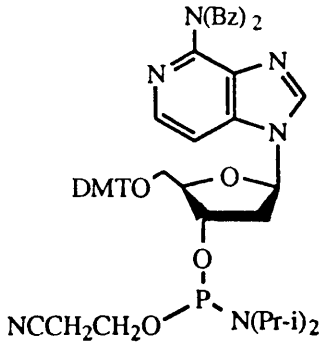
This approach consists of the deletion of one of the potential contact sites of natural bases by replacement of an atom or a functional group, e.g. a ring nitrogen, an exocyclic amino group on G, C, A, and the methyl group of T. Ideally, the alteration should be subtle enough so that it only affects the protein contact at that particular point and does not cause major conformational changes in the DNA structure.

7-Deaza-2'-deoxyguanosine phosphoramidites (46a-b), in which a potential hydrogen bond acceptor was replaced by a carbon, were prepared and incorporated into

oligonucleotides by Seela and Driller (1985). The modified nucleoside (47) was obtained by phase-transfer glycosylation of 2-amino-4-methoxy-7H-pyrrolo(2, 3-d)pyrimidine (48) with 1-chloro-2'-deoxy-3,5-di-O-p-toluoyl-D-erythro-pentofuranose (49) followed by demethylation of methoxynucleoside (50). Treatment of (47) with isobutyric anhydride followed by selective deprotection with sodium hydroxide at 0°C resulted in the formation of the N^2-isobutyrylated compound (51). Treatment of (51) with 4, 4'-dimethoxytrityl chloride and then with methyl N, N-diisopropylchlorophosphoramidite or 2-cyanoethyl N, N-diisopropylchloro-phosphoramidite produced phosphoramidites (46a-b) (Figure 1.13) which were then incorporated into oligomers by automatic solid-phase synthesis. Subsequently, using a similar strategy, the oligonucleotides containing 7-deaza-2'-deoxyadenosine were also prepared via the phosphoramidites (52a-b) (Seela and Kehne, 1985).



(52) a. R = CH_3
b. R = CH_2CH_2CN



(53)

3-Deaza-2'-deoxyadenosine is another useful probe for studying DNA-protein interactions since it is able to maintain Watson-Crick hydrogen bonding pattern but removes an essential hydrogen bond acceptor (adenine N-3) from 2'-deoxyadenosine. Its preparation and incorporation into oligonucleotides has been described by Cosstick et al. (1990) in order to study the interaction of the Eco RV with DNA. 3-Deaza-2'-deoxyadenosine phosphoramidite (53) was prepared from 2, 6-dichloro-pyridine-N-oxide in 14 steps with overall yield of less than 3%. The modified base was

inserted in oligonucleotides containing the Eco RV endonuclease recognition sequence d(GATATC) by automated solid-phase synthesis. Melting temperature of the oligomers showed that the modified base had little effect on the thermal stability of the duplexes.

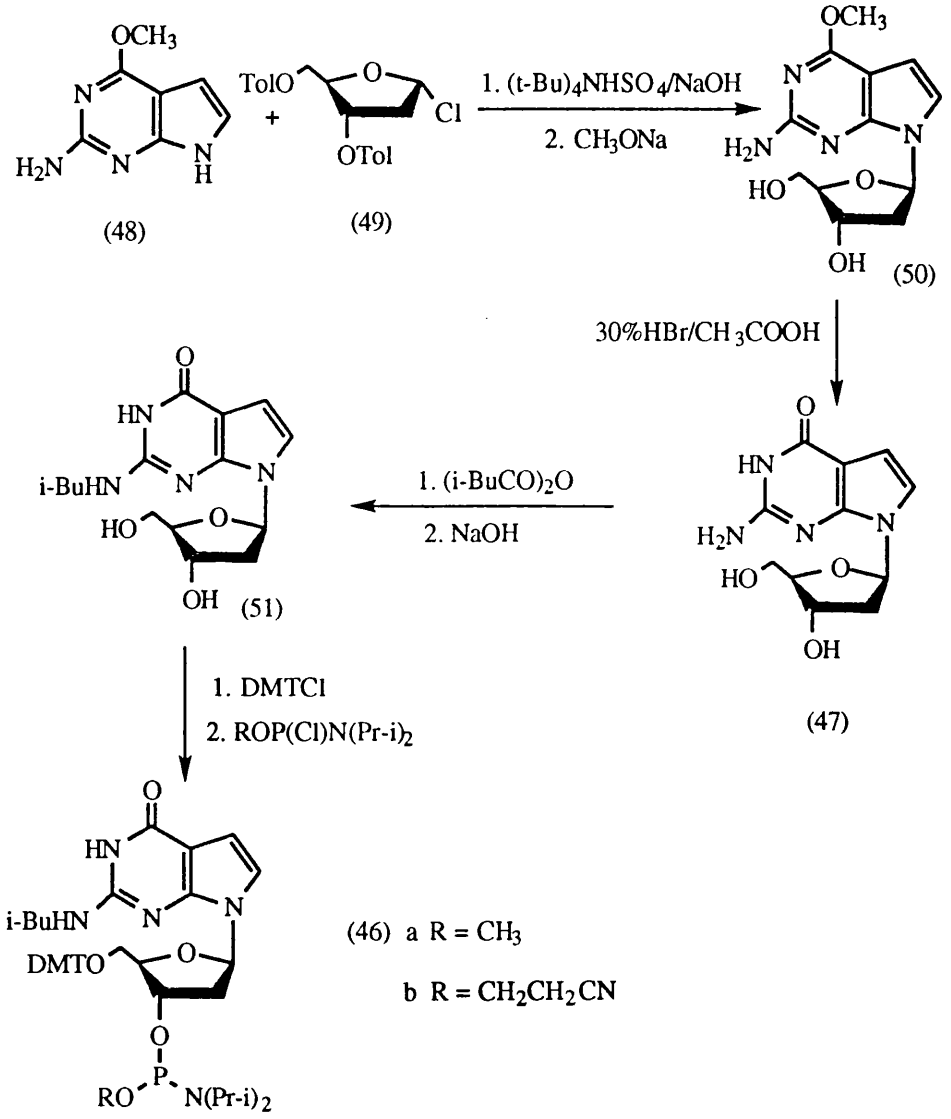


Figure 1.13. Chemical synthesis of 7-deaza-2'-deoxyguanosine phosphoramidites.

2-Aminopurine nucleoside can be considered as a derivative of 2'-deoxyguanosine from which the carbonyl at the 6-position has been removed. It is also a structural isomer of 2'-deoxyadenosine in which the amino group has been moved from the 6-position to the 2-position. A novel approach to the synthesis of the 2-aminopurine nucleoside and its corresponding phosphoramidite derivative, together

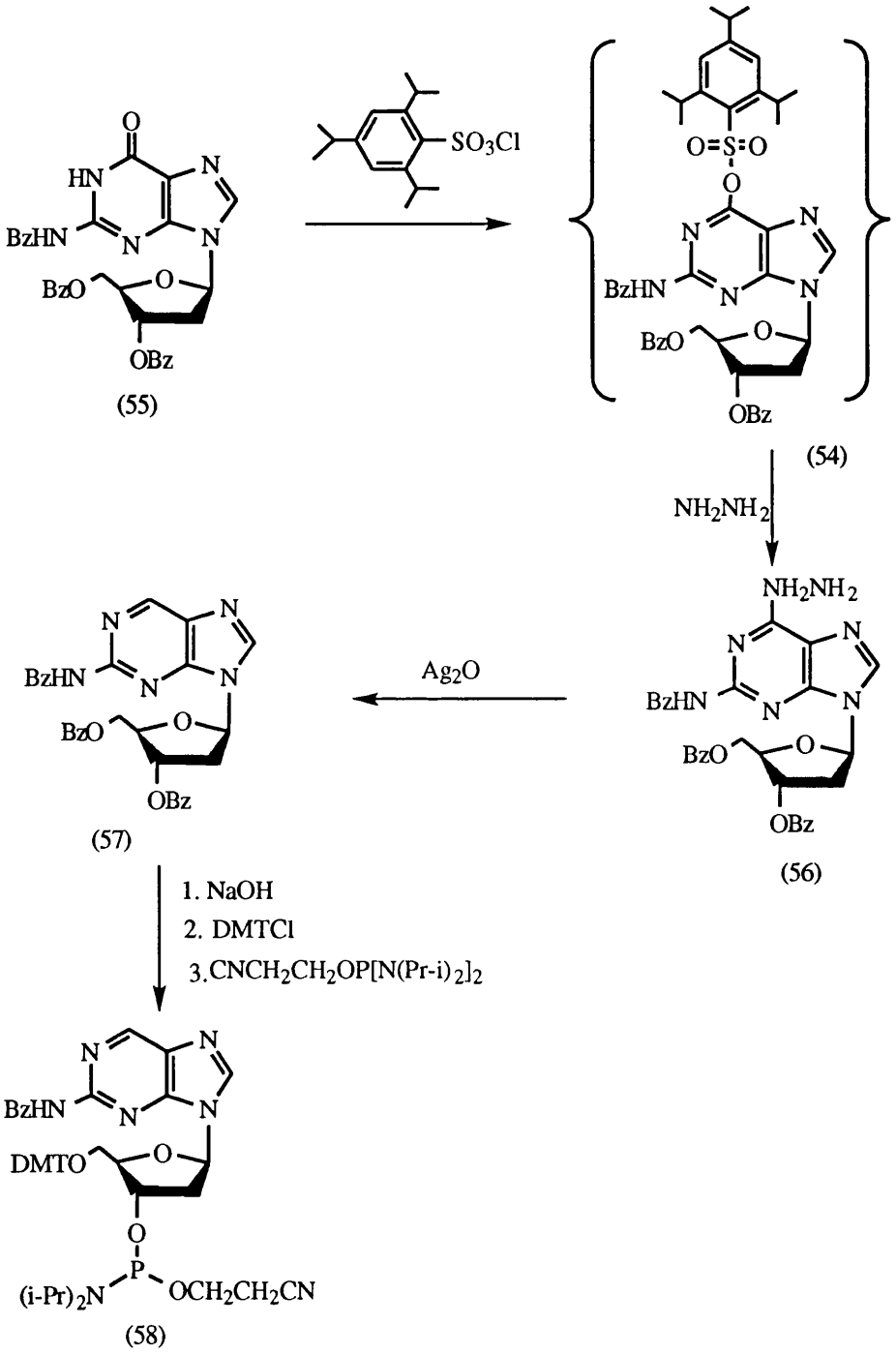
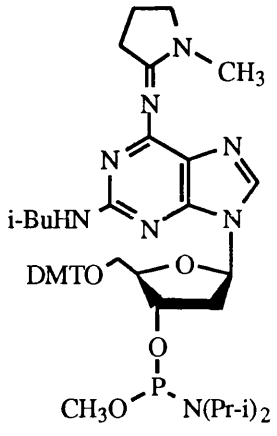


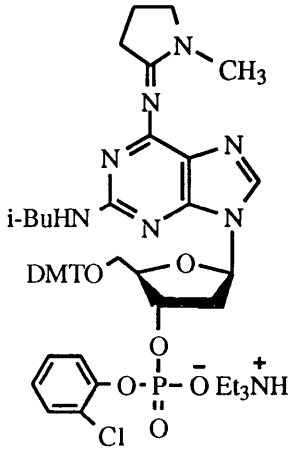
Figure 1.14. Chemical synthesis of phosphoramidite of 2-aminopurine nucleoside.

with its incorporation into oligonucleotides, has been reported by McLaughlin and co-workers (1988). The 6-sulfonated 2'-deoxyguanosine derivative (54), prepared from protected 2'-deoxyguanosine (55), was reacted with hydrazine to produce the 6-hydrazino derivative (56). Oxidation of (56) in the presence silver oxide formed the

protected 2-aminopurine nucleoside derivative (57). Deprotection of the carbohydrate residue and subsequent tritylation and phosphitylation produced the phosphoramidite (58) (Figure 1.14). It was incorporated into oligonucleotides on an automatic DNA synthesizer in exactly the same manner as the four common nucleoside phosphoramidite monomers with no observable difference in coupling efficiency.



(59)

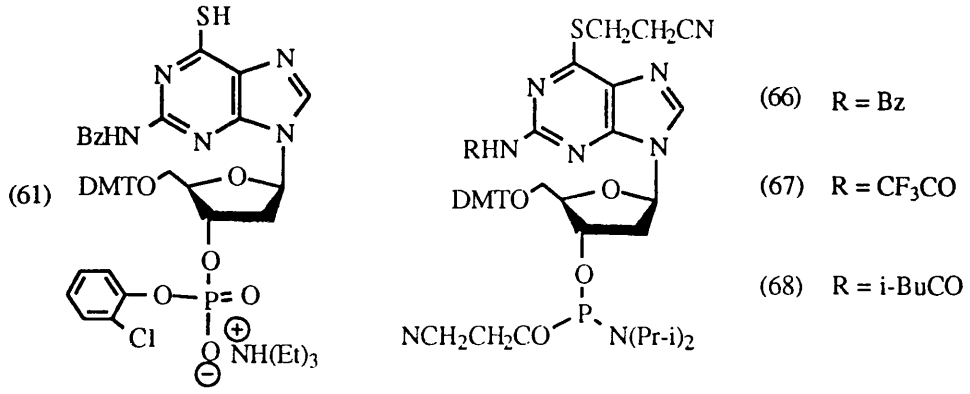


(60)

Synthesis of oligonucleotides containing 2, 6-diaminopurine, a structural guanine analogue, in which the 6-carbonyl group is replaced by an amino group, has been reported by Chollet et al. (1986). By insertion of phosphoramidite (59) or phosphotriester (60), oligomers up to 27 bases in length, containing up to five 2, 6-diaminopurines, were synthesized with coupling yields of 95-99%. However the two amino-protecting groups were difficult to remove and severe deprotection conditions, such as conc. ammonia at 65°C for 7 days or 0.1 M NaOH at 40°C for 72 hours were necessary. The thermal stabilities of the oligonucleotide duplexes containing 2, 6-diaminopurine showed that the introduction of 2, 6-diaminopurine into DNA sequences stabilized duplex formation with complementary sequences since an extra hydrogen bond is formed between the modified base and thymine.

6-Thio-2'-deoxyguanosine is a dG analogue in which the 6-carbonyl group is replaced by sulphur. Since the hydrogen bonds between DNA binding proteins and the 6-keto group of deoxyguanosine often contribute to specificity of action, replacement of

the 6-keto oxygen with sulphur should give a probe valuable in the study of DNA-protein interactions. The first report of chemical synthesis of oligonucleotides containing 6-thioguanine was from Rappaport (1988). 6-Thiodeoxyguanosine phosphotriester (61) was incorporated into oligomers using phosphotriester chemistry without the protection of the 6-thio function. Because of the great difficulty in removing the protecting groups on the bases and lack of evidence that loss of the sulphur had not taken place during deprotection, this method was soon superseded by other methods.



The first novel approach for the synthesis of 6-thioguanine containing oligomers was reported by Christopherson and Broom (1991) (Figure 1.15). Phenoxyacetylation of the exocyclic amino group of 6-thiodeoxyguanosine (62) with phenoxyacetyl chloride, after transient protection of the hydroxyl functions of the sugar, produced the N^2-protected derivative (63) which was then alkylated at S^6 with 3-bromopropionitrile to yield N^2, S^6-protected derivative (64). Treatment of the compound (64) with dimethoxytrityl chloride and then with 2-cyanoethyl N, N, N', N'-tetraisopropylphosphorodiamidite yielded the phosphoramidite (65). 0.01 M NaSH was included in the deprotecting reagent (1.0 M NaOH) to prevent the conversion of 6-thioguanine to guanine. Using a similar approach, phosphoramidites (66) (Waters and Connolly, 1992), (67) and (68) (Sudhakar *et al.*, 1992) were also prepared and incorporated into DNA. It must be pointed out that the trifluoroacetyl in (67) is not an ideal protecting group since the removal of this group requires severe conditions which

could destroy 6-thioguanine.

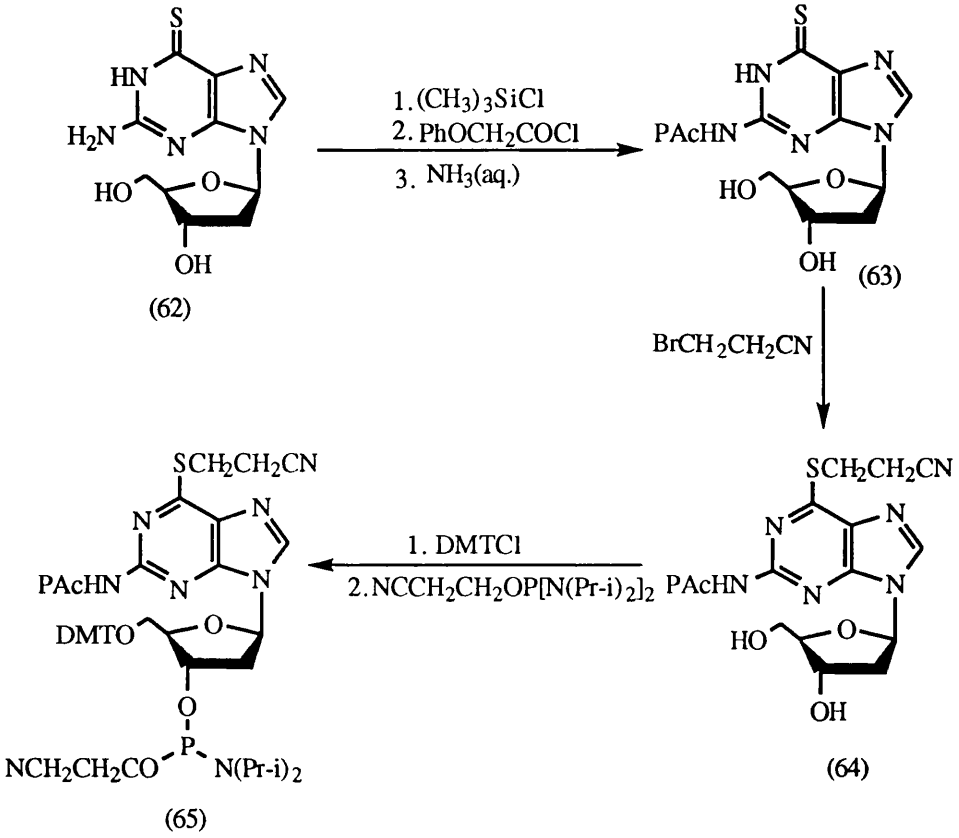


Figure 1.15. Chemical synthesis of 6-thio-2'-deoxyguanosine phosphoramidite.

For the pyrimidines, the base analogues prepared for study of DNA-protein interactions generally involve the deletion or alteration of the N^4-amino group, the O^4- and O^2-carbonyl groups and the C-5 methyl group. The deletion of the thymine methyl group can be achieved simply by replacement of thymidine with commercially available 2'-deoxyuridine. Other modifications can only be achieved by carefully designed modified base analogues.

Synthesis of oligonucleotides containing 4-thiothymidine has been reported by Connolly and Newman (1989) (Figure 1.16). Treatment of 3', 5'-protected thymidine (69) with Lawesson's reagent produced (70) which was subsequently deprotected and tritylated on 5'-OH group. The sulphur atom in (71) was protected with $-SCH_3$ group since it is nucleophilic and reactive toward reagents such as CH_3I. The $-SCH_3$ group

can be removed with dithiothreitol (DTT). Further reaction of (72) with 2-cyanoethyl N, N-diisopropylchlorophosphoramidite gave thiothymidine phosphoramidite (73) suitable for oligonucleotide synthesis (Figure 1.16). The phosphoramidite was incorporated into d(GACGATATCGTC), a self-complementary dodecamer containing the EcoRV recognition site (underlined) in place of the two T residues within this site with almost the same coupling efficiency as that on unmodified base addition. But, because of the instability of the S-SCH_3 bond to H^+ and I_2 during the elongation steps, the yield of oligonucleotides was only 10-15% compared to 60% for the unmodified dodecamer.

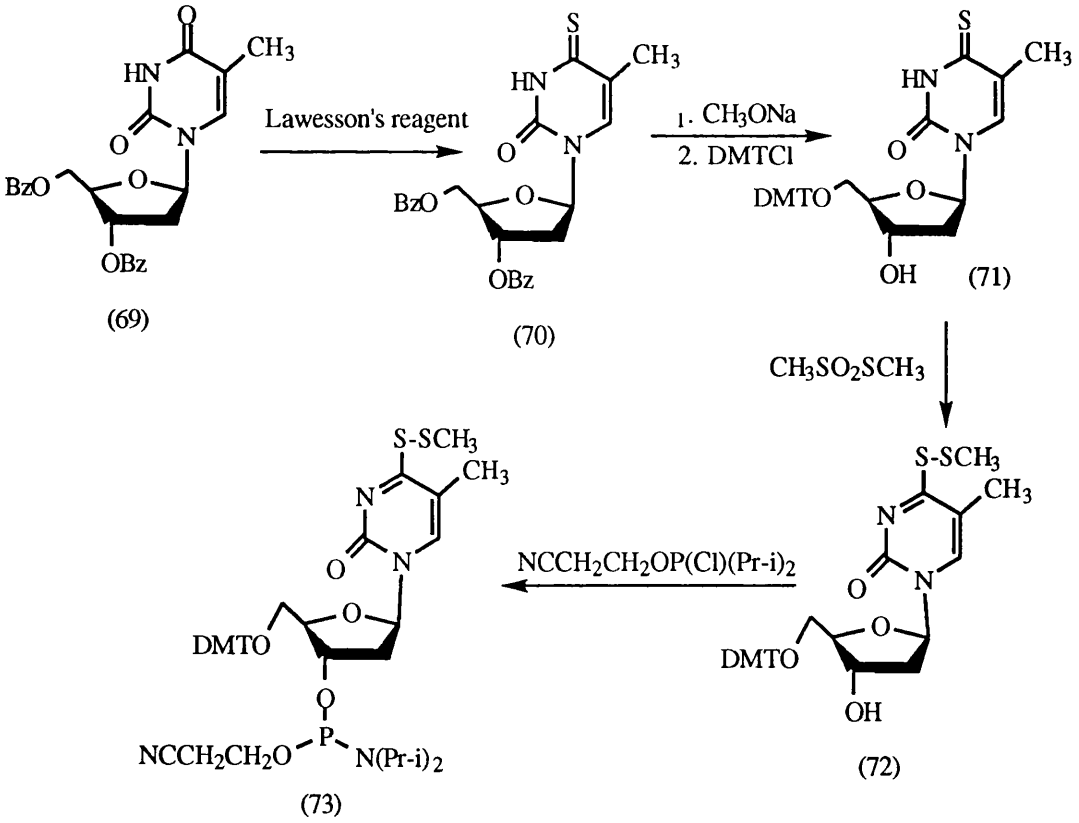
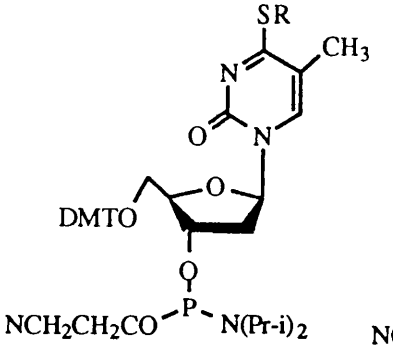


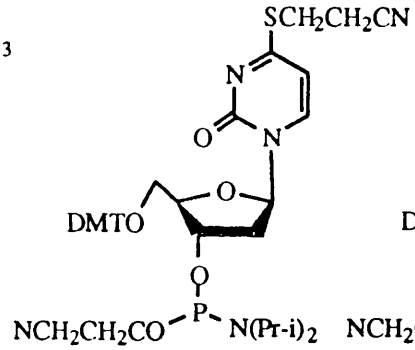
Figure 1.16. Chemical synthesis of phosphoramidite of 4-thiothymidine.

Since the -SCH_3 does not give appropriate protection, Nikforov and Connolly (1991) subsequently used aryl groups to protect the 4-thioketo function (74 and 75). The removal of phenyl or p-nitrophenyl group could be achieved by incubation of the

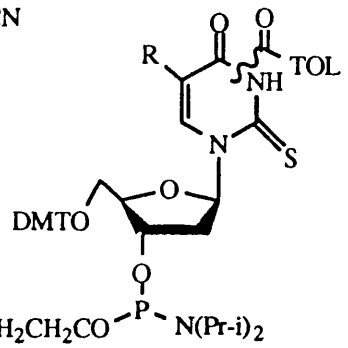
protected oligomers with a 0.3 M CH_3COSK solution in EtOH at 55°C. The purity and the yield of the oligonucleotides thus obtained were much greater than those obtained with -SCH_3 protection. Later, the cyanoethyl group was used to mask the sulphur atom in the 4-thiothymidine (76) (Nikiforov and Connolly, 1992b). Deprotection of S-cyanoethyl group was easily effected by treatment of protected oligomers with a 0.3 M solution of 1, 8-diazabicyclo (5, 4, 0)undec-7-ene (DBU) in dry acetonitrile at room temperature for 1 h. Using the same protection group, oligomers containing 4-thio-2'-deoxyuridine were also synthesized via the phosphoramidite (77) (Nikiforov and Connolly, 1992b).



(74). R = Ph, (75). R = P-NO_2Ph,
(76). R = $NCCH_2CH_2$



(77)



(84) R = CH_3
(85) R = H

Connolly and Newman (1989) also developed a method for synthesis of oligonucleotides containing 2-thiothymidine. The synthetic route to 2-thiothymidine phosphoramidite (78) is shown in Figure 1.17. The preparation started with the formation of 5'-iodo-5'-deoxy-3'-O-acetylthymidine (79) from 3'-O-acetylthymidine (80). The 5'-iodo derivative (79) was converted to the 2, 5'-cyclo derivative (81) with silver acetate which was then treated with H_2S to produce the 2-thiothymidine derivative (82). Reaction of (82) with 4, 4'-dimethoxyltrityl chloride (DMT-Cl) followed by deacetylation and phosphitylation formed phosphoramidite (78) (Figure 1.17). As the sulphur atom in 2-thiothymidine was expected to be much less reactive than that in 4-thiothymidine it was not protected.

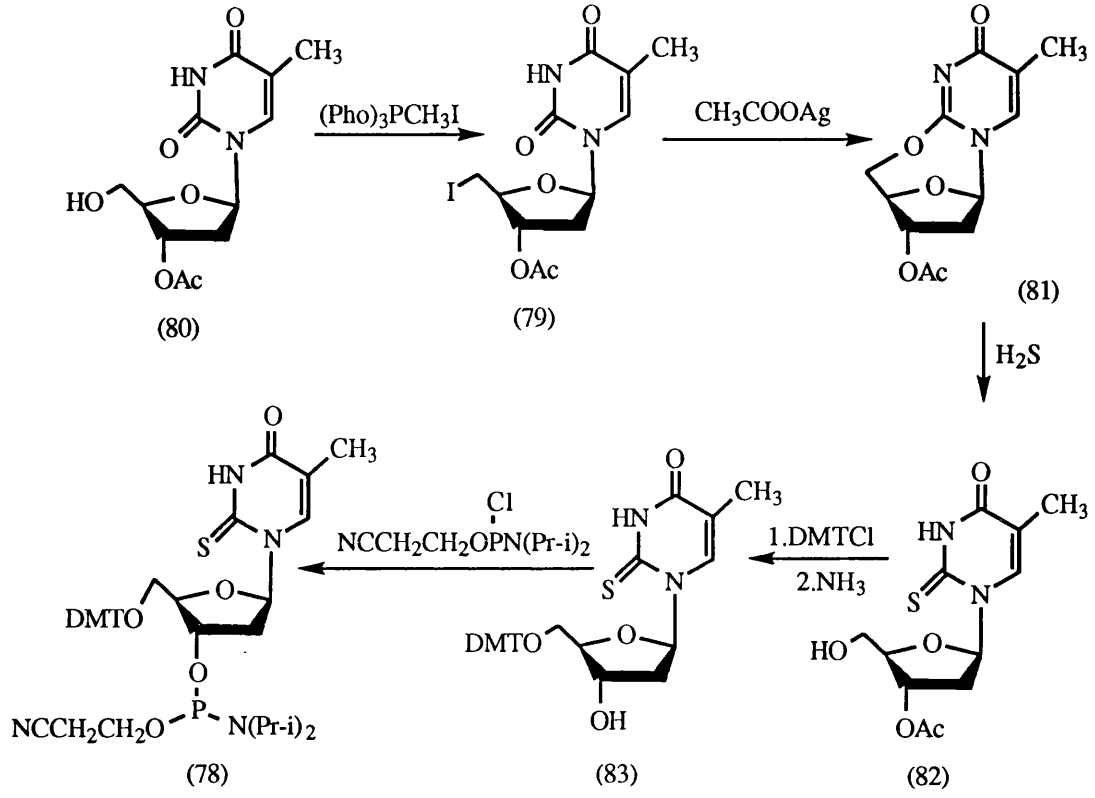


Figure 1.17. Chemical synthesis of 2-thiothymidine phosphoramidite.

Using the similar strategy, Rajur and McLaughlin (1992) also prepared oligonucleotides containing 2-thiothymidine . In contrast to the results of Connolly and Newman (1989), it was found that 2-thiothymidine was generally unstable to the oxidation conditions employed in DNA synthesis and that protection of sulphur atom at the 2-position is necessary. This has been firmly supported by a recent report of Kuimelis and Nambiar (1994). They demonstrated that pure oligomers containing 2-thiothymidine or 2-thiodeoxyuridine can be obtained only by the incorporation of the phosphoramidites (84) or (85) which were protected with toluoyl group at either N^3- or O^4-position of the thionucleoside. This protecting group was readily cleaved by standard post-synthetic ammonia treatment.

The synthesis of oligonucleotides containing 2-pyrimidinone nucleoside, in which the exocyclic amino group of deoxycytidine is deleted, was described by Gildea and McLaughlin (1989). Protected 2-pyrimidinone nucleoside (86) was obtained from

the protected deoxycytidine (87) via a hydrazino derivative (88). After removal of protection groups from (86), the pyrimidinone nucleoside (89) was reacted with 9-chloro-9-phenylxanthene and subsequently phosphitylated with 2-cyanoethyl N, N-diisopropylchlorophosphoramidite to produce (90) (Figure 1.18). Because of the lability of 2-pyrimidinone nucleoside to acid and base catalyzed hydrolysis, the more acid labile 9-phenylxanthene (Px) group was used in place of the more common DMT group. Self-complementary oligomers containing the Eco RI recognition site (GAATTC) substituted by the pyrimidinone residue were synthesized on a CPG solid support with no observable change in coupling efficiency compared to unmodified base additions. However, double HPLC isolations were necessary in order to adequately purify the modified oligomers.

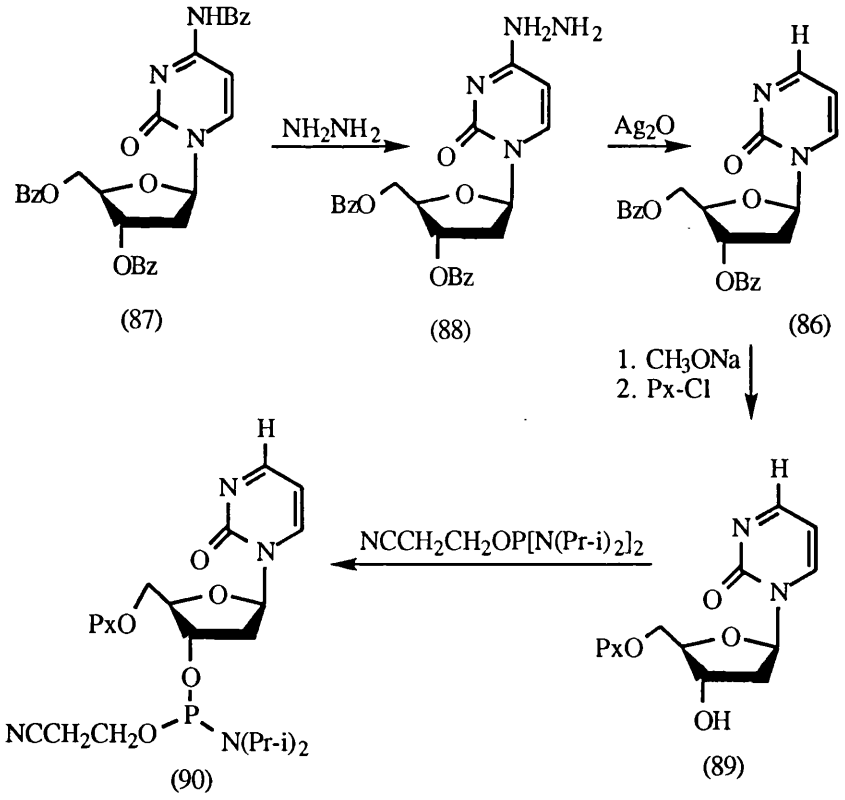


Figure 1.18. Chemical synthesis of phosphoramidite of 2-pyrimidinone nucleoside.

Oligonucleotides containing 5-methyl-2-pyrimidinone nucleoside, which can be considered as a thymidine analogue with the deletion of O^4-carbonyl group, were

prepared by using an approach similar to that of Gildea (see Figure 1.18) (Connolly and Newman, 1989). 5'-DMT-4-thiothymidine (72), prepared as in Figure 1.16, was converted into 4-hydrazino derivative (91) which was oxidized to 5-methyl-2-pyrimidinone nucleoside derivative (92) with silver oxide. Further reaction of (92) with 2-cyanoethyl N, N-diisopropylchlorophosphoramidite yielded phosphoramidite (93) (Figure 1.19) which was then incorporated into DNA.

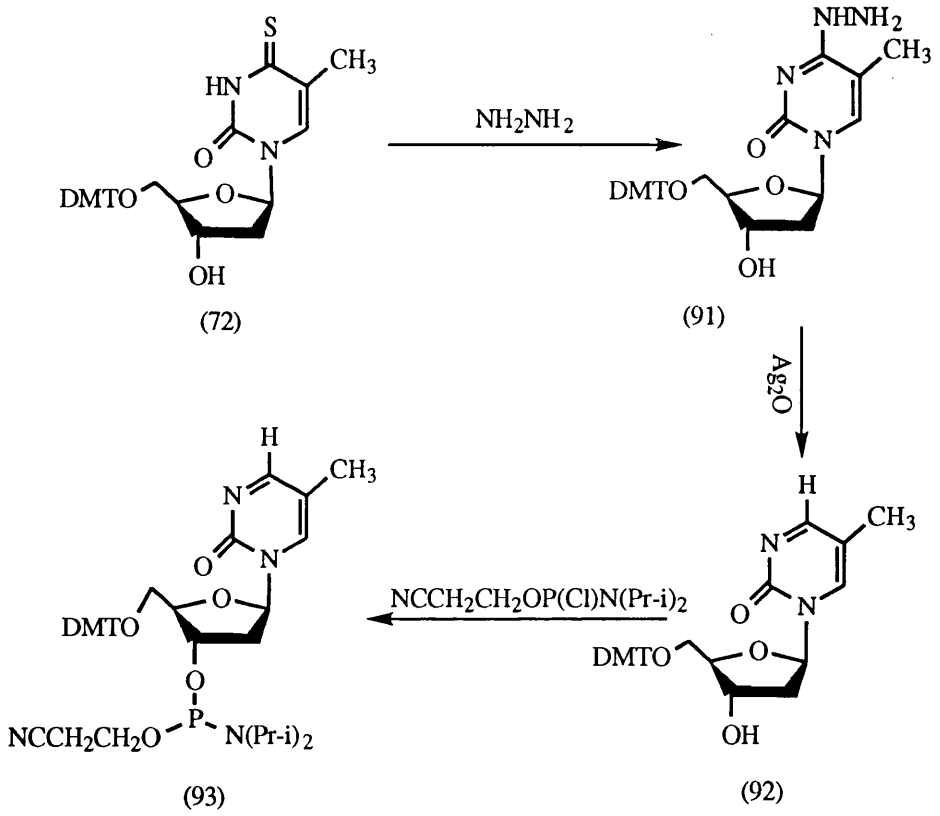


Figure 1.19. Chemical synthesis of phosphoramidite of 5-methyl-2-pyrimidinone nucleoside.

The oligonucleotides containing modified bases discussed above and other modified bases, as well as those commercially available, such as inosine, purine, N^6-methyladenine, 5-bromouracil, uracil, 5-bromocytosine and 5-methylcytosine, have been widely used to investigate DNA-protein interactions, especially DNA-restriction enzyme and DNA-repressor protein interactions. (for a review see Aiken and Gumport,

1991). It is impossible to discuss these investigations here in great detail. However the following examples will be mentioned in order to demonstrate the usefulness of modified base containing oligonucleotides in the study of DNA-protein interactions.

The Eco RI endonuclease is one of the most thoroughly studied restriction enzymes. It recognizes DNA duplexes containing d(GAATTC) and cleaves DNA between the G and the A. Oligonucleotides in which 7-deaza-2'-deoxyguanosine or 7-deaza-2'-deoxyadenosine replaced G or A in the recognition sequence d(GAATTC) were used as probes to investigate their cleavage by the enzyme in order to examine whether the 7-nitrogens of purine residues within the recognition sequence form a hydrogen bond with a particular amino acid of the enzyme as suggested by Rosenberg et al. (1984). The cleavage of the modified oligomers by the Eco RI endonuclease when guanine was replaced by 7-deazaguanine, or when one of the dA residues was replaced by 7-deazaadenine, was strongly decreased compared to the regular DNA-fragment, and replacement of both dAs completely prevented cleavage of the oligomer (Seela and Drill, 1986; Seela and Kehne, 1987). The results clearly supported the suggestion that 7-nitrogen of purines are proton acceptor sites for the endonuclease Eco RI and are both contacted by the protein.

The sequence-specific interaction of Eco RI with DNA has also been investigated by Brennan et al. (1986) with a series of self-complementary octanucleotides and by McLaughlin et al. (1987) with a series of self-complementary decanucleotides, in both cases, containing a base analogue that had one of the potential contact points removed. In similar experiments, Lesser et al. (1990) also looked at the effect of modified base substitutions in only one strand of the duplex recognition site. These experiments demonstrated that the 7-nitrogen and 6-carbonyl group of deoxyguanosine are both contacted by the protein but there is no direct contact to the 2-amino group of the deoxyguanosine. The amino group of dA in position 2 (i.e. G<u>A</u>ATTC) is contacted by the protein while same group in the third position is not important for the endonuclease reaction. Loss of the methyl group of the thymidine in position 4 has little effect on the cleavage but the methyl group of the thymidine at

position 5 makes contact with the protein and therefore it is important for the endonuclease reaction.

Another widely studied restriction endonuclease is Eco RV which recognizes the sequence d(GATATC) in DNA duplexes and cleaves between the two central residues. In order to study the role of each base in the recognition process, Fliess et al. (1988) used oligonucleotides containing modified bases (hydroxanthine, 6-methyladenine, 7-deazaadenine, uracil, 5-bromouracil, 5-methylcytosine, and 5-bromocytosine) each at a unique position in the Eco RV recognition sequence (GATATC). A kinetic comparison between cleavage of the parent and these sequences showed that several functional groups in the recognition sequence are essential for DNA cleavage by the enzyme. This enzyme recognizes its substrate presumably through hydrogen bonds to the exocyclic NH_2 group and the N-7 of the adenine in position 2, the exocyclic NH_2 groups of the adenine in position 4 and the cytosine in position 6, as well as through hydrophobic interaction with both thymidine residues. Since all these groups are in the major groove, and as the 2-NH_2 group of the guanine in position 1 is not essential for cleavage, it was concluded that Eco RV interacts with its substrate via the major groove. Similar investigations carried out by Mazzarelli et al. (1989) and Newman and Connolly (1990), and more recently by Waters and Connolly (1994) demonstrated that the 6-oxygen and 7-nitrogen of dG in the major groove are also vital for DNA recognition and hydrolysis.

The specific interaction of RNA polymerase with promoter DNA is an important step in the expression of genes. The insertion of modified nucleosides in oligonucleotides provides valuable approach to better understand the recognition of promoters by RNA polymerase in bacteria. To identify promoter function groups that effect transcription by RNA polymerase, two adjacent thymine residues were displaced by uracil at position 34 and 35 relative to the transcription initiation site of the bacteriophage lambda P_R promoter (Dubendorff *et al.*, 1987). It was found that the removal of either methyl group led to a 5-fold reduction in the rate of formation of a transcriptionally competent complex, and that the loss of the methyl group at both sites

generated an inactive promoter. Consequently, the methyl group of these specific thymine residues appears critical for the interaction of RNA polymerase with the lambda P_R promoter

In an attempt to better understand the interaction between the trp repressor and its operator sequence, Mazzarelli et al. (1992) prepared fourteen modified DNA sequences by insertion of seven different base analogues (purine, 5-methyl-2-pyrimidone, 2,6-diaminopurine, 7-deazaadenine, uracil, inosine and 5-methylcytosine) at positions -4/+4 and -5/+5 in the trp operator and measured the affinity between the trp repressor and these modified sequences. The results suggested that the carbonyl at dT_{+4} was critical for the formation of the high-affinity sequence-specific complex. Additionally, the thymine methyl group at dT_{+4} and the N-7 nitrogen of dA_{+5} appear to be critical contacts necessary for high-affinity binding by the repressor whereas the adenine amino group at dA_{-4} or dA_{+5} seemed to be not critical for binding.

1.4.1.ii. Photo-Cross-Linking Approach

Photochemical cross-linking has been widely used in the study of DNA-protein interactions. Because irradiation of DNA-protein complexes with UV light produces covalent linkages between amino acid residues and nucleic acid bases with minimal perturbation of the system being studied and cross-linking can involve only those amino acid residues at the interface of DNA-protein complexes, this technique can be used to identify the most important residues in the DNA-protein interaction. The basic procedures of this technique include UV irradiation of a DNA-protein complex, separation and proteolysis of the covalently linked complex, and isolation and sequencing of the peptide-DNA complex.

Covalent cross-linking of protein to normal DNA by short-wave (254 nm) UV irradiation has gained a great deal of popularity because of the simplicity of the procedure. The number of studies utilizing this technique are too numerous to mention here (for a review see: Williams and Konigsberg, 1991). But there are several difficulties if normal DNA is used as a substrate. Since far-UV light is structurally damaging to both nucleic acids and proteins, this approach may lead to cross-linking which lacks biological relevance. The second limitation is the low yield of cross-linking. Therefore, it would be preferable to carry out a photochemical cross-linking experiment with DNA specifically substituted with photo-active bases with a different UV absorption maxima from that of DNA (260 nm) and proteins (280 nm). This might minimize photodamage to DNA and proteins and increase cross-linking efficiency, an extra benefit of this approach is that analysis of cross-linked complexes is greatly facilitated since the site of adduct formation on DNA is already known.

5-Bromouracil is structurally similar to thymine, as the 5-substituent differs little in size with van der Waals radii of bromine and methyl at 1.95Å and 2.00Å, respectively (Hutchinson, 1973). Its phosphoramidite monomer (94) is commercially available and can be easily incorporated into DNA by standard automated synthesis. UV irradiation of 5-bromo-2'-deoxyuridine (5-BrdU)-substituted DNA (95) results in

photodissociation of bromine with the generation of a reactive free radical at C-5 of uracil (96). A properly positioned substituent within a bound protein (97) could accept the free radical and directly cross-link to the DNA to form a covalently joined complex (98) (Figure 1.20). Synthetic oligonucleotides containing this modified base have been used to identify amino acid residues at the interface of protein-DNA complexes on a number of occasions.

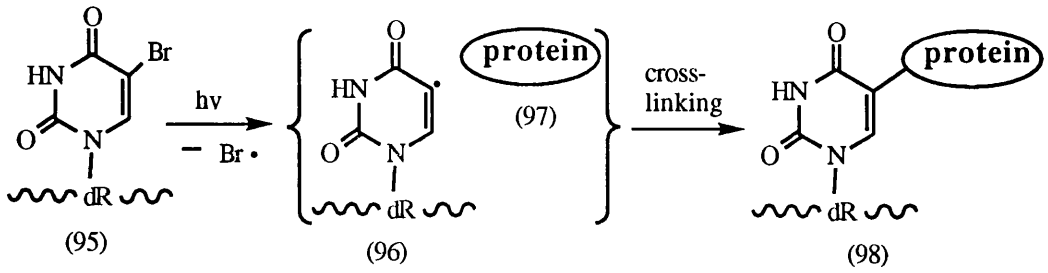
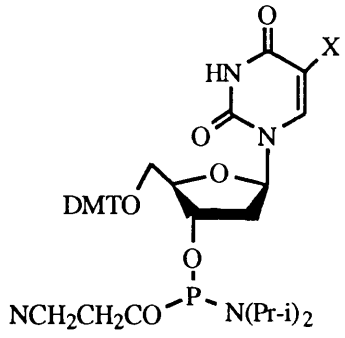


Figure 1.20. Proposed mechanism of photocross-linking of DNA containing 5-BrdU with a protein (Adapted from Blatter *et al.*, 1992)

In order to study interactions between the Eco RI or the Eco RV restriction endonucleases and DNA, Wolfes and coworkers (1986) synthesized self-complementary oligonucleotides containing 5-BrdU in various positions within and outside the recognition sequence. Upon irradiation by near UV light (>300nm) and in the absence of Mg^{2+} they were cross-linked to the enzymes. Although the yield of cross-linking was low, the site of cross-linking of EcoRI was determined to be close to Met-137. Similarly interactions between lac repressor and its operator DNA have been investigated by Matthews et al. (Allen *et al.*, 1991; Wick and Matthews, 1991). Twenty oligonucleotides with the substitution of 5-BrdU for the thymidines within the -6 to +25 region of the lac operator relative to the initiation site for transcription were synthesized. By irradiation of the DNA-lac repressor complexes, five sites of high efficiency cross-linking to the protein were identified. These were at positions +3, +4, +14, +18, and +19 (Wick and Matthews, 1991). Large scale irradiation of the complexes between repressor and each of the five singly substituted operator DNAs

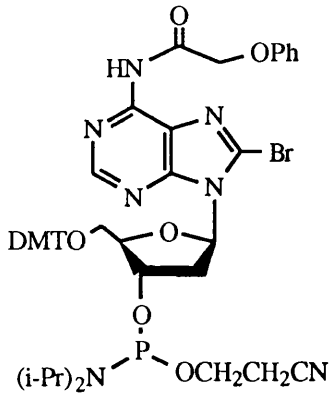
produced sufficient covalent complex for proteolysis, separation of the peptide-DNA, and sequencing of the DNA-bound peptide. The DNAs substituted with 5-BrdU for thymidine at positions +3, +18, and +19 produced cross-links to His-29; and a 5-BrdU at position +14 formed a cross-link to Tyr-17 (Allen *et al.*, 1991). These results indicated that His-29 is in close contact with thymidine at positions of +3, +18, and +19 and Tyr-17 with thymidine at position +14.

Blatter et al.(1992) identified an amino acid-base contact in the GCN4-DNA complex by using a similar approach. An oligonucleotide substituted with 5-BrdU for thymidine at position +3 formed a cross-link with Ala-238. This result was strikingly consistent with the crystallographic structure of the complex. Since this approach needs no prior information regarding the structure of the protein or the structure of the protein-DNA complex, it should be generalizable to DNA-binding proteins that interact with the DNA major groove.



(94) X = Br

(99) X = I



(100)

5-Iodouracil is another photoactive analogue of thymine. The van der Waals radius of iodine is 2.15Å, only 8% larger than the methyl group. Oligomers containing this photoactive base can be obtained by incorporation of its commercially available phosphoramidite (99). It has been shown that substitution of 5-iodo-2'-deoxyuridine (5-IdU) for T did not appreciably disturb the protein-DNA complex and produced higher cross-linking yields than did the substitution of BrdU for T (Willis *et al.*, 1993).

Although its mechanism of photocross-linking is less understood, it is generally assumed that it follows a mechanism similar to that of its bromo counterpart.

The photoactive 8-bromo-2'-deoxyadenosine (8-BrdA) has been synthesized and incorporated into oligonucleotides via the phosphoramidite (100) as a means to identify DNA-amino acid contacts in protein-DNA complexes by photocross-linking (Liu and Verdine, 1992). Preliminary experiments showed that the introduction of a single 8-BrdA residue to oligonucleotides containing a specific binding site of the transcription factor NF-kB did not significantly change the stability of the DNA duplexes or their binding affinity to the protein. Consequently, oligonucleotides containing this modified base may be useful in probing specific contacts in protein-DNA complexes.

4-Thiothymidine and 6-thioguanosine have an λmax at 340 nm and are photoactive at 340-350 nm which is well away from the usual absorption maxima of proteins (280 nm) and nucleic acids (260 nm). Their incorporation into DNA has been discussed earlier (see 1.4.1.i). Irradiation with long wavelength UV light can generate reactive species which will crosslink to proteins. Thus incorporation of these thionucleosides into DNA may allow identification of amino acid residues at the interface of protein-nucleic acid complexes. These bases were incorporated into synthetic oligonucleotides containing the recognition site (GATATC) of the Eco RV endonuclease and methyltransferase (Nikiforov and Connolly, 1992). Upon irradiation with long wavelength UV light (340-360 nm) these oligonucleotides were photochemically crosslinked to both enzymes. The yields were up to 35%. It is anticipated that oligonucleotides containing these two thiobases could find wide application in the study of DNA-binding proteins and related areas of biochemistry and molecular biology.

Oligonucleotides containing azidonucleotides have long been used for the study of protein-DNA interactions (for a recent review see: Sylvers and Wower, 1993). These oligomers are chemically inert until irradiated with long wavelength UV light,

upon which a chemically reactive nitrene is produced. The nitrene can then react rapidly and relatively nonspecifically with adjacent amino acid residues of a bound protein, resulting in covalent attachment of DNA to the protein. This relatively nonspecific reaction makes these azido-DNAs excellent probes for the investigation of DNA-protein interactions since cross-linking to every amino acid could take place and is not confined just to amino acids with some specific functional groups. To date, however, oligonucleotides carrying azidonucleoside in which the azido groups are directly on the nucleoside base have been synthesized exclusively by enzymatic incorporation of azidonucleotides into DNA and hence their synthesis and applications will not be discussed here. There have, however, been a few reports of the chemical synthesis of oligonucleotides containing azidonucleosides in which azido functions are connected to a nucleoside base via a linker.

Gibson and Benkovic (1987) described the synthesis of the nucleoside phosphoramidite (101) and its insertion into oligonucleotides. 5-Iododeoxyuridine (102) was coupled with propargyl phthalimide (103) in the presence of bis(triphenylphosphine)palladium dichloride and CuI to form deoxyuridine derivative (104). Reduction of the acetylenic bond in (104) by H_2 over Pd/C formed the nucleoside (105) which was then converted into phosphoramidite (101) by standard methods and used for automated DNA synthesis. The photoactive azido group was introduced at the oligomer, rather than monomer, level by treatment of deprotected oligomers (106) with the cross-linker (N-hydroxysuccinimidyl)-5-azido-2-nitrobenzoate (Figure 1.21). The azido-bearing oligonucleotides (107) have been used as primers for template-directed DNA synthesis with either the Klenow fragment of E. coli DNA polymerase I, bacteriophage T_4 DNA polymerase, or AMV reverse transcriptase. Brief illumination of the primer elongation mixture with 302 nm light formed covalent complexes between DNA and the polymerases. More detailed examination of photo-cross-linking of the Klenow fragment with the azido-DNA (Catalano *et al.*, 1990) revealed that the protein contacts between five and seven bases of duplex DNA. Proteolytic digestion of the cross-linked DNA-protein complexes

followed by isolating and sequencing DNA-labeled peptides showed that Tyr-766 was the site of cross-linking and thus the catalytic site in the polymerase was localized (Catalano *et al.*, 1990).

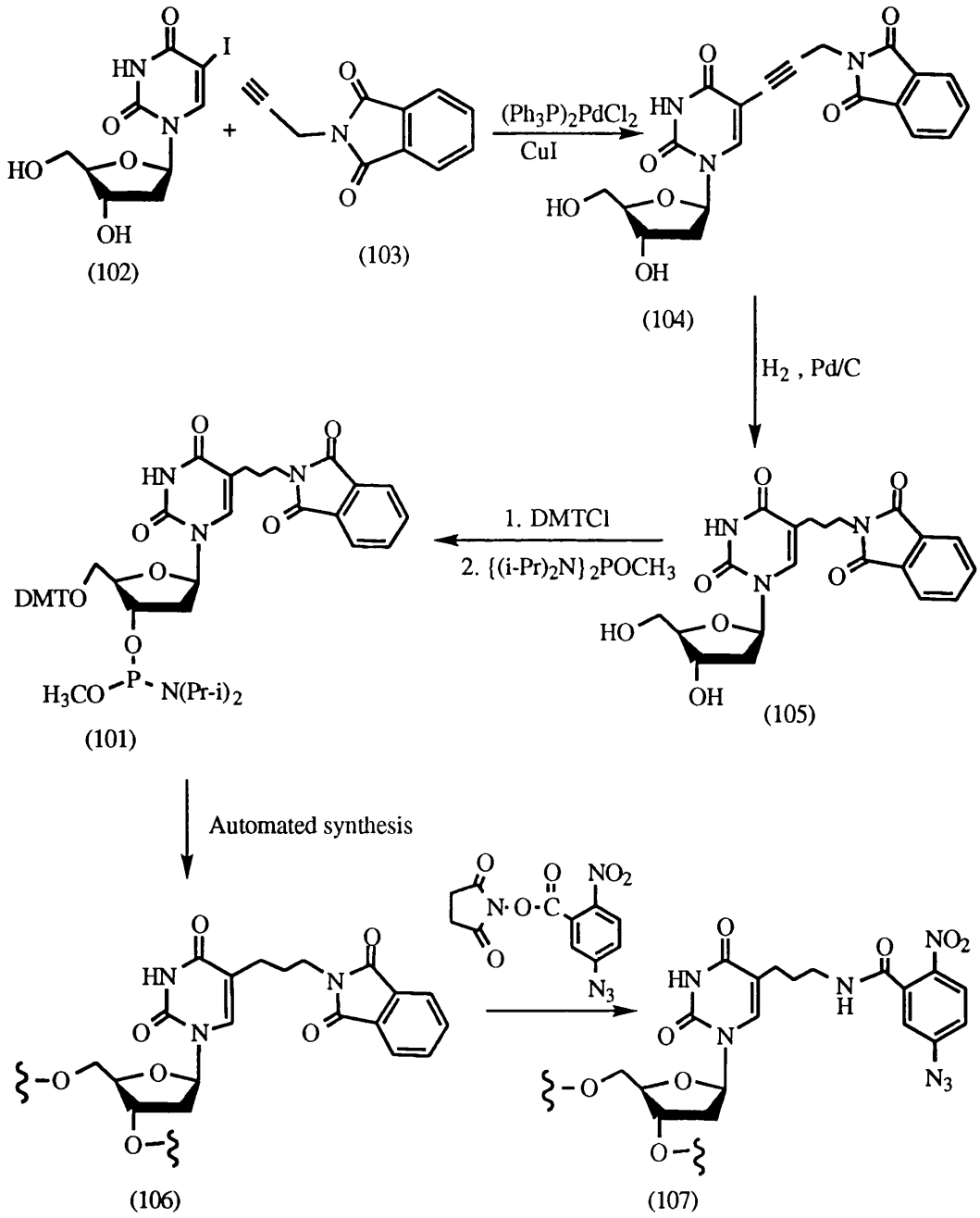


Figure 1.21. Chemical synthesis of azido-labeled oligonucleotide analogues on the 5-position of deoxyuridine via a linker.

Brandley and Hanna (1992) described the synthesis of

5-thiocyanato-2'-deoxyuridine phosphoramidite (108) for the derivatization of oligonucleotides. The phosphoramidite (108) was synthesized by direct thiocyanation of 2'-deoxyuridine (109) with SCNCl followed by tritylation and phosphitylation of the resulting compound (110). Modified oligonucleotides carrying 2'-deoxyuridine derivative (111) were deprotected under standard conditions without affecting the 5-thiocyanato function. Upon treatment with DTT at 55°C, it was reduced to a 5-mercapto function (112) which was subsequently reacted with p-azidophenylacyl bromide to produce oligonucleotides (113) suitable for the study of protein-DNA interactions (Figure 1.22)

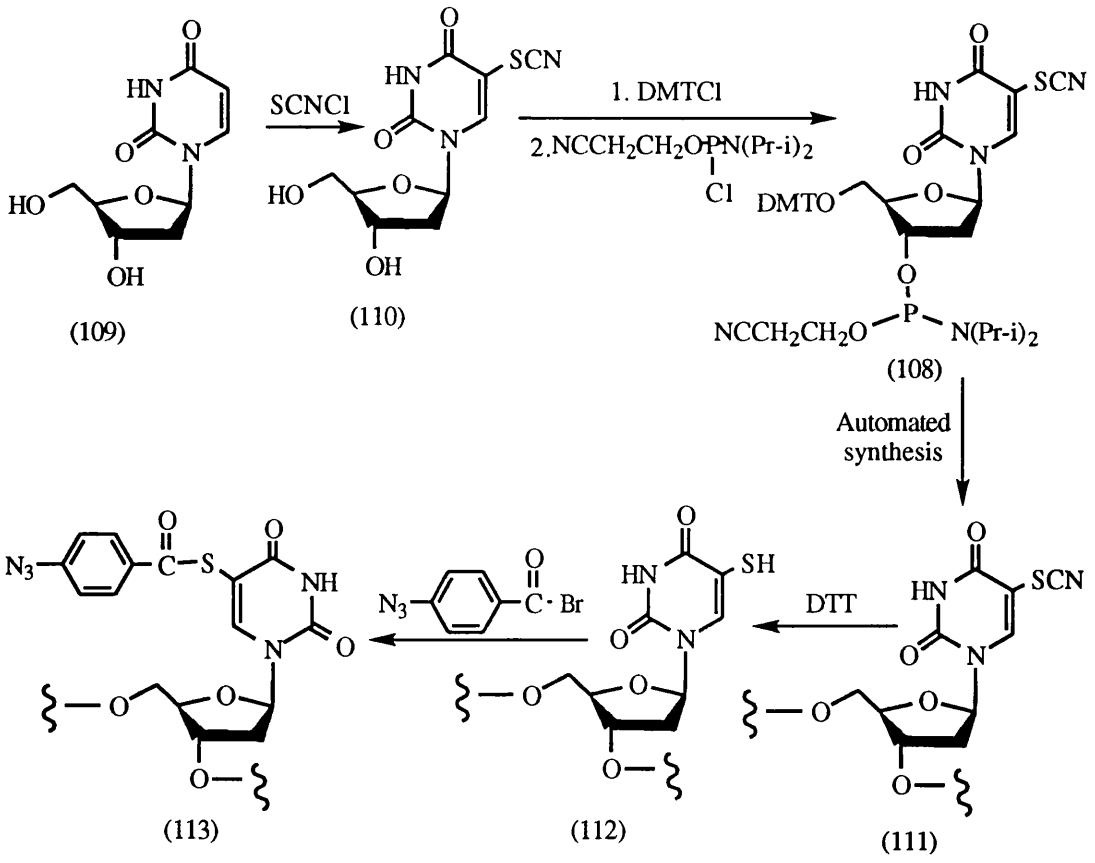


Figure 1.22. Chemical synthesis of oligomers containing a photoreactive aryl azide group linked to the 5-thiol group of 5-thiodeoxyuridine.

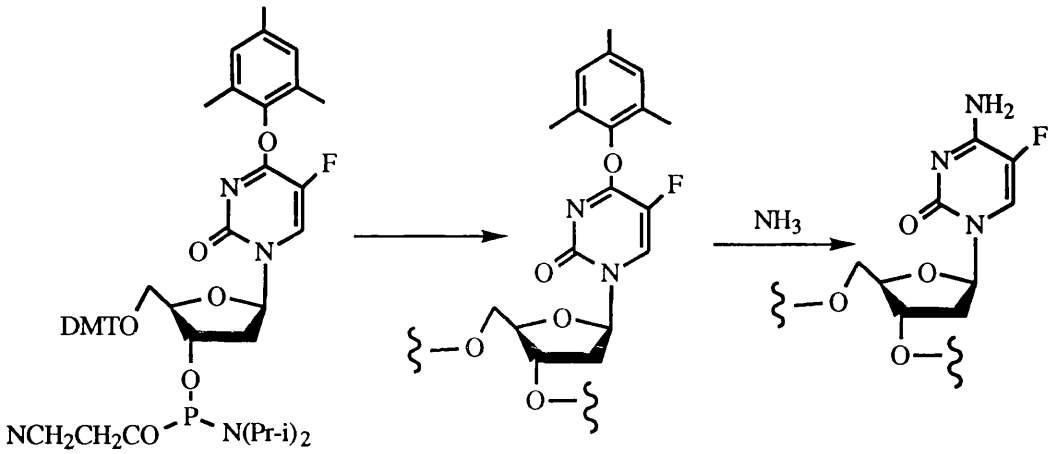
Laser cross-linking of protein-DNA complexes is a recently emerged new technique and becoming increasingly popular for the study of DNA-protein interactions

(for reviews see: Hockensmith *et al.*, 1991; Pashev *et al.*, 1991). It was first demonstrated in 1982 when the complex between E. Coli RNA polymerase and T7 phage DNA was covalently linked by a single 20 ns laser pulse (Harrison *et al.*, 1982). Since lasers are more powerful than traditional UV sources, using lasers as a source of monochromatic UV light means that the number of photons required for the cross-linking may be delivered in nano- or pico-second time intervals. The extremely short cross-linking time means that this technology can be used to study dynamic aspects of rapid protein-DNA interactions under conditions that freeze the "instantaneous" binding equilibria involved and hence to characterize thermodynamically and kinetically the sequential interactions that occur with protein-DNA complexes of biological interest. Besides the extremely short time of irradiation, the laser-induced reactions proceed via higher excited states of the nucleotide bases, which could sharply increase the efficiency of cross-linking. An excellent example is that irradiation of the complex between 5-iodo-2'-deoxyuridine-substituted DNA and bacteriophage R 17 coat protein with monochromatic 325 nm light from a helium cadmium laser led to regiospecific cross-linking at yields as high as 94% (Willis *et al.*, 1993). A more recent example is the identification of three point contacts in the *Oxytricha* teloretic protein-DNA complex by laser photo-cross-linking using 5-bromodeoxyuridine substituted oligomers (Hicke *et al.*, 1994).

1.4.1.iii. Chemical Cross-Linking Approach

The enzymatic addition of methyl groups to DNA is an essential element of genomic function in organisms ranging from bacteria to mammals. For example, in bacteria DNA methylation directs the mismatch repair and restriction-modification systems, which correct errors of replication and prevent transformation by non-self-DNA, respectively. In mammals, methylation of a cytidine residue within a specific gene can inhibit transcription of that gene directly by inhibiting the binding of transcription factors or by affecting the surrounding chromatin structure. The mechanism by which (cytosine-5)-methyl transferases transfer a methyl group from the

cofactor S-adenosyl-L-methionine (SAM) to duplex DNA can be explored with an oligonucleotide specifically modified with the mechanism-based inhibitors, such as 5-fluoro-2'-deoxycytidine, by the chemical cross-linking of DNA and protein.



(114) (115) (116)

Figure 1.23. Chemical synthesis of oligomers containing 5-fluorocytosine by ammonolysis of oligomers carrying 4-O-(2, 4, 6-trimethylphenyl)-5-flurouracil.

To determine the active site nucleophile of DNA (cytosine-5)-methyltransferase M.Hae III, an enzyme that recognizes dyad-symmetric site d(GGCC) and carries out two sequential methylations of specific cytosines, an oligonucleotide containing a 5-fluorocytosine at the expected M.Hae III methylation site (5'-CGCATAGG (5-F-DC)CATGACG-3') was prepared by the incorporation of the 4-O-(2,4,6 -trimethylphenyl)-5-fluoro-2'-deoxyuridine phosphoramidite (114) by conventional solid-phase synthesis (Chen *et al.*, 1991; MacMillan, 1992). Upon removal of the protecting groups with conc ammonia, the oligomer (115) containing 4-O-(2,4,6 -trimethylphenyl)-5-fluorouracil residue was transformed into an oligomer carrying 5-fluorocytosine (116) in quantitative yields (Figure 1.23). The purified oligonucleotide was annealed with a complementary strand carrying a 5-methylcytosine residue (3'-GCGTATC(5-Me-dC)GGTACTGC-5') and then incubated with M.Hae III along with the cofactor S-adenosyl-L-methionine. The cross-linked complex was purified and

submitted to proteolysis. Sequencing of the proteolytic fragments revealed that Cys-71 of M.Hae III formed a covalent bond with DNA during catalytic methyl transfer (Chen *et al.*, 1991).

The preparation of the oligonucleotides containing 5-F-dC was independently reported by Schmidt et al. (1992) to investigate the catalytic mechanism whereby the DNA cytosine methyltransferase of E.coli K-12 methylates cytosine residues at C-5 in a sequence-specific manner. Two alternative approaches have been developed, one of which proceeds via a fully protected phosphoramidite of 5-fluoro-2'-deoxycytidine (117) and the other via a fully protected phosphoramidite of 5-fluoro-4-methylmercapto-2'-deoxyuridine (118). Since the electron-withdrawing effect of the fluoro group increases lability of the benzoyl group bound to the exocyclic amino function of the 5-fluorocytosine ring, it can be removed completely by treatment with concentrated ammonia at room temperature overnight without damaging 5-fluoro-2'-deoxycytidine. The methylmercapto function of (117) was replaced by an amino group in the final ammonia treatment used for cleavage from support and base deprotection. Sequence-specific covalent cross-linking of the wild enzyme to synthetic oligonucleotides containing 5-fluoro-2'-deoxycytidine was demonstrated (Hanck *et al.*, 1993). However, the cross-linking reaction was totally abolished with a mutant enzyme in which the cysteine sulfhydryl function at position 177 was replaced by a serine hydroxyl group. These results strongly support a catalytic mechanism in which the sulfhydryl group of cysteine residue 177 undergoes Michael addition to the C^5-C^6 double bond, thus activating position C-5 of the substrate DNA cytosine residue for electrophilic attack by the methyl group donor SAM (Figure 1.24). Direct support for this mechanism has also been provided by the isolation and characterization of the covalent intermediates formed by the M. Eco RII (Friedman and Ansari, 1992) or the human methyltransferase (Smith *et al.*, 1992) with DNA.

It is worthwhile to point out that the 5-fluoro-2'-deoxycytidine can be incorporated into oligonucleotides with its phosphoramidite (119) without protecting its exocyclic amino function (Marasco and Sufrin, 1992).

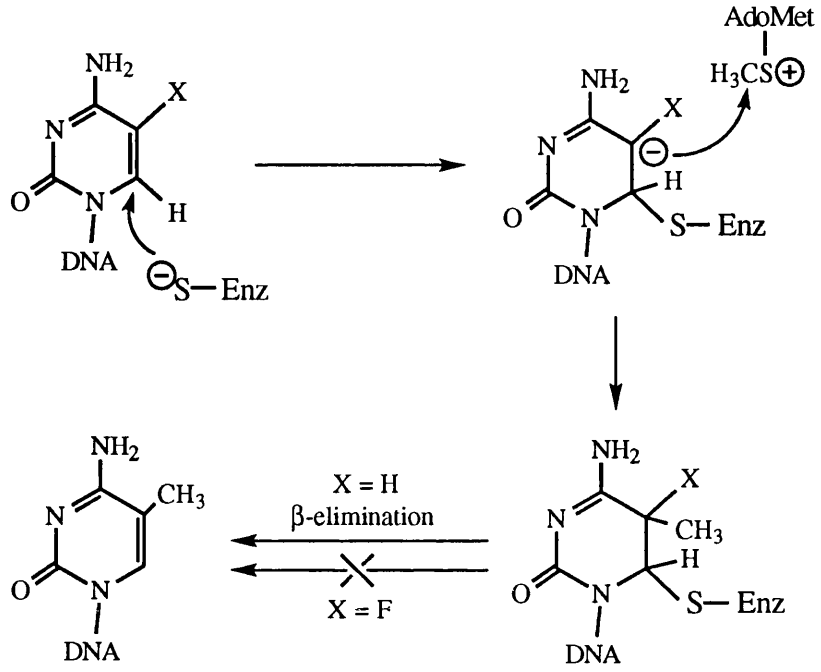
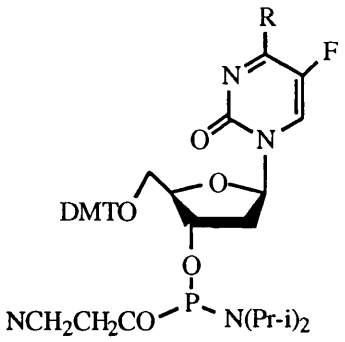


Figure 1.24. Proposed catalytic mechanism of enzymatic DNA cytosine-C^5 methyltransferase. If a 5-fluorocytosine residue (X = F) is attacked by the enzyme, the final step of the reaction is blocked and a stable covalent adduct of enzyme and substrate DNA results. (Adapted from Hanck *et al.*, 1993).



(117) R = NHBz

(118) R = SCH_3

(119) R = NH_2

Thanks partially to the availability of 5-FdC containing oligonucleotides, the structure of a chemically trapped covalent reaction intermediate has been determined by X-ray crystallography between the Hhal DNA cytosine-5-methyltransferase, S-adenosyl-L-homocysteine, and a 13-mer oligonucleotide duplex containing methylated 5-fluorocytosine at its target (Klimasauakas *et al.*, 1994). The most spectacular feature of the complex structure is the location of the target cytosine, which

is completely looped out from the DNA double helix and firmly bound in a pocket of the enzyme's catalytic domain (Figure 25). It is generally believed that this unexpected remarkable discovery not only solved the structure of this particular complex but also revealed insight into a problem confronted by many catalytic DNA-binding proteins: how to assemble an active site around a reactive centre that is deeply buried within the structure of canonical B-DNA (Verdine, 1994). The significance of this work has been further highlighted by the awarding of the Nobel prize to one of the main contributors of this discovery (Roberts, 1994).

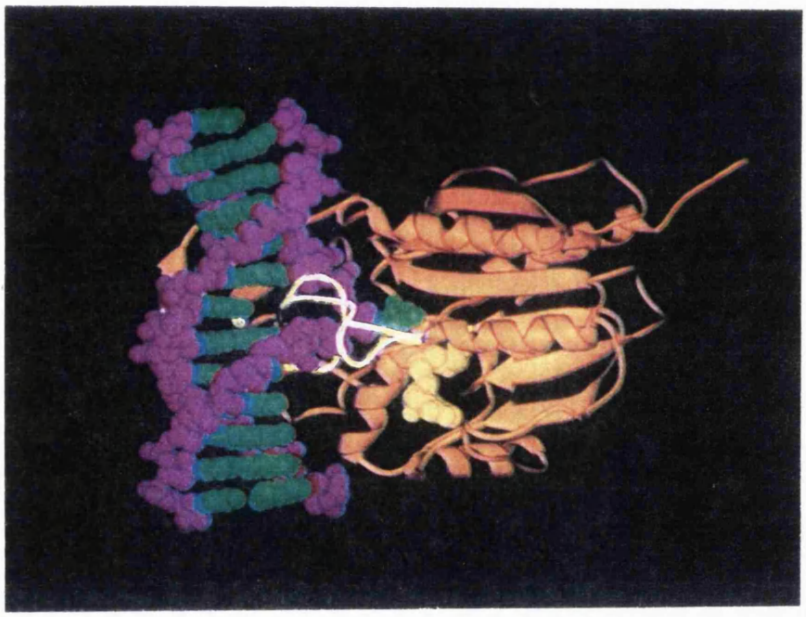


Figure 1.25. Graphic representation of the complex of M. Hhal covalently bound to a 13-mer DNA duplex containing methylated 5-fluorocytosine. The protein is in brown, the DNA backbone is in magenta and the DNA bases are in green, and the active-site loop and the two recognition loops in white. The end product of the reaction, S-adenosyl-L-homocysteine, is in yellow (Taken from Klimasauakas *et al.*, 1994).

1.4.1.iv. Site-Specific Cross-Linking of DNA

Site-specific cross-linking DNA is a potentially valuable tool for the study of the interactions between DNA and those proteins which require strand separation or local helical distortion of DNA to perform their functions.

Webb and Matteucci (1986a; 1986b) inserted the modified nucleoside phosphoramidite (124a-b) into oligonucleotides by solid-phase methods. The CPG-bound oligomers (125a-b) containing triazoyl nucleobases were converted to their corresponding ethyleneimino derivatives (126a-b) upon sequential treatment with ethyleneimine, DBU in acetonitrile, and finally conc ammonia (Figure 1.26). These oligomers formed cross-linked duplexes with their complementary oligonucleotides. These cross-linked duplexes have been used to determine the structural requirements for the exonuclease and polymerase activities of Klenow DNA polymerase (Cowart *et al.*, 1989). It was demonstrated that the polymerase action of Klenow fragment did not require that the DNA duplex undergo strand separation for activity, whereas the exonuclease site required that at least four base pairs of the primer strand must melt out in order for the enzyme to be able to remove a base pair through exonuclease activity. These results showed that the polymerase and exonuclease sites of the Klenow fragment are physically separate and have different substrate structural requirements.

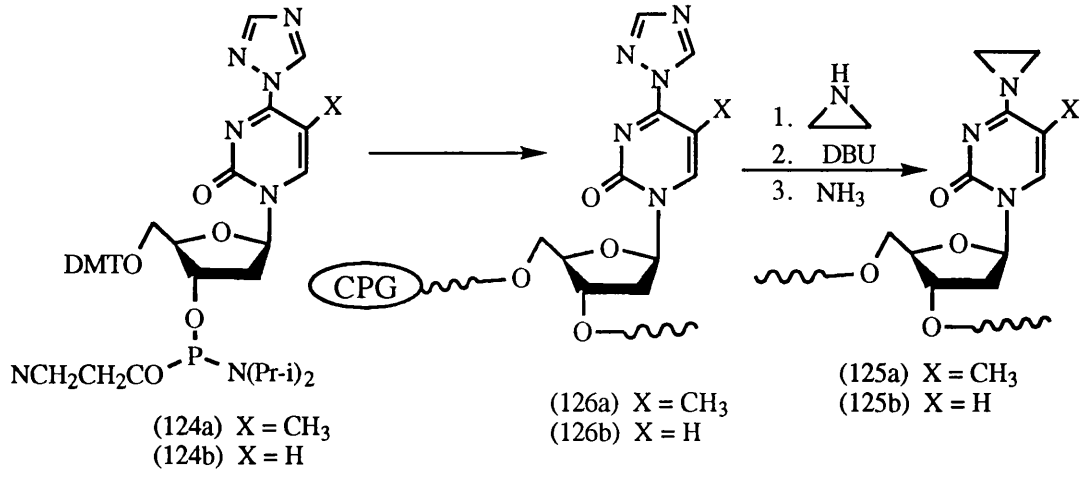


Figure 1.26. Chemical synthesis of oligonucleotides containing N^4, N^4-ethanocytosine or 5-methyl-N^4, N^4-ethanocytosine.

Site-specific cross-linking of DNA via formation of an interstrand disulphide bond between alkane thiol tethers attached to the exocyclic amino group of adenines located on the opposite strand of the DNA duplex has been described by Ferentz et al. (1991; 1993). In their method, oligonucleotides containing O^6-phenyl-2'-deoxyinosine (127) were treated with alkylamines to convert the O^6-phenyl-dI moiety to an adenine analogue with an alkane thiol tether on the amino group (128). After removal of the protecting group on the thiol function, the oligomers (129) were oxidized under aerobic conditions to form an inter-oligonucleotide disulphide bond (130) (Figure 1.27). NMR and CD spectroscopy indicated that these cross-linked duplexes retained the structure of native B-DNA. However, cross-linking increased the Tm of oligonucleotides by 15-21°C. Since these unstrained, cross-linked DNAs are both kinetically and thermodynamically resistant to reduction, they should facilitate studies of enzyme-mediated unpairing process such as transcription, replication, and recombination (Ferentz and Verdine, 1991; Ferentz *et al.*, 1993).

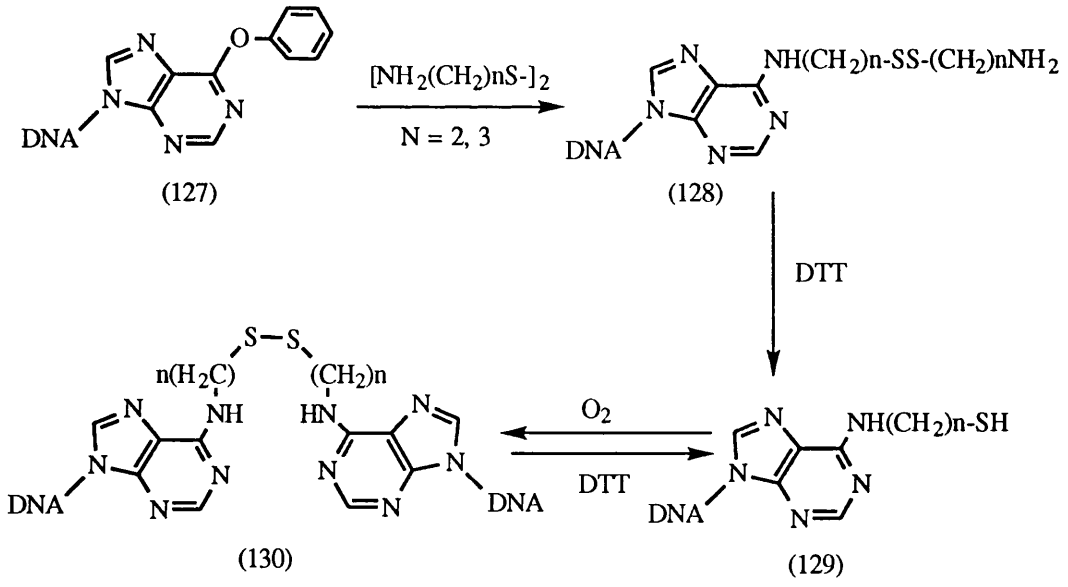
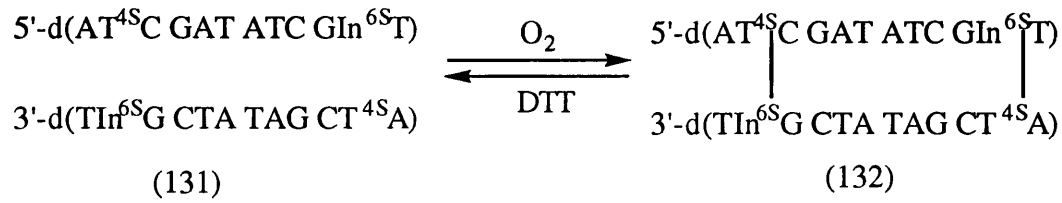


Figure 1.27. Chemical synthesis of the site-specific cross-linked duplex via formation of an interstrand disulphide bond.

Cross-linking of DNA through the formation a disulphide bridge directly between the bases has been reported by Milton et al. (1993). A self-complementary

DNA duplex (131) containing 6-thio-2'-deoxyinosine and 4-thiothymidine in 50 mM $NaHCO_3$ at 5°C underwent cross-linking under aerobic conditions to produce the duplex linked by two disulphide bonds (132) (Figure 1.28). The cross-linking was essentially complete after 5 days and could be reduced to the original oligomer by treatment with dithiothreitol.



(131) (132)

Figure 1.28. Cross-linking of DNA duplex by the formation of disulphide bond directly between thiobases.

1.4.1.v. Oligomers Labeled With NMR-Sensitive Atoms

NMR is a powerful technique for studying DNA-protein interactions. However, ^{1}H-NMR studies of DNA-protein complexes are complicated by the overlap of resonances as a consequence of the large number of peaks and broad lines, due to the size of the complexes. NMR spectra of protein-DNA complexes can be greatly simplified by the use of selective isotopic labeling of DNA.

A convenient synthesis of oligonucleotides in which the amino group of a cytosine was labeled with nitrogen-15 has been reported by Kellenbach (1991). Triazolo deoxyuridine phosphoramidite (124b) was incorporated into oligonucleotides and subsequently the modified oligomers were treated with concentrated ^{15}N-ammonia to give DNA segments containing $^{15}N^4$-cytosine residues. NMR spectra of the oligonucleotide corresponding to the glucocorticoid response element indicated that the ^{15}N-chemical shift of the exocyclic amino function of cytosine of about 98 ppm is different from the nitrogen chemical shifts of the protein backbone which typically resonate at 105-130 ppm. Those oligomers should consequently facilitate the study of protein-DNA interactions. Similarly, in order to resolve the issue of whether the

N-terminal arm of λ repressor contacts the nonconsensus operator half-site of OL1 operator DNA, which even X-ray crystallographic studies failed to resolve (Clarke *et al.*, 1991), two 21-mer oligomers containing the OL1 site-specifically labeled with $^{15}N^4$-cytosine at C-8 and C-8' positions respectively were synthesized and annealed to the complementary, nonmodified strands to furnish two duplex 21-mer oligonucleotides which differed only in the position of their ^{15}N label within the OL1 site (MacMillan *et al.*, 1993). NMR study of the binding complex between λ repressor and these duplexes showed that although the protein's N-terminal contacts the consensus operator half-site, as suggested by X-ray structure of the complex (Clarke *et al.*, 1991), it does not interact with the nonconsensus half-site.

The methyl group of thymine plays a vital role in the hydrophobic interaction between DNA and proteins. In ^1H-NMR spectra of protein-DNA complexes, however, it is difficult to observe cross-peaks between the thymine methyl group and aliphatic protein side-chains due to severe overlap of resonances in a crowded part of the spectrum. Isotope labeling of the methyl group with ^{13}C along with heteronuclear editing techniques will help to overcome these problems. In an effort to understand sequence-specific DNA recognition of the glucocorticoid receptor, Kellenbach et al. (1992) synthesized ^{13}C-labeled thymidine phosphoramidite and inserted it into an oligonucleotide corresponding to the stronger binding half-site of the consensus glucocorticoid response element. NMR of the complex between the labeled oligomer and the glucocorticoid receptor DNA-binding domain revealed a hydrophobic contact between the labeled thymidine group and the methyl group of a valine residue. These data indicated that stable isotope labeling of major groove functional groups of DNA can simplify the NMR interpretation of protein-DNA complexes and thus provide more insight into the mechanism underlying the sequence specificity of protein-DNA recognition.

1.4.2. Oligonucleotides for Study of DNA Damage Produced by Alkylating Agents

Over the past two decades, much effort has been directed toward the identification of the covalent chemical-DNA adducts and photoproducts formed in DNA of cells and tissues treated with oncogenes and carcinogens. However genotoxic agents such as chemical carcinogens and radiation commonly generate more than one, and usually many, different DNA modifications (Singer and Kasmierek, 1982). Thus it was very difficult to assess the impact of any individual adduct on DNA structure and the contribution of each adduct to the spectrum of mutations induced by chemical or radiation treatment, or to assess the role of each of DNA adducts in the cytotoxicity, or to identify the DNA repair proteins responsible for protecting cells from each specific form of DNA damage. However, the development of methods of synthesis of oligonucleotides containing DNA adducts at defined sites has made it possible to explore the contribution of individual DNA adducts to these biological effects.

DNA damage can be caused by DNA exposure to alkylating agents, UV light, reactive oxygen radicals, and polycyclic aromatic hydrocarbons. Because of our great interests on the DNA damage caused by alkylating agents, especially by the carcinogenic N-nitroso compounds, only the application of modified oligomers for the study of this kind of damage will be discussed.

The carcinogenicity and mutagenicity of N-nitroso compounds are believed to be result from their alkylation of DNA, in particular, alkylation of the O^6-position of guanine and O^4-position of thymine. These alkylated bases lose an imino proton at N-1 position of guanine or N-3 position of thymine and therefore there is a complete change in the pattern of hydrogen bonding in base-pairing and result in the production of GT-AT and AT-GC transition mutations (reviewed by Swann, 1990). To investigate the biological and structural role of these alkylated bases in chemical carcinogenesis, the incorporation of these analogues into oligonucleotides has been undertaken.

The first chemical synthesis of an oligonucleotide containing O^6-methylguanine was reported by Fowler et al. (1982) by using the phosphotriester procedure in

solution. Since their approach was difficult and tricky both in preparation of the modified nucleoside and in its incorporation into oligomers, it was superseded when Gaffney et al. (1984) reported the preparation of O^6-methylguanine containing oligomers by phosphoramidite chemistry. Their approach for making the modified nucleoside consisted of the sulphonation of the O^6-position of a fully protected 2'-deoxyguanosine (133) followed by successive displacement with an amine and then alkoxide ion. The modified nucleoside (134) was selectively deprotected and then sequentially treated with 4, 4'-dimethoxytrityl chloride and methyl N, N, N', N'-tetraisopropylphosphorodiamidite (Figure 1.29). The phosphoramidite (135) was

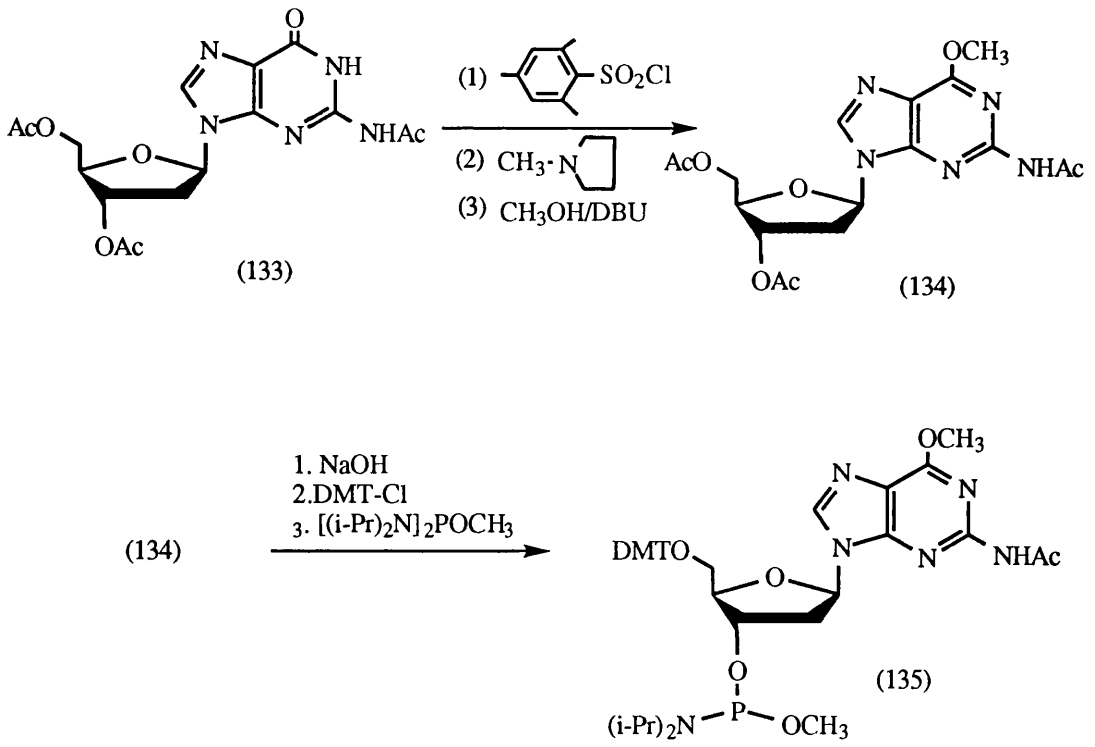
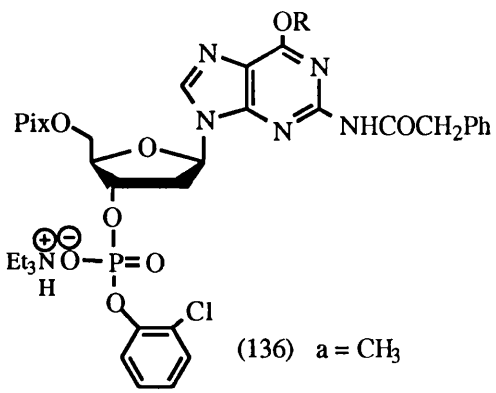


Figure 1.29. Chemical synthesis of O^6-methyl-2'-deoxyguanosine phosphoramidite.

incorporated into oligonucleotides with coupling efficiency of 91%. The acetyl group on the N^2-position of the modified base was removed by a 3-day treatment with methanol/DBU. To define the effects of O^6-alkylation of guanine on DNA structure and stability, a set of self-complementary dodecamers d(CGNGAATTC(O^6Me)GCG),

where N = A, C, G, or T, was prepared and used to study the thermostability of DNA duplexes containing O^6-methylG (Gaffney *et al.*, 1984). It was found that each sequence formed a stable duplex but with a Tm between 19°C and 26°C lower than the Tm of the corresponding parent oligomers with overall order of N = C > A > G > T. These data indicated that the O^6-methylation of guanine generated regions of localized instability in DNA regardless of the base opposite the lesion. Such instability may disrupt critical regulatory events and may be as significant as or more significant than is the mutation itself to the oncogenic process initiated by alkylating agents.

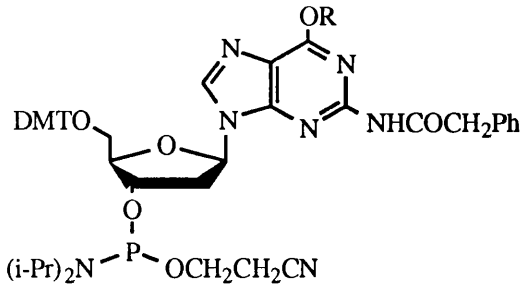
Using a similar approach for synthesis of O^6-alkylguanosine as above, Li and Swann (1989) successfully synthesized oligonucleotides containing O^6-methyl, O^6-ethyl-, and O^6-isopropylguanine. The modified monomers (136a-c) were incorporated into oligonucleotides by the phosphotriester chemistry in solution in order to obtain sufficient material needed for structure studies using 2D NMR and X-ray crystallography. The success of this approach for making oligomers containing O^6-alkylguanine with longer alkyl groups should mainly be attributed to the introduction of phenylacetyl group to protect the N^2-position of O^6-alkylguanine because it could be removed easily. Only a short (30 h) exposure to ammonia at room temperature was necessary for deblocking and therefore the risk of formation of 2, 6-diaminopurine by displacement of the alkoxy group from O^6-alkylguanine by ammonia was greatly reduced. The phenylacetyl group was later used for the protection of N^2-position of O^6-alkylguanine phosphoramidites (137a-c) for routine solid-phase synthesis of the oligomers containing O^6-alkylguanine (Smith *et al.*, 1990). The combined use of phenylacetyl-protected alkylguanine with the easily deblocked monomers, such as PAC phosphoramidites (Pharmacia), is now the preferred method for the synthesis (Smith *et al.*, 1990).



(136) a = CH_3

b = CH_3CH_2

c =$(CH_3)_2CH$



(137) a = CH_3

b = CH_3CH_2

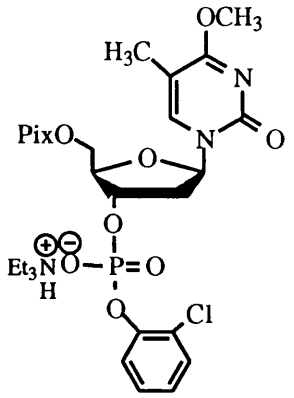
c =$(CH_3)_2CH$

Because O^4-alkylthymines are sensitive to acids, bases, and nucleophiles such as thiols or amines, their incorporation into synthetic DNA has been a difficult task. In spite of these problems, Li et al. (1987) successfully developed a procedure for the synthesis of oligonucleotides containing the mutagenic O^4-methylthymine lesion. The modified base was inserted in oligonucleotides of up to 12 bases in length by the phosphotriester approach in solution via the monomer (138). Methoxide ion (methanol/DBU) rather than ammonia was used to remove the N-acyl protecting groups from the oligonucleotide in order to overcome the problems arising from the susceptibility of O^4-methylthymine toward nucleophilic attack. Enzymic hydrolysis of the oligonucleotides and chromatography of the resultant nucleosides gave no evidence of impurities arising from side reactions during the condensation and deprotection procedures.

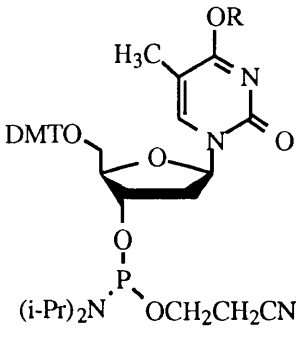
Fernandez-Forner *et al.* (1990) developed a procedure for the solid-phase synthesis of oligonucleotides containing O^4-ethylthymine. Their approach entailed the preparation and utilization of the O^4-ethylthymidine phosphoramidite (139b) in conjunction with nucleoside phosphoramidites protected with the p-nitrophenyl-ethoxycarbonyl group. Due to the lability of O^4-ethylthymine to ammonia, the modified oligomers were deprotected by treatment with 0.5 M DBU in pyridine to remove nucleobase and phosphate protecting groups and then with 0.5 M DBU in

ethanol/pyridine (1:1) to liberate the oligomer from the solid support.

A simpler method for the synthesis of oligonucleotides containing O^4-alkylthymine via nucleoside phosphoramidite derivatives (139a-b) was reported by Xu and Swann (1990). The phosphoramidites (139a-b) were inserted in oligonucleotides of up to 48 bases in length by routine solid-phase synthesis with coupling efficiency of more than 98%. The PAC monomers were used to allow the complete deprotection of oligomers with methanol/DBU or ethanol/DBU and avoid the use of ammonia which otherwise will attack the O^4-alkylthymines to form 5-methylcytosines.

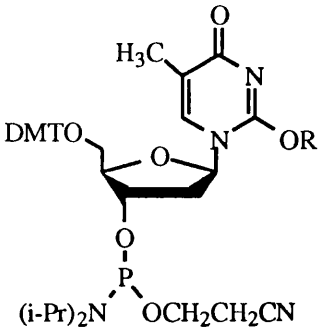


(138)



(139) a. R = Me

b. R = Et



(140) a. R = Me

b. R = Et

The synthesis of oligomers carrying O^2-alkylthymine, which is also produced in DNA after exposure to alkylnitrosoureas, was reported by Bhanot et al. (1992) and more recently by Xu and Swann (1994) by the insertion of the phosphoramidites (140a-b). The purified oligomers containing O^2-ethylthymine were used as template by T7 DNA polymerase to investigate the mutagenic specificity of O^2-ethylthymine *in vitro* (Bhanot *et al.*, 1992). DNA sequencing revealed that both dA and dT were incorporated opposite O^2-ethylthymine residue. While the incorporation of dA impeded the progress of DNA synthesis, the incorporation of dT residue resulted in efficient chain extension. These data provided a basis for understanding a molecular mechanism whereby

ethylating agents such as N-ethyl-N-nitrosourea can induce AT transversion mutagenesis and activate protooncogenes by an AT-TA transversion event in tumours induced by these agents.

Oligonucleotides containing O^6-alkylguanine or O^4-alkylthymine made by the methods above have been widely used to study the biological and structural effects of these lesions (Swann, 1990). The number of these studies are too numerous to list here, but their use to provide answers to two important questions is sufficiently interesting to be mentioned briefly.

It is now clear that the formation of O^6-alkylguanine in DNA is the most important factor in the carcinogenic action of nitrosamines. The O^6-alkylguanine is promutagenic, that is to say that when DNA synthesis occurs using DNA containing O^6-alkylG as template then T is incorporated opposite the alkylated guanine resulting, after second round of replication, in an AT pair in place of the original GC pair (i.e. a GC to AT transition mutation). It is the mutation in protooncogenes such as ras which appear to be crucial. The cell protects itself against the nitrosamines by a DNA repair enzyme O^6-alkylguanine-DNA-alkyltransferase so that there is a competition between the formation of the mutations and the repair. If the O^6-alkylguanine persists in specific codons of protooncogenes and then miscodes on replication there is a cancer; if the alkyltransferase removes the alkyl group before replication no harm occurs. Mammary tumours and skin tumours induced by N-methyl-N-nitrosourea have a G:C to A:T transition mutation which, it is believed, is caused by alkylation of the O^6-position of guanine. But this mutation only occurs in the second G of codon 12 (5'---GGA---3') of the H-ras. The selectivity for this second G has puzzled molecular biologists for years but despite many studies there was no explanation for it. There are three simple probable explanations: (a) the central G is a preferred position for alkylation; (b) the O^6-methylguanine at this position may not be repaired by the cell; and (c) an O^6-alkylguanine flanking by this sequence miscodes with much higher frequency than in other sequences. By measuring the rates of repair by the E.coli a d a

O^6-alkylguanine-DNA-alkyltransferase of an O^6-methylguanine in different positions of chemically synthesized DNA duplexes having the H-ras sequence, it was shown that the most important reason for this selectivity is that the O^6-methylguanine in the central position was by far the least well repaired (Georgiadis *et al.*, 1991).

For many years it was believed that the GC to AT transition mutations were caused by the formation of stable hydrogen bonds between O^6-alkylG and T. However, measurement of melting curve of DNA duplexes containing O^6-alkylguanine has shown that alkylG-C pair was more stable than alkylG-T pair (Li and Swann, 1989). This raised the question: why is T incorporated by the DNA polymerase opposite the O^6-alkylG. Two possible explanations based on structure studies of oligomer containing these modified bases have been proposed to account for the transition mutations (Li *et al.*, 1988; Kalnik *et al.*, 1989a; Kalnik *et al.*, 1989b). The first is that the polymerase mistakes O^6-alkylguanine in the template strand for adenine because of the physical similarity of these bases. The second is that the important factor in the miscoding is that the alkylG-T mispair retains the Waston-Crick alignment with N1 of the purine juxtaposed to N3 of the pyrimidine while the alkylG-C adopts a Wobble conformation which is accompanied by the distortion of the phosphodiester links on both 3' and 5' to the C. Very recently steady state and pre-steady state studies on the incorporation of thymine and cytosine opposite O^6-methylguanine in the template DNA strand by Klenow fragment of E.coli polymerase I showed that the rate of the formation of phosphodiester bond of thymine opposite O^6-methylguanine was four times of that of formation of the phosphodiester bond of cytosine opposite O^6-methylguanine (Tan *et al.*, 1994) which is consistent with the view that the effect of the phosphodiester bond on 3' to the incoming C is the critical factor in the mechanism of the mutagenesis.

1.4.3. As Antisense Oligonucleotides for Therapeutic Purposes

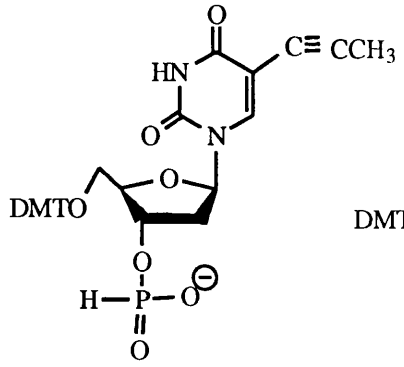
The use of synthetic antisense oligonucleotides for therapeutic purposes was

first proposed in 1978 by Zamecnik and Stephenson (1978). They were able to inhibit the growth of Rous Sarcoma virus in cell culture with a 13-mer oligonucleotide that was complementary to the RNA of this virus. The specific inhibition is based on Watson-Crick base pairing between the heterocyclic bases of the antisense oligonucleotide and of viral nucleic acid.

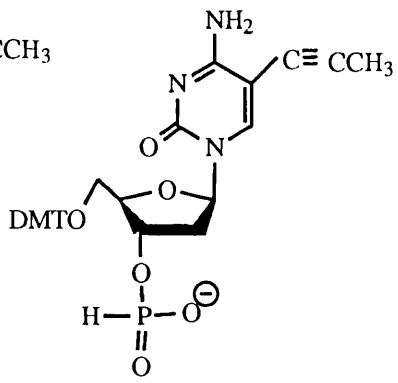
For the antisense oligonucleotide principle to be put into practice the nucleic acid derivatives employed must comply with the following requirements: (a) the complex formed between the oligonucleotide and its complementary target sequence must be sufficiently stable under physiological conditions; (b) the oligonucleotide must have a sufficiently long half-life under *in vivo* conditions for it to be able to display its desired action in the cell, it must therefore be resistant to enzymes (nucleases) that degrade nucleic acids; (c) the oligonucleotide must be able to pass through the cell membranes to reach its site of action; (d) the interaction between the oligonucleotide and its target sequence must be specific. Although the requirements for specificity and binding affinity could be achieved by unmodified oligonucleotides, adequate stability and sufficient passage through cell membranes can not be satisfactorily met by normal oligonucleotides. Therefore it is necessary for normal oligonucleotides to be chemically modified in a suitable manner in order to meet all these requirements. There have been several methods of modification of normal oligonucleotides for this purpose, including the modification of phosphodiester backbone, heterocyclic bases, and sugar (for reviews see: Uhlmann and Peyman, 1990; Englisch and Gauss, 1991; Milligan *et al.*, 1993; Varma, 1993).

Modifying the nucleoside bases is considered as a challenge since base pairing must be maintained through a combination of proper hydrogen bonding and base stacking interactions. However, the binding of antisense oligonucleotides could be increased by hydrophobic modifications at the 5-position of pyrimidines without affecting base pairing. A good example is the oligonucleotides containing C-5 propyne analogues of 2'-deoxyuridine and 2'-deoxycytidine which were synthesized by insertion of H-phosphonate monomers (141-142) (Froehler *et al.*, 1992). This

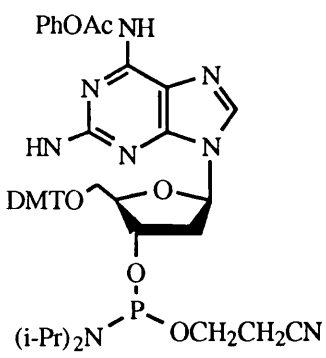
modification not only greatly increased DNA duplex stability as measured by Tm, but also enhanced the Tm of oligomers bound to RNA, as compared with oligomers containing T and 5-methyl-2'-deoxycytidine for both phosphodiester and phosphorothioate backbones. It was suggested that the increased duplex stability is because the propyne is planar with respect to the heterocycle and allows for increased base stacking; furthermore, the propyne is more hydrophobic than a methyl group to potentially allow for a further increase in the entropy of binding. This increased hydrophobicity may also facilitate passive diffusion of oligomers into cells, making these oligonucleotides especially useful in therapeutic applications (Froehler *et al.*, 1992). Indeed, phosphorothioate oligonucleotides containing these modified bases have shown gene-specific antisense inhibition at nanomolar concentrations of the oligonucleotides (Wagner *et al.*, 1993). Therefore, these oligonucleotides may have important applications in therapy and in studies of gene function.



(141)



(142)



(143)

The binding affinity of antisense oligonucleotides could be increased by introduction of modified bases that form more stable Watson-Crick base pairs with the complementary bases owing to additional hydrogen bonds. For example, diaminopurine forms three hydrogen bonds with thymine or uracil bases, whereas the natural partner adenine is able to form only two hydrogen bonds. Gryaznov and Schultz (1994) recently reported an efficient synthesis of a fully protected

phosphoramidite of 2, 6-diaminopurine nucleoside (143) and its incorporation into homo- and mixed-base oligonucleotides to investigate their binding affinities to complementary DNA and RNA. The substitution of 2, 6-diaminopurine for adenine improved the stability of hybrids formed with both DNA and RNA, especially the stability of duplexes formed with RNA. These oligomers appear to have great potential as antisense therapeutic and diagnostic agents which need binding of DNA to RNA targets where significant dA tracts might otherwise preclude strong binding (Gryaznov and Schultz, 1994).

1.4.4. As Non-isotopic Oligonucleotide Probes

Oligonucleotides are now widely used as probes for detection of specific genes and as primers for sequencing purposes. The most common method for the labeling of oligonucleotides has been the incorporation of the isotope ^{32}P. Although its sensitivity is high this isotope has a short half-life, and the technique has intrinsic hazards. Recent advances have shown that high sensitivity can be achieved by alternative markers, such as fluorophores, chromophores, and biotins. These markers can be attached to 3'- or 5'-terminus, modified phosphate groups and heterocyclic bases (Beaucage and Iyer, 1993). These non-radioactive probes have significant advantages over isotopically-labeled probes in being stable for extended periods of time and much safer to handle.

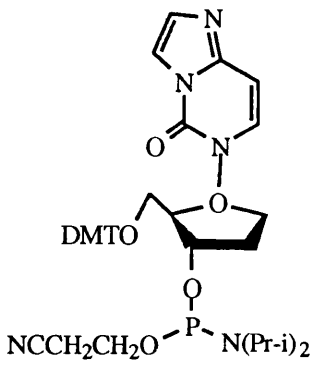
For the attachment of non-isotopic reporter groups to a base, the most widely used approach has been to incorporate a linker arm with a length of 3-12 atoms. The linker arm is attached to a DNA base on one end, and terminates in a primary reactive group such as an amino, hydroxyl, or sulphydryl function. Before oligomer synthesis, the reactive group is chemically protected to avoid side reaction with active phosphoramidite monomers. It is deprotected later to allow attachment of the desired reporter group.

The C-5 of pyrimidines is the position ideal for attachment of linker arms and reporter groups since this position is not involved in hydrogen bonding, and faces

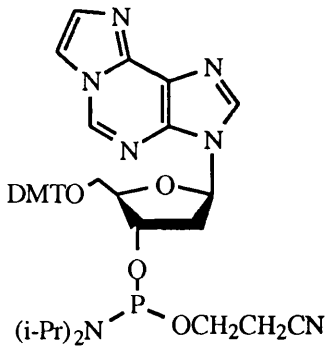
outward into the groove of a double-stranded DNA helix, and therefore reporter groups on this position have little effect on hybridization. The synthesis of this sort of oligonucleotides and their use as hybridization probes have been reviewed in a number of occasions (Ruth, 1991; Beaucage and Iyer, 1993).

For purine bases, the ideal location for attachment of reporter groups is the C-8 position of purine nucleosides since this position has steric tolerance in a double helix and is not involved in hydrogen bonding. Short fluorescent oligomers with a dansyl linked through C-8 of adenine by an aminoalkyl amino linker were synthesized using phosphotriester method (Singh *et al.*, 1990). A guanine derivative with anthracene attached through N^2[N^2-(anthracen-9-methyl)-2'-deoxyguanosine] was made and incorporated into synthetic oligomers using phosphoramidite chemistry (Casale and McLaughlin, 1990).

Some nucleoside analogues which are inherently fluorescent have also been prepared and used as DNA labeling. A recent example was the synthesis of 3, N^4-etheno cytidine and 1, N^6-etheno adenosine phosphoramidites (144, 145) and their incorporation into oligomers (Srivastava *et al.*, 1994). Their high fluorescent intensity (detectable below 1 x 10^{-9} M for adenosine site and below 1 x 10^{-7} M for cytidine site) was particularly useful. The usefulness of these modified oligomers as sequencing and amplification primers was demonstrated by PCR experiments (Srivastava *et al.*, 1994).



(144)



(145)

1.5. AIMS OF THIS PROJECT

With an efficient automated DNA synthesizer, synthesis of unmodified DNA has now become routine in many labs. However, methods for the synthesis of modified oligonucleotides are still developing and an experienced nucleic acid chemist is still required to design and synthesize modified oligonucleotides. As shown above, most methods for modified oligonucleotide synthesis follow the strategy presented in Figure 1.30a. A modified monomer is synthesized in protected form. This modified monomer, supplied to the automated synthesizer along with conventional reagents, is incorporated at a desired position during the oligonucleotide synthesis. The resin-bound, fully protected modified oligonucleotide thus obtained is then deprotected to yield the final product. The modified monomer must be stable to the chemical treatment involved in assembly, in particular, the oxidation of the phosphite ester, the deprotection of the 5'-OH, and finally the deprotection of the oligomer after synthesis. This strategy has several limitations: (a) each new modified base requires the synthesis of a modified monomer because these are normally not commercially available and even a minor change in the structure of modified base, for example from O^6-methylguanine to O^6-ethylguanine, needs multistep synthesis of a new monomer, followed by synthesis and purification of the new modified DNA; (b) a desirable modified monomer, which is made expensively and labour intensively, is often not stable under the conditions used in DNA synthesis and even a monomer which can survive the synthetic procedure may be damaged by post-synthesis procedures, such as deprotection by ammonolysis.

Because of the particular interest of our laboratory in the N-nitroso compounds, we have concentrated on methods for synthesis of DNA with modification on the 4-position of thymine and 6-position of guanine because O^4-alkylthymine and O^6-alkylguanine are the most important modified bases produced by the N-nitroso compounds. We wish to develop an alternative strategy for synthesis of oligonucleotides containing modified bases. This strategy is to incorporate a versatile monomer into DNA which combines the properties of stability to the normal procedure

of DNA synthesis with sufficient chemical reactivity to allow one to convert it into a number of desirable products after synthesis of the oligomer. It is termed the postsynthetic substitution strategy in this project and summarized in Figure 1.30b. Just as in the previous method (Figure 1.30a), this also employs a modified monomer in addition to conventional synthesis reagents. However in this case the modified monomer is a nucleoside derivative containing a leaving group, L, which is sufficiently stable to withstand the conditions of synthesis and can be displaced by nucleophiles after synthesis. The strategy has at least the following potential advantages: (a) a single synthesis of an oligomer containing the versatile base could provide a source of oligomers each containing a different modified base; (b) it offers the possibility of making DNA containing a labile or chemically reactive base; (c) special atoms, eg. NMR sensitive ^{17}O, ^{15}N,^{13}C or radioactive ^{35}S could be introduced by simple treatment with appropriate reagents at the last step. To develop this strategy for synthesis of oligonucleotides containing modified bases is the main aim of this project.

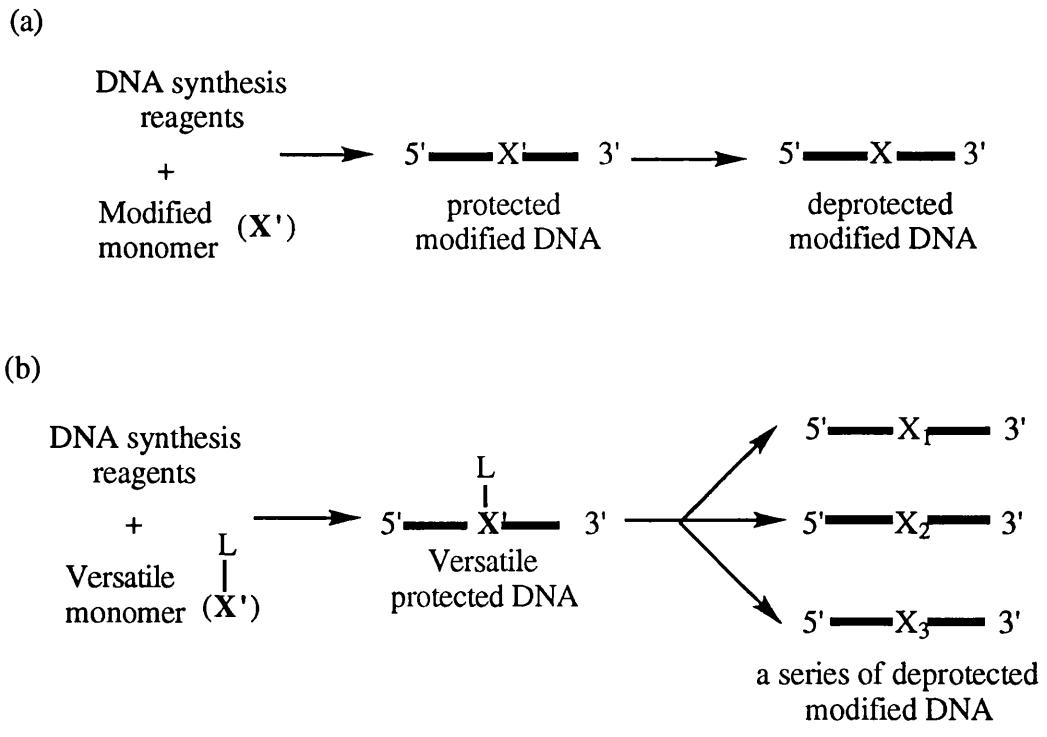


Figure 1.30. Strategies for modified DNA synthesis.

As discussed in section 1.4, one of the main applications of oligonucleotides containing modified bases is to investigate DNA-protein interactions. Because of the importance of the interaction in many biological processes, great efforts have been made to understand how these proteins interact with their DNA binding sites (for reviews see: Pabo, 1984; Harrison and Aggarwal, 1990; Aiken and Gumport, 1991; Pabo, 1992). One of the widely used approaches of studying DNA-protein interactions is the photo-cross-linking of a protein to its binding DNA, especially a DNA containing a photoactive base at a pre-determined site (see 1.4.1.ii).

λ Cro repressor is a sequence-specific DNA binding protein which plays an important role in the life style of bacteriophage λ. It competes with Cro repressor for binding to six specific operators (OR1, OR2, OR3, OL1, OL2, and OL3) by which switches λ phage growth from the lysogenic to the lytic mode (Ptashne, 1986). Although the model for its interaction with DNA was proposed as early as 1982 (Ohlendorf *et al.*, 1982) and has been a subject of intensive investigation since then (Hochschild *et al.*, 1986; Hochschild and Ptashne, 1986; Benson and Youderian, 1989; Metzler and Lu, 1989; Takeda *et al.*, 1989; Brennan *et al.*, 1990; Pendergrast *et al.*, 1992; Takeda *et al.*, 1992; Chen and Ebright, 1993) some of the proposed interactions between the amino acids of the protein and base pairs within the major groove of the DNA have been in dispute. For example, in the initial model (Ohlendorf *et al.*, 1982), thymine -2 is thought to make no contacts with Cro but Takeda et al. (1989) suggested that O-4 group of thymine -2 forms a hydrogen bond with Tyr-26. It was proposed in the initial model (Ohlendorf *et al.*, 1982), that Ser-28 contacts base pair T•A (3) and Lys-32 contacts base pairs G•C (4) and T•A (5). However, Ptashne and coworkers (Hochschild *et al.*, 1986; Hochschild and Ptashne, 1986) argued that Ser-28 contacts base pair G•C (4) and Lys-32 contacts base pair T•A (5) and G•C (6). On the other hand Takeda et al. (1989) and Benson and Youderian (1989) supported the proposed interactions in the initial model. Although the crystal structure of

Cro-DNA complex has been determined recently (Brennan *et al.*, 1990), it does not solve these arguments due to the low resolution of the structure.

During the period of this thesis, Nikiforov and Connolly (1992) demonstrated that both the Eco RV restriction endonuclease and Eco RV methylase can be cross-linked to their DNA binding site using oligonucleotides containing either 4-thiothymine or 6-thioguanine at specific positions. 4-Thiothymine and 6-thioguanine base analogues are the ideal base analogues for studying DNA-protein interactions. Structurally, the sulphur atom in the major groove of the DNA is only slightly larger than the native oxygen, but it otherwise chemically resembles oxygen. Therefore, the introduction of these thiobases into oligonucleotides should not appreciably perturb the binding interaction between the proteins and DNA. Furthermore, 4-thiothymidine and 6-thiodeoxyguanosine are photoactive at 340-350 nm wavelength, which is well away from the usual absorption maxima of protein (280 nm) and DNA (260 nm), thus cross-linking can be carried out at the wavelength which would not appreciably damage proteins and DNA. The successful development of efficient methods for synthesis of oligonucleotides containing 4-thiothymine or 6-thioguanine in the first part of this project has made it possible to study the interaction between Cro and its binding DNA using oligonucleotide containing these two photoactive bases with the photochemical cross-linking approach, and this is the second aim of this project.

CHAPTER 2

SYNTHESIS OF OLIGONUCLEOTIDES CONTAINING 4-SUBSTITUTED THYMINES BY INCORPORATION OF 4-TRIAZOLOTHYMINE INTO DNA AND POST-SYNTHETIC SUBSTITUTION OF THE TRIAZOLO GROUP

2.1. INTRODUCTION

Because of the particular interest of our lab in O^4-alkylthymine, which plays a prominent role in nitrosamine carcinogenesis (Singer, 1986), the development of a post-synthetic substitution strategy for the preparation of base-modified oligonucleotides was started with 4-substituted thymines. For this strategy to be successful the versatile nucleoside must meet at least the following two criteria: (a) it must be stable at the conditions of automated DNA synthesis; (b) it must undergo reasonably rapid and clean conversion to different modified thymine analogues upon treatment with different nucleophiles after DNA assembly.

The choice of a suitable versatile nucleoside derivative was facilitated by earlier work (Sung, 1981a; Sung, 1981b; Reese and Skone, 1984; Webb and Matteucci, 1986a; Li *et al.*, 1987; Fernandez-Forner *et al.*, 1990 ; Xu and Swann, 1990; Xu *et al.*, 1991) in which it was demonstrated that 4-triazolothymidine would undergo clean substitution upon treatment with different nucleophiles. 4-Triazolothymidine was first introduced by Sung in 1981 (Sung, 1981a) as a precursor for the synthesis of 2'-deoxy-5-methylcytidine and was soon incorporated into oligonucleotides, using phosphotriester chemistry in solution, to make oligomers containing 5-methylcytosine or thymine by post-synthetic treatment of the fully protected oligomers with aqueous ammonia or N^1, N^1, N^3, N^3-tetramethylguanidinium syn-4-nitrobenzaldoximate (TMG/Oximate) solution, respectively (Sung, 1981b). It was later used by Reese and Skone (1984) for the synthesis of O^4-substituted thymidines and in our lab for synthesis of O^4-alkylthymine phosphotriester (Li *et al.*, 1987) and phosphoramidite

(Xu and Swann, 1990) monomers to be used in oligonucleotide synthesis. More recently we have found that it can be converted easily into 4-thiothymidine at room temperature with thiolacetic acid (Xu *et al.*, 1991). Other previous papers also reported the preparation of 4-triazolothymidine phosphoramidite and its incorporation into oligonucleotides and the subsequent transformation of these to oligomers with a 4-modified base (Webb and Matteucci, 1986a; Fernandez-Forner *et al.*, 1990). Webb and Matteucci (1986a) used it to synthesize oligomers containing 5-methyl-N^4, N^4-ethanocytosine. The acid labile ethano group was successfully introduced at the oligomer level after removal of the DMT group protecting the 5'-OH, but it was unstable under the condition used for removing the conventional protecting groups from the bases. Similarly Fernandez-Forner et al. (1990) used 4-triazolothymine to make oligomers containing O^4-ethylthymine, but again had great difficulty deprotecting the final product. Although CH_3CH_2OH/DBU treatment transformed completely the 4-triazolothymine into the O^4-ethylthymine it only partially deprotected the conventional protecting groups of the bases. Conc. ammonia cannot be used for the further deprotection after CH_3CH_2OH/DBU treatment because of the sensitivity of O^4-ethylthymine to nucleophilic attack. The experience of these workers indicates that for the strategy to be successful one has to protect the other bases with groups which can be removed without destroying the desired product, and furthermore one needs a preparative technique to separate the desired oligomer from the impurities after synthesis. Fortunately, at the time when we started this work, there had been a considerable amount of work on the development of labile protecting groups for use in DNA synthesis. As a consequence, nucleoside phosphoramidites of the natural bases protected with the phenoxyacetyl group on adenine and guanine and the isobutyryl group on cytosine (Schulhof *et al.*, 1987; PAC amidites from Pharmacia) had been commercially available. These protecting groups can be completely removed from the bases by conc. aqueous ammonia for 5h at room temperature. So, in the work to be discussed in this chapter, "PAC amidites" were used, together with ion exchange high

performance chromatography at pH 12 for the separation and analysis of the desired products.

By incorporating 4-triazolothymine (T^{Tri}) into the dodecanucleotide AGCGAAXTCGCT (X standing for 4-triazolothymine) and subsequently by treating the oligomer with different nucleophiles, and using ion exchange HPLC at pH 12 for separation, we obtained, in high yield and purity, the parent oligomer containing thymine (T), and other five oligomers each containing a different modified base: O^4-methylthymine (T^{OMe}), O^4-ethylthymine (T^{OEt}), 5-methylcytosine (T^{NH2}), N^4-dimethylamino-5-methylcytosine [i.e. 4-(2, 2-dimethyl)hydrazino-5-methylpyrimid -2-one] (T^{DH}), or 4-thiothymine (T^{S}).

The work described in this chapter has been published [The Journal of Organic Chemistry, **57**, 3839-3845 (1992)].

2.2. MATERIALS AND METHODS

2.2.1. Chemicals and General Methods

The CPG-linked monomers and the chemicals used on the DNA synthesizer were obtained from Cruachem (Glasgow, Scotland) and the phosphoramidite monomers protected with phenoxyacetyl on the amino functions of adenine and guanine, and with isobutyryl on the amino function of cytosine (PAC amidites) were from Pharmacia; 5-methylcytosine cyanoethylphosphoramidite monomer protected with a benzoyl group on the 4-amino position was from Glen Research Corporation. Anhydrous methanol (MeOH, 99+%, Gold Label), 1, 1-dimethylhydrazine (98%), 1, 8-diazabicyclo[5.4.0]undec-7-ene (DBU, 96%) and thiolacetic acid (96%) were from Aldrich. Absolute ethanol (99.7+%, Analar) and conc. aqueous ammonia (d=0.88, Aristar) were from BDH. Acetonitrile (HPLC Grade, Rathburn) was dried with molecular sieve 4A at least overnight. All other chemicals were from either Aldrich or Sigma. All chemicals and solvents, unless stated otherwise, were used directly without further purification. The water content of anhydrous solvents was checked by Karl Fischer titration. Snake venom phosphodiesterase I (Crotalus durissus) was from Sigma and alkaline phosphatase from Boehringer Mannheim. T4 polynucleotide kinase was from Amersham. Gamma-[^{32}P-ATP] was from Du Pont. ^{1}H NMR spectra were obtained with a Varian VXR-400 with tetramethylsilane as internal standard and DMSO-d_6 as solvent. Reverse phase HPLC was carried out on a Gilson 320, with a 620 Datamaster for integration and Shimadzu SPD6A UV spectrophotometric detector, using a Waters 8NVC18 4μ column. Gradients were formed from 0.05 M aqueous KH_2PO_4 (pH 4.5) (buffer A) and 0.05 M aqueous KH_2PO_4 (pH 4.5) containing 33% CH_3CN (buffer B) at a flow rate of 3 ml/min. Fast protein liquid chromatography (FPLC) was carried out on a Dionex BIOLC system with a Dionex variable wavelength detector using a Pharmacia monoQ HR5-5 column. Gradients were formed from 0.4 M NaCl, 0.01 M NaOH aqueous solution (pH 12) (buffer C) and 0.8 M NaCl, 0.01 M NaOH aqueous solution (pH 12) (buffer D) or 1.2 M NaCl, 0.01 M NaOH aqueous

solution (pH 12) (buffer E) at a flow rate of 1 ml/min. All the buffers were made using water purified by a milli RO-milli Q water purification system (Millipore) and filtered using 0.4μ nylon membrane filters (Millipore). High performance thin layer chromatography (HPTLC) was carried out on Merck Kieselgel 60 F_{254} aluminium backed TLC sheets developed with 2.5% $CH_3OH/CHCl_3$.

2.2.2. Preparation of 5'-O-(4, 4'-dimethoxytriphenylmethyl)-2'-deoxy -4-triazolothymidine-3'-O-(2-cyanoethyl-N,N-di-isopropylamino)-phosphoramidite (T^{tri} Monomer)

This was prepared by Dr Xu using a modification of a previous procedure (Webb and Matteucci, 1986b) and is now commercially available from Glen Research (USA). 1, 2, 4-Triazole (1.38 g, 20 mmole) was suspended in dry CH_3CN (25 ml) at O°C (ice bath) and 0.4 ml of $POCl_3$ was slowly added with rapid stirring. Triethylamine (3 ml) was added dropwise and the suspension left stirring for 30 min. 5'-O-(4,4'-dimethoxytriphenylmethyl)thymidine-3'-O-(2-cyanoethyl-N,N-di-isopropyl-amino)phosphoramidite (**146**) (250 mg, 0.335 mmole) in dry CH_3CN (5 ml) was added over 20 min and stirring continued for an hour. The reaction was stopped with saturated aqueous $NaHCO_3$ (30 ml), then extracted with CH_2Cl_2 (50 ml). The organic layer was washed with saturated aqueous $NaHCO_3$ (30 ml), saturated aqueous NaCl (30 ml), then dried (Na_2SO_4), evaporated under reduced pressure into a small volume, co-evaporated with toluene twice, then precipitated from toluene into cold n-pentane (salt-ice bath). The resulting white precipitate, washed twice with fresh n-pentane, was dissolved in anhydrous benzene and lyophilized to give a white powder (260 mg, 97%).

2.2.3. Synthesis of Oligonucleotides

Oligonucleotides were synthesized on a Cruachem PS200 automatic DNA synthesizer (Cruachem Ltd., Glasgow, Scotland) using PAC amidites of the normal

bases without changing any part of the program set by the manufacturer while modified oligomers were synthesized with modification of the program for normal oligomer synthesis as follows.

2.2.3.i. Oligonucleotides Containing Triazolothymine Bound to CPG-Support

The portion of the oligonucleotide 3' to the 4-triazolothymine was synthesized on the machine, then the T^{Tri} monomer was added manually. In both 1.0 μmol and 0.2 μmol scale 10 mg of the monomer in a 2 ml conical glass vial sealed with septum top (Wheaton reacti-vial) was dissolved in 0.1 ml of anhydrous CH_3CN and 0.1 ml of 0.5 M tetrazole in anhydrous CH_3CN was added. The mixture was transferred into a Hamilton gas tight syringe (250 μl). The bottom end of the DNA column was disconnected from the machine and the mixture of monomer and tetrazole was injected from the gas tight syringe to the DNA column. The syringe was used to draw the solution in and out of the cartridge several times over a period of 3 min, then the cartridge was immediately reconnected to the synthesizer to complete the synthesis of the rest portion of the oligonucleotide with final DMT group left on. The yield of each coupling reaction was assessed by measuring the amount of 5'-protecting group (DMT) released by dichloroacetic acid. Dichloroacetic acid wash containing released DMT was collected and mixed with 0.1 M p-toluene sulphonic acid in acetonitrile and the absorbance measured at 495 nm.

2.2.3.ii. Oligonucleotides Containing 5-Methylcytosine (5-Me-C)

These oligomers were also synthesized from the commercial 5-methylcytosine monomer (Glen Research Corporation) using the same procedure as above and deprotected with conc aqueous ammonia at 55°C overnight and purified with Nensorb Prep cartridge (Du Pont, USA) as described in section 2.2.5.

2.2.4. Preparation of Modified Oligomers from the Oligomer Containing Triazolothymine

2.2.4.i. Oligomers Containing O^4-methylthymine (T^{OMe}) and O^4-ethylthymine (T^{OEt})

a. Optimization of substitution

This was carried out by Dr Xu. CPG-support (5 mg) bearing the 12 mer oligomer containing T^{tri}, with the 5'-DMT still on, was put into Eppendorf tubes and either MeOH/DBU (1 ml; 9:1, v/v) alone, or EtOH/DBU (1 ml; 9:1 v/v) with 10 mg of cetyltrimethylammonium bromide was added to each tube and the mixture left at 25°C for 12, 24, 48, and 68 h respectively. The solution was neutralized (90 µl 50% aqueous acetic acid, 1.5 equivalent to DBU) and immediately passed through a Dowex 50 x 8, Na^+ form, 400 mesh ion exchange column (10 ml wet volume). The column was eluted with water and collected in 1 ml fractions. The absorbance of each fraction was measured at 260 nm and the oligomers were usually found in fractions 4 to 6. The fractions containing the desired oligomer were pooled together and separated from failure sequences and the DMT group finally removed using a Nensorb Prep cartridge as described in section 2.2.5. The oligomers were hydrolysed enzymically and the base composition measured as described in section 2.2.6. The extent of substitution by MeOH or EtOH was assessed by comparison of the percentage of 4-triazolothymidine (Rt 15.9 min) with that of O^4-methylthymidine (Rt 15.2 min) or with that of O^4-ethylthymidine (Rt 19.4 min) using reverse phase HPLC.

b. Synthesis of O^4-methylthymine and O^4-ethylthymine oligomers

The T^{Tri} oligomer was treated with either methanol/DBU at 25°C overnight or ethanol/DBU for 2 days. The solutions were neutralized and immediately passed through a Dowex 50 column and purified with a Nensorb Prep Cartridge. Highly pure oligonucleotides for base composition analysis or melting curve measurement were

obtained by FPLC as described in section 2.2.5.

2.2.4.ii. Oligomers Containing 4-Thiothymine (T^{SH})

a. Optimization of thiation

To determine the optimum condition for thiation, 0.5 ml of CH_3COSH/CH_3CN (10/90, v/v) was added to each of four Eppendorf tubes containing 3 mg of the CPG-support bearing the 12 mer T^{Tri} oligomer, with the 5'-DMT removed, and left at 25°C for 4, 8, 12 and 24 h respectively. Then each sample was washed with CH_3CN (5 x 1 ml) and treated with MeOH/DBU (0.5 ml; 90/10, v/v) overnight to cleave the oligomer from the support and to remove the protecting groups. The samples were neutralized with acetic acid and passed through a Dowex 50 column as described for the T^{OMe} oligomers. The products were separated from failure sequences with a Nensorb Prep cartridge and further separated by FPLC as described in section 2.2.5. The extent of thiation was calculated by comparing the integrated absorbance of the T^{SH} oligomer with that of T^{OMe} oligomer which had been formed from substitution of any remaining T^{Tri} during the deprotection.

b. Synthesis of 4-thiothymine oligomer

The T^{Tri} oligomer from above was treated with CH_3COSH/CH_3CN (1 ml; 10%, v/v) for 24 hrs, then washed with CH_3CN (5 x 1 ml). MeOH/DBU (1 ml; 90/10, v/v) was added and left overnight. The product was neutralized with acetic acid and passed through the Dowex 50 column as above and purified with Nensorb Prep cartridge and then FPLC as described in section 2.2.5.

2.2.4.iii. Oligomers Containing Thymine (T)

The above T^{Tri} oligomer was treated with 0.5 M aqueous NaOH (24 h, 25°C). The deprotected and substituted oligomer was purified with a Nensorb Prep cartridge

2.2.4.iv. Oligomers Containing 5-Methylcytosine (5-Me-C)

The T^{Tri} oligomer was treated with conc. aqueous ammonia (d=0.880) overnight at 25°C. The resulting oligomer was purified with a Nensorb Prep cartridge.

2.2.4.v. Oligomers Containing N^4-dimethylamino-5-methylcytosine (T^{DH})

For comparison Table 2.1 also contains data on DNA containing N^4-dimethylamino-5-methylcytosine made by Dr Xu by treating the T^{Tri} oligomer with $NH_2N(Me)_2/CH_3CN$ (10/90, v/v; 3 h at 25°C). 1, 1-Dimethylhydrazine was removed by washing with CH_3CN (5 x 1 ml). The oligomer was cleaved from the support and the protecting groups removed with aqueous 0.5 M NaOH (1 ml; 24 h, 25°C). The resulting oligomer was purified with a Nensorb Prep cartridge. The oligonucleotide with high purity for base analysis and melting curve measurement was obtained by FPLC.

2.2.5. Purification of Oligonucleotides

Oligomers deprotected with conc. aqueous ammonia, or with 0.5 M NaOH, or fractions containing the desired oligomers from a Dowex 50 column, were separated from failure sequences and the DMT group finally removed using a Nensorb Prep cartridge. The cartridge connected to a peristaltic pump was first washed and activated by 10 ml of methanol and then with 5 ml of 0.1 M triethyl ammonium acetate (TEAA, pH 7). After the crude oligomer had been loaded on the cartridge, the cartridge was eluted with 10 ml of 10% (v/v) acetonitrile in 0.1 M TEAA (pH 7) to remove the failure sequences. The DMT group was removed by eluting the cartridge with 20 ml of 0.5% trifluoroacetic acid in TEAA (0.1 M) and finally the fully deprotected oligomer was eluted with 10 ml of 35% methanol in water and collected in 1 ml fractions. The fractions containing the desired oligonucleotide were located by measuring their absorbance at 260 nm. Highly pure oligonucleotides for base composition analysis or melting curve measurement were obtained by FPLC. The column was eluted with

buffer C for 2 min, then with a linear gradient from 100% buffer C to 85% buffer C and 15% buffer E over 3 min, and then to 60% buffer C and 40% buffer E over the following 20 min at a flow rate of 1 ml/min. The desired peak was collected and immediately neutralized with 1 M acetic acid and desalted with a Waters Sep-pak C_{18} cartridge. A cartridge attached to a peristaltic pump was first activated and equilibrated by passing CH_3CN (10 ml) through the cartridge followed by 10 ml of water. After application of the samples, the cartridge was washed with 15 ml of water and eluted with 5 ml of CH_3CN/H_2O (7:3, v/v) and collected in 1 ml fractions and absorbance measured at 260 nm. The majority (~90%) of the oligomer was found in the first three fractions.

2.2.6. Base Composition Analysis of Oligonucleotides

Base composition analysis of the oligonucleotides made in this chapter was carried out by Dr Xu. In general 0.5 A_{260} units of an oligomer was dissolved in 160 µl H_2O and 20 µl 600 mM Tris·HCl, 60 mM $MgCl_2$, pH 8.5. Snake venom phosphodiesterase I (10 µl, 10 µg protein) was added and the mixture incubated (37°C, 30 min), then alkaline phosphatase (10 µl, 5 µg protein) was added and incubation continued for 30 min. The nucleosides were separated by HPLC using 96.5% buffer A and 3.5% buffer B for the first 8 min, then with a linear gradient from 3.5% to 50% of buffer B over the following 15 min. The eluate was monitored at 260 nm, but for the T^{SH} oligomer the first 13 min of the run was monitored at 260 nm (for the detection of dC, dG, T and dA) and the remainder at 335 nm for the detection of T^{SH}. The relative amount of each nucleoside was measured by integration of the absorbance of each peak and comparison of the peak areas with those of a mixture of standard bases. Retention times were: dC, 1.9 min; 5-Me-dC, 3.7 min; dI (from enzymatic hydrolysis of dA), 4.8 min; dG, 5.5 min; T, 6.5 min; dA, 12 min; T^{DH}, 14.8 min; T^{OMe}, 15.2 min; T^{SH}, 15.6 min; T^{OEt}, 19.4 min. T^{DH} [1-(2'-deoxyribofuranosyl)-4-(2, 2-dimethyl)hydrazino-5-methyl-pyrimid-2-one] as a standard base for base analysis was prepared by treating

3', 5'-O-di-t-butyldimethylsilyl-4-(1, 2, 4-triazolo)-thymidine, prepared as before (Xu and Swann, 1990), in CH_3CN with 1, 1-dimethylhydrazine at room temperature overnight. The mixture was purified by silica gel column. The main compound (Rf: 0.5, $CH_3OH/CHCl_3$ 5:95) was confirmed to be 1-(2'-deoxy-3', 5'-O-di-t-butyldimethylsilyl-ribofuranosyl)-4-(2,2-dimethyl)hydrazino-5-methyl-pyrimid-2-one by NMR spectroscopy; ^1H NMR data (in DMSO-d_6): 0.07 (12H, 2 s, 3' and 5'-Si(CH_3)2-R), 0.86 (18H, 2 s, 3'- and 5'-Si(R)$_2$-(CH_3)$_3$), 2.04 (2H, m, 2'- and 2"-H), 2.10 (3H, s, 5-CH_3), 3.05 (6H, s, N-N(CH_3)$_2$), 3.73 (2H, m, 5'-H), 3.79 (1H, m, 4'-H), 4.33 (1H, m, 3'-H), 6.14 (1H, t, 1'-H) and 7.37 (1H, s, 6H). This was desilylated with tetrabutylammonium fluoride in tetrahydrofuran to give the desired modified nucleoside (T^{DH}); ^1H NMR data (in DMSO-d_6): 1.95-2.07 (2H, m, 2'- and 2"-H), 2.11 (3H, s, 5-CH_3), 3.06 (6H, s, N-N(CH_3)$_2$), 3.45-3.59 (2H, m, 5'-H), 3.71 (1H, m, 4'-H), 4.27 (1H, m, 3'-H), 6.09 (1H, t, 1'-H) and 7.75 (1H, s, 6H). The other modified nucleosides used as reference samples in the base analysis had previously been made in our lab (Xu and Swann, 1990; Xu *et al.*, 1991).

2.2.7. DNA Synthesis on a Template Containing N^4-dimethylamino-5-methylcytosine (T^{DH})

2.2.7.i. 5'-^{32}P Labeling

A mixture (20 µl) containing 10 µM primer DNA, 10 mM $MgCl_2$, 70 mM Tris-HCl (pH 7.6), 1 mM spermidine, 0.5 µM [γ-^{32}P] ATP (3000 Ci/mmol), and 4 units of T_4 polynucleotide kinase was incubated at 37°C for 30 min and then at 70°C for 5 min to inactivate the kinase.

2.2.7.ii. Duplex Preparation

A solution (100 µl) containing 1 µM 5'-end-labeled primer DNA, 9 µM cold

primer DNA, 10 μM cold template DNA, 5 mM $MgCl_2$, 50 mM Tris·HCl (pH 7.4) was hybridized by heating at 100°C for 2 min and then cooling to room temperature over 2h.

2.2.7.iii. Elongation

A reaction mixture (5 μl) containing 1 μM duplex, 150 μM dNTP (dATP or dGTP), 5 mM $MgCl_2$ and 0.2 μM of E. coli DNA polymerase I (Klenow fragment) in 50 mM Tris·HCl (pH 7.4) was incubated at room temperature for 30 sec and quenched directly with 5 μl of denaturing gel loading buffer (80% formamide and 0.1% each of bromophenol blue and xylene cyanol FF in electrophoresis buffer).

2.2.7.iv. Gel Electrophoresis

The elongated mixtures (2 μl each) obtained above were loaded on a 20% polyacrylamide gel (32 x 18 x 0.06 cm) containing 7 M urea. The buffer for electrophoresis was 45 mM Tris-borate, 10 mM EDTA at PH 8.0 (TBE buffer). Electrophoresis was carried out at 2000 V, 30 W until the two dyes were about 10 cm apart (1.5-2 h). The gel was transferred onto a piece of 3M paper, dried at 75°C under vacuum and autoradiographed. The amount of elongated oligomer was determined by cutting the gel into sections followed by scintillation counting in 3 ml of scintillation fluid.

2.2.8. Melting Curve Measurement

Non-self-complementary oligomers (Table 2.2) were annealed with an equimolar quantity of complementary strand in 0.1 M Hepes, pH 7.5, 0.2 M NaCl and 0.02 M $MgCl_2$. The temperature dependent change in absorbance at 260 nm was followed using a CARY3 spectrophotometer connected to a Cary temperature controller (Varian Techtron Pty Ltd, Australia). The temperature was increased by 1°C/min. The Tm values were determined as the maximum values of the first derivative graph of the absorbance vs temperature graph.

2.3. RESULTS AND DISCUSSION

2.3.1. Synthesis of Versatile Thymine Monomer

4-Triazolothymine phosphoramidite monomer (147) was prepared from commercially available counterpart thymine phosphoramidite (146) as described by Webb and Matteucci (1986a) (Figure 2.1). The thymidine phosphoramidite reacted with triazolating agent, prepared from phosphoryl chloride, 1, 2, 4-triazole, and triethylamine, in acetonitrile solution at room temperature {97% yield with the purity of >95% as assessed by ^{31}P NMR (149.45 and 150.13 ppm in $CDCl_3$)}. HPTLC (in 2.5% $CH_3OH/CHCl_3$) showed only two bright spots (phosphoramidite steric isomers) with Rfs (0.25 and 0.37) which differed from the starting material (usual UV absorption spots) with Rfs (0.35 and 0.42).

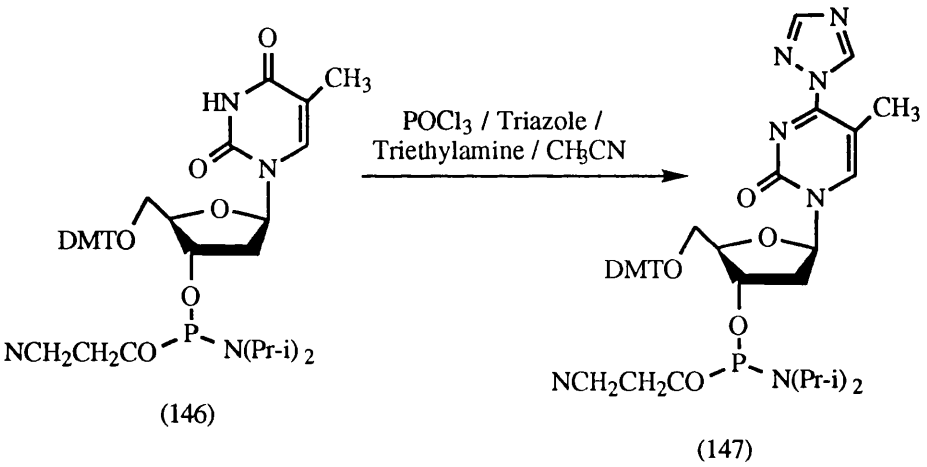


Figure 2.1. Chemical synthesis of triazolothymidine phosphoramidite from its thymidine counterpart.

2.3.2. Synthesis and Conversion of Oligonucleotides

As a model, a self-complementary dodecamer AGCGAAXTCGCT was assembled (**X** standing for 4-triazolothymine). The triazolothymidine was incorporated into the oligomer in the same way as the four natural nucleosides, except the coupling time was extended from 0.5 min (for normal nucleosides) to 3 min. The coupling

efficiency was similar to that of unmodified bases. The modified monomer was injected into the reaction column with a 250 µl Hamilton syringe, instead of being injected by DNA synthesizer, to save the valuable modified monomer. The protected oligomer (148) still attached to the CPG-support was treated at 25°C with different reagents to cleave the oligomer from the CPG support, to remove the protecting groups from the oligomer, and to displace the triazolo group in the oligomer to produce a series of modified oligomers (149-154) (Figure 2.2). Although we were not able to check the stability of triazolothymidine during oligonucleotide synthesis, comparison of the chromatographic profile of the 12 mer containing O^4-methylthymine (Figure 2.6a) or 5-methylcytosine (Figure 2.6b) with that of the oligomer containing thymine (Figure 2.6c) demonstrates that the 4-triazolothymine residue is sufficiently stable during the oligomer assembly.

The conversion of the 4-triazolothymine (T^{Tri}) oligomer into the oligomers containing modified bases involves two major reactions: substitution of the triazolo group by nucleophiles and deprotection of the oligomer. Experiments were done to follow the progress of these reactions and to discover the optimum conditions for the subsequent synthesis of oligomers.

2.3.2.i. Oligomers Containing O^4-methylthymine (T^{OMe}) or O^4-ethylthymine (T^{OEt})

The T^{Tri} oligomer was exposed to MeOH/DBU. The deprotection is complete within 5 h (Figure 2.3B) and the substitution was finished within 12 h. So overnight MeOH/DBU treatment was generally employed in the production of the methylated oligomer. This gave essentially pure T^{OMe} oligomer (Figure 2.6a). Base analysis of the main product confirmed the correct composition.

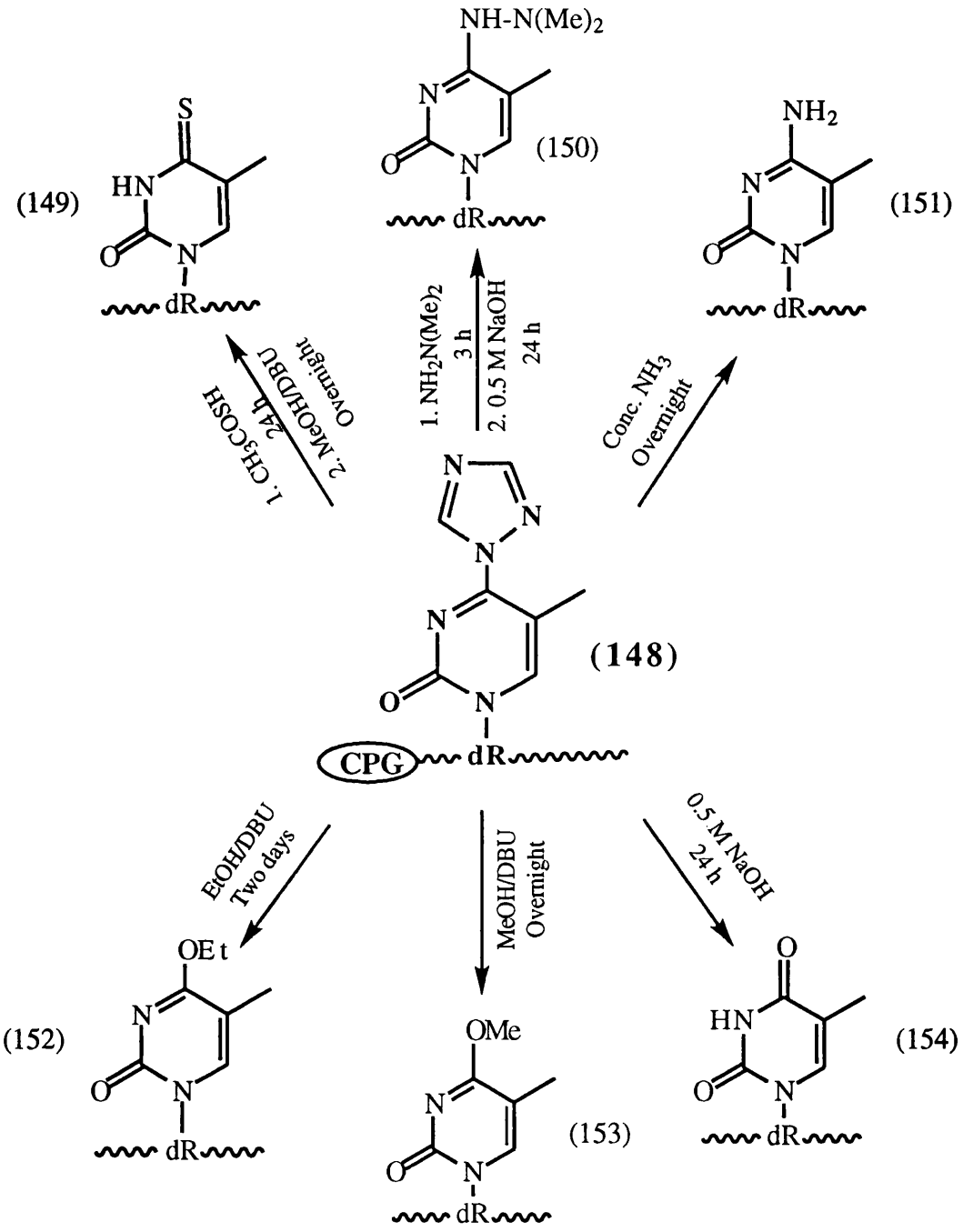


Figure 2.2. The post-synthetic conversion of the 4-triazolothymidine oligomer into a series of oligonucleotides containing modified thymines.

Because ethoxide is less nucleophilic than methoxide, a more prolonged exposure to EtOH/DBU is needed. A lipophilic counter ion, cetyltrimethylammonium, was added to the reaction mixture to keep the partially deprotected oligomer in solution (Xu and Swann, 1990). Less than 10% of the T^{Tri} residue was left unsubstituted after

12 h and the substitution was finished within 24 h. All protecting groups were cleaved within 36 h (Figure 2.3D), so two days EtOH/DBU treatment was generally employed. An essentially pure oligomer was easily obtained after Nensorb column purification and the correct composition was confirmed by base analysis. It is worth mentioning that some commercial anhydrous ethanol contains a few percent of methanol. Use of this must be avoided since as the methoxide is more nucleophilic and therefore substantial amount of the methylated oligomer will be produced even though the percentage of methanol is low.

O^4-alkylthymine is produced by the alkylation of DNA by the N-nitroso compounds and is believed to play an important role in their carcinogenic and mutagenic properties (Singer, 1986). Oligomers containing T^{OMe} or T^{OEt} have previously been synthesized by incorporation of T^{OMe} or T^{OEt} monomer either using the phosphotriester method (Li *et al.*, 1987) or the phosphoramidite approach (Fernandez-Forner *et al.*, 1990; Xu and Swann, 1990), but the present method can provide T^{OMe} and T^{OEt} oligomers in a single synthesis without the need to prepare T^{OMe} and T^{OEt} monomers. Comparison of the results here with those of Fernadez-Forner et al. (1990) shows the advantage of the labile base-protecting groups used here. Fernandez-Forner et al. showed that the conventional base protecting groups (benzoyl on adenine and cytosine, and isobutyryl on guanine) are unsuitable for the synthesis of oligomers containing O^4-ethylthymine, and although they introduced a new protecting group (p-nitrophenylethoxycarbonyl) for this purpose it also was difficult to remove.

2.3.2.ii. Oligomers Containing Thymine (T)

24 h treatment with 0.5 M aqueous NaOH was sufficient for substitution of the triazolo group and for deprotection of the other bases (Figure 2.3C). Base analysis of the main product gave only four unmodified bases (dA, dG, dC and T) and a more prolonged exposure (3 days) to 0.5 M NaOH caused no obvious destruction of the oligomer. Only deprotected thymine containing oligomer and no deprotected

4-triazolothymine containing oligomer was formed during the course of the reaction. Although oligomers containing thymine can be easily prepared from thymidine monomer, this substitution allows a reference sample containing thymine to be obtained during the same synthesis as the modified oligomer, and would also allow ^{17}O to be introduced from ^{17}O-water for NMR studies.

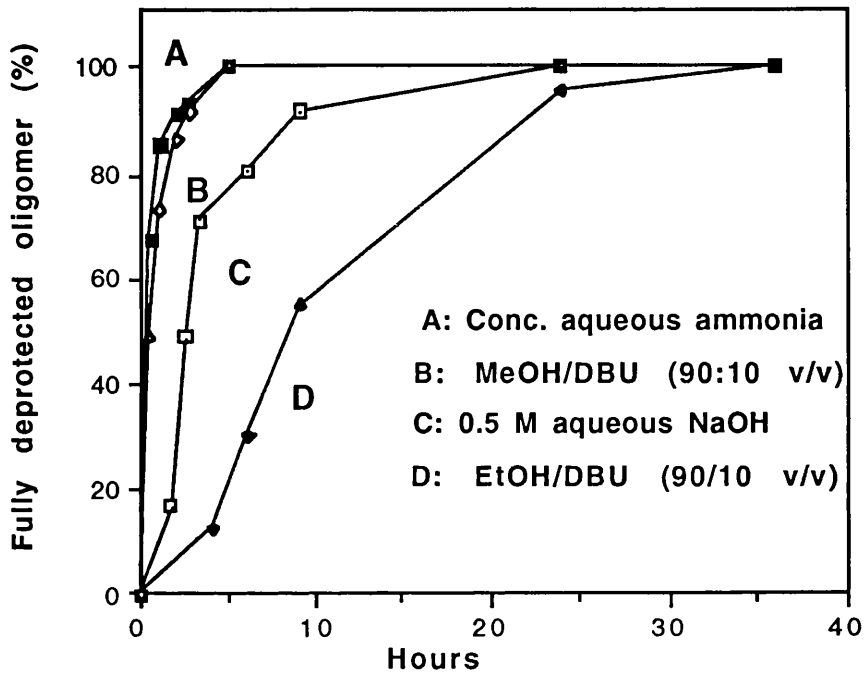


Figure 2.3. The time courses of deprotection of an oligomer AGCGAAXTCGCT (X: T^{Tri}) protected with PAC groups. The CPG-support bearing the protected T^{Tri} 12 mer, with the 5'-DMT removed, was treated at 25°C with conc. ammonia, MeOH/DBU, EtOH/DBU or 0.5 M aqueous NaOH respectively. The course of the deprotection was followed by FPLC on a mono Q column.

2.3.2.iii. Oligomers Containing 5-Methylcytosine (5-Me-C)

The T^{Tri} oligomer treated with conc. aqueous ammonia overnight gave pure 5-methylcytosine containing oligomer (Figure 2.6b). The correct composition of the oligomer was confirmed by base analysis. Deprotection was complete within 5 h (Figure 2.3A), and the longer treatment was adopted simply because it is routinely used for deprotecting unmodified oligomers. Although 5-methylcytosine containing oligomers can be obtained by incorporating the commercially available monomer, the

present procedure can provide 5-methylcytosine containing oligomer as well as the parent containing thymine in a single synthesis. More importantly, it is possible to introduce NMR sensitive ^{15}N from easily available ^{15}N-ammonia. Comparison of the oligomer made from T^{Tri} oligomer with the same sequence made from commercial 5-methylcytosine monomer shows that the purity was similar, but the yield from the T^{Tri} oligomer was better (Table 2.1).

2.3.2.iv. Oligomers Containing N^4-Dimethylamino-5-methylcytosine (T^{DH})

This was chosen as an example of the application of this strategy to the synthesis of oligomers containing labile amino-derivatives in the 4-position of pyrimidines and because of its potential use for study of base-pairing properties (see 2.3.5).

The T^{Tri} oligomer was treated with 1, 1-dimethylhydrazine/CH_3CN for 3 h to substitute the triazolo group. As the oligomer was still bound to the support the excess reagent was washed off with CH_3CN, then the oligomer was cleaved from the CPG-support and deblocked with 0.5 M aqueous NaOH for 24 h. This gave the desired oligomer (Figure 2.5a). The yield was slightly less than those of the other oligomers (Table 2.1). One possible explanation is that during the substitution step dimethylhydrazine cut the oligomer from CPG-support and this oligomer was washed off before the deprotection step. Aqueous NaOH was employed rather than conc. ammonia for deblocking because NaOH converts any remaining 4-triazolothymine into thymine, and thymine containing oligomer can be separated easily from the desired modified oligomer by FPLC (Figure 2.4a). Base analysis of the main product confirmed the presence of the modified base (Figure 2.4b). As well as the four peaks corresponding to the four unmodified nucleosides (dA, dG, dC and T) there was one peak which co-eluted with authentic 1-(2'-deoxyribofuranosyl)-4-(2, 2-dimethyl)hydrazino-5-methylpyrimid-2-one.

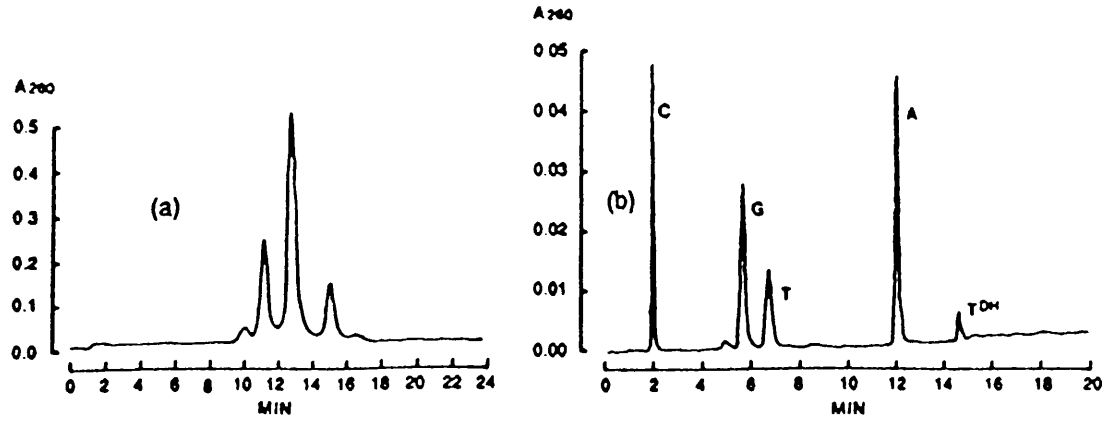


Figure 2.4. FPLC profiles of crude AGCGAAT^{DH}TCGCT (a) and HPLC base analysis of the purified oligomer (b). FPLC was performed on a Mono Q column using 0.4 M NaCl, 0.01 M NaOH for 2 min, then 0.8 M NaCl, 0.01 M NaOH increasing to 30% over 3 min, then to 70% over the following 20 min at a flow rate of 1 ml/min. HPLC was performed on an 8NVC18 4μ reverse phase column using 0.05 M KH_2PO_4 (pH 4.5) and 0.05 M KH_2PO_4 (pH 4.5) containing 33% CH_3CN at a flow rate of 3 ml/min using the gradient described in section 2.3.4.

2.3.2.v. Oligomers Containing 4-Thiothymine (T^{SH})

The T^{Tri} oligomer was treated with CH_3COSH/CH_3CN, which replaced the triazolo group with an SH group. The possible mechanism of this substitution is that the triazolo group was first protonated then attacked at the C-4 position by thio-acetate ion followed by loss of acetyl group to form 4-thiothymine. The acetyl group is lost possibly because the thio-keto form is more stable than the 4-S-acetyl form (Xu *et al.*, 1991). Interestingly thiolacetic acid does not remove the 5'-DMT group. Thiation was finished within 12 h, but overnight treatment with CH_3COSH/CH_3CN was generally used. As the oligomer was still attached to the CPG-support, the excess CH_3COSH was washed off with CH_3CN. For deprotection, either MeOH/DBU or conc. aqueous ammonia can be used. In agreement with a previous report (Connolly and Newman, 1989) it was observed that the 4-thiothymine in the dodecamer was converted to 5-methylcytosine by prolonged exposure to conc. aqueous ammonia. Therefore

MeOH/DBU (16 h, 25°C) is to be preferred (Figure 2.6d). The present method offers an easy way to produce T^{SH} oligomers with a good yield and without the need to prepare the 4-thiothymidine monomer.

UV spectrum of the T^{SH} oligomer (Figure 2.5), together with that of the control 12 mer and the free 4-thiothymidine, shows an absorption at 335 nm due to the presence of the modified base. This clearly confirmed its incorporation into the DNA. This feature is particular useful for a quick check of the presence of the modified base by UV spectrophotometry and for detecting the oligomer in chromatography at 330-340 nm, at which normal oligomers are transparent.

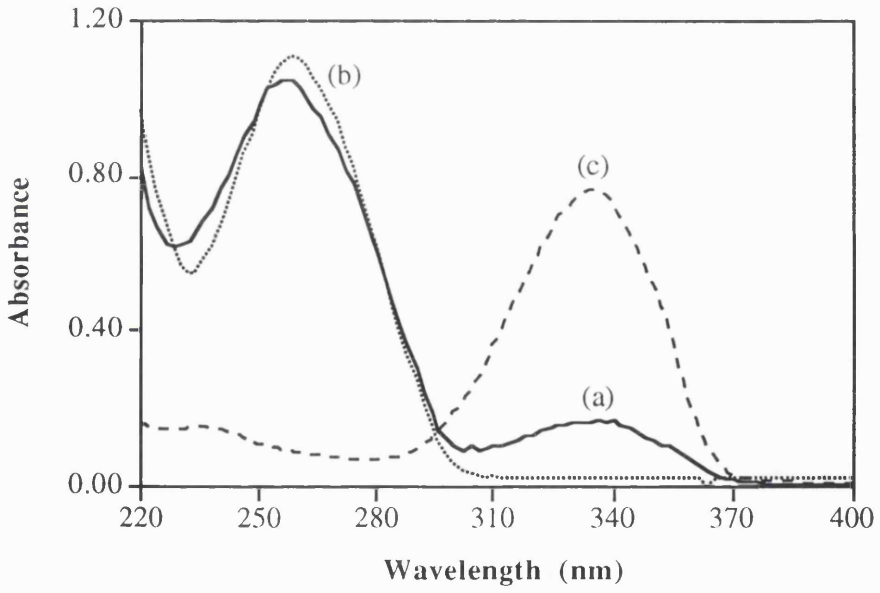


Figure 2.5. UV spectra of (AGCGAAT^{SH}TCGCT) (a), (AGCGAATTCGCT) (b), and 4-thiothymidine (c).

Comparison of these results with the previous reports (Connolly and Newman, 1989; Nikiforov and Connolly, 1991; 1992b) shows the advantages of the post-synthetic substitute strategy. Originally T^{SH} oligomers were prepared by incorporation of a 4-thiothymine phosphoramidite monomer in which the sulphur atom was protected with $-SCH_3$ group (Connolly and Newman, 1989). Unfortunately, this

monomer was not very stable towards some of the reagents used in oligonucleotide synthesis and there was only a moderate yield (10-15 %) of T^{SH} oligomer. For this reason the same authors later used the p-nitrophenyl group (Nikiforov and Connolly, 1991) or 2-cyanoethyl group (Nikiforov and Connolly, 1992b) to protect the sulphur .

Tale 2.1. The yield and purity of synthetic oligonucleotides*.

Oligomer	Scale	Yield (OD)	Purity (%)
AGCGAATTCGCT	1/5 x 1μmole	12.1	> 90
AGCGAATTCGCT	0.2 μmole	14.0	> 90
----AT^{OMe}----	1/5 x 1μmole	12.4	> 90
----AT^{OEt}----	1/5 x 1μmole	10.6	> 90
----AT^{NH2}----	1/5 x 1μmole	12.4	> 90
----AT^{NH2}----	0.2μmole	7.0	> 90
----AT^{DH}----	1/5 x 1μmole	8.0	60
----AT^{SH}----	1/5 x 1μmole	9.3	70

*Yield refers to the amount of oligomer (A_{260} units) recovered from the Nensorb cartridge. Purity is the purity of oligomers after Nensorb cartridge purification. The 5/1 x 1 mmol scale syntheses were carried out by substitution of 1/5th of contents of a 1μmol cartridge containing a synthetic 4-triazolothymine oligomer. The 0.2μmol scale syntheses were done for comparison, and in these only commercially available monomers were used.

2.3.3. Purification of Oligonucleotides

For purification, T^{NH2}, T^{DH} or T oligomers were passed through a Nensorb cartridge to remove the failure sequences since these sequences lack the 5'-DMT group and are eluted from the cartridge with a low percent (10%) of aqueous acetonitrile solution while the correct sequence still binds to the cartridge since it has the DMT group and binds relatively stronger than the failure sequences. T^{OR} or T^{SH} oligomers were first passed through a Dowex ion exchange column to remove DBU, followed by

a Nensorb cartridge. These steps resulted in reasonably pure products (Figure 2.4a and Figure 2.6) with high yields (Table 2.1). More highly purified oligomers could be obtained by FPLC. Under basic conditions, the oligomers containing O^4-alkylthymine (Figure 2.6a), or 5-methylcytosine (Figure 2.6b) were eluted earlier than their parent oligomer (Figure 2.6c) possibly because they lose one negative charge due to lack of the imino proton at the 3-position of pyrimidines. Because 4-thiothymine has a lower pKa than thymine (the pKa of uridine is 9.3 and that of 4-thiouridine is 8.2) (Saenger, 1984), the T^{SH} oligomer (Figure 2.7d) was eluted later and easily separated from the parent containing thymine and 4-substituted thymine containing oligomers. The nucleoside O^4-methylthymidine is very slowly converted to thymidine in alkaline solution (Xu and Swann, 1990), but the O^4-alkylthymine residue in the oligomer is more resistant to destruction by alkaline condition, perhaps because of steric hindrance by the neighbouring nucleotides. This allows one to use FPLC under basic conditions for the purification. After purification followed by immediate neutralization, the purified oligomer was checked again by FPLC and still gave a single peak.

2.3.4. Base Composition Analysis

Although sequences of normal DNA synthesized by a highly reliable automated DNA synthesizer are rarely checked, it is important to determine the base composition of DNA sequences containing modified bases since some base analogues could be potentially damaged during DNA assembly or in the process of post-synthetic treatments.

Base composition analysis was carried out by totally digesting the oligonucleotides with snake venom phosphodiesterase, which cut the oligonucleotides into individual nucleotides, and then with alkaline phosphatase to remove the phosphate groups from the nucleotides to form the nucleosides. All of the modified nucleosides were detected at 260 nm except T^{SH} being detected at 330 nm since its λmax is different from that of other nucleosides. However, separation of this modified nucleoside from

the four normal nucleosides was good enough to enable the UV detector to be switched from 260 nm to 335 nm to detect its elution without the need to perform a separate HPLC run.

In summary, by combination of the direct base-substitution at the oligomer level and separation by FPLC under the basic condition, this method provided a general and easy method for synthesis and purification of oligonucleotides containing different modified thymines avoiding the tedious steps of preparation of modified monomers. The method also provides potential for preparation of oligomers containing modified bases which are not stable during the assembly steps and for easily introducing atoms such as radioactive ^{35}S from ^{35}S thiolacetic acid or NMR sensitive atoms such as ^{13}C from ^{13}C-methanol, ^{15}N from ^{15}N-ammonia, and ^{17}O from ^{17}O-water.

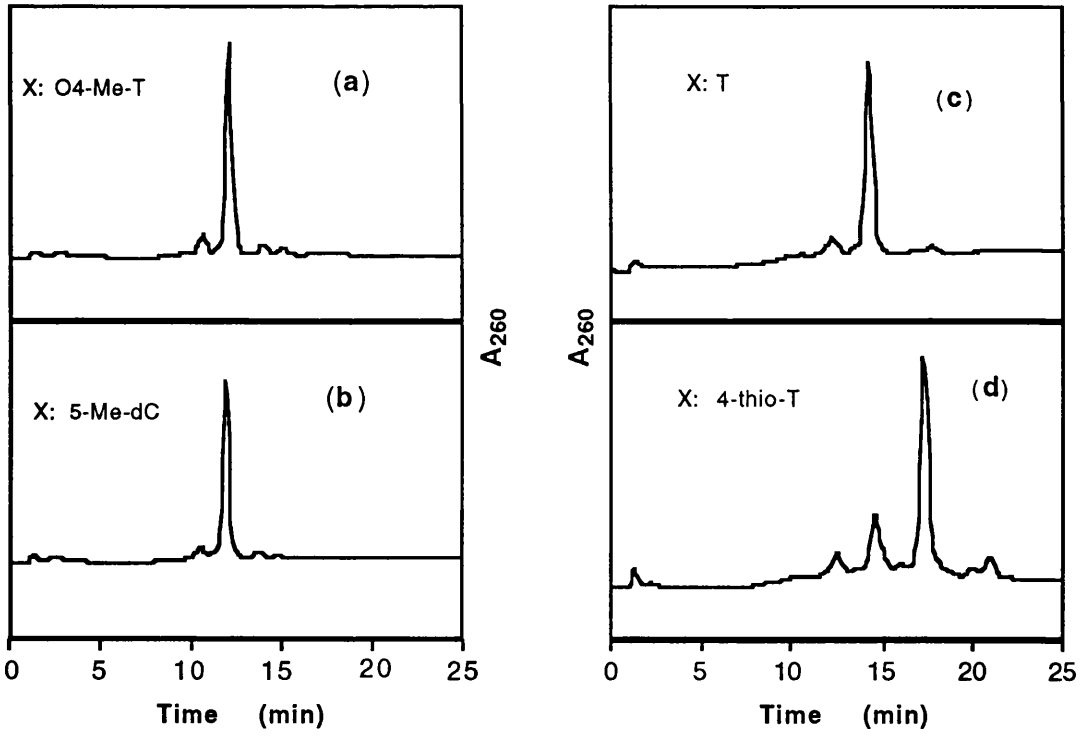


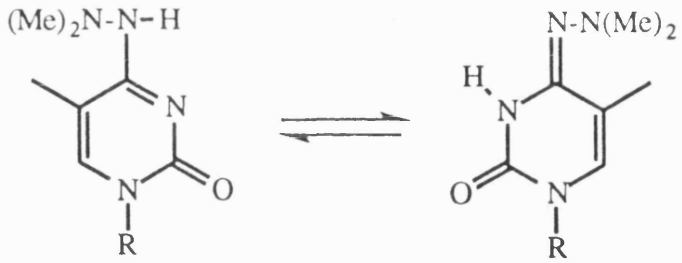
Figure 2.6. FPLC profiles of crude modified oligomers AGCGAAXTCGCT, (a) X = T^{OMe}, (b) X = 5-Me-dC, (c) X = T, and (d) X = T^{SH}. The oligomers were chromatographed on a Mono Q column using 0.4 M NaCl, 0.01 M NaOH for 2 min, then 1.2 M NaCl, 0.01 M NaOH increasing to 15 % over 3 min, then to 40% over the following 20 min at a flow rate of 1 ml/min.

2.3.5. Base-pairing Properties of Oligomers Containing N^4-dimethylamino-5-methylcytosine (T^{DH})

T^{DH} was made because it was expected to exist in about equimolar proportions of the amino and imino form (Figure 2.7) and thus might base-pair equally with adenine or guanine. This expectation was based on the observation that in the series cytosine, N^4-aminocytosine and N^4-hydroxy(or N^4-methoxy)cytosine, the ratio of amino to imino form is 10^4, 10-0.1 (depending on solvent) and 0.1-0.03 respectively, presumably reflecting increasing electronegativity in the series H, NH_2, and OH (or OCH_3) (Brown *et al.*, 1968; Anand *et al.*, 1987; Kong Thoo Lin and Brown, 1989). Little is known in this respect about DNA containing N^4-aminocytosine or N^4-alkylaminocytosine because the necessary oligomers have not been available. However melting temperature studies on a self-complementary 12-mer AGCGAAT^{DH}TCGCT in 0.1 M Hepes pH 7.5, 0.2 M NaCl and 0.02 M $MgCl_2$ solution did not show a distinct transition temperature suggesting that the self-complementary oligomer cannot form double stranded DNA, or even a hairpin structure, under these experimental conditions. So non-self-complementary sequences containing N^4-dimethylamino-5-methylcytosine were prepared and Tm values of their duplexes measured. The non-self-complementary duplexes containing T^{DH}:G or T^{DH}:A base pairs had sharp transition temperatures like the controls containing C:G or T:A pairs (Figure 2.9), but in both cases the Tm was 17-19°C lower than the control (Table 2.2). The depression of Tm was nearly equal to the effect of mismatch base-pairing (Table 2.2). This indicates that T^{DH} is not a good substitute for either thymine or cytosine. Possibly this is a consequence of steric hindrance by the dimethylamino group.

In order to further evaluate the base-pairing properties of T^{DH} synthetic 20 mers (CGCTCTTACAXGTATCGGAT, where X: T or T^{DH}) were synthesized and used as

templates for DNA synthesis by the Klenow fragment of E. coli DNA polymerase I. The ^{32}P labeled complementary strand ATCCGATAC, as primer, was annealed to the template, and the Klenow fragment with dATP or dGTP added. The reaction mixtures were analyzed by gel electrophoresis. It was found that T^{DH} did not block DNA synthesis but the elongation of the primer with both G and A appears to proceed at a rate similar to that for the mismatch incorporation of G opposite T in the template strand (Figure 2. 9) These results again show that the base-pairing property of T^{DH} oligomer is similar to the mismatch one and T^{DH} has no obvious preferable base-pairing with adenine or guanine. These results are consistent with the Tm measurement and confirm that T^{DH} does not replace either C or T.



(155) (156)

Figure 2.7. Tautomeric isomers of N^4-dimethylamino-5-methylcytosine (T^{DH})

Table 2.3. Tm values of DNA duplexes
5' CAG GAA TXC GC 3'
3' GTC GTT AYG CG 5'

X	Y	Tm (°C)	X	Y	Tm (°C)
T	A	58.7	C	G	59.2
TDH	A	42	C	A	42
TDH	G	40.6			

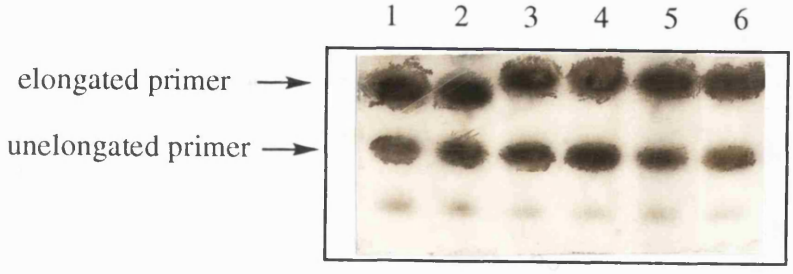


Figure 2.9. Autoradiograph showing incorporation of dATP or dGTP opposite T^{DH} in the template strands. Two bands are shown in each lane, representing the unelongated primer (low band) and the elongated primer as indicated by the labels. Lane 1-2: incorporation of dATP opposite T^{DH} in the template strand, lane 3-4: incorporation of dGTP opposite T^{DH} in the template strand, and lane 5-6: incorporation of dGTP opposite T in the template strand.

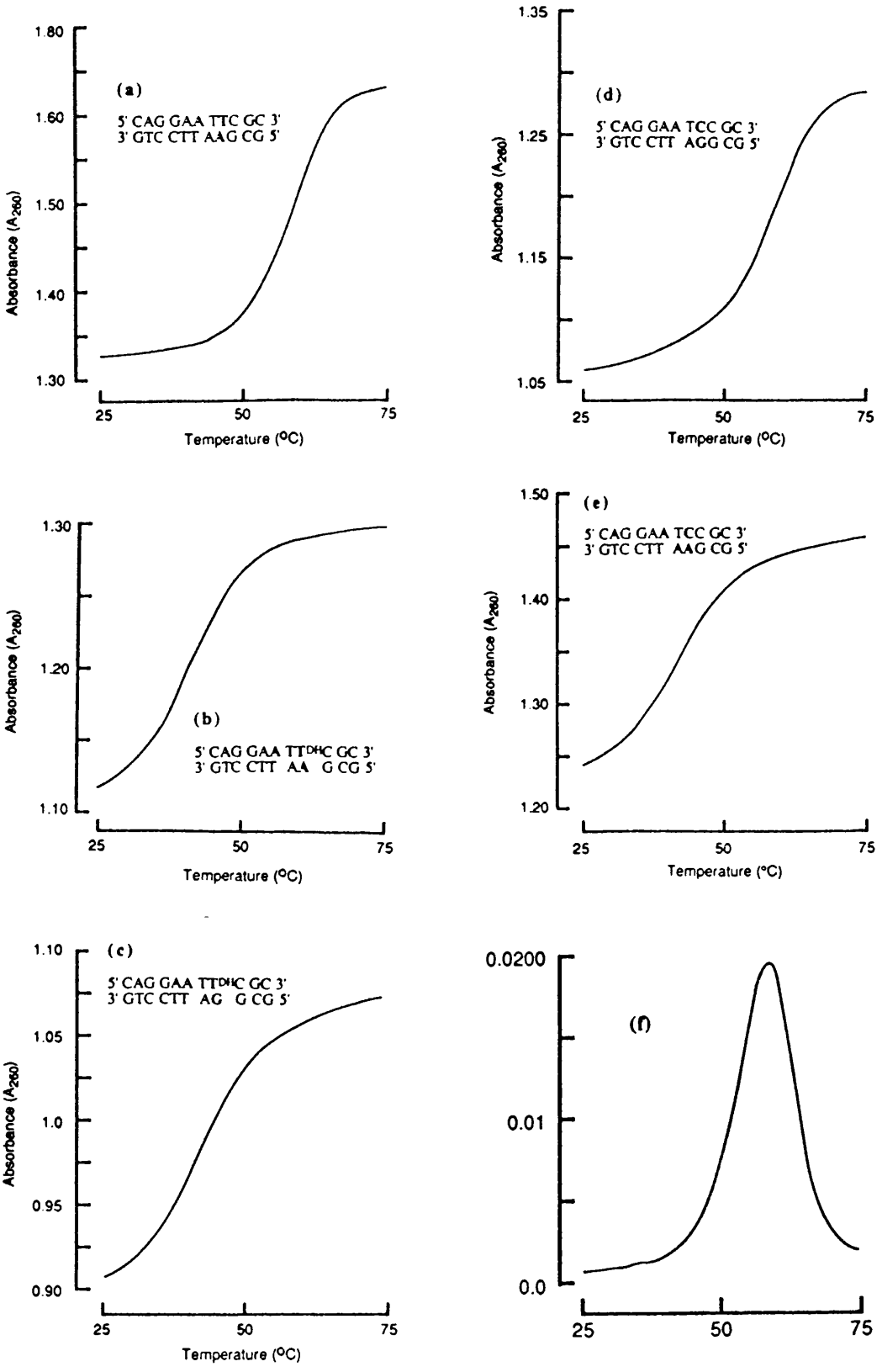


Figure 2.8. Tm curves of DNA duplexes (a-e), and the first derivative of the Tm curve (a) [(f)].

CHAPTER 3

SYNTHESIS OF OLIGONUCLEOTIDES CONTAINING LABILE 4-SUBSTITUTED THYMINES

3.1. INTRODUCTION

Although the post-synthetic substitution approach discussed in Chapter 2 and by others (Webb and Matteucci, 1986a; Ferentz and Verdine, 1991; Gao *et al.*, 1992; Ferentz *et al.*, 1993) offers the possibility of making DNA containing a labile base which would be unstable to the conditions of DNA assembly it is still difficult to apply this strategy to the preparation of oligomers containing a modified base which is unstable to deprotection procedures such as treatment by conc aqueous ammonia. One example, which has been mentioned in chapter 2, is the synthesis of oligonucleotides containing 5-methyl-N, N-ethanocytidine (Webb and Metteucci, 1986a). The acid labile ethyleneimino group, which could be destroyed by acid treatment for deblocking the 5'-DMT group in DNA synthesis, was successfully introduced into oligonucleotides by incorporation of 4-triazolothymidine into the oligomers and subsequent treatment of the oligomers with aziridine. However, there was then great difficulty in removing the protecting groups from the bases since 5-methyl-N^4, N^4-ethanocytosine is not stable under the normal conditions used for DNA deprotection. Although the 9-fluorenylmethoxycarbonyl (fmoc) group was successfully used for protecting the amino functions of adenine and cytosine and was removed from adenine and cytosine by a mild treatment with 1 M DBU in CH_3CN (10 min, room temperature), the fmoc group was unsuitable for protection of guanine and it was impossible to make oligomers containing both guanine and the ethanocytosine. We therefore turned our attention to further developing the post-synthetic strategy by preparing a versatile oligonucleotide which is fully cleaved from the support, deprotected and purified but with the leaving group, L, intact. We expected that this sort of versatile DNA could be used for making oligonucleotides containing a very labile base (such as N^4, N^4-ethanocytosine) since no any further post-synthetic treatments to the oligomer would be

needed after the substitution of "L" by a nucleophile. In this chapter the preparation of such a versatile DNA and its use for making oligomers containing labile 4-modified thymines is discussed. An attempt (MacMillan and Verdine, 1991) to make a similar versatile oligonucleotide and to use it for the synthesis of DNA containing modifications on the 4-position of a cytosine was not entirely successful because the leaving group 2, 4, 6-trimethylphenoxide was too stable and harsh condition (65°C, 14 h) required for its substitution was not ideal for the preparation of oligomers containing labile bases.

3.2. MATERIALS AND METHODS

3.2.1. Chemicals and Enzymes

The chemicals and enzymes used in this study were obtained as described in section 2.2.1 except the CPG-linked monomers and phosphoramidite monomers, protected with t-butylphenoxyacetyl on the amino functions of adenine, guanine and cytosine (Expedite amidites), and capping reagent t-butylphenoxyacetic anhydride were from Millipore; and 4-triazolothymidine phosphoramidite was from Glen Research.

3.2.2. Synthesis, Conversion, and Purification of Oligonucleotides

3.2.2.i. Oligomers Containing 4-Triazolothymine (T^{tri}) Bound to CPG-support

Synthesis of T^{tri} oligomers was carried out on an ABI 391 DNA synthesizer by incorporating 4-triazolothymidine monomer as described in section 2.2.3, except using Expendite amidites of normal bases and t-butylphenoxyacetic anhydride as the capping reagent.

3.2.2.ii. Oligomers Containing 4-Phenylthiothymine (T^{SPh})

The oligomer containing 4-triazolothymine prepared as above was treated with 0.1 M benzenethiol in the presence of diisopropylethylamine in acetonitrile overnight

while still attached to the CPG. The excess reagent was washed off with acetonitrile (5 x 1 ml), and protecting groups then removed with conc. aqueous ammonia at room temperature for 2 hours. The deprotected oligomers were purified with Nensorb cartridge as described in section 2.2.5 and were further purified by reverse phase HPLC using a Water 8NVC18 4μ column. The column was eluted with a linear gradient from 100% of 0.05 M aqueous KH_2PO_4 (pH 6.3) (buffer M) to 50% of buffer M and 50% of 0.05 M aqueous KH_2PO_4 (pH 6.3) in 33% CH_3CN (buffer N) at a flow rate of 1.5 ml/min over 30 minutes. The desired peak was collected and desalted with NAP-10 column.

3.2.2.iii. Oligomers Containing 5-Methyl-N^4, N^4-ethanocytosine (T^e) or 4-Azidothymine (T^{N_3})

The 5-methyl-N^4, N^4-ethanocytosine containing oligomers were prepared by treatment of the oligomer containing 4-phenylthiothymine with 20% aqueous aziridine solution for 20 min at room temperature and purified by HPLC as above. The oligomer containing 4-azidothymine were made by reaction of the T^{SPh} containing oligomer with NaN_3 at 35°C overnight in DMF/H_2O (1:1). The resulting oligomers were desalted, then purified by HPLC as above.

3.2.3. Base Composition Analysis of Oligonucleotides

Base composition analysis of the oligonucleotides was carried out essentially as described in section 2.2.6. but the digestion of the T^e containing oligomers was carried out at room temperature and the time for each digestion was extended to 2 hours. The nucleosides were separated by HPLC as described in 4.2.2.ii. and the column was eluted with 93% of buffer M and 7% of buffer N for the first 8 min, then with a linear gradient from 7% of buffer N to 35% of buffer N over the following 17 min and then to 100% of buffer N over the next 5 min and remaining at this concentration for further 10 min. The eluate was monitored at 260 nm, but for the oligomers containing

4-phenylthiothymine or 5-methyl-N^4, N^4-ethanocytosine it was first monitored at 260 nm to detect the four natural nucleosides and then at 315 nm for detection of 4-phenylthiothymidine or at 295 nm for detection of the ethanocytidine. The retention times were dC, 3.3 min; dG, 7.48 min; T, 10.37 min; dA, 14.90 min; T^{N3}, 16.80 min; T^{e}, 18.17 min; T^{SPh}, 34.13 min.

3.2.4. 4-Triazolothymidine

This was prepared based on the method of Webb and Matteucci (1986a) and used for the preparation of the modified nucleosides as the reference samples in the base composition analysis.

To a suspension of 1, 2, 4-triazole (13.8 g, 0.2 mol) in dry CH_3CN (250 ml) cooled in an ice bath was slowly added of $POCl_3$ (4 ml) with stirring. Triethylamine (30 ml) was added to the solution and the mixture was stirred for 30 min and then left in an ice bath.

Thymidine (2.43 g, 10 mmol) suspended in dry CH_3CN (100 ml) was treated with N, N-dimethylaminotrimethylsilane (10 ml, 62.4 mmol) with good stirring. After 45 min at room temperature the solution was concentrated under reduced pressure and redissolved in CH_3CN (10 ml). This solution was added to the phosphoryl tri-triazolide reagent prepared above. After being stirred for 2 h at room temperature the mixture was poured into 5% $NaHCO_3$ (500 ml) and extracted with CH_2Cl_2 (2 x 300 ml). The organic washes were concentrated and redissolved in 100 ml EtOH. This solution was treated with 80% CH_3COOH/H_2O (100 ml) for 2 h at room temperature with stirring to remove the trimethylsilyl group. The solution was reduced to a small volume and redissolved in C_2H_5OH (100 ml), concentrated to a small volume, and left in a fridge overnight for crystallization. The crystalline product was isolated by filtration and dried under vacuum. Yield 2.5 g (49.2%). ^1H NMR data (in DMSO-d_6): 2.11 (1H, m, 2'-H), 2.30 (3H, s, 5-CH_3), 2.34 (1H, m, 2''-H), 3.62-3.68 (2H, m, 5'-H), 3.89 (1H, m, 4'-H), 4.24 (1H, m, 3'-H), 5.22 (1H, t, 5'-OH), 5.31 (1H, d, 3'-OH), 6.10

(1H, t, 1'-H), 8.35 (1H, s, triazole H), 8.60 (1H, s, 6-H), 9.28 (1H, s, triazole H). UV: λ_{max} 245 nm and 324 nm.

3.2.5. 4-Phenylthiothymidine, 5-Methyl-N^4, N^4-ethanocytidine, and 4-Azidothymidine.

These modified nucleosides, used as the reference in the base analysis, were prepared by treating 4-triazolothymidine, prepared above, at room temperature with 0.1 M C_6H_5SH in the presence of diisopropylethylamine (2h), 20% aqueous aziridine (5 min) or NaN_3 in DMF/H_2O (1:1) (overnight) respectively. 5-Methyl-N^4, N^4-ethanocytidine and 4-phenylthiothymidine were purified by column chromatography and 4-azidothymidine by HPLC. They were characterized by ^{1}H NMR. ^{1}H NMR data of 4-azidothymidine (in DMSO-d_6): 2.17 (1H, m, 2'-H), 2.44 (1H, m, 2''-H), 2.70 (3H, s, 5-CH_3), 3.81-3.92 (2H, m, 5'-H), 4.06 (1H, m, 4'-H), 4.51 (1H, m, 3'-H), 6.54 (1H, t, 1'-H), 8.21 (1H, s, 6-H). UV: λ_{max} = 253 nm, 275 nm (shoulder). ^{1}H NMR data of 4-(phenyl)thiothymidine (in DMSO-d_6): 1.50 (3H, s, 5-CH_3), 2.30 (1H, m, 2'-H), 2.65 (1H, m, 2''-H), 3.18-3.41 (2H, m, 5'-H), 3.98 (1H, m, 4'-H), 4.31 (1H, m, 3'-H), 6.13 (1H, t, 1'-H), 7.25-7.38 (5H, m, aromatic H), 8.31 (1H, s, 6-H). UV: λ_{max} = 315 nm. ^{1}H NMR data of 5-methyl-N^4, N^4-ethanocytidine: 1.36 (3H, s, 5-CH_3), 1.48 (4H, s, N^4, N^4: CH_2-CH_2), 1.91 (1H, m, 2'-H), 2.32 (1H, m, 2''-H), 3.55-3.63 (2H, m, 5'-H), 3.75 (1H, m, 4'-H), 4.13 (1H, m, 3'-H), 6.39 (1H, t, 1'-H), 8.21 (1H, s, 6-H). UV: λ_{max} = 297 nm.

3.2.6. Ethylenimine (Aziridine)

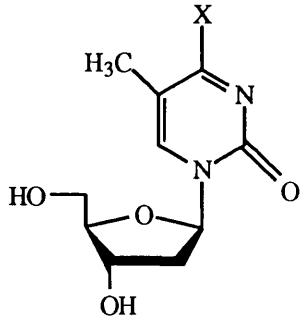
This was prepared using the method of Allen et al. (Allen *et al.*, 1963). β-aminoethyl sulphuric acid (141 g, 1 mol) was mixed with 40% NaOH solution (308

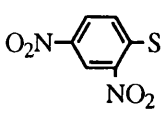
ml) in a 1 litre three-neck flask. The mixture was heated with a free flame until it just began to boil. The distillate was collected in an ice-cooled flask through a water-cooled condenser connected to the reaction flask. The heating was resumed when the initial reaction had subsided and about 130 ml of distillate was collected. Potassium hydroxide (120 g) was added gradually to the chilled distillate, whereup the imine separated as an upper layer. The upper layer was recovered and kept overnight in a fridge over 25 g of KOH (pellets). The crude imine was decanted from the hydroxide and distilled from a fresh 15 g of KOH. The fraction boiling at 50-100°C was collected and dried over 6 g of KOH. The crude ethylenimine was separated and dried over 6 g of fresh KOH. The crude compound was then decanted from the KOH and redistilled from 5 g of KOH, yield 11.1 g (25.6%) (b.p. 55-57°C).

3.3. RESULTS AND DISCUSSIONS

For the strategy to be successful, the thymine analogue, which is to be converted to a 4-modified thymine at the oligomer level at the final stage, must contain at the 4-position a group which is stable during DNA synthesis and to the procedures used to remove protecting groups after synthesis but is still easily displaceable upon treatment with a number of nucleophiles under mild conditions. Leaving groups previously used in the post-synthetic substitution approach for the synthesis of oligomers carrying pyrimidines were either too reactive to survive during the deprotection procedures (Webb and Matteucci, 1986a; Xu *et al.*, 1992b) or too stable to be substituted by nucleophiles under mild conditions (MacMillan and Verdine, 1991). Since previously we successfully used 2, 4-dinitrophenylthio group at the 6-position of purine in a post-substitution strategy for making oligomers carrying modified purines (Xu *et al.*, 1992a; Xu *et al.*, 1992b), initially an attempt was made to use 4-(2, 4-dinitrophenyl)thiothymidine (157) as the versatile nucleoside. However, preliminary experiment at the nucleoside level showed that the dinitrophenylthio group at the 4-position of the thymine was too reactive to withstand even a brief treatment (30 min) of conc. aqueous ammonia at room temperature. This is understandable since

leaving groups are normally more active at the 4-position of thymine than at the 6-position of purine counterparts. We therefore decided to use phenylthio as a leaving group. Compared with 2, 4-dinitrophenylthio this group is less reactive since it does not contain two electro-withdrawing nitro groups. Preliminary results in the nucleoside level demonstrated that 4-phenylthiothymidine (158) was relatively stable toward the ammonia treatment at room temperature. No obvious substitution was observed by TLC after 2 h, though, as observed in an attempt of making oligomers containing 4-thiothymine from (158) (Nikiforov *et al.*, 1991), long exposure will cause the displacement of phenylthio group by ammonia. More importantly, it underwent clean and quick substitution upon treatment with a number of nucleophiles under mild conditions. Since it has been established in the previous work (see Chapter 2) that 4-triazolothymine can be incorporated into oligonucleotides and quickly substituted by a range of nucleophiles at the oligomer level, and it was also shown in our experiment that 4-triazolo group could be quickly displaced by the phenylthio group, we built the versatile oligonucleotide by using 4-triazolothymidine monomer (147), which can be prepared in one step from thymidine phosphoramidite and is now commercially available (Glen Research, USA), to avoid multiple chemical synthetic steps for preparing 4-phenylthiothymidine monomer. In order to reduce the possibility of ammonolysis of 4-phenylthiothymine, the monomers protected with the labile t-butylphenoxyacetyl group on the amino functions of adenine, guanine and cytosine (Expedite monomer, Millipore) were used. These monomers can be deprotected completely with conc. aqueous ammonia at room temperature within two hours (Sinha *et al.*, 1993).



	X		X
(157)		(162)	
(158)		(163)	N_3

4-Triazolothymidine phosphoramidite (147) was incorporated into a dodecamer GCTACT^{Tri}GACTGC (159) as described previously (see section 2.2.3). T-butylphenoxyacetic anhydride, instead of acetic anhydride, was used as the capping reagent to prevent the transamidation of N-t-butylphenoxyacetyl protected nucleosides during the capping reaction with acetyl group, as this transamidation has been observed in the study of the stability of N-t-butylphenoxyacetyl protected nucleosides towards chemical reagents used in oligonucleotide synthesis (Sinha *et al.*, 1993). The protected oligomer (159) was then treated with 0.1 M benzenthiol in the presence of diisopropylethylamine in CH_3CN overnight at room temperature to substitute the triazolo group. As the oligomer (160) was still attached to controlled-pore-glass (CPG) support, the excess reagents were washed off with CH_3CN, and then treated with conc. aqueous ammonia to fully cleave and deprotect the oligomer (Figure 3.1). The crude oligomer was purified and DMT group removed by Nensorb cartridge. Highly pure oligomer (161) was obtained by reverse-phase HPLC purification because the desired oligomer was well separated from impurities (mainly the oligomer containing 5-methylcytosine) due to the strong hydrophobic effect of the phenyl group (Figure 3.2a). The oligomer thus obtained was fully deprotected, highly purified and ready for use to make modified oligomers. To confirm the presence of the versatile nucleoside (158), base analysis was carried out by totally digesting the purified oligomer to nucleosides. The HPLC profile of the digest showed one extra peak (Figure 3.2b), which was confirmed to be the 4-phenylthiothymidine (158) by coinjection of the digested sample with the authentic modified nucleoside. For the detection of the 4-phenylthiothymidine the wavelength of the detector was changed, after the elution of the 4 natural nucleosides, from 260 nm to 315 nm which is the λ_{max} of the modified nucleoside.

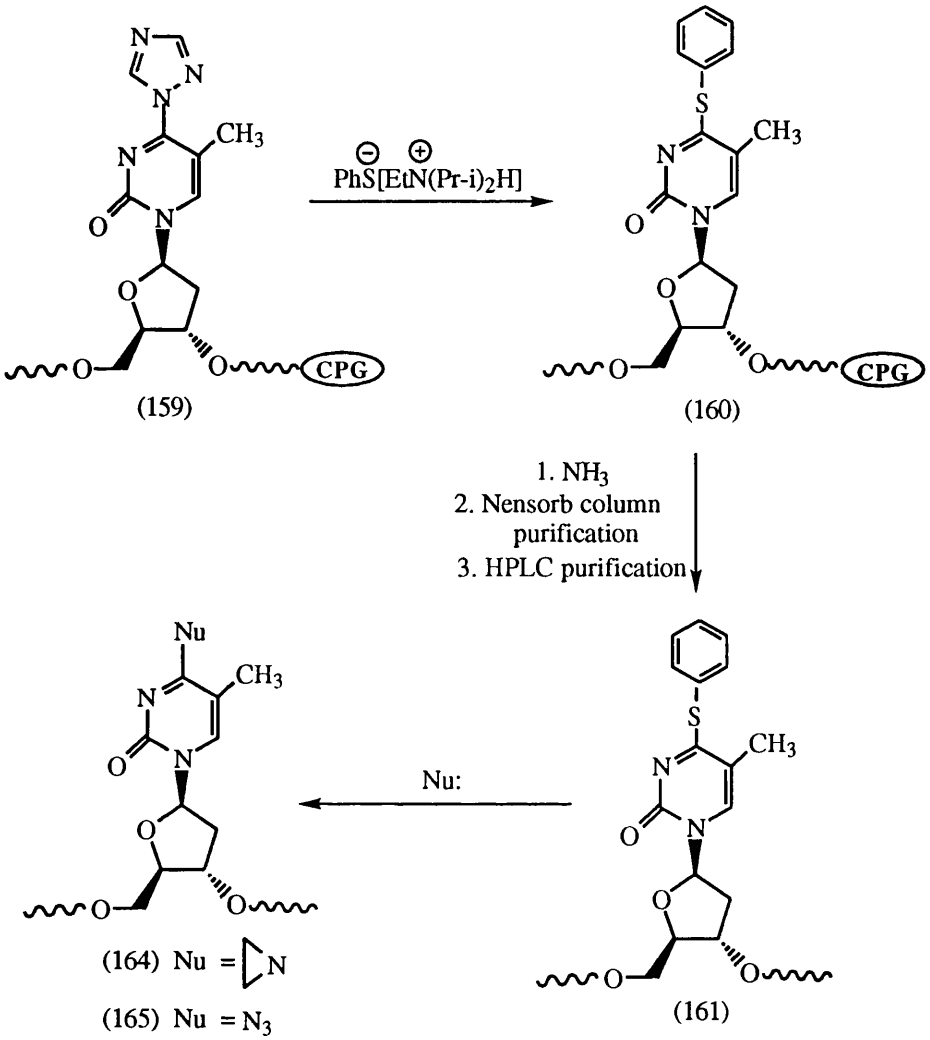


Figure 3.1. The post-synthetic conversion of the versatile oligomer (159) into different oligonucleotides containing modified thymines.

To demonstrate the advantages of the versatile oligonucleotide (161), we chose two labile nucleoside analogues 5-methyl-N^4, N^4-ethanocytidine (T^e) (162) and 4-azido-5-methyl-2-pyrimidinone-1-β-(2'-deoxyriboside) (T^{N_3}) (163) as the target molecules.

For the preparation of the oligomer containing (162), the versatile oligonucleotide (161) was treated with aqueous aziridine solution at room temperature for 20 min. HPLC showed the substitution finished completely within that period and produced the essentially pure modified oligomer (164) (Figure 3.2c). The oligomer was confirmed by enzymatic hydrolysis followed by HPLC chromatography of the digested

solution (Figure 3.2d).

The oligomers containing 5-methyl-N^4, N^4-ethanocytidine have been synthesized by Webb and Matteucci (1986a) by incorporation of 4-triazolothymidine into the oligomer and subsequently treating the oligomer with aziridine. To avoid the damage of the modified base by the ammonia deprotection step, dA and dC phosphoramidites protected on the exocyclic amino functions with 9-fluorenylmethoxycarbonyl (fmoc) group, which can be removed rapidly with 1 M DBU in acetonitrile, were prepared and used for the oligomer synthesis. Since the protection of exocyclic amino group of deoxyguanosine with fmoc group did not give the desired product, their method did not allowed the preparation of oligomer containing the modified base with the four natural bases of DNA (Webb and Metteucci, 1986a).

The T^e containing oligonucleotides can form site-specific cross-linked DNA duplexes upon hybridization with the complementary sequences (Webb and Matteucci, 1986a). Covalently linked DNA duplexes are a valuable tool for the study of protein-DNA interactions since many DNA-binding proteins require strand dissociation or local helical distortion to perform their functions. For example, cross-linked duplexes formed by T^e oligomers with their complementary sequences have been successfully used to determine the structural requirements for the exonuclease and polymerase activities of prokaryotic and phage DNA polymerases (Cowart *et al.*, 1989). Another potential application of the T^e oligomers is for antisense oligonucleotides since the ability to form a covalent bond with their complementary targets could make them more efficient as antisense oligonucleotides. Aziridine derivatives of guanine (N^6, N^6-ethano-2, 6-diaminopurine) and adenine (N^6, N^6-ethanoadenine) have also been incorporated into oligonucleotides because of their potential biological applications (Matteucci and Webb, 1987; Cowart and Benkovic, 1991).

The oligomer containing (163) was prepared simply by treating (161) with

NaN_3 in 50% DMF/H_2O at 35°C overnight. Essentially pure oligomer was obtained after removing the excess NaN_3 by NAP-10 column (Figure 3.2e). The presence of the modified nucleoside was confirmed by base analysis (Figure 3.2f). The desired oligomer was not obtained by reacting the oligomer (160) or (161) with NaN_3 followed by ammonia deprotection, possibly due to the instability of the modified base to the ammonia treatment.

Oligonucleotides containing azidonucleosides have been widely used as probes for investigating the interaction of nucleic acids with proteins (Sylvers and Wower, 1993). However, so far, they have been prepared exclusively by the enzymatic incorporation and there is no report of their chemical synthesis in the literature (Sylvers and Wower, 1993). Phosphoramidite monomer, the most widely used chemistry for DNA synthesis, is incompatible with azides due to reactivity with the trivalent phosphorus (Sylvers and Wower, 1993).

The above results clearly demonstrated the advantages of using the versatile oligomer (161) for making oligonucleotides containing labile modified bases. Although our experiments were carried out only with the thymidine derivative, it is reasonable to assume that its uridine analogue would behave similarly, and hence it is possible to synthesize the oligomers containing labile deoxyuridine or deoxycytidine analogues by a similar approach.

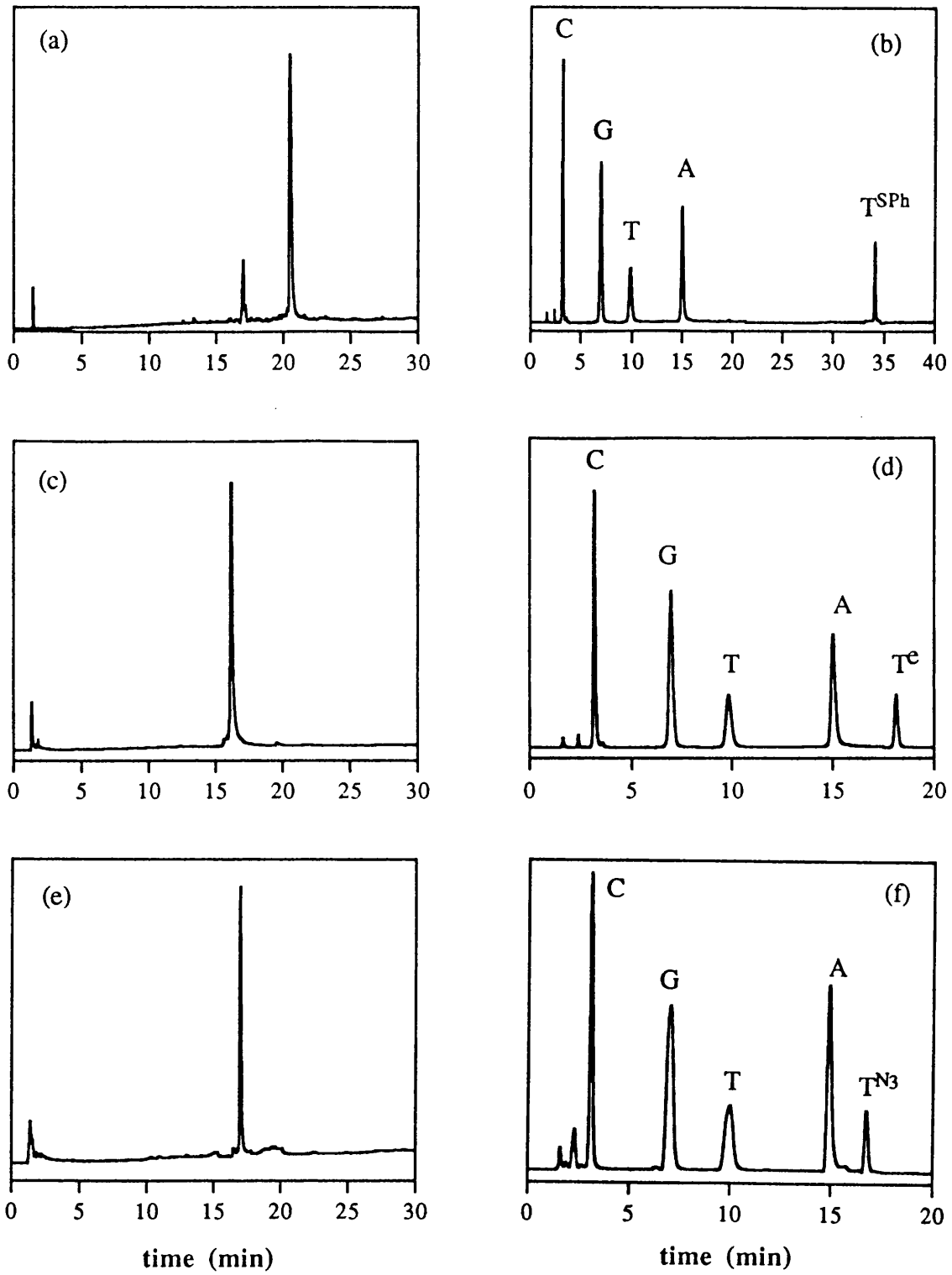


Figure 3.2. HPLC traces of crude oligonucleotides containing T^{SPh} (a), T^{e} (c), or T^{N_3} (e); and of base composition analysis of oligomers containing T^{SPh} (b), Te (d), or T^{N_3} (f).

CHAPTER 4

SYNTHESIS OF OLIGONUCLEOTIDES CONTAINING 6-SUBSTITUTED GUANINES

4.1. INTRODUCTION

Although the methods for the synthesis of oligomers containing modified pyrimidines by post-synthetic substitution strategy have been developed in this project and by others (Webb and Matteucci, 1986; Fernandez-Forner *et al.*, 1990; MacMillan and Verdine, 1991; Chen *et al.*, 1991), little had been done on modified purines by the time we started this work. Because of our particular interest in DNA containing O^6-alkylguanine which, like O^4-alkylthymine, plays an important role in nitrosamine carcinogenesis, and the importance of modified purines for investigation of DNA-protein interaction and for other applications (see Chapter 1), we turned our attention towards developing a similar strategy for modified purines. Although, at the time when this project was started, synthesis of oligomers containing some modified purines, such as 6-thioguanine (Rappaport, 1988), 2, 6-diaminopurine (Chollet *et al.*, 1986), O^6-alkylguanine (Li and Swann, 1989; Smith *et al.*, 1990) had been reported, some of the procedures were very difficult and tedious to perform. For example, the method reported by Rappaport (1988) for the synthesis of oligonucleotides containing 6-thioguanine used the phosphotriester procedure with the conventional acyl groups protected on the bases. It took 5 days to cleave the oligomer from its support and remove the protecting groups on the phosphate with tetramethylguanidine and E-2-nitrobenzaldoxime in pyridine and another 7 days to remove protecting groups from bases with NH_3 in anhydrous methanol. All these operations were carried out under nitrogen in anhydrous conditions. Most other oligomers containing modified guanines were made by a route in which a modified nucleoside was synthesized, converted to the corresponding phosphoramidite or phosphotriester monomer, and then incorporated into the oligomer. It is obvious that such a route is limited to the incorporation of relatively stable monomers which can withstand the conditions of

DNA synthesis.

It is more difficult to develop a post-synthetic substitution strategy for purines than for pyrimidines. From a chemical point of view, guanine and adenine nucleosides are more difficult to work with due to their poor solubility in most of organic solvents, their sensitivity to acid, and the need for protecting groups on the exocyclic amino functions. This need causes extra difficulties for developing post-substitution strategy to modified purines since some modified purines are not stable under the harsh conditions needed for deprotection, and therefore protecting groups must be used which match the nature of the modified bases to be introduced. Furthermore, one particular problem with 6-O-substituted guanine is that the substitution makes it much more difficult to remove the protecting group from the 2-amino function (Gaffney *et al.*, 1984b) and therefore long and drastic conditions are needed to remove it which increases the probability of substitution or removal of the 6-O-modifications. However, despite these difficulties, this chapter discusses the successful development of a method for oligomers containing guanine modified at the 6-position. The versatile guanine monomer N^2-phenylacetyl-6-(2, 4-dinitrophenyl)thio-2'-deoxyguanosine-3'-phosphor-amidite (185) was prepared (Figure 4.3) and incorporated into oligomers. The oligomer containing this versatile guanine ($G^{S\emptyset}$) was converted into oligomers containing 6-thioguanine (G^{S}), 2, 6-diaminopurine (G^{NH2}), 2-amino-6-methyl-aminopurine (G^{NMe}), O^6-methylguanine (G^{OMe}) or guanine (G) by treatment with appropriate reagents after synthesis (Figure 4.6). It is worth noting that as the amino, alkylamino, alkoxy and oxy functional groups at the 6-position are introduced after synthesis the method could be specially useful for the introduction of ^{15}N, ^{13}C, ^{17}O from [^{15}N]-NH_3, [^{13}C]-MeOH and [^{17}O]-H_2O for NMR. Furthermore the method can be used for construction of oligomers containing other chemically reactive guanine derivatives which could be destroyed during the DNA assembly.

The work described in this chapter has been published [Tetrahedron, **48**, 1729-1740 (1992)].

4.2. MATERIALS AND METHODS

4.2.1. Chemicals and Enzymes

Chemicals and enzymes were obtained and treated, as necessary, as described in section 2.2.1.

4.2.2. Chromatography

Reverse phase HPLC and fast protein liquid chromatography (FPLC) were carried out as described in section 2.2.1. High performance thin layer chromatography (HPTLC) and thin layer chromatography (TLC) were carried out on Merck Kieselgel 60 F_{254} aluminum backed TLC sheets developed with 2.5% $CH_3OH/CHCl_3$ (solvent A),or 5% $CH_3OH/CHCl_3$ (solvent B) or CH_2Cl_2 /ethylacetate/diisopropylamine (75:25:2, v/v/v, solvent C).

4.2.3. Synthesis of 5'-O-(4, 4'-dimethoxytriphenylmethyl)-N^2-isobutyryl-2'-deoxy-6-(2, 4-dinitrophenyl)thioguanosine-3'-O-(2-cyanoethyl-N, N-diisopropylamino)phosphoramidite (177)

4.2.3.i. 5'-O-dimethoxytriphenylmethyl-N^2-isobutyryl-2'-deoxy-6-thioguanosine (173)

5'-Dimethoxytriphenylmethyl-N^2-isobutyryl-2'-deoxyguanosine (172) (1.28 g, 2 mmol) was dissolved in anhydrous CH_2Cl_2 (25 ml) and triethylamine (1.2 ml), 2-mesitylenesulphonyl chloride (0.88 g, 4 mmol), and 4-(dimethylamino)pyridine (10 mg) added sequentially. After 1 h stirring at room temperature, the solution was cooled with an ice-bath and N-methylpyrrolidine (1 ml) slowly added. The solution was left stirring for 20 min in the ice-bath and then for another 30 min at room temperature. Thiolacetic acid (1 ml) in CH_2Cl_2 (5 ml) was added dropwise and stirred for another 1 h. The reaction mixture was washed with saturated aqueous $NaHCO_3$ (2 x 25 ml) and

then with saturated aqueous NaCl (2 x 25 ml). The organic layer was dried over anhydrous Na_2SO_4 and evaporated to solid, and purified by column chromatography on Kieselgel 60 H using CH_2Cl_2 followed by a gradient of 0.5% to 1.5% $CH_3OH/CHCl_3$ (v/v) with a few drops of pyridine added. The fractions containing the desired product were pooled and evaporated under reduced pressure to give a slightly yellow solid (0.53 g, 40.0%).^{1}H NMR data (in DMSO-d_6): 1.13 (6H, d, -C(CH_3)$_2$- of isobutyl), 2.37-2.78 (3H, m, 2'-H, 2"-H and -CH- of isobutyl), 3.09-3.19 (2H, m, 5'-H), 3.71 (6H,s, OCH_3 of DMT), 3.95 (1H, m, 4'-H), 4.40 (1H, m, 3'-H), 5.35 (1H, d, 3'-OH, ex), 6.25 (1H, t, 1'-H), 6.78-7.40 (13H, m, aromatic-H), 8.30 (1H, s, 8-H), 11.93 (1H, s, 2-NH, ex) and 13.41 (1H, s, 6-SH). UV: λmax=332.0 nm (in MeOH)

4.2.3.ii. 5'-O-dimethoxytriphenylmethyl-N2-isobutyryl-2'-deoxy-6-(2, 4-dinitrophenyl) thioguanosine (176)

To a solution of compound (**173**) (0.33 g, 0.5 mmol) in anhydrous CH_3CN (25 ml) was added triethylamine (0.5 ml) and 2, 4-dinitrofluorobenzene (0.11 g, 0.6 mmol). After 40 min, TLC showed that the starting material was entirely converted to a new compound with higher Rf (solvent A), which was yellow under visible light. The reaction mixture was concentrated under reduced pressure into a small volume and diluted with $CHCl_3$ (25 ml). The solution was washed with saturated aqueous $NaHCO_3$ (25 ml), then with saturated aqueous NaCl (25 ml). The organic layer was dried over anhydrous Na_2SO_4, evaporated to give a yellow solid, and purified by silica gel column chromatography using CH_2Cl_2 followed by 1% CH_3OH/CH_2Cl_2 (v/v) with a few drops of pyridine added. The fractions containing the desired product were combined and evaporated to a yellow solid (0.38 g, 92%). ^{1}H NMR data (in DMSO-d_6): 1.03 (6H, d, -C(CH_3)$_2$- of isobutyl), 2.37-2.90 (3H, m, 2'-H, 2"-H and -CH- of isobutyl), 3.12 (2H, m, 5'-H), 3.70 (6H, s, OCH_3 of DMT), 3.98 (1H, m,

4'-H), 4.55 (1H, m, 3'-H), 5.31 (1H, d, 3'-OH, ex), 6.40 (1H, t, 1'-H), 6.73-7.36 (13H, m, aromatic-H), 8.09 (1H, d, 6-H of 2,4-dinitrophenyl), 8.29 (1H, m, 5-H of 2,4-dinitrophenyl), 8.54 (1H, s, 8-H), 8.90 (1H, d, 3-H of 2,4-dinitrophenyl) and 10.52 (1H, s, 2-NH, ex) UV: λmax = 348.0 nm (in MeOH)

4.2.3.iii. 5'-O-dimethoxytriphenylmethyl-N^2-isobutyryl-2'-deoxy-6-(2, 4-dinitrophenyl)thioguanosine-3'-O-(2-cyanoethyl-N, N-diisopropylamino) phosphoramidite (177)

Compound (176) (250 mg, 0.3 mmol) was dissolved in 2 ml dry THF and 0.25 ml of N, N-diisopropylethylamine added. The solution was cooled in an ice-bath and 150 µl of 2-cyanoethyl-N, N-diisopropylchlorophosphoramidite in 1 ml of dry THF was added dropwise under stirring. After 10 min, the ice-bath was removed and the reaction left stirring at room temperature until TLC showed the complete conversion of starting material into two compounds with higher Rfs in solvent B. The mixture was diluted with ethyl acetate (25 ml) and washed with saturated aqueous NaCl (2 x 25 ml). The organic layer was dried over Na_2SO_4 and evaporated to give a yellow solid, which was purified by column chromatography eluted with CH_2Cl_2/ethyl acetate/ diisopropylamine (85:15:1, v/v/v). The fractions containing the desired product were combined and evaporated to a yellow solid, redissolved in benzene and lyophilized to give a yellow powder (190 mg, 62.1%). ^{31}P NMR data (in $CDCl_3$): 149.41 and 149.16.

4.2.4. Synthesis of 5'-O-(4, 4'-dimethoxytriphenylmethyl)-N^2-phenylacetyl-2'-deoxy-6-(2, 4-dinitrophenyl)thioguanosine-3'-O-(2-cyanoethyl-N, N-diisopropylamino)phosphoramidite (185)

4.2.4.i. 3', 5'-Diacetyl-N^2-phenylacetyl-2'-deoxyguanosine (180)

The synthesis of 3', 5'-diacetyl-N^2-phenylacetyl-2'-deoxyguanosine was

carried out by Dr Xu using the method of Li and Swann (Li and Swann, 1989). Its structure was confirmed by NMR spectroscopy. ^{1}H NMR data (in DMSO-d_6): 2.02-2.08 (6H, 2 s, 3' and 5' -$COCH_3$), 2.54-2.97 (2H, m, 2'-H and 2"-H), 3.80 (2H, s, -CH_2- of phenylacetyl),4.21-4.23(3H, m, 4'-H and 5'-H), 5.30 (1H, m, 3'-H), 6.23 (1H, t, 1'-H), 7.26-7.34 (5H, m, aromatic -H of phenylacetyl),8.25 (1H, s, 8-H), 11.95 (2H, s, 2-NH, ex, and 1-NH).

4.2.4.ii. 3', 5'-Diacetyl-N2-phenylacetyl-2'-deoxy-6-thioguanosine (181)

This experiment was done by Dr Xu. 3', 5'-Diacetyl-N2-phenylacetyl-2'-deoxyguanosine (180) (2.8 g, 6 mmol) was dissolved in anhydrous CH_2Cl_2 (50 ml), and triethylamine (3 ml), 2-mesitylenesulphonyl chloride (2.0 g, 9 mmol) and 4-(dimethylamino)pyridine (20 mg) sequentially added. After 1 h stirring at room temperature, the solution was cooled in an ice-bath and N-methylpyrrolidine (5 ml) in CH_2CH_2 (5 ml) was added slowly. The solution was left stirring for 30 min in the ice-bath and then for another 30 min at room temperature. Thiolacetic acid (5 ml) in CH_2Cl_2 (10 ml) was added dropwise and stirred for another 1 h. The reaction mixture was washed with saturated aqueous $NaHCO_3$ (2 x 75 ml) and then with saturated aqueous NaCl (2 x 75 ml). The organic layer was dried (Na_2SO_4) and evaporated to solid, and purified by column chromatography using CH_2Cl_2 followed by a gradient from 0.5% to 2% $CH_3OH/CHCl_3$ (v/v). The fractions containing the desired product were pooled and evaporated under reduced pressure to give a slightly yellow solid, yield 2.3 g (80%).^{1}H NMR data (in DMSO-d_6): 2.02-2.08 (6H, 2 s, 3' and 5' -$COCH_3$), 2.56-2.98 (2H, m, 2'-H and 2"-H), 3.84 (2H, s, -CH_2- of phenylacetyl),4.19-4.25(3H, m, 4'-H and 5'-H), 5.32 (1H, m, 3'-H), 6.23 (1H, t, 1'-H), 7.27-7.34 (5H, m, aromatic -H of phenylacetyl),8.43 (1H, s, 8-H), 12.23 (1H, s, 2-NH, ex) and 13.27 (1H, s, 6-SH). UV: λmax = 332 nm (in MeOH)

3.2.4.iii. N^2-phenylacetyl-2'-deoxy-6-thioguanosine (182)

This experiment was carried out by Dr Xu. Pyridine (15 ml) and NaOH solution (2.5 M, 9 ml) were added to 3', 5'-diacetyl-N^2-phenylacetyl-2'-deoxy-6-thioguanosine (181) (1.9 g, 4 mmol), and the solution was vigorously stirred at room temperature for 6 min. Dowex 50-X8 in pyridinium form (20 ml) was then added to neutralize the solution. The Dowex was filtered and washed with pyridine. The filtrate was evaporated under vacuum to a foam and then was purified by column chromatograph on Kieselgel 60 H using a gradient of increasing ethanol in chloroform. The fractions containing desired product were pooled and evaporated to give a slightly yellow solid, yield 1.25 g (80%). Its structure was confirmed by NMR spectroscopy; ^{1}H NMR data (in DMSO-d_6): 2.57-2.99 (2H, m, 2'-H and 2"-H), 3.83 (2H, s, -CH_2- of phenylacetyl), 3.54 (2H, m, 5'-H), 3.83 (1H, m, 4'-H), 4.37 (1H, m, 3'-H), 4.98 (1H, t. 5'-OH, ex), 5.33 (1H, d, 3'-OH, ex), 6.20 (1H, t, 1'-H), 7.27-7.35 (5H, m, aromatic -H of phenylacetyl),8.41 (1H, s, 8-H), 12.23 (1H, s, 2-NH, ex) and 13.26 (1H, s, 6-SH); UV: λmax=331.0 nm (in MeOH).

4.2.4.iv. N^2-phenylacetyl-2'-deoxy-6-(2, 4-dinitrophenyl)thioguanosine(183)

The purified N^2-phenylacetyl-2'-deoxy-6-thioguanosine (**182**) was converted to compound (**183**) using a similar procedure as for compound (180). The identity was confirmed by NMR spectroscopy; ^{1}H NMR data (in DMSO-d_6): 2.30-2.72 (2H, m, 2'-H and 2"-H), 3.55 (2H, m, 5'-H), 3.71 (2H, s, -CH_2- of phenylacetyl), 3.85 (1H, m, 4'-H), 4.41 (1H, m, 3'-H), 4.90 (1H, t. 5'-OH, ex), 5.34 (1H, d, 3'-OH, ex), 6.34 (1H, t, 1'-H), 7.27 (5H, m, aromatic -H of phenylacetyl), 8.12 (1H, d, 6-H of 2,4-dinitrophenyl), 8.36 (1H, m, 5-H of 2,4-dinitrophenyl), 8.68 (1H, s, 8-H), 8.88 (1H, d, 3-H of 2,4-dinitrophenyl) and 10.88 (1H, s, 2-NH, ex).

4.2.4.v. 5'-O-(4, 4'-dimethoxytriphenylmethyl)-N^2-phenylacetyl-2'-deoxy-6-(2, 4-dinitrophenyl)thioguanosine-3'-O-(2-cyanoethyl-N, N-diisopropylamino)-phosphoramidite (185)

N^2-phenylacetyl-2'-deoxy-6-(2, 4-dinitrophenyl)thioguanosine (183) (0.58 g, 1 mmol) was treated with dimethyltrityl (DMT) chloride (1.1 equivalents) in dried pyridine (10 ml) overnight. The reaction solution was reduced to a small volume, then taken into $CHCl_3$ (40 ml), washed with saturated aqueous $NaHCO_3$ (3 x 40 ml). The organic layer was separated and dried over anhydrous Na_2SO_4. The Na_2SO_4 was filtered and washed with $CHCl_3$ (2 x 5 ml). The filtrate was concentrated and coevaporared with toluene to give a yellow foam, then purified by column chromatography with $CHCl_3$ as eluent to give the corresponding 5'-O-DMT derivative, yield 0.63 g, (74.8%). The resulting compound (184) was confirmed by NMR spectroscopy; ^1H NMR data (in DMSO-d_6): 2.35-2.89 (2H, m, 2'-H and 2"-H), 3.08-3.24 (2H, m, 5'-H), 3.71 (8H, s, -CH_2- of phenylacetyl and CH_3O of DMT), 3.96 (1H, m, 4'-H), 4.51 (1H, m, 3'-H), 5.38 (1H, d, 3'-OH, ex), 6.38 (1H, t, 1'-H), 6.66-7.35 (18H, m, aromatic-H of phenylacetyl and of DMT), 8.04 (1H, d, 6-H of 2,4-dinitrophenyl), 8.25 (1H, m, 5-H of 2,4-dinitrophenyl), 8.55 (1H, s, 8-H), 8.87 (1H, d, 3-H of 2,4-dinitrophenyl) and 10.82 (1H, s, 2-NH, ex). The compound (184) was then converted into compound (185) using the same procedure as for compound (177) (section 4.2.3.iii). ^{31}P NMR data (in $CDCl_3$): 149.41 and 149.16.

4.2.5. Stability Tests

4.2.5.i. Stability of Compound (183) and (184) Towards the Conditions Used in Oligonucleotide Assembly

Compound (183) (20 mg) was dissolved in 3% dichloroacetic acid/dichloroethane and changes in the solution monitored by TLC ($CH_3OH/CHCl_3$,

10: 90, v/v) over 24 h at room temperature. Compound (184) (20 mg) was dissolved respectively in a) 4.4% N-methylimidazole in THF; b) acetic anhydride/ lutidine/THF (1:1:8); and c) 0.1 M iodine in THF/pyridine/water (40:9:1). Changes in the solution were monitored by TLC ($CH_3OH/CHCl_3$, 5: 95, v/v) over 24 h at room temperature.

4.2.5.ii. Stability of 2'-Deoxy-6-thioguanosine Towards Conc. Ammonia at 25°C and 55°C

2'-Deoxy-6-thioguanosine (10 mg) was dissolved in 1 ml of conc. ammonia and the solution was kept at 25°C for 3 days and at 55°C for a day respectively. The reaction course was monitored by TLC ($CH_3OH/CHCl_3$,15: 85, v/v).

4.2.6. Removal of N^2-phenylacetyl Group from N^2-phenylacetyl-6-thiodeoxyguanosine (182)

The sample was treated with aqueous ammonia in a sealed Eppendorf tube at 25°C. The aliquots were taken at different time intervals and analyzed by reverse-phase HPLC (Figure 4.5). The column was eluted with 93% buffer A and 7% buffer B for the first 8 min, then with a linear gradient from 7% to 80% of buffer B over the following 7 min, and then remaining in this concentration for another 5 min. The amount of 6-thiodeoxyguanosine and its N-2 protected derivative was measured from the absorbance at 335 nm of each chromatographic peak.

4.2.7. Synthesis of Oligonucleotides

Oligonucleotides were synthesized on a Cruachem PS200 automatic DNA synthesizer (Cruachem Ltd., Glasgow, Scotland) using PAC amidites of the normal bases. The general procedure was carried out as described in section 2.2.3.

4.2.8. Conversion and Purification of Oligonucleotides

4.2.8.i. Preparation of Oligomers Containing O^6-methylguanine (G^{OMe})

CPG-support bearing the 12 mer $G^{S\emptyset}$ oligomer, with the 5'-DMT still on (CPG-$G^{S\emptyset}$ oligomer) was put into Eppendorf tubes and MeOH/DBU (1ml; 9:1, v/v) was added and left at 25°C for 2 days. The solution was neutralized (90 µl 50% aqueous acetic acid,1.5 equivalent to DBU) and immediately passed through a Dowex 50 x 8, Na^+ form, 400 mesh ion exchange column (10 ml wet volume) eluted with water and collected in 1 ml fractions. The oligomers were usually found in fractions 4 to 6 by measuring A_{260}. The oligonucleotides were separated from failure sequences and the DMT group finally removed using a Nensorb Prep cartridge (Du Pont company) as described in section 2.2.5. Further purification of the modified oligomer by FPLC was carried out on a Dionex BIOLC system with a Dionex variable wavelength detector using Pharmacia monoQ HR 5-5 column. The column was eluted with 100% of buffer C for 2 min, then with a linear gradient from 100% of buffer C to 90% of buffer C and 10% of buffer E over the following 3 min, then from 10% to 30% of buffer E over the next 20 min, and then remaining in this buffer condition for another 3 min.The desired peak was collected, immediately neutralized with 1 M CH_3COOH, and desalted as described in section 2.2.5.

4.2.8.ii. Preparation of Oligomers Containing Guanine from the $G^{S\emptyset}$ Oligomer

The CPG-$G^{S\emptyset}$ oligomer was treated with 0.5 M aqueous NaOH (25°C, 2 day). The deprotected and substituted oligomer was purified with a Nensorb Prep cartridge and FPLC as above.

4.2.8.iii. Preparation of Oligomers Containing 6-Thioguanine (G^S)

The CPG-$G^{S\emptyset}$ oligomer was treated with 10% mercaptoethanol in conc. ammonia for 2 day at 25°C, then the product was purified by Nensorb cartridge and

FPLC as described above.

4.2.8.iv. Preparation of Oligomers Containing 2, 6-Diaminopurine (G^{NH2})

The CPG-$G^{S\emptyset}$ oligomer was treated with conc. aqueous ammonia (d=0.88) containing 65 mM tetramethylguanidine and 75 mM 2-nitrobenzaldoxime for 2 days at 25°C. The resulting oligomer was purified with a Nensorb Prep cartridge and further with FPLC.

4.2.8.v. Preparation of Oligomers Containing 2-Amino-6-methylaminopurine (G^{NMe})

The CPG-$G^{S\emptyset}$ oligomer was treated with 40% aqueous methylamine for 2 days at 25°C. The resulting oligomer was purified with a Nensorb Prep cartridge and further with FPLC.

4.2.9. Base Analysis

The purity of the oligomers was assessed by chromatographic separation and quantitation of the nucleosides obtained from enzymic digestion (see section 2.2.6). The nucleosides were analyzed by HPLC using 93% buffer A and 7% buffer B for the first 7 min, then with a linear gradient from 7% to 80% of buffer B over the following 10 min, then remaining at this concentration for another 5 min in order to detect any remaining un-substituted $G^{S\emptyset}$ nucleoside. The eluate was generally monitored at 260 nm, but for the G^{S} oligomer an extra run was monitored at 335 nm for the detection of 6-thiodeoxyguanosine. Retention times were: dC:1.5 min; dI (from enzymatic hydrolysis of dA): 3.5 min; dG: 3.9 min; T: 5.5 min; G^{S}: 5.7 min; G^{NH2}: 7.0 min; dA: 9.5 min; G^{OMe}: 11.7 min; G^{NMe}: 12.5 min; $G^{S\emptyset}$: 20.2 min.

4.2.10. Melting Curve Measurement

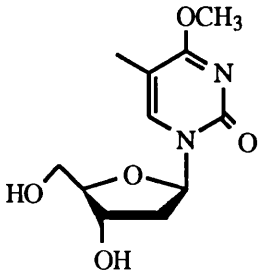
This was carried out as described in section 2.2.8.

4.3. RESULTS AND DISCUSSIONS

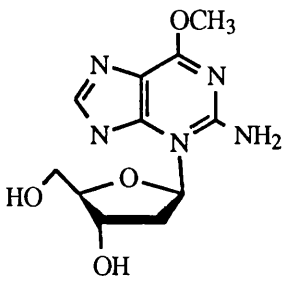
4.3.1. Design and Preparation of Versatile Guanine Monomer

For the strategy to be successful, the monomer must contain at the 6-position of guanine a group which is sufficiently stable to withstand the conditions of DNA synthesis but is also sufficiently reactive to be substituted by nucleophiles after synthesis. Comparing a similar functional group at the 4-position of thymine and at the 6-position of guanine, one finds that the latter is more stable towards nucleophilic reagents (eg. MeOH/DBU, conc. ammonia), which makes substitution more difficult. For example, in conc ammonia the half-life of O^4-methylthymidine at room temperature is approximately 10 hours with 5-methylcytidine as the main product (Xu and Swann, 1990), and consequently concentrated ammonia cannot be used for the deprotection of oligonucleotides containing this modified base; in contrast, oligomers containing O^6-methyldeoxyguanosine (167) can be safely deblocked with ammonia at room temperature without detectable formation of 2, 6-diaminopurine (Smith *et al.*, 1990). Since 2'-deoxy-6-mesitylenesulphonylguanosine (2'-deoxy-6-Ms-guanosine) (168) can be converted easily into O^6-alkylguanosine by the successive treatments with N-methylpyrrolidine, then with alcohol/DBU (Reese and Skone, 1984), initially an attempt had been made by Dr Xu to use 6-Ms-guanine as a versatile guanine. 6-Ms-guanine monomer (169) was prepared in one step from commercial guanine phosphoramidite monomer and incorporated into oligomers with satisfactory coupling yield. However, the oligomers containing 6-Ms-guanine did not give a satisfactory amount of the expected 6-substituents. A possible explanation is that 6-mesitylenesulphonyl group of guanine was unstable to the conditions of DNA synthesis or this group was less reactive at the oligomer level. So our attention was turned towards construction of another versatile guanine monomer. Concurrently with this work simple methods for direct thiation of the 4-position of pyrimidine nucleosides and the 6-position of deoxyguanosine were independently developed by us (Xu *et al.*, 1991) and others (Kung and Jones, 1991; Nikiforov *et al.*, 1991). Because the C-S

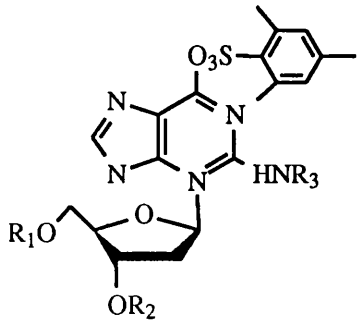
bond is more easily split than C-O, it seemed reasonable to design a versatile monomer based on thioguanine rather than on guanine itself.



(166)



(167)

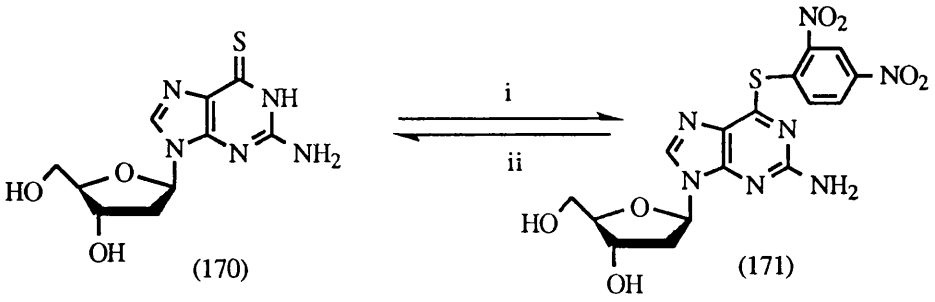


(168) $R_1 = R_2 = R_3 = H$

(169) R_1 = DMT, $R_2 = NCCH_2CH_2O\text{-}P(\text{i-Pr})_2N$, $R_3 = COCH_2OPh$.

The 6-thioketo function of 6-thiodeoxyguanosine is easily oxidized by many oxidation reagents, such as iodine which is used when oligonucleotides are synthesized by phosphoramidite chemistry. Furthermore the sulphur is nucleophilic; for example it can react with reagents such as methyl iodide forming alkylated derivatives. Thus there is a possibility of side reactions with activated nucleoside phosphoramidites during DNA synthesis with 6-thiodeoxyguanosine. Therefore the 6-thioketo group must be protected with a suitable group to prevent it from oxidation and side reactions. Such a protecting group must be stable during DNA synthesis but also should make the 6-substituent a better leaving group. In the only published paper on the synthesis of 6-thioguanine containing oligomers at the time when we started this work (Rappaport, 1988) the 6-thio group was not directly protected, but the author assumed that the thio group had been automatically protected by mesitylene sulphonyl group during its condensation steps with sulphonyl triazole derivative, a coupling agent used in the phosphotriester method. However, our preliminary experiments failed to obtain the expected product when 5'-O-DMT-N^2-isobutyryl-6-thiodeoxyguanosine (173) was treated with 2-mesitylene sulphonyl chloride. We therefore turned to the 2,

4-dinitrophenyl group for protection of 6-thio function of thioguanine. Preliminary experiments showed that this protecting group could easily be put onto the 6-thio group of 6-thiodeoxyguanosine (170) with 2, 4-dinitrofluorobenzene, and removed selectively from (171) with 2-mercaptoethanol (Figure 4.1), which makes synthesis of 6-thioguanine containing oligomers possible; and it also makes the 6-substituent as a whole a better leaving group, providing the possibility of producing other modified oligomers.



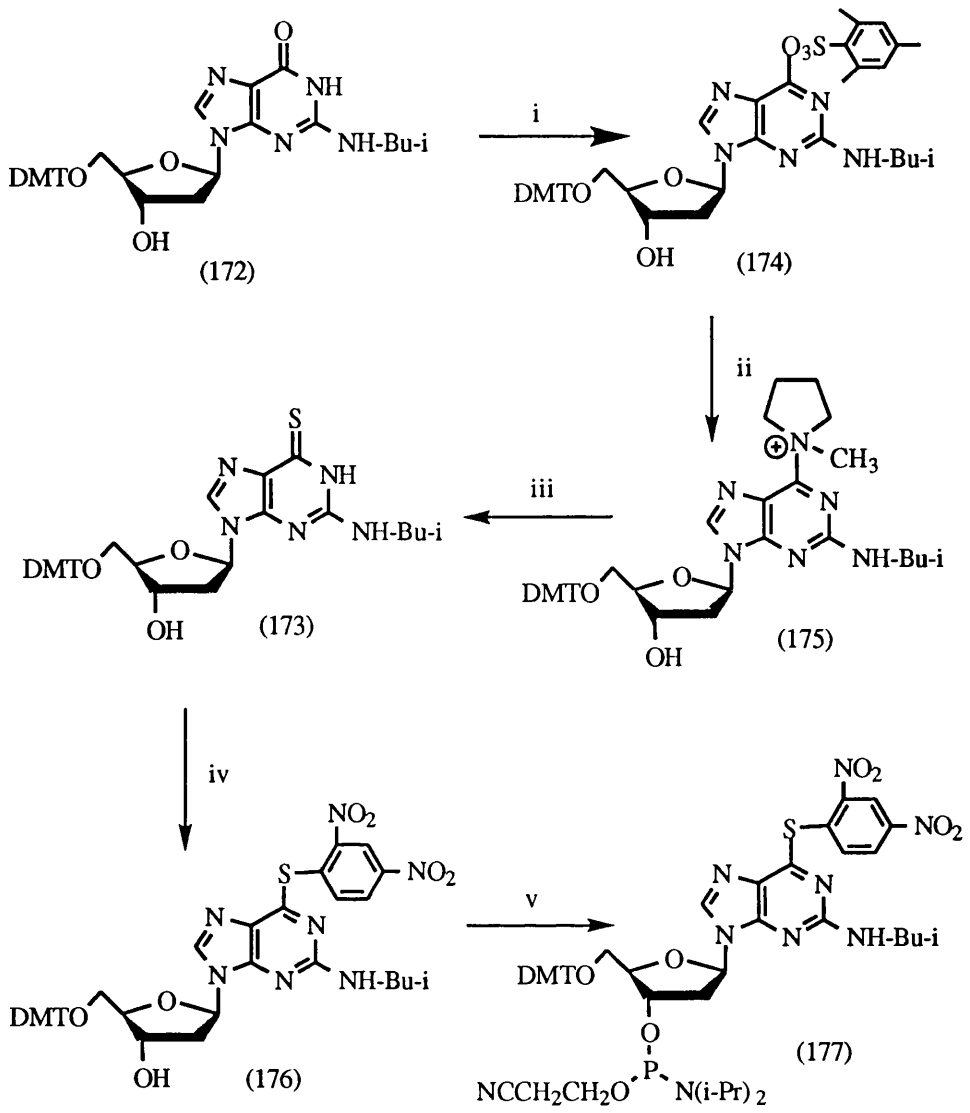
i: 2, 4-dinitrofluorobenzene/triethylamine. ii: 2-mercaptoethanol

Figure 4.1. Reaction scheme for (i) protecting the 6-thio function of 6-thiodeoxyguanosine with 2, 4-dinitrophenyl group; and (ii) removing the protecting group from 6-(2, 4-dinitrophenyl)thiodeoxyguanosine with 2-mercaptoethanol.

Another important point is the choice of protecting group on the 2-amino function of guanine. At first, the conventional protecting group, isobutyryl, was used because the nucleoside, 5'-O-dimethoxytrityl-N^2-isobutyryl- 2'-deoxyguanosine (172, Figure 4.2), is commercially available. It was converted into its 6-thio-analogue (173) by one-pot reaction. The 6-position of the compound (172) was first sulphonated with 2-mesitylenesulphonyl chloride (Ms-Cl), the resulting Ms derivative (174) was displaced by N-methylpyrrolidine, and the quaternary amino derivative (175) was then substituted with thiolacetic acid forming the 6-thio-analogue (173). Although the 6-keto and 3'-OH can both be sulphonated by Ms-Cl the former is more nucleophilic under these conditions and therefore the main product was O^6-sulphonated. Furthermore the Ms group on the 6-position can be substituted by N-methylpyrrolidine, then by

thiolacetic acid whereas that group on the 3'-position is unreactive with both reagents. The desired product (173) can be separated from impurities by a single silica gel column chromatography, and then converted into the 5' and N^2-protected 6-(2, 4-dinitrophenyl)thiodeoxyguanosine (176) with high yield by treatment with 2, 4-dinitrofluorobenzene in the presence of triethylamine (1h, 25°C). There was no detectable formation of 3'-O-(2, 4-dinitro)-phenyl derivative. Compound (176) was then converted to the phosphoramidite monomer (177) by treatment with 2-cyanoethyl-N, N-diisopropylchlorophosphoramidite in the presence of N, N-diisopropylethylamine. The monomer was incorporated into a dodecamer AGCYAATTCGCT (Y: $G^{S\emptyset}$) with satisfactory coupling yield (>98%) and then converted into the desired modified guanine (G^S, G^{OMe}) by treatment with the appropriate reagents.

The above procedure to prepare the versatile monomer has the advantage of a straightforward synthetic route and good overall yield (note that the first three steps were carried out in one-pot reaction). However, removal of the isobutyryl group protecting the N^2-position of this modified guanine required much longer exposure to deblocking reagents (eg. 3 days, at room temperature in conc. ammonia) than was required for removal of the other bases (PAC protection groups, 5 h at room temperature in conc ammonia). This prolonged deprotection step increases the probability of substituting the 6-modification by an amino function. To overcome this, the 6-(2, 4-dinitrophenyl)thioguanine monomer with a more base-labile group, the phenylacetyl group, on the N^2-position was prepared as illustrated in Figure 4.3 and used in all the following experiments.



i: mesitylenesulphonyl chloride. ii: N-methylpyrrolidine. iii: thiolacetic acid.
iv: 2,4-dinitrofluorobenzene. v: phosphitylating reagent.

Figure 4.2. The synthetic route for N^2-isobutyryl-6-(2, 4-dinitrophenyl)thiodeoxy-guanosine phosphoramidite.

2'-Deoxyguanosine (178) was first treated with acetic anhydride and the resulting product 3', 5'-diacetyl-2'-deoxyguanosine (179) was further treated with phenylacetic anhydride to produce a fully protected 2'-deoxyguanosine derivative (180). This compound was converted into the 6-thio analogue (181) with isolated yield of 70-80% by reaction with mesitylene sulphonyl chloride, then with N-methylpyrrolidine, followed by thiolacetic acid. The 6-thio analogue (181) was

selectively deacylated with NaOH in pyridine solution. This step inevitably leads to some loss of the phenylacetyl group from the N^2-position and therefore it is essential to control the reaction time. Li and Swann (1989) deacylated the O^6-alkyl-2'-deoxyguanosine analogue with NaOH in 6 min. We adopted this time scale for deacylation of the 6-thio analogue and no detectable (on TLC) deprotection on N^2-position was observed. The resulting N^2-phenylacetyl-2'-deoxy-6-thioguanosine (182) reacted with 2,4-dinitrofluorobenzene in the presence of triethylamine to give compound (183) with isolated yields of 80-90%. The reaction was again highly regioselective because the 6-thioketo group of (182) is much more nucleophilic than 3' and 5' hydroxyl groups. Using standard methods, compound (183) reacted with DMT-Cl/pyridine, then with phosphitylating reagent to give the designed phosphoramidite monomer (185).

All these intermediates were characterized by ^{1}H-NMR and gave satisfactory data. The peak at 342 nm in the UV spectrum of 6-thiodeoxyguanosine at pH=7 (Figure 4.4)) is characteristic of the C=S function and provides a means of checking for its presence after thiolation. The UV spectra of 5'-O-DMT-N^2-isobutyryl-2'-deoxy-6-thioguanosine (173) and 3', 5'-diacetyl-N^2-phenylacetyl-2'-deoxy-6-thioguanosine (181) confirmed the presence of C=S group by their λ_{max} at 332 nm.

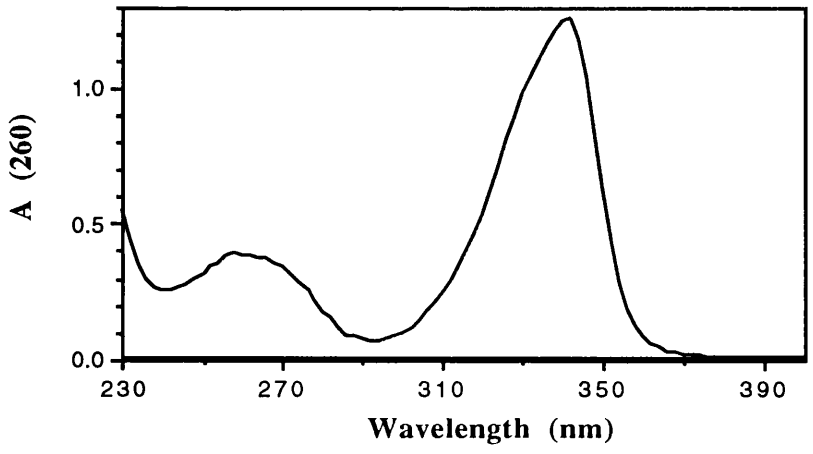
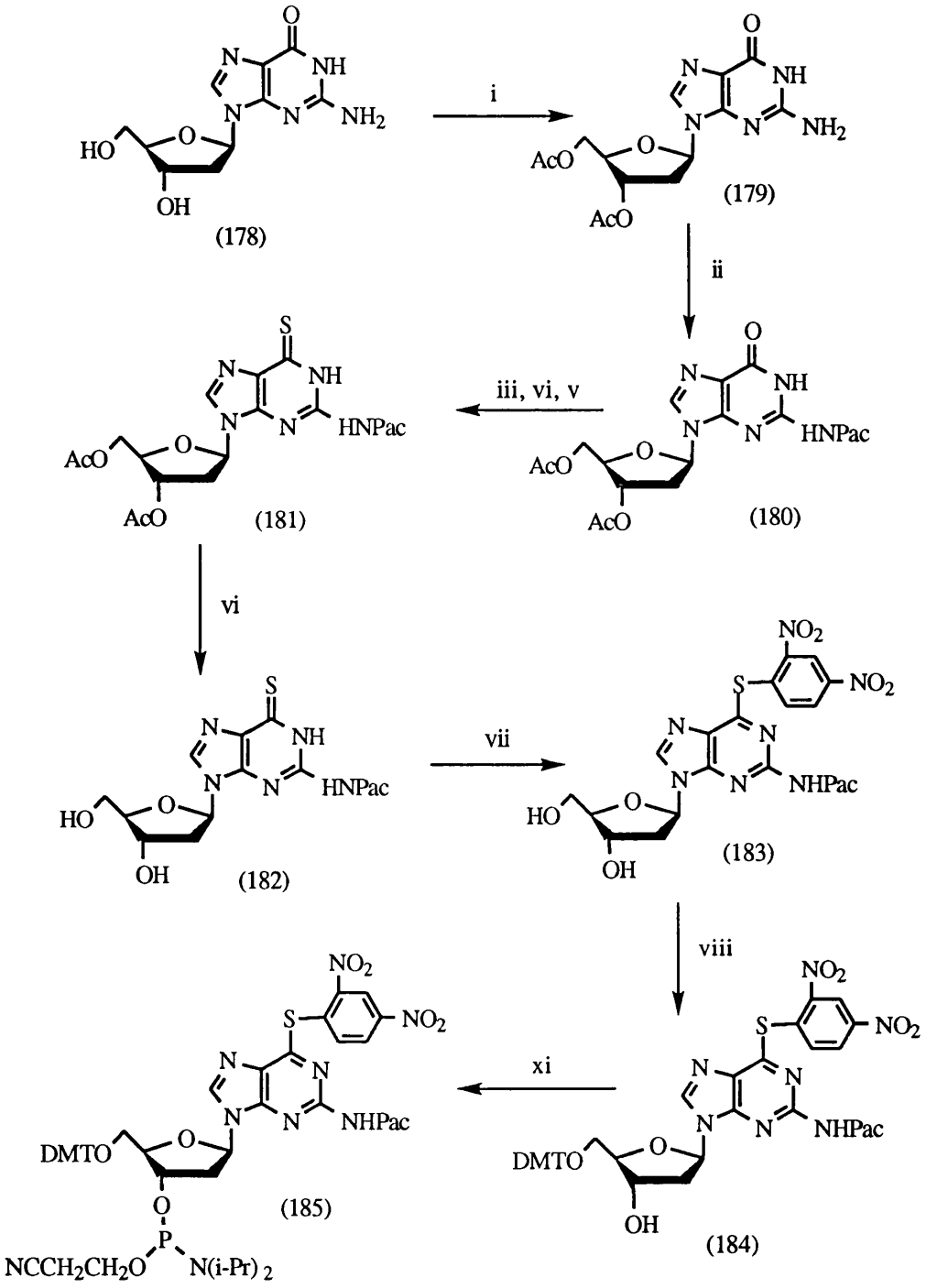


Figure 4.4. The UV spectrum of thiodeoxyguanosine at PH 7.



i: acetic anhydride. ii: phenylacetic anhydride
iii: mesitylenesulphonyl chloride. iv: N-methylpyrrolidine.
v: thiolacetic acid. vi: 2 M NaOH. vii: 2,4-dinitrofluorobenzene.
viii: 4,4'-dimethoxytriphenylmethyl chloride/pyridine.
ix: phosphitylating reagent. Ac: acetyl. Pac: phenylacetyl.

Figure 4.3. The synthetic route for N^2-phenylacetyl-6-(2, 4-dinitrophenyl)thiodeoxy-guanosine phosphoramidite.

4.3.2. Removal of N^2-phenylacetyl Group from Compound (182)

The phenylacetyl group used for the protection of the N^2-position of O^6-alkyl-2'-deoxyguanosine monomers has a half-life of less than 1 hour in aqueous conc. ammonia at 22°C (Li and Swann, 1989; Smith *et al.*, 1990). To investigate its removal from N^2-phenylacetyl-6-thiodeoxyguanosine (182) was treated with conc. ammonia in a sealed Eppendorf tube at 25°C and the deblocking course was analyzed by reverse phase HPLC at different time intervals (Figure 4.5). The amount of 6-thio-2'-deoxyguanosine and its N^2-protected analogue were measured from the absorbance of each chromatographic peak at 335 nm. The half-life for removal of the phenylacetyl group (182) was 1.2 hour at 25°C.

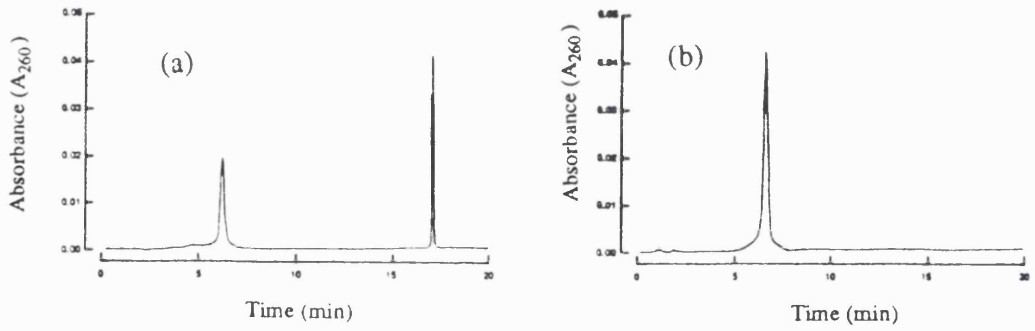


Figure 4.5. Reverse-phase HPLC showing the rate of removal of the phenylacetyl group from N^2-phenylacetyl-6-thiodeoxyguanosine. (a) is the HPLC profile of the mixture after 1 h reaction showing that less than half the N^2-protected nucleoside was converted into the deprotected nucleoside; (b) is the HPLC profile after 17 h reaction showing complete removal of the N^2-protecting group. The chromatographic conditions are as described in section 4.2.6.

4.3.3. The Stability of 2'-Deoxy-6-(2,4-Dinitrophenyl)thioguanosine Monomer to the Conditions of DNA Synthesis

The stability of 6-(2, 4-dinitrophenyl)thioguanine monomer to the reagents used

in oligonucleotide synthesis was studied before the monomer was used for DNA synthesis. Compound (183) was dissolved in 3% dichloroacetic acid/dichloroethane (deblocking reagent); and compound (184) was dissolved in acetic anhydride/lutidine/THF (1:1:8) (capping reagent A), 4.4% N-methylimidazole in THF (capping reagent B); and 0.1 M iodine in THF/pyridine/water (40:9:1) (oxidation reagent). By monitoring changes in these solutions, it was shown that compound (184) was stable towards the reagents used in oligomer assembly for at least 24 h at room temperature, and compound (183) was stable towards 3% dichloroacetic acid in dichloroethane at room temperature for at least 3 h. It was also found that the 2, 4-dinitrophenyl group protecting the 6-thioketo function can be cut off completely by 1 M mercaptoethanol in CH_3CN in the presence of triethylamine within 30 min. These results suggested that the monomer (185) would be stable during the synthesis, and that the protecting group easily removed after synthesis.

4.3.4. Synthesis and Conversion of Oligonucleotides

The versatile monomer (185) was incorporated into oligonucleotides by the DNA synthesizer without changing the normal program except that (185) was injected manually with 250 μl gas tight syringe and coupling time for (185) was 3 mins. The coupling yield was calculated to be about 98% by measuring the amount of 5'-protecting group DMT released by dichloroacetic acid. The two steps needed to get the desired modified oligomers, substitution and deprotection, were investigated in detail. In all of the syntheses in this work the normal bases were protected with labile groups (Schulhof *et al.*, 1987) (PAC monomers, Pharmacia) and the conditions under which these protecting groups can be removed by conc. ammonia, alcohol/DBU and aqueous alkali at room temperature have been determined in Chapter 2. So the main work was the substitution and deprotection of the modified guanine. After synthesis the CPG-support bearing the synthetic 12 mer (AGCG$^{S\emptyset}$AATTCGCT) with 5'-DMT still attached was divided into parts, each of which was treated with appropriate chemical

reagents forming a series of different modified oligomers (Figure 4.6).

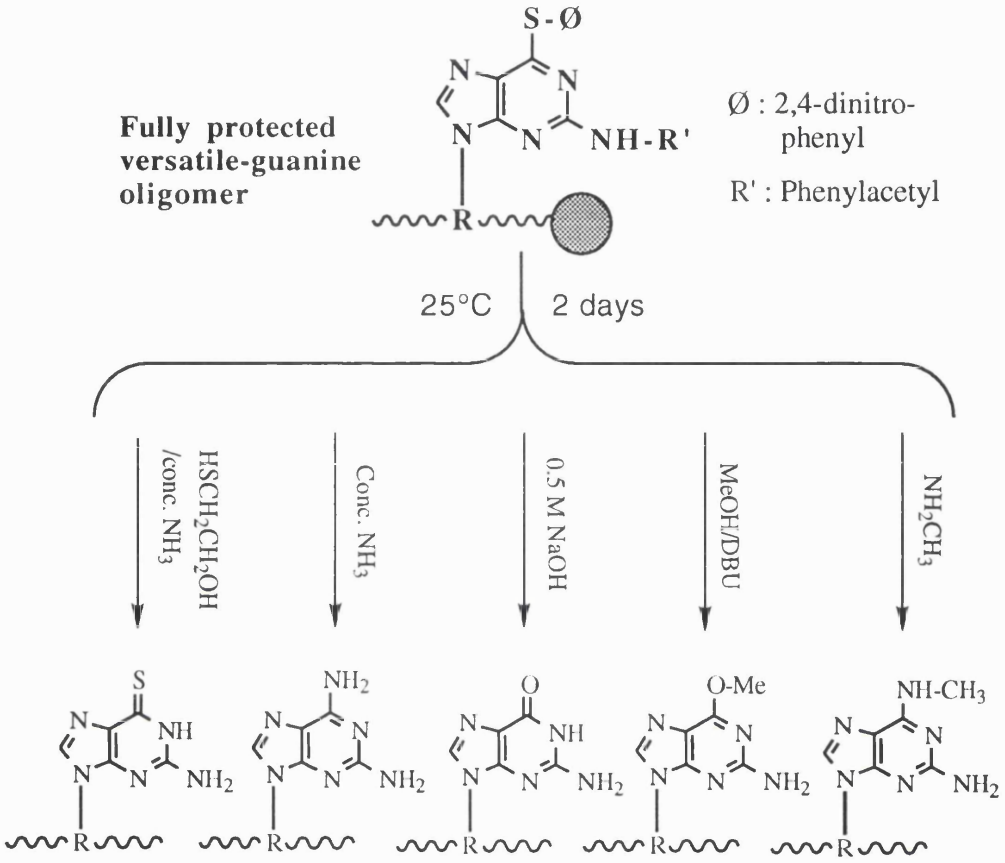


Figure 4.6. Conversion of the oligonucleotide containing $G^{S\emptyset}$ into a series of oligonucleotides containing different modified guanines.

4.3.4.i. Oligomers Containing 6-Thioguanine (G^{S})

For several reasons synthesis of oligonucleotides containing 6-thioguanine is of great interest. 6-Thioguanine has been used as an antitumour drug for more than thirty years and it has been suggested that its mode of action involves incorporation of thioguanine into DNA (LePage and Jones, 1961; Christie *et al.*, 1984). Thus the properties of DNA containing 6-thioguanine are potentially relevant to cancer chemotherapy. Furthermore since hydrogen bonds between DNA binding proteins and the 6-keto oxygen of deoxyguanosine contribute to the specificity of protein-DNA

interaction, G^S oligomers are useful probes for the study of DNA-protein interactions. Finally oligonucleotides containing 6-thioguanine are photoreactive and they can be used for photoaffinity labelling of DNA binding proteins (Nikiforov and Connolly, 1992).

Treatment of the protected and CPG bound oligomer containing N^2-phenylacetyl-6-(2,4-dinitrophenyl)thioguanine (CPG-$G^{S\emptyset}$ oligomer) with mercaptoethanol at room temperature for 3 h completely cleaved the 2, 4-dinitrophenyl group from sulphur. The course of this deprotection can be monitored by eye since the removal of this group changed the colour of the support from yellow to nearly white. As this treatment did not cleave the oligomer from the support, the excess reagent was washed off with acetonitrile. Conc. ammonia was then employed to cut all other protecting groups and produced an oligomer containing 6-thioguanine. It has been reported that 2'-deoxy-6-thioguanosine is labile to aqueous alkali (Rappaport, 1988), but we found that 2'-deoxy-6-thioguanosine and G^S oligomer are stable to conc. ammonia for at least 3 days at room temperature, but unstable at 55°C. Therefore we used conc. ammonia at 25°C to deprotect oligomers containing 6-thioguanine. Further optimization revealed that removal of 2, 4-dinitrophenyl was very rapid in ammonia containing 10% of mercaptoethanol so that this mixture solution at room temperature was added for two days to give the desired product. The crude oligomer was purified in the same way as normal oligomer by Nensorb cartridge and its purity was good enough for many biological applications (Figure 4.10e). Highly pure G^S oligomer can be directly obtained by FPLC (Figure 4.10j), in which the desired G^S oligomer was well separated from impurities.

The UV spectrum of the thioguanine containing oligomer is shown in Figure 4.7, together with that of the control 12 mer and the free 6-thio-2'-deoxyguanosine. The absorption band at ~340 nm due to the presence of the modified base is clearly visible. Like the oligomers containing 4-thiothymine, this extra band is particularly useful for a quick check of the presence and the integrity of the modified base by UV

spectrophotometry and for the detecting the thio-oligomers in chromatography by their absorption at a wavelength at which normal oligomers have no absorption.

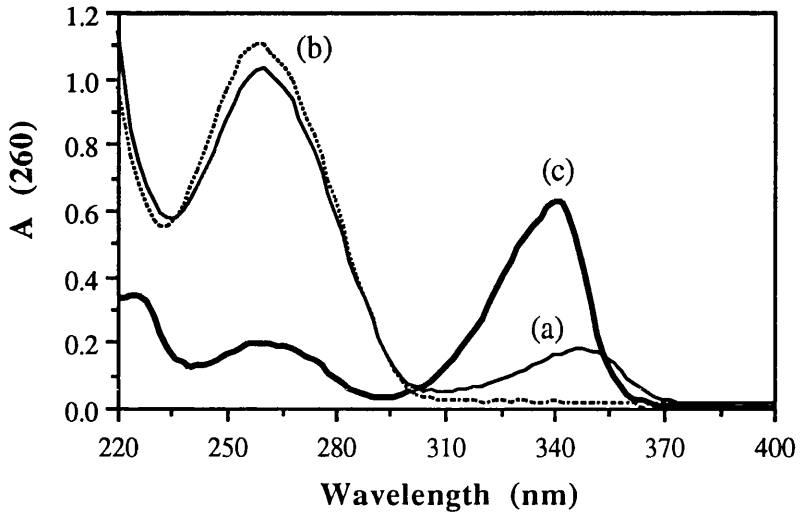


Figure 4.7. UV spectrum of d(AGCG^{S}AATTCGCT) (a), of d(AGCGAATTCGCT) (b), and of 6-thiodeoxyguanosine (c).

The present method provides a much simpler and more efficient procedure to obtain G^{S} oligomers than the only method (Rappaport, 1988) which had been published at the time when this work was started, where G^{S} oligomers was prepared by the phosphotriester procedure with the conventional acyl groups protected on the bases. The greatest difficulty in that paper was the deprotection step, where 1 M benzenethiol in pyridine (8 h) was first used for removal of the group on 6-thio position, tetramethylguanidine and E-2-nitrobenzaldoxime in pyridine (5 days) then used to cleave the oligomer from its support and to remove the protecting groups on the phosphates, then NH_3 in anhydrous methanol (7 days) used to remove protecting groups from bases. All these operations were carried out under nitrogen in anhydrous condition (Rappaport, 1988). A significant omission from that report was any evidence to show that all the final product contained thioguanine and that loss of the sulphur had not taken place since the 6-thioketo function was not protected. When we were

preparing the manuscript of this work for publication, an oligomer containing G^S (AAACG^{S}TTT) was synthesized using the 2-cyanoethyl group to protect the thio function of 6-thioguanine phosphoramidite (Christopherson and Broom, 1991), but there was no direct evidence that under their separating system (reverse phase HPLC, pH=7.6) the G^S oligomer can be separated from oligomers containing G. So it is still not certain that the G^S oligomer was free from the parent. We have used an anion exchange column at pH 12 to give good separation of G^S oligomer from other oligomers of the same length (Figure 4.10). More publications have recently appeared for the synthesis of oligomer containing thioguanine (Sudhakar *et al.*, 1992; Waters and Connolly, 1992). Those reports used the conventional approach for oligonucleotide synthesis (refer to Figure 1.30a). Waters and Connolly (1992) protected the 2-amino function with a benzoyl group which was removed from the oligomer by 4 h treatment with conc ammonia at 55°C which inevitably led to the formation of 2, 6-diaminopurine. Sudhakar et al. (1992) used trifluoroacetyl or isobutyryl groups for 2-amino protection which are difficult to remove.

4.3.4.ii. Oligomers Containing 2, 6-Diaminopurine (G^{NH2})

Initially the CPG-$G^{S\emptyset}$ oligomer was treated with conc. ammonia at 55°C overnight to cleave the oligomer from the CPG, remove the protecting groups and substitute the 2, 4-dinitrobenzenethio group, but besides the desired G^{NH2} oligomer a substantial amount of oligomers containing G^S and guanine itself was obtained. Presumably the G^S oligomer was formed because the nucleophile ammonia or hydroxide ion attacked the 1-position of the dinitrophenyl group leaving sulphur in the 6-position of guanine (routes c and d, Figure 4.8), and the parent oligomer was produced by the attack of hydroxide ion on 6-carbon of guanine (route b, Figure 4.8). The amount of G^S oligomer was reduced dramatically by lowering the temperature, and treatment with conc. ammonia at 25°C (2 days) gave the desired G^{NH2} oligomer with negligible amount of G^S oligomer. However there was still an unacceptable amount of

oligomer containing guanine resulting from the attack by hydroxide ion. It has been previously observed that conc. ammonia containing tetramethylguanidine (TMG) and E-2-nitrobenaldoxime removed the protecting groups from oligonucleotides faster than ammonia alone (Li and Swann, 1989), suggesting that TMG and the oxime must favour the role of ammonia as a nucleophile. So treatment of the CPG-$G^{S\emptyset}$ oligomer at 25°C with conc. ammonia containing TMG/oxime for 2 days produced G^{NH2} oligomer with negligible amount of oligomer containing G^{S} and G after Nensorb column purification (Figure 4.10a).

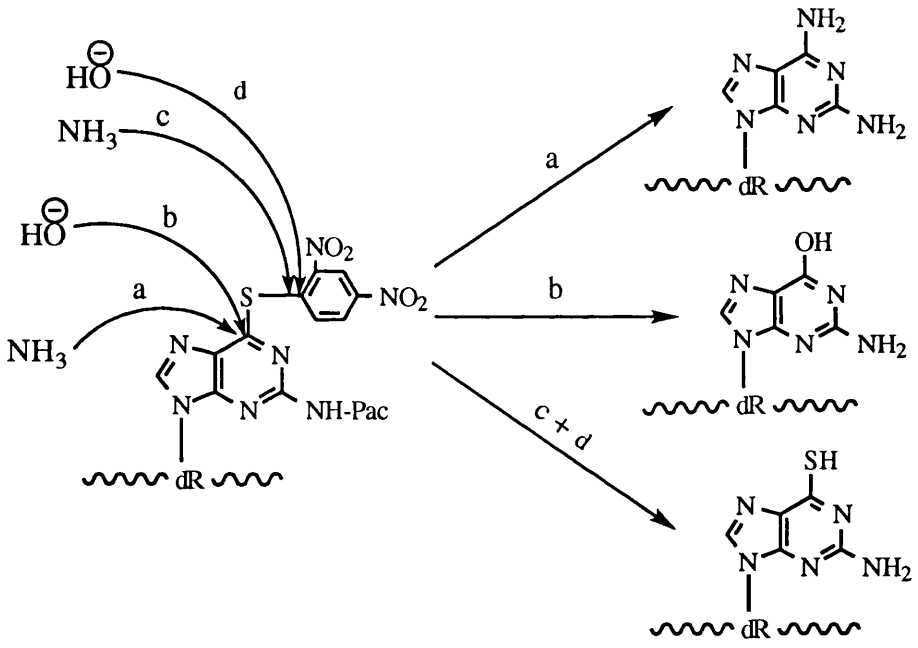


Figure 4.8. Possible reaction routes of the versatile oligomer containing 6-(2,4-dinitrophenyl)thioguanine with conc. ammonia at 55°C.

Interest in synthetic oligomer containing 2, 6-diaminopurine has been greatly stimulated by the finding that S-2L cyanophage DNA contains 2, 6-diaminopurine instead of adenine (Kirnos *et al.*, 1977) and G^{NH2} is believed to have an extra hydrogen bond in base-pairing with thymine (Lamm *et al.*, 1991; Sproat *et al.*, 1991). Preparation of short oligomers containing G^{NH2}, in which an N^2, N^6-diacyl-protected monomer was used, has been reported (Gaffney *et al.*, 1984a; Chollet *et al.*, 1986),

but the procedures are very difficult not only in preparation of the modified monomer but also in assembly and deprotection of the oligomer. Due to the extreme acid lability of the glycosidic linkage of the 2, 6-di-N-acyl-derivative, the aprotic acid, $ZnBr_2$, was used for detritylation, which inevitably led to incomplete detritylation (Gaffney *et al.*, 1984a). To circumvent this, Chollet et al. (1986) used 1-methyl-2, 2-diethoxypyrrolidine for protection of the 6-amino group to reduce the depurination. However, deprotection in conc. ammonia at 65°C for 5-7 days was still necessary because of the slow hydrolysis of the two amino-protecting groups on the modified base. It is clear that these methods are far from satisfaction and needed to be considerably improved, and in their later paper Chollet et al. (1988) used a DNA polymerase to incorporate the triphosphate of 2, 6-diaminopurine nucleoside to produce G^{NH2} oligomers rather than their published chemical approach. The present method provides a more simple and practical way to prepare oligomers containing 2, 6-diaminopurine than previous methods (Gaffney *et al.*, 1984a; Chollet *et al.*, 1986) and furthermore allows one to introduce ^{15}N using easily available ^{15}N ammonia.

4.3.4.iii. Oligomers Containing 2-Amino-6-methylaminopurine (G^{NMe})

The CPG-$G^{SØ}$ oligomer was treated with methylamine (25°C, 2 days) to effect both substitution and deprotection and give an oligomer containing 2-amino-6-methylaminopurine. This gives a good example of the post-synthetic substitution to make oligomers containing different alkylamines. Because alkylamines are generally good nucleophiles so that they can easily be introduced into 6-position of guanine at the oligomer level by the present method. It is noteworthy that an oligomer containing a bifunctional group, eg. ethylene diamine could be useful for DNA crosslinking or attachment of non-radioactive labeling reagents.

4.3.4.iv. Oligomers Containing O^6-methylguanine (G^{OMe})

There has been tremendous interest in the synthesis and investigation of

oligomers containing O^6-methylguanine because this base is believed to play the most important role in the carcinogenic action of N-nitroso compounds (Swann, 1990). Although several methods for the synthesis of oligomers containing O^6-methylguanine have been developed in our lab (Li and Swann, 1989; Smith *et al.*, 1990) and by others (Fowler *et al.*, 1982; Gaffney *et al.* , 1984b), post-synthetic substitution provides an alternative with the advantage that NMR sensitive group such as ^{13}C from ^{13}C-CH_3OH could be introduced at the last step. Treatment of the CPG-$G^{S\emptyset}$ oligomer with MeOH/DBU (25°C, 2 days) effected both substitution and deprotection to give the G^{OMe} oligomer. It is interesting to note that when a higher temperature (35°C) was employed overnight, a greater percentage of G^S oligomer was produced (Figure 4.9). Both 2, 4-dinitrophenyl and 2, 4-dinitrobenzenethio groups are good leaving groups and presumably higher temperature favours the substitution of the former group to give G^S oligomer. Oligomers prepared by this method are completely identical to ones prepared by the previous methods (Li and Swann, 1989; Smith *et al.*, 1990).

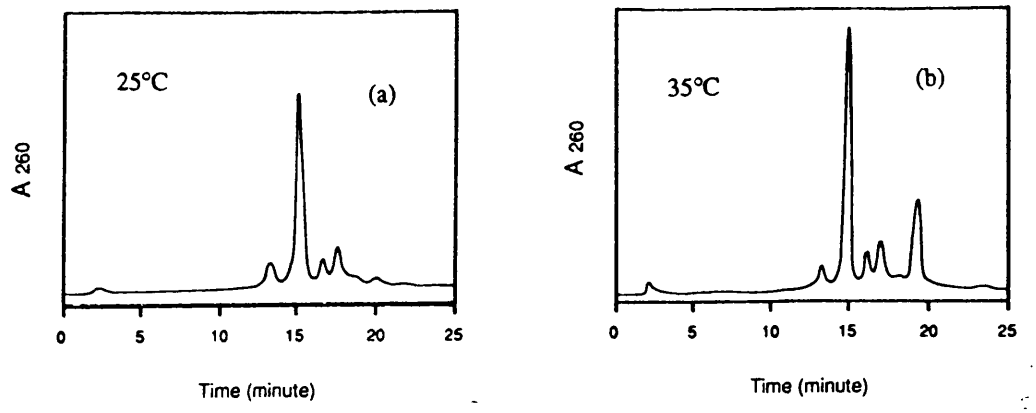


Figure 4.9. FPLC profiles of unpurified 12 mer AGC G^{OMe}AA TTC GCT prepared by post-synthetic substitution with MeOH/DBU at 25°C (a), and at 35°C (b). The Figure b shows that if the substitution is carried out at slightly higher temperature (35°C), then a greater amount of oligomer containing 6-thioguanine (the peak eluting at 19 min) will be produced. The chromatographic conditions are as described in section 4.2.8.

4.3.4.v. Oligomers Containing Guanine Itself

Treatment of the CPG-$G^{S\emptyset}$ oligomer with 0.5 M NaOH at 25°C for 2 days gave an oligomer containing guanine. This parent oligomer was purified by the usual method. Compared with conventional methods, the present method has two advantages: i) the parent oligomer, as a reference sample, can be obtained during the same synthesis as the modified oligomer; ii) NMR sensitive ^{17}O could be easily introduced at last step from ^{17}O-water.

4.3.5 Purification of Oligonucleotides

Oligonucleotides containing thioguanine, 2, 6-diaminopurine, 2-amino-6-methylaminopurine or guanine were first purified with Nensorb cartridge as described in section 2.2.5. O^6-methylguanine containing oligomers were first passed through a Dowex ion exchange column to remove the DBU and then purified by Nensorb cartridge. Oligomers containing 2, 6-diaminopurine were passed through the cartridge without pre-extraction of the oxime/TMG mixture. This mixture was removed as monitored visually by the migration of the orange coloured mixture anions, together with the failure sequences by 20 ml 10% acetonitrile solution (10 ml is normally used for normal oligomers). The purity of the oligomers after Nensorb column purification were good enough for most biological applications. (Figure 4.10a-e). Highly pure oligomers were obtained by further purification with FPLC (Figure 4.10f-j).

Under the basic condition (pH 12) of the anion exchange chromatography, the oligomers containing O^6-methylguanine, 2-amino-6-methylaminopurine or 2, 6-diaminopurine can be well separated from the parent and the G^S oligomers (Figure 4.10). At pH 12, guanine and thioguanine resides would be negatively charged at the N-1 position but as O^6-methylguanine, 2-amino-6-methylaminopurine or 2, 6-diaminopurine do not have this imino proton, the total number of negative charges on the oligomer containing these modified bases is one less than that of their corresponding parent oligomer and the G^S oligomer. Therefore they have shorter

retention time than the parent and G^S oligomers.

The behaviour of the G^S oligomer in anion exchange FPLC at pH=12 is very interesting. Although both guanine base and its thio-analogue have an imino proton which ionizes under the condition (pH=12) of FPLC, the retention time of the thio oligomer was longer than that of its parent oligomer. Figure 4.10 shows that a 12-mer oligomer containing 6-thioguanine can be well separated not only from those containing the modified guanines, which lack an imino proton on the modified bases, but also from its parent oligomer, which has the same negative charges as the thio oligomer (Xu and Swann, 1992). The different retention time between the parent oligomer and its thio counterpart is probably caused by two effects: the thio-base is more acidic than is the oxo analogue, for example the pKa of N-1 imino proton of thiopurine is 7.37 while that of hypoxanthine is 8.94 (Maunter, 1956); and 2'-deoxy-6-thioguanosine is more lipophilic than 2-deoxyguanosine. This increased lipophilicity was observed on TLC in the experiments for the thiolation of deoxyguanosine derivatives, where the thio-analogues have higher mobility than their deoxyguanosine derivatives.

4.3.6. Base Composition Analysis of Oligonucleotides

All of the analogue containing oligonucleotides gave the expected compositions (Table 4.1). The eluate was generally monitored at 260 nm, but for the G^S oligomer a separate HPLC run was monitored at 335 nm to detect the 2'-deoxy-6-thioguanosine as shown in Figure 4.11.

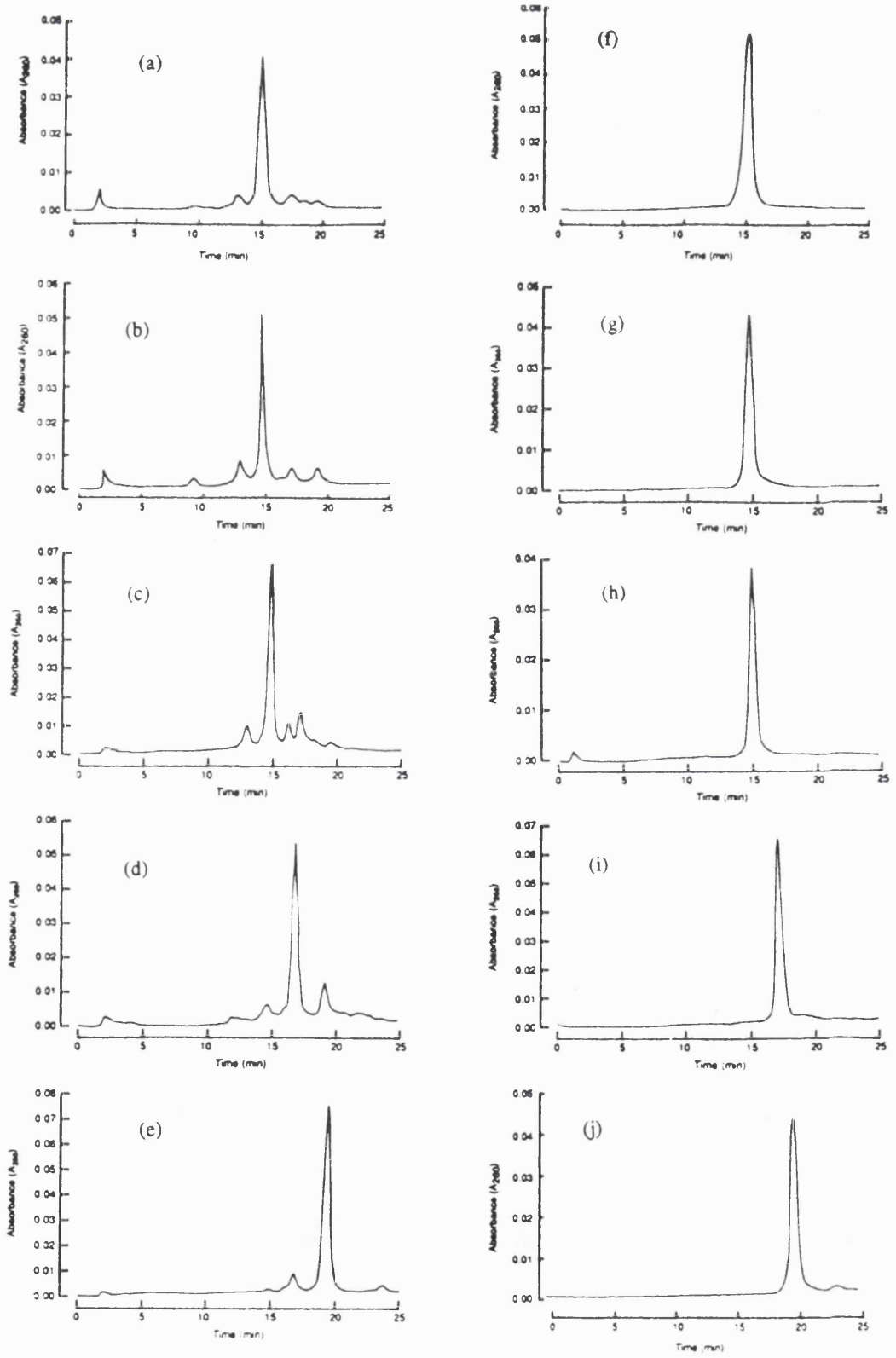


Figure 4.10. FPLC profiles of Nensorb column purified (a-e) and FPLC purified (f-j) 12 mer (AGCYAATTCGCT) prepared by post-synthetic substitution. (a) and (f): Y = G^{NH2}, (b) and (g): Y = G^{NHMe}, (c) and (h): Y = G^{OMe}, (d) and (i): Y = G, (e) and (j): Y = G^{S}. The chromatographic conditions are as described in section 4.2.8.

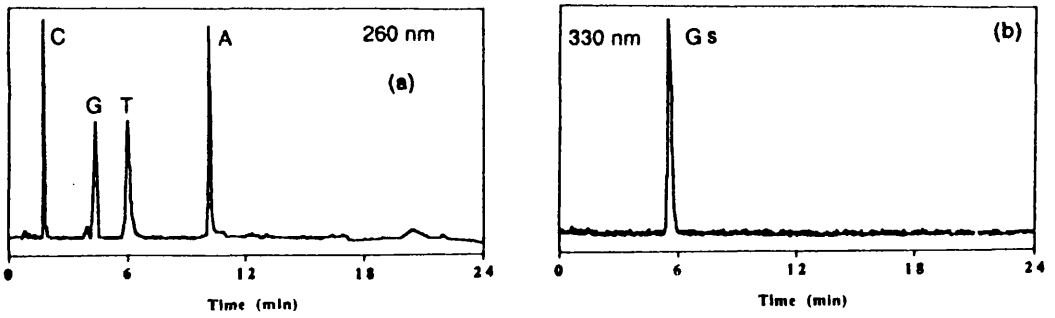


Figure 4.11. HPLC profile of the base analysis of FPLC purified 12 mer (AGCG^{S}AATTCGCT) measured at 260 nm (a) and at 335 nm (b). The chromatographic conditions are as described in section 4.2.9.

Table 4.1. Base composition of the FPLC purified oligonucleotides determined as described in section 4.2.9.

Oligonucleotide	Expected Base Composition	Determined Base Composition
AGCG^{S}AATTCGCT	$C_{3.0}$, $G_{2.0}$, $T_{3.0}$, $A_{3.0}$, $G^{S}_{1.0}$	$C_{2.9}$, $G_{2.1}$, $T_{3.0}$, $A_{2.8}$, $G^{S}_{1.1}$
AGCG^{OMe}AATTCGCT	$C_{3.0}$, $G_{2.0}$, $T_{3.0}$, $A_{3.0}$, $G^{OMe}_{1.0}$	$C_{2.8}$, $G_{2.2}$, $T_{3.0}$, $A_{2.9}$, $G^{OMe}_{0.9}$
AGCG^{NH_2}AATTCGCT	$C_{3.0}$, $G_{2.0}$, $T_{3.0}$, $A_{3.0}$, $G^{NH2}_{0.9}$	$C_{3.0}$, $G_{2.1}$, $T_{2.8}$, $A_{2.7}$, $G^{NH2}_{0.8}$
AGCG^{NMe}AATTCGCT	$C_{3.0}$, $G_{2.0}$, $T_{3.0}$, $A_{3.0}$, $G^{NMe}_{1.0}$	$C_{2.9}$, $G_{1.9}$, $T_{2.8}$, $A_{3.0}$, $G^{NMe}_{1.0}$
AGCGAATTCGCT	$C_{3.0}$, $G_{3.0}$, $T_{3.0}$, $A_{3.0}$	$C_{2.9}$, $G_{3.1}$, $T_{3.0}$, $A_{2.8}$

4.3.7. Stability of DNA Duplexes Containing 6-Thioguanine and 4-Thiothymine

The thermal melting profiles of DNA duplexes containing 6-thioguanine were measured and compared with control DNA, and with DNA containing 4-thiothymine. Although all these duplexes showed a distinct transition temperature (Figure 4.12) the presence of 6-thioguanine produced a much greater depression of the melting temperature (Tm) than the presence of 4-thiothymine. The Tm of a self-complementary duplex (AGCG^{S}AATTCGCT) containing two 6-thioguanine: C base pairs was 12.8 °C lower than that of the control DNA (Table 4.2). The Tm of a non

self-complementary duplex (CAGG^{S}AATTCGC) containing one 6-thioguanine: C pair was 6.9 °C lower than that of the analogous control. The presence of 4-thiothymine produced less instability. The Tm of a self-complementary duplex (AGCGAAT^{S}TCGCT) containing two 4-thiothymine: A pairs was only 1 °C lower than the control. This result is consistent with previous results (Connolly and Newman, 1989). However a non self-complementary duplex (CAGGAATT^{S}CGC) containing only one 4-thiothymine: A pair had a Tm 3 °C lower than the control. A very similar duplex destabilisation caused by incorporation of 6-thioguanine into oligomers has also been reported by Waters et al. (1992). The Tm of a self-complementary duplex (GACG^{S}ATATCGTC) was 11°C lower than that of its parent duplex. As the 6-position of guanine forms a Watson-Crick hydrogen bond with the 4-HN_2 group of cytosine this destabilisation is probably partially caused by a weaker 6-thioguanine: cytosine base pair as compared with a guanine: cytosine base pair. This has been confirmed very recently by temperature dependence NMR and UV studies of a self-complementary octamer containing thioguanine (AAACG^{S}TTT) (Christopherson and Broom, 1994) which indicated that the thioguanine: C base pair is indeed weaker than the G: C base pair. More importantly, the presence of the thioguanine in the octamer also affects the overall structure of the entire duplex, especially its neighbouring bases (Christopherson and Broom, 1994). It is the combination of these two effects that contributes the abnormal depression of the melting temperature of G^{S} containing oligomers.

The reason is not clear yet why the incorporation of thiothymine into oligonucleotides has little effect on the stability of DNA duplex. A possible explanation is that, unlike 6-thioguanine, the presence of 4-thiothymine in DNA duplex only weakens the thiothymine: A base pair but has little effect to the entire structure of the DNA duplex.

Although 6-thioguanine and 6-mercaptopurine, which is converted to 6-thioguanine *in vivo*, are widely used in cancer therapy, their mode of action at the molecular level is still not very clear. The cytotoxicity occurs at cell division and it

seems possible that the distortion of the DNA structure induced by the presence of G^S, which can be seen here as a reduction in Tm, plays a role in the cytotoxicity of these drugs.

Table 4.2. The melting temperature (Tm) values of DNA duplexes containing 6-thioguanine and 4-thiothymine and their differences from the parent ones

Self-complementary Duplexes	Tm	ΔTm	Non-self-complementary Duplexes	Tm	ΔTm
5' AGC GAA TTC GCT 3' TCG CTT AAG CGA	64.5	0	5' CAG GAA TTC GC 3' GTC CTT AAG CG	58.7	0
5' AGC G^SAA TTC GCT 3' TCG C TT AAG ^{S}CGA	51.2	12.8	5' CAG G^SAA TTC GC 3' GTC C TT AAG CG	51.8	6.9
5' AGC GAA T^{S}TC GCT 3' TCG CTT^S A AG CGA	63.5	1	5' CAG GAA TT ^{S}C GC 3' GTC CTT AA G CG	55.7	3

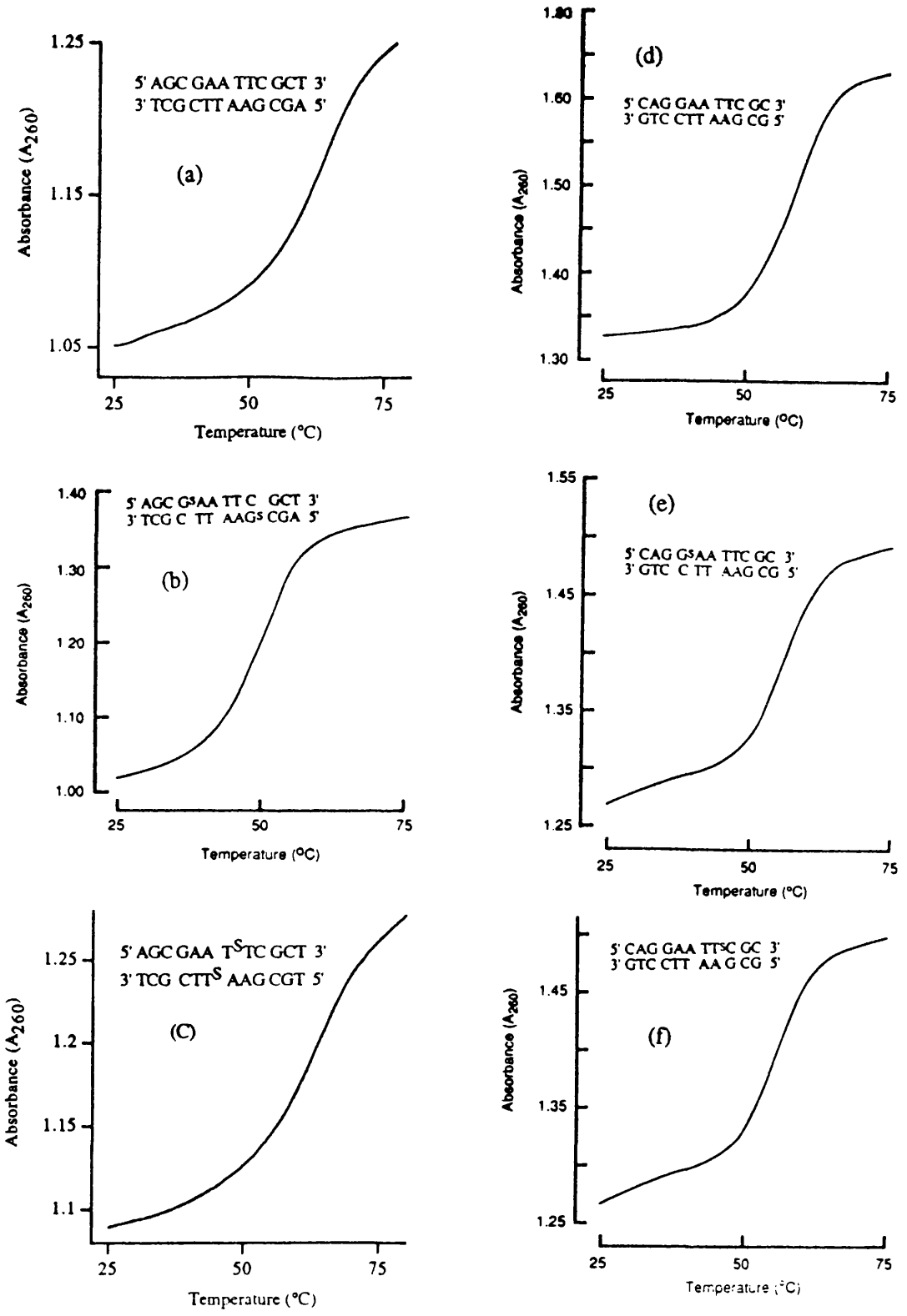


Figure 4.12. Tm curves of DNA duplexes containing 6-thioguanine and 4-thiothymine and their parent duplexes.

CHAPTER 5

STUDIES ON THE INTERACTIONS BETWEEN λ-PHAGE CRO REPRESSOR AND ITS OPERATOR DNA BY PHOTOCHEMICAL CROSS-LINKING

5.1. INTRODUCTION

DNA-binding proteins play a central role in many biological processes, and great efforts have been made to understand how these proteins interact with their DNA (for reviews see: Pabo, 1984; Harrison and Aggarwal, 1990; Aiken and Gumport, 1991; Pabo, 1992). The purpose of the experiments described in this chapter was to study the interactions between the Cro repressor of λ-phage (referred to as Cro or Cro repressor or Cro protein) and its operator DNA by photochemical cross-linking using oligonucleotides containing photoactive thiobases.

Cro is a sequence-specific DNA binding protein which, together with λ repressor, controls the life style of bacteriophage λ by competitively binding to six specific operators (OR1, OR2, OR3, OL1, OL2, and OL3) (Ptashne, 1986). The sequence of the six operators (Table 5.1) are similar and partly palindromic, especially at their ends. The palindromic nature of these sequences is important because both Cro and repressor are dimers, and the palindromic parts of these sequences provide them with two almost identical recognition sites one to each subunit of the dimer.

Cro repressor is a small protein. It has only 66 amino acid residues with a monomer molecular weight of 7,351 and has methionine as its amino and alanine as its carboxyl termini (Hsiang *et al.*, 1977) (Figure 5.1). Its crystal structure was solved in 1981 at 2.8 Å resolution (Anderson *et al.*, 1981) (Figure 5.2a). It consists of three strands of antiparallel β-sheet (residues 2-6, 39-45, and 48-55) and three α-helices (residues 7-14, 15-23, and 27-36). The α–helices are not packed against each other in

the usual way, i.e. in an antiparallel fashion with a short connecting loop, but, instead, the α_2 and α_3 helices form a unique arrangement joined by a loop region. This constitutes the helix-turn-helix DNA-binding motif which plays a central role in the binding of Cro repressor with its operator DNA, and also exists in many other bacterial and bacteriophage regulatory proteins, such as λ–repressor, phage 434 repressor, and catabolite gene activator protein (CAP) (Harrison and Aggarwal, 1990).

Cro forms a stable dimer in solution (Figure 5.2b). The two subunits are held together mainly by interactions between β_3 strands from each subunit. These two strands are arranged in an antiparallel fashion and are hold together by hydrogen bonds and hydrophobic interactions so that the three-stranded β sheets of the monomers form a six-stranded antiparallel β sheet in the dimer

In 1981 the sequence of λ operators was known; and the crystal structure of Cro was also known but this structure was the protein alone and therefore the interaction with the DNA was predicted from molecular modelling rather than being studied directly. Thus using a computer graphics facility with energy minimization, Matthews and colleagues (Anderson *et al.*, 1981; Ohlendorf *et al.*, 1982) proposed the first model of Cro interacting with its operator DNA. In the Cro dimer the α_3 helix in the helix-turn-helix motif of each subunit is at the ends of the elongated dimeric molecule (Figure 5.2b) and is separated by a distance of 34 Å (Anderson *et al.*, 1981), the same distance as the separation between neighbouring major grooves of a B-form DNA double helix, thus if the α_3 helix of one subunit binds into the major groove of B-DNA, the α_3 helix of the other subunit can bind into the major groove one turn further along the DNA molecule. By model building on a graphic display, Matthews and co-workers (Ohlendorf *et al.*, 1982) demonstrated that the two α_3 helices of the Cro dimer could fit very well into the major groove of DNA (Figure 5.3). In this

structure, the α_3 helices provide many of the sequence-specific interactions via hydrogen bonds from the amino acid side chains to the exposed parts of the base-pairs. The hydrogen bonds and van der Waals interactions between the Cro monomer and the operators were also defined in the structure (Ohlendorf *et al.*, 1982).

Since the structure of Cro was first reported (Anderson *et al.*, 1981) and the model for its interaction with DNA proposed (Ohlendorf *et al.*, 1982) , this protein has been the subject of intensive investigation. (e. g. refs; Hochschild *et al.*, 1986; Hochschild and Ptashne, 1986; Benson and Youderian, 1989; Metzler and Lu, 1989; Takeda *et al.*, 1989; Brennan *et al.*, 1990; Pendergrast *et al.*, 1992; Takeda *et al.*, 1992; Chen and Ebright, 1993). These studies have supported the essential features of the model (Ohlendorf *et al.*, 1982). In particular, the α_3 recognition helix of the protein fits into the major groove of the DNA and recognizes a sequence-specific site by direct interactions with the edges of the base pairs (Brennan et al., 1990). However some of proposed interactions between the amino acids of the protein and base pairs within the major groove of the DNA have been in dispute. For example, in the Ohlendorf's model, thymine -2 is thought to make no contacts with Cro, but Takeda et al. (Takeda *et al.*, 1989) suggested that O-4 of thymine-2 forms a hydrogen bond with Tyr-26. Also Ohlendorf et al. proposed that Ser-28 contacts base pair T•A (3) and Lys-32 contacts base pairs G•C (4) and T•A (5). However, based on a set of genetic and chemical modification data obtained from the binding studies *in vivo* of wild and mutant Cro protein to various operator sequences, Ptashne and coworkers (Hochschild *et al.*, 1986; Hochschild and Ptashne, 1986) argued that Ser-28 contacts base pair G•C (4) and Lys-32 contacts base pair T•A (5) and G•C (6); on the other hand Takeda et al. (Takeda *et al.*, 1989) and Benson and Youderian (1989) supported the proposed interactions in the Ohlendorf's model by *in vitro* and *in vivo* binding data. Although the crystal structure of a Cro-DNA complex has been determined recently (Brennan *et al.*, 1990), it does not solve these arguments due to the low resolution of the structure.

Photo-cross-linking of a DNA-protein complex offers a means of probing the

interaction interface between the two molecules since irradiation of DNA-protein complex with UV light produces covalent linkages between amino acid residues and nucleic acid bases and cross-linking can only take place on those amino acid residues that are at the interface of the protein-DNA complex. This technique has been widely used to study DNA-protein interactions (for reviews see: Williams and Konigsberg, 1991; Hockensmith *et al.*, 1991).

However the usual procedure of cross-linking of protein to normal DNA by short-wave (254 nm) UV light has many disadvantages, notably, damage to both DNA and proteins by far UV light and the very low yield of cross-linking. Therefore, one would prefer to use a DNA specifically containing a photoactive base analogue absorbing at longer wavelengths to minimize photodamage to DNA and proteins and to increase cross-linking efficiency. More importantly, analysis of the cross-linked complex would be facilitated since the cross-link would involve only one base in a known position in the DNA sequence. Ideally, a photoactive base incorporated into DNA should provide minimal structural perturbation and exhibit high photoreactivity toward amino acids. Two bases, in particular, 5-bromouracil (Wolfes *et al.*, 1986; Allen *et al.*, 1991; Wick and Matthews, 1991; Blatter *et al.*, 1992; Hicke *et al.*, 1994) and 4-thiouracil (Bartholomew *et al.*, 1990; Favre, 1990; Rinke-Appel *et al.*, 1993), have previously been used for this purpose.

Nikiforov and Connolly (1992) cross-linked both the Eco RV restriction endonuclease and Eco RV methylase to their DNA binding site using oligonucleotides containing either 4-thiothymine or 6-thioguanine at specific positions. 4-Thiothymine and 6-thioguanine possess several desirable properties for studying DNA-protein interactions. The sulphur atom in the major groove of the DNA is only slightly larger than oxygen, but it otherwise chemically resembles oxygen. Like oxygen it has two unpaired electron-pairs. One of these can base pair with cytosine, while the other is available to interact with proteins in the major groove. Therefore, the introduction of these thiobases into oligonucleotides should not appreciably perturb the binding interaction between the proteins and DNA. Furthermore, 4-thiothymidine and

6-thiodeoxyguanosine have an λmax at 340 nm, and are photoactive at 340-350 nm wavelength, which is well away from the usual absorption maxima of protein (280 nm) and DNA (260 nm), thus cross-linking can be carried out at a wavelength which would not appreciably damage proteins and DNA. Upon irradiation with near UV light (340-360 nm) they become chemically reactive through an unknown mechanism, and are capable of covalently linking to a bound protein (Nikiforov and Connolly, 1992).

Our development of efficient methods to incorporate modified bases, including 4-thiothymine and 6-thioguanine, into oligonucleotides offered the probability of using photochemical cross-linking to study Cro-DNA interactions. The results presented in this chapter show that oligonucleotides containing these thiobases can indeed cross-link to Cro protein upon irradiation with long-wave UV light (~360 nm), and the cross-linking results are consistent with the Cro-DNA interaction model proposed by Ohlendorf et al. (1982) as revised later by Takeda et al. (1989) and further revised by Brennan et al. (1990). The results also support the suggestion that cross-linking can take place only on those amino acid residues at the interface of DNA-protein complexes. However our attempt to identify the amino acids covalently linked to DNA was unsuccessful, mainly due to the instability of cross-linked Cro-DNA complex.

Table 5.1. The six operators found in the λ phage DNA

OR1	T	A	C	C	T	C	T	G	G	C	G	G	T	G	A	T	A
	A	T	G	G	A	G	A	C	C	G	C	C	A	C	T	A	T
OR2	T	A	A	C	A	C	C	G	T	G	C	G	T	G	T	T	G
	A	T	T	G	T	G	T	C	A	C	G	C	A	C	A	A	C
OR3	T	A	T	C	A	C	C	G	C	A	A	G	G	G	A	T	A
	A	T	A	G	T	G	G	C	G	T	T	C	C	C	T	A	T
OL1	T	A	T	C	A	C	C	G	C	C	A	G	T	G	G	T	A
	A	T	A	G	T	G	G	C	G	G	T	C	A	C	C	A	T
OL2	C	A	A	C	A	C	C	G	G	C	A	G	A	G	A	T	A
	G	T	T	G	T	G	G	C	C	G	T	C	T	C	T	A	T
OL3	T	A	T	C	A	C	C	G	C	A	G	A	T	G	G	T	T
	A	T	A	G	T	G	G	C	G	T	C	T	A	C	C	A	A

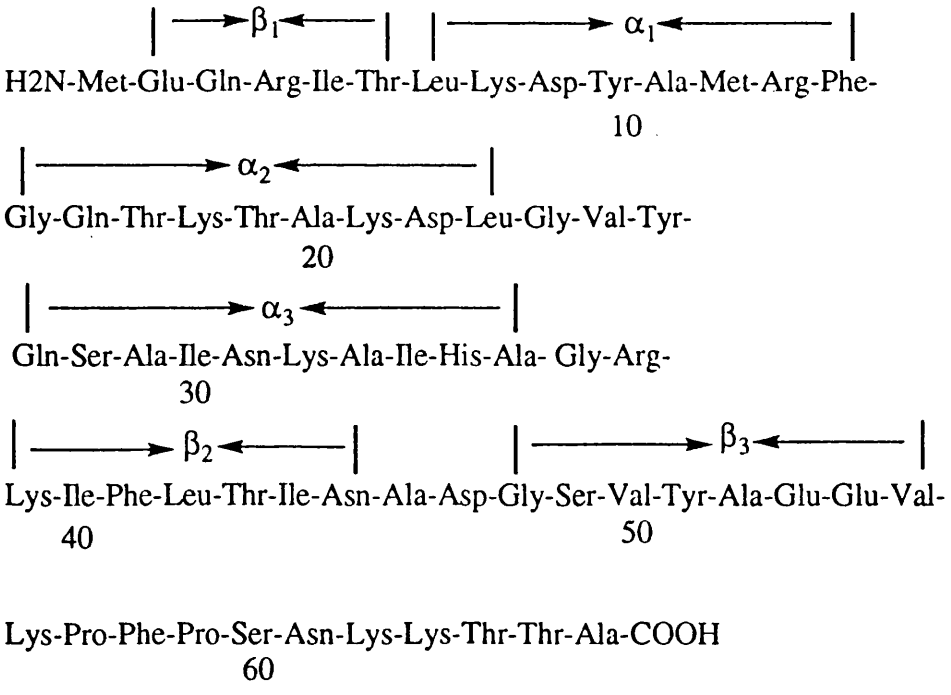


Figure 5.1. Amino acid sequence of Cro protein. α_1, α_2, and α_3 indicate three α-helical region (residue 7-14, 15-23, and 27-36), and β_1, β_2, and β_3 indicate three β–sheet region (residues 2-6, 39-45, and 48-55)

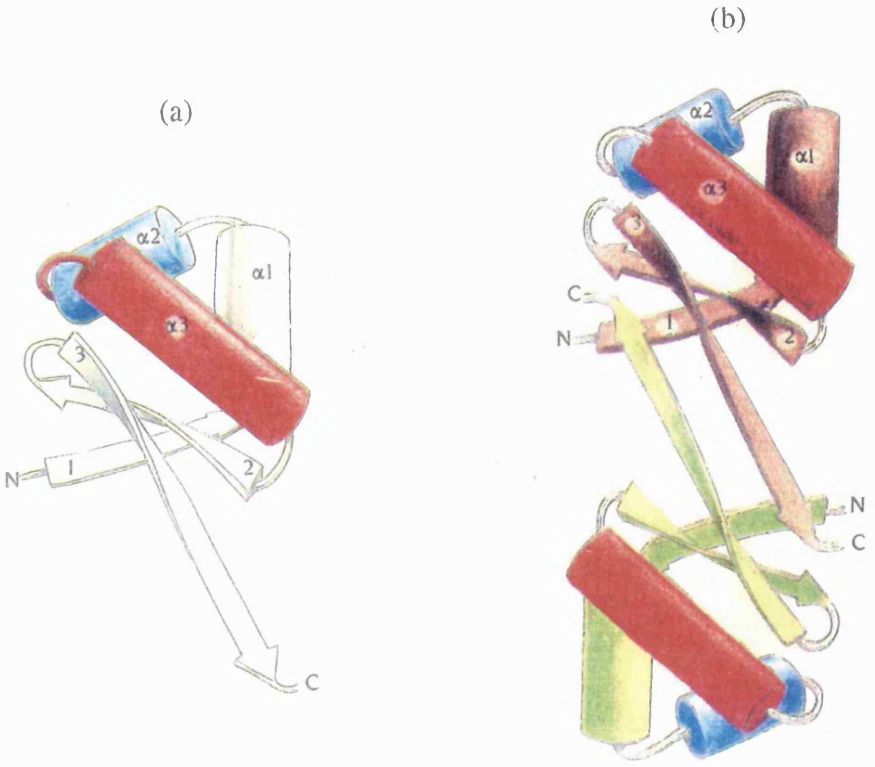


Figure 5.2. Three dimensional structure of Cro repressor. (a) Structure of Cro monomer. α_2 and α_3 helices that form the helix-turn-helix motif are coloured blue and red respectively. (b) Structure of Cro dimer. One subunit is coloured green and the other is coloured brown. The main dimer interactions are between β_3 from each subunit. α helices 2 and 3, the helix-turn-helix motif, are blue and red, respectively, in both subunits (Taken from Branden and Tooze, 1991).

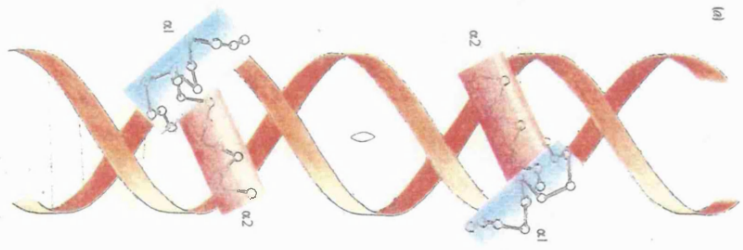


Figure 5.3. A schematic diagram of the interaction of Cro with DNA showing that a pair of symmetry-related helices of the Cro dimer fit nearly into neighbouring major grooves of DNA (Taken from Branden and Tooze, 1991).

5.2. MATERIALS AND METHODS

5.2.1. Materials and Enzymes

E. coli K12 W 3110 (pcro1), which contains a plasmid DNA pcro1 containing cro gene under control of the lac operator and which will overexpress Cro when induced with IPTG, was a kind gift from Dr Takeda (USA). Lysozyme and DNase were obtained from Sigma. Bacto-trypton was from Difco Laboratories, and Bacto-yeast extract was from Beta Lab. Chymotrypsin was from Sigma, and modified trypsin from Promega. All other chemicals and solvents were obtained as described in sections 2.2.1 and 3.2.1, and were used directly without further purification.

5.2.2. High Performance Liquid Chromatography

Reverse phase HPLC and fast protein liquid chromatography (FPLC) were carried out essentially as described in section 2.2.1.

5.2.3. Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis was carried out using either Mighty small II or SE-600 slab gel electrophoresis unit (Hoefer Scientific Instruments).

5.2.3.i. Non-denaturing Polyacrylamide Gel Electrophoresis for DNA-protein Binding Assay (Gel Retardation Assay)

This was carried out essentially as described by Sambrook et al. (1989).

A 8% acrylamide gel solution in 1 x TBE buffer (89 mM tris-borate, 2 mM EDTA) was degassed under water pump vacuum for 5 min, and polymerization was initiated by adding the appropriate amount of 10% ammonium persulphate (200 µl/100 ml gel solution) and N, N, N', N'-tetramethylenediamine (TEMED, 100µl/100 ml gel solution). The gel was polymerized for 60 min at room temperature.

After the gel had been run with 1 x TBE buffer for 30 min, the samples were

loaded and electrophoresis was carried out at 50 V (constant voltage) for the Mighty small II gel electrophoresis unit [gel size: 6 cm (h) x 8 cm (w)] or 100 V for the SE-600 gel electrophoresis unit [gel size: 13.5 cm (h) x 14 cm (w)]. After the run, the gel was transferred onto a piece of 3M paper, dried at 75°C under vacuum and autoradiographed at -70°C. The appropriate gel bands were cut into sections, and their radioactivity was counted in 3 ml of scintillation fluid using Packard Tri-Carb 460 CD liquid scintillation counter.

5.2.3.ii. SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The SDS polyacrylamide gel electrophoresis was carried out with discontinuous buffer system as described in a Hoefer's technical bulletin: "The electrophoretic separation of low molecular weight polypeptides in polyacrylamide gels". Working solutions were as follows.

Separating gel solution:

20% T 0.5% C, 10% glycerol (v/v), 0.1% SDS (w/v), 0.75 M Tris (pH 9.3).

Stacking gel solution:

10% T 4.8% C, 10% glycerol (v/v), 0.1% SDS (w/v), 0.125 M Tris (pH 6.8).

Tank buffer:

0.38 M glycine, 0.05 M tris-HCl, 0.1% SDS (w/v), pH 8.4-8.6.

Sample treatment buffer:

0.0625 M Tris (pH 6.8), 1% SDS (w/v), 15% glycerol (v/v),

15 mM Dithiothreitol (DTT), 0.1% bromophenol blue.

Running gel overlay:

0.75 M Tris (pH 9.3), 0.1% SDS (w/v).

Staining solution:

0.025% Coomassie Blue R-250, 40% CH_3COOH (v/v), 7% CH_3OH.

Destaining solution:

7% CH_3COOH (v/v), 5% CH_3OH.

Gel preparation

A separating gel solution was degassed under water pump vacuum, and the polymerization was initiated by adding appropriate amount of freshly prepared 10% ammonium persulphate and TEMED (see section 5.2.3.i). The gel solution was gently poured into the sandwich and covered with running gel overlay solution. The gel was polymerized overnight.

After removing the overlay from the separating gel, the stacking gel was poured and allowed to polymerize for at least 1.5 h.

Loading and running gels

Samples were mixed with an equal volume of sample treatment buffer and loaded onto the wells. Electrophoresis was carried out at 15 mA (constant current) until the dye front had entered the separating gel and was then increased to 25 mA when the SE-600 gel electrophoresis unit was used (gel size: 13.5 cm x 14 cm x 1mm). Half of these currents were applied to the Mighty small II gel electrophoresis unit (gel size: 6 cm x 8 cm x 1mm) The gels were then stained with Coomassie Blue or silver as described below, or dried and autoradiographed as described in section 5.2.3.i.

Coomassie Blue staining of the gels

After the run, the gel was put into the staining solution and gently shaken overnight. Destaining was carried out at room temperature until the required intensity of colour of the bands was reached. The gel was then transferred to 7% CH_3COOH for storage. The stained gel was placed on a light box and photographed with a Polaroid instant sheet film.

Silver staining of the gels

This was carried out according to maker's instructions using the Silver Stain Plus kit (Bio-Rad).

5.2.4. Preparation and Characterization of Oligonucleotides

5.2.4.i. Synthesis and Purification of Oligonucleotides

Oligonucleotides containing 4-thiothymine or 6-thioguanine were synthesized on a DNA synthesizer as described in the previous chapters. After treatment with appropriate reagents, the oligomers were separated from the failure sequences using a Nensorb cartridge and further purified by FPLC as described (section 2.2.5) except the column was eluted with buffer C (0.4 M NaCl, 0.01 M NaOH aqueous solution) for 2 min, then with a linear gradient from 100% buffer C to 80% buffer C and 20% buffer E (1.2 M NaCl, 0.01 M NaOH aqueous solution) over 3 min, and then to 50% of buffer C and 50% of buffer E over the following 20 min for the oligomer having the sequence of the top strand of the OR3 operator (see Figure 5.4); or with buffer C for 2 min, then with a linear gradient from 100% buffer C to 75% buffer C and 25% buffer E over 3 min, and then to 45 % of buffer C and 55% of buffer E over the following 20 min for the oligomers having the sequence of the bottom strand (see Figure 5.4). The desired peak was collected and neutralized with 1 M acetic acid, and desalted with Waters Sep-pak C_{18} cartridge as described (section 2.2.5). The purity of the oligomers was checked by re-injecting the desalted samples into FPLC.

5.2.4.ii. Base Analysis and Melting Temperature (Tm) Measurement of Oligomers

Base composition analysis of the oligonucleotides was carried out as described in sections 2.2.6 and 4.2.9. The Tm of the oligonucleotides was measured as described in section 2.2.8 except in 10 mM Tris, pH 7.3, 10 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl.

5.2.4.iii. Circular Dichroism (CD) Spectroscopy of Oligonucleotides

The measurement of CD spectra of the oligonucleotides was carried out by Dr Drake's team (Chemistry, UCL) in the UV region in 10 mM K-phosphate, pH 7.3, 10 mM $MgCl_2$, 100 mM KCl at room temperature.

5.2.5. Preparation of Double Stranded Oligonucleotides

5.2.5.i. 5'-^{32}P-labeling of Oligonucleotides

Oligonucleotides (50 pmol) were 5'-^{32}P-labeled using [γ-^{32}P]-ATP (50 μCi) and T4 polynucleotide kinase (20 units) at 37°C in 70 mM Tris (pH 7.6), 10 mM $MgCl_2$, and 1 mM spermidine in a volume of 20 μl. After 30 min incubation, excess ATP was removed by passing the mixture through a Biogel P-6 column (0.5 cm x 6 cm). The eluate was collected in 3 drop fractions. The labeled oligonucleotide was located by counting the radioactivity of each fraction (by counting 2 μl of eluate from each fraction). The fractions containing labeled oligomers were pooled and dried down in a Speed-Vac evaporator under vacuum.

5.2.5.ii. Strand Hybridization.

The single-stranded, ^{32}-P labelled 21-mers containing modified bases were mixed with their non-labelled complementary strands in 10 mM $MgCl_2$, 0.2 mM EDTA, and 100 mM KCl. Samples were heated at 50°C for 2 min and cooled down to room temperature over a period of 2 h. In order to ensure that the ^{32}-P labeled strand would quantitatively form double stranded DNA, 20% of excess of complementary strand was added.

5.2.6. λ Cro Preparation.

The preparation of Cro protein was based on the protocol provided by Dr Takeda in USA. All experiments were carried out in the cold room (4°C) except the cell growth. Cells were grown at 30°C. The buffers were as follows.

Buffer F: 10% sucrose, 50 mM Tris, pH 7.4, 0.1 mM EDTA, 50 mM KCl.

Buffer G: 10 mM K-phosphate, pH 6.4, 0.1 mM EDTA, 5% glycerol.

Buffer H: 0.1 M K-phosphate, pH 7.0, 0.1 mM EDTA, 5% glycerol.

Buffer I: 0.1 M K-phosphate, pH 7.0, 5% glycerol.

Buffer J: 10 mM Tris, pH 7.4, 0.1 mM EDTA, 0.2 M KCl, 5% glycerol.

5.2.6.i. Cell Growth.

Luria-Bertani (LB) medium:

LB medium (pH 7.0) (Sambrook *et al.*, 1989) containing 1% Bacto-trypton, 0.5% Bacto-yeast extract, and 1% NaCl was autoclaved at 120°C for 20 min. After the medium had cooled to room temperature, ampicillin solution (50 mg/ml) was added to give a final concentration of 5µg per ml.

Preparation of an inoculum for large scale cell growth.

E. coli strains K12 W 3110 transformed with plasmid DNA pcro1 (lyophilized powder) was first grown in 10 ml LB medium at 30°C for 48 hours. The fresh bacteria was then inoculated onto a LB medium agar plate and incubated overnight. A single colony from the plate was transferred into two glass universal bottles each containing 10 ml LB medium and incubated at 30°C overnight. Sterile glycerol was added to the culture to give the final concentration of 20% (v/v), and was kept in a freezer at -70°C and used as the seeds for large scale cell growth.

To check that this culture contained the correct plasmid and would express Cro, LB medium (100 ml) was inoculated with 2 ml of the culture obtained above and incubated at 30°C. The cells were induced with IPTG and harvested as described below. The cell-free extract was prepared as described in section 5.2.6.ii and examined in SDS gel as described in section 5.2.3.ii.

Large scale cell growth.

The cell seed (2 ml) was inoculated into 100 ml of LB medium in a 500 ml conical flask, and incubated overnight at 30°C. The fresh culture was then transferred

evenly into five 3 litre conical flasks each containing 500 ml of LB medium. The inoculated medium was incubated at 30°C, and the cell density determined by measuring the absorbance at 600 nm. When the OD_{600} reached 1.0, isopropylthiogalactoside (IPTG) was added (final concentration of 1 g/litre) and incubation continued for another 3 h.

Bacterial cells were harvested by centrifugation in a Sorvall RC-5B centrifuge at 4°C and frozen immediately at -70°C.

5.2.6.ii. Cell Lysis and Crude Extract Preparation

Frozen cells (10 g) were homogenized at low speed in 120 ml of buffer F. Lysozyme (15 mg) was added and incubated for 30 min on ice-bath with occasional stirring. The temperature was raised to 30°C to complete lysis. The solution was cooled on ice and 2 mg DNase and 0.3 ml 1 M $MgCl_2$ added. When the viscosity of the solution had decreased KCl (0.4 g) was added. The extract was then centrifuged (50k rpm, 2 h; Ti 70.1 rotor). The supernatant was diluted to an ionic strength equivalent to 0.2 M KCl with buffer G by measuring its conductivity.

5.2.6.iii. Phosphocellulose Column Chromatography of Cro Repressor

Dry phosphocellulose (20 g) was suspended in 0.5 N NaOH (500 ml) and left for 5 min. The supernatant was decanted off and the phosphocellulose was washed with water until the pH was below 11. The phosphocellulose was resuspended in 0.5 N HCl (500 ml) and left for 5 min. The acid was removed by filtration and the solid was washed with water until the pH was above 3. The solid was resuspended in buffer G containing 0.2 M KCl. The suspension was degassed under water pump vacuum and poured into a glass column (1.6 cm x 20 cm).

After the phosphocellulose column had been equilibrated with buffer G containing 0.2 M KCl, the crude extract obtained above was loaded onto the column and washed through with 100 ml of buffer G containing 0.2 M KCl to remove proteins that did not bind to the column. The column was then washed with a linear 0.2-1.0 M

KCl gradient in buffer G (100 ml each) to elute out the proteins bound to the column (including Cro protein). The gradient was run at a rate of 30 ml/h by pumping the buffer with a peristaltic pump, and the eluate was collected in 3 ml fractions using a fraction collector and the A_{280} of each fraction measured. Fractions in the main peak were pooled and gently stirred on an ice bath while ammonium sulphate was added to 60% saturation (39 g per 100 ml) to precipitate the protein. After being stirred for 60 min on ice-bath, the solution was centrifuged at 12,000 rpm for 40 min in Sorvall RC-5B centrifuge. The precipitate was isolated by carefully decanting the supernatant.

5.2.6.iv. Gel Filtration Column Purification of Cro Protein

Sephadex G-75 (30 g) was hydrated in H_2O for 60 min and then washed with HPLC water several times, and finally resuspended in HPLC water to give approximately 300 ml swollen gel. The gel was degassed using a water pump and poured into a glass column (1.5 cm x 100 cm), letting the liquid run out of the bottom under gravity and continuously topping up the column with gel until the desired height of gel was reached.

After equilibrating the column with buffer H, the protein precipitate from the phosphocellulose column was redissolved in 4 ml of buffer H and loaded onto the column. The column was eluted with buffer H (30 ml/h) by pumping the buffer with a peristaltic pump, and the eluate was collected in 10 ml fractions. A_{280} of each fraction was measured. Fractions containing the Cro repressor protein were identified on SDS gels then pooled and subjected to hydroxyapatite column purification.

5.2.6.v. Hydroxyapatite Column Chromatography of λ Cro Protein

Hydroxyapatite (Bio Gel HT, 45 ml) was washed with buffer I (0.1 M K-phosphate, pH 7.0, 5% glycerol) (2 x 50 ml) and resuspended in the same buffer to give a the final volume of 60 ml. After being degassed under water pump vacuum, the suspension was poured into a plastic column (2.5 cm x 6 cm).

After the column had been pre-equilibrated with 100 ml of buffer I, the fractions containing the Cro repressor protein from the gel filtration column purification were loaded directly onto the column. The column was eluted with a linear 0.1 M-1.0 M K-phosphate gradient in buffer I at a flow rate of 20 ml per hour by pumping the buffer into the column with a peristaltic pump. The eluate was collected in 4 ml fractions, and A_{280} measured. Fractions containing Cro protein were identified by SDS gel electrophoresis (section 5.2.3). These fractions were pooled and concentrated with ultrafiltration using an Amicon ultrafiltration (Stirred cells, Series 800). The concentrated protein was dialyzed against buffer J and then against same buffer containing 50% glycerol and stored at -20°C.

5.2.6.vi. λ Cro Concentration Determination

The protein concentration was determined using the absorbance of the solution at 277 nm [a 1 mg/ml solution has an $A_{277} = 0.74$ (Takeda *et al.*, 1986)].

5.2.7. Equilibrium Repressor-Operator Binding Assay

0.1 nM ^{32}-P labeled 21-mer duplexes and various concentrations of Cro protein were incubated together on ice for 30 min in 50 µl of binding buffer (10 mM Tris, pH 7.3, 10 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl, 150 µg/ml BSA) then 20 µl of the mixture electrophoresed on non-denaturing polyacrylamide gels (section 5.2.3.i). After autoradiography, the gel bands containing Cro-DNA complex and free DNA were excised and counted. The equilibrium dissociation constants (K_d) were determined as follows:

One step binding of Cro repressor (R) to operator DNA (O) is expressed as:

$$R + O = RO \qquad (1)$$

The equilibrium dissociation constant is written as:

$$K_d = [R]\,[O] / [RO] = [R]\,([O]_0 - [RO]) / [RO] \qquad (2)$$

Where $[O]_0$ is the initial concentration of (O).

Thus $[R][O]_0 - [R][RO] = K_d[RO]$ (3)

i.e. $[O]_0 = \{[RO][R] + K_d[RO]\} / [R] = [RO] + K_d[RO] / [R]$ (4)

From equation (4) we obtain

$[RO] = [O]_0 - Kd[RO] / [R]$ (5)

The graph of [RO] vs [RO] /[R] should be linear with a slope = $-K_d$.

5.2.8. Photo-cross-linking of Cro to Oligonucleotides Containing Modified Bases

^{32}-P labeled 21 bp operator DNA (0.5 μM) containing either 4-thiothymine or 6-thioguanine at a single site in either strand was incubated with Cro protein (1.5 μM) in the binding buffer (30 min, 0°C). The samples were transferred into glass capillary tubes (ID=1.4 mm, OD= 1.8 mm) and irradiated with a long wavelength (360 nm) UV lamp (Model UVL-56, Ultra-Violet Products USA) at a distance of 1.5 cm at room temperature. Samples were removed at different time intervals and stored in the dark for analysis by SDS polyacrylamide gel electrophoresis as described in section 5.2.3. Crosslinked DNA-protein complexes were identified by autoradiography of dried gels and quantified by scintillation counting of excised gel slices.

5.2.9. Isolation of Cross-linked Complexes

Two methods, FPLC and preparative SDS gel electrophoresis, were used for isolating cross-linked complexes from free DNA and protein. FPLC separation was done using a Pharmacia MonoQ HR 5-5 column eluted (0.5 ml/min) with buffer K (0.4 M NaCl, 20 mM Tris, pH 7.4) for two min, then with a linear gradient from 100% of buffer K to 95% of buffer K and 5% of buffer L (1.2 NaCl, 20 mM Tris, pH 7.4) over the next 3 min, and then to 65% of buffer K and 35% of buffer L over the following 40 min. The eluate was collected in 0.5 ml fractions, and the radioactivity of 5 μl from

each fraction measured with a liquid scintillation counter. The fractions containing cross-linked complex were located by analysis of the radiolabelled fractions in SDS polyacrylamide gel electrophoresis. These fractions were pooled and desalted with NAP-25 column. The volume of the desalted sample was reduced by lyophilization.

Before isolation of cross-linked species with SDS gel the UV irradiated samples were concentrated using Amicon ultrafiltration (Stirred cells, Series 800). Then 20% SDS, 0.15 M DTT, and glycerol were added to concentrations of 0.5% (SDS), 7.5 mM (DTT), and 5% (glycerol) respectively. This mixture was electrophoresed on a preparative SDS polyacrylamide gel as described in section 5.2.3 to separate cross-linked complex from free DNA and protein. Bands were visualized by autoradiography of the wet gel. The band containing cross-linked complex was excised and cut into small pieces, and the complex was electroeluted into a buffer (25 mM Tris, 192 mM glycine, 0.1% SDS) with an electroelutor (Model 422 Electro-Eluter, Bio-Rad) according to the maker's instructions. The eluate was dialyzed (SpectraPor 6, 3,500 MW cut-off) against water.

5.2.10. Proteolytic Digestion of Cross-linked Complexes

The isolated cross-linked complexes were treated with chymotrypsin or modified trypsin (Promega) which is resistant to autolysis. Concentrated solutions of the enzymes were prepared, and an appropriate amount added to the cross-linked material in 0.1 M NH_4HCO_3, pH 7.8 or 50 mM Tris-HCl, 1 mM $CaCl_2$, pH 7.6. After incubation (8 h, room temperature), a second addition was made from a fresh solution of the enzyme and digestion was continued for 16 h. Proteolysis was stopped by freezing the mixture in liquid nitrogen. The extent of proteolysis was determined by analysis of a small fraction of the sample in a SDS polyacrylamide gel as described in section 5.2.3. The band was visualized by autoradiography of the dried gel.

5.2.11. HPLC Purification of Peptide-DNA Complexes

The peptide-DNA complexes were separated from non-cross-linked species by

ion exchange HPLC as described in section 5.2.9 except the column was eluted (0.5 ml/min) with buffer K (0.4 M NaCl, 20 mM Tris, pH 7.4) for two min, then with a linear gradient from 100% of buffer K to 95% of buffer K and 5% of buffer L (1.2 M NaCl, 20 mM Tris, pH 7.4) over the next 3 min, and then to 55% of buffer K and 45% of buffer L over the following 40 min.. Fractions (0.25 ml) were collected and counted in a scintillation counter to determine the fractions containing radiolabeled species. These fractions were pooled and dialyzed (SpectroPor 6, 3,500 MW cut-off) against water then concentrated by lyophilization.

5.2.12. Peptide Sequence Analysis

This was carried out by Dr Coles (Biochemistry, UCL) by automated Edman degradation, performed on an Applied Biosystems 470A sequencer using manufacturer's methods and protocols.

5.3. RESULTS

5.3.1. Preparation and Characterization of Oligonucleotides

Although λ operators are 17 base-pairs long we synthesized all operators as 21-mers since the Cro appears to interact with DNA over 19 to 20 bp (Ohlendorf *et al.*, 1982; Takeda *et al.*, 1986). Because Cro binds most strongly to the OR3 operator DNA, each synthesized oligonucleotide contains the sequence of the OR3 operator with two extra bp at each end (Figure 5.4a). These 4 bp are those found in bacteriophage λ DNA sequence (Johnson *et al.*, 1980). For reference, a 22 bp DNA sequence was also made containing the cyclic AMP receptor protein (CRP) binding site (Figure 5.4b). This has no 3 successive base-pairs in common with the OR3 operator.

The oligonucleotides were first purified using Nensorb columns to remove failure sequences, then further purified by FPLC. To ensure that the modified bases had been incorporated into the oligonucleotides without alteration base composition analysis of the oligonucleotides was undertaken. Base analysis of two sample oligonucleotides (Figure 5.5) showed that 4-thiothymidine and 6-thiodeoxyguanosine were present without any modification. Since the modified bases in these two oligomers are located near the 3'-end of the sequences and therefore the base analogues were subject to longer exposure to the reagents than those in other sequences, it is reasonably to assume that the thiobases in other sequences were also intact.

Since Cro binds double stranded DNA it is important to check that all of the oligonucleotides formed stable duplex under the DNA-protein binding assay conditions or the conditions for DNA-Cro cross-linking. This was done by measuring the melting temperature of an oligonucleotide in a buffer similar to the binding buffer. The melting curves of two typical operator DNAs containing 4-thiothymine (T_{+3}/B) or 6-thioguanine (T/B_{-4}), together with that of the control DNA, are given in Figure 5.6, and the melting temperature of all the DNA duplexes are given in Table 5.2. All of the Tm's of the oligomer duplexs are greater than 60°C and would therefore be double

stranded at the conditions under which DNA-protein binding assay and photo-cross-linking were carried out.

Cro protein binds B-form DNA (Brennan *et al.*, 1990). Therefore measurement of circular dichroism (CD) spectra of the oligonucleotides was undertaken. A typical right handed B-form DNA structure has a large negative peak at around 250 nm and a smaller positive peak at around 280 nm (Fairall *et al.*, 1989). The x-axis crossover point between the positive and negative peaks is at around 270 nm.

CD spectra of the oligonucleotides were recorded in a similar buffer to the one used for DNA-protein binding assay. As expected, the parent DNA duplex (T/B) produces a spectrum typical of B-form DNA (Figure 5.7), and as the operators containing 4-thiothymine ($T/B_{+3'}$) or 6-thioguanine (T_{+8}/B) gave almost identical spectra these were also B-form duplexes. All of them had positive maxima at around 280 nm and negative maxima at around 250 nm and all have their crossover at around 265 nm. The exceptionally high λmax of 6-thiodeoxyguanosine and 4-thiothymidine at 340 nm showed up as an entirely separate peak in the CD spectrum with a λmax at ~345 nm.

(a)

```
          +1 +2 +3 +4 +5 +6 +7 +8 +9 -8' -7' -6' -5' -4' -3' -2' -1'
(T)  T  C T A T  C A C  C G C A A  G G G  A T  A  A  A
(B)  A  G A T A  G T G  G C G T T  C C C  T A  T  T  T
          -1 -2 -3 -4 -5 -6 -7 -8 -9 +8' +7' +6' +5' +4' +3' +2' +1'
```

(b)

```
T A A T G T G A G T T A G C T C A C T C A T
A T T A C A C T C A A T C G A G T G A G T A
```

Figure 5.4. Sequences of DNA duplex used in this experiment. (a) Sequence of DNA containing OR3 operator. The two strands have been designated top strand (T) and bottom strand (B) and base pairs in the operator have been numbered within each half-site from 1 to 9 starting at the distal base pair (Ohlendorf *et al.*, 1982) as shown. The sequence of OR3 operator is bold. (b) Sequence of CRP binding site.

Table 5.2. Tm values, equilibrium dissociation constants, and cross-linking efficiency of the operator DNAs.

[a]DNA	Tm (°C)	[b]K_d (nM)	[c]X-link (%)	Sequence of strands containing modified bases
T/B	67.9	0.77		
T_{+1}/B (T^S)	67.3	0.86	2.4	5' TC T^S ATC ACC GCA AGG GAT AAA 3'
T_{+3}/B (T^S)	66.6	1.31	2.2	5' TCT AT^{S}C ACC GCA AGG GAT AAA 3'
T_{+8}/B (G^S)	64.3	1.21	nil	5' TCT ATC ACC G^{S}CA AGG GAT AAA 3'
$T_{-6'}$/B (G^S)	64.7	1.43	9.0	5' TCT ATC ACC GCA AG^{S}G GAT AAA 3'
$T_{-4'}$/B (G^S)	63.9	1.58	3.0	5' TCT ATC ACC GCA AGG G^{S}AT AAA 3'
$T_{-2'}$/B (T^S)	66.9	1.11	nil	5' TCT ATC ACC GCA AGG GAT^S AAA 3'
T/B_{-2} (T^S)	67.7	0.98	nil	3' AGA T^{S}AG TGG CGT TCC CTA TTT 5'
T/B_{-4} (G^S)	65.0	1.74	3.4	3' AGA TAG^S TGG CGT TCC CTA TTT 5'
T/B_{-5} (T^S)	66.5	1.05	8.6	3' AGA TAG T^{S}GG CGT TCC CTA TTT 5'
T/B_{-6} (G^S)	64.4	1.62	7.2	3' AGA TAG TG^{S}G CGT TCC CTA TTT 5'
T/B_{-7} (G^S)	63.3	1.58	7.2	3' AGA TAG TGG^S CGT TCC CTA TTT 5'
T/B_{-9} (G^S)	64.6	1.31	nil	3' AGA TAG TGG CG^{S}T TCC CTA TTT 5'
$T/B_{+8'}$ (T^S)	66.3	0.95	nil	3' AGA TAG TGG CGT^S TCC CTA TTT 5'
$T/B_{+3'}$ (T^S)	65.8	1.45	2.6	3' AGA TAG TGG CGT TCC CT^{S}A TTT 5'
$T/B_{+1'}$ (T^S)	67.5	0.91	3.1	3' AGA TAG TGG CGT TCC CTA T^{S}TT 5'

[a]T/B refers to the double-stranded nonsubstituted operator DNA. Each substituted operator sequence is referred to by the strand T (top) and B (bottom), the site of the substitution is indicated with subscript numbers indicated, and the type of substitution is shown in parentheses.

[b]K_d values were determined as described in section 5.2.7.

[c]Fraction of DNA in the cross-linked complex relative to the total DNA. With the sequences reported as 'nil' there was no detectable cross-linking.

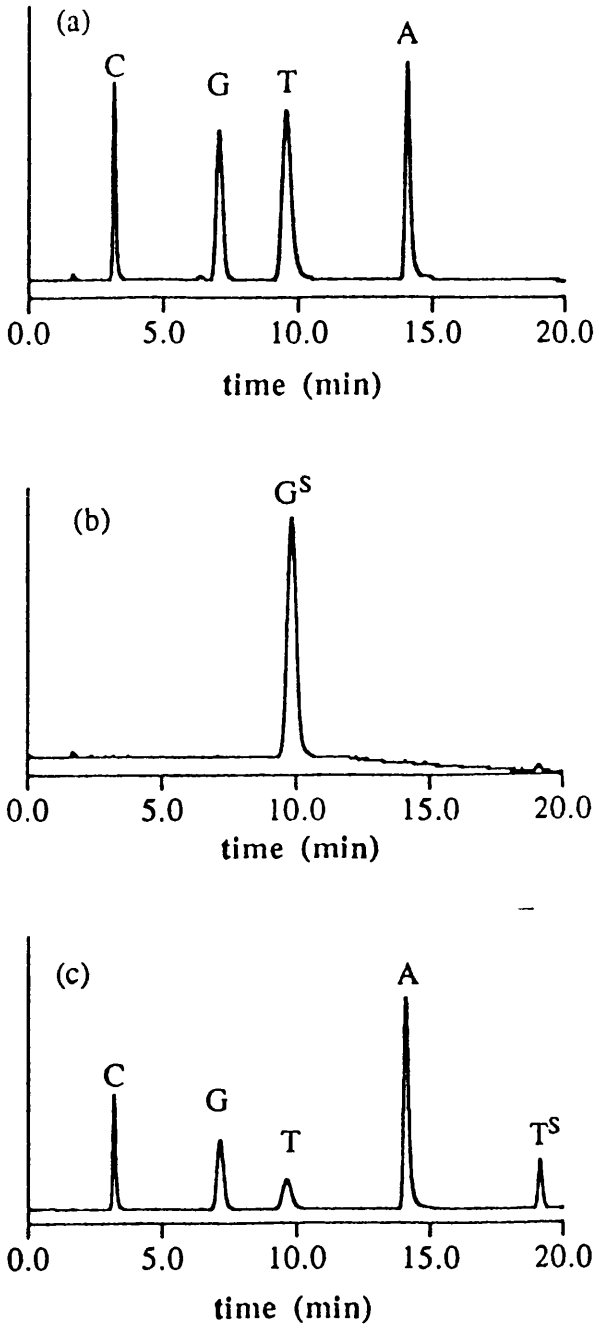


Figure 5.5. Base composition analysis by HPLC of the purified oligomers 5'-TTTA TCCCTTGCGGTG^{S}ATAGA-3' measured at 260 nm (a) and 340 nm (b) (base composition expected: $T_{8.0}$, $C_{4.0}$, $G_{4.0}$, $A_{4.0}$, $G^S_{1.0}$, found: $T_{8.5}$, $C_{4.1}$, $G_{4.0}$, $A_{3.7}$, $G^S_{1.1}$), and 5'-TCTATCACCGCAAGGGAT^{S}AAA-3' measured at 260 nm for four normal bases and then at 340 nm for 4-thiothymidine in one run (c) (base composition expected: $T_{3.0}$, $C_{5.0}$, $G_{4.0}$, $A_{8.0}$, $T^S_{1.0}$, found: $T_{3.0}$, $C_{5.2}$, $G_{4.2}$, $A_{7.6}$, $T^S_{1.1}$).

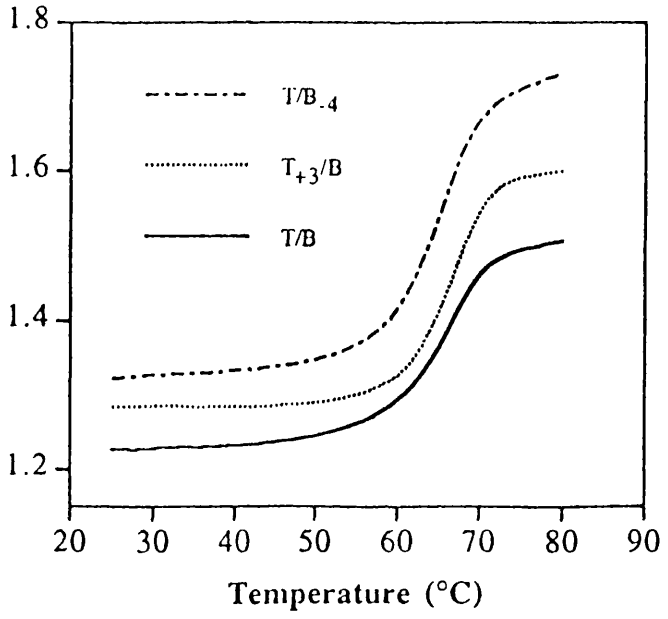


Figure 5.6. Tm curves of DNA duplexes T/B, T_{+3}/B, and T/B_{-4}.

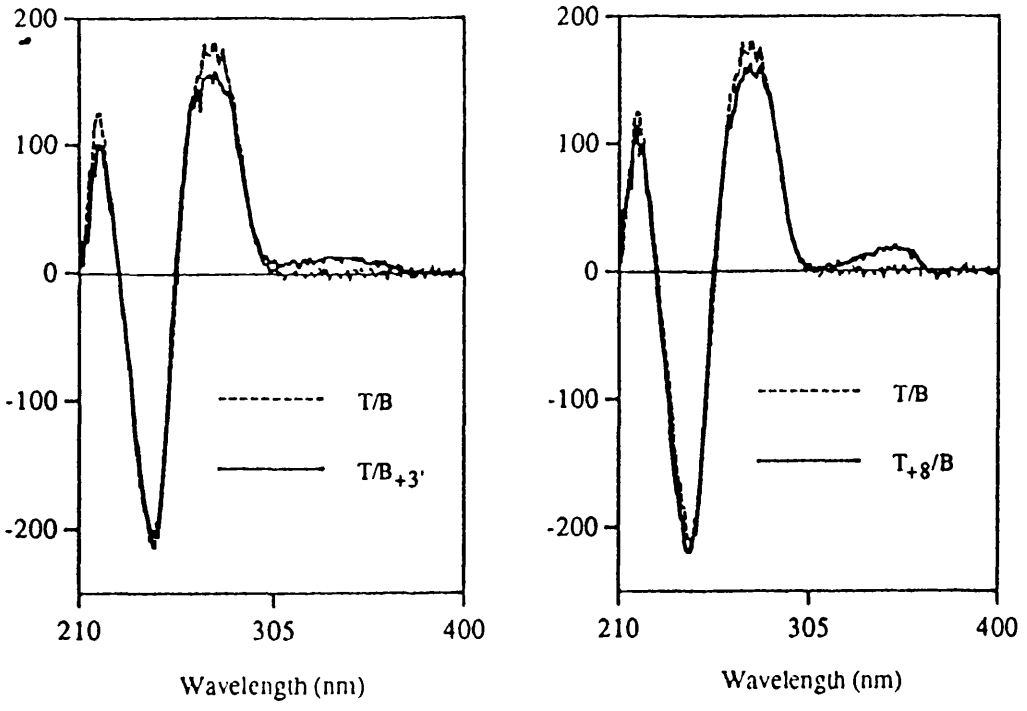


Figure 5.7. CD spectra of operator DNAs T/B, T/$B_{+3'}$, and T_{+8}/B.

5.3.2. Preparation of Cro Protein

The preparation of Cro protein was carried out using the protocol used by Dr Takeda (personal communication), one of the leading investigators of Cro protein (see section 5.2.6). The Cro gene in the plasmid is under control of the lac operator. Before inoculating the large scale broth two small samples of the inoculation were grown up, one with added ITPG to switch on any genes controlled by the lac operator, and the other sample with no added inducer. Electrophoresis of crude extract showed a band of about 7.5 KD (Figure 5.8) in the induced but not in the non-induced culture. This band was the desired Cro protein. The cells were then grown in large scale and harvested by centrifugation and frozen immediately at -70°C.

The Cro protein was purified as described in section 5.2.6. The first step was phosphocellulose column chromatography. At the salt concentration used (0.2 M KCl), the negatively charged phosphocellulose matrix does not bind any of the cellular nucleic acids or the majority of the cellular proteins and so they are washed through the column with wash buffer. Proteins that bind negatively charged compounds, including Cro protein were retained in the column during the wash process. These column bound proteins were then eluted off the column by increasing the potassium chloride concentration linearly from 0.2 M to 1 M. A_{280} of each fraction was measured (Figure 5.9). The main peak (fractions 25 to 35) was pooled, and proteins were precipitated with ammonium sulphate (60% saturation).

The protein was further purified by gel filtration on sephadex G-75. Fractions were collected and A_{280} measured. The results (Figure 5.10) showed one main peak preceded by one smaller peak. Analysis of the fractions by SDS polyacrylamide gel electrophoresis (Figure 5.8b) revealed that the main peak was Cro protein. The smaller, early eluting peak consisted of proteins with larger molecular weights. Fractions from the main peak were pooled and further purified by hydroxyapatite column chromatography. The absorbance (A_{280}) of the fractions showed one main peak followed by a much smaller peak (Figure 5.11). SDS gel electrophoresis showed that fractions 16 to 23, which contained the vast majority of the protein, were Cro protein

(Figure 5.12). These fractions were pooled, concentrated by ultrafiltration and then dialyzed. The protein was stored at -20°C in buffer J containing 50% of glycerol.

The concentration of Cro (Cro concentration is given for the protein dimer since this is the active form of the protein) was determined by measuring the absorbance at 277 nm of the solution (a 1mg/ml solution has an $A_{277} = 0.74$ (Takeda *et al.*, 1986)). The concentration of the stock solution was 211μM. 1 mL of this solution was obtained, so in total 3.1 mg of the protein was obtained. The protein bound sequence-specifically to the OR3 operator. The activity of the protein was about 50% as determined by gel retardation assay.

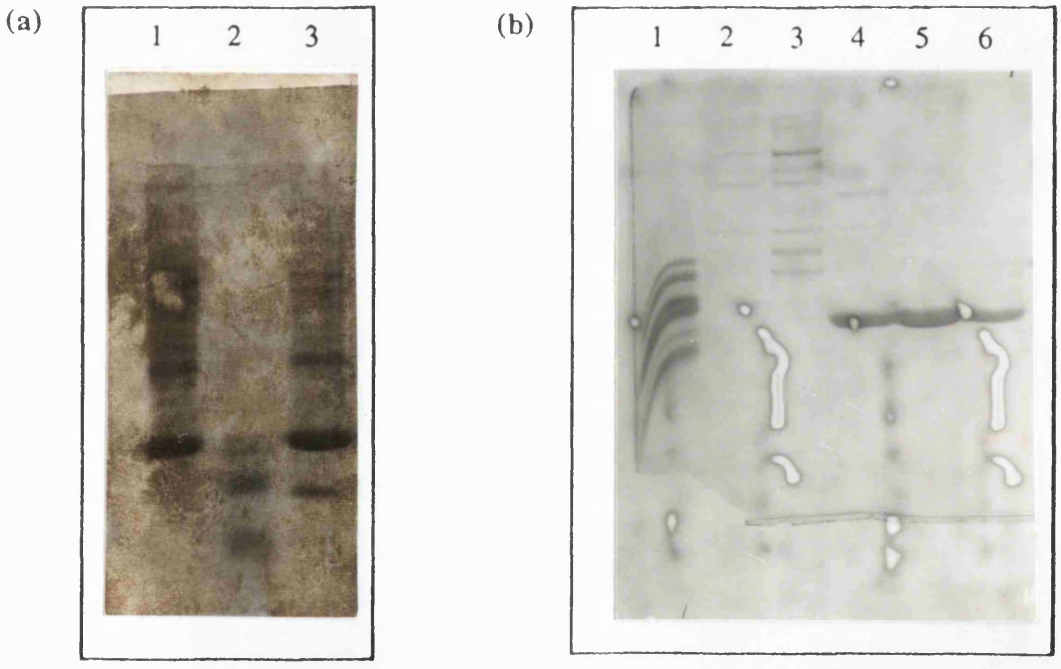


Figure 5.8. SDS-PAGE of crude Cro protein. (a) Crude extracts of cells carrying plasmid DNA pcro1. Lane 1, crude extract of uninduced cells; lane 2, protein molecular weight marker (MW are 16.9, 14.4, 10.6, 8.1, 6.2, 3.5, and 2.5. The last two proteins were merged in the gel); lane 3, crude extract of induced cells. (b) SDS-PAGE of selected fractions from the gel filtration purification of Cro protein. Lane 1, protein MW marker as in the Figure (a); lane 2-6 were fractions of 6, 7, 11, 12, and 13, respectively. The gels were Coomassie Blue stained as described in section 5.2.3.

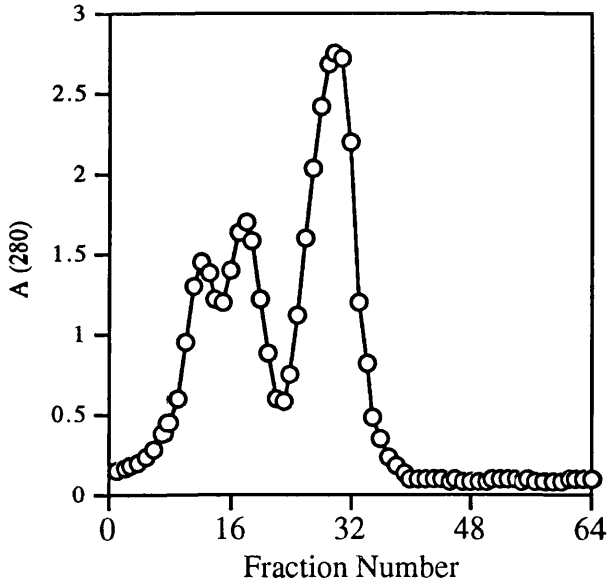


Figure 5.9. UV absorbance of fractions collected from phosphocellulose cdumm purification.

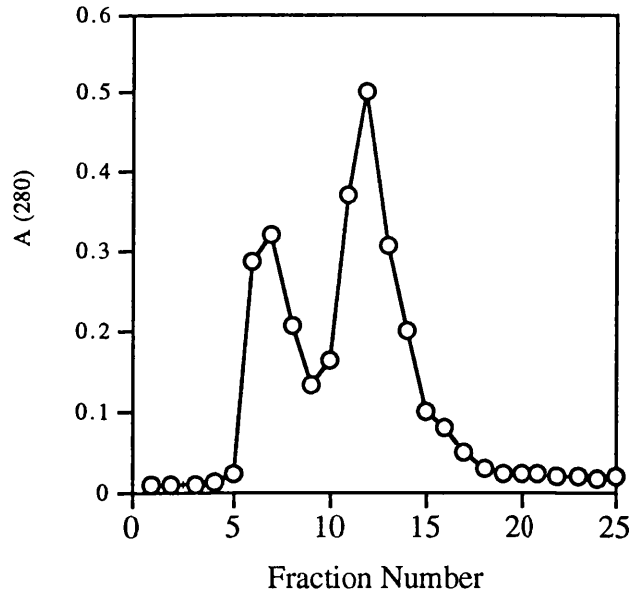


Figure 5.10. UV absorbance of fractions collected from the gel filtration of Cro pnteim.

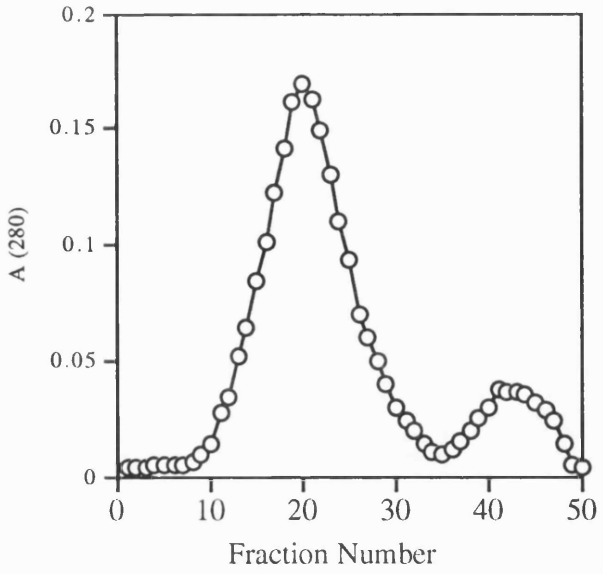


Figure 5.11. UV absorbance of the fractions collected from the hydroxyapatite column purification of Cro protein.

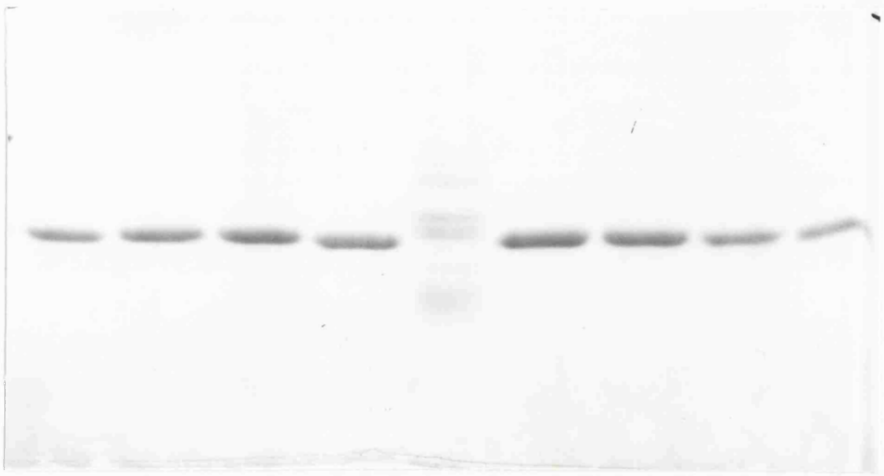


Figure 5.12. SDS-PAGE of fractions from the hydroxyapatite column purification of Cro protein. Lane 1-4 and 6-9, fraction 16-19 and 20-23, respectively; lane 5, protein MW maker as in the Figure 5.8. The gel was Coomassie Blue stained as described in section 5.2.3.

5.3.3. Equilibrium Binding Studies

The effect of the presence of thio-bases on the affinity of the repressor for the double-stranded operators was assessed by gel retardation. The Cro repressor-operator

complexes moved more slowly than DNA. After autoradiography (Figure 5.13) gel bands containing complex and free operator were excised and the radioactivity measured. The K_d (Table 5.2) was determined from the graphs of [Cro-DNA] as function of [Cro-DNA]/[Cro] (Figure 5.14). The data showed that the presence of the thiobases within the operator sequence had minimal effect on the equilibrium dissociation constant, indicating that the incorporation of the thio-bases into the operator did not appreciably perturb the binding interaction between the protein and DNA. However the protein bound to the oligonucleotides containing thioguanine with a slightly lower binding affinity than to those containing thiothymine. This result parallels the observation that the presence of thioguanine in DNA produces a greater decrease in Tm than does the presence of thiothymine.

5.3.4. Cro-Operator DNA Cross-Linking Studies

The ability of each substituted operator to covalently cross-link with Cro was assessed by irradiation with 360 nm light of a mixture of the protein and the operator DNA at room temperature (Cro-DNA complex is stable at room temperature (Takeda *et al.*, 1992)) followed by SDS-PAGE of the irradiated samples. Samples were not heated before loading on the gel because it was found that this heating breaks the cross-linked complex. UV irradiation produced an approximate 20 KD ^{32}P-labeled complex (Figure 5.15). This is the expected size of one molecule of Cro (MW 7,350) cross-linked to a 21 bp oligonucleotide (MW 12,600).

There were remarkable differences in the efficiency of crosslinking to the modified bases incorporated at different positions in the operator (Table 5.2). Operators substituted by a thiobase at positions +1, +3, -4, -6, (lane T_{+1}/B, T_{+3}/B, T/B$_{-4}$, and T/B$_{-6}$, respectively, Figure 5.15), and at their symmetry-related positions +1', +3', -4', -6' (lane T/B$_{+1'}$, T/B$_{+3'}$, $T_{-4'}$/B, and $T_{-6'}$/B, respectively, Figure 5.15) all cross-linked to Cro protein. In addition, cross-link was also detected with operators containing thiobase at positions -5 (lane T/B$_{-5}$) and -7 (lane T/B$_{-7}$). In contrast, no

cross-link was observed with thiobase substituted operators at position -2 (lane T/B_{-2}) and its symmetry-related position -2' (lane $T_{-2'}/B$), as well as at positions +8, +8', -9 (lane T_{+8}/B, $B/_{+8'}$, and T/B_{-9}, respectively). These results were reproducible and observed in experiments carried out at different times. Thiobases at the positions believed to contact with Cro protein in the Cro-DNA interaction model were cross-linked, while those where cross-linking was not observed are those postulated not to interact with Cro repressor in the model (Ohlendorf *et al.*, 1982; Takeda *et al.*, 1989; Brennan *et al.*, 1990).

To ensure that the DNA cross-link was specific for Cro, and not another protein in the binding buffer, the concentration of Cro was titrated into otherwise identical mixtures with the operator DNA, T/B_{-6}, followed by UV irradiation and SDS-PAGE (Figure 5.16). The yield of cross-linked product was strictly dependent on the presence and amount of Cro in the reaction mixtures.

Additional control experiments established (Figure 5.17) that the formation of the crosslink was dependent upon UV irradiation , as incubation of Cro with the operator DNA T/B_{-6} without UV irradiation did not lead to any new bands (lane 2, Figure 5.17), and also upon the presence of the modified base in an oligonucleotide since irradiation of the Cro with the control DNA (T/B) did not lead to the formation of any new ^{32}P-labeled complexes (lane 3), and required specific Cro-DNA complex formation as cross-linking was not observed when a 22 bp DNA duplex containing thiothymine with sequence of the CRP binding site but lacking a binding site for Cro was irradiated (lane 4). Furthermore, cross-linking was severely inhibited by 4-fold excess of T/B operator (lane 6) but was not inhibited by a 4-fold excess of CRP binding site (lane 7). Thus cross-linking of Cro to its operators is sequence-specific.

To characterize the kinetics of appearance of the cross-linked product several time course experiments were performed. Figure 5.18 shows a typical result. The yield of cross-linked product increased with time, reached the maximum at about one hour, and then subsequently decreased, indicating that the product itself is labile to UV. The

profile of each reaction was similar, but the time scale varied slightly depending on the position of the photoactive base in the operator.

To assess the UV damage to Cro protein the UV irradiated complex was analyzed with gel retardation assay to see whether the protein was still bound to the DNA. Figure 5.19 shows that after irradiation the protein was still bound to DNA, indicating that the protein retained its binding activity and long wavelength UV irradiation caused no obvious damage to the protein within the period tested (4 h). This is expected since the λmax of proteins is 280 nm, well away from the wavelength (360 nm) used in our experiments and shows the advantages of using photoactive bases with a λmax at longer wavelength than those of DNA and proteins.

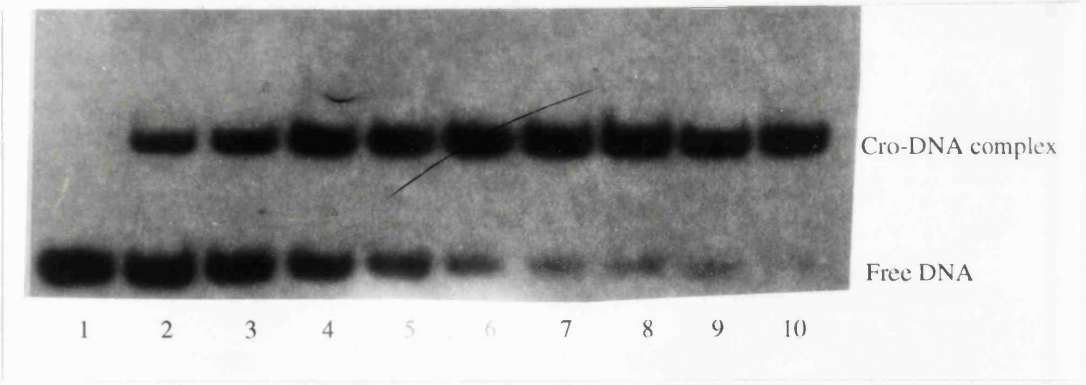


Figure 5.13. Autoradiogram of gel retardation analysis of Cro protein for determination of equilibrium constants. Lane 1-10, operator OR3 with increasing amount of Cro protein. The protein concentrations were (from lane 1, nM) 0, 0.4, 0.6, 0.9, 1, 1.5, 2.0, 2.5, 3, 9. Bands corresponding to operator and Cro repressor-operator complex were excised and counted.

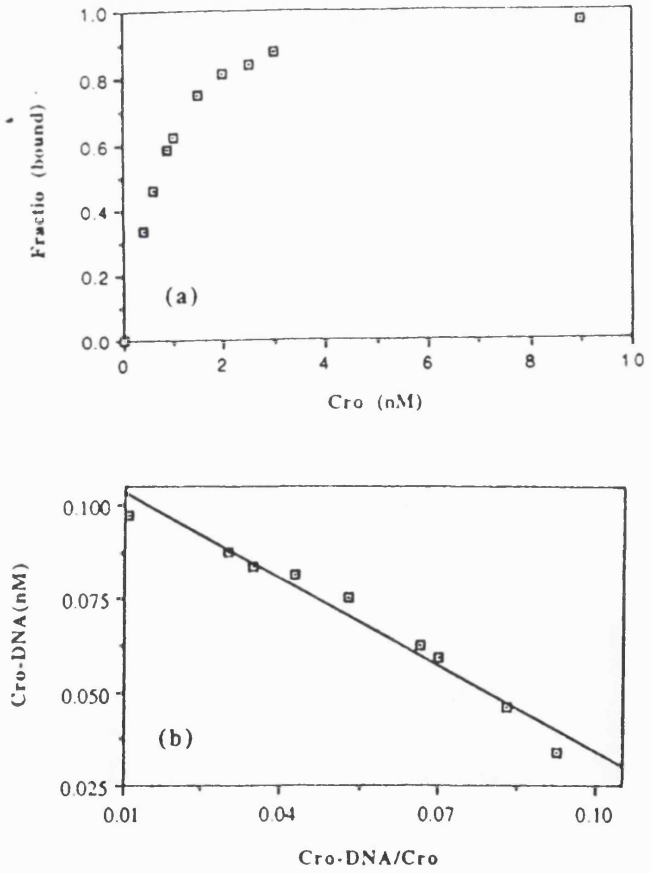


Figure 5.14. Determination of the equilibrium dissociation constant (K_d). (a) DNA binding of Cro protein as a function of Cro concentration. (b) Graph of [Cro-DNA] as a function of [Cro-DNA]/[Cro].The K_d was calculated from the slope of the straight line (K_d = -slope). Binding assay was carried out as described in section 5.3.3.

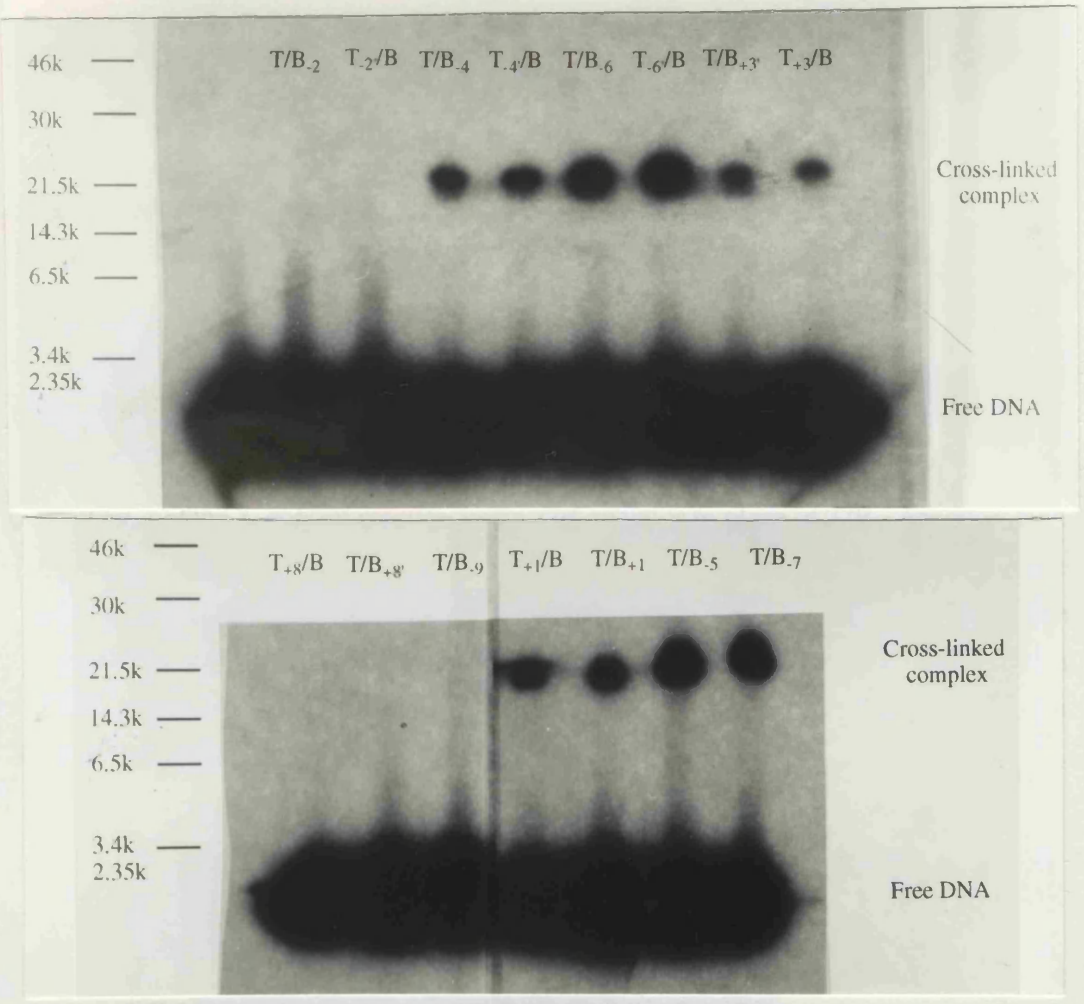


Figure 5.15. Analysis of cross-linking formation by SDS-PAGE between the Cro and the operator DNAs. Lane 1, free DNA operator (T/B); other lanes, cross-linking with operator DNAs as indicated. Experiment conditions were as described in section 5.2.8.

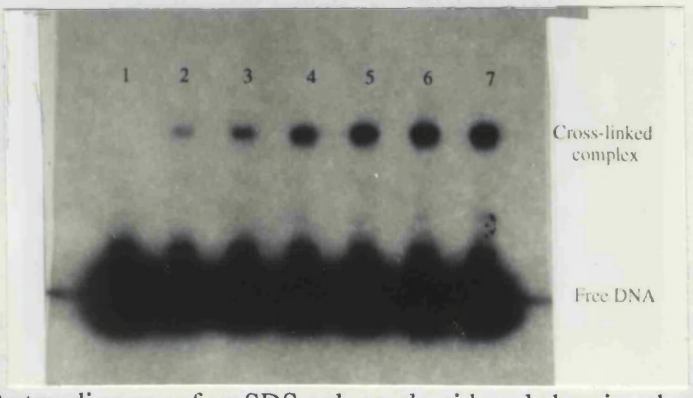


Figure 5.16. Autoradiogram of an SDS polyacrylamide gel showing that the efficiency of photocross-linking of DNA (T/B_{-6}) to Cro protein is dependent upon the concentration of the protein. Reaction mixture contained equal amounts of DNA (0.5 uM) and 2-fold serial dilutions of Cro protein as indicated. Samples were irradiated with UV lamp for 60 min.

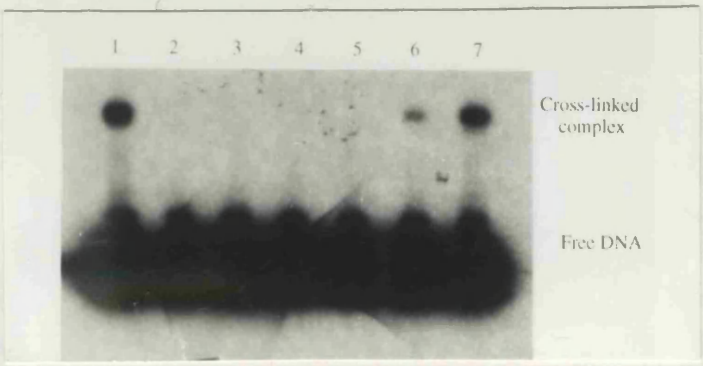


Figure 5.17. Control experiment of cross-linking between Cro and DNA. Lane 1, cross-linking reaction of Cro with T/B$_{-6}$; lane 2, control reaction omitting UV irradiation; lane 3, control reaction using T/B operator in place of T/B$_{-6}$ operator; lane 4, control reaction using CRP binding site 22 bp thiothymine containing DNA in place of T/B$_{-6}$ operator; lane 5, control reaction omitting Cro; lane 6, control reaction in the presence of a 4-fold excess of T/B DNA; lane 7, control reaction in the presence of a 4-fold excess of CRP binding site 22 bp DNA. The samples were irradiated for 60 min and analyzed by SDS-PAGE as described in section 5.2.3.

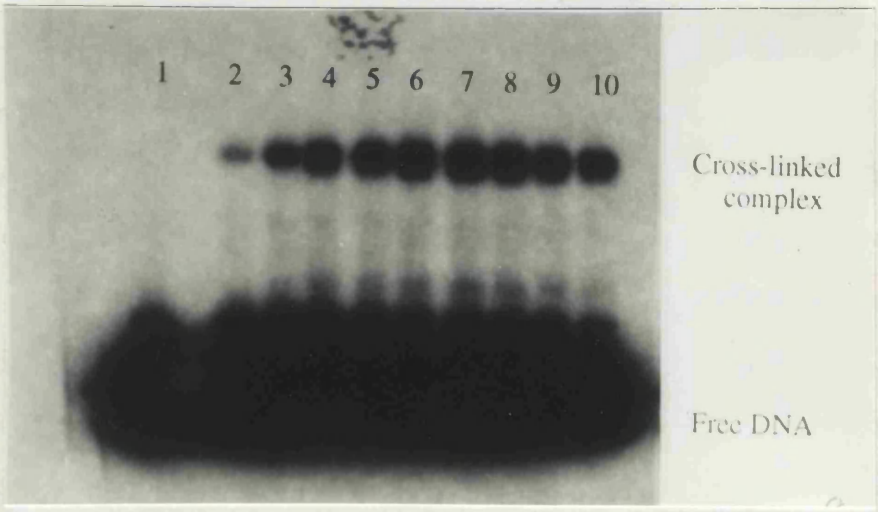


Figure 5.18. Time course of cross-link formation between T/B$_{-6}$ and Cro. Aliquots were taken at different time intervals from the same reaction mixture after UV irradiation. The time course was (from lane 1 to 10) 0, 2', 4', 8', 16', 32', 64', 2h, 4h, 8h, respectively.

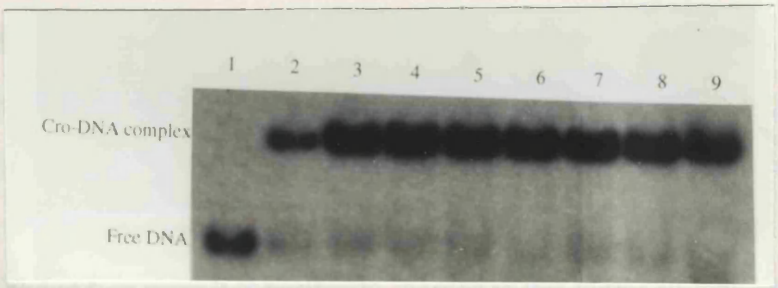


Figure 5.19. Gel retardation assay of UV irradiated samples showing that Cro protein was still bound to DNA after being irradiated for 4 h. Lane 1, free DNA; lane 2, DNA-Cro complex; lane 3-9, aliquots taken from the reaction mixture after being irradiated for 4', 8', 16', 32', 64', 2h, 4h, respectively.

5.3.5. Isolation of Cross-Linked Complexes.

After cross-linking Cro protein with its operator DNAs, we attempted to identify the amino acid residues covalently linked to the DNA. The basic procedure involved proteolysis of the cross-linked complex, separation of the digested DNA-peptides, and determination of the sequence of the peptide covalently linked to the DNA. The protocols for these experiments were set up using the cross-linked complex of Cro-T/B$_{-6}$ in which thioguanine was the photoactive base in the DNA. The operator was irradiated in the presence of Cro protein under large scale conditions (2 ml of the complex were irradiated), which it was hoped, would yield sufficient material to identify the amino acids in the protein sequences that cross-linked to the modified base in the operator. BSA was not included because its presence could complicate the following purification steps. Two methods, preparative SDS-PAGE and FPLC, were used for isolation of the protein-DNA complex and both provided good separation between the cross-linked complex and free DNA and protein. For SDS-PAGE, the irradiated sample was first concentrated with Amicon ultrafiltration and then loaded onto a preparative SDS gel and electrophoresed under denaturing conditions. An autoradiogram of the wet gel (Figure 5.20) shows good isolation of the cross-linked complex. The cross-linked band was excised and electroeluted from the polyacrylamide gel, and the eluate was dialyzed. The recovery was 39% (Table 5.3) for four steps (concentration/electrophoresis/ electroelution/dialysis) calculated from the total amount

of radiolabel in the eluted complex.

The different species (free DNA and protein, and cross-linked complex) in the irradiated sample have different binding affinities to mono Q anion-exchange matrices at the buffer system used (pH 7.4, Tris). It was reasoned that free Cro protein (pI=10.58) would be eluted first since it is positively charged in this buffer system, and therefore should not bind to the positively charged column packing, the cross-linked complex should elute second as Cro contains 5 positive charges, and the total number of negative charge of the Cro-DNA complex is five less than that of free DNA, and free DNA would eluted last since it contains the largest number of negative charges. As expected, the cross-linked complex was eluted earlier than the free DNA and well separated from other undesired species (Figure 5.21). The radioactivity profile (Figure 5.22) indicated that the radioactivity was presented in peak 1, 3, and 4. The fractions from peak 3 (Figure 5.21) were desalted and concentrated. The recovery of the complex up to this stage was 44% for three steps (FPLC separation/desalting/ concentration).

To confirm that peak 3 was a covalently-linked complex, the fractions collected from this peak, together with other radiolabeled fractions, were analyzed by SDS-PAGE (Figure 5.23). As expected, peak 4 contained mainly the free DNA, peak 1 contained species with high mobility in the SDS gel, but unexpectedly peak 3 gave two bands. The upper band had an apparent molecular mass of 20 KD which corresponds to the predicted size of Cro bound to a 21 bp oligonucleotide, and the lower band had the mobility of a double-stranded 21-mer indicating that this band was free DNA.

There are two possible explanations for these two bands. First peak 3 (Figure 5.21) could be a mixture of covalently cross-linked and non-covalently bound complexes, and second, the covalently cross-linked complex is not stable and the bond linking two species was broken during the handling and storage. To test the first possibility, the non irradiated complex containing the same components as that used for cross-linking was subject to FPLC. The Figure 5.24 showed no peak corresponding to peak 3 in Figure 5.21. This ruled out the first possibility and indicated that the

formation of the second band most likely resulted from the breakage of the cross-linked complex during the handling and storage. This breakage was also observed in the complex isolated by SDS-PAGE (lane 4, Figure 5.23). Further experiments showed that the cross-link was unstable to acid such as trifluoroacetic acid, and was labile under alkaline condition, and the breakage of the complex was observed even in NH_4HCO_3 solution (pH 8.0). The complex was reasonably stable at -20°C, but decomposition of the complex was observed at room temperature. Furthermore the complex was moderately more stable in Tris buffer, especially in the presence of DTT, than in NH_4HCO_3 or K-phosphate buffers. Similar instabilities were also observed with other cross-linked Cro-DNA complexes.

5.3.6. Attempts to Isolate and Identify the Peptide Cross-linked to DNA.

To cleave Cro in the complex so that the amino acids involved in the cross-link could be identified by sequencing the DNA-bound peptides, the isolated complex was subjected to extensive proteolysis. Preliminary tests showed that both chymotrypsin and modified trypsin would digest Cro. Therefore, these two proteases were used for digestion. Both produced a new band migrated with a mobility similar to free DNA (Figure 5.25), a result which seems to confirm that digestion was sufficient to generate a small peptide attached to DNA.

Following digestion with proteases, our attempt to isolate the DNA-peptide complex was made using FPLC. It was reasoned that DNA with a small attached peptide should bind similarly strongly to the anion-exchange column (Mono Q) like free DNA, thus allowing it to be separated readily from the early-eluting, non-DNA bound peptides. Chromatography of the digested sample on Mono Q column resulted in elution of a peak at identical position to the corresponding free DNA (Figure 5.26). This peak was desalted, concentrated by freeze drying and subjected to peptide sequencing. Unfortunately, the sequencing did not show any signals of amino acids. Three attempts were made with the Cro-T/B_{-6} complex, sadly none of these attempts

were successful. The attempt to identify the cross-link point of Cro-T/B$_{-5}$ complex, in which the photoactive base was thiothymine, also turned out to be unsuccessful. These results seemed to indicate that peak 2 in Figure 5.26 was not a peptide-DNA species, but a free DNA produced from the decomposition of covalently-linked Cro-DNA and peptide-DNA species.

The yields of DNA-peptide through the different stages of purification were followed by radiolabel in the DNA, and the data for two thio-base substituted DNA, calculated from the total amount of radiolabel in DNA recovered in each step, are shown in Table 5.3. The recovery was reasonable at each step, and FPLC purification gave slightly better yields than the SDS-PAGE. It should be pointed out that the surfaces of tubes, vessels, and other containers should be silanized if possible because it reduces the loss of the samples due to its adsorption to surfaces. However radiolabel, that could not be solubilized by extensive washing, was still detected on such surfaces.

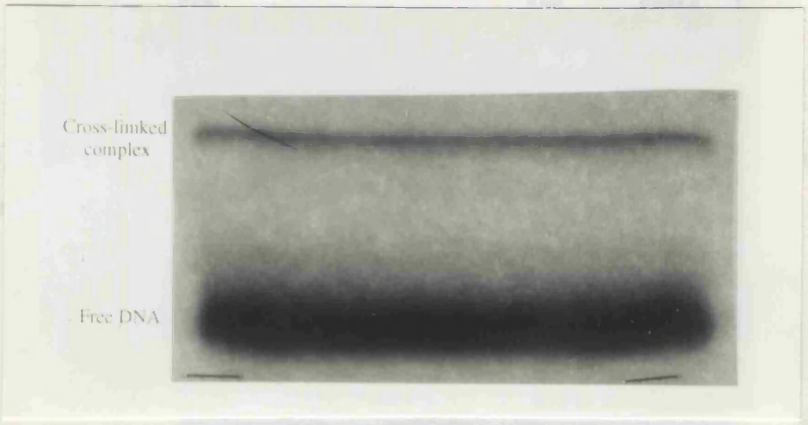


Figure 5.20. Isolation of cross-linked Cro-T/B$_{-6}$ operator complex with preparative SDS-PAGE. Following irradiation and concentration, the sample was electrophoresed on a 20 % SDS polyacrylamide gel to resolve free DNA and protein from the cross-linked species. The bands were visualized by autoradiography of the wet gel and removed from the gel for further processing.

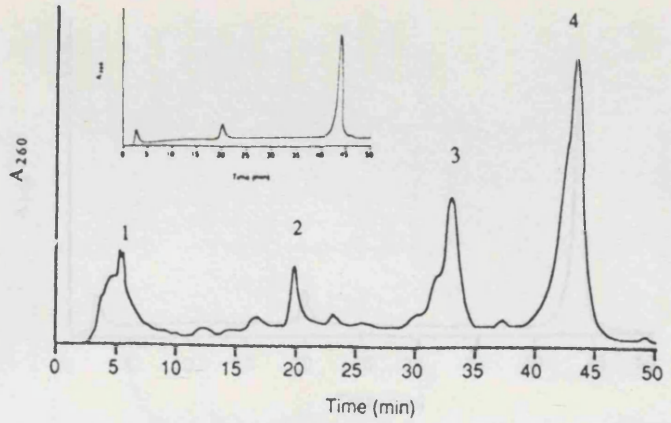


Figure 5.21. FPLC trace of UV irradiated Cro-T/B$_{-6}$ complex (Peak 2 was the excess of complementary strand of the oligomer containing thiobases. Inserted fig. was the HPLC trace of the T/B-6 operator).

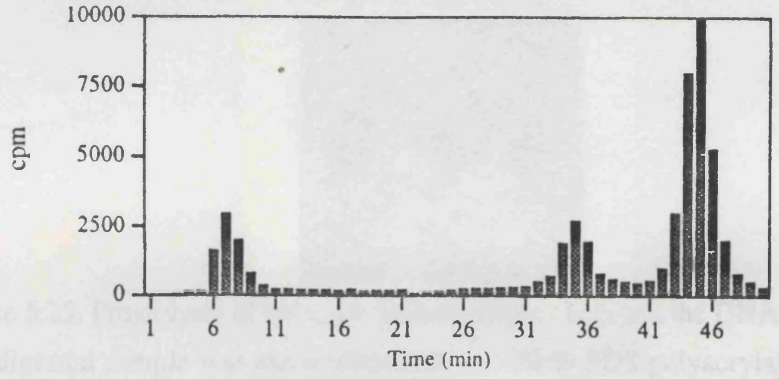


Figure 5.22. Radiolabel profile of the fractions collected from the FPLC isolation of the cross-linked complex.

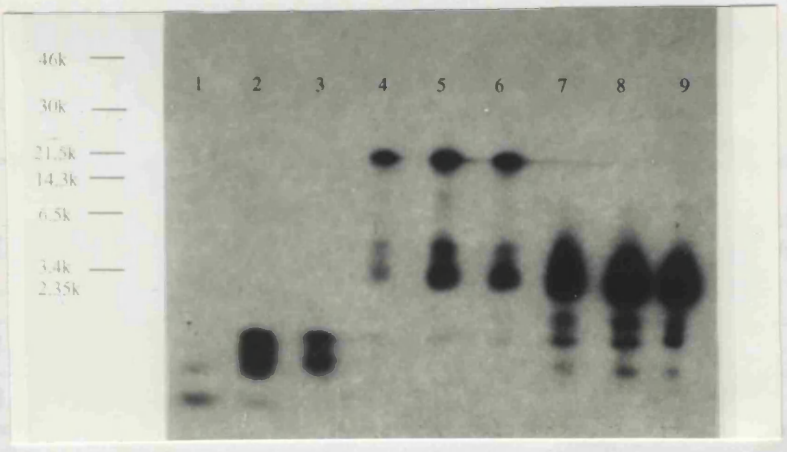


Figure 5.23. SDS-PAGE analysis of the selected fractions from Figure 5.21. Lane 1-3, fractions from the peak 1; lane 4, cross-linked complex separated from SDS-PAGE; lane 5-6, fractions from the peak 3; lane 7-9, fractions from the peak 4.

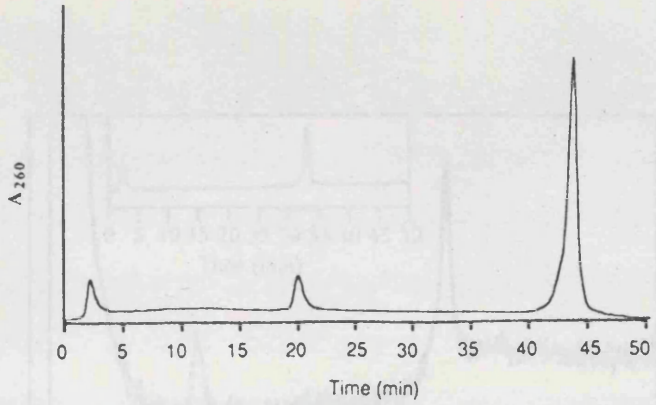


Figure 5.24. FPLC trace of non-irradiated complex containing same components as that used for UV irradiation showing no corresponding peak to the peak 3 in Figure 5.21.

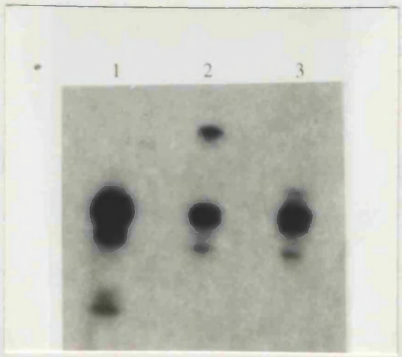


Figure 5.25. Proteolysis of the cross-linked complex between the DNA T/B$_{-6}$ and Cro. The digested sample was electrophoresed on a 20 % SDS polyacrylamide gel which was analyzed by autoradiography. Lane 1, digested mixture with modified trypsin; lane 2, mixture of free DNA and X-linked Cro-DNA; lane 3, digested complex with chymotrypsin.

Table 5.3. Recovery* of cross-linked complexes.

DNA	Separation of Cro-DNA complex		Separation of peptide-DNA complex
	concentration/ SDS gel/ elution/dialysis	FPLC/ desalting/ concentration	
T/B$_{-6}$	38.5%	44.2%	20.1%
T/B$_{-5}$	32.1%	40.3%	18.6%

*Recovery was calculated from the total amount of radiolabel in DNA recovered at each step, with the irradiation step constituting 100% recovery. The maximum theoretical recovery corresponds to the efficiency of cross-linking at each site : T/B-6, 7.2%; T/B-5, 8.6%.

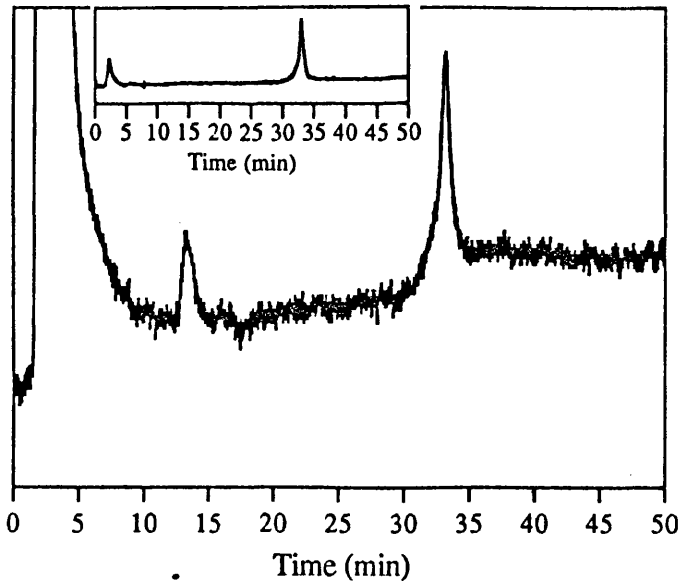


Figure 5.26. FPLC trace of the digested sample (Inserted fig for the free DNA).

5.4. DISCUSSIONS.

The cross-linking of Cro protein to DNA containing 4-thiothymine or 6-thioguanine at a single site was chosen as a means of probing the interaction points between the protein and the operator. To ascertain that the introduction of the modified bases into the DNA operator did not affect appreciably the binding affinity between the two molecules the ability of the substituted operators to bind Cro was measured under equilibrium conditions using the gel retardation method. Compared to other commonly used assays for protein-DNA interactions, such as filter binding assay (Kim *et al.*, 1987), this technique is quicker, simpler and more reliable (Carey, 1991). The results (Table 5.1) shows that incorporation of the modified bases at a single site within the operator minimally affects the equilibrium binding affinity for Cro, demonstrating that, structurally, 4-thiothymine and 6-thioguanine are good candidates as photoactive bases in the photochemical cross-linking approach of studying DNA-protein interactions.

After irradiation covalently cross-linked protein-DNA complexes were easily identified in an SDS-PAGE system since these species have lower mobility than free DNA and protein. The specificity of cross-linking was established by the following observations. Cross-linking was not observed when the mixture of Cro protein and a DNA containing a CRP binding site while lacking Cro binding site was irradiated, and excess specific DNA reduced cross-linking while excess nonspecific oligonucleotide had no effect (Figure 5.17). Additionally, cross-linking depended upon the presence and the amount of the Cro protein in the reaction mixture (Figure 5.16), indicating cross-linking involves Cro protein.

Numerous studies have indicated that only the native protein complexed with nucleic acids can be cross-linked. That cross-linking can take place only on amino acid residues that are at the interface of the protein-DNA complex has been supported by studies on proteins whose three-dimensional structures have been solved and whose cross-linking sites have been identified (for a review see Williams and Konigsberg, 1991). For example, cysteine 33 in the bacteriophage fd gene 5 protein, that was cross-linked to fdDNA, is located within a predicted DNA-binding groove (Brayer and

Mcpherson, 1983); the three amino acids, serine 80, isoleucine 81, and threonine 82, which were cross-linked in the ribonuclease A-(pUp) complex, are known to constitute part of the binding site for the pyrimidine ring (Havron and Sperling, 1977). A more recent example is the alanine 238 in the GCN4 protein that cross-linked to 5-position of the bottom strand thymine at position +3 of its 9-base-pair DNA binding site. The crystallographic structure of GCN4-DNA complex revealed that this amino acid residue is indeed in van der Waals contact with the 5-methyl group of the thymine (Blatter *et al.*, 1992).

Absence of cross-linking could be ascribed to either no contact or insufficiently close proximity between amino acid residues and DNA bases, or, to low reactivity of the amino acid residue contacting the DNA. Unfortunately, no data are currently available on the relative reactivities of the 20 different amino acids to a photochemically excited base (Williams and Konigsberg, 1991). Although some model studies have been carried out on amino-acid-nucleotide mixtures, these model studies, in general, seem to be of little use in this regard. For example, while Schott and Shetlar (1974) failed to observe any cross-linking between phenylalanine and thymine when solutions of the free base and the free amino acid were irradiated, phenylalanine in three DNA-binding proteins has been identified by cross-linking to the bound DNA (Merrill *et al.*, 1984; Merrill *et al.*, 1988; Shamoo *et al.*, 1988). Therefore, as pointed out by Williams and Konigsberg (1991) and supported by model systems studied in attempting to understand the chemistry of the photo-cross-linking (Hockensmith *et al.*,1991), in the absence of any relevant data, it seems reasonable to assume that any of the 20 amino acids present in proteins can, in principle, be photo-cross-linked to nucleic acids.

Given the assumptions above, our results of cross-linking between Cro and DNA are consistent with the prediction of the initial model (Ohlendorf *et al.*, 1982) for Cro binding as later revised by Takeda et al.(1989) and further revised by Brennan et al.(1990) (Figure 5.27).

In the model (Figure 5.27), both guanine-7 and guanine-6 are predicted to

contact with Arg-38 by forming three hydrogen bonds between the guanidinium group of arginine residue with O-6 and N-7 of guanine-6 and N-7 of guanine-7, respectively. These predications are consistent with our observations that operators containing 6-thioguanine at -7 or -6 position both were cross-linked to Cro protein.

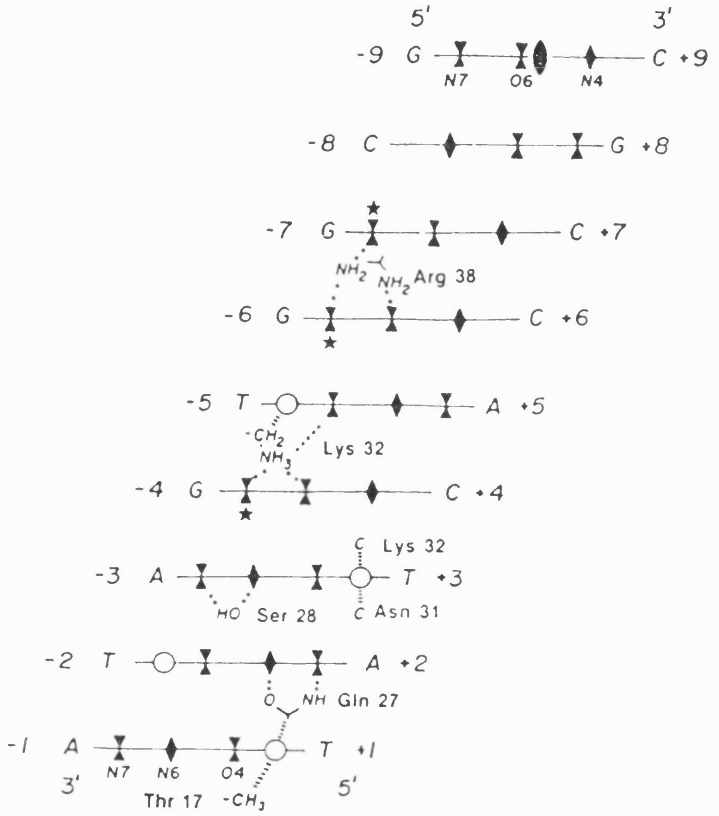


Figure 5.27. Schematic representation of the presumed sequence-specific interactions between Cro and the parts of base pairs exposed within the major groove of the DNA. The dyad symbol within the topmost base pair indicates the center of the overall 17-base-pair binding region. The symbols are as follows: ⧗ , hydrogen bond acceptor; ⧫, hydrogen bond donor; ○ , methyl group of thymine; ★, guanine N7 which is protected from methylation when Cro is bound. Presumed hydrogen bonds between Cro side chains and the bases are indicated (····) and van der Waals contacts between Cro and the thymine methyl groups are shown (IIII). (Taken from Takeda *et al.* 1992).

Lys-32 is postulated to make four specific interactions with thymine -5 and guanine -4, its ε-amino group participating in the three hydrogen bonds with N-7 and

O-6 groups of guanine at -4, and O-4 group of thymine at -5, and its hydrophobic side chain making a van der Waals contact with the methyl group of the thymine -5 (Figure 5.27). These postulations are consistent with our cross-linking results that DNA containing either 4-thiothymine at -5 or 6-thioguanine at -4 produced cross-linked complex with Cro upon UV irradiation.

Thymine +3 is thought to make van der Waals contacts via its 5-methyl group with Asn-31 and Lys-32. This is in agreement with the cross-linking results observed here which imply that thymine +3 seemed to contact with Cro protein.

In the initial model (Ohlendorf *et al.*, 1982), thymine-2 is suggested to make no contacts with Cro. However, later Takeda et al. (1989) suggested that the O-4 group of thymine -2 probably forms a hydrogen bond with Tyr-26. The Cro-DNA crystal structure study (Brennan *et al.*, 1990) was in favour of the initial model, indicating that Tyr-26 is not involved in any direct contact to the operator. The experiment presented here further supported the initial model since no cross-link was observed between thiothymine -2 and Cro.

Thymine +1 was predicted in the model (Figure 5.27) to make van der Waals contacts with Glu-27 and Thr-17. In agreement with this prediction, our experiment also showed cross-linking formation at this base site, indicating thymine +1 contacts Cro protein.

As originally proposed by Ohlendorf et al (1982) and supported by the crystal structure studies of Cro-DNA (Brennan *et al.*, 1990), for the most part of the DNA operator, the conformation of those bases involved in protein recognition is typical B-form. In the middle of the operator, however, the DNA is bent, and functional groups of the central three bp of the operator do not interact with Cro protein. Figure 5.15 shows that no cross-linking were formed in the middle of three base-pairs, supporting the above predictions.

Since Cro dimer has twofold symmetry (Figure 5.2) and the six operators also have approximate twofold symmetry, it was predicted in the initial model (Anderson *et al.*, 1981; Ohlendorf *et al.*, 1982) that each subunit of Cro dimer makes similar

contacts with the two halves of pseudo-symmetric operators. Takeda et al. (1989) also expected, from the results of systematic base substitution, that Cro dimer binds operator DNA by aligning the twofold symmetry of the protein to the pseudo-twofold symmetry of the operator. The recent crystal structure of Cro-DNA complex (Brennan *et al.*, 1990), albeit at low resolution, provided further support to these predictions. In our investigation, similar cross-linking results were observed between operators containing a thiobase at symmetrically-related sites. Thus cross-linking was observed in the symmetrically-related positions of +1 and +1', +3 and +3', -4 and -4', and -6 and -6', while no cross-linking was produced at -2 and -2' positions (Figure 5.15). These results imply that Cro dimer has similar contacts to the two halves of the operator. However, some differences did exist in the photoreactivity (reflecting in the cross-linking yield) between the symmetrically related bases. Although it cannot be excluded that these differences in photoreactivity are reflections of DNA-sequence related effects, such as differences in electron resonance in the thiobase ring caused by stacking interactions at different positions in the operator, or that it was simply an experimental error, it is possible that this pattern of reactivity reflects an asymmetric binding of Cro dimer to the pseudo twofold symmetry of the OR3 operator. The base sequence of the two halves of the OR3 operator differs at position 5, 7, and 8. Because of the different functional groups presented in the major groove, these differences in base sequence could lead directly to differences in the contacts. Although the crystal structure of Cro-DNA complex did not reveal these asymmetric interactions due to low resolution of the structure (Brennan *et al.*, 1990), the 1.8Å crystal structure of λ repressor-OL1 complex has revealed that, despite important similarities, the repressor dimer makes substantially different contacts on the two halves of the OL1 operator (Beamer and Pabo, 1992)

The main challenge to the proposed model (Figure 5.27) came from Ptashne and co-workers, (Hochschild *et al.*, 1986; Hochschild and Ptashne, 1986). Based on a set of genetic and chemical modification data obtained from the binding studies *in vivo*

of wild and mutant Cro protein to various operator sequences, they suggested that a subset of the specific contacts that Cro makes with the λ operator is identical to a subset of the contacts made by λ repressor, i.e. Ser-28 interacts with guanine at -4 rather than adenine at -3 as the model of Figure 5.27 suggested, and Lys-32 interacts with guanosine at -6 and probably the 5th base pair but not with guanine -4 as the original model suggested. Although the crystal structure of Cro-DNA complex was determined recently (Brennan *et al.*, 1990), due to low resolution (3.9Å) of the structure it failed to resolve unequivocally the differences between the model in Figure 5.27 and the model proposed by Hochschild and colleagues. Brennan et al. speculated that Ser-28 contacts base-pair 3, 4, or both and Lys-32 is closest to base-pair 5, probably also contacts base-pair 6 and might reach base-pair 4. One of initial aims of this study was to resolve these arguments. Cross-linking Cro protein to T/B-4 and T/B-6 respectively, and identifying the amino acids cross-linked to these two bases could resolve this dispute. For example, identification of Ser-28 and Lys-32 as the amino acids cross-linked to G-4 and G-6 respectively would strongly support Ptashne and co-workers' argument. The failure to achieve this aim can mainly be ascribed to the instability of the cross-linked complex. The complexes are labile to both acidic and alkaline conditions, and gradually break down even at room temperature. In an attempt to identify the cross-link point, Cahill et al. also observed similar instability of the cross-linked complex formed upon UV irradiation between serum response factor (SRF) and oligonucleotides encoding the c-fos serum response element containing 6-thioguanine. Identification of the amino acid residues cross-linked to the 6-thioguanine in the oligonucleotides was unsuccessful despite tremendous efforts, mainly due to the instability of the cross-linked bond (Cahill, M. A. et al. personal communications). Although cross-linking of oligonucleotides containing 4-thiothymine or 6-thioguanine to the Eco RV restriction endonuclease and modification methylase was reported three years ago, so far the identification of the amino acid residues involved in the cross-links has not yet been reported. The work reported in this chapter also underlines the

difficulties involved in the identification of the cross-link points between proteins and the thiobase in the oligonucleotides. Although it is possible that a protocol might eventually be established where the cross-link is stable to the manipulation required to identify the peptide attached to the DNA, that goal has not yet been achieved.

Although the mechanism of crosslinking between thiobases in DNA and protein is unknown, the photochemistry of 4-thiopyrimidine nucleoside is well documented (for a review see Favre, 1990; Nikiforov and Connolly 1992). Briefly, two reaction pathways predominate during the photoreactions of 4-thiopyrimidines. Under aerobic conditions 4-thiopyrimidines are oxidized to their corresponding 4-sulphonate derivatives. These derivatives are extremely reactive and react with nucleophiles with displacement of the 4-sulphonate group, thus, the nucleophiles become attached to the C-4 position of the pyrimidine ring and the sulphonate is lost. Under anaerobic conditions a radical mechanism produces cross-linking to the C-6 atom of the pyrimidine ring with concomitant reduction of the 5, 6-double bond. Much less is known about the photochemistry of 6-thiodeoxyguanosine, but it also is suggested to form sulphonate derivatives that can react with nucleophiles. However, it is far from clear whether these reaction pathways represent the mechanisms of photocross-linking of proteins to their corresponding DNA.

It would help us to understand the mechanism of the photocross-linking if one could find out whether the sulphur atom forms part of the cross-linking. One approach to answer this question would be to use oligomers containing ^{35}S-labeled thymine at the 4-position or guanine at the 6-position. For example, absence of the band of cross-linked complex in the autoradiograph of SDS-PAGE of UV irradiated reaction mixture containing Cro and ^{35}S-labeled operator T/B$_{-5}$ (this operator has been established to be able to cross-link to Cro protein in this project) would mean that the sulphur is lost during photoreaction and therefore it does not form a part of the cross-linking. Such result would provide support to the mechanism under aerobic condition as described above. Our lab is currently developing the chemical methods for synthesis of oligonucleotides containing ^{35}S-labeled 4-thiothymine and 6-thioguanine.

REFERENCES

Adams, S. P., Kavka, K. S., Wykes, E. J., Holder, S. B. and Galluppi, G. R. (1983)

J. Am. Chem. Soc. **105**, 661-663.

Agarwal, K. L., Yamazaki, A., Cashion, P. J. and Khonara, H. J. (1972)
Angew. Chem. Int. Ed. Engl. **11**, 451-459.

Aiken, C. R. and Gumport, R. (1991) *Methods in Enzymology* **208**, 433-457.

Allen, C. F. H., Spangler, F. W. and Webster, E. R. (1963). *Organic Synthesis*
New York, John Wiley and Son, Inc. 433-434.

Allen, T. D., Wick, K. L. and Matthews, K. S. (1991) *J. Biol. Chem.*
266, 6113-6119.

Alvarado-Urbina, G., Sathe, G. M., Liu, W.-C., Gillen, M. F., Duck, P. D.,
Bender, R. and Ogilvie, K. K. (1981) *Science* **214**, 270-274.

Anand, N. N., Brown, D. M. and Salisbury, S. A. (1987) *Nucleic Acids Research*
15, 8167-8176.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. (1981) *Nature*
290, 754-758.

Andrus, A., Efcavitch, J. W., McBride, L. J. and Giusti, B. (1988)
Tetrahedron Letters **29**, 861-864.

Arnhelm, N. and Levenson, C. H. (1990) *Chem. Eng. News* **68**, 36.

Bahl, C. P., Wu, R., Itakura, K., Katagiri, N. and Narang, S. A. (1976)
Proc. Nat. Acad. Sci. USA **73**, 91-94.

Bartholomew, B., Meares, C. F. and Dahmus, M. E. (1990) *J. Biol. Chem.*
265, 3731-3737.

Basu, A. K. and Essigmann, J. M. (1988) *Chemical Research in Toxicology* **1**, 1-18.

Beamer, L. J. and Pabo, C. L. (1992) *J. Mol. Biol.* **227**, 177-196.

Beaucage, S. L. and Caruthers, M. H. (1981) *Tetrahedron Letters* **22**, 1859-1862.

Beaucage, S. L. and Iyer, R. P. (1993) *Tetrahedron* **49**, 1925-1963.

Benson, N. and Youderian, P. (1989) *Genetics* **121**, 5-12.

Bhanot, O. S., Grevatt, P. C., Donahue, J. M., Gabrielides, C. N. and Solomon, C. N. (1992) *Nucleic Acids Research* **20**, 587-594.

Blatter, E. E., Ebright, Y. W. and Ebright, R. H. (1992) *Nature* **359**, 650-653.

Bradley, D. H. and Hanna, M. M. (1992) *Tetrahedron Letters* **33**, 6223-6226.

Branden, C. and Tooze, J. (1991) *Introduction to Protein Structure* New York and London, Garland Publishing, Inc.

Brayer, G. D. and Mcpherson, A. (1983) *J. Mol. Biol.* **169**, 565-596.

Brennan, C. A., Van Cleve, M. D. and Gumport, R. I. (1986) *J. Biol. Chem.* **261**, 7279-7286.

Brennan, R. G., Roderick, S. L., Takeda, Y. and Matthews, B. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8165-8169.

Brown, D. M., Hewlins, M. J. E. and Schell, P. (1968) *J. Chem. Soc. Chem. Commun.* 1925-1929.

Calladine, C. R. and Drew, H. R. (1992) *Understanding DNA* London, Academic Press.

Carey, J. (1991) *Methods in Enzymology* **208**, 103-117.

Casale, R. and McLaughlin, L. W. (1990) *J. Am. Chem. Soc.* **112**, 5264-5271.

Catalano, C. E., Allen, D. J. and Benkovic, S. J. (1990) *Biochemistry* **29**, 3612-3621.

Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S. and Verdine, G. L. (1991) *Biochemistry* **30**, 11018-11025.

Chen, Y. and Ebright, R. H. (1993) *J. Mol. Biol.* **230**, 453-460.

Chollet, A., Chollet-Damerius, A. and Kawashima, E. H. (1986) *Chemica Scripta* **26**, 37-40.

Chollet, A. and Kawashima, E. (1988) *Nucleic Acids Research* **16**, 305-317.

Christie, N. T., Drake, S., Meyn, R. E. and Nelson, J. A. (1984) *Cancer Research* **44**, 3665-3671.

Christopherson, M. S. and Broom, A. D. (1991) *Nucleic Acids Research* **19**, 5719-5724.

Christopherson, M. S. and Broom, A. D. (1994) *Nucleosides & Nucleotides* **13**, 351-368.

Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C. and Pabo, C. O. (1991) *Science* **254**, 267-270.

Connolly, B. and Newman, P. (1989) *Nucleic Acids Research* **17**, 4957-4974.

Cosstick, R., Li, X., Tuli, D., William, D. M., Connolly, B. A. and Newman, P. C. (1990) *Nucleic Acids Research* **18**, 4771-4778.

Cowart, M., Gibson, K. J., Allen, D. J. and Benkovic, S. J. (1989) *Biochemistry* **28**, 1975-1983.

Cowart, M. and Benkovic, S. J. (1991) *Biochemistry* **30**, 788-796.

Dickerson, R. E. and Drew, H. R. (1981) *J. Mol. Biol.* **149**, 761-786.

Dubendorff, J. W., deHaseth, P. L., Rosendal, M. R. and Caruthers, M. H. (1987) *J. Biol. Chem.* **262**, 892-898.

Eckstein, F., Ed. (1991) *Oligonucleotides and Anologues, A Practical Approach* Oxford, Oxford University Press.

Edge, M. D., Greene, A. R., Heathcliffe, G. R., Meacock, P. A., Schuch, W., Scanlon, D. B., Atkinson, T. C., Newton, C. R. and Markhan, A. F. (1981) *Nature* **292**, 756-762.

Efcavitch, J. W., McBride, L. J. and Eadie, J. S. (1987). *Biophosphate and Their Analogues--Synthesis, Structure, Metabolism and Activity* Amsterdam, Elsevier. 65-70.

Engels, J. W. and Uhlman, E. (1989) *Angew. Chem. Int. Ed. Engl.* 716-734.

Englisch, U. and Gauss, D. H. (1991) *Angew. Chem. Int. Ed. Engl.* 613-629.

Fairall, L., Martin, S. and Rhodes, D. (1989) *The EMBO Journal* **8**, 1809-1817.

Favre, A. (1990) *Bioorganic Photochemistry* New York, John Wiley and Sons. 379-425.

Ferentz, A. E. and Verdine, G. L. (1991) *J. Am. Chem. Soc.* **113**, 4000-4002.

Ferentz, A. E., Keating, T. A. and Verdine, G. L. (1993) *J. Am. Chem. Soc.* **115**, 9006-9014.

Fernandez-Forner, D., Palom, Y., Ikuta, S., Pederoso, E. and Eritja, R. (1990) *Nucleic Acids Research* **18**, 5729-5734.

Fliess, A.; Woltes, H.; Seela, F.; Pingoud, A. (1988) *Nucleic Acids Research* **16**, 11781-11793.

Fowler, K. W., Buchi, G. and Essigmann, J. M. (1982) *J. Am. Chem. Soc.* **104**, 1050-1054.

Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B-C., Greene, P., Boyer, H. W. and Rosenberg, J. M. (1984) *Nature* **309**, 327-331.

Friedman, S. T. and Ansari, N. (1992) *Nucleic Acids Research* **20**, 3241-3248.

Froehler, B. C., Ng, P. G. and Matteucci, M. D. (1986) *Nucleic Acids Research* **14**, 5399-5407.

Froehler, B. C., Wadwani, S., Terhorst, T. J. and Gerrard, S. R. (1992) *Tetrahedron Letters* **33**, 5307-5310.

Gaffney, B. L., Marky, L. A. and Jones, R. A. (1984a) *Tetrahedron* **40**, 3-13.

Gaffney, B. L., Marky, L. A. and Jones, R. A. (1984b) *Biochemistry* **23**, 5686-5691.

Gao, X. (1985) *Nucleic Acids Research* **13**, 573-584.

Gao, H., Fathi, R., Gaffney, B. L., Goswami, B., Kung, P.-P., Rhee, Y., Jin, R. and Jone, R. A. (1992) *J. Org. Chem.* **57**, 6954-6959.

Garegg, P. J., Regberg, T., Strawinski, J. and Stromberg, R. (1985) *Chemica Scripta* **25**, 280-282.

Georgiadis, P., Smith, C. A. and Swann, P. F. (1991) *Cancer Research* 5843-5850.

Gibson, K. J. and Benkovic, S. J. (1987) *Nucleic Acids Research* **15**, 6455-6467.

Gildea, B. and McLaughlin, L. W. (1989) *Nucleic Acids Research* **17**, 2261-2281.

Gilham, P. T. and Khorana, H. G. (1958) *J. Am. Chem. Soc.* **80**, 6212-6222.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A. D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 106-110.

Groger, G., Ramalho-Ortigao, F. and Seliger, H. (1988) *Nucleic Acids Research* **16**, 7763-7771.

Gryaznov, S. and Schultz, R. G. (1994) *Tetrahedron Letters* **35**, 2489-2492.

Hall, R. H., Todd, A. and Webb, R. F. (1957) *J. Chem. Soc.* 3291-3296.

Hanck, T., Schmidt, S. and Fritz, H.-J. (1993) *Nucleic Acids Research* **21**, 303-309.

Harrison, C. A., Turner, P. H. and Hinkle, D. C. (1982) *Nucleic Acids Research* **10**, 2399-2414.

Harrison, S. C. and Aggarwal, A. K. (1990) *Annu. Rev. Biochem.* **59**, 933-969.

Havron, A. and Sperling, J. (1977) *Biochemistry* **16**, 5631-5635.

Hicke, B. J., Willis, M. C., Koch, T. H. and Cech, T. R. (1994) *Biochemistry* **33**, 3364-3373.

Hochschild, A., Douhan III, J. and Ptashne, M. (1986) *Cell* **47**, 807-816.

Hochschild, A. and Ptashne, M. (1986) *Cell* **44**, 925-933.

Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R., Evertsz, E. M. and von Hippel, P. H. (1991) *Methods in Enzymology* **208**, 211-236.

Hsiang, M. W., Cole, R. D., Takeda, Y. T. and Echols, H. (1977) *Nature* **270**, 275-277.

Hutchinson, F. (1973) *Quarterly Reviews of Biophysics* **6**, 201-246.

Johnson, A. D., Pabo, C. O. and Sauer, R. T. (1980) *Bacteriophage λ repressor and Cro Protein: Reactions with Operator DNA* Academic Press. Inc.

Josepheson, S., Lagerholm, E. and Palm, G. (1984) *Acta Chemica Scandinavica B* **38**, 539-545.

Kalnik, M. W., Li, B. F. L., Swann, P. F. and Patel, D. J. (1989a) *Biochemistry* **28**, 6182-6192.

Kalnik, M. W., Li, B. F. L., Swann, P. F. and Patel, D. J. (1989b) *Biochemistry* **28**, 6170-6181.

Kellenbach, E. R., van den Elst, H., Boelens, R., van der Marel, G. A., van Boom, J. H. and Kaptein, R. (1991) *Rec. Trav. Chim. Pays-Bas* **110**, 387-388.

Kellenbach, E. R., Remerowski, M. L., Eib, D., Boelens, R., Van der Marel, G. A., van den Elst, H., van Boom, J. H. and Kaptein, R. (1992) *Nucleic Acids Research* **20**, 653-657.

Kennard, O. and Hunter, W. N. (1989) *Quarterly Reviews of Biophysics* **22**, 327-379.

Khorana, H. G. (1968) *Pure and Applied Chemistry* 17349-17381,

Kim, J. G., Takeda, Y., Matthews, B. W. and Anderson, W. F. (1987) *J. Mol. Biol.* **196**, 149-158.

Kirnos, M. D., Khudyakov, T. Y., Alexandrushkina, N. I. and Vanyyushin, B. F. (1977) *Nature* **270**, 369-370.

Klimasauakas, S., Kumar, S., Roberts, R. J. and Cheng, X. (1994) *Cell* **76**, 357-369.

Kong Thoo Lin, P and Brown, D. M. (1989) *Nucleic Acids Research* **17**, 10373-10383.

Koster, H. (1972) *Tetrahedron Letters* 1527-1530.

Kuimelis, R. G. and Nambiar, K. P. (1994) *Nucleic Acids Research* **22**, 1429-1436.

Kung, P.-P. and Jones, R. A. (1991) *Tetrahedron Lettters* **32**, 3919-3922.

Lamm, G. M., Blencowe, B. J., Sproat, B. S., Iribarren, A. M., Ryder, U. and Lamond, A. I. (1991) *Nucleic Acids Research* **19**, 3193-3198.

Leatherbarrow, R. J. and Fersht, A. R. (1986) *Protein Engineering* **1**, 7-16.

LePage, G. A. and Jones, M. (1961) *Cancer Research* **21**, 1590-1594.

Lesser, D. R., Kurpiewski, M. R. and Jen-Jacobson, L. (1990) *Science* **250**, 776-786.

Letsinger, R. L. and Mahadevan, V. (1966) *J. Am. Chem. Soc.* **88**, 5319-5324.

Letsinger, R. L. and Ogilvie, K. K. (1967) *J. Am. Chem. Soc.* **89**, 4801-4803.

Letsinger, R. L. and Ogilvie, K. K. (1969) *J. Am. Chem. Soc.* **91**, 3350-3355.

Letsinger, R. L., Finnan, J. L., Heavner, G. A. and Lunsford, W. B. (1975) *J. Am. Chem. Soc.* **97**, 3278-3279.

Letsinger, R. L. and Lunsford, W. B. (1976) *J. Am. Chem. Soc.* **98**, 3655-3661.

Li, B. F. L., Reese, C. B. and Swann, P. F. (1987) *Biochemistry* **26**, 1086-1093.

Li, B. F. L., Swann, P. F., Kalnik, M. W., Kouchakdjian, M. and Patel, D. J. (1988) *NMR Spectroscopy in Drug Research* Copenhagen, Munksgaard. 309-340.

Li, B. F. L. and Swann, P. F. (1989) *Biochemistry* **28**, 5779-5786.

Liu, J. and Verdine, G. L. (1992) *Tetrahedron Letters* **33**, 4265-4268.

Lohrmann, R., Soll, D., Hayatsu, H. and Ihtsuka, E. (1966) *J. Am. Chem. Soc.* **88**, 819-829.

MacMillan, A. M. and Verdine, G. L. (1991) *Tetrahedron* **47**, 2603-2616.

MacMillan, A. M., Chen, L. and Verdine, G. L. (1992) *J. Org. Chem.* **57**, 2989-2991.

MacMillan, A. M., Lee, R. J. and Verdine, G. L. (1993) *J. Am. Chem. Soc.* **115**, 4921-4922.

Marasco, C. J. and Sufrin, J. R. (1992) *J. Org. Chem.* **57**, 6363-6365.

Marshall, W. S. and caruthers, M. H. (1993) *Science* **259**, 1564-1570.

Matteucci, M. D. and Caruthers, M. H. (1980) *Tetrahedron Letters* **21**, 719-722.

Matteucci, M. D. and Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103**, 3185-3191.

Matteucci, M. D. and Webb, T. R. (1987) *Tetrahedron Letters* **28**, 2469-2472.

Maunter, H. G. (1956) *J. Am. Chem. Soc.* **78**, 5292-5294.

Mazzarelli, J., Scholtissek, S. and McLaughlin, L. W. (1989) *Biochemistry* **28**, 4616-4622.

Mazzarelli, J. M., Rajur, S. B., Iadarola, P. L. and Maclaughlin, L.W. (1992) *Biochemistry* **31**, 5925-5936.

McBride, L. J. and Caruthers, M. H. (1983) *Tetrahedron Letters* **24**, 245-248.

McLaughlin, L. W., Benseler, F., Graeser, E., Piel, N. and Scholtissek, S. (1987) *Biochemistry* **26**, 7238-7245.

McLaughlin, L. W., Leong, T., Benseler, F. and Piel, N. (1988) *Nucleic Acids Research* **16**, 5631-5644.

Merrifield, R. B. (1962) *Feb. Proc.* **21**, 412.

Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85**, 2149-2154.

Merrifield, R. B. (1965) *Science* **150**, 178-185.

Merrill, B. M., Williams, K. R., Chase, J. W. and Konigsberg, W. H. (1984) *J. Biol. Chem.* **259**, 10850-10856.

Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H. and Williams, K. R. (1988) *J. Biol. Chem.* **263**, 3307-3313.

Metzler, W. J. and Lu, P. (1989) *J. Mol. Biol.* **205**, 149-169.

Michelson, A. M. and Todd, A. R. (1955) *J. Chem. Soc.* 2632-2638.

Milligan, J. F., Matteucci, M. D. and Martin, J. C. (1993) *J. Med. Chem.* **36**, 1923-1937.

Milton, J., Connolly, B. A., Nikiforov, T. T. and Cosstick, R. (1993) *J. Chem. Soc. Chem. Commun.* 779-780.

Newman, P. C., Williams, D. M., Cosstick, R., Seela, F. and Connoly, B. A. (1990) *Biochemistry* **29**, 9902-9910.

Nikiforov, T. T. and Connlly, B. A. (1991) *Tetrahedron Letters* **32**, 3851-3854.

Nikiforov, T. T. and Connolly, B. A. (1992a) *Nucleic Acids Research* **20**, 1209-1214.

Nikiforov, T. T. and Connolly, B. A. (1992b) *Tetragedron Letters* **33**, 2379-2382.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. and Metthews, B. W. (1982) *Nature* **298**, 718-723.

Pabo, C. O. (1984) *Annu. Rev. Biochem.* **53**, 293-321.

Pabo, C. O. (1992) *Annu. Rev. Biochem.* **61**, 1053-1095.

Pashev, L. G., Dimitrov, S. I. and Angelov, D. (1991) *Trends Biochem. Sci.* **16**, 323-326.

Pendergrast, P. S., Chen, Y., Ebright, Y. W. and Ebright, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10287-10291.

Ptashne, M. (1986) *A genetic swich* New York, Blackwell.

Rajur, S. B. and McLaughlin, L. W. (1992) *Tetrahedron Letters* **33**, 6081-6084.

Rappaport, H. P. (1988) *Nucleic Acids Research* **16**, 7253-7267.

Reese, C. B. and Zard, L. (1981) *Nucleic Acids Research* **9**, 4611-4626.

Reese, C. B. and Skone, P. A. (1984) *J. Chem. Soc. Perkin Trans.* **1**, 1263-1271.

Rinke-Appel, J., Junke, N., Birmacombe, R., Dokudovskaya, S., Dontsova, O. and Bogdanov, A. (1993) *Nucleic Acids Research* **21**, 2853-2859.

Roberts, J. R. (1994) *Angew. Chem. Int. Ed. Engl.* **33**, 1222-1228.

Ruth, J. L. (1991) *Oligonucleotides and Analogues* Oxford, Oxford University Press, 255-282

Saenger, W. (1984) *Principles of Nucleic Acid Structure* New York, Springer-Verlag.
Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning* New York, Cold Spring Harbor Laboratory Press.

Schmidt, S., Pein, C.-D., Fritz, H.-J. and Cech, D. (1992) *Nucleic Acids Research* **20**, 2421-2426.

Schott, H. N. and Shetalar, M. D. (1974) *Biochem. Biophy. Res. Commun.* **59**, 1112-1116.

Schulhof, J. C., Molko, D. and Teoule, R. (1987a) *Nucleic Acids Research* **15**, 397-416.

Schulhof, J. C., Molko, D. and Teoule, R. (1987b) *Tetrahedron Letters* **28**, 51-54.

Seela, F. and Driller, H. (1985) *Nucleic Acids Research* **13**, 911-926.

Seela, F. and Kehne, A. (1985) *Biochemistry* **24**, 7556-7561.

Seela, F. and Drill, H. (1986) *Nucleic Acids Research* **14**, 2319-2332.

Seela, F. and Kehne, A. (1987) *Biochemistry* **26**, 2232-2238.

Shaltiel, S. (1967) *Biochem. Biophy. Res. Commu.* **29**, 178-183.

Shamoo, Y., Williams, K. R. and Konigsberg, W. H. (1988) *Proteins: Struct. Funct. Genet.* **4**, 1-6.

Singer, B. and Kasmierek, J. T. (1982) *Annu. Rev. Biochem.* **52**, 655-693.

Singer, B. (1986) *Cancer Research* **46**, 4879-4885.

Sinha, N. D., Biernat, J. and Koster, H. (1983) *Tetrahedron Letters* **24**, 5845-5864.

Sinha, N. D., Biernat, J. and Koster, H. (1984) *Nucleic Acids Research* **12**, 4539-4557.

Sinha, N., Davis, P., Usman, N., Perez, J., Hodge, R., Kremsky, J. and Casale, R. (1993) *Biochimie* **75**, 13-23.

Singh, D., Kumar, V. and Ganesh, K. N. (1990) *Nucleic Acids Research* **18,** 3339-3345.

Smith, M., Rammler, D. H., Goldberg, I. H. and Khorana, H. G. (1962) *J. Am. Chem. Soc.* **84**, 430-440.

Smith, C. A., Xu, Y.-Z. and Swann, P. F. (1990) *Carcinogenesis* **11**, 811-816.

Smith, S. S., Kaplan, B. E., Sowers, L. C. and Newman, E. M. (1992) *Proc. Natl. Acad. Sci. UAS* **89**, 4744-4748.

Sproat, B. S. and Gait, M. J. (1984) *Oligonucleotide Synthesis* Oxford, Washington DC, IRL Press. 83-116.

Sproat, B. S., Irabarren, A. M., Garcia, R. G. and Beijer, B. (1991) *Nucleic Acids Research* **19**, 733-738.

Srivastava, S. C., Raza, S. K. and Misra, R. (1994) *Nucleic Acids Research* **22**, 1296-1304.

Sudhakar, T., Jayaraman, K. and Durland, R. H. (1992) *Tetrahedron Letters* **33**, 7651-7654.

Sung, W. L. (1981a) *J. Chem. Soc. Chem. Commu.* 1089.

Sung, W. L. (1981b) *Nucleic Acids Research* **9**, 6139-6151.

Swann, P. F. (1990) *Mut. Res.* **233**, 81-94.

Sylvers, L. A. and Wower, J. (1993) *Bioconjugate Chemistry* **4**, 411-418.

Takeda, Y., Kim, J. G., Caday, K. D., Steers, E., Ohlendorf, D. H., Anderson, W.

F. and Mattews, B. W. (1986) *J. Biol. Chem.* **261**, 8608-8616.

Takeda, Y., Sarai, A. and Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 439-443.

Takeda, Y., Ross, P. D. and Mudd, C. P. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8180-8184.

Tan, H.-B., Swann, P. F. and Chance, E. M. (1994) *Biochemistry* **33**, 5335-5346.

Uhlmann, E. and Peyman, A. (1990) *Chem. Rev.* **90**, 544-585.

Varma, R. S. (1993) *SYNLETT* 621-637.

Verdine, G. L. (1994) *Cell* **76**, 197-200.

Vu, H., McCollum, C., Jacobson, K., Theisen, P., Vinayak, R., Spiess, E. and Andrus, A. (1990) *Tetrahedron Letters* **31**, 7269-7272.

Wagner, R. W., Matteucci, M. D., Lewis, J. G., Gutierrez, A. J., Moulds, C. and Froehler, B. C. (1993) *Science* **260**, 1510-1513.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawfoed, J., L,, van Boom, J. H., Van der Marel, G. and Rich, A. (1979) *Nature* **282**, 680-686.

Wang, A. H.-J., Quigley, G. J. and Kolpak, F. J. (1981) *Science* **211**, 171-176.

Waters, T. R. and Connolly, B. A. (1992) *Nucleosides & Nucleotides* **11**, 985-998.

Waters, T. R. and Connolly, B. A. (1994) *Biochemistry* **33**, 1812-1819.

Watson, J. D. and Crick, E. H. C. (1953) *Nature* **171**, 737-738.

Webb, T. R. and Matteucci, M. D. (1986a) *Nucleic Acids Research* **14**, 7661-7674.

Webb, T. R. and Matteucci, M. D. (1986b) *J. Am. Chem. Soc.* **108**, 2764-2765.

Wick, K. L. and Matthews, K. S. (1991) *J. Biol. Chem.* **266**, 6106-6112.

Williams, K. R. and Konigsberg, W. H. (1991) *Methods in Enzymology* **208**, 516-539.

Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R. and Koch, T. H. (1993) *Science* **262**, 1255-1257.

Wolfes, H., Fliess, A., Winkler, F. and Pingoud, A. (1986) *Eur. J. Biochem.* **159**, 267-273.

Xu, Y.-Z. and Swann, P. F. (1990) *Nucleic Acids Research* **18**, 4061-4065.

Xu, Y.-Z., Zheng, Q. and Swann, P. F. (1991) *Tetrahedron Letters* **32**, 2817-2820.

Xu, Y.-Z., Zheng, Q. and Swann, P. F. (1992a) *Tetrahedron Letters* **33**, 5837-5840.

Xu, Y.-Z., Zheng, Q. and Swann, P. F. (1992b) *Tetrahedron* **48**, 1729-1740.

Xu, Y.-Z. and Swann, P. F. (1992) *Analytical Biochemistry* **204**, 185-189.

Xu, Y.-Z. and Swann, P. F. (1994) *Tetrahedron Letters* **35**, 303-306.

Zamecnik, P. C. and Stephenson, M. L. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 280-284.

Tetrahedron Letters, Vol.32, No.24, pp 2817-2820, 1991
Printed in Great Britain

0040-4039/91 $3.00 + .00
Pergamon Press plc

Simple Synthesis of 4-Thiothymidine, 4-Thiouridine and 6-Thio-2'-deoxyguanosine

Yao-Zhong Xu*, Qinguo Zheng and Peter F. Swann

CRC Nitrosamine-Induced Cancer Research Group
Department of Biochemistry and Molecular Biology, University College London
Gower Street, London WC1E 6BT, England

key words: thionucleosides; 4-thiothymidine; 4-thiouridine; 6-thio-2'-deoxyguanosine; thiation

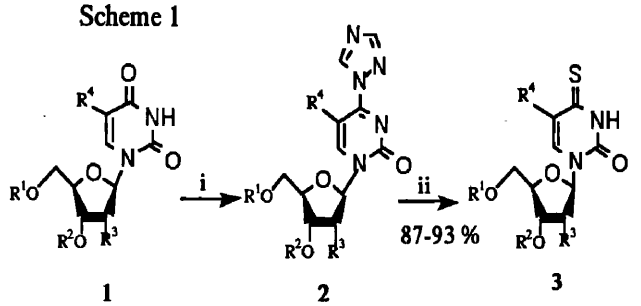
Abstract: *4-triazolo-pyrimidine nucleosides and 6-O-(mesitylenesulfonyl)-2'-deoxyguanosine, when treated with thiolacetic acid at room temperature, gave the corresponding 4-thiopyrimidine nucleosides and 6-thio-2'-deoxyguanosine with high yields (86-93%). Possible mechanisms are discussed.*

4-Thiothymidine and 4-thiouridine can be prepared by treatment of the corresponding protected pyrimidine nucleosides with phosphorus pentasulphide[1] or Lawesson's reagent [2] in appropriate solvents at reflux, but unfortunately the methods are not suitable for acid-labile and temperature-sensitive compounds. For example under these conditions the trityl group will be removed from trityl-protected nucleosides and 2'-deoxyguanosine will depurinate. For this reason, 6-thio-2'-deoxyguanosine is usually made by condensing protected 2'-deoxyribofuranosyl chloride with protected 6-chloroguanine to give a mixture of the α and β-anomers of 6-chloro-2'-deoxyguanosine. The desired β-anomer is isolated by fractional crystallization and treated with hydrogen sulfide (H_2S), or its sodium salt, to give 6-thio-2'-deoxyguanosine[3,4,5]. The method is tedious and difficult to perform, and the overall yield is poor. In the course of synthesis of oligodeoxyribonucleotides containing 4-thiothymine and 6-thioguanine[6], we have found an easy synthesis of 4-thiothymidine and 6-thio-2'-deoxguanosine and of the ribonucleoside, 4-thiouridine, in very mild conditions with high yields.

4-THIOPYRIMIDINE NUCLEOSIDES:

Protected nucleosides, with silyl, DMT (dimethoxytrityl) and acyl protecting groups on the sugar moiety (**1a-c**), were converted into their 4-triazolo-derivatives by reaction with $POCl_3$ and 1,2,4-triazole in quantitative yield[7]. Reaction of these with thiolacetic acid at room temperature gave the protected 4-thio-nucleosides in very

Scheme 1



	R^1	R^2	R^3	R^4
a.	Si ;	Si;	H;	CH_3
b.	DMT;	Si;	H;	CH_3
c.	Bz;	Bz;	BzO;	H

Si: t-butyldimethylsilyl;
DMT: dimethoxytrityl;
Bz: benzoyl;

i: $POCl_3$ / triethylamine / triazole; ii CH_3COSH;

high yields (87-93%) (Scheme 1). The thio-derivatives have been isolated and their structures confirmed by NMR, U.V. (in particular the λmax at 326 - 335nm showing the presence of sulphur) and by comparison with published data. The rate of the reaction between the 4-triazolides and thiolacetic acid increased when the acidity of the medium was increased, and reduced with basicity of the medium. The reaction took place much faster in protic solvents (e.g. CH_3OH) than in aprotic solvent (e.g. CH_3CN). Unexpectedly none of the products had the acetyl group on the 4-S position[8]. When the reaction course was followed by TLC, only 4-thiopyrimidine and no 4-*S*-acetyl-pyrimidine nucleoside was observed as an intermediate, and the same product was obtained using thiolbenzoic acid instead of thiolacetic acid. Although the mechanism has not yet been clarified, these observations suggest that protonation of the triazolo group was followed by an attack at C-4 by thio-acetate ion. Possibly the acetyl group is lost because the thio-keto form is more stable than the 4-*S*-acetyl form. However cytidine, treated with thiolacetic acid in pyridine at 50°C for 4 hrs, quantitatively gave *N*-acetylcytidine[9] rather than 4-thiouridine. A possible explanation of this observation, which offers an alternative method for selective acylation of the 4-amino group of cytidine[10], is that the amino group of cytidine is a better nucleophile and less good leaving group than the triazolo group.

It is worth noting that the very acid-labile DMT protecting group (in compounds **1-3b**), which is commonly removed by treatment with 80% acetic acid at 20°C for 20 min, is not removed by the treatment with thiolacetic acid in CH_3CN or CH_2Cl_2 even though thiolacetic acid is a stronger acid (pKa = 3.62) than acetic acid (pKa =4.76)[11]. This suggests that thiolacetic acid plays a nucleophilic rather than acidic role in these aprotic solvents[12]. The stability of the DMT protecting group to these conditions is important because it is widely used in chemical synthesis of DNA or RNA oligomers.

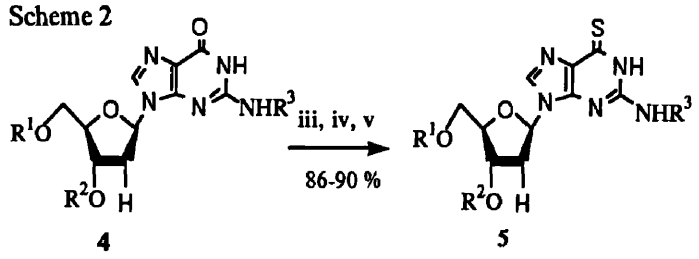
A typical thiation procedure was as follows: to a stirring solution of the protected 4-triazolo pyrimidine nucleosides **2a-c** (1 mmole) in 20 ml of CH_3CN was added 1 ml of thiolacetic acid at room temperature. The reaction was left stirring overnight and the starting material was completely converted into a new spot with higher R_f by TLC [in 5 % CH_3OH:$CHCl_3$, with visualization under UV lamp, or by treatment with anisaldehyde / conc. H_2SO_4 / ethanol (5:5:90, v/v/v) and heating]. The reaction solution was diluted with CH_2Cl_2 (50 ml), washed with saturated aqueous $NaHCO_3$ (2 x 50 ml), then with saturated aqueous NaCl (50 ml). The organic layer was dried (Na_2SO_4) and evaporated under reduced pressure to give a slightly-yellow solid (one spot on TLC). Pure compounds (**3a-c**) were obtained by silica gel column chromatography and characterized by NMR and UV[13].

6-THIO-2'-DEOXYGUANOSINE:

Protected 2'-deoxyguanosines, with acyl and silyl protecting groups on the 5', 3' and N-2 position (**4d-e**), were treated with mesitylenesulphonyl chloride to form 6-*O*-mesitylenesulphonyl derivatives, then with *N*-methylpyrrolidine[7,14], followed by thiolacetic acid. This gave the protected 6-thio-2'-deoxyguanosines (**5d-e**) in high yields (86-90%) (Scheme 2). In this case also the product was the 6-thio-keto rather than the 6-*S*-acetyl derivative[8]. Presumeably the 6-thio-keto form is more stable than the 6-*S*-acetyl form. By deprotection, compound **5d** has been quantitatively converted into free 6-thio-2'-deoxyguanosine, a potential drug[15].

A typical procedure was as follows: The nucleosides were protected by published procedures. 1 ml of triethylamine, 20 mg of dimethylaminopyridine and 1.0 g of mesitylenesulphonyl chloride were sequentially

Scheme 2



4 5

	R^1	R^2	R^3
d.	i-Bu;	i-Bu;	i-Bu;
e.	Si;	Si;	PAc;

i-Bu: iso-butyryl;
Si: t-butyldimethylsilyl;
PAc: phenylacetyl

iii: mesitylenesulphonyl chloride; iv: *N*-methyl pyrrolidine v: CH_3COSH

added into the stirring CH_2Cl_2 (50 ml) solution of the protected nucleosides **4d-e** (2 mmole) at room temperature. Stirring was continued until all the nucleoside had become a new compound with higher R_f (TLC). Then the solution was cooled in an ice-bath and a mixture of *N*-methylpyrrolidine (2 ml) and CH_2Cl_2 (5 ml) was added dropwise. After TLC showed that the material had changed into a compound with very low R_f (5% CH_3OH:$CHCl_3$), a mixture of thiolacetic acid (2 ml) and CH_2Cl_2 (5 ml) was added dropwise. After 30 min, when TLC showed that the low R_f compound had completely disappeared, the reaction solution was poured into 50 ml of 0.5 M aqueous KH_2PO_4 (pH 6.3). The organic layer was washed with saturated aqueous $NaHCO_3$ (2 x 50 ml), then with saturated aqueous NaCl (50 ml), dried (Na_2SO_4), then evaporated under reduced pressure to give an oily residue which was precipitated from toluene into n-pentane to give a slightly-yellow powder. Pure compounds **5d-e** were obtained by silica gel column chromatography and characterized by NMR and UV[16].

The present method for transformation of nucleosides to thio-nucleosides has at least two advantages over phosphorus pentasulphide or Lawesson's reagent: the reactions occur at room temperature with high yields; and after reaction the reagent, thiolacetic acid, is converted into the acetate ion which is easy to remove. The method could be used for direct incorporation of radioactive ^{35}S from readily available ^{35}S-thiolacetic acid into nucleosides, nucleotides and DNA-oligomers and the method might be suitable for thiation of other temperature- and acid-sensitive heterocyclic compounds.

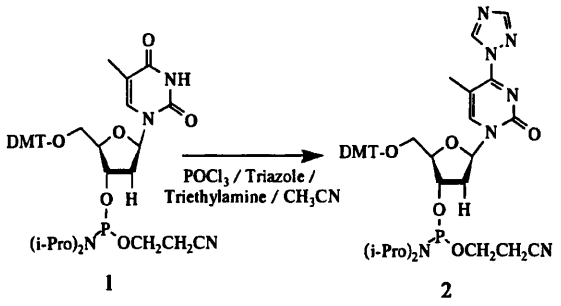
ACKNOWLEDGEMENTS:

We wish to thank the Cancer Research Campaign for most generous support of this work.

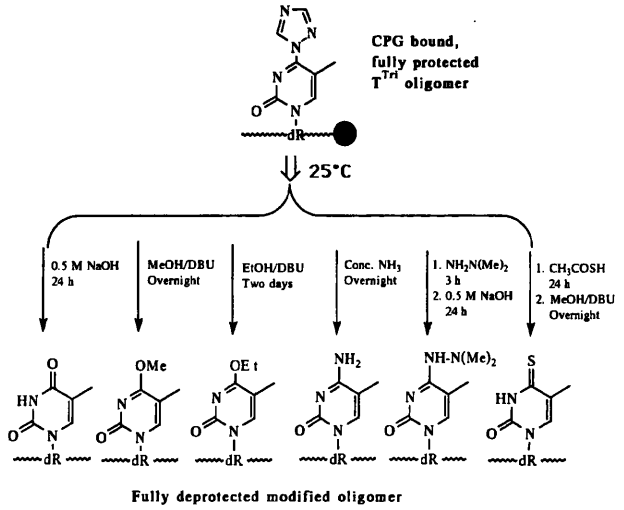
REFERENCES AND NOTES:

1. Fox, J.J.; Van Pragg, D.; Wempen, I.; Doerr, I.L.; Cheong, L.; Knoll, J.E.; Eidinoff, M.L.; Bendich, A. and Brown, G.B. *J. Am. Chem. Soc.* **1959**, *81*, 178-187; Falco, E.A.; Otter, B.A. and Fox, J.J. *J. Org. Chem.* **1970**, *35*, 2326-2330.
2. Connolly, B.A. and Newman, P.C. *Nucleic Acids Research* **1989**, *7*, 4957-4974.
3. Rappaport, H.P. *Nucleic Acids Research* **1988**, *6*,7253-7267.
4. Iwamoto, R.H.; Acton, E.M. and Goodman, L. *J. Med. Chem.* **1963**, *6*, 684-688.
5. Townsend, L.B. in *Nucleoside Analogues --- Chemistry, Biology, and Medical Applications*; ed by Walker, R.T.; Clercq, E.D. and Eckstein, F. Plenum Press, New York and London **1979**, pp.193-223. and references therein
6. Manuscript in preparation
7. Reese, C.B. and Skone, P.A. *J. Chem. Soc. Perkin Trans.* **1984**, *1*, 1263-1271; Xu, Y.-Z. and Swann, P.F. *Nucleic Acids Research* **1990**, *8*, 4061-4065.

Scheme I



Scheme II



phosphoramidite monomers to be used in oligonucleotide synthesis,[7] and very recently we have found that it can be easily converted into 4-thiothymidine at room temperature with thiolacetic acid.[8] Two previous papers have also reported the preparation of oligomers containing 4-triazolothymine with the subsequent transformation of these to oligomers with a new 4-modified base.[9,10] Webb and Matteucci[9] used it to synthesize oligomers containing 5-methyl-N^4,N^4-ethanocytosine. The acid-labile ethano group was successfully introduced after removal of the 4,4'-dimethoxytriphenylmethyl (DMT) group protecting the 5'-OH, but there was then great difficulty in removing the protecting groups from the bases. The 9-fluorenylmethoxycarbonyl (Fmoc) group was successfully removed from adenine and cytosine, but the Fmoc group was found unsuitable for guanine and it was impossible to make oligomers containing both guanine and the ethanocytosine. Similarly Fernandez-Forner et al.[10] have used it to make oligomers containing O^4-ethylthymine, but again had considerable difficulty deprotecting the final product. The experience of these workers indicates that for the strategy to be successful one has to match the deprotection procedure to the nature of the base which has been introduced, and furthermore one has to have an analytical and preparative technique which will allow the separation of the desired product from the impurities. In this regard we used the labile protecting groups,[11] i.e., the phenoxyacetyl group on adenine and guanine and the isobutyryl group on cytosine, together with ion-exchange chromatography[12] at pH 12 for separation and analysis of the desired products.

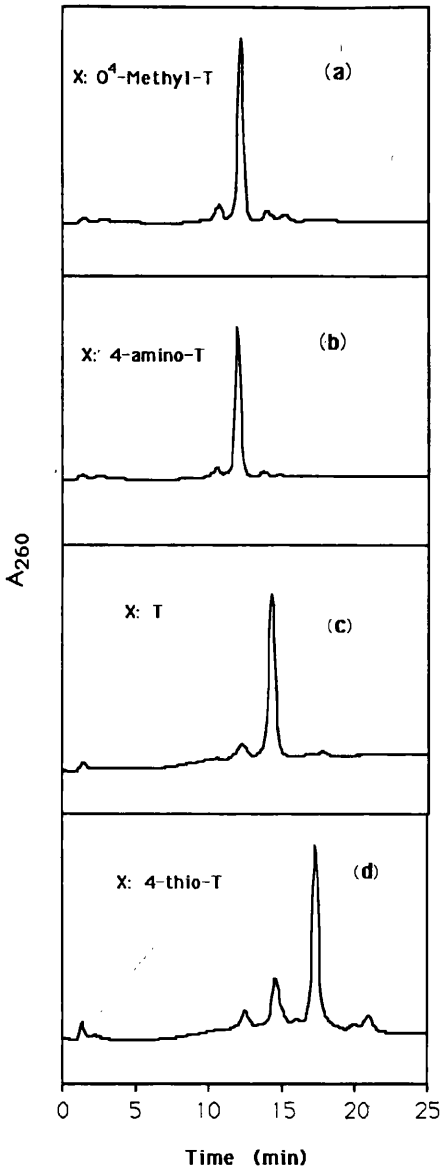


Figure 1. FPLC profiles of crude modified oligomers AGCGAAXTCGCT (X: T^{OMe}, T^{NH2}, T, and T^{S}). The oligomers were chromatographed on a Mono Q column using 0.4 M NaCl, 0.01 M NaOH for 2 min, and then 1.2 M NaCl, 0.01 M NaOH increasing to 15% over 3 min, and then to 40% over the following 20 min at a flow rate of 1 mL/min.

Results and Discussion

4-Triazolothymine phosphoramidite monomer (compound 2) was prepared from commercially available compound 1 (Scheme I) in high yield (97%) as described earlier.[9] The purity assessed by HPTLC and ^{31}P NMR was >95%. The modified monomer was incorporated into the DNA-oligomer with a coupling yield >98%.

As a model, a self-complementary dodecamer AGCGAAXTCGCT was assembled (X standing for 4-triazolothymine), and then the protected oligomer still attached to the controlled-pore-glass (CPG) support was treated at 25 °C with different reagents to substitute 4-triazolothymine in the oligomer to cleave the oligomer from the CPG support and to remove the protecting groups from

(7) Xu, Y.-Z.; Swann, P. F. *Nucleic Acids Res.* **1990**, *18*, 4061–4065.
(8) Xu, Y.-Z.; Zheng, Q.; Swann, P. F. *Tetrahedron Lett.* **1991**, *32*, 2817–2820.
(9) Webb, T. R.; Matteucci, M. K. *Nucleic Acids Res.* **1986**, *14*, 7661–7674.
(10) Fernandez-Forner, D.; Palom, Y.; Ikuta, S.; Pedroso, E.; Eritja, R. *Nucleic Acids Res.* **1990**, *18*, 5729–5734.
(11) Schulhof, J. C.; Molko, D.; Teoule, R. *Tetrahedron Lett.* **1987**, *28*, 51–54.
(12) Xu, Y.-Z.; Swann, P. F. *Anal. Biochem.*, in press.

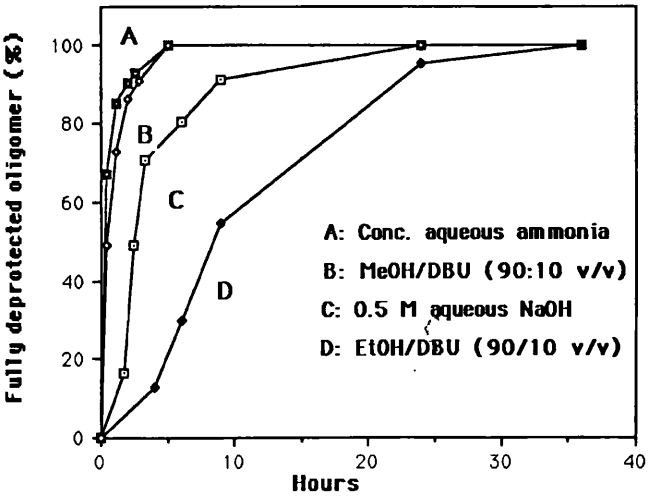


Figure 2. Time courses of deprotection of an oligomer AGCGAAXTCGCT (X: T^{Tri}) protected with PAC groups. The CPG support bearing the protected T^{Tri} 12 mer, with the 5'-DMT removed, was treated at 25 °C with concd ammonia, MeOH/DBU, EtOH/DBU, or 0.5 M aqueous NaOH, respectively. The course of the deprotection was followed by FPLC on a mono Q column.

the oligomer, thus to produce modified oligomers (Scheme II). Although we were not able to check the stability of compound **2** during oligonucleotide synthesis, comparison of the chromatographic profile of the 12 mer containing O^4-methylthymine (Figure 1a) or 5-methylcytosine (Figure 1b) with that of the oligomer containing thymine (Figure 1c) (all of these oligomers were produced by postsynthetic substitution) shows that the 4-triazolothymine residue is sufficiently stable during the synthesis of the oligomer.

The conversion of the 4-triazolothymine (T^{Tri}) oligomer into the modified base oligomer involves two major reactions: substitution and deprotection. Experiments were done to follow the progress of these reactions and to discover the optimum conditions for the subsequent synthesis of oligomers.

a. O^4-Methylthymine. The fully protected T^{Tri} 12 mer was exposed to MeOH/DBU at 25 °C. The cleavage from the CPG support and the deprotection are complete within 5 h (Figure 2B), and the substitution was finished within 12 h. Therefore, overnight MeOH/DBU treatment was generally employed in the production of the methylated oligomer. This gave essentially pure T^{OMe} oligomer (Figure 1a). Base analysis of the main product confirmed the correct composition.

b. O^4-Ethylthymine. Because ethoxide is less nucleophilic than methoxide, a more prolonged exposure to EtOH/DBU at 25 °C is needed. A lipophilic counterion, cetyltrimethylammonium, was added to the reaction mixture to keep the partially deprotected oligomer in solution.[7] Less than 10% of the T^{Tri} residue was left unsubstituted after 12 h, and the substitution was finished within 24 h. All protecting groups were cleaved within 36 h (Figure 2D), so 2 days of EtOH/DBU treatment was generally employed. An essentially pure oligomer was easily obtained, and the correct composition was confirmed by base analysis. It is worth mentioning that most commercial anhydrous ethanol contains quite a high percentage of methanol. This must be avoided otherwise methylated oligomers will be produced to contaminate ethylated products.

We have previously synthesized oligomers containing T^{OMe} or T^{OEt} by incorporation of either T^{OMe} or T^{OEt} monomer[7] but the present method can provide T^{OMe} and T^{OEt} oligomers in a single synthesis without the need to prepare T^{OMe} and T^{OEt} monomers. Comparison of our previous results,[7] and the results reported now, with those of Fernandez-Forner et al.[10] shows the advantage of the labile base-protecting groups used here. Fernandez-Forner et al. showed that the conventional base protecting groups (benzoyl on adenine and cytosine and isobutyryl on guanine) are unsuitable for the synthesis of oligomers containing O^4-ethylthymine, and although they employed a new protecting group (*p*-nitrophenylethyl) for this purpose it also was very difficult to remove.

Table I. Yield and Purity of Synthetic Oligomers[a]

oligomer	scale	yield (OD)	purity (%)
AGCGAATTCGCT	$^1/_5 \times 1$ μmol	12.1	>90
AGCGAATTCGCT	0.2 μmol	14.0	>90
---AT^{OMe}---	$^1/_5 \times 1$ μmol	12.4	>90
---AT^{OEt}---	$^1/_5 \times 1$ μmol	10.6	>90
---AT^{NH2}---	$^1/_5 \times 1$ μmol	12.4	>90
---AT^{NH2}---	0.2 μmol	7.0	>90
---AT^{DH}---	$^1/_5 \times 1$ μmol	8.0	60
---AT^{S}---	$^1/_5 \times 1$ μmol	9.3	70

[a] The full sequence is shown for the parent dodecamer. Only the central dimer containing the modified thymine is shown for the modified oligomers. Yield refers to the amount of oligomer (A_{260} units) recovered from the Nensorb cartridge. Purity is the purity, estimated from the FPLC profile, of the oligomers recovered from the Nensorb cartridge. The $^1/_5 \times 1$ μmol scale syntheses were carried out by substitution of $^1/_5$th of the contents of a 1 μmol cartridge containing a synthetic 4-triazolothymine oligomer. The 0.2 μmol scale syntheses were done for comparison, and in these only commercially available monomers were used.

c. Thymine. Treatment of the T^{Tri} 12 mer with 0.5 M aqueous NaOH at 25 °C for 24 h was sufficient for cleavage of the oligomer from the CPG support, for substitution of the triazolo group, and for deprotection of the other bases (Figure 2C). Base analysis of the main product (Figure 1c) gave only four unmodified bases (dA, dG, dC, and T), and a more prolonged exposure (3 days) to 0.5 M NaOH caused no obvious destruction of the oligomer. Only deprotected thymine-containing oligomer and no deprotected 4-triazolothymine-containing oligomer was formed during the course of the reaction. Although oligomers containing thymine can be easily prepared from thymine monomer, this substitution allows a reference sample containing thymine to be obtained during the same synthesis as the modified oligomer and would also allow ^{17}O to be introduced from ^{17}O-water for NMR studies.

d. 5-Methylcytosine. The T^{Tri} 12 mer was treated with concd aqueous ammonia (overnight, 25 °C), which cleaved the oligomer from the CPG support, substituted the 4-triazolo group of the base, removed the protecting groups, and gave pure 5-methylcytosine-containing oligomer (Figure 1b). The correct composition of the oligomer was confirmed by base analysis. Deprotection is complete within 5 h (Figure 2A), and the longer treatment (overnight) was adopted simply because it is routinely used for deprotecting unmodified oligomers. Although 5-methylcytosine monomer is commercially available, the present procedure can provide 5-methylcytosine-containing oligomer as well as the parent containing thymine in a single synthesis. Furthermore, it is possible to introduce NMR-sensitive ^{15}N from easily available ^{15}N-ammonia. Comparison of the oligomer containing 5-methylcytosine made from the T^{Tri} oligomer with the same sequence made from commercial 5-methylcytosine monomer shows that the purity was similar, but the yield from the T^{Tri} oligomer was slightly better (Table I).

e. N^4-(Dimethylamino)-5-methylcytosine [4-(2,2-Dimethylhydrazino)-5-methylpyrimid-2-one]. This was chosen as an example of the application of this strategy to the synthesis of oligomers containing labile amino de-

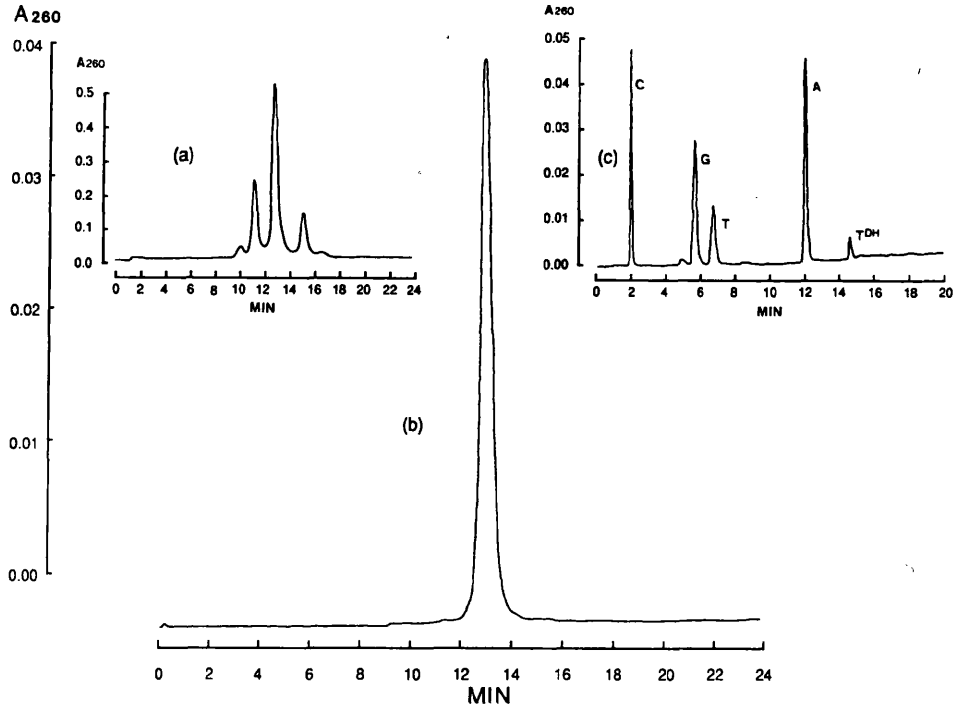


Figure 3. FPLC profiles of crude (a) and purified (b) AGCGAAT^{DH}TCGCT and HPLC base analysis (c) of the purified oligomer. FPLC was performed on a Mono Q column using 0.4 M NaCl, 0.01 M NaOH for 2 min, then 0.8 M NaCl, 0.01 M NaOH increasing to 30% over 3 min, then to 70% over the following 20 min at a flow rate of 1 mL/min. HPLC was performed on an 8MBC18 10-μm reversed-phase column using 0.05 M KH_2PO_4 (pH 4.5) and 0.05 M KH_2PO_4 (pH 4.5) containing 33% CH_3CN at a flow rate of 3 mL/min.

rivatives in the 4-position of pyrimidines. To our best knowledge, no paper has previously been published describing the chemical synthesis of an oligonucleotide containing such a labile hydrazino derivative.

The T^{Tri} 12 mer was treated with 1,1-dimethylhydrazine/CH_3CN for 3 h at 25 °C to substitute the triazolo group. Because the substituted 12 mer was still attached to the CPG support, the excess 1,1-dimethylhydrazine could be washed off with CH_3CN, and then the oligomer was cleaved from the CPG support and deblocked with 0.5 M aqueous NaOH for 24 h at 25 °C. This gave the desired oligomer (t_R = 13 min) (Figure 3a and b). The yield was slightly less than those of the other oligomers (Table I). One possible explanation is that during the substitution step the dimethylhydrazine, as a weak base, might slightly cleave the oligomer from the CPG support and this cleaved oligomer was washed off before the deprotection step. Because this is the first report of chemically synthesized oligomer containing a hydrazino derivative, the HPLC profile of the nucleoside digest is also presented (Figure 3c). As well as the four peaks corresponding to the four unmodified nucleosides (dA, dG, dC, and T) there was one peak which coeluted with authentic 1-(2′-deoxyribofuranosyl)-4-(2,2-dimethylhydrazino)-5-methylpyrimid-2-one. Aqueous NaOH was employed rather than concd NH_3 for deblocking because NaOH converts any remaining 4-triazolothymine into thymine, and the thymine-containing oligomer (t_R = 15 min) can be separated easily from the desired modified oligomer (t_R = 13 min) by FPLC (Figure 3b). From the results described above, we believe that other amino derivatives could be introduced in a similar way.

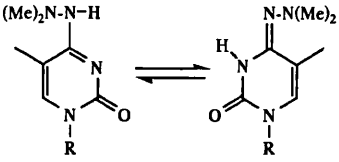
f. 4-Thiothymine. The T^{Tri} 12 mer was treated with CH_3COSH/CH_3CN at 25 °C, which replaced the triazolo group with an SH group.[8] Interestingly, this treatment does not remove the 5′-DMT group, which is particularly useful for later separation of failure sequences from desired oligomer. Thiation was finished within 12 h, but overnight treatment with CH_3COSH/CH_3CN was generally used. As the oligomer was still attached to the CPG support, the excess CH_3COSH was washed off with CH_3CN. For deprotection, either MeOH/DBU or concd aqueous NH_3 can be used. In agreement with a previous report[2] we observed that the 4-thiothymine in the dodecamer was converted to 5-methylcytosine by prolonged exposure to concd aqueous NH_3. Therefore, MeOH/DBU (16 h, 25 °C) is preferred (Figure 1d). This conversion, at the oligomer level, of thiothymine into other modified thymine derivatives by nucleophiles, like ammonia, may be useful for preparing olgomers containing other 4-modified pyrimidines, for example, those containing an aziridine group. The present method offers an easy way to produce T^S oligomers with a good yield and without the need to prepare a 4-thiothymine monomer.

Comparison of these results with the two previous reports[2,13] shows the advantages of the postsynthetic substitution strategy. Originally, T^S oligomers were prepared by incorporation of a 4-thiothymine phosphoramidite monomer in which the sulfur atom was protected with a methylsulfenyl (-SCH_3) group.[2] Unfortunately, this monomer was not very stable toward some of the reagents used in oligonucleotide synthesis and there was only a moderate yield (10–15%) of T^S oligomer. For this reason the same authors very recently used the *p*-nitrophenyl group to protect the sulfur.[13]

For purification, the resulting T^{NH2}, T^{DH}, or T oligomers with DMT groups still remaining on their 5′ ends were passed through a Nensorb Prep cartridge to remove the failure sequences which have no DMT groups on their 5′ ends, and the resulting T^{OR} or T^S oligomers with DMT groups still remaining on their 5′ ends were passed through Dowex ion-exchange column to remove DBU, followed by Nensorb Prep cartridge to remove the failure sequences. These steps result in reasonably pure products (cf. Figures 1 and 3) with high yields (Table I). More highly purified

(13) Nikiforov, T. T.; Connolly, B. A. *Tetrahedron Lett.* **1991**, *32*, 3851–3854.

Scheme III



Amino form Imino form

oligomers could be obtained by FPLC. Under basic conditions, the oligomers containing O^4-alkylthymine (Figure 1a), 5-methylcytosine (Figure 1b), or N^4-substituted 5-methylcytosine (Figure 3a) were eluted earlier than their parent oligomer containing thymine (Figure 1c) because they have one less negative charge due to lack of an imino proton at the 3-position.[12] Because 4-thiothymine has a lower pK_a than thymine,[14] the T^S oligomer (Figure 1d) was eluted later and easily separated from the parent containing thymine. The nucleoside O^4-methylthymidine is very slowly converted to thymidine in alkaline solution,[7] but the O^4-alkylthymine residue in the oligomer is more resistant to destruction by alkali, perhaps because of steric hindrance by the neighboring nucleotides. This allows one to use FPLC under basic conditions for the purifications. After purification followed by immediate neutralization, the purified oligomer was checked again by FPLC and still gave a single peak.

By combination of the direct base substitution at the oligomer level and separation by FPLC under basic conditions, we are able to provide a general and easy method for synthesis and purification of oligodeoxyribonucleotides containing different modified thymines avoiding the tedious steps of preparation of modified monomers. The method also provides potential for preparation of oligomers containing modified bases which are not stable during the assembly steps and for easily introducing atoms such as radioactive ^{35}S from ^{35}S-thiolacetic acid or NMR sensitive atoms such as ^{13}C from ^{13}C-methanol, ^{15}N from ^{15}N-ammonia, and ^{17}O from ^{17}O-water.

Base-Pairing Properties of Oligomers Containing N^4-(Dimethylamino)-5-methylcytosine [4-(2,2-Dimethylhydrazino)-5-methylpyrimidin-2-one] (T^{DH}). T^{DH} was made because it was expected to exist in about equimolar proportions of the amino and imino form (Scheme III) and thus might base-pair equally with adenine or guanine. This expectation was based on the observation that in the series cytosine, N^4-aminocytosine, and N^4-hydroxy(or N^4-methoxy)cytosine the ratio of amino to imino form is 10^4, 10–0.1 (depending on solvent), and 0.1–0.03, respectively, presumably reflecting increasing electronegativity in the series H, NH_2, and OH (or OCH_3).[15-17] Little is known in this respect about DNA containing N^4-aminocytosine[18] or N^4-(alkylamino)cytosine because the necessary oligomers have not been available. For this reason we developed the method for preparation of T^{DH} oligomers. However, melting temperature studies on a self-complementary 12-mer AGCGAAT^{DH}TCGCT in 0.1 M Hepes pH 7.5, 0.2 M NaCl and 0.02 M $MgCl_2$ solution, did not show a distinct transition temperature indicating that the self-complementary oligomer cannot form double-stranded DNA, or even a hairpin structure, under these experimental conditions. Non-self-complementary duplexes containing T^{DH} paired to guanine or to adenine had sharp transition temperatures like the controls containing G:C or A:T pairs, but in both cases the Tm was 17–19 °C lower than the control (Table II). The depression of Tm was nearly equal to the effect of mismatch base-pairing (Table II). This indicates that T^{DH} is not a good substitute for either thymine or cytosine. Possibly this is a consequence of steric hindrance by the dimethylamino group.

Table II. Tm Values of DNA Duplexes Containing the Modified Base

5' CAG GAA TXC GC 3'
3' GTC CTT AYG CG 5'

X	Y	Tm
T	A	58.7
T^{DH}	A	42
T^{DH}	G	40.6
C	G	59.2
C	A	42

In order to further evaluate the base-pairing properties of T^{DH} a synthetic 20 mer CGCTCTTA-CAT^{DH}GTATCGGAT, was used as template for DNA synthesis by the Klenow fragment of E. coli DNA polymerase I. The ^{32}P labeled complementary strand ATCCGATAC, as primer, was annealed to the template, and Klenow fragment with dATP or dGTP added. Preliminary experiments show that T^{DH} does not block DNA synthesis but the elongation of the primer with both G and A appears to proceed at a rate similar to that previously reported[19] for the mismatch incorporation of G opposite T in the template strand. These results are consistent with the Tm measurement and confirm that T^{DH} does not replace either C or T.

Experimental Section

Chemicals and Enzymes. The CPG-linked monomers and the chemicals used on the synthesizer were obtained from Cruachem (Glasgow, Scotland) and the monomers of (2-cyanoethyl)phosphoramidites protected with phenoxyacetyl on the amino functions of adenine and guanine and with isobutyryl on the amino function of cytosine ("PAC amidites") were from Pharmacia; 5-methylcytosine (cyanoethyl)phosphoramidite monomer protected with a benzoyl group on the 4-amino position was from Glen Research Corporation. Anhydrous methanol (MeOH, 99+%, Gold Label), 1,1-dimethylhydrazine (98%) (NB this is a suspected carcinogen), 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU, 96%), and thiolacetic acid (96%) were from Aldrich. Absolute ethanol (99.7+%, Analar) and concd aqueous ammonia (d = 0.880, Aristar) were from BDH. Acetonitrile (HPLC Grade, Rathburn) was dried with molecular sieves (4A) at least overnight. All other chemicals were from either Aldrich or Sigma. All chemicals and solvents, unless stated otherwise, were used directly without further purification. The water content of anhydrous solvents was checked by Karl Fischer titration. Snake venom phosphodiesterase I (*crotalus durissus*) was from Sigma and alkaline phosphatase from Boehringer Mannheim.

Chromatography and Purification. Reversed-phase HPLC for base analysis was carried out on a Gilson 320, with a 620 Datamaster for integration and Shimadzu SPD6A UV spectrophotometetric detector, using a Waters 8MBC18 10μ column. Gradients were formed from 0.05 M aqueous KH_2PO_4 (pH 4.5) (buffer A) and 0.05 M aqueous KH_2PO_4 (pH 4.5) containing 33% CH_3CN (buffer B) at a flow rate of 3 mL/min. Fast protein liquid

(14) The pK_a of uridine is 9.3, and the pK_a of 4-thiouridine is 8.2. (Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: Berlin, 1984; p 111.)

(15) Brown, D. M.; Hewlins, M. J. E.; Schell, P. *J. Chem. Soc. C.* **1968**, 1925–1929.

(16) Anand, N. N.; Brown, D. M.; Salisbury, S. A. *Nucleic Acids Res.* **1987**, *15*, 8167–8176.

(17) Kong Thoo Lin, P.; Brown, D. M. *Nucleic Acids Res.* **1989**, *17*, 10373–10383.

(18) Negishi, K.; Harada, C.; Ohara, Y.; Oohara, K.; Nitta, N.; Hayatsu, H. *Nucleic Acids Res.* **1983**, *11*, 5223–5233.

(19) Kuchta, R. D.; Benkovic, P. A.; Benkovic, S. J. *Biochemistry* **1988**, *27*, 6716–6725.

chromatography (FPLC) was carried out on a Dionex BIOLC system with a Dionex variable-wavelength detector using a Pharmacia mono Q HR5/5 column. Gradients were formed from 0.4 M NaCl, 0.01 M NaOH, aqueous solution (pH 12) (buffer C) and 0.8 M NaCl, 0.01 M NaOH, aqueous solution (pH 12) (buffer D) or 1.2 M NaCl, 0.01 M NaOH, aqueous solution (pH 12) (buffer E) at a flow rate of 1 mL/min. High-performance thin-layer chromatography (HPTLC) was carried out on Merck Kieselgel 60 F_{254} aluminum-backed TLC sheets developed with 2.5% $CH_3OH/CHCl_3$. Nensorb Prep Cartridges (NEN Research Products, Du Pont Co., Boston, MA 02118, USA) were used according to the maker's instructions.

Base Analysis. The purity of the oligomers was assessed by analysis of a nucleoside digest. In general, 0.5 A_{260} unit of an oligomer was dissolved in 160 μL of H_2O and 20 μL of 600 mM Tris-HCl, 60 mM $MgCl_2$, pH 8.5. Snake venom phosphodiesterase I (10 μL, 10 μg of protein) was added and the mixture incubated (37 °C, 30 min), and then alkaline phosphatase (10 μL, 5 μg of protein) was added and incubation continued for 30 min. The deoxyribonucleosides were separated by HPLC using 96.5% buffer A and 3.5% buffer B for first 8 min and then with a linear gradient from 3.5% to 50% of buffer B over the following 15 min. The chromatography was monitored at 260 nm, but for the T^S oligomer the first 13 min of the run was monitored at 260 nm for the detection of dC, dG, T, and dA and the remainder at 335 nm for the detection of T^S. The amount of each nucleoside was measured by integration of the absorbance of each peak.[2,7,20] Retention times were as follows: dC, 1.9 min; 5-Me-dC, 3.7 min; dI (from enzymatic deamination of dA), 4.8 min; dG, 5.5 min; T, 6.5 min; dA, 12 min; T^{DH}, 14.8 min; T^{OMe}, 15.2 min; T^S, 15.6 min; T^{Tri}, 15.9 min; T^{OEt}, 19.4 min. Authentic compounds used for reference were prepared as follows: T^{OMe} and T^{OEt} as before;[7] T^S as before;[8] and T^{Tri} as literature.[9] T^{DH} [1-(2′-deoxyribofuransoyl)-4-(2,2-dimethylhydrazino)-5-methylpyrimid-2-one] was prepared as follows: to an CH_3CN (40 mL) solution of 3′,5′-*O*-bis(*tert*-butyldimethylsilyl)-4-(1,2,4-triazolo)thymidine (1.02 g, 2 mmol), prepared as before,[7] was added 1,1-dimethylhydrazine (600 μL, 10 mmol) at room temperature. After the mixture was stirred overnight, TLC ($CH_3OH/CHCl_3$ (5:95)) showed that most of the starting material had disappeared. Additional 1,1-dimethylhydrazine (400 μL) was added, and stirring continued for 5 h. TLC showed that besides the main spot (R_f 0.5), there were two UV absorbing spots (R_f 0.7, small amount, and R_f 0.25, tiny amount). The reaction solution was concentrated and the residue diluted with ethyl acetate (100 mL) and washed with saturated aqueous $NaHCO_3$ (2 × 100 mL) and then with saturated aqueous NaCl. The organic layer was dried over Na_2SO_4, evaporated into an oily residue, and purified by silica gel column. The compound with R_f 0.7 was found to be 3′,5′-bis(*tert*-butyldimethylsilyl)thymidine, and the main compound (R_f 0.5) was confirmed to be 1-(2′-deoxy-3′,5′-*O*-bis(*tert*-butyldimethylsilyl)ribofuranosyl)-4-(2,2-dimethylhydrazino)-5-methylpyrimid-2-one by NMR spectroscopy: ^1H NMR data (in DMSO-d_6) 0.07 (2 s, 12 H, 3′- and 5′-$Si(CH_3)_2R$), 0.86 (2 s, 18 H, 3′- and 5′-$Si(R)_2(CH_3)_3$), 2.04 (m, 2 H, 2′- and 2″-H), 2.10 (s, 3 H, 5-CH_3), 3.05 (s, 6 H, $NN(CH_3)_2$), 3.73 (m, 2 H, 5′-H), 3.79 (m, 1 H, 4′-H), 4.33 (m, 1 H, 3′-H), 6.14 (t, 1 H, 1′-H), and 7.37 (s, 1 H, 6 H). This compound was desilylated with tetrabutylammonium fluoride in tetrahydrofuran,[21] and the resulting nucleoside (T^{DH}) was purified by silica gel column and crystallized from acetone/methanol (90/10, V/V): ^1H NMR data (in DMSO-d_6) 1.95–2.07 (m, 2 H, 2′- and 2″-H), 2.11 (s, 3 H, 5-CH_3), 3.06 (s, 6 H, $NN(CH_3)_2$), 3.45–3.59 (m, 2 H, 5′-H), 3.71 (m, 1 H, 4′-H), 4.27 (m, 1 H, 3′-H), 6.09 (t, 1 H, 1′-H), and 7.75 (s, 1 H, 6 H); UV (MeOH) λ_{max} 289, $A_{260/289}$ 0.653; mp 197 °C.

DNA Synthesis on a Template Containing N^4-(Dimethylamino)-5-methylcytosine [4-(2,2-Dimethylhydrazino)-5-methylpyrimidin-2-one] (T^{DH}).

5′-^{32}P Labeling. A mixture (20 μL) containing 10 μM primer DNA (5′-ATCCGATAC-3′), 10 mM $MgCl_2$, 0.5 μM [γ-^{32}P]-ATP (3000 Ci/mmol), and 4 units of T_4 polynucleotide kinase in 50 mM Tris-HCl (pH 7.4) was incubated at 37 °C for 30 min.

Duplex Preparation. A solution (100 μL) containing 1 μM 5′-end-labeled primer DNA, 9 μM cold primer DNA, 10 μM cold template DNA (5′-CGCTCTTACAT^{DH}GTATCGGAT-3′), 5 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.4) was hybridized by heating at 100 °C for 2 min and then cooling to room temperature over 2 h.

Elongation and Electrophoresis. A reaction mixture (5 μL) containing 1 mM duplex, 150 μM dNTP (dATP or dGTP), 5 mM $MgCl_2$, and 0.2 μM E. coli DNA polymerase I (Klenow fragment) in 50 mM Tris-HCl (pH 7.4) was incubated at room temperature for 30 s and quenched directly with 5 μL of denaturing gel loading buffer (80% formamide and 0.1% each of bromophenol blue and xylene cyanol FF in electrophoresis buffer). Aliquots (2 μL) were then loaded on a 20% polyacrylamide gel (32 × 18 × 0.06 cm) containing 7 M urea. Electrophoresis was carried out at 2000 V, 30 W, for 2 h. The gel was autoradiographed and the amount elongated determined by cutting the gel into sections followed by scintillation counting in 3 mL of scintillation fluid.

Melting Curve Measurement. Non-self-complementary oligomers (Table II) were annealed with an equimolar quantity of each complementary strand in 0.1 M Hepes, pH 7.5, 0.2 M NaCl, and 0.02 M $MgCl_2$. The temperature-dependent change in absorbance at 260 nm was followed using a CARY3 spectrophotometer connected to a Cary temperature controller (Varian Techtron Pty Ltd, Australia). The temperature was increased by 1 °C/min. The Tm values were determined as the maximum values of the first derivative graph of the absorbance vs temperature graph.

Preparation of T^{Tri} Monomer: 5′-*O*-[(4,4′-Dimethoxytriphenyl)methyl]-4-triazolothymidine 3′-*O*-(2-Cyanoethyl (*N*,*N*-diisopropylamino)phosphoramidite) (2). Compound **2** was prepared by modification of a previous procedure.[9] 1,2,4-Triazole (1.38 g, 20 mmol) was suspended in dry CH_3CN (25 mL) at 0 °C (ice bath), and 0.4 mL of $POCl_3$ was slowly added with rapid stirring. Triethylamine (3 mL) was added dropwise and the suspension left stirring for 30 min. 5′-*O*-[(4,4′-Dimethoxytriphenyl)methyl]thymidine 3′-*O*-(2-cyanoethyl (*N*,*N*-diisopropylamino)phosphoramidite) (250 mg, 0.335 mmol, compound **1**) in dry CH_3CN (5 mL) was added over 20 min and stirring continued for 1 h. The reaction was stopped with saturated aqueous $NaHCO_3$ (30 mL) and then extracted with CH_2Cl_2 (50 mL). The organic layer was washed with saturated aqueous $NaHCO_3$ (30 mL) and saturated aqueous NaCl (30 mL) and then dried (Na_2SO_4), evaporated under reduced pressure into a small volume, coevaporated with toluene twice, and then precipitated from toluene into cold *n*-pentane (salt–ice bath). The resulting white precipitate, washed twice with fresh *n*-pentane, was dissolved in anhydrous benzene and lyophilized to give a white powder (260 mg, 97%). Its purity was checked by ^{31}P NMR (149.45 and 150.13 ppm in $CDCl_3$) and by TLC in 2.5% $CH_3OH/CHCl_3$ on HPTLC aluminum sheets (Silica gel 60 F_{254}, from E. Merck) which showed only two bright spots (phosphoramidite steric isomers) with R_f's (0.25 and 0.37) differing from the starting material (compound **1**, usual UV absorbing spots) with R_f's (0.35 and 0.42).

Synthesis of Oligodeoxyribonucleotides. Oligonucleotides were synthesized on a Cruachem PS200 automatic DNA synthesizer (Cruachem Ltd., Glasgow, Scotland) using PAC amidities of the normal bases (see above). The portion of the oligonucleotide 3′ to the 4-triazolothymine was synthesized on the machine, and then the T^{Tri} monomer was added manually. In both 1.0- and 0.2-μm scale 10 mg of compound **2** in a 2-mL conical glass vial with septum top (Wheaton reactivial) was dissolved in 0.1 mL of anhydrous CH_3CN and 0.1 mL of 0.5 M tetrazole in anhydrous CH_3CN added. The bottom end of the cartridge was disconnected from the machine and the mixture of the monomer and tetrazole injected from a gas tight syringe. The syringe was used to draw the solution in and out of the cartridge several times over a period of 3 min, and then the cartridge was immediately reconnected to the synthesizer to complete the synthesis. The yield of each coupling reaction was assessed by measuring the amount of 5′-protecting group (DMT) released by dichloroacetic acid. For comparison, a 5-methylcytosine-containing oligomer was also synthesized manually from the commercial monomer (Glen Research Corporation) using the same procedure as above. The 5-methylcytosine oligomer was deprotected (concd aqueous NH_3,

(20) Smith, C. A.; Xu, Y.-Z.; Swann, P. F. *Carcinogenesis* **1990**, *11*, 811–816.

(21) Beaucage, S. L.; Ogilvie, K. K. *Tetrahedron Lett.* **1977**, *18*, 1691–1694.

55 °C, overnight) and purified with Nensorb Prep cartridge.

Preparation of Oligomers Containing O^4-Methylthymine (T^{OMe}) and O^4-Ethylthymine (T^{OEt}).

a. Optimization of Substitution. The CPG support (5 mg) bearing the fully protected T^{Tri} 12 mer (AGCGAAT^{Tri}TCGCT) was put into Eppendorf tubes and either MeOH/DBU (1 mL, 9:1, v/v) alone or EtOH/DBU (1 mL, 9:1 v/v) with 10 mg of cetyltrimethylammonium bromide were added to each tube and left at 25 °C, for 12, 24, 48, and 68 h, respectively. The solution was neutralized (90 μL of 50% aqueous acetic acid, 1.5 equiv to DBU) and immediately passed through a Dowex 50 × 8, Na^+ form, 400 mesh ion-exchange column (10 mL wet volume), eluted with water, and collected in 1-mL fractions. The oligomers were usually found in fractions 4–6 by measuring A_{260} nm. The oligonucleotides were separated from failure sequences and the DMT group finally removed using a Nensorb Prep cartridge. The resulting oligomers were hydrolyzed enzymically and the base composition measured (cf. base analysis). The extent of substitution by MeOH or EtOH was assessed by comparison of percentage of 4-triazolothymidine (t_R 15.9 min) with that of O^4-methylthymine (t_R 15.2 min) or with that of O^4-ethylthymine (t_R 19.4 min) using reversed-phase HPLC.

b. Synthesis of O^4-Methylthymine and O^4-Ethylthymine Oligomers. The fully protected T^{Tri} 12 mer described above while still attached to the CPG support was treated with either methanol/DBU or ethanol/DBU at 25 °C overnight or for 2 days, respectively. The solution was neutralized and immediately passed through Dowex 50 as described above. The oligonucleotides were purified with a Nensorb Prep Cartridge (yield in Table I).

Preparation of Oligomers Containing Thymine from the T^{Tri} Oligomer. The fully protected T^{Tri} 12 mer described above while still attached to the CPG support was treated with 0.5 M aqueous NaOH (24 h at 25 °C). This cleaved the oligomer from the CPG support, substituted the 4-triazolothymine, and removed all protecting groups except the DMT group at the 5′ end of the oligomer. The resulting oligomer was purified with a Nensorb Prep cartridge as above (yield in Table I).

Preparation of Oligomers Containing 5-Methylcytosine. The fully protected T^{Tri} 12 mer described above while still attached to the CPG support was treated with concd aqueous ammonia (d = 0.880, overnight, 25 °C). This cleaved the oligomer from the CPG support, substituted the 4-triazolothymine, and removed all protecting groups except 5′-DMT. The resulting oligomer was purified with a Nensorb Prep cartridge (yield in Table I).

Preparation of Oligomers Containing N^4-(Dimethylamino)-5-methylcytosine (T^{DH}). The fully protected T^{Tri} 12 mer described above while still attached to the CPG-support was treated with $NH_2N(Me)_2/CH_3CN$ (10/90, v/v, 3 h, 25 °C) to substitute the 4-triazolothymine. As the oligomer was still attached to the CPG support, it can be washed with CH_3CN (5 × 1 mL) to completely remove 1,1-dimethylhydrazine. The oligomer was then cleaved from the CPG support, and all protecting groups except 5′-DMT were removed with aqueous 0.5 M NaOH (1 mL, 24 h, 25 °C). The resulting oligomer was purified with a Nensorb Prep cartridge (yield in Table I). The FPLC profile of the product and the HPLC of the nucleoside digest is shown in Figure 3.

Preparation of Oligomers Containing 4-Thiothymine.

a. Optimization of Thiation. To determine the optimum condition for thiation, 0.5 mL of CH_3COSH/CH_3CN (10/90, v/v) was added to each of four Eppendorf tubes containing 3 mg of the CPG support bearing the protected T^{Tri} 12 mer, with the 5′-DMT removed, and left at 25 °C for 4, 8, 12, and 24 h, respectively. Then each sample was washed with CH_3CN (5 × 1 mL) and treated with MeOH/DBU (0.5 mL, 90/10, v/v) overnight to cleave the oligomer from the CPG support and to remove the protecting groups. The products were separated by FPLC and the extent of thiation was calculated by comparing the integrated absorbance (at 260 and 335 nm) of the T^S oligomer with that of T^{OMe} oligomer which had been formed from substitution of T^S and of any remaining T^{Tri} during the deprotection.

b. Synthesis of Oligomers Containing 4-Thiothymine. The fully protected T^{Tri} 12 mer described above while still attached to the CPG support was treated with CH_3COSH/CH_3CN (1 mL, 10 %, v/v) for 24 h to convert 4-triazolothymine in the oligomer into 4-thiothymine. As the oligomer was still attached to the CPG support, it can be washed with CH_3CN (5 × 1 mL) to completely remove CH_3COSH. MeOH/DBU (1 mL, 90/10, v/v) was added and left overnight for deprotection. The resulting product was then purified as described above for the T^{OMe} oligomer (yield in Table I).

Acknowledgment. We wish to thank the Cancer Research Campaign for their most generous support and Dr. Michael J. Gait for critically reading this manuscript. We are particularly grateful to Dr. D. M. Brown who suggested that 4-(2,2-dimethylhydrazino)-5-methylpyrimid-2-one might have interesting properties and to Dr. Paul Kong (Medical Research Council Laboratory of Molecular Biology, Cambridge) for advice about 1,1-dimethylhydrazine, Dr. Panogiotis Georgiadis for helpful discussions, and Ms. H. B. Tan for help in DNA elongation experiments.

Registry No. 1, 98796-51-1; 2, 101712-10-1; AGCGAAT^{Tri}CGCT, 141344-58-3; CGCTCTTACAT^{DH}GTATCGGAT, 141507-63-3; ATCCGATAC ^{32}P labeled, 141344-59-4; AGCGAATTCGCT, 130583-08-3; AGCGAAT^{OMe}TCGCT, 130583-09-4; AGCGAAT^{OEt}TCGCT, 130583-10-7; AGCGAAT^{NH_2}TCGCT, 141344-60-7; AGCGAAT^{DH}TCGCT, 141344-61-8; AGCGAAT^{S}TCGCT, 141096-11-9; 5′CAGGAATTCGC3′·3′GTCCTTAAGCG5′, 141076-26-8; 5′CAGGAATT^{DH}CGC3′·3′GTCCTTAAGCG5′, 141374-93-8; 5′CAGGAATT^{DH}CGC3′·3′GTCCTTAGGCG5′, 141344-64-1; 5′CAGGAATCCGC3′·3′GTCCTTAGGCG5′, 141344-66-3; 5′CAGGAATCCGC3′·3′GTCCTAAGCG5′, 141344-67-4; DNA polymerase, 9012-90-2; triazole, 288-88-0.