KINETIC STUDIES ON THE FIDELITY OF DNA REPLICATION INVOLVING DNA TEMPLATES CONTAINING *O*^{6}-METHYLGUANINE

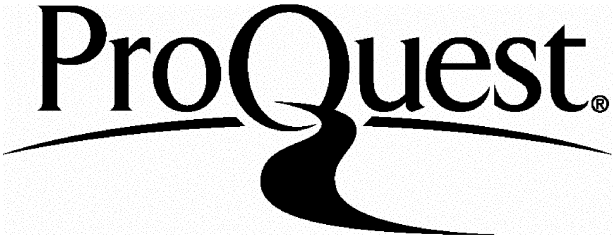
by

Hwee Boon Tan

A thesis submitted for the degree of Doctor of Philosophy

Department of Biochemistry and Molecular Biology
University College London

August 1993



To My Parents

KINETIC STUDIES ON THE FIDELITY OF DNA REPLICATION INVOLVING DNA TEMPLATES CONTAINING O^6-METHYLGUANINE

by Hwee Boon Tan

Abstract

Production by *N*-nitroso compounds of O^6-alkylguanine (O^6-alkylG) in DNA directs the misincorporation of thymine during DNA replication, leading to G:C to A:T transition mutations, despite the fact that DNA containing O^6-alkylG:T base-pairs is less stable than that containing O^6-alkylG:C pairs. In the work presented in this thesis, the kinetics of incorporation by Klenow fragment of *Escherichia coli* DNA polymerase I of thymine (T), and of cytosine (C), opposite O^6-meG in the template DNA strand were examined. Both T and C were incorporated opposite O^6-meG much slower than nucleotides forming regular A:T or G:C base pairs. Using an excess of Klenow over DNA and various concentrations of dTTP and dCTP, the progress of incorporation of a single nucleotide in a single catalytic cycle of a preformed Klenow-DNA complex was measured (pre-steady state kinetics). The results were consistent with the kinetic scheme:

1. polymerase-DNA binds dNTP;
2. conformational change in polymerase;
3. formation of phosphodiester between the dNTP and the 3'-OH of primer;
4. conformational change of polymerase;
5. release of pyrophsphate.

The results were analysed mathematically to identify the steps at which the rate constants differ significantly between the incorporation of T and C. The only significant difference was the 5-fold difference in the rates of formation of the phosphodiester bond (for dTTP, $k_{forward}$ = 3.9 s^{-1} and k_{back} = 1.9 s^{-1}; for dCTP, $k_{forward}$ = 0.7 s^{-1} and k_{back} = 0.9 s^{-1}). The equilibrium constants for each step suggest that the greatest change in the Gibbs' free energy occurs at the conformational change after polymerisation, and that while the formation of the phosphodiester bond to T is slightly exothermic, that to C is slightly endothermic. The K_ms calculated from the rate constants (K_m = 33.5 μM (24.0-46.7)* for both dTTP and dCTP [* 5% and 95% confidence limits]) were close to the approximate K_ms obtained from Michaelis-Menten analysis of the initial rates of pre-steady state polymerisation (K_m = 30-35 μM for T and C). The measured progress of independently determined steady state experiments (i.e. polymerisation under conditions of excess DNA

over Klenow) was close to that predicted from these calculated rate constants. The incorporation of the nucleotide following C in an O^6-meG:C pair was much slower than that following T in an O^6-meG:T pair. Taken with the available structural data (Kalnik et al., 1988a, b), this suggests that the discrimination in favour of the incorporation of T opposite O^6-meG arises mainly because the T:O^6-meG base-pair retains the Watson-Crick configuration (with the N1 of the purine juxtaposed to N3 of the pyrimidine), whereas the C:O^6-meG base-pair is a wobble base pair with a distorted phosphodiester link 3' to the C. The slow incorporation of C opposite O^6-meG, and of the next correct nucleotide following the incorporation of C, can be ascribed to the stereochemical problems encountered when forming the distorted phosphodiester links. The recent X-ray crystallography data (Beese et al., 1993) of a Klenow complexed with duplex DNA provided evidence that Klenow fragment interacts with the primer-template through the phosphodiester backbone, thus an incorporation event that produces a distortion in the phosphodiester backbone, such as the incorporation of C opposite O^6-meG, could very well reduce the rate of its incorporation.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Peter Swann, for his guidance and support throughout the course of this project. I am also grateful to Yao-Zhong Xu, who synthesized the protected O^6-meG monomer for DNA synthesis and for his advice in the synthesis and purification of oligonucleotides containing O^6-meG. I would also like to thank Ted Chance for his assistance with the analyses of the data and for allowing me to print this thesis with his computer and printer. This project was funded by the Cancer Research Campaign, and I am most grateful to them for their generous support.

I am delighted to have worked with the members of our lab, Raymond Mace, Panagiotis Georgiadis, Qinguo Zheng, Tim Waters, Felipe Ribeiro-Pinto, and David Moulton, who accepted me as an 'Honorary monk' in the all-male group despite my lack of knowledge in football. I am also happy to have known other members of the department, in particular Dr. Liz Shepard and members of her lab, who provided me with support and many cups of coffee when the experiments did not turn out right.

I am also grateful to Mr. Doug Sargent, Mr. Mike Larkum, and Mr. Brian Horn for building the rapid-quench and the quench flow apparatuses for my experiments.

Finally, I am greatly indebted to my parents and my brother, to Nick and his parents for their faith and moral support, especially towards the last year of the project. This thesis would have been much more difficult to write without their constant encouragement.

Table of Contents

Chapter 1. (Table of Contents Continued)

Appendix

List of Figures

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 3. (Continued)

Chapter 3. (Continued)

Chapter 3. (Continued)

Chapter 3. (Continued)

List of Tables

Chapter 1.

Chapter 2.

Chapter 3. (List of Tables Continued)

Chapter 3. (List of Tables Continued)

Abbreviations

AP	aminopurine
5-BrdU	5-bromodeoxyuridine
COSY	correlated spectra
CPG	controlled-pore glass
DMBA	dimethylbenzanthracene
DMTr	4,4'-dimethoxytriphenylmethyl
EDTA	ethylenediaminetetraacetic acid
ENU	*N*-ethyl-*N*-nitrosourea
FPLC	Fast Protein Liquid Chromatography™ (HPLC on Mono-Q HR 5-5 column from Pharmacia)
HPLC	high performance liquid chromatography
MMTr	4-monomethoxytriphenylmethyl
MNNG	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine
MNU	*N*-methyl-*N*-nitrosourea
MSNT	1-(mesitylenesulphonyl)-3-nitro-1,2,4-triazole
N7-alkylG	N7-alkylguanine
NDMA	*N*-nitrosodimethylamine
NDEA	*N*-nitrosodiethylamine
NMR	nuclear magnetic resonance
NOC	*N*-nitroso compounds
NOE	nuclear Overhauser enhancement
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
NNN	*N'*-nitrosonornicotine
O^6-meG	O^6-methylguanine
O^6-alkylG	O^6-alkylguanine
THF	tetrahydrofuran
T_m	melting temperature
TMG	N^1,N^1,N^3,N^3-tetramethylguanidinium
TSNA	tobacco-specific nitrosamines

The last thing that one knows in constructing a work is what to put first

Blaise Pascal

CHAPTER 1.

A General Introduction

Since the beginning of recorded history man has known of cancer. A disease that affects all animal species, even plants, it generally manifests itself late in life. As we have achieved advances in medicine, the average human life span has increased so that cancer is now a major killer in developed countries, responsible for one in five deaths on the average. Though cancer research is now an important area of research, and we have gained much insight into this field in the past few decades, a full understanding of the disease is still not within grasp. A diagnosis of cancer in a patient still inevitably casts a great shadow of despair as many forms of cancer remain at present terminal diseases.

Despite the enormous amount of work that has been carried out in cancer research, there is still a big gap of knowledge on the origin of human and animal cancers. However, chemical carcinogenesis is relatively better understood. Of the carcinogens known to man, the *N*-nitroso compounds constitute the largest group and their carcinogenicity has been well-studied. O^6-alkylguanine (O^6-alkylG) is a major promutagenic compound produced by the carcinogenic *N*-nitroso compounds. Its presence in DNA can lead to transition mutations because it directs the incorporation of thymine in the daughter strand during DNA synthesis. It does not do this by the formation of a stable H-bonded base-pair with thymine, contrary to the popular belief that in DNA synthesis "base selection is template-directed by the formation of the appropriate Watson-Crick base pair with the incoming deoxynucleoside triphosphate" (McHenry, 1988). The objective of this work is to examine this paradox in the hope that its resolution might add to our knowledge of the underlying factors in DNA replication as well as throwing light on the mechanism of action of the carcinogenic nitrosamines. At the time this project was started, this paradox had not been addressed mainly because the availability of an oligonucleotide containing O^6-methylguanine (O^6-meG) at a specific site within a sequence is a prerequisite for investigating the kinetics of mispairing opposite the promutagenic O^6-meG and until then, the chemical synthesis of DNA containing O^6-meG had been too difficult for the biochemist. A year or so after the project was begun, experiments with the same objectives were published (Singer et al., 1989 and Dosanjh et al., 1990), but the conclusions obtained from these studies were difficult to interpret and the actual mechanism of this misincorporation remained unknown.

This thesis describes the kinetics of incorporation of thymine and of cytosine opposite an O^6-meG in the template strand. But before the work is described, it is necessary to provide a general introduction to the topic. The first part of the introduction explains the relevance of *N*-nitroso compounds to cancer, how O^6-alkylG produced by the alkylation of DNA is involved in transition mutations and the importance of somatic cell mutation caused by replication of alkylated DNA in the activation of oncogenes. The aim of this project is to study the kinetics of formation of two mismatched base-pairs involving the promutagenic O^6-meG, i.e. the O^6-meG:T and the O^6-meG:C base-pairs. In order to understand how such mismatches can occur, we need to consider the structure of normal (Watson-Crick) and mismatched base-pairs, how base-pairing interactions are affected by the structure of the bases, and how these interactions might lead to the occurrence of transition mutations. These aspects are discussed in the next part of the introduction. Transition mutations are produced by errors in DNA replication *in vivo*, and the biological factors involved should not be overlooked in examining the occurrence of mutation of any kind. The remainder of the introduction describes how the DNA polymerase, in particular the Klenow fragment from *Escherichia coli* (which is the DNA polymerase used in this project), attains a high level of accuracy in copying DNA.

1.1 *N*-nitroso compounds and cancer

N-nitroso compounds- an historical perspective

N-nitroso compounds (NOCs) constitute the largest and most versatile group of chemical carcinogens known to man, capable of causing cancer in all animal species and in all organs. The structure of some members of NOCs is shown in Fig. 1.1. To date, more than 300 NOCs have been tested and found to be carcinogenic in animals.

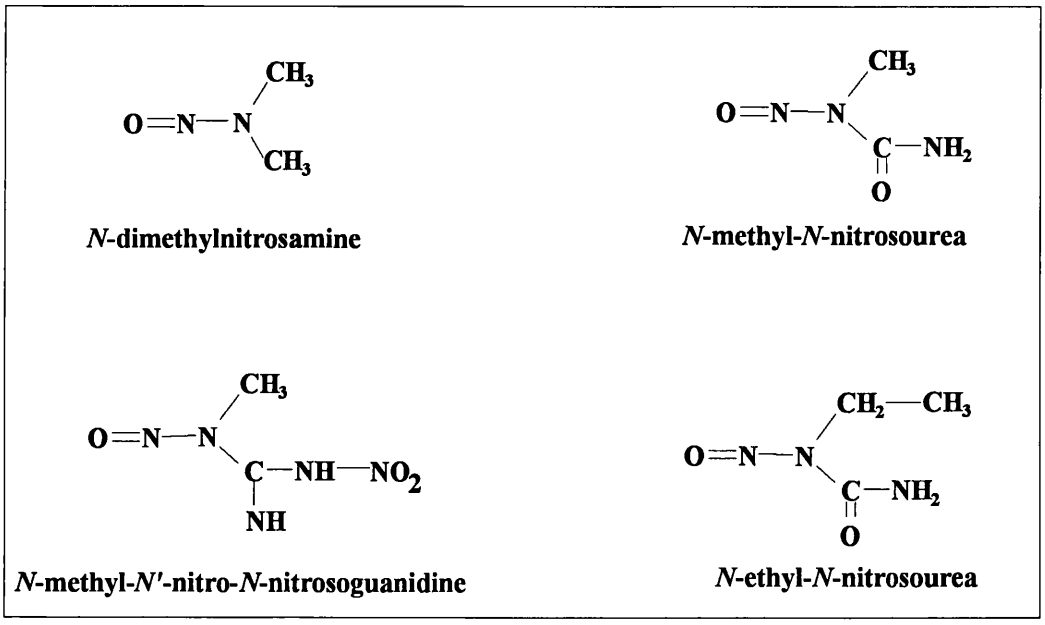


Figure 1.1. The chemical structures of some *N*-nitroso compounds.

Today, we know more about the mechanism of action of NOCs as initiators of cancer than any other carcinogens, and we can justifiably say that much progress has been made in this field since the first report in 1956 by Magee and Barnes that *N*-nitrosodimethylamine (NDMA) is carcinogenic. A brief history of research on NOCs is outlined in Table 1.

Table 1. Important landmarks in the history of research on the mechanism of action of *N*-nitroso compounds in chemical carcinogenesis.

1956	The carcinogenic action of the simplest nitrosamine, *N*-nitrosodimethylamine (NDMA), was discovered by Magee & Barnes, who showed that nearly all rats (19 out of 20) fed on a diet containing NDMA developed malignant tumours within less than a year.
1957	Lawley & Wallick showed that NDMA reacts with the bases in DNA, predominantly on the 7-position of guanine.
1958	Magee & Vandekar showed that the *N*-nitroso compounds are metabolised by the liver in the presence of oxygen, while other tissues tested were inactive.
1958	Loveless showed that ethyl methanesulphonate is mutagenic.
1959	Loveless published a series of papers concerned with mutations in T2 phage caused by alkylating agents, which later became crucial in developing the understanding of the action of *N*-nitroso compounds. The inactivation of T2 phage by different alkylating agents was investigated. Among the alkylating agents studied were methyl and ethyl methanesulphonate and diethyl sulphate. Methyl methanesulphonate was shown to be highly cytotoxic but non-mutagenic, whilst both ethyl methanesulphonate and diethyl sulphate were found to induce mutation. This led to the conclusion that ethylation induces mutation.
1962	Following the demonstration in 1959 by Magee & Barnes that a single dose of NDMA can produce kidney tumours in the rat, Druckrey & Steinhoff succeeded in producing liver tumours in guinea pigs with diethylnitrosamine. This was important because guinea pigs are well known for their high resistance to the induction of tumour by another class of carcinogens, the aromatic amines, and because this was the beginning of the demonstration that *N*-nitroso compounds are active in all organs in all species.
1962	Magee & Farber reported that a metabolite of NDMA methylates DNA and RNA. This was the first paper to report the reaction of a carcinogen metabolite with DNA and to produce evidence for a link between the two.

Table 1. Continued.

1962	Heath demonstrated that certain *N*-nitroso compounds require metabolic activation for their carcinogenic action.
1963	Krieg observed that the mutations caused by ethyl methanesulphonate were primarily G:C to A:T transition mutations.
1964	Brookes & Lawley studied the binding of carcinogenic and non-carcinogenic polycyclic hydrocarbons to DNA in the mouse skin. They showed that polycylclic aromatic hydrocarbons also produce metabolites which react with DNA and that there was a correlation between the carcinogenicity of each polycyclic hydrocarbon studied and the extent of reaction with DNA. Taken with the work on nitrosamines, this thus introduced the possibility that all carcinogens act in the same way.
1967	Druckrey et al. published an important review on the carcinogenic action of 65 different *N*-nitroso compounds in the rat. They observed that only those *N*-nitroso compounds that break down spontaneously (e.g. NMU), or that have a free hydrogen on the α-carbon and can therefore be broken down enzymically to an alkylating metabolite (e.g. NDMA), are carcinogenic.
1968	It was known that the major product from alkylation of DNA is N7-alkylG, yet some alkylating agents that are well capable of producing N7-alkylG are at best weak carcinogens. There appeared to be no correlation between alkylation of N7 of guanine and carcinogenesis. Thus, Swann & Magee set out to discover if a correlation exists between the ability of a compound to produce N7-alkylG and its potency as a carcinogen. They tested the carcinogenicity of two alkylating agents (methyl methanesulphonate and dimethyl sulphate) and two nitrosamines (NMU and NDMA). All four compounds produced 7-methylG in the nucleic acids of the rat, but only the nitrosamines were carcinogenic in the rat.
1969	Loveless noticed a parallel between the carcinogenicity of the methylating and ethylating nitrosamides and methyl and ethyl methanesulphonate, and the ability of these compounds to produce G:C to A:T transition mutations in T-even phage. He found that carcinogenic alkylating agents and *N*-nitroso compounds alkylate the 6-*O*-position of guanine while the poorly carcinogenic alkylating agents such as methyl methanesulphonate do not share this property. The major promutagenic lesion in *N*-nitrosamine carcinogenesis was thus identified as O^6-alkylG.

Table 1. Continued.

1970	Miller drew together the British and the American work and proposed that all chemical carcinogens share the property of reacting with DNA.
1971	Swann & Magee showed ethyl methanesulphonate to be carcinogenic and capable of producing kidney tumours indistinguishable from those produced by NDMA or NMU. This showed that the *N*-nitroso compounds must owe their carcinogenic activity to their ability to alkylate a cellular component.
1973	O^4-alkylthymine, another promutagenic lesion, was identified in DNA after reaction with NMU and it was suggested that this alkylated base would miscode with guanine (Lawley et al).
1973	Gerchman & Ludlum showed that the O^6-meG in template RNA directed the misincorporation of uracil by RNA polymerase. This provided evidence that O^6-meG in template RNA could lead to G to A transition mutations.
1974	Goth & Rajewsky showed that the liver, which is refractory to the carcinogenic action of *N*-nitroso compounds, contains an enzyme system to remove O^6-alkylG, whereas the brain which is very susceptible to the carcinogenic action of *N*-nitroso compounds, does not. This suggested that organ specificity might also depend upon the ability to repair damaged DNA.
1976	Abbott and Saffhill showed that the presence of O^6-meG in DNA leads to misincorporation during DNA synthesis.
1977	Coulondre and Miller reported that virtually all the mutations produced by MNNG are G:C to A:T transition mutations, thus establishing the pre-eminence of the 6-*O*-alkylation of guanine in the mutagenic activity of the alkylating *N*-nitroso compounds.
1985	The Ha-*ras* oncogene was shown to be activated *in vivo* by a G:C to A:T point mutation in the 12th codon in NMU-induced mammary tumours (Zarbl et al).
1990	Activation by point mutation of Ha-*ras* was shown to precede the onset of neoplasia (Kumar et al., and Miyamoto et al), which is consistent with the idea that the mutation is an (the) initiating event in nitrosamine carcinogenesis.

Bioactivation of *N*-nitroso compounds

N-nitroso compounds are the *N*-nitroso derivatives of amines or amides. They can be divided into two groups, those which are chemically reactive under physiological conditions and can spontaneously form active alkylating agents (Fig. 1.2), e.g. *N*-methyl-*N*-nitrosourea and *N*-methyl-*N*-nitro-*N*'-nitrosoguanidine (Archer, 1989), and those such as the dialkylnitrosamines and cyclic nitrosamines, which require metabolic activation to form active alkylating species.

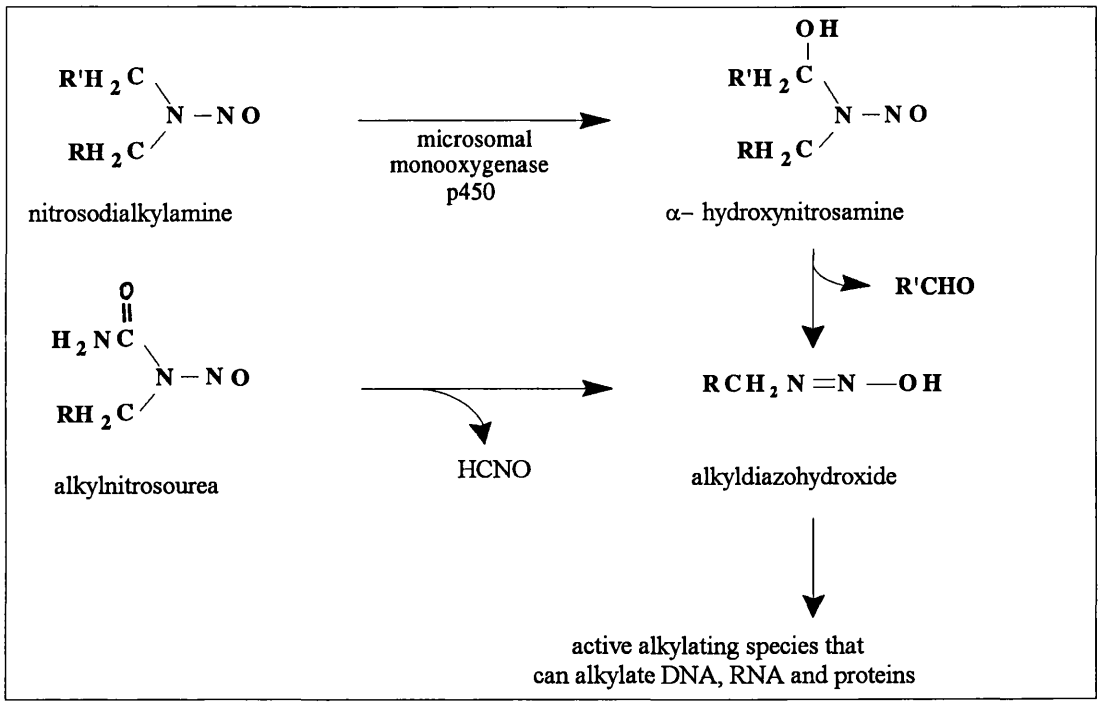


Figure 1.2. Bioactivation of NOCs. Nitrosamines require metabolic conversion to α-hydroxynitrosamines before they can form a chemically active alkylating species.

The difference in chemical stability of nitrosamines and nitrosamides is reflected in the different organ and tissue specificities of these compounds. Whereas nitrosamides have a wide spectrum of target tissues and also tend to produce tumours at or near the site of application, nitrosamines only produce tumours in organs that can metabolise these compounds, which may be far removed from the site of administration. This activation pathway for aliphatic and cyclic nitrosamines has been confirmed recently (reviewed by Archer, 1989). The first step in the metabolic pathway involves hydroxylation of one of the two carbons adjacent to the *N*-nitroso group (α-hydroxylation), a reaction which is catalysed by the cytochrome P-450 enzyme system. This membrane-bound enzyme system consists of a family of isoenzymes which hydroxylate a

number of chemicals foreign to the organism, including drugs and carcinogens, as part of a detoxification process. The P-450 responsible for the metabolism of dimethylnitrosamine is 2EI, the ethanol inducible enzyme (discussed by Yang et al., 1991).

Human exposure to *N*-nitroso compounds

Since the discovery of the *N*-nitroso compounds, and in particular since the occurrence of widespread poisoning of sheep in Norway fed on fish meal preserved with nitrite (Ender et al., 1964), there has been considerable interest in the possibility that some human cancers may be the result of exposure to *N*-nitroso compounds. This exposure can come from two sources, endogenously synthesized nitrosamines and pre-formed nitrosamines in the diet (e.g. tobacco smoke, certain kinds of pickles and preserved meats and etc.)

Endogenous formation of NOCs can occur by several means in the human body:

(1) Nitrosation reactions in the stomach are generally regarded as the most important (Bartsch et al., 1989). This results from the ingestion of nitrite or nitrate and secondary or tertiary amines. Nitrite is chemically more reactive than nitrate, but both are incapable of nitrosating amines; they can however be converted to oxides of nitrogen (N_2O_3, N_2O_4) which can nitrosate amines and amides to form *N*-nitroso compounds by the direct transfer of a nitric oxide radical (NO•) to the amine or amide. As these nitrosation reactions are catalysed by acid, the acidic environment of the stomach is well suited to such reactions (reviewed by Leaf et al., 1989).

(2) Besides chemical nitrosation reactions, bacteria that possess the nitrate reductase enzyme also promote the formation of *N*-nitroso compounds in the stomach and urinary bladder by first reducing nitrate to nitrite and then converting the latter to a nitrosating species such as N_2O_3 or N_2O_4 (reviewed by Leaf et al., 1989). Microbial formation of *N*-nitroso compounds has been demonstrated in human beings (Bartsch et al., 1986). Various strains of bacteria were obtained from human sources (urinary infections, wounds, hemocultures, nasotracheal secretions, vagina and faeces) and tested for their ability to produce *N*-nitrosomorpholine (NMOR) from nitrite and morpholine at neutral pH, and a wide spectrum of nitrosation rates was displayed by different strains of bacteria.

(3) Activated macrophages can also *N*-nitrosate secondary amines (reviewed by Marletta, 1989). Following the discovery that people with infections have increased excretion of nitrate, it is now widely accepted that an activated macrophage utilizes L-arginine as a precursor for the production of the cytotoxic nitric oxide; the nitric oxide is inactivated when it reacts with oxygen to produce nitrosating agents (reviewed by Marletta, 1989).

(4) Endothelial cells also produce nitric oxide from L-arginine when stimulated by bradykinin, and in this case the nitric oxide plays a role in intracellular signalling. Certain classes of neurons also appear to be capable of oxidising arginine (reviewed by Leaf et al., 1989).

General speculation about exposure to endogenously formed nitrosamines has tended to centre on cancer of the stomach, oesophagus and bladder. Endogenous nitrosation probably represents the greater part of human exposure to *N*-nitroso compounds, but it is very difficult to assign a degree of importance to it due to the lack of sensitive and reliable methods to measure the amount of nitrosamines produced endogenously. The problem is made more difficult because endogenous nitrosation is affected by numerous factors such as diet, smoking and the presence of inhibitors like vitamin C and E. The NPRO test devised by Bartsch and Oshima and their co-workers is probably the best method available today for measuring endogenous nitrosation. It is based on the detection of *N*-nitrosoproline (NPRO) and other *N*-nitrosamino acids (NAA) in the urine after the subject has been given a dose of L-proline (see Bartsch et al., 1989). This method has the limitation that it measures only nitrosation in the stomach but not endogenous nitrosation catalysed by macrophages and endothelial cells. Nonetheless, the NPRO test has been applied in several clinical studies where the endogenous nitrosation of people living in high- and low-incidence areas for stomach cancer in Japan and Poland, of people from high- and low-incidence areas for oesophageal cancer in China, and of European patients who suffer from urinary bladder infections. The results showed that people living in the high-risk areas have higher levels of endogenous nitrosation (Bartsch et al., 1989), supporting the theory that high levels of endogenous nitrosation can increase the risk of cancer of the stomach, oesophagus and urinary bladder.

Human exposure to pre-formed nitrosamines in the diet is easier to measure, and some progress has been made with tobacco smoke, chewing tobacco and snuff, notably by Hecht and Hoffmann. There is strong data to support the causal relationship between cigarette smoking and cancer of the lung, trachea, oesophagus, pancreas and nasal cavity (reviewed by Hecht and Hoffmann, 1989). The evidence linking snuff dipping and cancer of the oral cavity is also particularly strong. Tobacco chewing and snuff dipping have also been suggested to increase the risk of cancer of the nasal cavity, pancreas, kidney and bladder (Hecht & Hoffmann, 1989).

Tobacco smoke, chewing tobacco and snuff all contain nicotine alkaloids, which can be nitrosated. The nitrosamines formed from the nitrosation of such tobacco alkaloids are referred to as the tobacco-specific nitrosamines (TSNAs). These have been tested in animals and found to affect the same organs that are at risk in individuals who smoke (reviewed by Hecht and Hoffmann, 1989). Of the known TSNAs, NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone) and NNN (*N*'-nitrosonornicotine) are the most potent carcinogens; they are the only

TSNAs known to cause cancer of the oral cavity in experimental animals, and NNK has been shown to be capable of causing lung cancer in all animal species tested (reviewed by Hecht & Hoffmann, 1989).

The discussion on the carcinogenicity of *N*-nitroso compounds and on human exposure to *N*-nitroso compounds clearly shows the importance of these compounds as carcinogens, and it is vital that research on this field moves forward so that we may learn the causes of cancer and improve the preventive measures and cures for it.

1.2 Alkylation by *N*-nitroso compounds

N-nitroso compounds are carcinogenic because they can spontaneously give rise to, or be metabolised to active alkylating species which react with numerous sites in protein molecules, RNA and DNA. The sites of alkylation on DNA include the ring nitrogens, the exocyclic oxygen atoms of the DNA bases and the oxygens of the phosphodiester links between nucleotides. At least twelve sites in DNA are targets for alkylation under physiological conditions (see Table 2).

Different alkylating agents have different carcinogenic activities; even members from the same group (e.g. *N*-methyl-*N*-nitrosourea and *N*-ethyl-*N*-nitrosourea) show different mutagenic properties (Singer, 1979). Generally the extent of reaction and the proportion of each adduct formed depend on the alkylating agent used and on the alkylating group. It was the difference in chemical reactivity of different alkylating agents with DNA that provided a clue to the role of the individual alkylated products in the carcinogenicity of *N*-nitroso compounds (Table 2). Initially adducts formed from reactions at N7 of guanine attracted the most attention because this is by far the most abundant class formed when DNA is treated with most alkylating agents. However, it then became apparent that the ability to form N7-alkylG does not correlate with the carcinogenicity of alkylating agents. A major breakthrough in alkylation mutagenesis came in 1969 when Loveless proposed that alkylation at the O^6-position of guanine is the most important adduct in the mutagenicity and carcinogenicity of *N*-nitrosamines and *N*-nitrosamides. More support for this theory came when Lawley and Thatcher (1970) detected O^6-meG in DNA after its treatment with the carcinogenic and mutagenic MNNG, but not after treatment with the non-carcinogenic and poorly mutagenic dimethyl sulphate. These two alkylating agents are different in their chemical activity in that MNNG tends to react more at the exocyclic atoms of bases (including the O^6-position of guanine) whereas dimethyl sulphate tends to alkylate ring nitrogens (e.g. N7 of guanine). Loveless (1969) had suggested that O^6-alkylG would mispair with thymine instead of cytosine during DNA replication and that this would explain the G:C to A:T transition mutations he had observed. The process is shown schematically in Figure 1.3.

Direct evidence for the miscoding properties of O^6-meG was obtained from the experiments of Gerchman and Ludlum in 1973 in which they showed that O^6-meG can mispair with thymine during *in vitro* replication by DNA polymerase. Further evidence that O^6-alkylG can cause G:C to A:T transition mutations *in vivo* came from site-directed mutagenesis work by Loechler et al. (1984). They inserted O^6-meG into a unique site in M13 DNA, and used this modified M13 phage to transfect *E. coli*. After replication in *E.coli*, the progeny M13 DNA was examined for mutations at the specific site of insertion of O^6-meG. The only mutations they found at this site were G:C to A:T transition mutations.

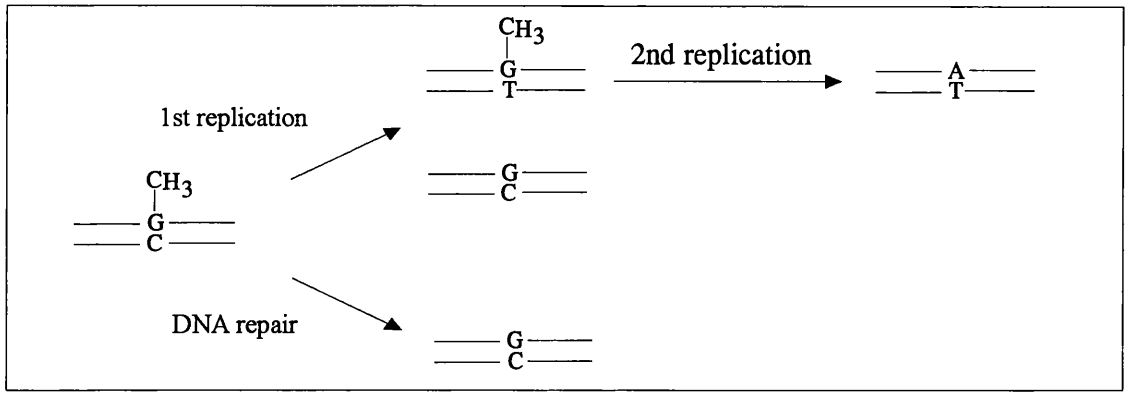
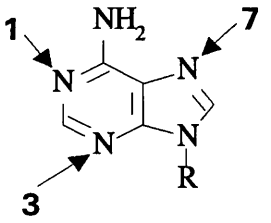
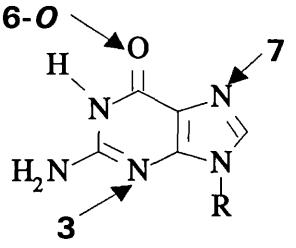
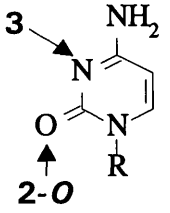
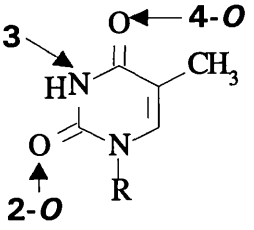
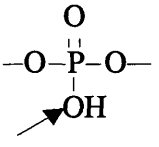


Figure 1.3. Mechanism of mutagenesis by O^6-meG.

O^6-meG is not the only promutagenic lesion in alkylation mutagenesis. O^4-alkylthymine (O^4-alkylT) is also promutagenic because it can mispair with guanine instead of adenine during DNA replication. O^4-alkylT was first identified in 1973 by Lawley and his co-workers. Its significance in alkylation mutagenesis escaped notice at first because it is only produced in very minute amounts by the methylating *N*-nitroso compounds such as NMU and MNNG, compared with the total alkylation sites in DNA (Table 2). Like O^6-alkylG, the miscoding properties of O^4-alkylT have been verified using both RNA and DNA polymerases of bacterial or mammalian origin (reviewed by Saffhill et al., 1985).

In *in vitro* replication assays using RNA polymerase, O^2-alkylT appeared to be promutagenic (reviewed by Saffhill et al., 1985); but when the same experiments were performed with DNA polymerase in place of RNA polymerase, it produced much less mutations (~ 10% of that produced by the same level of O^4-alkylT; reviewed by Saffhill et al., 1985). The relevance of O^2-alkylT to mutagenesis and cell toxicity is as yet unknown, but it is known that the presence of O^2-alkylT in the template strand tends to inhibit DNA synthesis (Bhanot et al., 1992). Incorporation of dAMP opposite a template O^2-ethylthymine was found to impede DNA synthesis, whereas a dTMP incorporated opposite a template O^2-ethylT was easily extended, suggesting a role for O^2-ethylT in A:T to T:A transversion mutagenesis by ethylating agents (Grevatt et al., 1992).

Table 2. Reaction of *N*-methyl-*N*-nitrosourea (MNU), *N*-ethyl-*N*-nitrosourea (ENU), and methyl methanesulphonate (MMS) with DNA *in vitro*. (Data taken from Singer and Grunberger 1983).

	% Total Alkylation			
	MNU	ENU	MMS	
N1-	1.3	0.2	3.8	**Adenine** 
N3-	9.0	4.0	10.4	
N7-	1.7	0.3	1.9	
O^6-	6.3	7.8	0.25	**Guanine** 
N3-	0.8	0.6	0.6	
N7-	67	11.5	85	
O^2-	0.1	3.5	-	**Cytosine** 
N3-	0.6	0.22	-	
O^2-	0.11	7.4	-	**Thymine** 
O^4-	0.4	2.5	-	
N3-	0.3	0.8	0.08	
phosphate oxygen	16	52	0.8	**Phosphate** 

3-methylcytosine, 3-ethylthymine, 3-methylthymine, 1-methylguanine and 1-methyladenine are error-prone only during RNA replication and not during *in vitro* DNA replication (reviewed by Saffhill et al., 1985). The differences observed during replication of nucleic acids containing these modified bases with DNA and RNA polymerases could be due to the differences in fidelities of the two groups of enzymes. DNA polymerases are in general very accurate in replicating DNA, with error frequencies in the range of 10^{-9} to 10^{-12} errors made per nucleotide incorporated. RNA polymerases are much less accurate. It has been known that in the presence of Mn^{2+} *E. coli* DNA polymerase I can be induced to misincorporate, producing an error spectrum similar to those given by RNA polymerases.

The N3- and N7-alkylpurines also induce mutation, but by a completely different route. These products depurinate at a rate some 10^6-10^7 times faster than normal bases as a consequence of the instability of their glycosidic bonds and the action of DNA glycosylases. The apurinic sites produced as a result of depurination prevents normal DNA replication since the incoming nucleoside triphosphate cannot form a base-pair at these lesions. In the absence of a suitable substrate, an adenine is usually inserted at an apurinic site (discussed Laval et al., 1990).

1.3 Repair of O^6-methylguanine lesions in DNA

O^6-alkylG in DNA can cause mutation and lead to lethal consequences unless it is repaired. The majority of the O^6-alkylG lesions are repaired by a group of enzymes called the O^6-alkylG-DNA alkyltransferase. The alkyltransferase repairs O^6-alkylG lesions by transferring the 6-*O*-alkyl group to a cysteine residue in its active site, restoring the DNA back to its normal state. Once it has repaired an O^6-alkylG lesion, the alkyltransferase becomes inactivated because the cysteine-acceptor site cannot be regenerated, and in this respect the alkyltransferase cannot be considered as a true enzyme. If many O^6-alkylG lesions are present, the number that will be repaired depends on the number of alkyltransferase molecules present (reviewed by Pegg & Byers, 1992).

E. coli has two alkyltransferases, one known as the ogt protein that is constitutively expressed (ca. 200 copies per cell), and another called the ada protein that is expressed only when induced as part of the adaptive response to alkylating agents (up to 10 000 copies per cell, Topal, 1988). The adaptive response is so-called because if *E.coli* is first exposed to a very low, non-toxic dose of alkylating agent, e.g. MNNG, it becomes more resistant to the cytotoxic and mutagenic effects of a subsequent higher dose of the carcinogen. In contrast, only one alkyltransferase has so far been identified in mammalian cells. The mammalian alkyltransferase does not appear to be inducible (reviewed by Pegg & Byers, 1992); the constitutive level varies from tissue to tissue but is generally higher than that in *E. coli*. The highest alkyltransferase activity is found in the liver and spleen, and the lowest in the brain and mammary glands, and this may be important in determining the susceptibility of a particular tissue to carcinogenesis by *N*-nitroso compounds (Pegg & Byers, 1992).

The ogt and ada proteins and the mammalian alkyltransferase all share a high degree of sequence homology, especially in the sequence around the cysteine-acceptor site (reviewed by Pegg & Byers, 1992). However, while both the ogt and ada proteins can repair O^6-alkylG and O^4-methylthymine, the mammalian alkyltransferase repairs only O^6-alkylG but not O^4-methylthymine. All three alkyltransferases mentioned here are able to repair 6-*O*-methyl, -ethyl, -n-propyl, and other alkyl groups, although the activity of all three decreases as the size of the adduct increases (reviewed by Pegg & Byers, 1992).

Several human tumour cell lines have been found to lack alkyltransferase activity and as a consequence, these cell lines (known as Mer- or Mex-) are very sensitive to mutation by *N*-nitroso compounds (reviewed by Pegg & Byers, 1992). The alkyltransferase should also protect cells against mutation caused by endogenously formed *N*-nitroso compounds. In a recent study

by Aquilina et al. (1992), the spontaneous mutations occurring in the *aprt* gene in two CHO cell lines, one which lacks the alkyltransferase and the other which has this repair enzyme, were compared. The results suggested that the alkyltransferase protects the CHO cells against spontaneous G:C to A:T transition mutations. A possible source of these spontaneous G:C to A:T transitions is the endogenous production of O^6-alkylG, which suggests that endogenous nitrosation could play a role in cancer.

The post-replicative mismatch system also appears to reduce the number G to A transition mutations caused by alkylating agents (Sibghat-Ullah and Day, 1992). G:T mismatches occur spontaneously as a result of the deamination of 5-methylcytosine. This is repaired by the G:T mismatch system which restores the G:T pair back to the normal G:C base-pair. Both *E. coli* and human possess this capacity to repair spontaneous G:T mismatches in which the two phosphodiesters immediately 5' and 3' of the mismatched T are broken so that the T may be removed, leaving a single-nucleotide gap that is repaired by polymerase β in mammals (discussed by Sibghat-Ullah & Day, 1992). Recent experiments by Sibghat-Ullah & Day (1992) have shown that cell free extracts from Mer- and Mer+ human cell lines were able to remove the T from an oligodeoxynucleotide duplex containing an O^6-meG:T base-pair by incision of the phosphodiesters immediately 3' and 5' to the T, just like the incisions made by the G:T mismatch repair system when it repairs a G:T mismatch. When the experiments were repeated with an oligodeoxynucleotide duplex containing an O^6-meG:C base-pair, no incision was observed.

O^6-meG is cytotoxic as well as mutagenic. The mutagenic potential of O^6-meG has been known for a long time, but its cytotoxic properties, mediated through gross chromosomal damage instead of spot mutations, have only been generally accepted recently. Cell lines which are deficient in alkyltransferase activity but are tolerant to the cytotoxic effects of O^6-meG have been isolated by selecting MNNG-resistant clones, which shows that tolerance to the cytotoxic effects of alkylating agents is not related to the repair of O^6-alkylG (reviewed by Karran and Bignami, 1992). O^6-alkylG can induce strand breaks in DNA, and this could be due to the activity of the G:T mismatch repair system which has been proposed to recognise O^6-meG:T base-pairs (Sibghat-Ullah & Day, 1992). Because the G:T mismatch repair system can recognise O^6-meG:T base-pairs and selectively remove the thymine in such a mispair, the gap left after mismatch repair will have to be filled in by some DNA polymerase. If another thymine is filled in this gap, which is not unlikely since DNA polymerase I (which is responsible for repair in *E. coli*) prefers to insert thymine rather than cytosine opposite a template O^6-meG, the resulting O^6-meG:T base-pair could be subjected to "mismatch repair" again. This repeated "mismatch repair" and insertion of a thymine in the gap left by mismatch repair, referred to as repair synthesis, is a futile process and can go on forever in theory. It has been suggested that repair synthesis could eventually lead to strand breaks in DNA which inhibit initiation of replication in the cell cycle

(reviewed by Karran & Bignami, 1992). Very recently, human and CHO cell lines which are resistant to the cytotoxic effects but not the mutagenic effects of MNNG have been reported to be deficient in G:T mismatch repair (Branch et al., 1993), which lends support to the model of methylation tolerance mentioned. These cell lines also show increased mutation frequencies because of the G:T mismatch repair deficiency. This has serious implications in the context of carcinogenesis induced by the N-nitroso compounds because in the presence of alkylating agents such as MNNG, the cells deficient in G:T mismatch repair will have increased chances of survival, and these cells are more likely to contain transition mutations due to the lack of an efficient G:T mismatch repair system which could result in a selection of transformed cells for survival by alkylating agents.

1.4 Relevance of nitrosamines to oncogene activation

Oncogene activation and chemical carcinogenesis

An important milestone in the history of cancer research is the discovery of oncogenes, as this gave insight into the genetic events in carcinogenesis. Oncogenes were first discovered in retroviruses when it was noticed that some retroviruses can induce cancer in animals upon infection, or confer transforming properties on infected cells in culture. The genes that were responsible for inducing cancer were given the name 'oncogene'. It was later shown that the retroviral oncogenes have counterparts in cells of higher vertebrates, and so the two groups of oncogenes are called c-*onc* and v-*onc* respectively to identify the origin of the oncogenes. It soon transpired that the retroviral oncogenes actually came from cellular origins: they were acquired by the retroviruses when they infected the cells of higher vertebrates.

Obviously oncogenes cannot be oncogenic in normal cells, and it was noticed that some retroviral oncogenes differ from their cellular counterparts only by point mutations. To avoid confusion, the non-oncogenic forms of these cellular genes prior to activation are called proto-oncogenes. Generally, proteins coded for by proto-oncogenes are usually involved in signal transduction pathways for cell division and/or cell differentiation, and are therefore in a good position to contribute towards the neoplastic phenotype when altered in either their function or their expression. Point mutation is not the only mechanism whereby a proto-oncogene can be activated; deletion of DNA and gene rearrangement can also act as activating events for proto-oncogenes.

In discussions about cancer cells, a normal cell is said to be transformed when it acquires properties conferred upon them by the introduction of activated oncogenes. To date, the most effective transforming combinations of oncogenes involve the *ras* family of nuclear oncogenes. The *ras* gene family comprises the Harvey, Kirsten and N-*ras* genes; the proteins encoded by *ras* are regarded as signal transducers that are responsible for relaying signals from growth factor receptors to the interior of the cell. Activated *ras* has been found in many human and animal tumours, and most of the *ras*-activation results from point mutations at one of the several possible positions within the coding sequence, notably codons 12, 13 and 61 of the *ras* gene (Balmain & Brown, 1988). *Ras* gene mutations can be found in a variety of tumour types, although the incidence varies greatly as measured using the NIH/3T3 transformation assay (Table 3). Overall, about 15% of all the human tumours examined contained activated-*ras*, suggesting an important role of these genes in human carcinogenesis.

Table 3. Some *ras* mutations in chemically induced animal tumours (from Balmain & Brown, 1988; Topal, 1988 and Belinsky et al., 1989).

Carcinogen	Tumour	Oncogene	Codon	Mutation	Occurrence
	RAT				
MNU	mammary	H-*ras*	codon 12	GGA to GAA	61/61
NDMA	oesophagus	H-*ras*	codon 12	GGA to GAA	18/18
DMBA	mammary	H-*ras*	codon 61	CAA to CNA	
	MOUSE				
MNU	thymus	K-*ras*	codon 12	GGT to GAT	5/18
MNU	thymus	K-*ras*	codon 13	GGC to GAC	1/18
MNU	T-lymphoma	K-, N-*ras*	codon 12	GGA to GAA	85%
MNNG	skin papilloma	H-*ras*	codon 12	GGA to GAA	11/11
NDEA	hepatoma	H-*ras*	codon 61	CAA to AAA	6/16
NDEA	hepatoma	H-*ras*	codon 61	CAA to CTA	4/16
NDEA	hepatoma	H-*ras*	codon 61	CAA to CGA	3/16
NNK	A/J lung	K-*ras*	codon 12	GGT to GAA	7/11
NNK	A/J lung	K-*ras*	codon 61	CAA to CGA	2/11
NDMA	A/J lung	K-*ras*	codon 12	GGT to GAA	7/10
NDMA	A/J lung	K-*ras*	codon 61	CAA to CGA	3/10
DMBA	skin carcinoma	H-*ras*	codon 61	CAA to CTA	45/45

Ras-activation by *N*-nitroso compounds

Cancer involves several stages, at the least initiation and promotion. Initiation is irreversible: an epidermal cell once initiated can develop into a tumour regardless of whether promotion is begun immediately or forty weeks after initiation. This must mean that initiation involves inheritable genetic changes, for in the course of forty weeks the epidermis would have undergone many cell divisions.

N-nitrosamines are complete carcinogens capable of producing cancer without further treatment of the animal. For example, a single intravenous dose of *N*-methyl-*N*-nitrosourea, which has a half life of about 60 seconds, will induce benign or malignant tumours in every animal. To some extent the mechanism of initiation by *N*-nitroso compounds is understood. Nitrosamines are metabolized to alkylating species which react with DNA, and initiation by *N*-

nitroso compounds is probably a consequence of mutations caused by alkylation at the O^6-position of guanine and O^4-position of thymine by metabolites of the nitroso compounds.

Tumours induced in the skin, breast, liver and other tissues in both experimental animals and humans often contain mutant *ras* genes that are capable of transforming NIH-3T3 cells (Table 3). The normal and oncogenic forms of the *ras*-encoded p21 protein are often due to single point mutations in codons 12, 13 and 61 of *ras* (Reddy et al., 1982; Tabin et al., 1982), and a link between carcinogenic alkylating agents and *ras* mutation was established in 1983 when Sukumar et al. induced mammary tumours in rats by NMU and showed that the malignancy was caused by single G to A point mutations in codon 12 of the Ha-*ras*-1 locus.

Other carcinogens were similarly used to induce tumours and the resulting tumours analysed for *ras*-activation. It was found that the activating mutation always reflects the known DNA-binding and mutagenic characteristics of each carcinogen (Balmain & Brown, 1988; Belinsky et al., 1989). The majority of the *ras* genes found in tumours induced by *N*-nitroso compounds are activated by the same G^{35} to A^{35} transition mutation in codon 12. This is in agreement with the theory that the G^{35} is alkylated by a metabolite of *N*-nitroso compounds to form O^6-alkylG, which subsequently leads to transition mutations during DNA replication because of its miscoding properties.

O^4-alkylT also appears to act in the same way as O^6-alkylG. In A/J mouse tumours induced by NNK or dimethylnitrosamine, a small number of activated-*ras* genes were found to involve an A to G transition mutation in codon 61 (Belinsky et al., 1989). Since, O^4-alkylT prefers to form a base-pair with guanine rather than adenine, this result implicates the formation of O^4-alkylT:G base-pairs during DNA replication in the activation of *ras*.

Activation of *ras* was originally thought to occur at a late stage of carcinogenesis, as the NIH/3T3 cells used in transformation experiments to test the oncogenic potential of *ras* were already in a 'premalignant' state and as such required only one additional (late) event to progress to malignancy. However activated *ras* has been observed in some well-defined pre-cancerous lesions, indicating that the activated *ras* may play a role in the early stages of carcinogenesis (Bos, 1989). This has now been confirmed in mouse mammary carcinogenesis induced by NMU (Kumar et al., 1990 and Miyamoto, 1990). As early as two weeks after NMU treatment, activated *ras* oncogene was detected in what appeared to be normal mammary tissues in the mouse, and the onset of neoplasia was observed two months after this.

1.5 The mutagenic properties of alkylated bases

N-nitroso compounds are carcinogenic because they can be activated to form active alkylating species that alkylate DNA, producing the promutagenic lesions O^6-alkylG and O^4-alkylT. These adducts are promutagenic because they direct the misincorporation of thymine and guanine respectively during DNA replication, giving rise to transition mutations. How do these transition mutations occur? Misincorporation during DNA synthesis depends on two components: the structure of the DNA and how the DNA polymerases decide which nucelotide to incorporate opposite an alkylated base in the template DNA. To understand misincorporation opposite an alkylated base during DNA replication we must first look at the base-pairing properties of complementary and non-complementary bases (this section), then we must consider the DNA polymerases and the process of polymerisation and see how these properties affect the occurrence of transition mutations (next section).

The base-pairing properties of complementary and non-complementary bases and DNA replication

The acid-base behaviour of a base determines the tautomeric structure and the ability of a base to accept and donate protons in the formation of H-bonds, which is the key feature of base:base interactions. The pK_a values of the proton donors and acceptors on these bases are shown in Table 4. Species with high pK_a values (e.g. N1 of guanine and N3 of thymine/uracil) tend to be protonated and are thus able to donate protons in the formation of H-bonds. On the other hand, species with low pK_a values (e.g. N1 of adenine, N3 of cytosine and the ring nitrogens) are not protonated and are therefore proton acceptors in the formation of H-bonds.

Table 4. pK_a values for the normal bases occurring in nucelotides.

Base of 5' nucleotide	Site of protonation	pK_a
Adenosine	N1	3.88
Cytosine	N3	4.56
Guanine	N7	3.6
Guanine	N1	10.00
Thymine	N3	10.47
Uracil	N3	10.06

H-bonding affinities reflect the charge densities. The amino protons (e.g. 6-amino of A, 2-amino of G and 4-amino of C) have an average charge density of + 0.22 e and are very good proton donors. The amide-like protons (e.g. N1 of G and N3 of T/U) are slightly less acidic, with an average charge density of + 0.19 e; they too, are good H-bond donors. The ring nitrogens and the electron pairs of keto oxygens are basic in nature and provide good proton acceptors. Therefore electrophilic attacks by protons or alkylating agents occur primarily at the ring nitrogens and the keto oxygens.

Chemical modifications may also directly influence the pK_a values of the proton donors/acceptors on the bases. The analogues of thymine are a good example. Thymine and uracil are structurally identical in all respects except the 5-substituent. Thymine has a 5-methyl substituent while uracil has a proton on C5. Uracil is quite acidic, with a N3 pK_a of 9.4, but the electron-withdrawing 5-methyl substituent of thymine makes it slightly more acidic than uracil (N3 pK_a becomes 9.9). If this electron-withdrawing 5-methyl group is replaced by an electron-donating bromine, as in 5-bromouracil, then the base becomes quite basic, lowering the pK_a of N3 to 7.8.

The amide-like protons on the bases can migrate from the nitrogens to the neighbouring keto oxygens or other free nitrogens within the same molecule (Fig.1.4). This phenomenon is known as tautomerisation. A change in the position of H-atoms e.g. from N1 to the 6-*O*-position of guanine, can have profound consequences because it alters the H-bonding characteristics of the base; in this example a keto group with acceptor properties transforms into an enol donor group. Similarly, if a proton from the 6-amino group of adenine migrates to N1, then a proton-accepting imino group is formed at the 6-position of adenine. However, if this imino proton is rotated, it can also serve as a donor (Figure 1.4). By tautomeric changes, uracil and guanine in the enol form can simulate cytosine and adenine respectively, and cytosine and adenine in the imino form may substitute for uracil and guanine (Fig. 1.4). Since the formation of a complementary base-pair is much dependent on the proton acceptor/donor pattern of each base, keto-enol and amino-imino tautomerisation is likely to exert a great influence over the formation of base-pairs. Fortunately, under physiological conditions, the keto and amino tautomers predominate (> 99.99%), which limits the number of different base-pairs formed under such circumstances. This tautomeric equilibrium can be altered by radical changes in pH or chemical modifications of the bases.

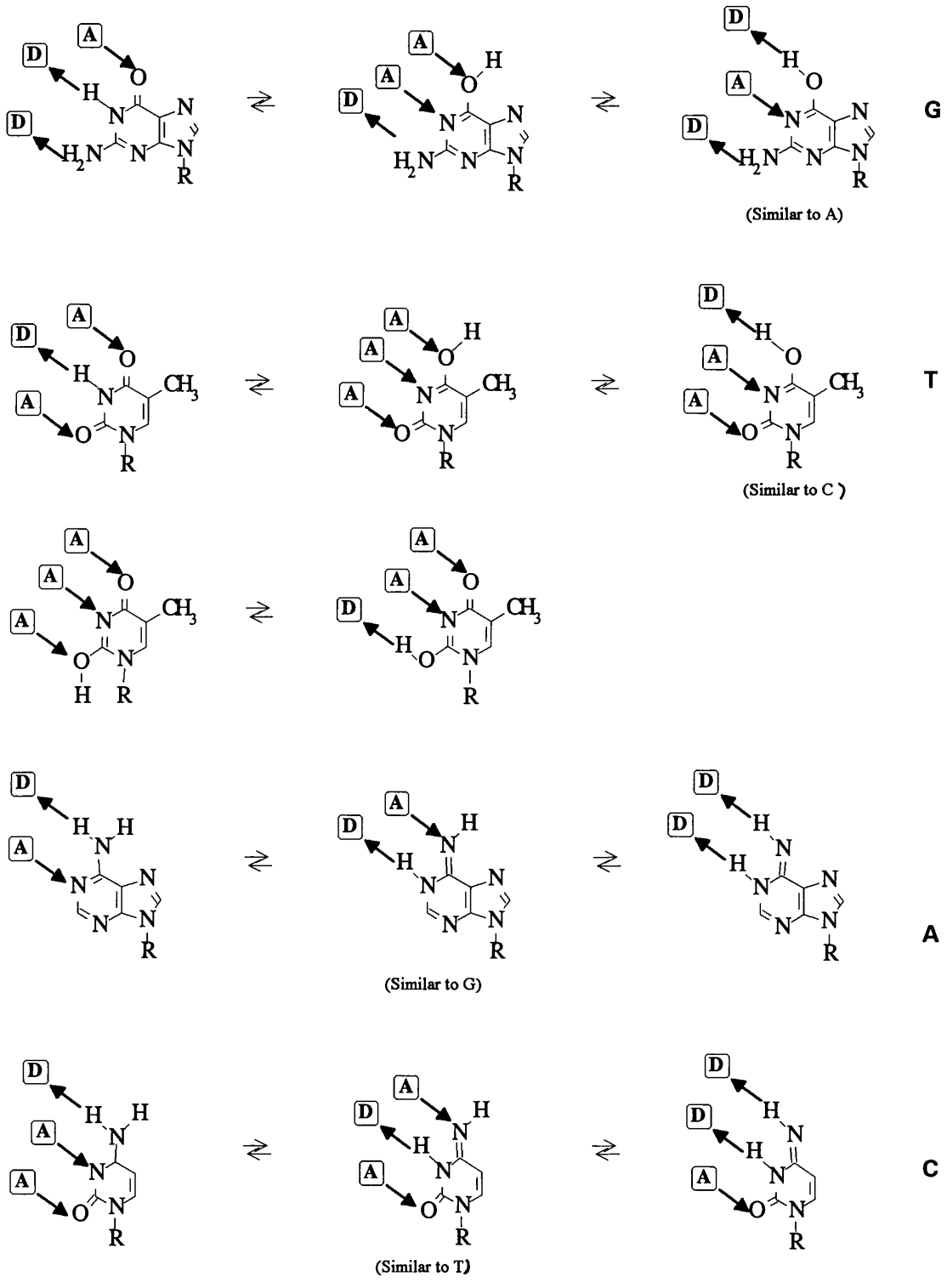


Figure 1.4. Keto-enol and amino-imino tautomerism in nucleoside bases. Arrows denoted A and D symbolize acceptor and donor sites for H-bonding. In the enol forms, the H-bonding patterns of G becomes equivalent to A, and T to C. In the imino forms, the H-bonding patterns of A is equivalent to G, and C to T. The situation changes, however, if the imino or enol groups rotate, giving rise to a diversity of H-bonding possibilities (taken from Saenger, 1984).

The Watson-Crick concept of complementarity of base-pairing

In 1953, Watson and Crick described the three-dimensional structure of DNA and proposed a mechanism for its biological replication based on the complementarity of base-pairing (Watson & Crick, 1953a, b). This was an important milestone in the history of biology, as it paved the way for molecular biology and enabled us to understand life in molecular terms, and the authors truly deserved the Nobel prize they were awarded for this notable contribution towards our understanding of the life sciences. The structures of the Watson-Crick A:T and G:C base-pairs are shown in Fig. 1.5. There are two H-bonds in an A:T pair and three in a G:C pair. The specific formation of H-bonds between complementary bases provides a basis for DNA replication.

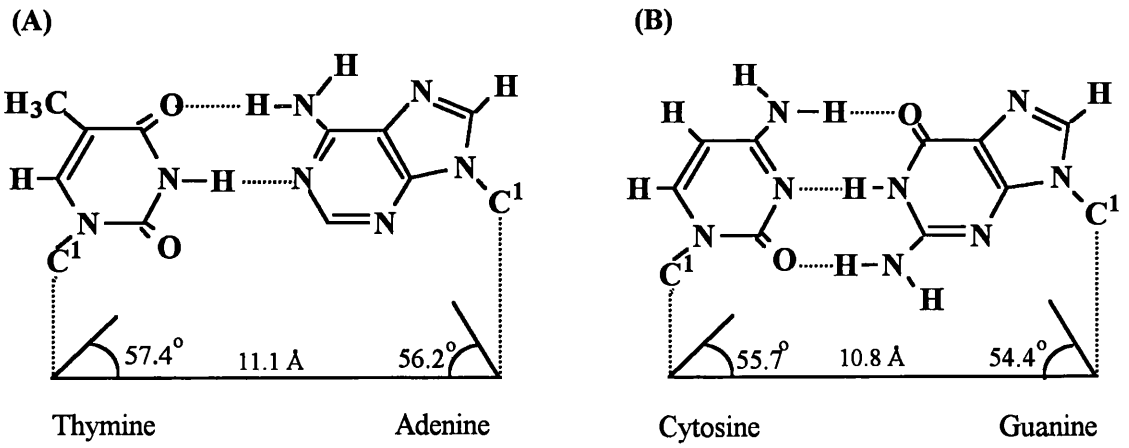


Figure 1.5. (A) and (B): Watson-Crick base-pairing for T:A and C:G respectively.

An important feature of a planar Watson-Crick base-pair is that N1 of the purine is always juxtaposed to N3 of the pyrimidine; the length of the H-bonds between the two bases are on average 2.8 to 2.95 Å apart. This geometry gives a C1'-C1' distance of 10.90 ± 0.15 Å with an angle of 55 ± 2° between the two glycosidic bonds for both the A:T and the G:C base-pairs.

Because the C1'-C1' distance and the angle between the two glycosidic bonds for all the Watson-Crick base-pairs are almost the same, the four base-pair combinations, A:T, T:A, G:C, and C:G can all be built into the same regular frame work of the DNA double helix (Fig.

1.5). This gives a considerable degree of uniformity to the sugar-phosphate backbone of the DNA double helix. If the C1'-C1' distance or the angle between the two glycosidic bonds of a base-pair is significantly different from those of a Watson-Crick base-pair, the double helix will either fail to accommodate such a base-pair or its sugar-phosphate backbone will have to be distorted in order to accommodate this base-pair. Such structural differences in DNA are potentially very significant because they can affect protein-DNA interactions, and in the context of DNA replication, the structural uniformity of the sugar-phosphate backbone may provide a basis for maintaining the fidelity of DNA synthesis. More detailed discussion is presented in a later section.

The origin of spontaneous mutations

If we accept that accurate replication of DNA is principally dependent upon the Watson-Crick concept of base complementarity, it then follows that substitution mutations occur as consequences of mispairing events. Two kinds of substitution mutations are found; a transition involves the replacement of a purine by another purine or a pyrimidine by a different pyrimidine, while in a transversion, a purine replaces a pyrimidine or vice versa. In the present work, attention is focused on transition mutations only.

Once the complementarity of base-pairing was accepted as the basis of biological DNA replication, many people were ready to believe that the formation of H-bonds between complementary base-pairs is the dominant basis for maintaining high fidelity during DNA replication. Thus, there was general agreement that thermodynamic parameters govern the accuracy of biological DNA synthesis. Based on this, the tautomer hypothesis was put forward by Watson and Crick (1953b) as an explanation for the spontaneous occurrence of transition mutations: some H-atoms on each of the nucleic acid bases can tautomerize, giving rise to keto or enol forms and amino or imino forms (Figure 1.4). As discussed above, the keto and amino tautomers predominate, with the rare tautomers present in less than 0.01% of the total base population. The predominant tautomers form normal Watson-Crick base-pairs with the complementary bases, while the rare tautomers can form non-Watson-Crick base-pairs and thus give rise to spontaneous transition mutations.

While plausible, the suggestion by Watson and Crick that substitution mutations arise as a result of the existence of rare tautomeric forms of bases is unsatisfactory mainly because it cannot explain the frequencies of spontaneous substitution mutation *in vivo* (approx. 10^{-9}-10^{-12} error per nucleotide incorporated) simply from the frequencies of unfavoured tautomers in

solution (estimated at 10^{-4}-10^{-5}). One has to remember that DNA replication is a biological process, and that while the physical properties and the isomeric equilibria of nucleotides are important in the formation of base-pairs, the role played by DNA polymerases in maintaining the fidelity of DNA synthesis must not be overlooked. Based on this, Topal and Fresco put forward a general base-pairing hypothesis to explain the occurrence of substitution mutations (1976), which proposes that the level at which mispairing occurs depends on two things: (1) the equilibrium constants for the tautomerisation of the keto-enol and amino-imino forms of the bases, and (2) proofreading during DNA replication which limits the level of mispairing to levels observed *in vitro*.

In many biological systems, DNA replication involves two steps, an incorporation step and a proofreading step that is designed to enhance fidelity. During incorporation, the DNA polymerase first binds the primer-template DNA, and then the incoming nucleoside triphosphate. Under the geometrical constraints imposed by these binding events, specific H-bonds are formed between the incoming nucleotide and the template base if the two bases are complementary. A phosphodiester bond is then formed between the α-phosphate of the incoming nucleotide and the 3'-hydroxyl of the growing chain. In this way, the polymerase incorporates any nucleotide that is capable of forming a complementary base-pair with Watson-Crick geometry and discriminates against all others. Most of the base-pairs formed will be Watson-Crick base-pairs as 99.99 % of the isomers are in the keto or amino form. The minor tautomers can also form base-pairs that are compatible with the steric constraints of a double helix, but because the H-bonding properties of the minor tautomers are altered, these base-pairs are mismatches or mispairs (Fig. 1.6). Topal and Fresco suggested four possible complementary purine-pyrimidine pairs involving rare tautomeric forms in which the Watson-Crick geometry is still retained with N1 of the purine juxtaposed to N3 of the pyrimidine and the glycosyl bond separation distance remains unchanged. During DNA replication, such base-pairs involving rare tautomeric forms can thus be formed, resulting in the formation of a mismatch during DNA replication. At this stage, the level of mispairing can then be estimated as follows,

level of mispairing = frequency of incoming dNTP being an unfavoured tautomer +
frequency of template base being an unfavoured monomer
$= 10^{-4} + 10^{-4}$

Upon formation of the phosphodiester bond but before the next nucleoside triphosphate binds to the polymerase-DNA complex, there is an opportunity for the removal of any mismatches in the so-called proofreading step. This is catalysed by a 3'→5' exonuclease, which may be a part of the polymerase enzyme, or it may be entirely separate. During this proofreading step, the newly synthesized base-pair becomes re-exposed to solvent so that there is renewed opportunity for tautomerisation to take place. Thus the total fidelity is estimated as the product

of the frequency of mispairing during synthesis and during proofreading, i.e. 10^{-8}-10^{-10} (Topal & Fresco, 1976). This theory can be tested with systems lacking a 3'→5' exonuclease. In such systems, according to Topal and Fresco, the rate of transition mutations should be that expected from the incorporation step alone, i.e. around 10^{-4}. This was shown to be indeed the case (Topal & Fresco, 1976).

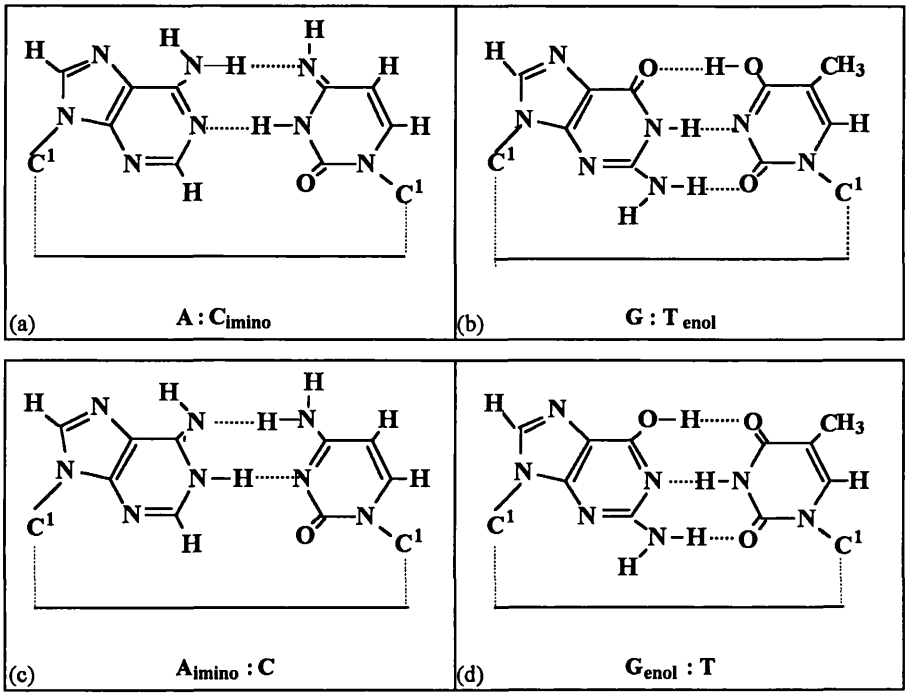


Figure 1.6. The four possible mismatched base-pairs that involve rare tautomeric forms. Tautomerism involving only one proton transfer gives rise to the following: (a) an $A{=}C_{imino}$ base-pair; (b) a $G{\equiv}T_{enol}$ base-pair; (c) an $A_{imino}{=}C$ base-pair; and (d) a $G_{enol}{\equiv}T$ base-pair.

The tautomer hypothesis is a beautiful theory and is widely accepted, but it is very difficult to prove or disprove. A mismatched base in DNA may have been incorporated as a rare tautomer, but this rare tautomer could very quickly revert to the favoured tautomer during the so- called 'breathing of DNA', and as a result escape detection. To this day, the tautomer hypothesis remains the preferred explanation for the occurrence of transition mutations, even though there is evidence that some transitions occur via other means. The mutagenic phenomenon of 5-bromouracil (5-BrU) provides an interesting example. Although uracil does not occur naturally in DNA, it is structurally very similar to thymine, the only difference being the absence of a methyl substituent at C5 in uracil, and as this group is not involved in H-bonding between Watson-Crick base-pairs, the A:U base-pair is very similar to the A:T base-pair. 5-bromodeoxyuracil (5-BrdU) has a 5-bromo group and this turns it into a potent mutagen.

Compelling evidence that 5-BrdU induces mutation came from the studies of Benzer and Freese in 1958 which compared the rates of spontaneous mutation and mutation in the presence of 5-BrdU in phage T4 and found the latter to be significantly higher (1000 fold higher) than the rate of spontaneous mutation. A year later, Freese showed that unlike spontaneous mutations, mutations induced by 5-BrdU can be induced to revert, which suggested that 5-BrdU causes mainly transition mutations (1959).

Several models of mispairing involving 5-BrdU have been proposed (Fig. 1.7). The first model was based firmly on the tautomer hypothesis. In this model, the rare enol form of BrdU (or thymine) forms a base-pair which retains Watson-Crick geometry with guanine but not with adenine (Fig. 1.7(b)). Spectroscopic investigations (Katritzky & Waring, 1962) showed that the presence of bromine on the 5-position of uracil enhances enolization. Compared with uracil, 5-BrdU has approximately ten times more of the enol form. Because of the increased proportion of the enol tautomer, there is a parallel increase in the occurrence of the 5-BrdU:G base-pairs that would subsequently lead to transition mutations (Topal & Fresco, 1976). However, due to the low abundance of the enol isomer, the tautomer of BrdU mispairing is difficult to test.

In a second model, the ionization model suggested by Lawley and Brookes (1962), the electronegative bromine atom in 5-BrdU induces ionization at N3, so that at physiological pH a significant proportion of 5-BrdU exists as ionized species. Using spectroscopic methods, these authors found the pK_a of ionization at N3 for 5-BrdU and thymine to be 8.1 and 9.8 respectively. At neutral pH, 7.4% of 5-BrdU, compared with 0.16% of thymine, is ionized. The anionic form of 5-BrdU is thought to form a base-pair that retains Watson-Crick geometry with guanine using two H- bonds (Fig. 1.7c)). Support for this model came from the work of Driggers & Beattie (1988), who examined the effects of increasing pH on the rates of formation of 5-BrdU:G base-pairs during primer elongation by several DNA polymerases; they found that when 5-BrdU was present in the template strand, increasing the pH facilitated the formation of 5-BrdU:G base-pairs by DNA polymerases, consistent with the ionization model for 5-BrdU:G mispairing.

A third model, which is a less extreme version of the ionization model, was proposed when it transpired that the electron-withdrawing 5-bromo substituent of 5-BrdU may diminish or even abolish the capacity of O^4 of 5-BrdU to act as a proton acceptor, and consequently, O^2 is thought to be the preferred proton acceptor. This enables 5-BrdU to form a wobble base-pair with guanine (Fig. 1.7(d)) without evoking the tautomer hypothesis. In thymine, the 5-methyl group is an electron-donating group, which pushes electrons towards O^4, making it a good proton-acceptor (reviewed by Saenger, 1984).

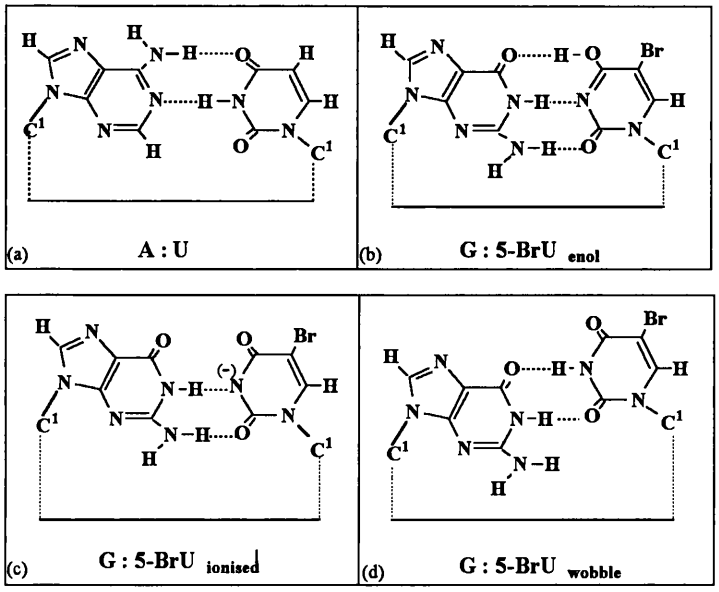


Figure 1.7. Schematic diagrams for the G:5-BrU base-pair.
(a) a normal A:U Watson-Crick base-pair;
(b) a G:5-BrU base-pair involving a 5-BrU enol-tautomer;
(c) a G:5-BrU base-pair involving an ionised 5-BrU according to Lawley & Brookes (1962); and
(d) a G:5-BrU wobble base-pair.

The fourth model is the base-stacking model, where the 5-BrdU:G base-pair is again thought to adopt a wobble conformation (Fig. 1.7(d)), similar in structure to the U:G base-pair found in several tRNA crystal structures, and the G:T base-pair studied in several oligonucleotides both in the solid state and in NMR solution studies (discussed by Brown et al., 1986). Stacking interactions are mainly due to dispersion and dipole-induced dipole forces and hydrophobic bonding, with the exocyclic groups (C=O and C-NH_2) or ring nitrogens located over the heterocycle of the adjacent base (Saenger, 1984). The 5-BrdU:G base-pair is believed to be more stable than the T:G mispair because of strengthened base-stacking interactions conferred by the 5-bromo substituent of 5-BrdU. The observed closeness between the bromine atom and neighbouring pyrimidine rings in crystals of oligonucleotides containing 5-BrdU gives support to this model (Sternglatz & Bugg, 1975). In addition, it is well known that 5-BrdU-induced mutations show a strong tendency to occur in certain 'hot spots' of particular nucleotide sequence. Although nearest neighbour stacking interactions do not play a primary role in base selection during DNA synthesis, they affect the occurrence of mispairing events to some extent. During polymerisation, it has been shown by Driggers & Beattie (1988) that although increasing the pH enhances the rates of incorporation of guanine opposite template 5-BrdU by DNA polymerases, it has no effects on the rates of misincorporation of 5-BrdU opposite template guanine residues. This could be explained by the fact that stabilization by stacking of a template 5-BrdU is stronger than that for an incoming 5-BrdU nucleotide. However, for DNA polymerases with an associated 3'→5' exonuclease, the observed sequence-dependent mispairing by 5-BrdU may perhaps be a result of sequence-dependent activity of the 3'→5' exonuclease.

Although there is some experimental basis for each of these proposals, conclusive evidence is still lacking. Many of the studies of base-pair mismatches using X-ray crystallography, NMR and spectroscopy use pre-formed mispairs and as such, can only provide information about fixed structures with the lowest free energy. On the other hand, mutagenesis is a biological process, and in order to understand how mispairs are incorporated, it is necessary to examine the dynamics of the incorporation process, to look into the interplay of the chemical, physical as well as the biological factors involved in DNA replication. We need to understand how H-bonding is affected by modifications of the functional groups on the bases, the geometrical constraints imposed on base-pairing and the mechanisms and kinetics of polymerisation and proofreading or even repairs, before we can safely decide which factor(s) is vital in maintaining fidelity during DNA synthesis.

Base-pairing properties of O^6-alkylguanine and O^4-alkylthymine: is H-bonding a crucial determinant of fidelity of DNA synthesis ?

O^6-alkylation of guanine residues in DNA can induce mutations through the formation of O^6-alkylG:T mispairs during DNA replication. O^6-alkylguanine can be regarded as a purine 'frozen' in the enol form. According to the tautomer hypothesis, O^6-alkylG can form a mismatched base-pair with thymine, in the same way that an enol tautomer of guanine can form a base-pair with thymine (Fig. 1.8(b)). Such an O^6-alkylG:T base-pair will still retain Watson-Crick geometry and involves two H-bonds, from N1 of the alkylG to N3 of thymine and from 2-amino of O^6-alkylG to O^2 of thymine. The O^6-alkylG:C base-pair was originally thought to have only one H-bond (Fig. 1.8(d)), which made it thermodynamically less stable than the O^6-alkylG:T base-pair (Fig. 1.8(c)). These structures appeared to provide a satisfactory explanation for the preferential incorporation of thymine over cytosine opposite O^6-alkylG during the replication of DNA (see Goth & Rajewsky, 1974b). For more than ten years, the miscoding properties of O^6-alkylG was believed to be the result of the formation of a stable mispair between O^6-alkylG and thymine. A similar scheme was conceived for base-pairs involving O^4-alkylT (Fig. 1.9c & d).

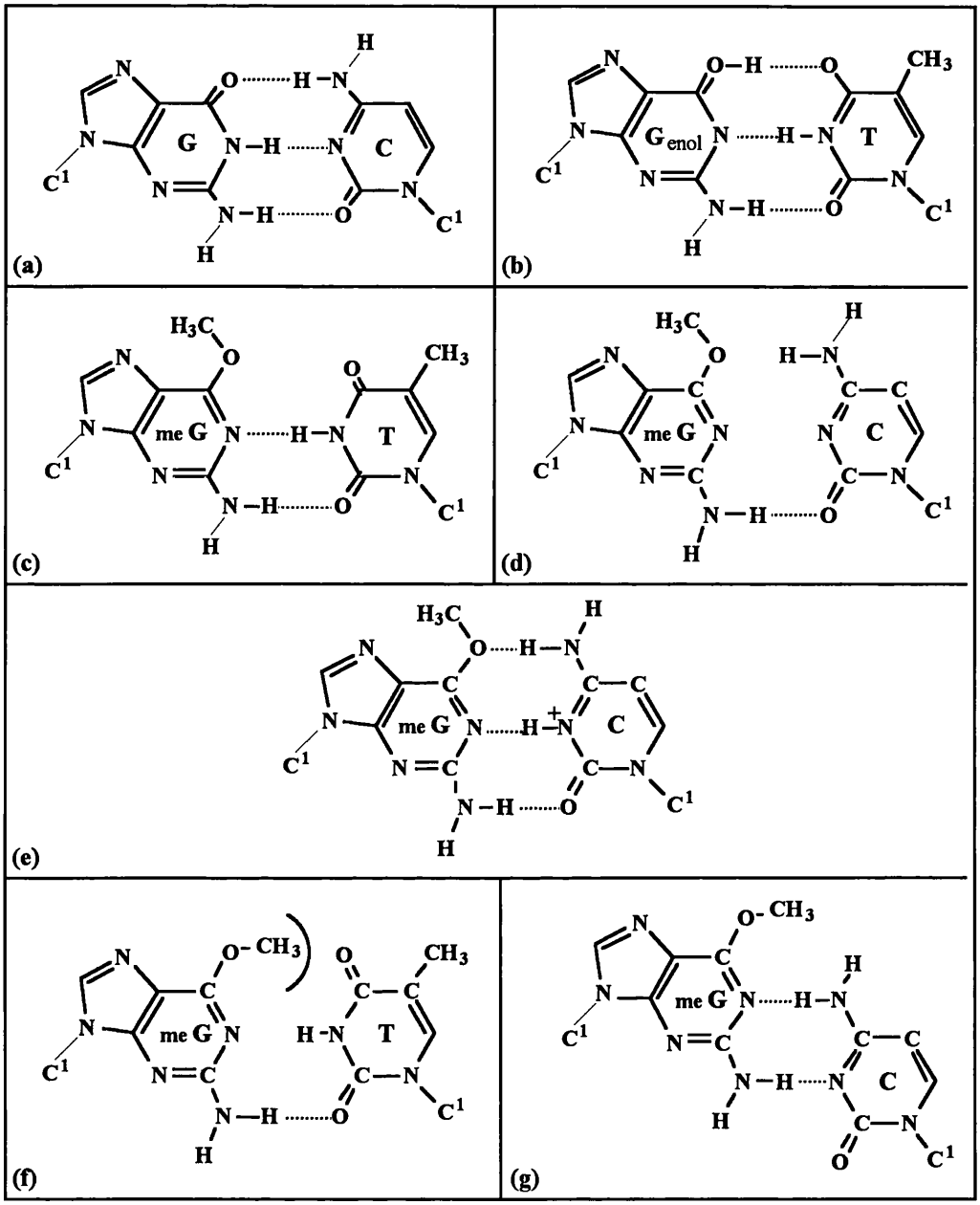


Figure 1.8. Postulated structures for the base-pairs between O^6-meG and thymine or cytosine. (a) a normal G:C Watson-Crick base-pair; (b) a G_{enol}:T base-pair; (c) the original meG:T base-pair; (d) the original meG:C base-pair; (e) the meG:C base-pair proposed by Williams & Shaw (1987) which involves a protonated C; (f) and (g) the meG:T and meG:C base-pairs predicted by Kalnik et al. (1989a, b) based on 2-D NMR studies.

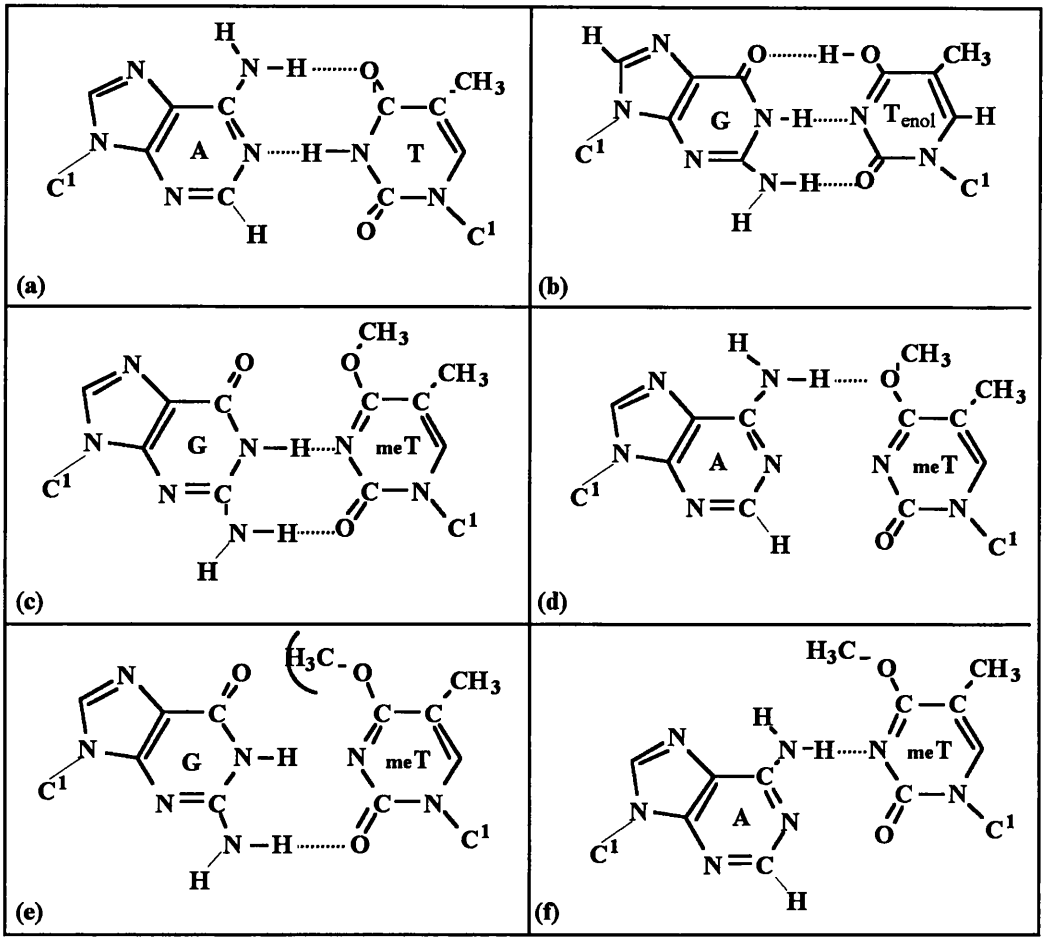


Figure 1.9. Postulated structures for the base-pairs between O^4-methylthymine and guanine or adenine. (a) a normal A:T Watson-Crick base-pair; (b) a G:T_{enol} base-pair; (c) and (d) the G:meT and A:meT base-pairs proposed by Lawley (1984); (e) and (f) the G:meT and A:meT base-pairs predicted by Kalnik et al. (1988a, b) based on 2-D NMR studies.

A crucial aspect in the base-pairing properties of O^6-alkylG is the conformation of the alkyl group at O^6. This O^6-alkyl group lies in the plane of the purine ring, either *syn* or *anti* to N1 (Fig. 1.10). The original O^6-alkylG:T structure requires the O^6-alkyl group to project into the major groove, i.e. the alkyl group is in the *anti* conformation to N1 (Fig. 1.8(c)). As such, the alkyl group does not interfere with the formation of H-bonds between the normal proton-donor and acceptor sites used for base-pairing. If the *O*-alkyl group adopts a *syn* conformation, then it would impose steric constraints that may very well disrupt the H-bonding potential. The same argument applies to base-pairs involving O^4-alkylT; the *anti*, but not the *syn*, orientation of the O^4-alkyl group permits the formation of H-bonds between the acceptor and donor sites of the bases (Fig. 1.10).

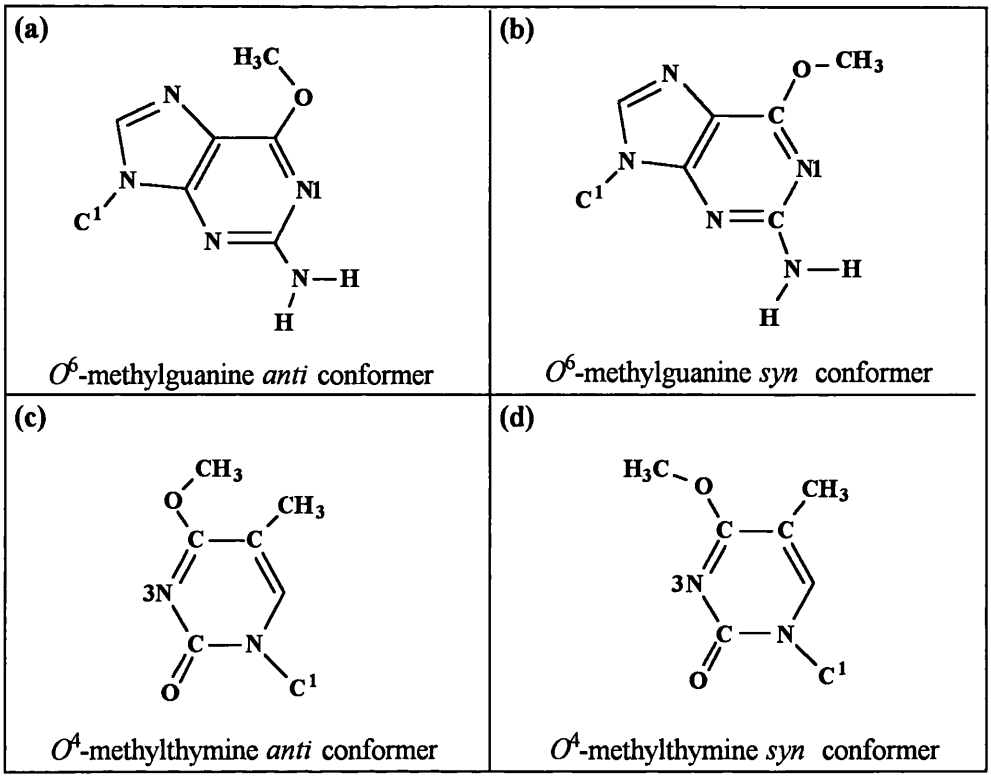


Figure 1.10. Schematic diagrams showing the *anti* and *syn* conformations of the methoxy group in O^6-meG and O^4-methylthymine. (a) and (b) the respective *anti* and *syn* conformation of the 6-*O*-methyl group of O^6-meG; and (c) and (d) the respective *anti* and *syn* conformation of the 4-*O*-methyl group of O^4-methylthymine.

Because of its importance, the question of the conformation of the alkyl group in O^6-alkylG has been considered by many authors. Several methods have been used to resolve this question, and they can be generally grouped into theoretical, NMR and X-ray crystallography studies. The first studies were theoretical studies and the initial calculations showed the *anti* conformation of the 6-*O*-alkyl group to be more stable than the *syn* conformation, with a difference of 2 kcal/mol; rotation about the C6-O^6 bond could occur but this would require an energy input of approximately 2.5 kcal/mol (Psoda, 1981 and Pohorille & Loew, 1987). This is in excellent agreement with the original O^6-meG base-pairing schemes which require the 6-*O*-methyl group of O^6-meG to adopt an *anti* conformation with respect to N1. However, a subsequent theoretical study by Pedersen et al. (1988) suggested that the opposite might be true. Using quantum and molecular mechanical calculations, these authors found the *syn*-, rather than the *anti*-, conformer of the 9-methyl derivative of O^6-meG to be lower in energy.

Results of X-ray crystallography studies by several groups also gave conflicting results. Crystals of O^6-methylguanosine gave X-ray diffraction data which suggested that the 6-

O-methyl group in O^6-methylguanosine lies preferentially *syn* to N1 (Parthasarathy & Fridey, 1986). The results also suggested that the C6-O^6 bond is too short to allow for rotation and therefore remains in the plane of the base. Two years later (1988), Yamagata et al. confirmed the observation that the 6-*O*-methyl group in O^6-methylguanosine, O^6-methyldeoxyguanosine and the 9-methyl derivative of O^6-meG is actually *syn* to N1 and contained in the same plane as the base.

At this point, virtually all of the crystal structures of the O^6-meG derivatives and calculations suggest that the 6-*O*-alkyl group is *syn* to N1. This is not in the conformation necessary for the original O^6-alkylG base-pairing structures. The only way the authors of these papers could escape the consequences of their calculations, that is to say that the original O^6-alkylG base-pairing structures were wrong, was to suggest that their results are not applicable to O^6-meG in a DNA duplex. Attempts were made to reconcile the conflicting views by the proposal that in the isolated O^6-alkylG molecule, the 6-*O*-alkyl group is *syn* and periplanar to N1, but when present in a DNA duplex, it rotates from the energetically more stable *syn* conformation to the less favourable *anti* conformation, and although there is restricted rotation about the C6-O^6 bond due to its short bond length (Parthasarathy & Fridey, 1986; Yamagata, 1988), the energy required to overcome this rotation (estimated at 2 to 2.7 kcal/mol; Parthasarathy & Fridey, 1986; Yamagata, 1988) could come from interactions with the DNA polymerase (Parthasarathy & Fridey, 1987) or from the formation of H-bonds (Yamagata et al., 1988).

Pedersen et al. (1988) also reported similar findings. From calculations performed on 9-methyl-O^6-meG, they found the *syn*-conformation of the 6-*O*-methyl group to be energetically more stable and estimated the energy barrier to rotation by the C6-O^6 bond at a level higher than that reported by Parthasarathy & Fridey (1986) and Yamagata (1988), at 7.93 or 9.19 kcal/mol depending on the method of calculation used. This makes the rotation about the C6-O^6 bond in an isolated O^6-meG molecule energetically expensive. But in a DNA duplex, the situation is different. Based on calculations performed on a pentamer duplex containing an O^6-meG base-paired with C or T, they concluded that an *anti*-conformation of the 6-*O*-methyl group in O^6-meG is energetically more stable, and that O^6-meG:T and O^6-meG:C base-pairs formed with the *anti*-conformers are 4-5 kcal/mol more stable than the same base-pairs formed with the *syn*-conformers. On the other hand, a more recent study which also used molecular modelling techniques to estimate the energies of the *syn*- and *anti*-conformers of O^6-meG in duplex DNA predicted that the two conformers will have similar energies (the difference is less than 1 kcal/mol; Loechler, 1991). This would suggest that it is energetically possible for the 6-*O*-alkyl group to rotate from a *syn*- to an *anti*-conformation within duplex DNA.

A similar situation was found with O^4-alkylT. X-ray crystallographic studies were also carried out on O^4-methylthymine and the 4-*O*-methyl group was shown to adopt a *syn* conformation (Brennan et al., 1986). A *syn* orientation of the 4-*O*-methyl group would prevent the formation of H-bonds between O^4-methylthymine and thymine/cytosine, and so the authors suggested that the 4-*O*-methyl group in O^4-methylthymine rotates from the *syn* to the *anti* conformation when the alkylated base is inserted into a DNA double helix. The energy required for this rotation is compensated by H-bond formation and improved base-stacking interactions. In a recent study by Cruzeiro-Hansson et al. (1992), the molecular dynamics of O^4-methylthymine in DNA was simulated, and the simulation showed the 4-*O*-methyl group of a O^4-methylthymine in DNA to be in a *syn*-conformation relative to N3.

Other evidence also challenged the original O^6-alkylG base-pairing structures. The original structures for mispairing by O^6-alkylG would suggest that an O^6-alkylG:T base-pair is more stable than an O^6-alkylG:C base-pair because the O^6-alkylG:T base-pair has two H-bonds whereas the O^6-alkylG:C base-pair has only one H-bond. Based on this, Nagata et al. (1982) used molecular orbital calculations and predicted that the O^6-alkylG:T base-pair should be as stable as normal Watson-Crick base-pairs. This was later disproved by melting point studies on DNA duplexes containing O^6-alkylG paired with adenine, guanine, cytosine or thymine (Gaffney et al., 1984; Gaffney & Jones, 1989). If one could measure the thermodynamic stability of the O^6-alkylG:T and O^6-alkylG:C base-pairs, this could provide information on the base-pairing structures of O^6-alkylG. The measurement of DNA melting temperatures (T_m) is the most commonly used method of studying the thermodynamic stability of base-pairs. Using a set of self-complementary dodecanucleotides, d(CGNGAATTCmeGCG) where N is A, G, C or T, and meG is O^6-methylgaunine, Gaffney et al. (1984) showed that the T_m values of these duplexes containing O^6-methylgaunine are 19-26 °C lower than the T_m of the parent duplex d(CGCGAATTCGCG). The most stable duplex was the sequence with O^6-meG base-paired with cytosine, and the least stable duplex had the O^6-meG base-paired with thymine. The overall order of melting sequence was N= T > G > A > C.

A further study was carried out by Gaffney & Jones (1989) using a set of ten non-self-complementary oligodeoxynucleotides, d(GGTTXTTGG) and d(CCAAYAACC) where X and Y are A, C, G, T or O^6-meG, to measure the stability of each duplex (again by optical melting methods), which was then related to the enthalpy, entropy and Gibb's free energy changes associated with the melting of the duplex through the Arrhenius equation: $\Delta G = \Delta H - T\Delta S = -RT\ln K$, where ΔG, ΔH and ΔS are the Gibb's free energy, enthalpy and entropy changes respectively for for the formation of a duplex from single DNA strands, T is the absolute temperature, R is the gas constant and K is the equilibrium constant for the formation of a duplex from single DNA strands. ***Sequences containing the O^6-meG:C base-pair were found***

to be more stable than corresponding sequences containing O^6-meG:T base-pairs. The O^6-meG base-pairs have a relatively narrow range of thermal stabilities; the difference in Gibb's free energy between O^6-meG:C and O^6-meG:T base-pairs is only 0.6 kcal/mol. Thus, on purely energetic grounds, the O^6-meG:C base-pair is more stable than the O^6-meG:T base-pair, contrary to the original base-pairing schemes for O^6-meG.

The discovery that the O^6-alkylG:T base-pair is energetically less stable than the O^6-alkylG:C base-pair meant that the original O^6-alkylG base-pairing schemes were incorrect. This caused much confusion. In an attempt to provide a solution, new base-pairing schemes for O^6-alkylG were proposed by Williams & Shaw (1987). The base-pairing interactions of O^6-meG with thymine and cytosine in chloroform were investigated with ^{1}H-NMR. At neutral pH, a 1:1 complex between O^6-methyldeoxyguanine and deoxythymine was formed. The ^{1}H-NMR spectra for this complex was taken as evidence for the O^6-alkylG:T base-pair originally suggested (Fig. 1.8(c)). There was no evidence for pairing between O^6-alkyldeoxyG and deoxycytosine at neutral pH, but if perchloric acid was added to protonate the cytosine, a base-pair involving three H-bonds (Fig. 1.8(e)) was formed. This was taken as an explanation for the observation that O^6-meG:C base-pairs are more stable than O^6-meG:T base-pairs.

Many of the studies mentioned above were conducted at the base or nucleoside level which may not give a true picture of the base-pairing properties of O^6-alkylG in DNA. To gain better insight, oligonucleotides containing O^6-alkylG should be used in structural studies. This remained a stumbling block for many years because the chemical synthesis of DNA containing O^6-alkylG was too difficult a task for the biochemist. Fortunately, research by the laboratories of Gaffney and Jones, and Li and Swann have led to successful synthesis of oligonucleotides containing O^6-alkylG in large enough quantities for NMR studies. The first studies of O^6-alkylG were carried out by Patel et al. (1986a, b; reviewed by Li et al., 1988 and Swann, 1990) on a set of dodecamers containing O^6-meG base-paired with cytosine or thymine. These were the same oligonucleotides used by Gaffney et al. (1984) for optical melting point studies. The sequences were: d(C1-G2-C3-G4-A5-A6-T7-T8-C9-meG10-C11-G12) and
d(C1-G2-T3-G4-A5-A6-T7-T8-C9-meG10-C11-G12) .
The bases are numbered 1 to 12 in the same way that Patel et al. did (1986a, b).

To work out the three dimensional structure of these 12-mers, the protons in these duplexes were assigned according to NMR spectra recorded in H_2O (for the exchangeable protons which may be used for H-bonding) and in D_2O (for the non-exchangeable protons on the bases and sugars which provide information on stacking interactions and conformation of the glycosidic bonds). ^{31}P NMR spectra provided information on the phosphodiester bonds, while two-dimensional proton correlated spectra (COSY) provided data on how the protons were

connected (via through-bond connections). The Nuclear Overhauser enhancement (NOE) spectra were used to work out the spatial relationship between protons that are placed within a distance of 4 Å or less of each other; this is based on the principle that when a proton is given energy at its NMR frequency, it becomes energized but quickly loses this energy to other protons in the vicinity via through-space interactions.

Based on these initial NMR studies, structures for O^6-meG base-pairing were proposed (Patel et al., 1986a, b) which were somewhat conjectural with respect to the conformation of the 6-*O*-methyl group. This was because O^6-alkylG is fairly unsuitable for NMR studies due to a lack of imino protons, and the alkyl protons of the O^6-alkylG do not give an informative NMR signal. Although NOEs from the amino protons of the 6-*O*-methyl group to H5 (strong) and H6 (weak) of the adjacent C9, this was insufficient to define whether the 6-*O*-methyl group was *anti* or *syn* to N1. All this data could suggest was that the 6-*O*-methyl group lies in the major groove. However, the observation of a strong NOE from the 6-*O*-methyl group to the 5-methyl group of the pairing T in the O^6-meG:T duplex studied provided a clue to the *syn* conformation of the 6-*O*-methyl group with respect to N1; if the conformation of the 6-O-methyl group had been *anti* to N1, it would have been too distant from the 5-methyl of the pairing partner T to provide a strong NOE signal.

Further NMR studies on the structure of O^6-alkylG base-pairing were carried out on another set of dodecamers containing O^6-ethylguanine base-paired with C or with T:

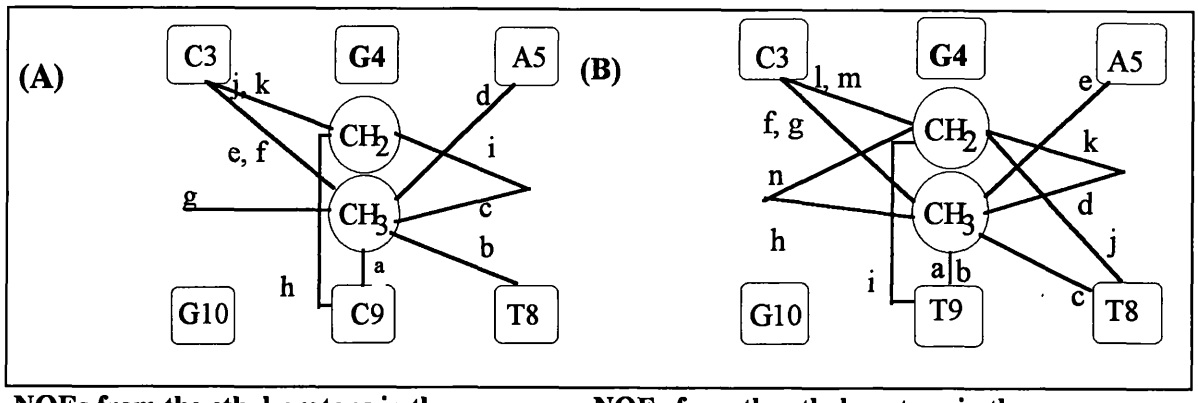
d(C1-G2-C3-etG4-A5-G6-C7-T8-C9-G10-C11-G12) and

d(C1-G2-C3-etG4-A5-G6-C7-T8-T9-G10-C11-G12).

(Kalnik et al. 1989a, b, also reviewed by Li et al., 1988 and later by Swann, 1990). These studies were able to provide information on the conformation of the 6-*O*-ethyl group because technical advancements had made it possible to measure the NOE between the 6-*O*-ethyl protons and the imino protons. Many NOE signals were detected which involved the ethyl protons of the O^6-ethylguanine, and this allowed a clear definition of the position of the 6-*O*-ethyl group. In particular, in the O^6-etG:C duplex, there was a strong NOE between the -CH_3 ethyl protons and 5-methyl of T8 in the opposite strand, which clearly showed that the ***6-O-ethyl group was syn to N1*** (Figure 1.11a).

Another significant result from the studies of Kalnik et al. (1989a, b) relates to the 4-amino protons of the cytosine that was base-paired with the O^6-etG. In a G:C base-pair, one of the two protons on the 4-amino of cytosine is H-bonded to O^6- of guanine while the other is free (Figure 1.8a). Thus, two resonances are detectable from the 4-amino protons in a G:C base-pair, one due to the H-bonded proton (8-8.5 ppm) and another due to the free proton (6.4-7 ppm). In the O^6-etG:C base-pair however, only a single peak at 7.78 ppm was detected for the 4-amino

protons of cytosine. This suggested that in this base-pair, one of the 4-amino protons of C was H-bonded but rotation about the C-NH_2 bond was rapid.



NOEs from the ethyl protons in the O^6-etG:C base-pair in DNA.

- **NOEs from the -CH_3 protons**

(a) amino protons of C9
(b) methyl protons of T8
(c) imino proton of T8
(d) H-bonded amino proton of A5
(e) H-bonded and exposed amino protons of C3
(f) H5 and H6 of C3
(g) imino proton of G10

- **NOEs from the -CH_2- protons**

(h) amino protons of C9
(i) imino proton of T8
(j) H-bonded amino proton of C3
(k) H5 of C3 (weak)

NOEs from the ethyl protons in the O^6-etG:T base-pair in DNA.

- **NOEs from the -CH_3 protons**

(a) imino protons of T9
(b) methyl group of T9 (weak)
(c) methyl group of T8
(d) imino proton of T8
(e) H-bonded and exposed amino protons of A5
(f) H-bonded and exposed amino protons of C3
(g) H5 and H6 of C3
(h) imino proton of G10

- **NOEs from the -CH_2- protons**

(i) imino protons of T9
(j) methyl group of T8
(k) imino proton of T8
(l) H-bonded amino proton of C3
(m) H5 of C3
(n) imino proton of G10

Figure 1.11. NOEs of the ethyl protons observed in the NMR spectrum of DNA duplexes containing an O^6-etG base-paired with thymine or cytosine (Kalnik et al., 1989a, b). The presence of an NOE indicates that the two protons are less than 4.5 Å apart, and these NOEs allow the position of the ethyl group relative to the surrounding structures to be defined. In particular, the strong NOE between -CH_3 of the ethyl protons and the 5-methyl protons of T9 in the opposite strand clearly show that the ethyl group must be *syn* to N1 of the O^6-etG.

Based on the results of these NMR studies, ***a wobble structure for the base-pairing of O^6-alkylG and cytosine was proposed*** (Kalnik et al., 1989a; see Fig. 1.8(g)). The 6-*O*-alkyl group is *syn* to N1 of the alkylG, and the O^6-alkylG:C pair is stabilized by two H-bonds. In order to accommodate the 6-*O*-alkyl group, the bases are shifted laterally. In normal Watson-

Crick base-pairs, N1 of the purine is juxtaposed to N3 of the pyrimidine, but in the O^6-alkylG:C base-pair, the purine is shifted towards the major groove of the DNA and N1 of the purine is juxtaposed, and H-bonded, to the amino group of C4 of the cytosine.

In the duplex containing the O^6-etG:T base-pair, there were strong NOE signals from the $-CH_2-$ and the $-CH_3$ ethyl protons to 5-methyl of T8 in the opposite strand (Figure 1.11). Again this clearly illustrated that the 6-*O*-ethyl group was *syn* to N1. Furthermore, in a normal Watson-Crick A:T base-pair, the N3 imino proton of the T is H-bonded to N1 of A, and this N3 imino proton of T resonates at 14 ppm. In the O^6-ethylG:T base-pair, the N3 imino proton of the T was found to resonate at 8.85 ppm, which led to the conclusion that in the O^6-etG:T base-pair, the N3 imino proton is not H-bonded. The proposed structure for the O^6-alkylG:T base-pair (Kalnik et al., 1989b; see Fig. 1.8f) thus has only one H-bond. There is no bonding between N3 of T and N1 of the alkylG because the bases are forced apart in order to accommodate the 6-*O*-alkyl group. ***The significance of the structure for the O^6-alkylG:T base-pair is that it retains a Watson-Crick alignment with the N1 of the alkylG juxtaposed to N3 of the thymine, in contrast to the wobble structure proposed for the O^6-alkylG:C base-pair.***

NMR studies on DNA duplexes containing O^4-meT (Kalnik et al., 1988a, b) also showed the 4-*O*-methyl group to be *syn* to N3 of the alkylated thymine. In general, the structures of DNA containing O^4-alkylT is known with greater certainty than that of DNA containing O^6-alkylG. This is because O^4-alkylT is more satisfactory from the point of view of NMR. The only protons on O^6-alkylG that give a good NMR signal are the alkyl protons; the amino protons do not give a sharp signal. By contrast, in O^4-alkylT, one has the alkyl protons, the C5-methyl protons and the C6 proton. The strength of the NOE between the alkyl protons and the C5-methyl protons is particularly informative in deciding whether the 4-*O*-alkyl group is *syn* or *anti* to N3 of the alkylT.

The NMR studies on DNA containing O^4-meT base-pairs (Kalnik et al., 1988a, b; reviewed by Swann, 1990) revealed that there is a striking resemblance between the structures involving O^6-alkylG and the structures for O^4-meT. The O^4-meT:G base-pair retains Watson-Crick alignment with N1 of the purine juxtaposed to N3 of the pyrimidine, but the bases are somewhat pushed apart so as to accommodate the bulky 4-*O*-methyl group in the *syn*-conformation (Figure 1.9e). Thus, only one H-bond could be formed in the O^4-meT:G base-pair, between the 2-amino of guanine and the 2-*O* of O^4-meT. On the other hand, the O^4-meT:A base-pair has a wobble conformation (Fig.1.9f), with N3 of the pyrimidine H-bonded to the 6-amino group of adenine.

The NMR structures proposed by Patel and his colleagues for the pairing of O^6-alkylG are in agreement with the thermodynamic measurements made by Gaffney and Jones (1989). The

O^6-alkylG:T base-pair has one H-bond whereas the O^6-alkylG:C base-pair has two, which made it energetically more stable than the former base-pair. However, three X-ray crystallographic studies (Ginell et al.,1990; Leonard et al.,1990 and Sriram et al., 1992) were published after the NMR studies by Patel and his colleagues, which threw the subject into some confusion. The first of these studies was reported by Ginell et al. (1990). Crystals of the self-complementary d(CGCmeGCG) were grown and the analysis showed the DNA to be in the left-handed Z form. The O^6-meG:C base-paired in this crystal had a Watson-Crick type geometry, which is in contrast to the Wobble structure for the O^6-alkylG:C base-pairing in right-handed B form DNA deduced from solution NMR studies by Patel and his colleagues. Within the same year, Leonard et al. (1990) reported findings which were also based on X-ray crystallography studies, that although the O^6-meG:T base-pair has a Watson-Crick type geometry, it is stabilized by not one but two H-bonds and the 6-O-methyl group is not *syn* but *anti* to N1. The same study also reported detection of two conformations of the O^6-meG:C base-pair: a Wobble structure and a Watson-Crick structure involving a protonated cytosine. The third X-ray crystallography study by Sriram et al. (1992) was carried out on an oligonucleotide duplex in the B-form which contained two O^6-ethylG:C base-pairs stabilized by the minor groove-binding drugs Hoechst 33258 and Hoechst 33342. The two O^6-etlG:C base-pairs within the same duplex was found to exist in different conformations. One had a Wobble structure with the 6-*O*-ethyl group *syn* to N1 while the other had bifurcated H-bond pairing and the 6-*O*-ethyl group was *anti* to N1; the authors attributed these different structures to differences in the local environment.

The different structures deduced from solution NMR studies and X-ray crystallography studies may be due to the fact that the former studies were carried out in solution and the latter in crystals. Since the ultimate aim of these structural studies is to see how the structures of the O^6-alkylG:T and O^6-alkylG:C base-pairs could affect the frequency of mutation and how these lesions might be recognized by the repair enzymes in the cell, it is probably more appropriate to carry out the studies in solution rather than in crystals. Very recently Goswami and Jones (1993) have shown unequivocally that the NMR derived structures of the O^6-alkylG:T base-pair is correct. They synthesized oligonucleotides containing O^6-meG with an ^{15}N label on either N1 or N2 of the alkylated guanine, and by monitoring the ^{15}N NMR spectra of these oligonucleotides, they were able to show that only the N2 of O^6-meG was H-bonded to thymine and that N1 was H-bonded to water.

The work detailed above has considered the base-pairing properties of the O^6-alkylG. The question is how do these studies relate to the coding properties of O^6-alkylG in DNA synthesis? Virtually all discussion of biological DNA synthesis emphasizes the role of the free energy of base-pairing in fidelity. Yet, here is a paradox with the alkylated bases forming an energetically less stable base-pair during DNA synthesis. This means that the preference for

misincorporation of thymine over cytosine opposite O^6-alkylG during DNA replication cannot be explained entirely in terms of thermodynamic stability and H-bonding. Why then does O^6-alkylG miscode? The peaks in the ^{31}P spectrum of oligodeoxynucleotide duplexes containing O^6-etG:C or O^6-etG:T pairs provided a clue to the answer. These peaks were assigned by Kalnik et al. (1989a, b; see Table 5), and the assignment shows that the greater difference in chemical shift relative to the control duplex containing G:C pairs, in the position of O^6-ethylG pairs, was the phosphodiesters 3' and 5' to the C in the O^6-ethylG:C base-pair. Unfortunately, the exact reason for the differences in the chemical shift of individual phosphodiesters is unknown, but Gorenstein (1992) has suggested that they are a reflexion of the changes in the torsion angles of the phosphodiester backbone. ***Thus, the O^6-etG:C base-pair, although thermodynamically more stable than the O^6-etG:T base-pair, produces more distortion of the DNA structure. This distortion is in the DNA strand opposite the O^6-etG, thus incorporation of cytosine during DNA replication opposite a template O^6-etG would have to accommodate this distortion.***

Table 5. Comparison of the ^{31}P chemical shifts of the phosphodiesters around the alkylated or mismatched base-pair in DNA duplexes containing O^6-etG:C, O^6-etG:T, or G:T base-pairs. In the table below the chemical shift is given as δ ppm, and the difference between the chemical shift of each phosphodiester in the mismatched duplexes and the equivalent phosphodiester in the perfectly base paired duplex containing a G:C pair at position 4 is given as Δ. The general structure of the oligomers was C1 .G2 .C3 .*G4.A5 .G6 .C7 .T8 . X9.G10.C11.G12
G12.C11.G10. X9.T8 .C7 .G6 .A5 .*G4.C3 .G2 .C1

The modified base pair was *G4.X9. The largest differences have been underlined. Note that these occur in the mismatches with a 'wobble' conformation i.e. in the G:T duplex and the etG:C duplex (taken from Kalnik et al., 1989a, b).

Phosphodiester	G:C duplex	etG:C duplex		G:T duplex		etG:T duplex	
	δ ppm	δ ppm	Δ	δ ppm	Δ	δ ppm	Δ
C3--*G4	-4.08	-3.97	0.11	-4.07	0.01	-3.86	0.22
*G4--A5	-4.15	-4.16	-0.04	-4.58	<u>-0.43</u>	-4.23	-0.08
T8--X9	-4.46	-3.87	<u>0.59</u>	-4.57	-0.11	-4.28	0.18
X9--G10	-3.95	-4.45	<u>-0.50</u>	-3.54	<u>0.41</u>	-4.16	-0.21

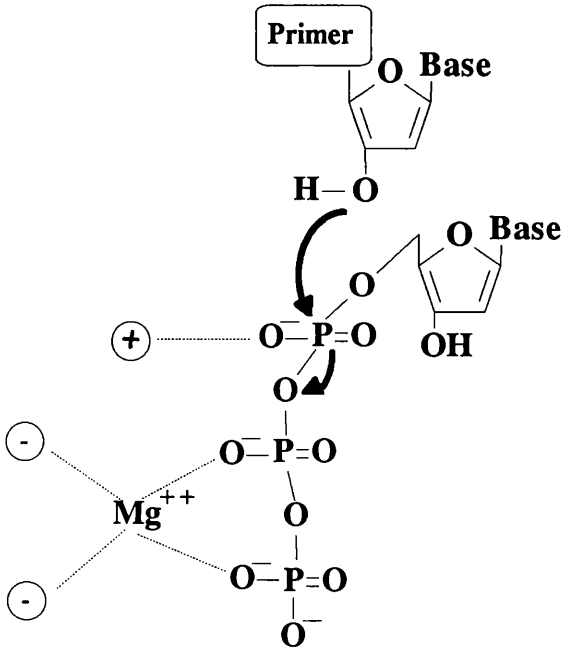
1.6 DNA polymerases and fidelity of DNA synthesis

Transition mutations occur as a result of error during DNA replication. To identify the mechanism by which O^6-alkylG miscodes, it is necessary to examine the biological factors involved in the replication process as well as the structural factors. Hence, the next part of the introduction describes the structure and function of the Klenow fragment of DNA polymerase I from *E. coli*, which is the DNA polymerase used in the work described in this thesis.

Replication of DNA is one of the most accurate biochemical processes known to man. Life itself, and its continuity, depends on the genetic code. Small changes such as a single base change can lead to catastrophic or even lethal consequences. Thus, from bacteria to man, when genetic information is passed down from one generation to another, a high degree of accuracy in copying DNA is crucial in order to maintain species homogeneity. On the other hand, an infinitesimal but well-defined error frequency is required for evolution. In biological systems, the error frequency in DNA synthesis is in the order of 10^{-9}-10^{-11} errors per nucleotide incorporated per generation (Drake, 1969). Organisms with small genomes can afford to be more error-tolerant but cells with large genomes must possess means to ensure a high fidelity of DNA replication.

The first step towards DNA polymerisation is the binding of substrates, and there is an obligatory sequential order of substrate-binding. The polymerase binds the primer-template DNA first and then the dNTP. Polymerisation is the formation of a phosphodiester bond between the 3'-end of the primer and the α-phosphate of the incoming dNTP, and it is well known that magnesium is required for this reaction. In Klenow fragment, magnesium is required for polymerisation but it is not essential for the binding of the dNTP; however, the dNTP usually binds as a Mg-β-γ bidentate chelate (Fig. 1.12). If a complementary nucleotide is bound, H-bonds are formed between the complementary bases, followed by a nucleophilic attack on the α-phosphate of the dNTP by the 3'-OH of the primer terminus and subsequently a phosphodiester bond is formed. The magnesium ion promotes the leaving of the pyrophosphate by the formation of a Mg-β-γ bidentate chelate (reviewed by Kornberg & Baker, 1992). Photoaffinity labelling experiments using 5'-[(fluorosulphonyl)benzoyl]deoxyadenosine triphosphate, a dATP analogue with an affinity label attached to the triphosphate, have shown that Arg 682 is probably involved in the binding of the β- and/or the γ-phosphate of the dNTP (Pandey & Modak, 1987). The roles played by magnesium and Arg 682 in relation to the β- and γ-phosphates remain unknown at present; however, molecular modelling studies by Yadav et al. (1992) showed that there is no direct contact between Arg 682 and the dNTP, and the authors suggested that Arg 682 may be used to accept the leaving pyrophosphate group after the phosphodiester bond is formed.

Figure 1.12. Binding of magnesium to the dNTP binding site in Klenow. The diagram shows Mg^{2+}.dNTP binding as a β-γ bidentate chelate, and nucleophilic attack by the 3'-OH of the primer terminus on the α-phosphate of the dNTP. The positive and negative charges shown are presumed to reside in the enzyme surface. After the phosphodiester bond has been formed, it has been speculated that the magnesium ion can promote the leaving of the pyrophosphate group by the formation of a Mg-β-γ bidentate chelate (Yadav et al., 1992) (Diagram taken from Kornberg & Baker, 1992).



Klenow fragment of *E. coli* DNA polymerase I

The kinetic studies in this Ph.D. project which will be described later were carried out using the Klenow fragment of *E. coli* DNA polymerase I. Its structure and mechanism of action will therefore be described in some detail in the remainder of this section. DNA polymerase I of *E. coli* was the first template-directed DNA synthesizing enzyme to be purified. Although the main function of pol I in the cell is one of repair rather than replication, nevertheless its simplicity makes it a good candidate for molecular investigations of the extremely complicated process of DNA replication.

Pol I is encoded by the locus *pol* A, and consists of a single polypeptide chain of 103 kDa. It can be cleaved by proteolytic treatment to give two fragments, a large fragment (68 kDa) called the Klenow fragment and a smaller fragment (36 kDa) which forms the N-terminal of pol I. Klenow fragment has two catalytic activities, a 5'→3' polymerase activity and a 3'→5' proofreading exonucleolytic activity. The smaller proteolytic fragment of pol I has the 5'→3' exonucleolytic activity (reviewed by Kornberg & Baker, 1992).

The crystal structure of Klenow fragment has been well investigated (Ollis et al., 1985; Derbyshire et al., 1988; Freemont et al., 1988; Beese et al., 1993). The 605 amino acid chain is folded into two domains; the bigger domain (residues 521-928) contains the polymerase activity

while the other (residues 324-517) contains the 3' to 5' exonuclease activity (Ollis et al., 1985). These two domains are shown in red and green respectively in Fig. 1.13.

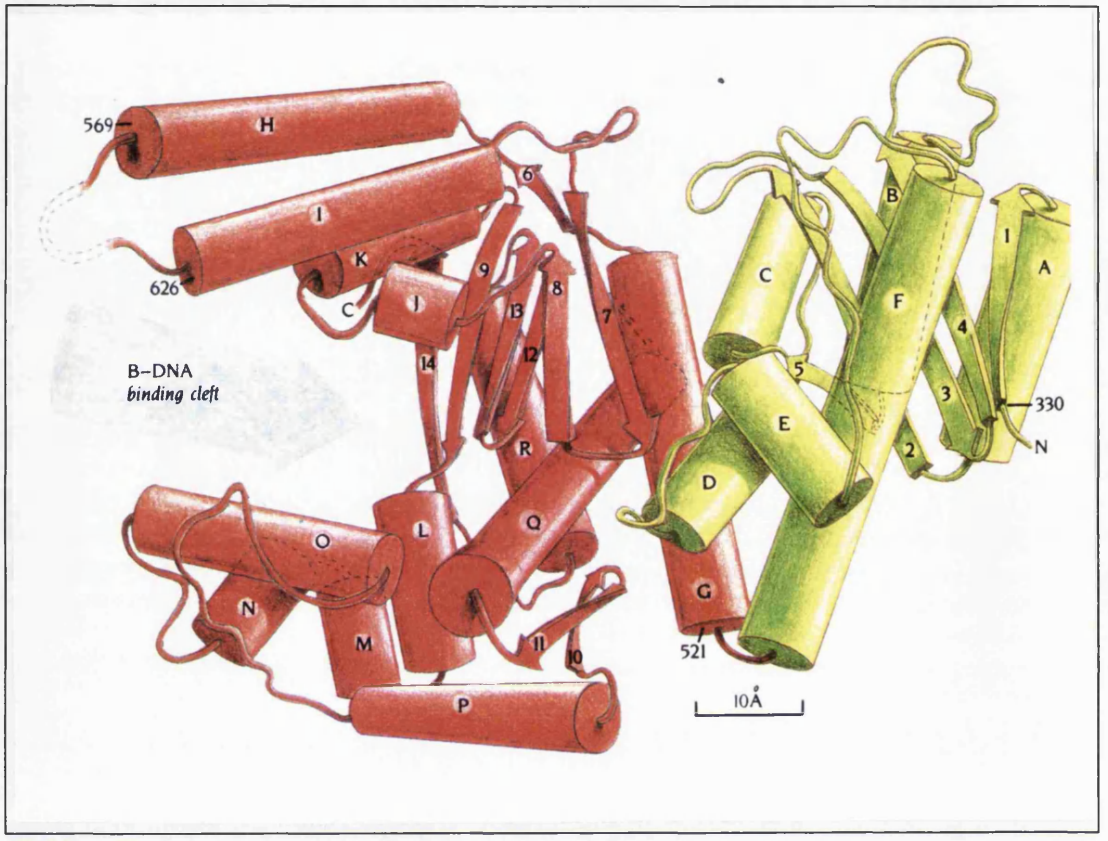


Figure 1.13. The secondary structure of Klenow fragment. α-helices are represented by cylinders and β-sheets by arrows. The polymerase and 3' to 5' exonuclease domains are shown in red and green respectively. The DNA binding cleft is indicated. The partially disordered sub-domain of 50 amino acids mentioned in the text is represented with broken lines. (taken from "An Introduction to Protein Structure", by C. Brandon & J. Tooze, 1991, Hall & Chapman).

The C-terminal domain containing the polymerase activity has a very deep cleft that was measured at approximately 20-24 Å wide and 25-35 Å deep (Ollis et al., 1985). The width of B-DNA is 24 Å, thus the shape and size of the cleft suggests that it binds double stranded B-DNA, although several places in the cleft might be too tight to accommodate this. In Fig. 1.13, one can see that helices H and I protrude to form a wall of the cleft, and these two helices are joined by 50 amino acids that seem to form a partially disordered subdomain, represented by broken lines in the figure. In the absence of structural data (the studies by Ollis et al. were carried out on a crystal of Klenow complexed with dTMP, not duplex DNA), no function could be assigned to this disordered subdomain. Nevertheless, the authors proposed that this partially

disordered subdomain could act as a lid to close off the cleft after duplex DNA has bound, so that the enzyme is less likely to dissociate from the DNA which results in processive polymerisation. However, very recently, a crystal was obtained which was complexed with duplex DNA, and in this crystal, the region previously referred to as the disordered subdomain became more ordered, forming two new helices H_1 and H_2 linked by two loops (Beese et al., 1993; see Fig. 1.18).

The helices J and K also protrude somewhat into the DNA binding cleft, and it was suggested that these helices probably serve as guides for positioning the DNA relative to the polymerase (Ollis et al., 1985). Thus, it is envisaged that during DNA synthesis, the polymerase would follow a helical path along the DNA. Evidence for this was obtained in a later experiment using fluorescence spectroscopy of Klenow fragment bound to duplex DNA containing a fluorophore (Allen & Benkovic, 1989; discussed in greater detail later). Molecular modelling based on the known charges on the amino acid side chains also predicted that virtually all of the positive charges of Klenow lies within the DNA binding cleft and follows a helical path, while the negative charges are spread all over the enzyme (Yadav et al., 1992).

The concentration of positive charges from amino acid side chains within the cleft suggests that the interaction between the DNA binding cleft and duplex DNA is through the phosphate backbone, which is reasonable since the enzyme has to bind duplex DNA irrespective of its sequence. This has been observed recently in the crystal structure of Klenow complexed with duplex DNA (Beese et al., 1993). Many of the residues within the cleft that were observed to interact with the DNA phosphate backbone in this crystal were either highly conserved within the Pol I family or found to have an amide side chain. These residues are located near or on helices I and J (which interacted with the primer strand), and on helix Q (interacting with the template strand). Helices H_1 and H_2, which were formed from a disordered region after Klenow fragment has bound duplex DNA, also contained highly conserved residues that were observed to interact with both the primer and template DNA strands. The position of these residues provided valuable information in elucidating the structure of the polymerase when it binds duplex DNA (discussed later).

The observation that interaction between the polymerase and DNA is mainly through the phosphate backbone, especially the interaction with the primer strand, particularly is important because it explains how the enzyme is sensitive to changes in the structure of the phosphodiester backbone. This could provide a basis for recognition by the enzyme of a mismatched base-pair that produces a distortion in the phosphodiester backbone. In the context of replication of DNA containing O^6-alkylG:T or O^6-alkylG:C base-pairs, which has been considered earlier, incorporation of a thymine opposite the alkylated guanine results in

minimal distortion of the phosphodiester backbone, but incorporation of a cytosine opposite the alkylated guanine produces a significant distortion of the phosphodiester bond in the primer strand 5' to the cytosine. If this distortion in the phosphodiester backbone can be detected by Klenow, the kinetics of incorporation of a cytosine opposite an O^6-alkylG, and of a subsequent incorporation following that of the cytosine will be affected. Whereas with the incorporation of thymine opposite a template O^6-alkylG, the distortion of the phosphodiester bond produced is perhaps small enough not to disturb its interaction with the enzyme to a great extent.

The approximate position of the 3' end of the primer strand in the large domain of Klenow has been determined by several experiments. DNA foot printing using DNase I showed that when Klenow fragment is bound at a primer terminus, it covers about eight base-pairs of duplex DNA upstream of the primer terminus (Joyce et al., 1986). Studies of interactions between Klenow fragment and fluorescent oligonucleotides also supported this model (Allen et al., 1989).

Much effort has been spent on the identification of the polymerase active site in Klenow fragment. There is a single site in pol I for which all dNTPs compete. In the absence of a primer-template, all dNTPs are bound with similar affinity. Inorganic pyrophosphate, dNMPs and dNDPs do not bind to or compete for this dNTP binding site (reviewed by Kornberg & Baker, 1992). There is evidence that the dNTP binding site lies within the DNA binding cleft of the large fragment of Klenow. DNA cross-linking experiments have been performed to identify the crucial residues involved in dNTP binding. In one such experiment using UV light, dTTP has been cross linked to His 881, which is found on helix R (Pandey et al., 1987). In another experiment, pyridoxal phosphate, which binds competitively to the dNTP binding site, has been found to interact with Lys 758, located on helix O which forms a wall of the DNA binding cleft (Fig. 1.13) (Basu et al., 1988). Tyr 766, at the C-terminal of helix O, is also involved in dNTP binding as revealed by photoaffinity-labelling experiments by Rush & Konigsberg (1990). The dNTP analogue 8-azido-dATP was cross-linked to Klenow fragment and the labelled peptide subsequently sequenced. Besides binding dNTP, Tyr 766 has also been found to bind the primer terminus in further studies using azido-DNA and fluorescent DNA probes (Catalano et al., 1990). Recent molecular modelling studies place Tyr 766 within a short distance between the sugar moieties of the primer terminus and of the dNTP (Yadav et al., 1992). This position of Tyr 766 makes it well suited to the polymerisation reaction where the primer terminus must surely be very near the dNTP to enable phosphodiester bond formation between these two molecules. Kinetic studies of a mutant DNA pol I in which Tyr 766 is replaced by serine or phenylalanine have demonstrated reduced fidelity of DNA synthesis (Carroll et al., 1991). It is speculated that the phenolic ring of Tyr 766 might be used to align the α-phosphorus of the dNTP with respect to the 3'-OH of the primer terminus, facilitating nucleophilic attack by the primer terminus.

Because of their importance, Tyr 766 and Gly 767 are strictly conserved in homologous polymerases from other species like phages T5 and T7, *Streptococcus pneumoniae*, and *Thermus aquaticus*, as are several other residues in this region.

The N-terminal smaller domain (residues 324-517) of Klenow fragment has a core of β -pleated sheets sandwiched between α-helices (Fig. 1.13). It can bind two divalent metal ions (Mg, Zn or Mn) and a dNMP molecule. The dNMP binding site has been well-studied. In a structural study of dTMP complexed with Klenow in the presence of divalent cations (Ollis et al., 1985), it was revealed that the phosphate and the ribose moieties of the dTMP are buried in the interior of the protein, leaving only the H-bonding groups on the base exposed to water (Fig. 1.14). Binding of dNMP abolishes the 3'→5' exonuclease activity, therefore the dNMP binding site must correspond to the 3' primer terminus in the exonuclease active site, and the phosphate of the dNMP represents the position of the phosphodiester bond to be cleaved by the exonuclease. It has been further demonstrated that the exonuclease site can bind single-stranded DNA of four nucleotides in length, which covers a distance of some 15 Å from the 3' terminus (Freemont et al., 1988).

Magnesium is required for the 3' to 5' exonuclease activity. The structural basis for this is shown by the work of Ollis et al. (1985), Derbyshire et al. and Freemont et al. (1988), and more recently, Beese and Steitz (1991). Two divalent cations are involved in the 3' to 5' exonuclease reaction. Binding of the dNMP to the exonuclease site requires interaction with a magnesium or zinc cation at site A (see Fig. 1.15); the divalent cation is coordinated by the carboxylate groups of Asp 355, Glu 357 and Asp 501, with the phosphate group of the dNMP providing the fourth ligand (Ollis et al., 1985). The second divalent cation is located between the phosphate of the dNMP and the carboxylate of Asp 424. The functions of these two divalent cations were examined by site-directed mutagenesis studies (Derbyshire et al., 1988). Two mutant forms of Klenow were produced, one was a double mutant with Asp 355 and Glu 357 replaced with Ala, the other mutant protein had Ala 424 instead of Asp 424. Both mutant proteins had levels of polymerase activity similar to those of the wild-type protein, but very low levels of exonuclease activity. The double mutant was unable to bind either dTMP or divalent metal ions, thus the conclusion must be that metal site A is important in binding of the substrate for the exonuclease reaction. The mutant which contained Ala 424 in place of Asp 424 could not bind a divalent cation either, but it bound dTMP even in the absence of a divalent cation. This suggested that metal site B is directly involved in the 3' to 5' exonuclease reaction (Derbyshire et al., 1988).

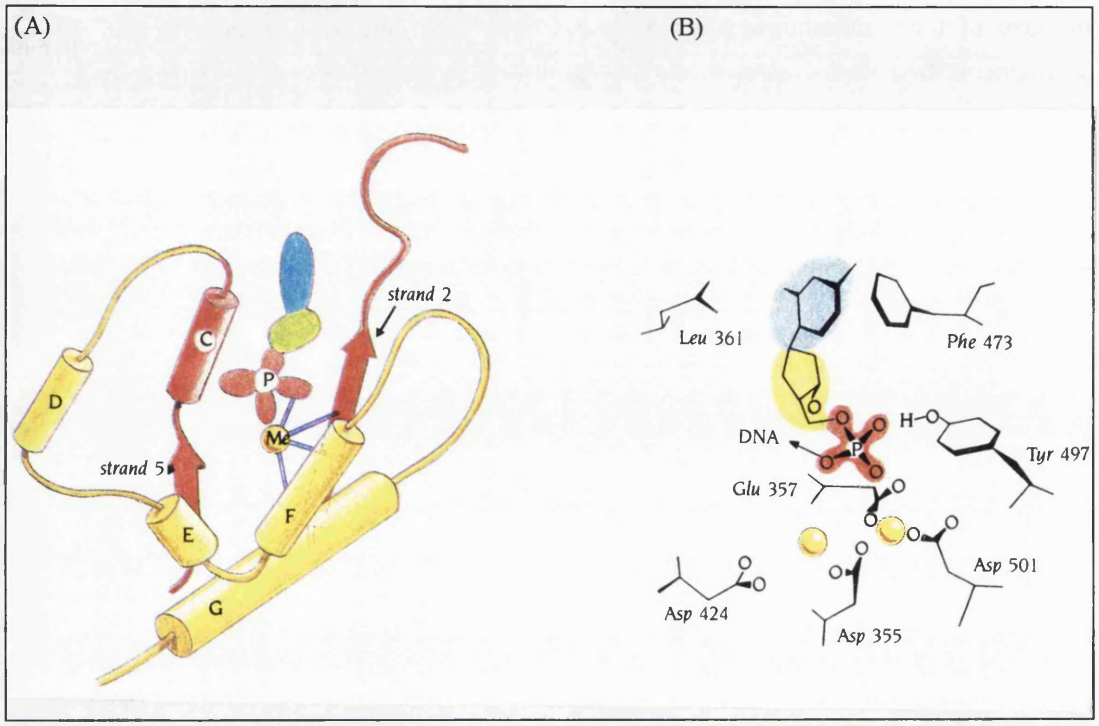


Figure 1.14. (A) Schematic drawing showing the binding of dNMP to the exonuclease site of Klenow fragment, based on a study by Ollis et al. (1985). The base of the dTMP is shown in blue, the ribose in green and the phosphate group in red. (B) Interactions of dTMP with various protein side chains in the Klenow-dTMP complex. The phosphate of the dNMP (red) interacts with a divalent cation (site A, discussed in the text and shown in the diagram as the left of the two yellow spheres) and the OH-group of Tyr 497. The 3'-OH of the dNMP is hidden and H-bonded to Thr 358 (not shown here), which may explain why dideoxynucleoside monophosphates cannot bind to the dNMP binding site. The base (blue) of the dNMP is surrounded by several hydrophobic protein side chains, most notably Leu 361 and Phe 473, interacting with each side of the base respectively. (taken from "An Introduction to Protein Structure", by C. Brandon & J. Tooze, 1991, Hall & Chapman).

Further support for this model of the exonuclease came from studies by the same laboratory (Freemont et al., 1988) using co-crystals prepared by diffusing single-stranded deoxytetranucleotides (dT_4) into crystals of Klenow fragment. In the Klenow-dT_4 complex, the single-stranded DNA was bound at the exonuclease site where only the two divalent cations at sites A and B and the OH of Tyr 497 were close enough to the phosphate to be removed to be directly involved in hydrolysis. Based on these observations, the authors proposed a mechanism for the 3'→5' exonuclease reaction which requires two divalent cations. The hydrolysis is elicited by a nucleophilic attack by a water molecule on the phosphorus of the phosphodiester bond to be

broken. Metal site A may facilitate this by guiding the attacking water molecule towards the best position for attack. When the water attacks the phosphorus, a transient pentacovalent intermediate is formed, and this intermediate is stabililized by metal site B. The reaction is complete when the phosphodiester bond is broken and the 3' terminal dNMP is released. Here, metal site B is also thought to promote the leaving of the dNMP (Fig. 1.15). A more recent X-ray crystallography study showed that in the 3' to 5' exonuclease reaction, Tyr 497, Glu 357, Asp 501 and Asp 355 bind to and position the two metal ions, the single-stranded DNA and the attacking water molecule, giving more support to the two-metal ion mechanism for the exonuclease reaction (Beese & Steitz, 1991).

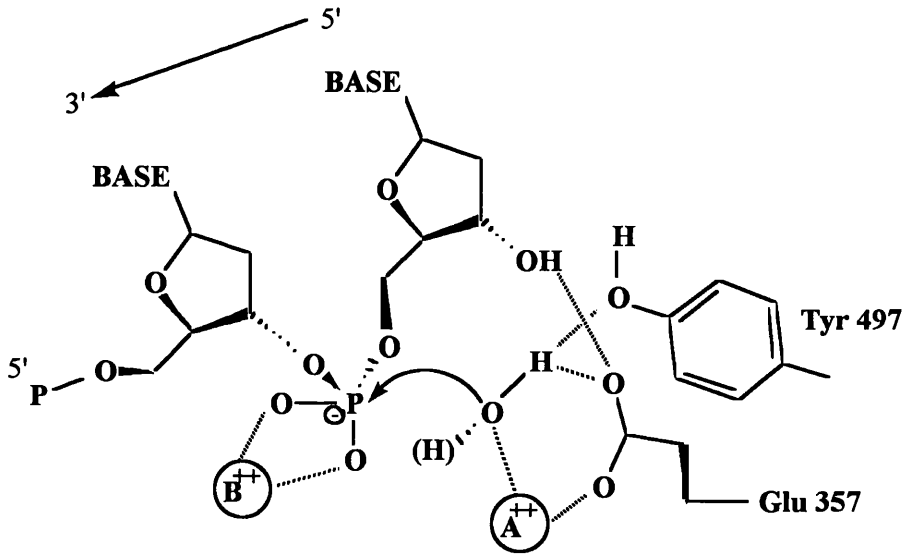


Figure 1.15. The two metal ion mechanism for the 3' to 5' exonuclease reaction proposed by Freemont et al. (1988). Two divalent metal ions are represented by A and B. Metal ion B is proposed to stabilize the transient pentacovalent species and the leaving of a 3' terminal dNMP (left residue in the diagram), whereas metal ion A may guide the water molecule towards the position that is best suited for attacking the phosphorus of the phosphodiester bond to be broken. The metal ions are separated by a distance of 3.9 Å (adapted from Freemont et al., 1988 and Beese & Steitz, 1991).

The editing function of Klenow fragment

The polymerase and the 3'→5' exonuclease activities of Klenow fragment reside on separate domains and the two active sites are approximately 30 Å apart, equivalent to the distance covered by 8 base-pairs. Direct support for the 8 base-pair distance between the polymerase and exonuclease active sites in Klenow fragment and the associated melting of the

primer terminus came from fluorescence energy transfer experiments (Allen & Benkovic, 1989). An oligonucleotide containing a fluorescent residue was prepared and annealed with its complementary strand. The fluorescence emission from this labelled duplex increases upon binding to specific positions in Klenow. By varying the position of the label within the duplex DNA and observing the emission, points of strong enzyme-DNA interactions could be elucidated. In this way, evidence for the 8 base-pair distance between the polymerase and exonuclease active sites in Klenow fragment and the associated melting of the primer terminus was derived.

If the polymerase and exonuclease active sites are separated by 30 Å, how do they coordinate with each other? A 'slide-and melt' model has been proposed where DNA can move between the two active sites by sliding along a helical path over a distance equivalent to 8 base-pairs (Joyce & Steitz, 1987 and Joyce et al., 1988). An intelligent editor is able to distinguish between correct and incorrect base-pairs. Whether a mispair will escape exonucleolytic hydrolysis and become fixed as a mutation in DNA depends on the balance between the polymerase and the exonuclease rates. Effective editing does not require absolute discrimination between correct and incorrect base-pairs, and pol I is known to be capable of extending incorrectly base-paired primer termini and of hydrolysing correctly base-paired termini. The role of Klenow as an editor is aptly described by Joyce et al. (1988): all that is required of the editor is that "the relevant reaction rates for correct and incorrect DNA termini be biased such that a mispaired terminus would be less likely to persist in DNA and become fixed as a mutation."

The 3'→5' exonuclease site contains a binding site for single-stranded DNA of four nucleotides in length (Freemont et al., 1988). On the other hand, the DNA binding cleft in the polymerase can only accommodate double-stranded DNA. In order for the DNA to move from the polymerase site and bind to the exonuclease site, the terminal 3-4 base-pairs have to be melted. The slide-and-melt model proposes that this could provide the physical basis for the editing decision. A mismatched base-pair is thermodynamically less stable than a correct base-pair, and would therefore cost less energy to melt, which increases the chances of it being excised by the exonuclease. ***Thus, any factor favouring melting at the 3' primer terminus and enhance its binding to the exonuclease site will increase the efficiency of proofreading***.

To obtain more evidence for the 'slide-and-melt' model of editing, the binding preference of duplexes containing matched and mismatched primer termini were examined in fluorescence energy transfer experiments (Catalano et al., 1990). The results showed that a duplex containing correct base-pairs at its 3' primer terminus binds predominantly at the polymerase site, and that exonucleolytic hydrolysis of mismatched primer termini is very fast when compared to that of a correctly base-paired primer terminus. This suggests that the rate-limiting step in the exonuclease activity of the enzyme is the translocation of the primer terminus

from polymerase to exonuclease active sites. This is in agreement with previous work carried out in the same laboratory (Kuchta et al., 1988a), which showed that the addition of a nucleotide onto a correctly base-paired terminus is rapid (50 s^{-1}) while the elongation of a mismatched primer terminus is considerably slower (ranging from 10^{-5} s^{-1} to 10^{-2} s^{-1}, depending on the mismatch). This pause in the polymerisation onto a mismatched primer terminus may allow time for melting and translocation to the exonuclease site where the mismatched base-pair is removed. After that, the primer terminus returns to the polymerase active site and polymerisation is resumed. It has been shown that in the absence of mismatched bases in duplex DNA, 88% of the DNA resides in the polymerase active site (Guest et al., 1991).

Structural evidence for Klenow as an editor was not available until very recently, when crystals of Klenow fragment complexed with duplex DNA were available (Beese et al., 1993). In the absence of structural data, it was initially thought that when duplex DNA binds to Klenow fragment, it enters the polymerase active site from the polymerase domain and binds directly in the DNA binding cleft (Joyce & Steitz, 1987). The recent study by Beese et al. proposed that the duplex DNA actually enters from the direction opposite to that originally proposed, that is the DNA enters Klenow fragment from an entrance adjacent to the 3' to 5' exonuclease site (Fig. 1.17) so that when it first enters, the DNA is perpendicular to the DNA binding cleft in the polymerase domain. After the DNA has bound, there is a conformational change which involves helices H and I. The disordered domain between helix H and I also becomes ordered as it forms two helices H_1 and H_2 (Fig. 1.18). The four helices H, I, H_1 and H_2, referred to as the 'thumb-like' protrusion by the authors in their paper, move towards the exonuclease and thus create a second cleft (cleft 2 in Fig. 1.18). This second cleft that is created as a result of the conformational change is perpendicular to the DNA binding cleft in the polymerase domain (cleft 1 in the Fig. 1.18), which means that for the DNA to reach the polymerase active site, the DNA will have to be bent by about 80 degrees. Thus, the single-stranded region of the template strand will be accommodated in the DNA binding cleft in the polymerase, while the 3' end of the primer strand has access to both the polymerase and exonuclease active sites (Fig. 1.17). The 'slide-and-melt' model is still applicable here, as access of the primer to the polymerase active site requires it to bend towards the polymerase active site, but the presence of a mismatch at the primer terminus could cause melting and prevent bending or reverse the direction of bending such that the primer terminus is placed towards the exonuclease site (Beese et al., 1993).

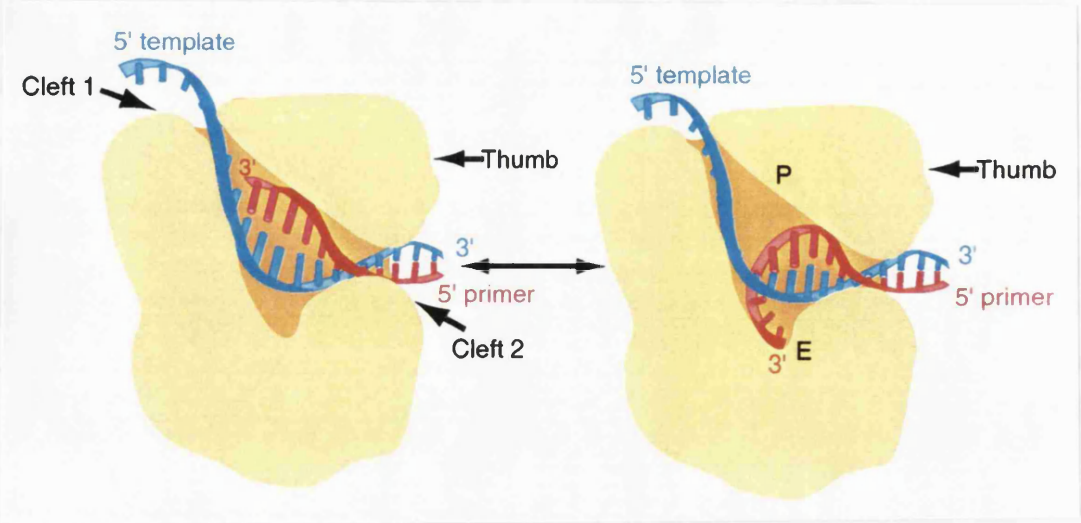


Figure 1.17. The model for DNA bound at the polymerase active site P (left) and at the 3' to 5' exonuclease site E (right) proposed by Beese et al. (1993). The primer and template strands are shown in red and blue respectively. The 3' terminus of the primer is proposed to shuttle between each active sites according to the 'slide-and-melt' model described by Joyce & Steitz (1987). Polymerisation and exonucleolytic excision can occur with or without dissociation of the enzyme from the DNA (modified from Beese et al., 1993).

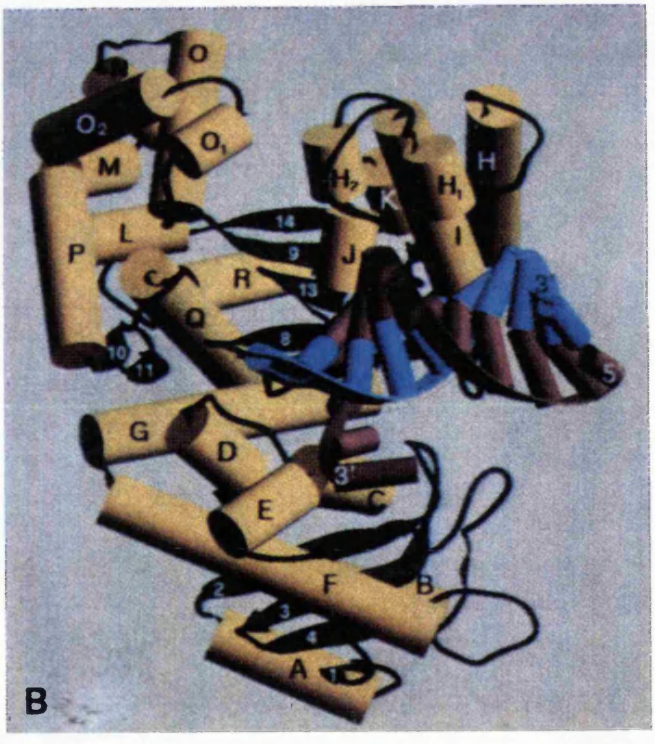


Figure 1.18. The secondary structure of Klenow shown in the same orientation as Figure 1.17. The helices and sheets are named as in Figure 1.13. The new regions of secondary structure (helices H_1, H_2, O_1 and O_2) were added on the basis of new X-ray crystallographic evidence reported by Beese et al. (1993) (diagram taken from Beese et al., 1993).

So far, the dissociation of the DNA from the enzyme has not been discussed. After a mismatched nucleotide has been incorporated, it could either become fixed as a mutation into DNA by polymerisation of the following base, or be translocated to the 3' to 5' exonuclease site to be removed, or the Klenow can dissociate from the DNA altogether so that the DNA may bind at the exonuclease site of another Klenow/exonuclease where it can be removed. The first option is unlikely (discussed earlier). Thus, after an incorrect nucleotide has been incorporated, the enzyme could have an intramolecular or an intermolecular pathway of editing, and the choice will depend on the relative rates of dissociation and translocation. The rates of Klenow-DNA dissociation have been measured and were found to be approximately 0.2 s^{-1} for correctly base-paired DNA and 3 s^{-1} for DNA containing a 3'-terminal mismatch (Kuchta et al., 1988). Thus Klenow fragment appears to edit its errors by a predominantly intermolecular pathway. *In vivo*, it is even possible that such mispaired termini could be removed by other cellular nucleases in addition to, or even instead of, the 3'→5' exonuclease of pol I.

The arrangement of the polymerase and exonuclease on separate domains as seen in Klenow fragment may be general among DNA polymerases. This arrangement is analogous to the situation in DNA polymerase III of *E. coli* where the polymerase and the 3' to 5' exonuclease activities are found in separate subunits, namely α and ε subunits respectively of the core enzyme. The many differences between pol I and pol III make it unlikely that the separation of the two active sites in both enzymes is merely an accident of evolution. Direct evidence for the sliding of DNA between the two catalytic sites has also been obtained for T7 DNA polymerase (Donlin et al., 1991). In eukaryotes, with the exception of pol δ, ε, and γ, proofreading by a 3'→ 5' exonuclease during DNA replication has been difficult to establish, possibly due to problems in purification of these polymerases. DNA polymerase α, the main enzyme responsible for replication of DNA in eukaryotes, is usually devoid of proofreading exonucleolytic activity when purified. However, there are a few reports of isolated pol α that do contain exonucleolytic activity (reviewed by Carroll & Benkovic, 1990). Pol β, the eukaryotic polymerase responsible for repairs, is also usually isolated without proofreading exonucleolytic activity, but it can associate with DNase V which contains both 3'→5' and 5'→3' exonucleases. It has also been shown that the ε subunit of *E. coli* pol III can cooperate with calf thymus DNA polymerase α (which lacks a proofreading exonuclease) to increase the fidelity of DNA synthesis (Perrino & Loeb, 1989). It is possible that in mammalian cells, a mechanism could exist where DNA polymerisation is carried out by pol α and when a mismatched base-pair is incorporated, it is excised by a separate exonuclease.

Another prokaryotic DNA polymerase, that from phage T7, must be briefly mentioned here because the kinetics of its mechanism of action will be considered later. The T7 DNA polymerase is a heterodimer consisting of the phage T7 gene 5 protein (80 kDa) and the host *E.*

coli thioredoxin protein (12 kDa). Alone, the gene 5 protein is inactive but when it is associated with the host thioredoxin, it has two catalytic activities, a 5'→3' polymerase activity and a proofreading 3'→5' exonuclease activity. Thioredoxin is a ubiquitous protein involved in a variety of cellular reactions through their redox active disulphide bridge, although this redox activity is irrelevant to the activation of T7 DNA polymerase. With respect to T7 DNA polymerase, thioredoxin is regarded as an accessory protein whose function is to increase the DNA-binding affinity of the T7 gene 5 protein, making the T7 DNA polymerase more of a processive enzyme than a distributive one, i.e. making T7 DNA polymerase incorporate more nucleotides each time it asssociates with a primer-template. Although T7 DNA polymerase lacks the 5'→3' exonuclease, this exonuclease is encoded by a separate gene in T7, gene 6 which is just downstream of gene 5. Together the products of genes 5 and 6 approximate that of *E. coli* pol I in mass.

The homology between the amino acid sequence of T7 polymerase and that of *E. coli* pol I is quite striking. There are several segments that are homologous, and eight of these homologous segments constitute most of the structures that line the DNA binding cleft (including the flexible sub-domain) in Klenow fragment. The extensive homology between these two DNA polymerases suggests that the two enzymes must have evolved from a common protein, the likelihood being that the phage acquired the bacterial enzyme at some stage in evolution.

CHAPTER 2.

Chemical synthesis and purification of regular oligodeoxynucleotides and of oligodeoxynucleotides containing O^6-methylguanine

2.1 Introduction

The oligonucleotides containing O^6-methylguanine (O^6-meG) used in the studies described in this thesis were synthesized chemically by the phosphoramidite method. Since the chemical syntheses of oligodeoxynucleotides containing O^6-meG can be carried out using either phosphotriester or phosphoramidite chemistry (Gaffney et al., 1984; Li & Swann, 1989; Smith et al., 1990), the chemistry of the phosphotriester and the phosphoramidite methods of DNA synthesis will be described here. But before the chemistry of the coupling reaction is described, the protection of the reactive groups will be considered first.

The main reactive centres in DNA are the OH groups at the 3' and 5' positions of the ribose, and the exocyclic amino groups of dA, dC and dG. To prevent unwanted reactions at these positions during synthesis of the oligonucleotide, these chemically reactive centres have to be protected.

(1) **Protection of the 5'-OH group**: The 5'-OH position is protected by a dimethoxytrityl group (DMTr). The acid-labile DMTr group can be easily cleaved with a mild acid such as dichloroacetic acid (DCA) after each synthesis cycle. Mild conditions are preferred because of the possibility of hydrolysis of the phosphodiester backbone and depurination under strong acidic conditions. The use of DMTr has another advantage in that the DMTr cation released upon treatment of the protected nucleoside with DCA is orange in colour (λ_{max} = 495 nm), and this can be used to monitor the progress and yield of synthesis.

(2) **Protection of the exocyclic amino groups**: The standard exocyclic amino protecting groups introduced by Khorana over twenty years ago are still in use today. They are benzoyl for the 6-amino group of dA and the 4-amino group of dC, and isobutyryl for the 2-amino of dG. Thymine does not require protection because it does not carry any exocyclic amino groups. These acyl protecting groups are stable under the conditions of synthesis and can be removed with ammonia after synthesis is complete.

(3) **Protection of the oxyanion of the 3'-phosphate**: Nowadays, the oxyanion of the 3'-phosphate group is protected by a cyanoethyl group, which is stable during synthesis and yet easily cleaved with ammonia after synthesis.

DNA consists of a long chain of deoxyribonucleotides joined together by 3'→5' phosphodiester bonds. The key step in the synthesis of oligonucleotides is therefore the specific formation of the internucleotide 3'→5' phosphodiester bonds. Several chemical methods for synthesizing oligonucleotides have been developed and their names reflect the chemistry used. The phosphodiester method devised in the 1960s is one of the earliest methods. This was succeeded by the phosphotriester method. Since the main cause of the low yields of the original phosphodiester method lies in the presence of an unprotected OH-group on the phosphate, protecting groups were designed for this free OH-group in order to minimise side reactions. Aryl protecting groups are now commonly used for this purpose. The aryl protecting group is sufficiently stable during synthesis and after this is complete, it can be cleaved by treatment with *syn*-2-nitrobenzaldoximate or 2-pyridinecarbaldoximate ions.

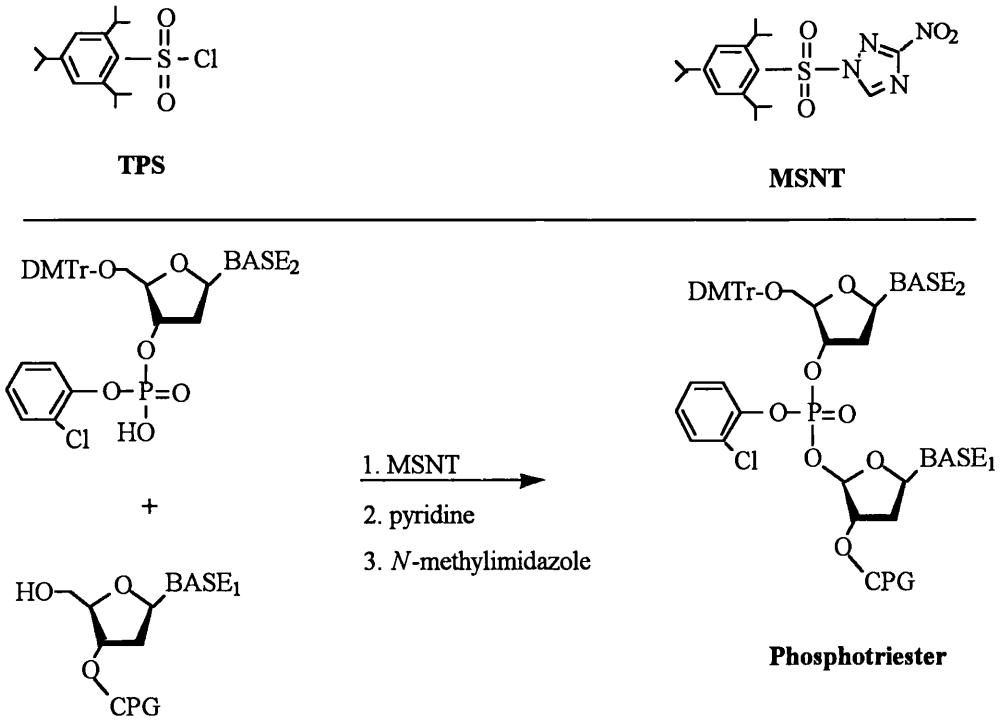


Figure 2.1. Chemistry of the phosphotriester method for DNA synthesis.

The phosphotriester syntheses were carried out either in solution or in solid-phase. In the late 1970s, oligonucleotide synthesis underwent a revolutionary change following the introduction of solid-phase synthesis by Bruce Merrifield. In solid-phase DNA synthesis, the 5'-MMTr protecting group of the first nucleotide on the 3'-end of the chain is replaced by a solid-support. As the oligonucleotide chain is anchored to a solid support, the unreacted material can simply be flushed away, eliminating the need for laborious chromatographic purification.

The phosphoramidite (phosphite triester) method

In both phosphodiester and phosphotriester methods, the phosphorus atom is pentavalent (P(V)), which is a more stable and hence less reactive state than the trivalent state (P(III)). If the pentavalent phosphorus is replaced by a trivalent phosphorus, then the coupling reaction would be faster. This is the rationale behind the phosphite triester method, also commonly known as the phosphoramidite method. In this method, the coupling reaction takes place between the 5'-OH of a support-linked nucleoside and a 2-cyanoethyl 5'-DMTr-(N-acylated)-deoxynucleoside 3'O-CN,N-diisopropylamino phosphite. The phosphite is activated by tetrazole (Fig. 2.2). The product of the coupling reaction is a dinucleotide phosphite, which has to be oxidised with aqueous iodine to a phosphotriester before the next nucleotide is added.

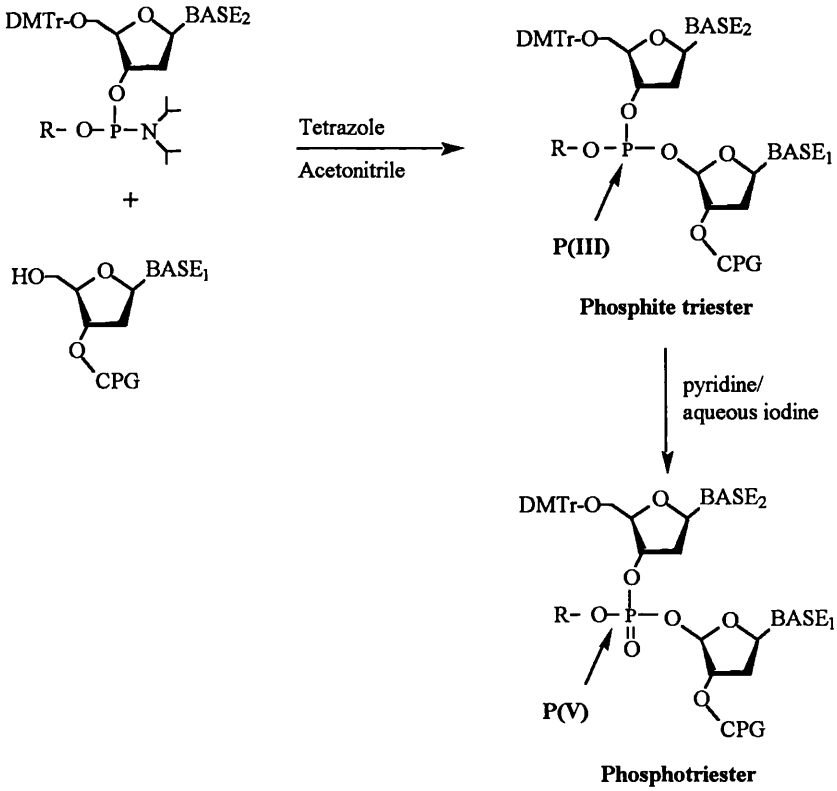


Figure 2.2. Chemistry of the phosphoramidite (phosphite triester) method of DNA synthesis.

With the phosphoramidite method, the coupling reaction rate is very fast; in a matter of a few seconds the reaction is complete and with the arrival of automated DNA synthesizers, an oligonucleotide of 15 residues long would only take less than two hours to synthesize. The efficiency of coupling is also very high, with an average of greater than 98 %.

The basic steps involved in a cycle of nucleotide addition by the phosphoramidite method is shown in Fig. 2.3. Notice that after the coupling reaction, a capping step is incorporated into the synthesis cycle to prevent any unreacted 5'-OH on the support-bound nucleoside from participating in further coupling reactions. In solution synthesis, this step is not required because the unreacted starting material would have been removed during chromatography. In solid-phase synthesis, chromatography is no longer required and any unreacted starting material will be retained with the desired product oligonucleotide on the support, and these can take part in further reactions. As synthesis proceeds, a complex series of truncated sequences will accumulate, the majority of which would only be one nucleotide shorter than the correct product. These impurities will have properties similar to the product oligonucleotide which presents a problem in purification. The capping step (with acetic anhydride) is therefore designed to avoid such problems by stopping the failure sequences from further reactions.

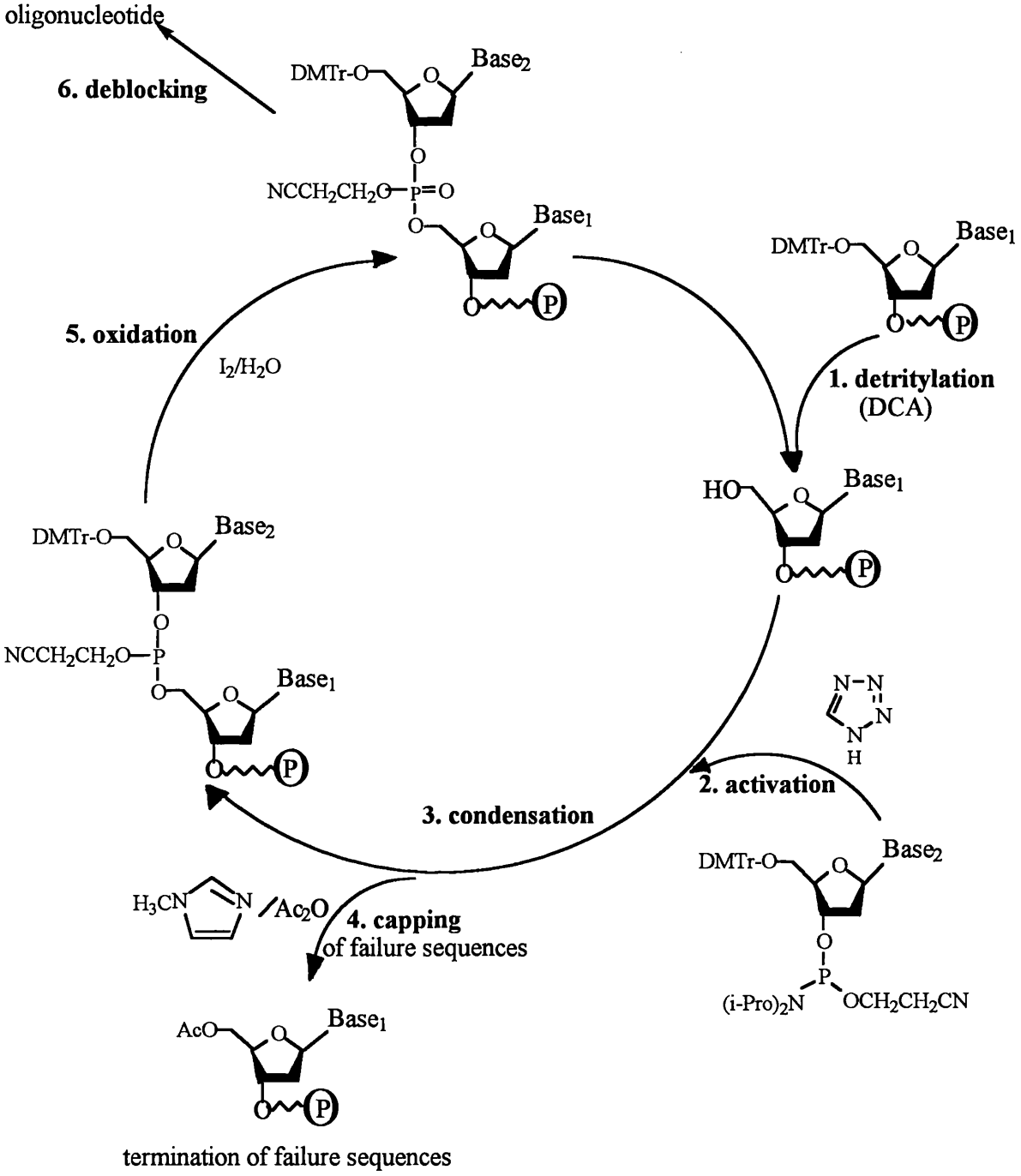


Figure 2.3. DNA synthesis cycle using phosphoramidite chemistry.

Deprotection of oligonucleotides containing only normal bases

When synthesis is complete, the oligonucleotide is deprotected with concentrated aqueous ammonia. This removes the phosphate-protecting group (2-cyanoethyl) and at the same time the link between the oligonucleotide and the controlled-pore glass support is also cleaved. The regime for deprotection with ammonia depends on the exocyclic amino-protecting groups. Of the commonly used acyl protecting groups, the protecting group that is most difficult to remove is the benzoyl group protecting the 6-amino group of adenine. For complete deprotection, it is necessary to leave the protected oligonucleotide in concentrated ammonia (> 88%) for at least 5 hours at 55 °C.

Chemical synthesis and deprotection of oligodeoxynucleotides containing O^6-methylguanine

In order to synthesize oligonucleotides containing O^6-alkylG, we need to consider the following points:

- the synthesis of a monomer of the modified base for phosphotriester or phosphoramidite synthesis;
- the stability of the modified base during synthesis, deprotection and purification;
- and finally, a method for measuring the purity of the oligonucleotide containing the modified base.

The most difficult part of the synthesis of oligonucleotides containing O^6-alkylG is the synthesis of the protected O^6-alkylG monomer. Once the monomer is made, assembly of the oligonucleotide is relatively simple. The portion of the oligonucleotide 3' to the O^6-alkylG residue is first synthesized automatically on a DNA synthesizer using standard 2-cyanoethyl phosphoramidite chemistry. The 5'-DMTr group is removed and the column containing the oligonucleotide is disconnected from the synthesizer. The protected O^6-alkylG monomer is then added manually via a clean and dry gas-tight syringe. The O^6-alkylG monomer can be added using either phosphotriester or phosphoramidite chemistry. However, with phosphoramidite chemistry the synthesis process is much simpler, faster, and the yield better. After the O^6-alkylG monomer has been added, the column is reconnected to the synthesizer and synthesis resumed using phosphoramidite chemistry.

The O^6-meG monomer used in the synthesis of oligonucleotides containing O^6-meG for the studies in this Ph.D. project was an N2-phenylacetyl-O^6-meG phosphoramidite monomer. Conventionally, the 2-amino of guanine monomers for use in DNA synthesis is protected with isobutyryl. However, the presence of an 6-*O*-alkyl group in O^6-alkylG renders this isobutyryl

protecting group very considerably more resistant to removal by ammonia, thus it is necessary to use severe ammonia deblocking conditions for complete deprotection (72 hours at 65 °C, Borowy-Borowski & Chambers, 1987). This can be potentially catastrophic. Many impurities are produced during prolonged ammonia treatment at high temperatures, in particular 2,6-diaminopurine, formed by the displacement of the alkoxy group in O^6-alkylG by ammonia. This poses a grave problem because 2,6-diaminopurine base-pairs with thymine so that DNA containing this impurity will have altered base-pairing properties from DNA containing O^6-alkylG.

There are two ways to avoid contamination with 2,6-diaminopurine. One is to deprotect with an agent other than ammonia. Thus, Gaffney et al. (1984) deblocked oligonucleotides containing O^6-meG with methoxide ions. However, this method cannot be applied to oligonucleotides containing guanine alkylated with longer chains because the methoxide may displace the 6-alkoxy group with a 6-methoxy group. For example, if one uses methoxide to deblock oligonucleotides containing N2-phenylacetyl-O^6-ethylG, the resulting oligonucleotide will have an O^6-meG in place of the desired O^6-ethylG. An alternative method is to replace the isobutyryl protecting group for the 2-amino of the O^6-alkylG with one that is more labile so that prolonged ammonia treatment at high temperatures can be avoided. For example, Gaffney et al. (1984) used an acetyl group for this protection and carried out the deprotection successfully. More recently, Li & Swann (1989) used the phenylacetyl group to protect N2 of the O^6-alkylG. This phenylacetyl protecting group has, at the nucleoside level, a half-life in ammonia of only 48 minutes at room temperature. Thus, deprotection of the N2-phenylacetyl protected O^6-alkylG is in theory faster than the deprotection of benzoyl protected adenine and cytosine, or of isobutyryl protected guanine.

Oligonucleotides synthesized by the phosphotriester method are normally deprotected by oximate ion and ammonia; the oximate ion is needed to deprotect the *O*-chlorophenyl protecting group on the 3'-phosphate used on the phosphotriester monomers while the ammonia removes the protecting groups on the bases. On the other hand, the phosphoramidite monomers have a 2-cyanoethyl protecting group on the phosphite, which is removed by ammonia treatment. Thus, deprotection of oligonucleotides synthesized by 2-cynoethyl phosphoramidite monomers require only ammonia. Li and Swann (1989) reported no difficulty in deprotecting oligonucleotides containing N2-phenylacetyl-O^6-meG synthesized by the phosphotriester method with a solution of N^1,N^1,N^3,N^3-tetramethylguanidinium salt (TMG) of *syn*-2-nitrobenzaldoxime and ammonia. However, very considerable difficulties were encountered when I attempted to deprotect oligonucleotides containing N2-phenylacetyl-O^6-meG synthesized by the phosphoramidite method. The deprotection of oligonucleotides containing O^6-meG added as a phosphoramidite was initially carried out in ammonia alone, and base analyses of the

oligonucleotides deprotected in this way showed that the normal bases were not completely deprotected by this ammonia treatment. It became clear that the oximate ion also deprotects the normal bases. Eventually, a mixture of ammonia and oximate ion was used to deprotect oligonucleotides containing O^6-meG synthesized by the phosphoramidite method. Using this method, Smith et al. (1990) successfully synthesized oligonucleotides containing O^6-methyl-, O^6-ethyl- and O^6-isobutyrylguanine with negligible formation of 2,6-diaminopurine.

Recently, phenoxyacetyl protected adenine and guanine monomers have become commercially available from Pharmacia; these monomers are very easily deblocked with ammonia (1 hour, room temperature) and are therefore the ideal choice for synthesis of oligonucleotides containing O^6-alkylG.

In the earlier syntheses, the 5'-DMTr group on the O^6-alkylG monomer was removed with a Lewis acid such as $ZnBr_2$ instead of the commonly used dichloroacetic acid for fear of depurination resulting from the susceptibility of O^6-alkylG to acid. Subsequent experiments have shown that this precaution is unnecessary if the 5'-DMTr deprotection is carried out under the normal conditions of synthesis (i.e. not more than 60-75 seconds exposure to 3 % DCA in anhydrous dichloroethane, Smith et al., 1990).

To complete the discussion on synthesis of oligonucleotides containing O^6-alkylG, a method that has been developed recently which involves the post-synthetic modification of a versatile monomer (Xu et al., 1992) will be mentioned briefly. This approach is different from the one described above in that a versatile base is incorporated into DNA during automatic synthesis and then converted into the desired modified base ***after*** synthesis by treatment with appropriate reagents. Using this method, oligonucleotides containing guanine modified at the 6-position with O-, S-, and N-derivatives including O^6-alkylG have been successfully synthesized (Xu et al., 1992). As mentioned before, the most difficult part of the synthesis of an oligodeoxynucleotide is the synthesis of the protected monomer itself. Thus, one can clearly see that the post-synthetic modification method has a great advantage in that a single synthesis of an oligonucleotide containing the versatile base provides a source of oligonucleotides each containing a different modified base. Another advantage of this method is that it also allows synthesis of oligonucleotides containing chemically reactive bases that would not normally survive the chemical treatments used in DNA synthesis.

Purification of oligonucleotides

There are two powerful methods in common use for the purification of oligonucleotides, gel electrophoresis and high performance liquid chromatography. The choice depends on the separation required, the length, quantity and recovery of the oligonucleotide to be purified.

Polyacrylamide gel electrophoresis separates oligonucleotides according to their unit charge difference. By varying the percentage of acrylamide in the gel, it is possible to resolve oligonucleotides of lengths ranging from 1 to more than 100 residues. However, this method is limited to a small scale (up to 1 mg) and recovery of the oligonucleotide from the gel is generally not very good.

With larger quantities of DNA (up to 50 mg), purification with high performance liquid chromatography (HPLC) is suitable. Two types of HPLC are commonly used, reversed phase HPLC and anion exchange. The former technique is effective in separating short oligonucleotides (up to 12 residues in length), but resolution is reduced as the length of the oligonucleotide increases and separation from impurities becomes more difficult in such cases. As reversed phase HPLC separates oligonucleotides according to their hydrophobicity, it is advantageous to leave the highly lipophilic 5'-DMTr-group on the last residue on the 5'-end of the oligonucleotide because this allows the desired oligonucleotide to be well separated from the shorter, non-DMTr-containing failure sequences. The 5'-DMTr-group can be removed after chromatography.

In our laboratory, oligonucleotides are frequently purified by anion exchange on a Pharmacia Mono-Q column, and separation of oligonucleotides of length up to 35-mers is good with very reproducible results with this column. The Dionex Nucleopac PA 100 column also gives a very good resolution of oligonucleotides.

2.2 Materials and methods

Chemical synthesis of oligodeoxynucleotides containing regular bases

Oligonucleotides were synthesized on a Cruachem PS200 automatic DNA synthesizer using 2-cyanoethyl phosphoramidite chemistry. The controlled pore glass (CPG)-linked monomers and the chemicals used on the synthesizer (DCA, tetrazole, acetic anhydride, lutidine, tetrahydrofuran (THF), *N*-methylimidazole, iodine and pyridine) were from Cruachem (Glasgow, Scotland). The 2-cyanoethyl phosphoramidites were from either Cruachem or Pharmacia. Acetonitrile (HPLC grade) was from Rathburn.

Particular care was taken to ensure that all the chemicals used in DNA synthesis were free of water. Gloves were worn when handling the chemicals and all receptacles for use on the synthesizer were carefully washed and oven-dried. Molecular sieves (4 Å) were added to the acetonitrile at least one day before use. The water content of the solvents and chemicals were checked by Karl-Fischer titration. The acetonitrile and tetrazole had less than 50 ppm water, while the other solvents had less than 200 ppm water. It is of utmost importance to remove traces of water as it reacts with the activated monomers and drastically reduces the yield of synthesis. The monomer and tetrazole solutions were prepared under nitrogen by injecting the appropriate amount of anhydrous acetonitrile into the sealed manufacturer's vials containing the reagents. The syringes and needles used for this purpose were washed with organic solvents and oven-dried before use.

Deprotection with ammonia

After synthesis was completed, the final DMTr-group was left on the oligonucleotide to assist in its separation from the failure sequences in a later step. The column (containing the unprotected oligonucleotide still attached to the CPG) was flushed with argon to remove excess solvent and then removed from the synthesizer. The CPG beads with the oligonucleotides still attached were transferred to a screw-capped micro-vial. Concentrated ammonia hydroxide (1-2 ml, sp. gr. 0.88, from BDH) was added and the vial was tightly sealed and left to incubate at 50° C for 16-20 hours. This ammonia treatment cleaved the oligonucleotides from the CPG beads and removed all the benzoyl and isobutyryl protecting group from the bases as well as the 2-cyanoethyl groups on the phosphates.

Purification of synthesized oligonucleotides

The DMTr-containing oligonucleotides were separated from failure sequences with a Du Pont NENSORB Prep cartridge according to the manufacturer's instructions (NEN Research Products, Du Pont Co., Boston, MA, USA). This purification is based on the selective elution of the non-tritylated failure sequences with a 10 % v/v acetonitrile in triethylammonium acetate buffer (TEAA, 0.1 M, pH 7). The DMTr-protected oligonucleotide of interest is retained on the hydrophobic column due to the increased hydrophobicity from the bulky DMTr-group. After the failure sequences have been eluted, the desired oligonucleotide is detritylated with 0.5 % trifluoroacetic acid in TEAA (0.1 M) and then eluted with 35 % v/v methanol in water.

The oligonucleotides were further purified on an ion exchange column (Pharmacia Mono Q HR 5-5) using the elution conditions shown in Table 6. The chromatography was carried out on a Dionex BIOLC system (Dionex Corp., Sunnyvale, CA, USA) with a Dionex variable wavelength detector. The buffers used were:

buffer A: aqueous 10 mM NaOH, 0.4 M NaCl (pH 12), and

buffer B: aqueous 10 mM NaOH, 0.8 M NaCl (pH 12).

The flow rate was 0.8 ml/min and chromatography progress was monitored at 260 nm. Up to 1 A_{260} unit of oligonucleotide was injected each time for preparative purposes. The fractions containing the purified oligonucleotide were collected and immediately neutralized with 0.1 M HCl. The oligonucleotides were then desalted on a NAP-10 column (Pharmacia) according to the maker's instructions.

Table 6. Buffers and elution gradients for the purification of oligonucleotides on a Pharmacia Mono Q HR 5-5 column.

Time (min)	Buffer A 0.4 M NaCl, 10 mM NaOH	Buffer B 0.8 M NaCl, 10 mM NaOH
0	100 %	0 %
2	100 %	0 %
5	60 %	40 %
25	0 %	100 %
32	100 %	0 %
35	100 %	0 %

Synthesis of oligomer containing O^6-meG using N2-phenylacetyl-O^6-methylguanine phosphoramidite monomer

The portion of the oligonucleotide 3' to the O^6-meG residue was synthesized using solid-phase phosphoramidite chemistry as described earlier. The DCA wash containing the DMTr cation released from the final 5'-OH was collected and used to quantitate the amount of free 5'-OH available for coupling with the N2-phenylacetyl-O^6-meG phosphoramidite monomer as follows: a sample (1 ml) of this DCA wash was mixed with 5 ml of a mixture of *p*-toluenesulphonic acid (0.1 M) in acetonitrile (1:1, v/v), and the absorbance read at 495 nm. Assuming a molar extinction coefficient for DMTr cation of 70 000 $M^{-1}cm^{-1}$, the amount of free 5'-OH terminals available for the next condensation was calculated. Ten equivalents of the O^6-meG monomer (relative to the amount of free 5'-OH available for coupling) was carefully weighed out into a 2 ml conical glass vial with septum top (Wheaton reacti-vial) and dissolved under dried nitrogen in anhydrous acetonitrile (0.1 ml) and 0.5 M tetrazole (0.1 ml). The column containing the 3'-portion of the oligonucleotide was dried by flushing with argon for a few minutes. The inlet end of the column was disconnected from the DNA synthesizer and the O^6-meG monomer/tetrazole mixture injected slowly into the column from a gas-tight syringe. With the same syringe, the mixture was drawn very slowly in and out of the column for thorough mixing over a period of 3 minutes. The column was then reconnected to the synthesizer to complete the synthesis. After washing with acetonitrile, the synthetic cycle was resumed at the capping step. The remainder of the sequence was synthesized automatically. The final DMTr group was left on the oligonucleotide at the end of synthesis to aid the subsequent purification of the oligomer.

Deprotection and purification of oligonucleotide containing O^6-meG

After synthesis, the CPG support with the oligonucleotide still attached was removed from the column and put into a small (25 ml) round-bottomed flask. 0.075 M *syn*-2-nitrobenzaldoxime and 0.0625 M TMG (both from Aldrich) dissolved in aqueous ammonia solution (2 ml, sp. gr. 0.88) was added and left at room temperature for 30 hours. The ammonia was then evaporated under reduced pressure and the oligonucleotide redissolved in triethylammonium acetate buffer (4 ml, 0.1M, pH 7.0).

The DMTr-containing oligonucleotide was partially purified and the DMTr-group removed with a Du Pont NENSORB Prep cartridge, further purified by FPLC on a Mono-Q HR 5-5 column (Pharmacia), neutralized with HCl and finally desalted on a NAP-10 column as with the normal oligonucleotides.

Characterisation of oligomer containing O^6-meG by base-composition analysis

The purity of the oligomer containing O^6-meG was assessed by base analysis. About 0.3-0.5 A_{260} units of the oligomer was completely digested to nucleosides by snake-venom phosphodiesterase (*Crotalus durissus*, from Boehringer-Mannheim) and calf intestine alkaline phosphatase (also from Boehringer-Mannheim). Snake-venom phosphodiesterase solution (10 µl, containing 10 µg enzyme, i.e. 0.05 units) was added to the oligonucleotide (0.5 A_{260} units in 120 µl, 60 mM Tris-HCl, 6 mM $MgCl_2$, pH 8.5) and incubated at 37 °C for 30 minutes. Alkaline phosphatase solution (10 µl, containing 5 µg enzyme, i.e. 5 units) was then added and left to incubate for a further 20 minutes. The enzymes were then heat-inactivated (80 °C, 3 minutes), and the samples kept on ice until ready for chromatographic separation.

The deoxynucleosides were separated by reversed-phase HPLC using a Waters Nova-pak phenyl (8MBPH) 4 µ cartridge. The HPLC was carried out on a Gilson 320 HPLC apparatus. Duplicate samples (60 µl) of the nucleosides were injected onto the column and eluted at 3 ml/min. with a gradient set up by two buffers (buffer A: 50 mM KH_2PO_4, pH 4.5; buffer B: 50 mM KH_2PO_4 in 35 % acetonitrile). Details of the chromatographic conditions are shown in Table 7. The eluate was monitored at 260 nm with a Shimadzu SPD6A UV spectrophotometric detector and the area under each peak was integrated with a Gilson 620 Datamaster module. Comparison of the peak areas on the chromatogram of the sample Fig. 2.5b with that of a standard mixture of nucleosides (Fig. 2.5a) gives the base ratio in the oligonucleotide (see legend).

Table 7. Buffers and elution gradients for the separation of nucleosides by reverse phase HPLC using a Waters Nova-pak phenyl (8MBPH) 4 µ cartridge.

Time (min)	Buffer A 50 mM KH_2PO_4, pH 4.5	Buffer B 50 mM KH_2PO_4 in 35 % acetonitrile
0	96 %	4 %
8	96 %	4 %
12	30 %	70 %
14	96 %	4 %
18	96 %	4 %

3.3 Results

The oligonucleotide containing O^6-meG, after purification, gave a single peak when chromatographed by anion exchange HPLC on a Pharmacia Mono-Q HR 5-5 column (Fig. 2.4).

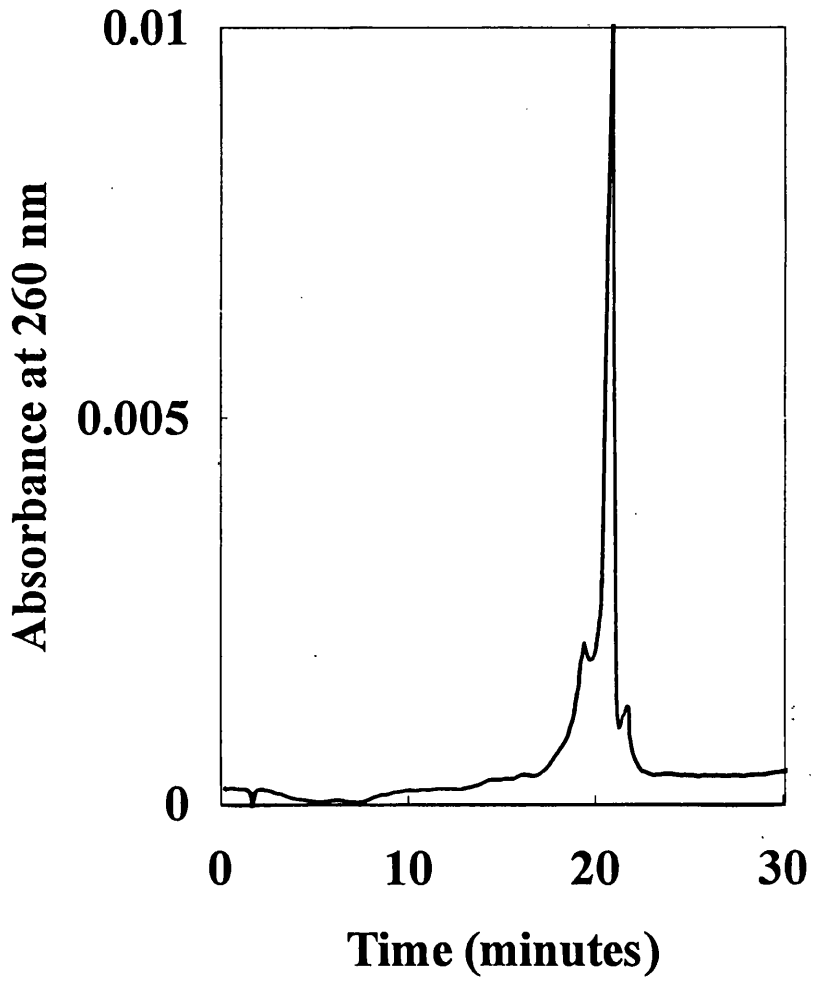


Figure 2.4. FPLC profile of the 20-mer oligonucleotide containing O^6-meG analysed on a Pharmacia Mono-Q HR 5-5 column. The gradient and buffers used for the chromatography are given in Table 6.

The separation of deoxynucleosides by reverse phase HPLC on a Waters Nova-pak phenyl (8MBPH) 4 μ cartridge is shown in Figure 2.5. Two HPLC profiles are shown, one of the separation of a standard mixture of deoxynucleosides and the other is the separation of the mixture of deoxynucleosides obtained from the enzyme-digest of the chemically synthesized oligomer containing O^6-meG. The area of each peak was integrated with a Gilson 620 data module, and comparison of the peak areas in the chromatogram of the enzyme-digest (Figure 2.5b) with that of the standard mixture of nucleosides (Figure 2.5a) gave the base ratio in the oligonucleotide. The ratio of bases in the oligonucleotide agreed with that expected from its sequence: 3'TAGGCTATCmeGTCATTCTCGC.

	Cytidine	Guanosine	Thymidine	Adenosine	O^6meGua
Expected base ratio:	6	3	7	3	1
Calculated base ratio:	6	3.2	7.3	2.8	1.2

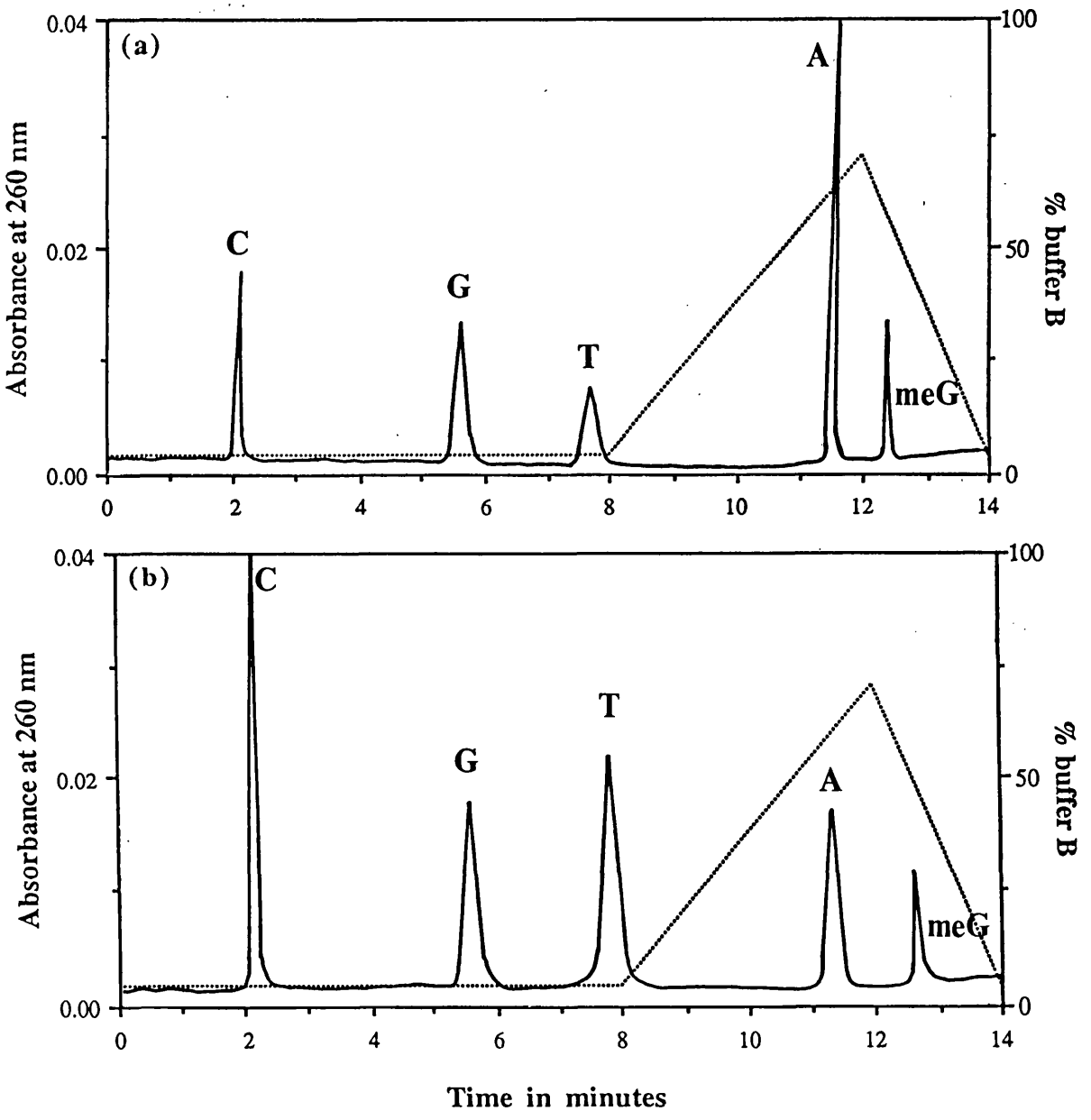


Figure 2.5. Reverse phase HPLC profiles showing the relative elution positions of a solution of standard nucleosides (a) and an enzyme digest of the chemically synthesized 20-mer containing O^6-meG (b). The gradient is represented by a broken line in each graph, with reference to the vertical axis on the right. The elution positions of the deoxynucleosides are indicated.

Chapter 3.

Kinetic analysis of the coding properties of O^6-methylguanine in DNA: the crucial role of the conformation of the phosphodiester bond

---

3.1 Introduction

This chapter describes the measurement of the rates of incorporation of thymine and of cytosine opposite O^6-methylguanine (O^6-meG) in a template DNA strand by the Klenow fragment of *E. coli* DNA polymerase I, and how the kinetic data obtained were subsequently analysed by computer non-linear regression to give a mechanistic kinetic scheme for these incorporations. An introduction to the problem of misincorporation by a template O^6-alkylguanine (O^6-alkylG) during DNA synthesis has been made in Chapter 1, so only a brief summary of this will be given here. The main part of this introduction is spent on the consideration of the various models which have been proposed to explain how DNA polymerases achieve the high level of fidelity of DNA synthesis observed. In addition, the methods used to study these models will also be discussed in detail because they played an important role in the elucidation of the models of fidelity of DNA synthesis. These methods are particularly important in the context of this thesis because different methods of analysis produce different results. A year or so after this Ph.D. project had started, Singer et al. (1989) published a paper which addressed the same question of how O^6-meG directs the misincorporation of thymine during DNA synthesis. These authors based their analyses on steady state experiments originally designed by Boosalis et al. (1987) (discussed later). Two years later, Dosanjh, Galeros, Goodman & Singer (1991) published another paper describing the effects of neighbouring bases on the incorporation of thymine and of cytosine opposite an O^6-meG in the template DNA strand, using techniques similar to those previously used by Singer et al. (1989). However, the results of this study differed from those reported by Singer et al. (1989; discussed in Section 3.5), and as will become apparent later, their results are difficult to interpret because of the methods they used in these studies. For this reason, a method different to that used by Singer et al.(1989) and Dosanjh et al. (1991) was chosen for this project, and the difference in the approach explains why the results from this study are so different from those of Singer et al. (1989) and Dosanjh et al. (1991).

As described in greater detail in Chapter 1, the promutagenic alkylated bases O^6-alkylG (Loveless, 1969) and O^4-alkylthymine (Lawley et al., 1973) which the *N*-nitroso compounds produce during their reaction with DNA, play the most important role in the carcinogenic action of these compounds, possibly through the induction of mutations in protooncogenes (Zarbl et al., 1985; Sukumar, 1990; Kumar et al., 1990). The presence of O^6-alkylG in DNA decreases the rate of DNA synthesis (Snow et al., 1984) and directs the misincorporation of thymine on replication of the alkylated DNA (Loveless, 1969; Gerchman & Ludlum, 1973; reviewed by Saffhill et al., 1985). It had been proposed, and generally accepted, that misincorporation of thymine opposite O^6-alkylG is the result of the formation of a more stable base-pair between O^6-alkylG and thymine than between O^6-alkylG and cytosine (Fig. 1.8 c & d).

The original structures for mispairing by O^6-alkylG would suggest that an O^6-alkylG:T base-pair is more stable than an O^6-alkylG:C base-pair because the O^6-alkylG:T base-pair has two H-bonds whereas the O^6-alkylG:C base-pair has only one H-bond. However melting studies of DNA duplexes containing O^6-meG show that O^6-meG:T base-pairs are energetically less stable than O^6-meG:C base-pairs (Gaffney et al., 1984; Gaffney & Jones, 1989). NMR studies of DNA duplexes containing either O^6-meG (Patel et al., 1986a, b) or O^6-ethylguanine (Kalnik et al., 1988a, b; reviewed by Li et al., 1988) also show that the O^6-alkylG:T pair is less H-bonded than the O^6-alkylG:C pair (Fig. 1.8 f & g). These NMR derived structures of the O^6-alkylG:T base-pair have been verified very recently by Goswami and Jones (1993). Thus O^6-alkylG like 2-aminopurine (Mhaskar & Goodman, 1984) and xanthine (Eritja et al., 1986) directs the incorporation of its less favoured partner. The objective of this work is to examine this paradox in the hope that its resolution might add to our knowledge of the underlying factors in DNA replication as well as throwing light on the mechanism of action of the carcinogenic nitrosamines.

The work in this thesis analysed the incorporation of thymine and of cytosine opposite O^6-meG in the template DNA strand, using steady state and pre-steady state methods with Klenow fragment of *E. coli* DNA polymerase I. The steady state assays were carried out with an excess of DNA primer/template over the polymerase and the reactions were initiated by the addition of enzyme to a solution of DNA and dNTP; this is similar to all classical studies of enzyme kinetics. The pre-steady state assays were performed with a slight excess of enzyme over DNA, and in these assays the enzyme-DNA complex was pre-formed in the presence of EDTA, the reactions were initiated by the addition of dNTP and magnesium, and a single catalytic cycle was followed. The results of the experiments were then analysed by detailed mathematical analysis ***to identify at which step or steps the rate constants differ between the incorporation of thymine and cytosine and thus lead to the preferential incorporation of thymine opposite***

O^6-meG. Kinetic mechanistic models for the incorporation of thymine and of cytosine opposite O^6-meG in the template DNA were solved by fitting the models to the experimental data by non-linear least squares analysis. This was carried out by a computer program called FACSIMILE which is specially designed to solve systems containing large numbers of differential equations by numerical integration and to perform non-linear least square (also known as non-linear regression) fit of the rate curves to experimental data in order to arrive at a solution that best fits the observed data. This will be discussed later and in greater detail in section 3.4.

The question of why DNA polymerases select thymine rather than cytosine for incorporation opposite O^6-meG in the template DNA strand is part of the general question of how DNA polymerases select the right base for incorporation into DNA. To understand this it is necessary to look at the mechanisms that contribute to fidelity of DNA synthesis. Replication of DNA is catalysed by DNA polymerases with speed and high fidelity. For example, *E. coli* is capable of adding 1000 base-pairs/second with an error frequency of only 10^{-8} to 10^{-10} error per nucleotide incorporated (Englisch et al., 1985). This remarkably high fidelity is the result of several error-avoiding mechanisms including base-selection and exonucleolytic proofreading.

Many models have been put forward to explain how DNA polymerases achieve a high degree of fidelity during polymerisation (Loeb & Reyland, 1987; Radman & Wagner, 1988; Echols & Goodman, 1991). In one of these models, a very simple view of the kinetics of polymerisation was perceived (Boosalis et al., 1987). The incorporation of a nucleotide into DNA by a polymerase-DNA complex was seen as a two-step reaction: the polymerase-DNA complex binds the nucleotide, followed by the formation of the phosphodiester bond and release of the pyrophosphate. In such a model, nucleotide selection can occur either by a better binding of the correct dNTP, or a faster rate of incorporation of the correct dNTP. Discrimination based on nucleotide binding is reflected in the K_ms of correct and incorrect dNTPs, while discrimination based on the rate of catalysis is reflected in the V_{max} values of the correct and incorrect dNTPs.

The K_m-discrimination model is essentially the same as the passive polymerase model proposed by Goodman and his co-workers (Goodman et al., 1983). As mentioned before, it is widely believed that differences in the energy of base-pairing between complementary and non-complementary nucleotides provide the basis for discrimination against selection of an incorrect nucleotide during DNA synthesis. Accordingly, Goodman et al. proposed that differences in H-bonding between competing base-pairs are the principal determinant of DNA replication fidelity. If the incoming dNTP is complementary to the template base, then it will be tightly bound. Otherwise, a mismatched dNTP will quickly dissociate from the polymerase-DNA complex. The

polymerase is described as passive because it does not in itself discriminate but merely exploits the differences in formation of H-bonds as a means of selection for the correct base.

However, the mutagenic properties of base analogues like O^6-meG, O^4-methylthymine (O^4-meT), and 2-aminopurine (2-AP) are in contradiction with the passive polymerase model. The miscoding properties of O^6-meG has already been discussed in Chapter 1. It directs the misincorporation of thymine rather than cytosine during DNA replication despite the greater stability of the O^6-meG:C base-pair over that of the O^6-meG:T base-pair. Likewise, O^4-meT prefers to base-pair with guanine instead of adenine in DNA synthesis even though the O^4-meT:A base-pair is thermodynamically more stable than the O^4-meT:G base-pair. For 2-aminopurine, DNA containing 2-AP:A wobble base-pairs was found to be more stable than that containing 2-AP:C protonated base-pairs, yet the polymerase inserts cytosine opposite 2-AP more efficiently than it does adenine (Mhaskar & Goodman, 1984; reviewed by Echols & Goodman, 1991). The same discrepancy also appears to be true of the base xanthine. Cytosine is more readily inserted than guanine opposite a template xanthine in spite of the greater stability of the X:G base-pair over that of the X:C base-pair (Eritja et al., 1986).

These phenomena described above suggest that DNA polymerases strongly prefer to insert nucleotides in the configuration closest to a Watson-Crick base-pair even when the base-pair may be aberrant. The failure of the simple view that the differences in the energy of base-pairing between complementary and non-complementary nucleotides is the major determinant of incorporation brings us to the V_{max}-discrimination model, which proposes that once dNTP is bound to a polymerase-DNA complex, there may be another discrimination step where the non-Watson-Crick geometry of a mismatched base-pair may slow down the rate of formation of the phosphodiester bond (discussed by Boosalis et al., 1987).

If we can assess the relative contribution of the K_m- and V_{max}-discrimination mechanisms, this will tell us whether the formation of H-bonds between complementary and non-complementary base-pairs is the major determinant of DNA synthesis fidelity and whether the geometry of the base-pair has significant effects on DNA synthesis. A simple way to tell whether a polymerase follows a K_m- or a V_{max}-discrimination model is to measure the steady state K_m and V_{max} values during polymerisation. A method was devised for this purpose by Boosalis et al. (1987) from the laboratory of Goodman; this method has been reviewed by Echols & Goodman (1991) and very recently by Goodman (1993). Boosalis et al. based their studies on a Michaelis-Menten model of DNA polymerisation and performed steady state polymerisation using an excess of DNA over polymerase to measure the K_m and V_{max} of polymerisation by polymerase α from *Drosophila*.

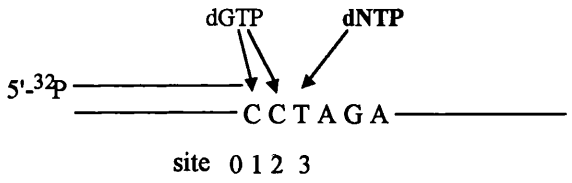
The model of polymerisation used was:

$$E.D + dNTP \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E.D.dNTP \xrightarrow{k_{cat}} E + D_1 + PPi$$

where E = enzyme, D = DNA, and D_1= elongated DNA (Boosalis et al., 1987). This kinetic scheme involves only two steps: binding of nucleotide to a polymerase-DNA complex followed by simultaneous catalysis and release of pyrophosphate.

In the method devised by Boosalis et al. (1987; see fig. 3.1), polymerisation of correct versus incorrect nucleotides is examined at a specific site. The primer is designed such that the target polymerisation site is located three bases downstream from the 3' terminus of the primer, so that before reaching the target site, the polymerase must first insert two correctly base-paired nucleotides at sites 1 and 2 (Fig. 3.1). The important point of this design is that the polymerisation event to be monitored (at site 3) starts with the polymerase already bound to the DNA, as required by the kinetic scheme above. The amount of enzyme bound to DNA (i.e. the E.D complex) before incorporation opposite site 3 is measured by the intensities of bands corresponding to incorporation at sites 2 and 3.

Figure 3.1. Schematic representation of primer-template designed for the gel fidelity assay devised by Boosalis et al. (1987).



The polymerisation reaction is carried out by pre-incubating the polymerase with a large excess of primer-template and then initiating the polymerisation reaction by adding dNTP and magnesium ions. To measure incorporation at site 3, incorporation at sites 1 and 2 have to be at maximal velocity, so a saturating concentration of the correctly paired dNTP for incorporation into these two sites is added. Boosalis et al. (1987) used 50 μM dGTP for incorporation opposite two cytosine bases at sites 1 and 2 (K_m for incorporation of correct dNTP is around 2-5 μM). After a suitable period of time, the reaction is quenched by EDTA. The primer is labelled with ^{32}P on its 5' terminus so that primers extended by one, two or three nucleotides can be separated by polyacrylamide gel electrophoresis and then visualized by autoradiography. The bands corresponding to each product of elongation can then be cut out and the amount of radioactivity quantitated by scintillation counting. Alternatively, the intensity of each band can be measured by densitometry. If the intensities of bands 2 and 3 are I_2 and I_3 respectively, then the ***relative rate*** of incorporation at site 3, v_{23}, is related to I_2 and I_3 by the following equation:

$$v_{23} = \frac{(I_2 + I_3)}{time} \cdot \frac{(I_3)}{(I_2)} \qquad \text{(i)}$$

To obtain an absolute value, I_2 and I_3 will have to be divided by the total intensities of all the bands ($I_0 + I_1 + I_2 + I_3$) and multiplied by the total amount of primer present.

A problem with this approach is that the enzyme could dissociate from the DNA at any time. For example, the enzyme may incorporate bases at sites 1 and 2 and then dissociate from the DNA before binding another DNA. Goodman recognised this problem of enzyme dissociation, and measured the ***relative rates*** of incorporation instead of the absolute rates to overcome this problem. The principle behind measurement of relative rates is that incorporation at site 3 can only take place when the enzyme is bound to the DNA which had already been elongated until site 2. In equation (i), the term (I_2 + I_3) represents the total amount of E.D complex which could incorporate a base at site 3; out of these E.D complexes, only a fraction will actually incorporate a base at site 3, and this fraction is given by the ratio of I_3/I_2. This consideration thus eliminates the problem of enzyme dissociation. However, the model assumes that once the enzyme dissociates from the DNA, it will only bind unelongated DNA but not the DNA which had already been elongated by 1, 2 or 3 bases, an assumption that is not valid. This has serious implications in the interpretation of the results and will be discussed later.

By obtaining the relative rates of incorporation at different concentrations of dNTP, K_m and V_{max} can be determined, for example from an Eadie-Hofstee plot of v against v/[dNTP]. Using this method, Boosalis et al. (1987) measured the K_m of incorporation of A, G, T and C opposite a template thymine at 4 μM, 4.2 mM, 10 mM and 100 mM respectively. Thus, there is a very big range of values of K_m for incorrect nucleotides, ranging from 1000- to 25000-times the control value depending on the nucleotide inserted. The effect of an incorrect nucleotide on V_{max} was found to be less dramatic; they found only a 4-fold reduction in the V_{max} when an incorrect nucleotide was incorporated. These results suggested that polymerase α from *Drosophila* maintains fidelity of DNA polymerisation mostly by K_m-discrimination.

According to the K_m-discrimination model, a polymerase selects the correct nucleotide by exploiting the differences in the energy of base-pairing between complementary and non-complementary nucleotides to provide a basis for discrimination against selection of an incorrect nucleotide during DNA synthesis. However, the free energy differences between matched and mismatched base-pairs are insufficient to account for the accuracy of DNA synthesis even by DNA polymerases lacking exonucleases (Preston et al., 1988). With a few exceptions in the literature, this difference in free energy is usually estimated at 0.2-0.4 kcal/mol for terminal base-pairs and 1-3 kcal/mol for internal base-pairs (reviewed by Echols & Goodman, 1991). If we write the Gibbs' free energy difference between the correct and incorrect base-pairs as ΔΔG,

at equilibrium and in an aqueous environment, $-\Delta\Delta G = -(\Delta G_{correct} - \Delta G_{incorrect}) = -RT\ ln$ (incorrect/correct). If ($-\Delta\Delta G$) is only 1- 2 kcal/mol at the most, this small free energy difference would only account for a level of discrimination (i.e. the ratio of incorrect:correct base incorporated) of 1 in 10 to 1 in 100 (Preston et al., 1988).

In an attempt to overcome this difficulty, the passive polymerase model was taken one step further by Goodman and his colleagues by assuming that the polymerase could magnify the free energy difference between a correct and an incorrect base-pair (Petruska et al., 1988). In an aqueous environment, the free energy differences between correct and incorrect base-pairs are small because of "entropy-enthalpy" compensation: the relatively large differences in enthalpy ($\Delta\Delta H$) are nearly cancelled by the large differences in entropy ($T\Delta\Delta S$) according to the equation $\Delta\Delta G = \Delta\Delta H - T\Delta\Delta S$. However, Petruska et al. (1988) argued that the situation in the polymerase may be very different if water is excluded from it. In this way, the free energy differences between a correct and incorrect base-pair would not be compensated for by a large difference in entropy.

Kinetic studies by El-Deiry et al. (1988) using DNA polymerase I from *E. coli* suggested that pol I relies on V_{max}-discrimination to achieve fidelity of DNA synthesis. Like Boosalis et al. (1987), they too measured K_m and V_{max} of incorporation using steady state methods. However, instead of the specially designed primer-template used by Boosalis et al. (1987), they used an oligo(dA)-poly(dT) primer-template for their studies, and did not pre-incubate the enzyme with the primer-template before the start of reaction. However, since the association between the enzyme and the primer-template is governed by diffusion, which is very fast, the fact that the polymerisation reactions in El-Deiry et al.'s experiments did not start with a polymerase-DNA complex did not really matter because the association would have occurred very rapidly in relation to the time period over which measurements were made. They measured the K_m and V_{max} of incorporation of dATP (correct nucleotide) and dGTP (mismatched nucleotide) into the oligo(dA)-poly(dT) primer-template. The results showed a 1600-fold difference in the V_{max} for insertion of dATP relative to that for the insertion of dGTP. The difference in the apparent K_m for the insertion of dATP relative to that for insertion of dGTP was only 6-fold. Thus, *E. coli* DNA polymerase I appeared to be a good example of a polymerase that discriminates against incorporation of an incorrect nucleotide by V_{max}-discrimination.

These studies on different DNA polymerases gave different conclusions about how DNA polymerases achieve fidelity during polymerisation, i.e. whether K_m or V_{max}-discrimination is more important in DNA polymerisation. This difference could at first be ascribed to the different behaviour of polymerases with dissimilar primer-templates. However,

there is a fundamental problem with these studies, and that is the fact that these studies were based on the assumption that DNA polymerisation follows Michaelis-Menten kinetics. A solution to this kind of kinetics is usually solved by steady state experiments because during the steady state, the flux of the reactants through each step of the reaction is constant, and this assumption allows the differential rate equations to be simplified so that a solution can be obtained algebraically. Hence, Boosalis et al. (1987) and El-Deiry et al. (1988) carried out steady state studies to analyse the kinetics of DNA polymerisation.

Steady state studies can only give limited insight into enzyme mechanisms, and as pointed out by Fersht (1985), the basic weakness of steady state studies is that "no direct information is obtained about the number of intermediates, so the minimum number is always assumed". As it turned out, the mechanism of polymerisation is more complex than was previously believed when the K_m- and V_{max}-discrimination models of DNA polymerisation were proposed. The K_m- and V_{max}-discrimination models were based on a Michaelis-Menten kinetic scheme, involving only two steps: binding of nucleotide followed by simultaneous catalysis and release of pyrophosphate (Boosalis et al., 1987). This is an over-simplistic view of the mechanism of DNA polymerisation, and has been proven so by the work on Klenow fragment and T7 DNA polymerase from the laboratories of Benkovic and of Johnson respectively, which showed that polymerisation of a nucleotide by a Klenow-DNA or a T7 DNA polymerase-DNA complex involves at least five discernible kinetic steps (Fig. 3.5). These experiments will be discussed in greater detail later. For the purpose of the present discussion, suffice it to say that the results of these experiments showing the complex mechanism of polymerisation have illustrated clearly that in the absence of mechanistic information, it was most unsatisfactory to propose the K_m- and V_{max}-discrimination models of DNA polymerisation on the assumption that the polymerisation obeys Michaelis-Menten kinetics. Even in a Michaelis-Menten type reaction, the so-called K_m is equal to the true binding constant only when the rate of the catalytic step is much slower than the rate of the substrate binding step, i.e. when $k_{cat} << k_2$ for the reaction scheme,

$$E.D + dNTP \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E.D.dNTP \xrightarrow{k_{cat}} E + D_1 + PPi$$

Thus, K_m-discrimination should not be misconstrued as necessarily meaning discrimination during the binding of dNTP. In the absence of mechanistic information, a general rule that is useful is to treat the K_m as "the overall dissociation constant of all enzyme-bound species" (Fersht, 1985). Similarly, V_{max}-discrimination may not necessarily reflect discrimination during the formation of phosphodiester bond.

Steady state techniques have the advantage that only standard laboratory equipment is needed, compared with pre-steady state methods which require devices for measuring rapid

reactions. More importantly, relatively little enzyme is needed for steady state studies and this is a great advantage considering the difficulties involved in obtaining large quantities of some polymerases, especially the protease-sensitive eukaryotic replication complexes (reviewed by Carroll & Benkovic, 1990). However, the results observed in steady state studies are very difficult to interpret. Polymerisation is a multistep process, involving first the binding of the primer-template and then a dNTP to the polymerase, followed by conformational changes and phosphodiester bond formation and etc. In steady state experiments involving the addition of a single nucleotide, the enzyme turns over several times as there is a large excess of DNA over polymerase. The polymerase has to dissociate from the product DNA after each polymerisation event and associate with another substrate molecule before a second round of polymerisation can take place. Therefore steady state polymerisation involves not one but many steps, and the measured steady state rate may be due to some rates involved in phosphodiester bond formation or enzyme conformational changes or dissociation of the enzyme from the product DNA or a combination of any of the three. The steady state rate does not provide information on each of these steps. It is only useful in supplying the velocity of the rate-determining step in the polymerisation reaction. In general steady state experiments will only allow detection of that intermediate which immediately precedes the rate-limiting step.

For the incorporation of a correct nucleotide into DNA by Klenow fragment and by T7 DNA polymerase, the rate-limiting step during steady state polymerisation with excess DNA over enzyme is due to the slow dissociation of the enzyme from the polymerase (Kuchta et al., 1987; Patel et al., 1991). This is seen very clearly in the biphasic kinetic profile for the incorporation of a single nucleotide into DNA by Klenow fragment, which has a fast initial rate (the pre-steady state burst) corresponding to the elongation of the pre-existing polymerase-DNA complex, followed by a slower linear rate when the rate of elongation becomes limited in part by the need for the enzyme to dissociate from the DNA after addition of the nucleotide (Kuchta et al., 1988; Patel et al., 1991; reviewed by Carroll & Benkovic, 1990; see Figure 3.3, page 85).

In this light, it is not surprising that the conclusions from the work of Boosalis et al. (1987) and El-Deiry et al. (1988) are different. In the studies by Boosalis et al., the design of the primer-template was such that at most three nucleotides were added during each encounter between the polymerase and the primer-template. As mentioned earlier, although the authors recognised this fact and claimed to have overcome this problem by making allowances for dissociation of polymerase from DNA in their calculations, the model that they proposed assumed that this dissociation is reversible only for the substrate DNA, but not for product DNA. No explanation for this was given, but presumably this assumption was made so that there would be less unknown rate constants and a solution could then be arrived at algebraically. Since the steady state rate of misincorporation carried out by Boosalis et al. (1987) used an

enzyme:DNA ratio of 1:30, their reactions were probably limited by the dissociation of enzyme from the DNA rather than the rate of polymerisation, and the rates measured were probably the rates of dissociation of the polymerase from the product DNA. As discussed by Wong et al. (1991) in their paper about the mechanism of misincorporation by T7 DNA polymerase, the value of 2.2 s^{-1} reported by Boosalis et al. (1987) for the V_{max} of polymerase α "seems much too slow for a polymerisation rate but falls, rather, in the expected range for a DNA dissociation rate".

The experiments by El-Deiry et al. (1988) were also carried out with excess DNA over enzyme (enzyme:DNA ratio was 0.17 nM: 285 nM), but because the primer-template used was $(dA)_{14}$-$(dT)_{350}$, the polymerisation carried out was less distributive than that carried out by Boosalis et al., i.e. the enzyme could incorporate more than one nucleotide before dissociating from the DNA. In this case, their polymerisation reaction was less likely to have been limited by an enzyme dissociation event, and the rates measured were more likely to be that of the slowest step before or during polymerisation. However, these rates could be a combination of rates from several steps and cannot be assigned to any step occurring in the polymerisation process. Thus, based on the results reported by El-Deiry et al. (1988), it is impossible to say whether discrimination by the polymerase occurs by nucleotide binding or by formation of phosphodiester bond.

For the reasons above, Benkovic and Johnson, working together at first and then separately, introduced pre-steady state methods to the studies of DNA polymerisation. The incorporation of a single nucleotide into a primer-template by a DNA polymerase is carried out in an instrument for rapid reactions. An excess of polymerase over DNA is used so that all the DNA is present as a polymerase-DNA complex before the nucleotide triphosphate is added to initiate the polymerisation. The catalytic actions of the Klenow fragment including both polymerisation and 3' to 5' exonucleolytic activity depend on the presence of magnesium ions. In order to form an enzyme-DNA complex prior to the start of polymerisation and to avoid exonuclease activity in this complex, EDTA is added with the polymerase and the DNA. Polymerisation is initiated by the simultaneous addition of magnesium ions and dNTP.

Using a combination of steady state, pre-steady state and some other techniques, the mechanism of polymerisation by Klenow fragment has been extensively studied by Benkovic and his co-workers (Bryant et al., 1983; Mizrahi et al., 1985; Kuchta et al., 1987, 1988; Dahlberg & Benkovic, 1991; Eger & Benkovic, 1992) while that of T7 DNA polymerase has been investigated by Johnson and his co-workers (Patel et al., and Wong et al., 1991). Since the enzyme used in the work in this thesis is Klenow fragment, a detailed description of how the kinetic schemes for correct incorporation and misincorporation by this enzyme is necessary. A

summary of the important discoveries in the evolution of the mechanistic kinetic scheme for the addition of a single nucleotide by Klenow fragment is given in Table 8 (page 90).

The steps involved in DNA polymerisation by *E. coli* DNA polymerase I were first investigated by Bryant et al. (1983). A $(dT)_{10}.(dA)_{1000}$ primer-template was used, and the authors established that there is an obligatory sequential order of binding of the substrates to the DNA polymerase. First, the polymerase binds the DNA to form a polymerase-DNA complex, and the polymerase-DNA complex then binds the dNTP. The free enzyme does not bind dNTP:

$$E + D \rightleftharpoons E.D \overset{dNTP}{\rightleftharpoons} E.D.dNTP$$

This obligatory sequential order of substrate binding was verified by isotope trapping experiments: the polymerase was pre-incubated with the primer-template, and [^{3}H]dTTP was added to initiate polymerisation, or challenger DNA (calf thymus DNA) was added either together with or before [^{3}H]dTTP. These studies confirmed that the polymerase must first bind the DNA before it binds dNTP. The processivity of pol I with the primer-template was measured at 50 ± 5 nucleotides added each time the enzyme binds to a DNA primer-template, by incorporating [^{3}H]dTTP into the preformed polymerase-DNA complex in the presence of cold competing DNA of random sequence (calf thymus DNA). Any polymerisation that took place in the polymerase.$(dT)_{10}.(dA)_{1000}$ complex would only last one processive cycle. Once the polymerase had dissociated from the primer-template, it would be trapped by the competing DNA and thus remained catalytically silent. A dissociation constant of 200 nM was obtained for the polymerase-DNA complex from a Scatchard plot. The rate of dissociation of the polymerase from the polymerase-DNA complex was measured at 0.25 s^{-1} by preincubating the enzyme with the DNA, adding the competing DNA at zero time, and then adding [^{3}H]dTTP and Mg^{2+} at various time intervals to quantitate the remaining amount of polymerase-DNA complexes. From the dissociation constant (200 nM) and the rate of dissociation (0.25 s^{-1}) of the polymerase-DNA complex, the rate of binding enzyme to DNA was calculated at 1.2×10^6 $M^{-1}s^{-1}$.

Having established the order of substrate binding and measured the rate and dissociation constants for the polymerase-DNA complex, Bryant et al. (1983) then used rapid quench techniques to study pre-steady state polymerisation. By using an enzyme:DNA ratio of 34:270, they observed a biphasic time course of [^{3}H]dTTP incorporation into DNA, and attempted to fit the data by computer-simulation to a simple model:

$$E.D + dNTP \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E.D.dNTP \xrightarrow{k_2} E.D_1.PP_i \xrightarrow{k_3} E + D_1 + PPi$$

The equilibration for E.D + dNTP $\rightleftharpoons$ E.D.dNTP was assumed to be rapid enough for it to be neglected. k_2 and k_3 were assigned values of 4.6 s^{-1} and 4 s^{-1} after computer simulation showed that these values produced the best fit to the experimental data. Since k_2 is only fractionally bigger than k_3, the authors proposed that there are ***two partially rate-limiting steps*** during polymerisation of DNA, ***one before and the other after phosphodiester bond formation*** (Bryant et al., 1983). The latter conclusion was deduced from experiments which showed that the formation of phosphodiester bond is not rate-limiting during polymerisation of a correct nucleotide (Bryant et al., 1983, discussed in the next paragraph). Since no chemistry could be assigned to the rate-limiting step occurring before phosphodiester bond formation, Bryant et al. suggested that this could be a conformational change involving the enzyme.DNA.dNTP ternary complex (1983).

The question of whether the rate of phosphodiester bond formation could be partially rate limiting during polymerisation of a correct nucleotide was investigated (Bryant et al., 1983). Pre-steady state polymerisation was carried out with both regular dTTP and with (S_p)-dTTPαS, and the lack of a significant difference in rate of polymerisation between these two nucleotides indicated that the formation of phosphodiester bond is not rate limiting. The rationale for these experiments with the thionucleotides is based on earlier work by Benkovic and Schray (1971). Because sulphur is less electronegative than oxygen, the phosphorus in a thiophosphate is less electrophilic and therefore less susceptible to nucleophilic attack than the corresponding phosphate containing only oxygen and phosphorus atoms. The work by Benkovic & Schray (1971) has shown that thiophosphotriesters are chemically less reactive than their phosphotriester counterparts in bimolecular nucleophilic displacement reactions by 30-100-fold. If this thio-substitution effect observed for phosphotriesters applies to phosphodiesters as well, and if the formation of phosphodiester bond is rate-limiting, then using dTTPαS in place of regular dTTP for DNA polymerisation would show a significant reduction in the rate of polymerisation (discussed by Bryant et al., 1983).

The technique of comparing the incorporation of regular dNTP with that of dTTPαS played a crucial role in the unravelling of the complex kinetics of DNA polymerase. In a subsequent paper in 1985, dTTPαS was again used for studies on the kinetics of pol I and of Klenow fragment (Mizrahi et al., 1985). Pre-steady state incorporation of [^{3}H]dTTP into a pre-existing enzyme-DNA complex (Klenow fragment-poly(dA).oligo(dT)) was compared with that of [^{35}S]dTTPαS. With dTTP, there was no pre-steady state burst, suggesting that there is only one rate limiting step before phosphodiester bond formation (Figure 3.2). With dTTPαS however, the initial rate of dTTPαS incorporation was the same as that for dTTP incorporation under the same experimental conditions, which proved that formation of the phosphodiester bond is not rate-limiting. However, the rate of incorporation of dTTPαS was observed to decrease

gradually with time. The authors suggested that this was because incorporation of dTTPαS into DNA might destabilize the DNA and thus affect the processivity of polymerisation. Measurement of the processivities revealed that for pol I, the processivity was reduced from 50 nucleotide additions per association of enzyme with DNA for regular dTTP to < 4 nucleotide additions per association of enzyme with DNA for dTTPαS.

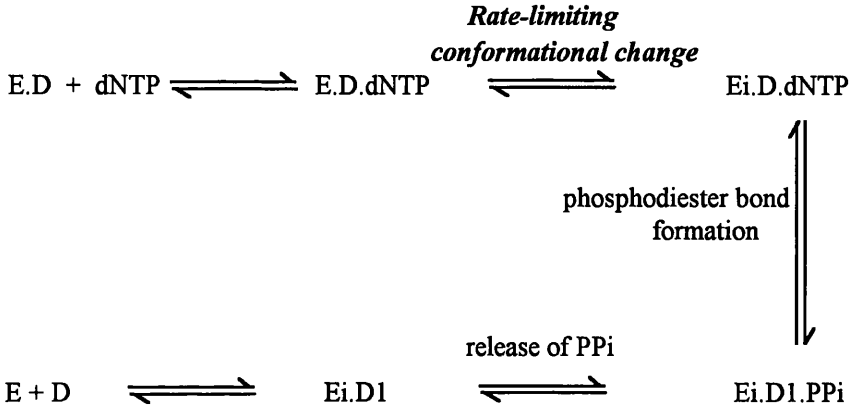
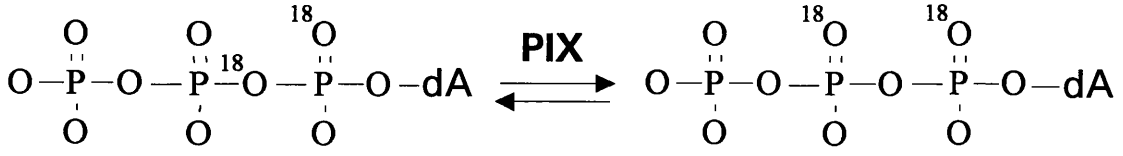


Figure 3.2. Reaction scheme for polymerisation of a correct nucleotide showing a rate-limiting conformational change occurring before the formation of the phosphodiester bond. Ei denotes the enzyme after it has undergone the conformational change. The release of pyrophosphate comes after the formation of the phosphodiester bond, and the rate of this step was shown to be very fast by positional isotope exchange experiments discussed in the text (reaction scheme from Mizrahi et al., 1985).

Having confirmed that there is a rate-limiting step which cannot be ascribed to any chemistry between dNTP binding and phosphodiester bond formation, Mizrahi et al. (1985) included this step in the guise of a conformational change by the enzyme into the simple model proposed by Bryant et al. in 1983. After phosphodiester bond formation, the release of pyrophosphate and DNA was also proposed to occur in two steps, with the release of pyrophosphate preceding that of the primer-template (Figure 3.2). In order to see if the release of pyrophosphate is rate-limiting, they designed a PIX (positional isotope exchange) experiment to test this. Two ^{18}O isotope labels were introduced into the triphosphate moiety of dATP, in one of the oxygen atoms of the α-phosphate and in the oxygen linking the α- and β-phosphates. Release of pyrophosphate comes after phosphodiester bond formation; if the release is slow, then there is a likelihood that the pyrophosphate may attack the phosphate of the newly incorporated nucleotide, resulting in an exchange of the ^{18}O labels from an α,β-bridging configuration into the β-non-bridging configuration. Very little exchange was observed and it was concluded that the rate of pyrophosphate release must be very rapid.



The experiments performed by Bryant et al. (1983) and Mizrahi et al. (1985) were carried out with either a poly(dA).oligo(dT), or a poly(dAT), or a poly(dT).oligo(dA) primer-template. These primer-templates are unsuitable for experiments where only one nucleotide is added, which are necessary for measuring the individual rate constants of the polymerisation reaction pathway. Therefore, a set of primer-templates had to be designed which allowed only one nucleotide to be incorporated; Kuchta et al. (1987) synthesized a set of 3 DNA duplexes (referred to as Duplexes 1-3) for this purpose and carried out both steady state and pre-steady state experiments to analyse the kinetic mechanism of Klenow fragment:

Duplex 1 : 9/20-mer	TCGCAGCCG AGCGTCGGCAGGTTCCCAAA(5')
Duplex 2 : 13/20-mer	TCGCAGCCGTCCA AGCGTCGGCAGGTTCCCAAA(5')
Duplex 3 : 16/20-mer	TGCGTCCGGCGTAGAG CGCAGGCCGCATCTCCTAG(5')

Pre-steady state incorporation of dATP into Duplex 1 with an enzyme:DNA ratio of 60:510 showed a biphasic progress curve (Figure 3.3) where the rate of the second phase (slow) which represented the dissociation of Klenow from the product DNA was measured at 0.06 s^{-1}. The experiment was repeated with an enzyme:DNA ratio of 100:300 with various concentrations of dATP. Computer simulation suggested values of 50 s^{-1} for k_{cat} and 5.5 μM for the dissociation constant of dNTP-binding:

$$E + D \underset{0.06\ s^{-1}}{\overset{1.2 \times 10^7\ M^{-1}s^{-1}}{\rightleftharpoons}} E.D \xrightarrow{50\ s^{-1}} E.D1 \underset{1.2 \times 10^7\ M^{-1}s^{-1}}{\overset{0.06\ s^{-1}}{\rightleftharpoons}} E + D1$$

To test if the formation of phosphodiester bond is rate-limiting, pre-steady state polymerisation into Duplex 1 was carried out with dATPαS in place dATP, and it was found that the rate of incorporation was 3-7 fold slower (Kuchta et al., 1987). For the reasons discussed earlier, this small difference in rate of polymerisation between dATP and dATPαS was used as evidence that formation of phosphodiester bond is not rate limiting. Thus, the k_{cat} of 50

s^{-1} observed in the equation above must correspond to that of the rate-limiting conformational change that occurs just before phosphodiester bond formation (Figure 3.2, page 83).

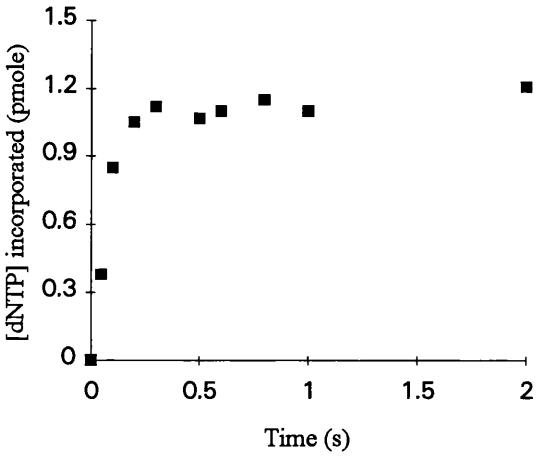


Figure 3.3. Graph showing pre-steady state incorporation of a correct dNTP into Duplex 1. The experiment was performed with 60 nM Klenow fragment and 510 nM DNA. The initial burst represents polymerisation by the pre-formed enzyme.DNA complex. The slower second phase represents the dissociation of Klenow from the product DNA , measured at 0.06 s^{-1} (Kuchta et al., 1987).

The dissociation constant governing the release of pyrophosphate was also measured (Kuchta et al., 1987). An [8-^{3}H]-dATP was first introduced into Duplex 1 by 3'-end labelling with Klenow fragment. The resulting duplex containing a ^{3}H-dAMP at its 3'-terminus was then purified and used in pre-steady state pyrophosphorolysis experiments. Pyrophosphorolysis is the reverse of polymerisation, involving the formation of dNTP from the 3' terminal dNMP of the primer strand and pyrophosphate. In Kuchta et al.'s experiments, pyrophosphorolysis was initiated by mixing a solution of the DNA containing a [^{3}H]-dA at the 3' end of the primer strand with a solution of Klenow fragment, pyrophosphate and magnesium ions. The fraction of [^{3}H] remaining in DNA was quantitated at different time points with various concentrations of pyrophosphate. The k_{cat} for pyrophosphorolysis was measured at 0.12 s^{-1} while the dissociation constant was estimated at 100 μM. The authors acknowledged that the k_{cat} of pyrophosphorolysis measured (0.12 s^{-1}) was lower than that predicted by computer simulation (0.2 s^{-1}) and attributed the discrepancy to "experimental uncertainty in the measurements".

Steady state polymerisation experiments were used to determine the dissociation constant for the Klenow-DNA complex. Polymerisation was initiated by adding a solution containing 50 nM Klenow and 25-200 nM Duplex 1 to a solution of α-^{32}P-dATP and Mg^{2+}, and the amount of ^{32}P incorporated was quantified at different times (15-90 s). The data were put into a Scatchard plot to determine the dissociation of the Klenow-DNA complex. In this way, the dissociation constant for enzyme from the primer-template DNA was found to be 5 nM, suggesting that in the association

$$E + D \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} E.D$$

k_{on} =1.2 $M^{-1}s^{-1}$, since k_{off} had previously been measured at 0.06 s^{-1} (see page 16).

The work by Kuchta et al. (1987) examined the kinetics of polymerisation of a correct nucleotide. Using similar methods, the kinetics of polymerisation of an incorrect nucleotide were resolved by the same laboratory (Kuchta et al., 1988) and the authors were then in a position to compare the kinetics of incorporation of a correct versus incorrect nucleotide. Based on these analyses, misincorporation by Klenow fragment was proposed to be the result of discrimination at three stages. During incorporation of a mismatched nucleotide, the K_m for the incorrect nucleotide was found to be only very slightly increased (8-21 μM compared with a K_m of 5 μM for incorporation of a correct nucleotide). The authors were unable to measure or calculate the rate of the conformational change which takes place just before phosphodiester bond formation, which is the rate-limiting step for incorporation of a ***correct*** nucleotide by Klenow fragment (Kuchta et al., 1987). However, the rate of phosphodiester bond formation for the incorporation of a ***mismatched nucleotide*** was found to be dramatically reduced, based on the large thio-effect (65-fold) observed when misincorporation was carried out with dNTP and with dNTPαS under similar conditions. It was therefore concluded that ***the first stage of discrimination against misincorporation comes primarily from a reduced rate of phosphodiester bond formation when an incorrect nucleotide is bound***; selection of the correct dNTP at the nucleotide-binding step appears to be of much less importance.

The second stage of discrimination against misincorporation was proposed to come from a second conformational change which occurs immediately after phosphodiester bond formation (Kuchta et al., 1988). The authors observed that in certain misincorporation events, there was evidence for dNTP turnover and yet they could not detect the accumulation of the expected products of misincorporation. The rate of the 3'→5' exonuclease activity was measured and found to be much too slow (1-10 x 10^{-3} s^{-1}) than the rate of dissociation of the mismatch-containing primer from Klenow (3 s^{-1}) to explain the phenomenon observed. In order to fit the data, extra time must be allowed for the 3'→5' exonuclease to remove the mismatched nucleotide. Therefore a second conformational change was proposed which takes place immediately after phosphodiester bond formation, and which is (partially) rate-limiting during misincorporation so that this allows the relatively slow exonuclease to remove the mismatch. The rate of this second conformational change was estimated at 0.0035 s^{-1} during misincorporation by computer simulation. During polymerisation of a correct nucleotide, this step is kinetically invisible because it is very fast, therefore the rate of the second conformational change must be greater than 50 s^{-1} (which is the k_{cat} for incorporation of nucleotide and therefore the rate of the slowest step during correct polymerisation; Kuchta et al., 1987) for incorporation of a correct nucleotide.

The third and final stage of discrimination against misincorporation by Klenow comes from the slower incorporation of the nucleotide following the mismatch (Kuchta et al., 1988). Pre-steady state studies of addition of the next correct nucleotide after a mismatch showed this to be considerably slower (k_{cat} = 10^{-5} to 1.4 x 10^{-2} s^{-1} depending on the mismatch) than addition onto a correctly base-paired primer terminus (k_{cat} = 50 s^{-1}, Kuchta et al., 1987). Again, this slower addition of the nucleotide following the mismatched nucleotide allows the 3' to 5' exonuclease to edit the mismatch.

The kinetic mechanisms proposed by Kuchta et al. (1987, 1988) were somewhat incomplete because some of the internal rate constants (i.e. for the steps E.D.N $\rightleftharpoons$ Ei.D.N $\rightleftharpoons$ Ei.D1.PP_i) were not determined. The simple kinetic scheme proposed by Bryant et al. (1983) contained two partially rate-limiting steps (page 81). The work by Mizrahi et al. (1985) and Kuchta et al. (1987, 1988) focused only on the first rate-limiting step. The existence of the second rate-limiting step was subsequently confirmed by pulse-chase experiments carried out by Dahlberg & Benkovic (1991) (Figure 3.4).

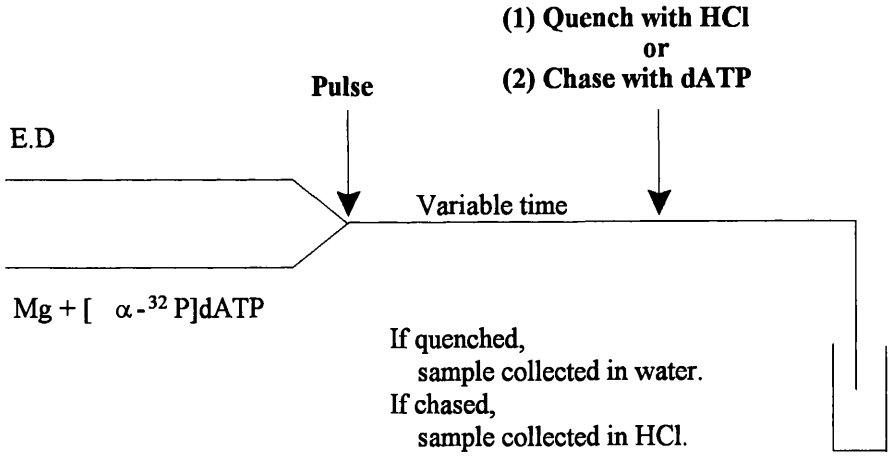


Figure 3.4. Pictorial representation of the pulse-chase experiments performed by Dahlberg & Benkovic (1991).

In a pulse-quench experiment (Figure 3.4), the amount of primer that been elongated within the reaction time was deduced from the of radioactivity incorporated into the primer-template. In a pulse-chase experiment, the reaction was chased by the addition of cold dNTP and then quenched 50 ms later. This permitted any enzyme-DNA-dNTP intermediate that had not been allowed time to form the phosphodiester bond in the pulse-quench experiment to do so in the pulse-chase experiment. The difference in amount of radioactivity incorporated between the pulse-chase and pulse-quench experiments represents the amount of that intermediate which had already bound the dNTP but had not had time to form the phosphodiester bond. The authors observed a 20 % difference in the amount of radioactivity incorporated in pulse-chase and pulse-

quenched experiments, which implied two possibilities: accumulation of the E.D.dNTP complex or the Ei.D.dNTP complex during polymerisation (Figure 3.2). Since the conformational change induced by the binding of dNTP is rate-limiting (50 s^{-1}, Kuchta et al., 1987; see Figure 3.2), very little of the E.D.dNTP complex would have been chased to completion during the chase time of 50 ms. Thus, most of the intermediate that could be chased was the Ei.D.dNTP complex, and in order for this to accumulate, there must be another slow step further down the polymerisation pathway. This was taken as evidence for the second conformational change. The forward and reverse rate constants for phosphodiester bond formation was not determined, but the equilibrium constant for this step was calculated as 4.

Thus, the rate of polymerisation of a correct nucleotide appears to be limited by both the first and second conformational changes. The k_{cat} for the first conformational change has been determined at 50 s^{-1} (Kuchta et al., 1987). If the net rate of polymerisation is known, then the rate of the second conformational change can be determined. By measuring the net flux of multiple consecutive polymerisation events, Dahlberg & Benkovic were thus able to calculate the forward rate constant for the second conformational change during polymerisation of a correct nucleotide at 15 s^{-1} (1991).

The incomplete kinetic scheme proposed by Kuchta et al. (1988) for misincorporation by Klenow was further refined by Eger & Benkovic in 1992 by the use of competition assays where a correct nucleotide competes directly with an incorrect nucleotide for incorporation into the same site in the primer-template. These experiments were carried out using excess Klenow over DNA at a fixed concentration of the correct nucleotide (radiolabelled on the α-phosphate) while the concentration of the incorrect nucleotide (cold) was varied. The high enzyme:DNA ratio ensured that all the DNA was in an enzyme-DNA complex. Short reaction times (the longest reaction time was 1 s) were chosen such that only negligible amount of the incorrect nucleotide would have been incorporated and it could be assumed that the reaction with incorrect nucleotides did not proceed as far as phosphodiester bond formation. This then eliminated the problem of separating the products of elongation by correct and incorrect dNTPs. The amount of correct nucleotide incorporated was measured and the data compared with computer simulations of product formation according to different reaction schemes. The reaction scheme that best fitted the data was finally chosen as the correct reaction scheme. Two reaction schemes were tested: one which included a conformational change after the incorrect nucleotide was bound, and one which did not have this conformational change. Computer simulations based on the first reaction scheme gave a better fitting to the experimental results, and calculated the forward rate constant for the conformational change induced by binding of the incorrect nucleotide at 50 s^{-1}. This is similar to the rate of conformational change when a correct nucleotide is bound as measured by Kuchta et al. (1987). Since the rate of incorporation of a mismatched nucleotide is

much slower than that for a correct dNTP, and the competition assays showed that the rate of the conformational change induced by the binding of a correct nucleotide is the same as that induced by the binding of an incorrect nucleotide, the obvious conclusion must be that the rate-limiting step during the incorporation of a mismatched nucleotide is the formation of the phosphodiester bond (Eger & Benkovic, 1992).

Eger & Benkovic also reported a second competition assay using exactly the same conditions as above except that the incorrect nucleotide was replaced by a phosphorothioate-analogue of the incorrect nucleotide. This experiment was carried out to see if the thio-substitution had any effects on the binding of the thio-nucleotide and on the subsequent conformational change induced by the binding of the thio-nucleotide. The authors were able to use the same kinetic constants obtained for competition with the incorrect nucleotide to fit the data for competition with the thio-analogue of the incorrect nucleotide, and therefore concluded that (1) thio-substitution has no effect on the binding of the nucleotide and (2) that the rate of the conformational change induced by the thio-nucleotide is the same as that induced by a correct nucleotide (50 s^{-1}). Therefore, the only step which can contribute to the slower incorporation of the thio-nucleotide must be the formation of the phosphodiester bond (Eger & Benkovic, 1992).

This result that the formation of a phosphodiester bond is rate-limiting during the incorporation of an incorrect nucleotide and of a thio-nucleotide was very important because in the previous paper by Kuchta et al. (1988), the evaluation of the reaction scheme for misincorporation depended upon the assumption that phosphodiester bond formation is rate-limiting during misincorporation. This assumption was based on earlier work by Benkovic and Schray (1971), and has been severely criticized by Herschlag et al. (1991) who asserted that the work by Benkovic and Schray was carried out on phosphate triesters, and could not be applied to the formation of phosphodiester bonds (discussed in greater detail later). This has very serious implications: if the effect of thio-substitution could not be used to determine the rate-limiting step during polymerisation, then the kinetic schemes that were previously proposed based on this thio-effect argument have not been rigorously tested. By obtaining evidence that the formation of phosphodiester bond is rate-limiting in the incorporation of a mismatched nucleotide and in the incorporation of a thio-analogue of the mismatched nucleotide, the authors have defended the reaction scheme for misincorporation proposed by Kuchta et al. (1988).

Although the competition experiments showed the formation of phosphodiester bond to be rate-limiting during misincorporation, the experiments were not designed to measure the velocity of this step. Instead, the rate of phosphodiester bond formation during misincorporation was determined from two pre-steady state experiments, from misincorporation experiments with excess Klenow over DNA at different concentrations of the incorrect nucleotide, and from

experiments designed to measure the rate of pyrophosphate release. Data from these experiments were then used in computer simulation to determine the rate of phosphodiester bond formation, which was found to be 2.5 x 10^{-2} s^{-1} for the forward reaction and 3 x 10^{-3} s^{-1} for the reverse reaction (Eger & Benkovic, 1992). Data from steady state misincorporation experiments carried out using excess Klenow over DNA and with different concentrations of the incorrect nucleotide were used to determine the forward rate of the second conformational change. This was determined at 6 x 10^{-3} s^{-1} (Eger & Benkovic, 1992). Finally, pyrophosphorolysis experiments were carried out in which a mutant Klenow fragment which was lacked the 3' to 5' exonuclease activity was incubated with pyrophasphate and a 5'-labelled primer-template containing a mismatched base at the 3'-end. Data from these pyrophosphorolysis experiments were then used to calculate the reverse rate of the second conformational change and both forward and reverse rate constants for pyrophosphate release by computer simulation (Eger & Benkovic, 1992).

A summary of the evolution of the mechanistic scheme for the incorporation of a nucleotide by Klenow is given in Table 8.

Table 8. The evolution of the mechanistic kinetic scheme for the incorporation of a single nucleotide by Klenow fragment of *E. coli* DNA polymerase I.

1983 Bryant et al. established the obligatory sequential order of substrate-binding in which DNA polymerase I binds the substrates: the polymerase has to bind the DNA primer-template first and the enzyme-DNA complex then binds the dNTP. The enzyme alone does not bind dNTP. A simple model for the polymerisation reaction was also proposed:

$$E.D + dNTP \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E.D.dNTP \xrightarrow{k_2} E.D_1.PP_i \xrightarrow{k_3} E + D_1 + PP_i$$

The reaction was thought to be rate-limited by two steps, since the values of 4.6 s^{-1} and 4 s^{-1} assigned to k_2 and k_3 were very close.

The pre-steady state rate of polymerisation of a dNTPαS was found to be the same as that for the incorporation of a correct dNTP. Based on earlier work by Benkovic & Schray which showed a 30-100-fold reduction in the rates of reaction of phosphate triester upon thio-substitution, the absence of an effect of the thio-substitution on the rate of polymerisation observed by Bryant et al. was taken as evidence that the formation of the phosphodiester bond is not rate-limiting in the incorporation of a correct dNTP. Thus, the first partially rate-limiting step in the simple model proposed by Bryant et al. was tentatively assigned as a conformational change since no chemistry could be ascribed to this step.

Table 8. Continued.

1985	By comparing the rates of incorporation of regular dNTPs and their phosphorothioate analogues, Mizrahi et al. confirmed that the formation of the phosphodiester bond is not rate-limiting. In the same report, Mizrahi et al. also carried out positional isotope exchange experiments to study the rate of pyrophosphate release, and concluded that this must be very rapid during the incorporation of a correct nucleotide.
1987	Again, by comparing the rates of incorporation of the regular dNTP with that of dNTPαS, Kuchta et al. found that the formation of phosphodiester bond is not rate-limiting during incorporation, and concluded that the conformational change before the formation of phosphodiester bond must be rate-limiting during the incorporation of a correct nucleotide.
1988	Kuchta et al. compared the kinetics of incorporation of a correct and a mismatched dNTP, and proposed that Klenow fragment discriminates against the incorporation of an incorrect nucleotide in three stages: • by greatly reducing the rate of phosphodiester bond formation for an incorrect nucleotide, • by slowing down the rate of the second conformational change that occurs after the formation of the phosphodiester, which allows time for the relatively slow 3' to 5' exonuclease to remove the mismatched nucleotide, and • by reducing the rate of addition of the nucleotide following the mismatch, which also allows the exonuclease more time to edit the mismatch.
1991	Using pulse-chase experiments, Dahlberg & Benkovic demonstrated the existence of the second conformational which occurs after phosphodiester bond formation but before pyrophosphate release during the incorporation of a correct nucleotide .
1991	Herschlag et al. criticised the use of the effect of thio-substitution on the rate of polymerisation as a diagnostic tool for determining the rate-limiting step during polymerisation, because this was based on work by Benkovic & Schray (1971) on phosphate ***triesters***. Herschlag et al. showed that thio-substitution only slows the rate of reactions with phosphate ***diesters*** by 4-11-fold. This has serious implications on the interpretation of the phosphorothioate data in the earlier work by Bryant et al. (1983), Mizrahi et al. (1985), and Kuchta et al. (1987 & 1988).

Table 8. Continued.

1992	Eger & Benkovic carried out competition experiments using a correct dNTP and an incorrect nucleotide, or the thio-analogue of an incorrect nucleotide, and carried out computer simulations to the results. They confirmed that the rate of phosphodiester bond formation for a mismatched nucleotide and for the thio-analogue of a bond formation for a mismatched nucleotide and for the thio-analogue of a mismatched nucleotide, is the rate-limiting step during misincorporation. This result was important because it defended the model for misincorporation proposed by Kuchta et al. (1988) against the crticism by Herschlag et al. (1991).

The results discussed above were carried out using Klenow fragment. Similar kinetic schemes for incorporation of correct and incorrect nucleotides by T7 DNA polymerase were solved by work from the laboratory of Johnson, using both steady state and pre-steady state methods similar to those used by Benkovic and his workers. This work was described in a series of papers published in 1991 (Patel et al., Wong et al., and Donlin et al.). Like Klenow fragment, the kinetic scheme for incorporation of nucleotide by T7 DNA polymerase has seven discernible kinetic steps (Fig. 3.5). In addition, the rate-limiting step during incorporation of a correct nucleotide was identified as the first conformational change involving the E.D.N ternary complex which precedes the formation of the phosphodiester bond (Patel et al., 1991). Unlike Klenow fragment, the rate of phosphodiester bond formation was found to be only partially rate-limiting during misincorporation; this conclusion was based on a moderate (17-19-fold) reduction in the rate of polymerisation when phosphorothioate analogues were used in the polymerisation assays (Wong et al., 1991). ***For T7 DNA polymerase, the conformational change induced by dNTP-binding remains the rate-limiting step during misincorporation***, although the magnitude of the rate of this step is much lower for a mismatch than for a correct nucleotide (300 s^{-1} for correct incorporation and 0.14 s^{-1} for misincorporation; Wong et al., 1991). The rate of the slow conformational change during misincorporation was determined from competition assays where incorrect nucleotides competed with the correct nucleotide for incorporation into the same site in the primer-template. Because of the importance of this conformational change (Fig. 3.5, step 3), the kinetic scheme for T7 DNA polymerase is described as an ***induced-fit mechanism*** (Wong et al., 1991). Addition of the nucleotide following a mismatch was found to be slower, just like Klenow, and coupled with a faster rate of transfer of a mismatch from the polymerase site to the exonuclease site, this provides more time for excision of the mismatch by the 3' to 5' exonuclease (Donlin et al., 1991).

Thus, the kinetic mechanisms whereby Klenow fragment and T7 DNA polymerase maintain a high fidelity during DNA synthesis were finally solved. In these kinetic schemes, the conformational changes in the polymerisation pathway play an important role in the way Klenow fragment and T7 DNA polymerase achieve high fidelity during DNA synthesis. Their identification has proven beyond doubt that the K_m- and V_{max}-discrimination models that are based on a two-step mechanism and proposed to explain how DNA polymerases achieve a high fidelity of DNA replication are over-simplified. The conformational change that occurs immediately before phosphodiester bond formation is the rate-limiting step during the incorporation of a correct nucleotide in the kinetic schemes of both Klenow and T7 DNA polymerase (Kuchta et al., 1987; Patel et al., 1991). During misincorporation, the kinetic schemes for the two enzymes are slightly different. With Klenow fragment, the slowest step during misincorporation is the formation of the phosphodiester bond. Although the rate of the second conformational change after phosphodiester bond formation is also reduced, this is only partially rate-limiting (Kuchta et al., 1988). During misincorporation by T7 DNA polymerase, the conformational change induced by dNTP-binding remains the rate-limiting step during misincorporation while the rate of phosphodiester bond is only partially rate-limiting (Wong et al., 1991).

E + D	$\underset{k_2}{\overset{k_1}{\rightleftharpoons}}$	**E.D**	(1) Enzyme binds DNA to form E.D complex
E.D + dNTP	$\underset{k_4}{\overset{k_3}{\rightleftharpoons}}$	**E.D.dNTP**	(2) E.D complex binds dNTP to form E.D.dNTP
E.D.dNTP	$\underset{k_6}{\overset{k_5}{\rightleftharpoons}}$	**Ei.D.dNTP**	(3) Enzyme in E.D.dNTP complex undergoes a conformational change (Ei)
Ei.D.dNTP	$\underset{k_8}{\overset{k_7}{\rightleftharpoons}}$	**Ei.D$_1$.PP$_i$**	(4) The phosphodiester bond is formed.
Ei.D$_1$.PP$_i$	$\underset{k_{10}}{\overset{k_9}{\rightleftharpoons}}$	**E.D$_1$.PP$_i$**	(5) Enzyme undergoes a second conformational change, returning to a 'relaxed' state
E.D$_1$.PP$_i$	$\underset{k_{12}}{\overset{k_{11}}{\rightleftharpoons}}$	**E.D$_1$ + PP$_i$**	(6) Pyrophosphate is released
E.D$_1$	$\underset{k_{14}}{\overset{k_{13}}{\rightleftharpoons}}$	**E + D$_1$**	(7) Enzyme dissociates from DNA

Figure 3.5. Kinetic scheme for the incorporation of a single nucleotide by Klenow fragment and T7 DNA polymerase. A summary for the evolution of this kinetic scheme is given in Table 8.

As mentioned earlier, the use of phosphorothioate nucleotide analogues have played a key part in solving the mechanistic kinetic scheme of polymerisation of correct and incorrect nucleotides by Klenow and by T7 DNA polymerase. If the observed thio-effect (i.e. ratio of the rate of polymerisation of regular dNTPs with that of the dNTPαS) was less than 30-100-fold, then the formation of phosphodiester bond was not considered to be rate-limiting. The rationale behind the method of studying the thio-effect is based on work on phosphate ***triester*** by Benkovic & Schray (1971). But recently, this assumption has been severely criticized by Herschlag et al. (1991), who argued that the effect of thio-substitution on phosphate triesters are different from that on phosphate ***diesters***. The same paper has shown that thio-substitution only slows the rate of reactions with phosphate diesters by 4-11-fold, and this result has serious implications on the interpretation of the phosphorothioate data:

(1) the small thio-effect (3-fold) observed by Kuchta et al. (1987) and Patel et al. (1991) for the incorporation of a correct nucleotide by Klenow fragment and T7 DNA polymerase respectively could have been due to a slow formation of phosphodiester bond;

(2) the large thio-effect (65-fold) observed by Kuchta et al. during the misincorporation by Klenow fragment, was probably due to a combination of steps rather than just the slower rate of phosphodiester formation alone, since thio-substitution only reduces the rate of reaction with phosphate diesters by 4-11-fold, which is too small to explain the 65-fold thio-effect observed;

(3) the moderate thio-effect (17-19-fold) observed during misincorporation T7 DNA polymerase (Wong et al., 1991) could have been due to several possibilities: slow binding of the dNTPαS, or slow conformational change of the ternary complex induced by the dNTPαS, or a slow rate of phosphodiester bond formation, or a combination of any of the three.

Wong et al. (1991) acknowledged the possibility that the phosphorothioate data might have been misinterpreted, and admitted in their discussion of the paper that the use of the phosphorothioate data as a probe for the rate-limiting step "cannot be regarded as a rigorous, quantitative analysis" and that they had only intended it "as an approximate means to probe the free energy surface of polymerisation".

If the magnitude of the thio-effect cannot be used to probe the rate-limiting step during polymerisation, this presents a great problem. It is fairly certain that the kinetic scheme for incorporation of a nucleotide by these two DNA polymerases involve at least two conformational changes, one before and the other after phosphodiester bond formation, but the difficulty in solving the reaction scheme lies in the question of which of the internal steps (i.e. the steps $E.D.N \rightleftharpoons Ei.D.N \rightleftharpoons Ei.D1.PP_i$) is the rate-limiting step? The rate of an internal step can only be measured if one is able to quantitate the amount of the intermediate(s) formed during the reaction. Since this cannot be done for the internal steps of the polymerisation pathway, it is not

possible to find a true solution through measurement simply because one can assign a wide range of values to these rate constants and they would all fit the data. For example, a fast forward rate can be offset by a fast reverse rate or a slower forward rate of the following step and the overall rate would still be the same. The problem can be solved if one knows which of the internal steps is rate-limiting and then use this as a constraint for the system. In the elucidation of the kinetic schemes for Klenow and for T7 DNA polymerase, the magnitude of the thio-effect was used as a diagnostic tool to probe the rate-limiting step, and the values of the internal rate constants were subsequently derived from computer simulations using reaction schemes which included the formation of phosphodiester bond as a rate-limiting step or the conformational change as a rate-limiting step, according to the observed thio-effect (Kuchta et al., 1987, 1988; Patel et al., 1991; Wong et al., 1991).

If the magnitude of the thio-effect cannot be used to provide information about the rate-limiting step, another constrain on the internal steps of the polymerisation reaction must be applied before the internal rate constants can be determined. Eger & Benkovic (1992) overcame this problem by using data from competition assays to determine the rate-limiting step. As mentioned earlier, they were able to show that the rate of binding an incorrect nucleotide, and the subsequent rate of the conformational change induced by this binding, were the same as the corresponding rates for the incorporation of a correct nucleotide, and the obvious conclusion from this must be that the rate-limiting step during the incorporation of an incorrect nucleotide must be the formation of the phosphodiester bond itself.

In this Ph.D. project, the initial analyses of the data for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand was carried out using the phosphorothioate data as a constraint for the internal steps of the polymerisation reaction. The report by Herschlag et al. (1991) was, unfortunately, noticed after this initial analysis had been completed. Thus, an alternative constraint for the system had to be applied. It was then that the data from the competition experiments were analysed. This analysis proved to be extremely informative in determining the internal rate-constants and allowed the mechanistic kinetic schemes for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand to be solved without having to use the phosphorothioate data.

Wong et al. (1991) and Eger & Benkovic (1992) also used competition assays in their investigations. Wong et al. (1991) used competition assays to measure the K_i (equilibrium constant for the inhibitor, in this case, the incorrect nucleotide) for inhibition of incorporation of a correct nucleotide by a mismatched nucleotide. The inhibition observed was only very slight even in the presence of a large excess of the incorrect nucleotide; for example, in one of their competition assays, the ratio of correct:incorrect dNTP was 10 μM:2 mM. This slight inhibition

by the incorrect nucleotide on the polymerisation of the correct nucleotide could be due to several possibilities: (1) the incorrect nucleotide does not bind very well to the enzyme-DNA complex, or (2) the rate of the first conformational change (step 3, Figure 3.5) is very slow, or (3) the rate of phosphodiester bond formation when an incorrect dNTP is bound is very slow (step 4, Figure 3.5). Unfortunately, apart from calculating the K_i, the results of the competition assay were not rigorously analysed and no information on the rate-limiting step during misincorporation was obtained.

Eger & Benkovic also carried out competition assays using both correct and incorrect dNTP to determine the rate constants for the binding of the incorrect dNTP and for the conformational change before the formation of phosphodiester bond (steps 2 and 3, Figure 3.5). Unlike Wong et al., who did not analyse the data from the competition assays, Eger & Benkovic performed ***computer simulations*** of product formation according to different models, and were thus able to show that the rate of the conformational change induced by the binding of a correct nucleotide is the same as that induced by the binding of an incorrect nucleotide. However, with a more powerful ***analytical tool which carries out parameter-fitting as well as simulation*** (explained in following paragraph and section 3.4), there is much more kinetic data that could be obtained from such competition experiments, and as will be shown in the analysis of the data from the competition assays in this project, the inclusion of the data from the competition assay in the calculations not only permitted the internal rate constants to be solved, it also allowed the detection of a previously invisible nucleotide exchange step in the polymerisation pathway. This is discussed in grater detail in section 3.4.

Many of the rate constants mentioned above were, in general, not measured but were obtained from computer analysis of the progress curves, which are the integration of the differential rate equations implied by the reaction scheme. The power of the computer program used depends on several factors:

- the amount of experimental data available,
- the number of rate equations the program can handle, and
- whether the program is only a simulation program or it can perform parameter-fitting as well as simulation. When a simulator program is used, the operator guesses some rate constants and uses these to generate progress curves which are then compared with the experimental data. On the other hand, a program which has both simulation and parameter-fitting facility contain iterative loops between each simulation, and these iterative loops compute the residual sum of squares (which is a measure of goodness-of-fit of the simulated curve to the observed data) at the end of each simulation, adjusts the rate constants and then carries out another round of simulation. This iterative process is known as parameter-fitting and it is repeated many times

until the residual sum of squares has reached a minimum, and that is when the best fit of the model to the data has been calculated.

The computer program used for the analysis of kinetic data by both Benkovic's and Johnson's groups (SIMUL or KINSIM, written by Barshop et al., 1983) was only a simulator program. Plainly, if the reaction scheme contains many steps, it would be humanly impossible to test the adequacy of a model and to fit the model to the observed data simply because there are too many combinations of rate constants to consider. There are two ways to solve a kinetic scheme which contains may steps. The first method is to fit one reaction at a time with a simulator program; when the rate constants for one reaction is solved, the reaction scheme can be expanded to include another step, and with more experimental data, the rate constants for the other steps in the expanded reaction scheme can then be calculated. The process is repeated until all the rate constants in the reaction scheme have been solved and the reaction scheme is able to fit all the data observed. This was the method used by the laboratories of Benkovic and of Johnson to solve the mechanistic kinetic schemes of Klenow fragment and T7 DNA polymerase respectively. The method is very tedious because it requires many experiments in order to solve the reaction scheme. In addition, with so many experiments, some with bigger experimental errors than others, the inconsistent experimental errors could make the fitting of the model quite difficult.

The alternative method to solve a kinetic scheme with many reactions is to use a program which has both simulating and parameter-fitting capabilities, as the program can then determine the best fit of the model to the data by systematically varying the rate constants. This was the method chosen to analyse the experimental data in this thesis, and one of the objectives of this thesis is to illustrate that the progress curves from pre-steady state incorporation experiments contain a wealth of kinetic information that could be used to calculate many rate constants by putting all the steps in the kinetic scheme into one computer program. This method of analysis eliminates the need for many experiments to solve the kinetic scheme which saves much time and labour. Section 3. 4 gives a detailed description of how the solution to the kinetic problems of insertion of thymine and cytosine opposite an O^6-meG in the template strand is solved by this analytical approach, and it is hoped that this will convince the reader that the analytical method used here is one of the quickest and most satisfactory available for solving kinetic problems in general.

3.2 Materials and methods

Enzymes and nucleotides. T4 polynucleotide kinase was obtained from Amersham International and snake venom phosphodiesterase (*Crotalus durissus*) and alkaline phosphatase from Boehringer Mannheim. FPLC-pure cloned Klenow fragment (1 unit = 2 pmole; 1 unit catalyzes the incorporation of 10 nmol of total deoxynucleotide into acid-insoluble product in 30 minutes at 37°C with poly(dA-dT).poly(dA-dT) as template), 2'-deoxynucleoside-5'-triphosphates (ultrapure grade) and the *S*-stereoisomers of 2'-deoxynucleoside 5'-O-(1-thiotriphosphate), [(S_p)-dNTPαS] were from Pharmacia. [γ-^{32}P]ATP was from New England Nuclear.

Synthesis of the DNA primer-templates. The sequences of the primer-templates used in this study are given in Table 9. Details of the syntheses are given in Chapter 2.

Table 9. Primer-template duplexes. Duplex A was used for studies of incorporation of cytosine or thymine opposite O^6-meG. Duplexes B, C and D were used to measure the rate of 3'→5' exonuclease activity, and the addition of the next correct nucleotide.

Duplex A	5'ATCCGATAG 3'TAGGCTATCmeGTCATTCTCGC
Duplex B	5'ATCCGATAGT 3'TAGGCTATCmeGTCATTCTCGC
Duplex C	5'ATCCGATAGC 3'TAGGCTATCmeGTCATTCTCGC
Duplex D	5'ATCCGATAGT 3'TAGGCTATCAGTATTCTCGC

Measurement of the concentration of each oligodeoxynucleotide. The concentration of each oligomer was calculated from the absorbance at 260 nm (A_{260}) of the oligomer, assuming that the molar extinction coefficient of the oligonucleotide is the sum of the molar extinction coefficients of each nucleotide (i.e. at neutral pH, ε^{dCMP} = 7050 $M^{-1}cm^{-1}$; ε^{dGMP} = 12010 $M^{-1}cm^{-1}$; ε^{dTMP} = 8400 $M^{-1}cm^{-1}$; ε^{dAMP} = 15200 $M^{-1}cm^{-1}$; $\varepsilon^{dO^6\text{-meGMP}}$ = 5300 $M^{-1}cm^{-1}$) multiplied by the number of times that nucleotide occurs in the oligonucleotide.

5' ^{32}P-end labelling of oligodeoxynucleotide primers. The primers were 5'-end-labelled to a specific activity of 50 Ci/mmol by incubating in 50 mM Tris.HCl (pH 7.4), 5 mM dithiothreitol, 1 mM spermidine, 10 mM $MgCl_2$, 10 units of T4 polynucleotide kinase, 1 nmol primer DNA, and 50 μCi of [γ-^{32}P]-ATP (3000 Ci/mmol) in a total volume of 10 μl at 37 °C. After 30 minutes, the reaction was terminated by heating to 70°C for 2 minutes.

Annealing oligodeoxynucleotide primer-templates. The ^{32}P-labelled primers were mixed with the complementary template DNA (to give a final DNA duplex concentration of 10 μM) in a solution containing 50 mM Tris.HCl (pH 7.4) and 5 mM $MgCl_2$ and annealed by heating the solution to 80 °C for 2 minutes and then allowing it to cool to room temperature over a period of 2-3 hours. To ensure that all the primer would be in double-stranded DNA, 20 % excess of the template strand was added.

Measurement of 3'→5' exonuclease activity. The rate of removal of nucleotides at the 3'-terminus of the primer by the 3'→5' exonuclease activity of Klenow fragment was measured by incubating at 30 °C 0.05 μM primer/template and 0.1 μM Klenow fragment in 50 mM Tris.HCl, pH 7.4 and 5 mM $MgCl_2$. After time intervals of 1-12 minutes, an aliquot was quenched into denaturing gel loading buffer (0.05% w/v bromophenol blue and 0.05% w/v xylene cyanole in formamide) and the products of exonuclease activity separated by the electrophoretic method described later.

Steady state incorporation of thymine or cytosine opposite *O*^{6}-meG in the template strand. Several sets of steady state experiments were performed. First, steady state time courses for the incorporation of thymine and of cytosine opposite *O*^{6}-meG in the template strand were measured. The reactions were carried out at 30 °C in a 200 μl volume which contained 20 μM dNTP (N= T or C), 1 μM primer-template, 0.109 μM Klenow fragment, 50 mM Tris.HCl (pH 7.4) and 5 mM $MgCl_2$. The reactions were initiated by the addition of Klenow fragment, and at time intervals from 30-420 seconds, a 20 μl aliquot was quenched into 10 μl of denaturing gel loading buffer (0.05% w/v bromophenol blue and 0.05% w/v xylene cyanole in formamide). The amount of primer elongated was measured by gel electrophoresis, autoradiography and scintillation counting, which is described later.

The next set of steady state experiments were also time course experiments, but with different amounts of Klenow fragment. The reaction conditions were the same as those described above except the concentration of Klenow in the reaction. Three concentrations of Klenow were used for each nucleotide(T and C): 0.109, 0.33 and 0.65 μM.

The effect of different concentrations of dNTP on the rate of reaction was investigated in the third set of steady state experiments. The amount of thymine or cytosine incorporated into the DNA was measured after 30 or 60 seconds after the reaction had been initiated by the addition of Klenow fragment. Each reaction was carried out at 30 °C with 1 μM primer-template, 0.109 μM Klenow fragment and 10, 40, 60, 80, 120, or 200 μM dNTP (N=T or C), in a buffer of 50 mM Tris.HCl (pH 7.4) and 5 mM $MgCl_2$.

The rapid-quench and quench-flow apparatus for pre-steady state experiments. Two rapid reaction apparatus were made for the pre-steady state experiments. The first one was a rapid-quench apparatus based on a design described by Fersht (1985), shown schematically in Figure 3.6. The reaction was initiated by mixing reactants A and B: 40 μl of each reactant was drawn from each reservoir into syringes A and B respectively (Hamilton gas tight syringes, 100 μl capacity) using a three-way valve (Hamilton HVP valve, "T" housing with 90° plug). Between each reaction, the reaction tube and four-way connector was flushed via the three-way tap that was connected to the third syringe (syringe C in Fig. 3.6). Using this three-way tap, the reaction tube was connected to a vacuum so that any remainder of the reaction mix could be sucked away by vacuum.The reaction tube was then flushed with water and air. On the first signal from the timer, the contents of syringes A and B were discharged through a four-way connector (Chromatronix) into the reaction tube. The length of the reaction tube was carefully measured so that it was just long enough to hold the reaction mixture when the contents from syringes A and B were discharged. On the second signal, 200 μl of air contained in syringe C (Hamilton gas tight syringe, 500 μl capacity) blew the reaction mix into the quench solution (0.3 M EDTA). The syringe plungers were driven by two hydraulic cylinders (Schrader Bellows model B3451 A, 1" stroke, 1" bore); plungers from syringes A and B were driven by hydraulic cylinder I and the plunger from syringe C was driven by hydraulic cylinder II. The compressed air supply (50 psi) driving these hydraulic cylinders was controlled by two Schrader solenoid valves type LB 43323 SA (obtained from Darenth Automation, Croydon, Surrey, England) triggered by a home-made timer.

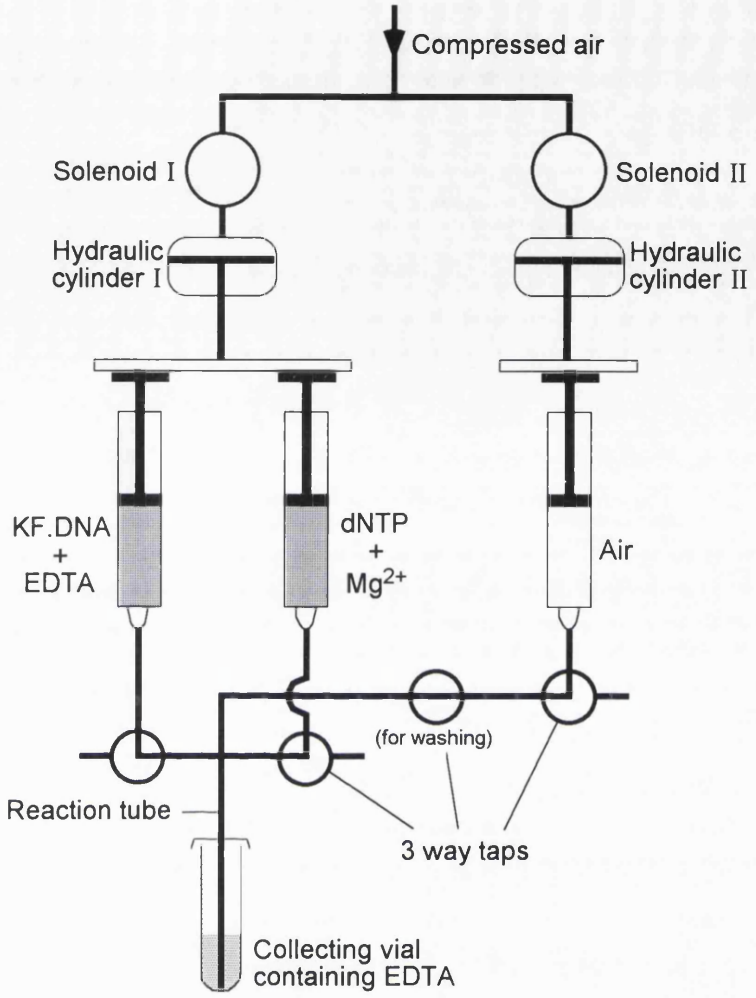


Figure 3.6. The rapid-quench apparatus built for the pre-steady state experiments described in this work.

The pre-steady state experiments described in the next section were carried out using this rapid-quench apparatus. The results obtained were reproducible and generally quite satisfactory. However, considerable difficulties were encountered in using this apparatus. First, the apparatus had to be precision made, the syringes must be absolutely vertical and the distance between the plunger at the start of the reaction and at the end of the reaction had to be very carefully measured, otherwise, the syringes would most certainly break under the great force applied by the plungers. Second, because the plungers come down with such a great force, after several firings, the screws holding the hydraulic cylinders that drive the plungers were loosen so that the position of the plunger at the start of the reaction moved slightly downwards. If uncorrected, this downward movement could cause the syringes to break. This problem was eventually overcome by putting in more screws and moving the position of the hydraulic cylinders so that the syringe plunger does not reach the very end of the syringe when the

hydraulic cylinders were fired. A third problem was that all the joints had to fit perfectly. There was a particular problem with the joint between the syringe tip and the screw that connects the syringe tip to the Hamilton valve blocks. Although these screws were made to fit syringes with luer tips, the fitting was unsatisfactory under high pressure because when the plungers were fired, a small amount of the reactant solution escaped from this joint and created an aerosol. Since one of the reactant solutions contained radioactivity, this presented a health hazard. Besides, the accuracy of the experiments would be affected as well. This problem was eventually overcome by the use of syringes with luer lock tips rather than luer tips. A fourth problem which could not be overcome, was that washing and refilling the syringes with reactants between each experiment required great concentration of the mind. When a hundred or more reactions had to be carried out on the same day, this could become quite tedious.

Because of the difficulties in operating the rapid-quench apparatus, it was decided that a second apparatus which is simpler to use would be built. Thus, the quench-flow apparatus was built. This was based on a design described by Johnson in Methods in Enzymology (1986) and is shown schematically in Fig. 3.7. Again, the reaction was initiated by mixing reactants A and B. Syringes A and B (Hamilton gas tight syringes, 250 µl) were each filled to maximum capacity with the respective reactants via a three-way valve (Hamilton HVP valve, "T" housing with 90 ° plug). A stepper motor (200 steps per revolution, type M41200-125B from Motori ed Apparecchiature Elettriche, Ivrea, Italy) was used to drive a ball screw (thread pitch 1.25 mm/s) connected to a plate that pushed the syringe plungers at a constant speed selected by the user (Fig. 3.7). The syringes were driven at a specified velocity to discharge the reactants at a certain speed through a four-way connector (Chromatronix) into a reaction tube. Reaction would take place as the two reactants were mixed and forced through the reaction tube. The end of the reaction tube was carefully placed in some quench solution (0.3 M EDTA), so that the reaction was quenched as soon as it reached the end of the reaction tube. The reaction time was determined by the time required for the reaction mix to flow through the length of the reaction tube, so this could be controlled by varying the flow rate through the reaction tube, and the volume of the reaction tube. For example, for a flow rate of 0.02 ml/sec and a reaction tube of 2 µl, the reaction time would be 100 msec.

With the quench-flow apparatus, it was possible to fill the syringes with a large volume of reactant (usually 250 µl was used) and then discharge only 30 µl for each reaction. This had the advantage that the syringes would only need refilling after every 7-8 reactions. However, this apparatus failed to give either reproducible or sensible results (see Fig. 3.21 and 3.22, page 124, Results section). The inherent problem with an apparatus of this kind is the difficulty of getting turbulent flow in the reaction tube. Even if the mixing is good when the two reactants meet head on (Fig. 3.7), if the subsequent flow in the reaction tube is laminar rather than turbulent, the

material in the centre of the tube will flow at a faster rate and thus have a shorter transit time than the peripheral material.

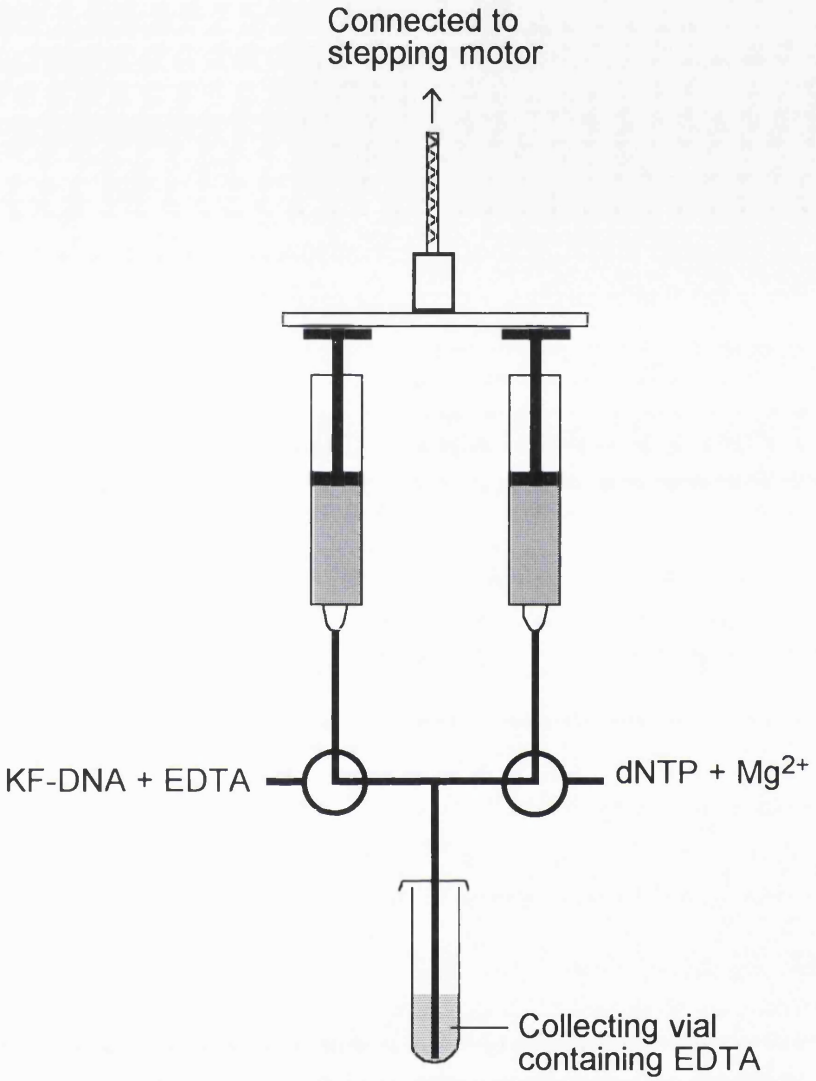


Figure 3.7. The rapid-flow apparatus built to carry out the pre-steady state experiments. This apparatus failed to give either reproducible or sensible results, thus all the pre-steady state experiments were carried out with the rapid-quench apparatus (Figure 3.6).

The basic work on laminar and turbulent flow in pipes was done by Reynolds in 1883. Reynolds's work was with pipes of fairly large diameter but for the sake of argument we can assume that it is valid for the small diameter tubes used in the quench-flow apparatus. Reynolds showed that the flow in a pipe is turbulent if the Reynolds number, N_R, is greater than 2000:

$$N_R = \frac{v.D.\rho}{\mu}$$

where v = velocity, D = diameter of the pipe, ρ = density of the fluid, and μ = viscosity. For water, $\rho = 1000$ kg/m^3 and $\mu = 1 \times 10^{-3}$ N.s/m^2 (the units of Newton are kg.m.s^{-2}). To maximise the velocity of the flow, a tubing with the smallest available internal diameter was used. This had an internal diameter of 0.005", i.e. 0.127 mm. In order to achieve turbulent flow, the critical velocity v_c in this tubing is therefore

$$v_c = \frac{N_R.\mu}{D.\rho} = \frac{2000.1 \times 10^{-3}}{0.127 \times 10^{-3}.1000} = 15.7 \text{ ms}^{-1}$$

The internal diameter of the tubing was 0.127 mm, so the volume of the tubing was 0.0127 µl/mm length. To achieve the critical velocity, one would need to pump at 157000 x 0.0127 µl/s, i.e. approximately 200 µl/s. I used 25 µl of A and 25 µl of B for each reaction, i.e. total reaction volume was 50 µl. This means that all the reactant had to be discharged in 250 ms.

The stepper motor used by Johnson (1986) had 200 steps/turn and a maximum speed of 5 turns/s. The stepping motor used in my quench-flow apparatus also had 200 steps/turn but a maximum speed of only 2 turns/s. In 250 ms this would move only half a turn, and for a thread pitch of 1.25 mm/turn, the syringe plungers would advance 0.625 mm in 250 ms. In order to discharge 25 µl from each syringe with an advance of 0.625 mm, one would need a syringe with a very large bore. For comparison, a 250 µl Hamilton syringe has a capacity of 3.85 µl/mm; in order to deliver 25 µl/0.625 mm, one would need to use a 2.6 ml syringe. Plainly the inherent inaccuracy of driving such a wide-bore syringe over such a short distance precludes their use. It is obvious that there is very little that one can do to avoid this problem without increasing the volume of the reaction. Yet it was impossible to increase the reaction volume to that required (at least 300 µl) because very large quantities of DNA, Klenow fragment and radioactivity would be required, and the length of the reaction tube would have to be ridiculously long.

The consideration of the problems with the quench-flow apparatus emphasizes the difficulty of rapid reactions. It was clear that this apparatus could not be used and consequently all the pre-steady state experiments were performed with the rapid-quench apparatus.

Pre-steady state incorporation of thymine and of cytosine opposite O^6-meG. Several pre-steady state experiments were performed and all of these experiments were carried out with a slight excess of enzyme over DNA to ensure that the catalytic cycle followed did not involve the dissociation of the enzyme from the product DNA. The first experiments were to measure the rates of incorporation of thymine and of cytosine opposite O^6-meG in the template strand: reservoir A was filled with 0.2 µM DNA primer/template, 0.24 µM Klenow fragment, 2.5 mM EDTA, and 50 mM Tris.HCl, pH 7.4 and reservoir B with 10-160 µM dTTP or dCTP, 12.5 mM $MgCl_2$, and 50 mM Tris.HCl, pH 7.4. Reagent A (40 µl) was mixed with reagent B (40 µl) at room temperature, thus during reaction, the concentrations of the reactants were 0.12 µM

Klenow, 0.1 μM DNA and 5-80 μM dNTP. After time intervals between 200 ms and 10 s (for dTTP) or 1 s and 15 s (for dCTP), the reaction was quenched into 40 μl of 0.3 M EDTA, pH 7.4. The elongation of the primer was measured as described below. All experiments were carried out in triplicates as a check on the reproducibility of the measurements.

The initial rates of incorporation of different concentrations of thymine and of cytosine opposite O^6-meG in the template strand were also measured. These assays were carried out with 0.12 μM Klenow fragment, 0.1 μM DNA and 5-100 μM dNTP (N= T or C). In order to measure the initial rates of incorporation, the reactions were quenched at a time when only ~15% of the primer would be elongated. The sampling times for different concentrations of dTTP and dCTP used are listed in Table 10.

Table 10. Sampling times for measuring the initial rates of incorporation at different [dNTP]s.

[dTTP] in μM	Sampling time (ms)	[dCTP] in μM	Sampling time (s)
5	1500	5	10
10	1000	10	5
20	750	20	5
30	500	30	1.5
50	250	50	1
100	250	100	1
150	250	150	1
200	250	200	1

Pre-steady state polymerisation using phosphorothioate nucleotides. As mentioned in the introduction to this chapter, the comparison of the rates of pre-steady state polymerisation of normal dNTP and of their phosphorothioate analogues were used extensively in the elucidation of the kinetic mechanisms for Klenow fragment (Bryant et al., 1983; Mizrahi et al., 1985; Kuchta et al., 1987 and 1988) and T7 DNA polymerase (Patel et al., 1991 and Wong et al., 1991). Thus, the incorporation of (S_p)-dTTPαS and of (S_p)-dCTPαS opposite O^6-meG in the template strand was measured to see if the formation of phosphodiester bond is rate limiting in the incorporation of thymine and cytosine opposite a template O^6-meG.

Rapid-quench kinetics of the incorporation opposite O^6-meG of the *S*-stereoisomer of the 2'-deoxynucleoside 5'-O-(1-thiotriphosphates), (S_p)-dTTPαS or (S_p)-dCTPαS, were investigated using 0.1 μM DNA, 0.12 μM Klenow fragment, and 40 μM dNTPαS. The reactions were initiated by the addition of magnesium and the (S_p)-dNTPαS to a solution containing EDTA and the Klenow-DNA complex as described in the previous paragraph. After a

reaction time of 400 ms -18 s, the reaction was quenched into 0.3M EDTA. The amount of DNA elongated was quantitated by gel electrophoresis followed by autoradiography and scintillation counting (described below).

The effect of thio-substitution was further investigated by a competition assay where both dTTP and (S_p)-dTTPαS were allowed to compete for incorporation by Klenow fragment opposite O^6-meG in the template strand. These experiments were carried out to see if the rate of phosphodiester bond formation is the only rate-limiting step during the incorporation of (S_p)-dTTPαS. The assays were carried out with 0.1 μM DNA, 0.12 μM Klenow fragment, 20 μM dTTP, and 0-40 μM (S_p)-dTTPαS. The reactions were initiated by the addition of magnesium and dTTP/(S_p)-dTTPαS to a solution containing EDTA and the Klenow-DNA complex, and were quenched into EDTA after 500 ms to 4 s. The amount of DNA elongated was quantitated by gel electrophoresis followed by scintillation counting (described below).

Addition of the next correct nucleotide after a thymine or a cytosine base-paired with O^6-meG in the template strand. DNA duplexes B, C, and D (Table 9) were used to study the rates of addition of the next correct nucleotide after a thymine (duplex B) or a cytosine (duplex C) that has been base-paired with O^6-meG in the template strand. Duplex D, which contained a 3'-terminal thymine in the primer strand paired with adenine in the template strand, was included as a control. A solution containing 50 μM dNTP (N= A for duplexes B and C, and N= C for duplex D), 1 μM primer/template, 0.02 μM Klenow fragment, 50 mM Tris.HCl, pH 7.4 and 5 mM $MgCl_2$ was incubated at 30 °C. After 15-120 s, the reaction was quenched into denaturing gel loading buffer (0.05% w/v bromophenol blue and 0.05% w/v xylene cyanole in formamide). The amount of primer elongated was measured as described below.

Pre-steady state experiments that measure the addition of the next correct nucleotide after a thymine or a cytosine base-paired with O^6-meG in the template strand have also been carried out. However, due to time constraints, the optimal concentrations of these polymerisation experiments have not been found. The experimental details of these experiments are given here to provide a guideline to future experiments in this area. Experiments to measure the pre-steady state rate of addition of the next correct nucleotide after a thymine or a cytosine base-paired with O^6-meG in the template strand were carried out in almost the same way as the pre-steady state experiments for measuring the rates of incorporation of cytosine or thymine opposite O^6-meG in the template strand with the rapid-quench apparatus. The only differences were that smaller concentrations of the next correct nucleotide and shorter reaction times were used. This was because the addition of a nucleotide following a thymine or cytosine base-paired with O^6-meG in the template strand was much faster than the incorporation of thymine or cytosine opposite O^6-

meG in the template strand. The assays were carried out with 0.1 μM DNA, 0.12 μM Klenow fragment and 2.5-10 μM dATP. The reactions were initiated by the addition of magnesium and dATP to a solution containing EDTA and the Klenow-DNA complex, and were quenched into EDTA after 200 ms to 1 s. The amount of DNA elongated was quantitated by gel electrophoresis followed by scintillation counting (described below). The results of these experiments showed that the reaction times used were too long: by the time the first observation was made (200 ms) the amount of primer elongated had already reached the maximum limit and there was no change in the level of product throughout the observation period of 200 ms - 1 s. Thus, any future experiments to study the pre-steady state incorporation of the addition of the next correct nucleotide after a thymine or a cytosine base-paired with O^6-meG in the template strand should use a shorter reaction time or lower concentrations of the next correct dNTP, or a combination of both.

Measurement of elongation by gel electrophoresis, autoradiography and scintillation counting. After the reaction, the elongated DNA was separated from the parent primer by electrophoresis using 20% polyacrylamide gels containing 7 M urea. The gels were then dried and the ^{32}P-labelled DNA located by autoradiography. The amount of radioactivity present in each band was quantitated by cutting the band out and scintillation counting in 4 ml of Ecoscint A (National Diagnostics).

In all the polymerisation experiments described, a 5'-^{32}P-labelled primer was used so that both unelongated and elongated primers could be seen on the gel. In theory, it is also possible to use an unlabelled primer and quantitate the amount of dNTP incorporated by using α -^{32}P-dNTP, in which case only the products of elongation can be visualized. This method has been tried and found to be unsatisfactory for several reasons. First, in order to accurately measure the amount of primer elongated, the same volume had to be loaded in each well during gel electrophoresis. Since very small volumes (< 4 μl) were loaded in each well, small pipetting errors could be magnified. Even with the use of syringes and tips specially designed for loading small volumes (1-3 μl), it was still difficult to minimise loading errors. Second, to measure small amounts of elongation (< 20% elongation), a high specific activity of the α-^{32}P-dNTP had to be used. This was not only a health hazard, but it also increased the background radioactivity of the gel which affected the accuracy of measurement, especially in cases where the amount of primer elongated was quite small.

There is an alternative method for measuring the amount of elongation which also uses cold primer-templates and α-^{32}P-dNTPs. This is the DE81 filter assay frequently used by workers from the laboratory of Benkovic. This method is very simple: an aliquot of the reaction mix is spotted onto a DE81 filter paper (based on DEAE-cellulose) which can selectively retain

oligonucleotides but not nucleotides after suitable washing procedures. However, when this method was used to separate cold oligonucleotides from α-^{32}P-dNTP, and to separate 5'-^{32}P-labelled oligonucleotides from cold dNTPs, satisfactory separation was not achieved despite numerous attempts. Thus, in all the polymerisation experiments carried out, 5'-^{32}P-labelled primers and cold dNTPs were used, and the elongation was quantified by gel electrophoresis followed by autoradiography and scintillation counting as described above.

Computer analysis of data to determine rate constants. Rate constants for the incorporation of a single nucleotide opposite O^6-meG in the template DNA strand were calculated by mathematical analysis of the results obtained (i.e. the observed extent of elongation as a function of time) to the reaction scheme depicted in Figure 3.5 using the computer programme FACSIMILE (United Kingdom Atomic Energy Authority, Harwell; available from ARC Scientific, 257 Woodstock Road, Oxford, OX2 7AE, England). FACSIMILE is not specifically written for enzyme kinetics but is a very powerful computer program for analysing data which depends on processes that can be described by differential equations, and offers particular advantages for this work which will be discussed below. The program will run on a mainframe computer or IBM 386 or 486 PCs. In this work it was run on an IBM PS2/30 upgraded from 80286 to 80386 SX (25 MHz) (SX Now, Kingston Technology Corporation, 17600 Newhope, Fountain Valley, CA 92708, USA) and fitted with an Intel 80387 SX maths coprocessor. The SX Now upgrade board contains a 16 kB memory cache which increases the speed of execution about three fold. FACSIMILE is written in FORTRAN so it requires an ANSI-FORTRAN 77 compiler (FTN 77/486, from Salford Software Marketing Ltd, Computing Centre, University of Salford, Salford M5 4WT, England), and the necessary run time software DBOS, Version 2.60.

Competition assays. The ultimate test for the mathematical model elucidated is to compare the predicted outcome with the observed kinetics. The FACSIMILE program was used to predict what would happen under two very different experimental conditions, and experiments were then carried out under these experimental conditions to see whether the predictions were correct. Two competition assays were carried out where a mixture of dTTP and dCTP were incubated with Klenow and DNA containing O^6-meG in the template strand. The two experimental conditions chosen, i.e. the ratio of the concentration of the two nucleotide triphosphates, were those which the computer predicted that the incorporation of thymine is approximately four times that of cytosine, and where the incorporation of thymine is similar to that of cytosine.

The two thymine/cytosine competition assays were carried out which used 5'-^{32}P-labelled DNA primer annealed with a complementary template strand (duplex A, Table 9) and cold dCTP and dTTP. In the first competition assay, 25 µl of solution A (0.20 µM DNA

primer/template, 0.24 μM Klenow fragment, 2.5 mM EDTA, and 50 mM Tris.HCl, pH 7.4) was mixed with 25 μl of solution B (40 μM dTTP, 40 μM dCTP, 12.5 mM $MgCl_2$, and 50 mM Tris.HCl, pH 7.4) at room temperature, and after time intervals between 500 ms and 4 s, the reactions were quenched into 50 μl of 0.3 M EDTA, pH 7.4.

In the second competition assay, 25 μl of solution A (0.20 μM DNA primer/template, 0.24 μM Klenow fragment, 2.5 mM EDTA, and 50 mM Tris.HCl, pH 7.4) was mixed with 25 μl of solution B (40 μM dTTP, 160 μM dCTP, 12.5 mM $MgCl_2$, and 50 mM Tris.HCl, pH 7.4) at room temperature. After time intervals between 200 ms and 4 s, the reactions were quenched into 50 μl of 0.3 M EDTA, pH 7.4.

There were substantial difficulties in separating the products of elongation because the primer elongated by a thymine and that elongated by a cytosine, hereafter referred to as the 9+T-mer and 9+C-mer respectively, differed by only one base. Although they had slightly different mobilities in the gel, the separation by gel electrophoresis was not good enough for accurate quantitation by scintillation counting. Thus, initially, a double label competition assay was carried out using α-^{32}P-dCTP and α-^{33}P-dTTP nucleotides (α-^{33}P-dTTP is available from NEN Du Pont and Amersham). It was not possible to use a ^{35}S-nucleotide because this would not incorporate well into DNA. ^{3}H-nucleotides were considered but rejected on the grounds that its long half-life could lead to disastrous results for the laboratory if contamination should occur. The β-radiation from ^{32}P and ^{33}P each has energy-emission at 1.7 MeV and 0.25 MeV respectively, thus it was possible to perform a dual-label scintillation counting by counting in 3-windows: from 0.05-0.15 MeV, from 0.15-1.7 MeV and from 0.05-1.7 MeV. A third of the counts from the ^{32}P is found in the 0.05-0.15 MeV window, but the inaccuracies from this could be reduced by using more ^{33}P than ^{32}P in each experiment (at least 4-times more ^{33}P than ^{32}P).

However, since the ratio of the concentrations of dNTP:DNA ranged from 20 μM dNTP:0.1μM DNA to 80 μM dNTP:0.1μM DNA, only a minute proportion of each nucleotide would be incorporated and this presented many problems. Calculations based on predicted incorporation showed that the specific activities of the dNTPs required for the experiments must be at least 10-20 μCi/nmol. Kuchta et al. (1988) used a specific activity of 200 μCi/nmol of dNTP for their misincorporation experiments. A comparison of the required specific activity with that of the α-^{33}P-dTTP stock from NEN (800 Ci/mmol i.e. 800 μCi/nmol) gives an idea of how high this was. Because the α-^{33}P-dTTP was a relatively new product, it was very expensive (approx. £250 excl. VAT for 250 μCi) and this limited the number of experiments that could be performed.

Another problem was that only a small volume from the reaction mix could be loaded onto the gel, thus the volume of the reaction mixture had to be reduced by ~ 10-fold. Since 0.3 M EDTA was used to quench the reaction, after evaporation the concentration of EDTA became 3 M and this affected gel electrophoresis. Denaturing gel loading buffer was not used for quenching the reaction because the effectiveness of formamide as a quench for rapid reactions was unknown and because evaporation of the formamide is exceedingly difficult.

Because of these numerous problems with the gel electrophoresis, it was decided that another method for separation of the 9+T-mer and 9+C-mer from the radiolabelled dNTPs should be used. In the first attempt gel filtration was used. Trials were carried out with Biogel-P4 and Biogel-P6 (Biorad) but they were unsuccessful because of the very large amounts of ^{32}P-dCTP and ^{33}P-dTTP and the very small amounts of labelled primers. Next, the separation was carried out by HPLC on a Dionex Nucleopac PA 100 column. The separation of the three oligonucleotides was good, but the reproducibility of the results were not satisfactory. Although it seemed probable that a good separation of the three oligonucleotides might be achieved by altering the conditions of anion exchange, this method was not pursued because it had been found that the 9+T-mer could be very well separated from the 9-mer and the 9+C-mer on a Pharmacia Mono-Q column (FPLC) with good reproducibility (Figure 3.8). It was therefore decided that the competition experiments should be carried out with a 5'-labelled primer and the reaction mix analysed by gel electrophoresis and FPLC. Since the method uses 5'-^{32}P-labelled DNA primer rather than α-^{32}P-labelled dNTP, the percentage of primer elongated by T and by C could be measured from the gels while the percentage of 9+T-mer could be determined from FPLC, and by subtraction, the percentage of 9+C-mer could be determined.

The buffers used for ion-exchange chromatography on a Pharmacia Mono-Q HR 5-5 column were: buffer A: aqueous 10 mM NaOH, 0.4 M NaCl (pH 12), and
buffer B: aqueous 10 mM NaOH, 0.8 M NaCl (pH 12).
The buffer gradients used are shown in Table 10. The flow rate was 0.7 ml/min. 10 µl of each reaction sample was injected for analysis and the eluate collected every 0.7 minute. The amount of radioactivity present infraction was quantitated by scintillation counting in 4 ml of Ecoscint A (National Diagnostics).

Table 11. Buffer gradients for the separation of elongation products of thymine/cytosine competition assays on a Pharmacia Mono Q HR 5-5 column.

Time in minutes	Buffer A (0.4 M NaCl, pH12)	Buffer B (0.8 M NaCl, pH12)
0	100%	0%
2	100%	0%
5	85%	15%
25	65%	35%
28	0%	100%
30	100%	0%

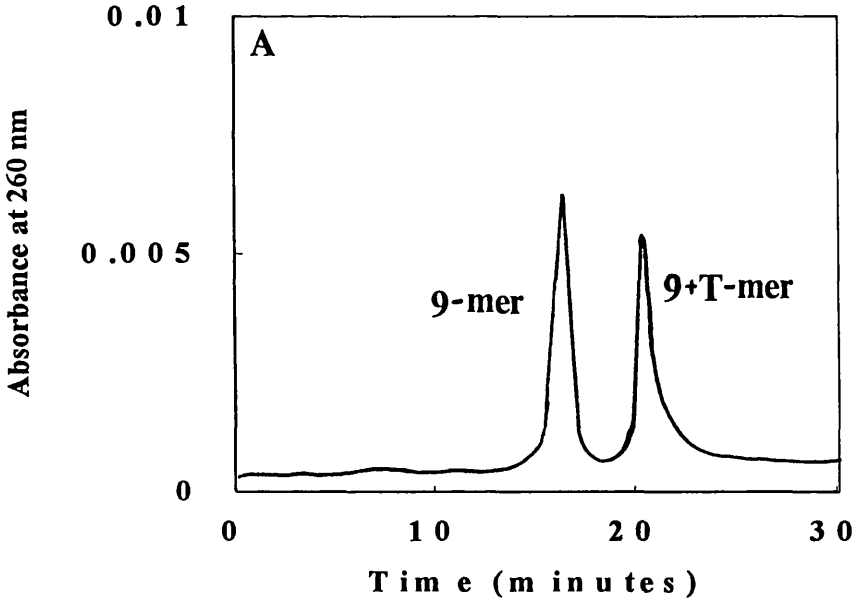
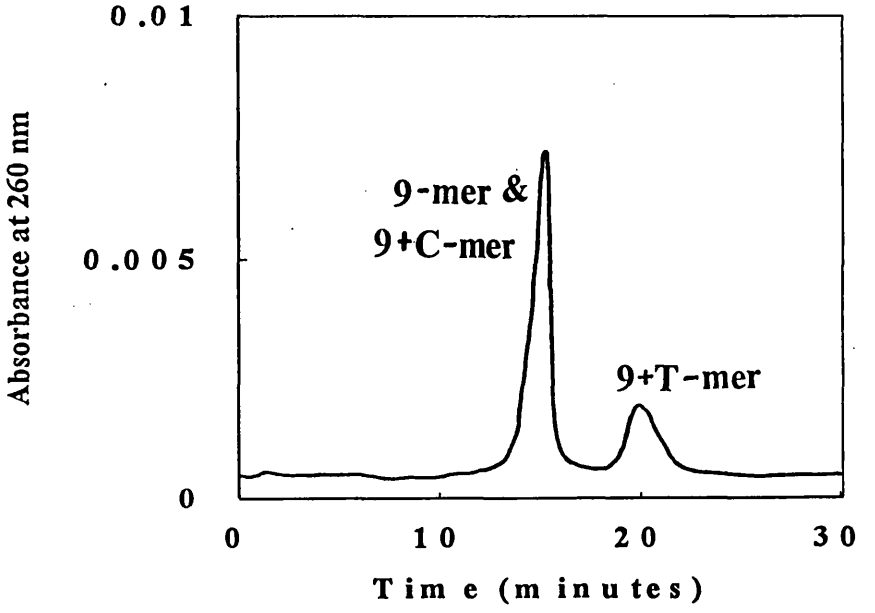


Figure 3.8. FPLC profiles (above and following page) showing the separation of the products of elongation obtained in the thymine/cytosine competition assays. Of the three oligonucleotides to be separated, the 9+T-mer was very well separated from the 9+C-mer and the unelongated 9-mer primer (Figures A and B), while the 9+C-mer was eluted almost immediately after the elution of the 9-mer (i.e. the unelongated primer) (Figures B and C). In (A) and (B), different amounts of DNA were coinjected for so that the area under each peak could be used as a further check for the elution positions of each primer. In (A), the ratio of the primers injected was 4:3 for the 9-mer and 9+T-mer. In (B), the ratio of the DNA injected was 2:1:1.5 for the 9-mer, 9+T-mer and 9+C-mer respectively.



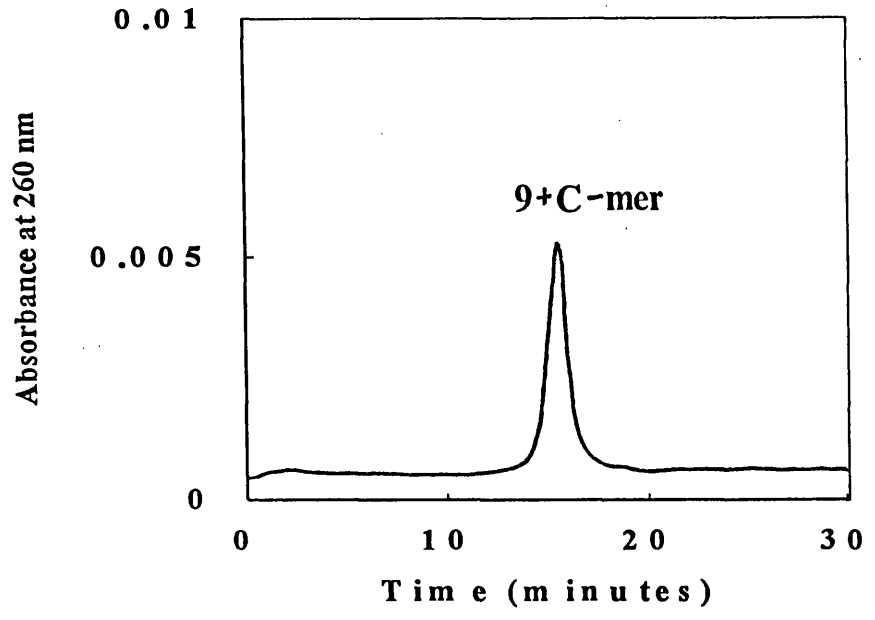


Figure 3.8 (B and C). Legend on previous page.

3.3 Results

3'→5' exonuclease activity on a primer with a 3'-terminal thymine or cytosine base-paired with O^6-meG in the template DNA strand. Figure 3.9 is an autoradiograph showing the degradation by Klenow fragment of a 5'-labelled 10-mer primer annealed with a complementary 20-mer template strand (duplex C in Table 9). Figure 3.10 shows the 3'→5' exonuclease activity of Klenow fragment on a primer with a 3'-terminal thymine or cytosine base paired with O^6-meG in the template DNA strand; a parallel experiment using a duplex containing a 3'-terminal thymine base paired with an adenosine in the template strand was included as a control. The exonuclease rates (legend to Figure 3.10) were calculated from the graphs of log (fraction of unshortened DNA) against time (Figure 3.10). Note that the exonuclease rates are much slower than the rates of polymerisation.

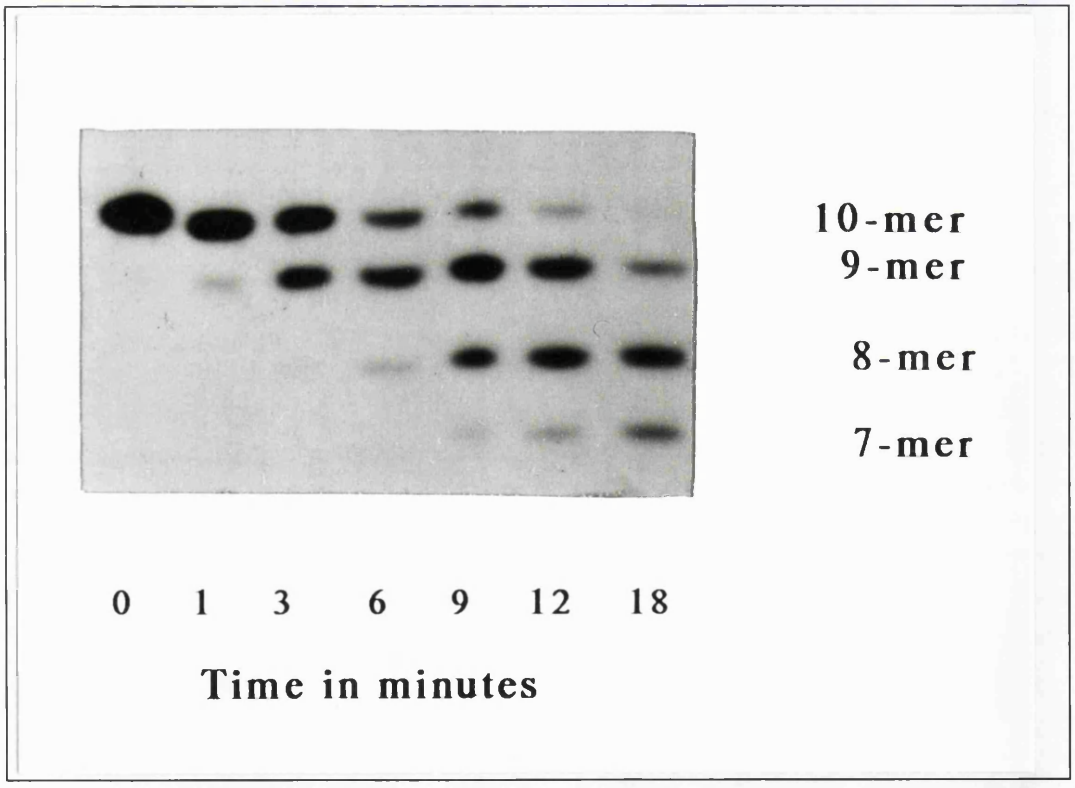


Figure 3.9. Autoradiograph showing Klenow 3' to 5' exonuclease activity at various time intervals on the primer/template 5'ATC CGA TAG C 3'
3'TAG GCT ATC meG GTC ATT CTC GC 5'

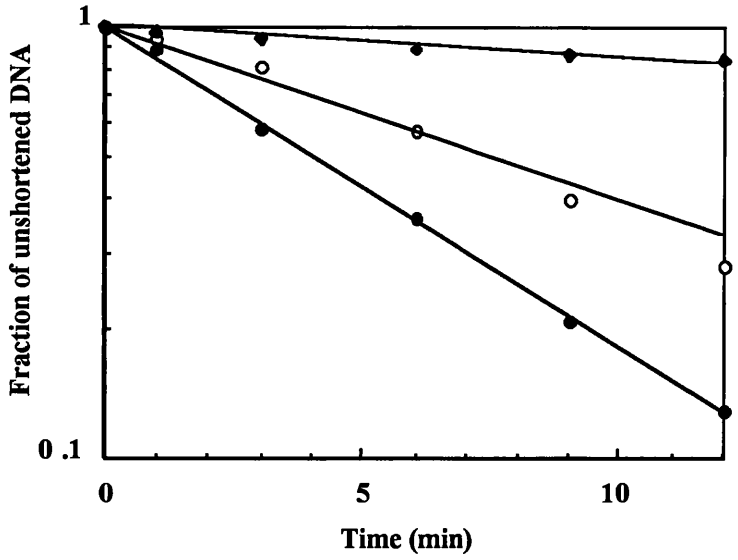


Figure 3.10. 3'→5' exonuclease rates with duplexes B, C and D (Table 9). k_{exo} for each DNA duplex was measured as described under Materials and Methods using 0.05 μM labelled DNA duplex and 0.1 μM Klenow fragment. In duplexes B (○) and C (●), the 3'-terminal base of the primer was either thymine or cytosine, paired with an O^6-meG in the template strand (k_{exo} = 1.6 x 10^{-3} s^{-1} and 2.8 x 10^{-3} s^{-1} respectively). In duplex D (♦), the 3' terminal thymine was paired with an adenine (k_{exo} = 6.5 x 10^{-4} s^{-1}).

Steady state incorporation of thymine or cytosine opposite O^6-meG in the template strand. These experiments were carried out with a ten-fold excess of DNA (1 μM) over Klenow fragment (0.1 μM) and 20 μM dNTP. The results are shown in Figures 3.11 to 3.13. Figure 3.11 shows the steady state time courses for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. Note that the graph for the incorporation of thymine opposite template O^6-meG does not extrapolate to zero on the vertical axis but intercepts it at approximately 0.05 μM at zero time (Fig. 3.11(b)). This initial rapid polymerization before the steady-state is achieved (the pre-steady state burst) represents polymerization of the DNA to which the enzyme was originally bound. The presence of this pre-steady state burst shows that the rate of the steady state is limited by a post-synthetic event rather than by the rate of polymerization. This is discussed at greater length below. For the incorporation of thymine, the steady state gradient (30-120 s) was calculated as 0.029 μmol/s/μmol enzyme and represents the lower limit for the value of the rate-limiting post-synthetic event. This is of similar magnitude to the dissociation rate of Klenow fragment from a 14/20-mer duplex DNA (0.06 s^{-1}, Kuchta et al., 1987).

In contrast to thymine there was no evidence for any pre-steady state burst with cytosine. This indicates that with cytosine the rate of polymerization is the predominant factor limiting the rate of the steady state, and that in this case the overall rate is not limited by the rate of some reaction occurring after the actual addition of the nucleotide to the primer strand. The steady state rate of incorporation of cytosine opposite template O^6-meG was calculated from the initial slope as 0.016 μmol/s/μmol enzyme.

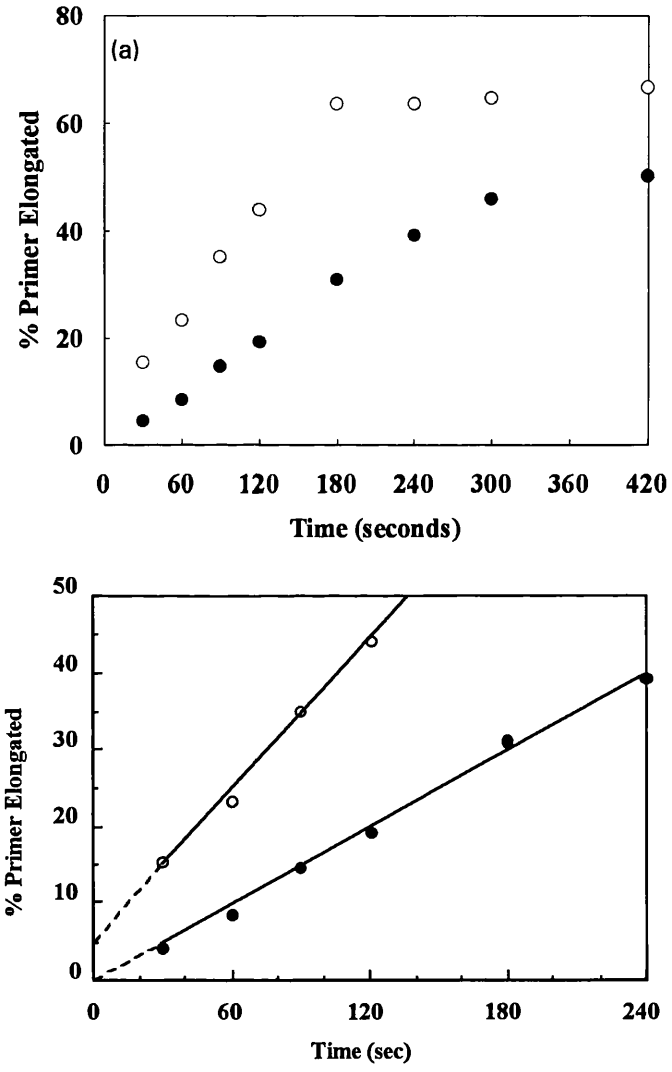


Figure 3.11 (a) shows the steady state time courses for the incorporation of thymine (○) and of cytosine (●) opposite O^6-meG in the template DNA. 0.1 μM Klenow fragment was incubated with 1 μM DNA duplex A and 20 μM dTTP or dCTP as described in the Materials and Methods section. The data points at time points between 30 seconds and 240 seconds are shown in (b) where each set of data points is joined by a straight line and then extrapolated back to the vertical axis.

Figures 3.12 (a) and (b) show the effect of different concentrations of Klenow on the rates of incorporation of thymine (a) and of cytosine (b) opposite O^6-meG in the template strand. With the incorporation of thymine, the reaction appeared to be near saturation when the E:DNA ratio was 0.327:1 and reached a plateau at approximately 93 % incorporation. The elongation did not reach 100% completion even with an E:DNA ratio of 0.654:1. These plateax were observed at 93-95 % completion at the higher enzyme concentrations probably because at that stage in the reaction, the rate of polymerisation was the same as the rate of the 3' to 5' exonuclease activity.

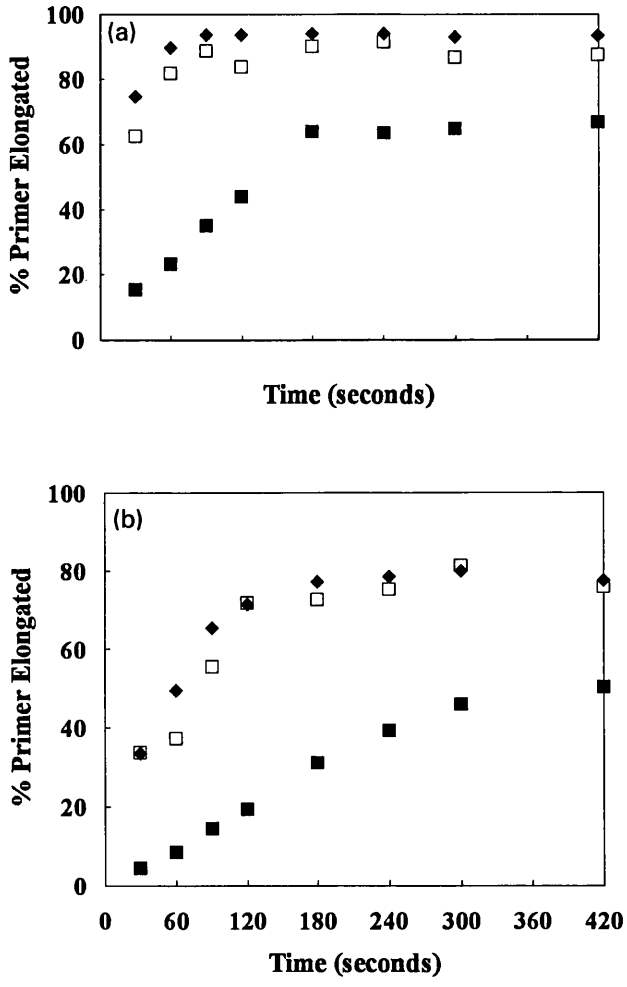


Figure 3.12. Steady state incorporation of thymine (a) or cytosine (b) opposite O^6-meG in the template strand. The assays were carried out with 1 μM DNA duplex A and 20 μM dTTP or dCTP at three enzyme concentrations: 0.109 μM (■), 0.327 μM (□) and 0.654 μM (♦) as described in the Materials and Methods section.

The rates of incorporation of cytosine opposite O^6-meG (Figure 3.12b) in the template strand were considerably slower than those for the incorporation of thymine. Although the reaction also appeared to near saturation when the enzyme:DNA ratio was 0.327:1, the maximum amount of primer elongated by cytosine was only about 80% observed (c.f. with > 90% for thymine).

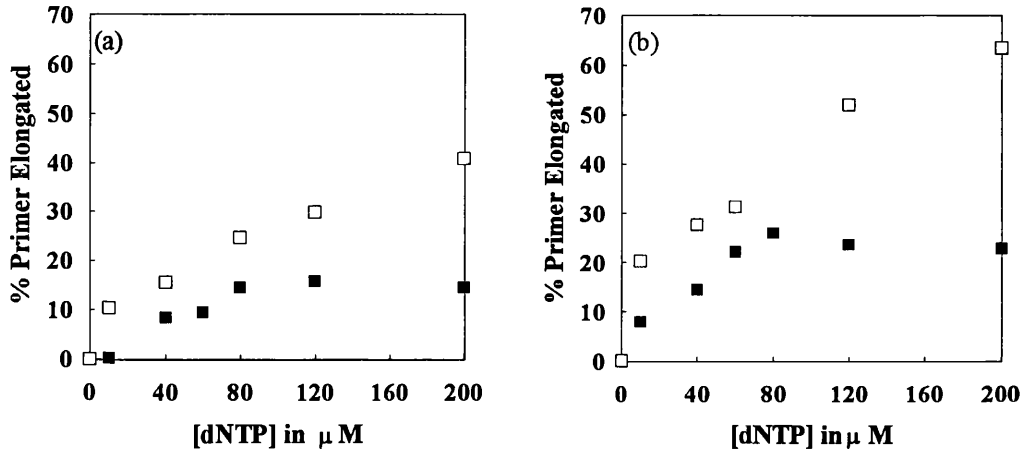


Figure. 3.13 Rates of incorporation of thymine (□) and of cytosine (■) opposite O^6-meG in the template strand at different concentrations of the nucleoside triphosphates (10 μM to 200 μM). Graphs (a) and (b) were respectively constructed from data points taken at 30 s and 60 s after the reaction has started. The experiments were performed with a polymerase:DNA ratio of 0.109 μM:1 μM.

Figure 3.13 shows the effect of different concentrations of dNTP on the steady state rate of incorporation of thymine and of cytosine opposite O^6-meG in the template strand under conditions of excess DNA over enzyme. The graphs for the incorporation of cytosine at 30 s and 60 s are both hyperbolic, with a K_m of about 30-35 μM. However, with the incorporation of thymine, the reaction profiles at 30 s and 60 s are not hyperbolic but appeared to consist of two linear phases. Figures 3.13(a) and (b) cannot be compared with the classical Michaelis-Menten graph of rate of reaction against concentration of substrate because the latter is constructed from initial reaction rates whereas in Figures 3.13(a) and (b), most of the rates measured were not initial rates.

Pre-steady state incorporation of thymine or cytosine opposite O^6-meG. Figure 3.14 is an autoradiograph showing the pre-steady state incorporation of 20 μM dTTP over a time course of 200 ms to 5 s. The primer-template used (Duplex A) was designed to allow the incorporation of a single nucleotide opposite the O^6-meG in the template strand under the experimental conditions. All the polymerisation assays carried out with thymine showed only one elongation product, i.e. a 10-mer formed by the incorporation of a single thymine nucleotide. No accumulation of an 11-mer was observed even at the highest dTTP concentration used (80 μM).

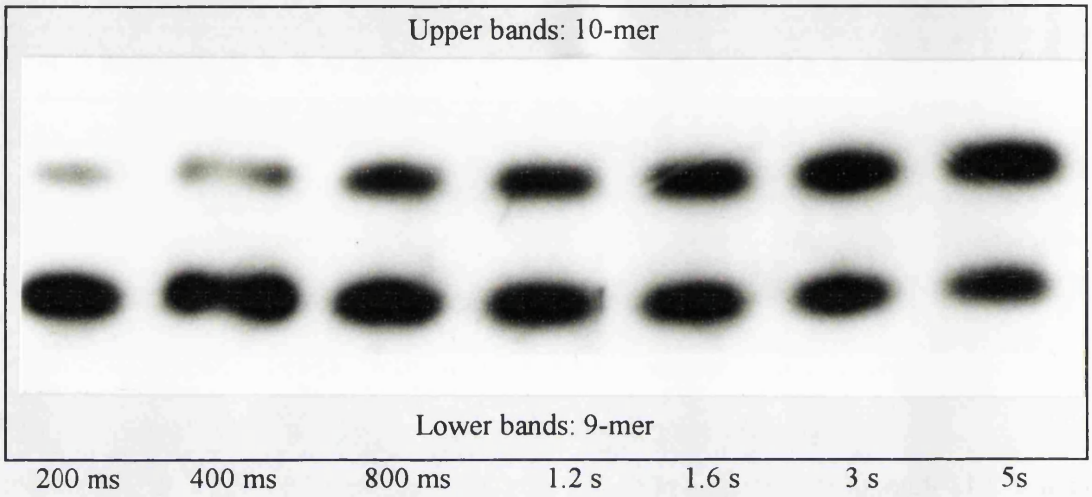


Figure 3.14. Autoradiograph showing the pre-steady state incorporation of 20 μM dTTP opposite O^6-meG in the template DNA strand. The dTTP was incorporated into a pre-formed Klenow-DNA complex over a time course of 200 ms - 5 s. The enzyme:DNA ratio used was 0.12 μM:0.1 μM. The lower and upper bands in each track represent the 5'-radiolabelled unelongated primer (9-mer) and the primer elongated by a thymine nucleotide (10-mer) respectively. Seven tracks of the gel are shown, corresponding to samples obtained at 200 ms, 400 ms, 800 ms, 1.2 s, 1,6 s, 3 s and 5 s respectively from left to right, as indicated by the labels next to the autoradiograph.

During pre-steady state incorporation of 80 μM of dCTP opposite O^6-meG in the template strand, a small amount (up to 6% of the total DNA at 30 s) of the product of elongation was further elongated to an 11-mer (Figure 3.15). This misincorporation of a further nucleotide (cytosine opposite a template thymine) was not detected following the incorporation of thymine opposite template O^6-meG. The possible importance of this is discussed in a later section.

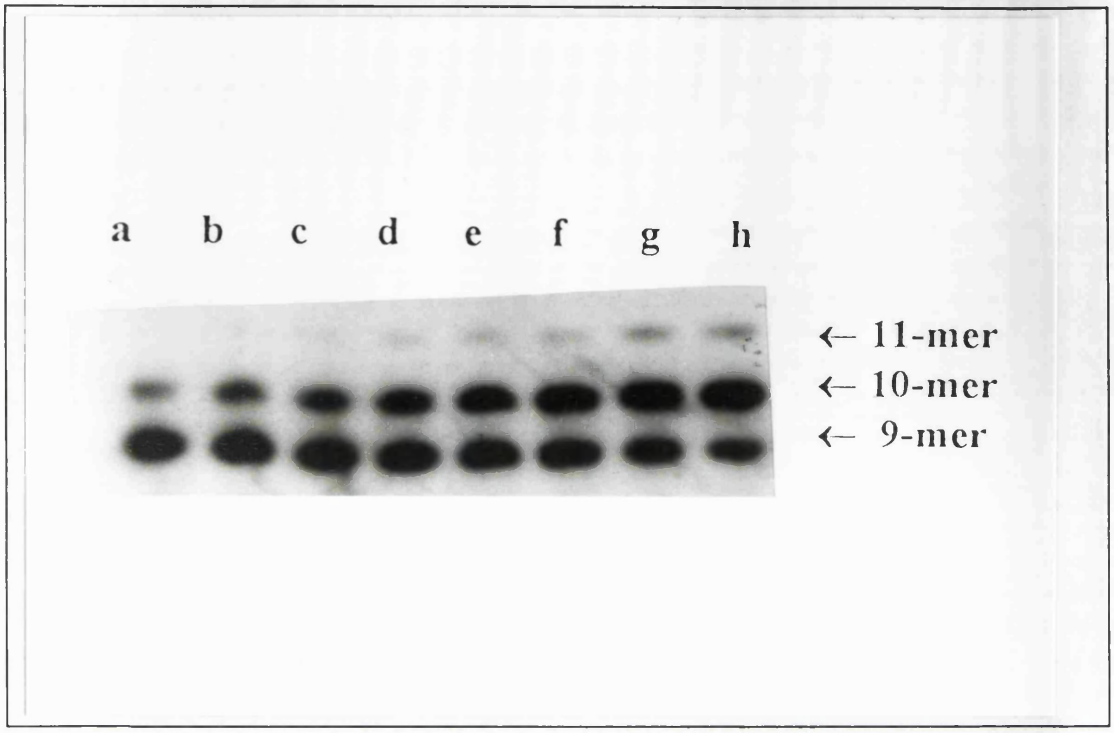


Figure 3.15. Autoradiograph showing the accumulation of a 10-mer and an 11-mer during the incorporation of dCTP opposite O^6-meG in the template strand. The experiment was carried out by preincubating 0.12 μM Klenow fragment with 0.1 μM DNA in the presence of EDTA, and the reaction was initiated by the addition of 80 μM dCTP and Mg^{2+}. Three bands are shown in each track, representing the unelongated primer (9-mer), a 10-mer and an 11-mer respectively as indicated by the labels. This is in contrast to the incorporation of thymine opposite O^6-meG in the template strand under similar conditions which showed no elongation of the primer to form an 11-mer, even though at each time point, the amount of 10-mer accumulated during the incorporation of thymine was significantly higher than the amount of 10-mer accumulated during the incorporation of cytosine. Lanes a to h on the auotoradiogrph are samples taken at 1 s, 2 s, 3 s, 5 s, 7 s, 10 s, 15 s and 30 s respectively.

The pre-steady state incorporation (i.e. incorporation on a pre-existing enzyme-DNA complex) of thymine and of cytosine opposite O^6-meG in the template strand are shown in Figures 3.16 and 3.17 respectively. These experiments were performed with 0.1 μM DNA and 0.12 μM Klenow fragment and 5, 10, 20, 40 and 80 μM dNTP. The amount of primer strand elongated was measured at intervals between 200 ms and 16 seconds. From Figures 3.16 and 3.17, it is obvious that thymine is incorporated at a faster rate than cytosine opposite O^6-meG in the template strand. Under similar conditions, the initial rate of incorporation of cytosine opposite O^6-meG in the template strand was roughly 4.7 times slower (approximately 15% elongated per sec) than that for the incorporation of thymine opposite O^6-meG.

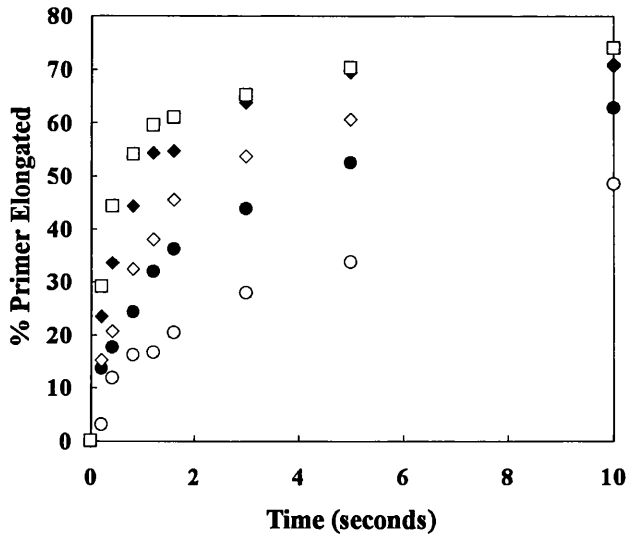


Figure 3.16. Pre-steady state incorporation of thymine opposite O^6-meG in the template DNA strand, using an excess of Klenow fragment over DNA. 0.12 μM Klenow fragment and 0.1 μM DNA duplex A were pre-incubated with EDTA and the reactions were initiated by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◊), 40 μM (♦) and 80 μM (□) dTTP.

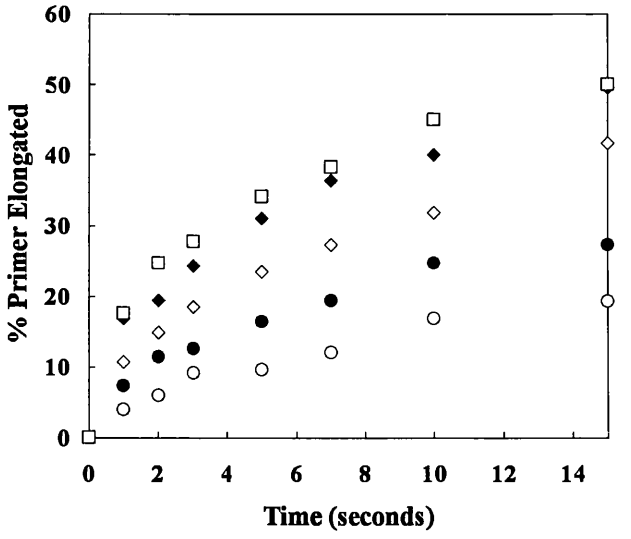


Figure 3.17. Pre-steady state incorporation of cytosine opposite O^6-meG in the template DNA using an excess of Klenow fragment over DNA. 0.12 μM Klenow fragment and 0.1 μM DNA duplex A were pre-incubated with EDTA and the reactions were initiated by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◊), 40 μM (♦) and 80 μM (□) dCTP.

Figures 3.18 and 3.19 show the initial rates of incorporation of thymine and of cytosine opposite O^6-meG in the template strand at 5 μM to 200 μM dNTP. These initial rates were measured at a time when only 15% of the primer was elongated (Table 10). With the incorporation of thymine, the shape of the graph is clearly hyperbolic with a K_m of 30-35 μM. The shape of this graph is in great contrast to that in Figures 3.13(a) and (b), which are also graphs of rates of elongation against the concentration of dTTP. The difference is explained by the way the rates of elongation were obtained. In Figure 3.13(a) and (b), the rates were obtained from steady state experiments which used excess DNA over enzyme (the ratio was 0.109 μM Klenow to 1 μM DNA). As explained previously, under these conditions the rates measured were not the rates of polymerisation but the rates of enzyme dissociating from the product DNA and binding another substrate DNA molecule. By contrast, the reaction rates used for constructing Figures 3.18 are the initial rates of incorporation of thymine into a preformed enzyme-DNA complex, and since these experiments were measured using a slight excess of enzyme over DNA, the reactions monitored did not involve the dissociation of the enzyme from the product DNA.

With the incorporation of cytosine, the graph of initial rate versus [dCTP] is also a hyperbola (Figure 3.19) with an approximate K_m of 30-35 μM. This is of similar magnitude as the approximate K_m obtained for the incorporation of thymine opposite O^6-meG in the template strand, but the maximal rate of incorporation reached for cytosine is nearly 5 times lower than that for thymine. As will be described later, similar values of K_m and V_{max} were also obtained from the rate constants determined by computer analysis of the pre-steady state data shown in Figures 3.16 and 3.17.

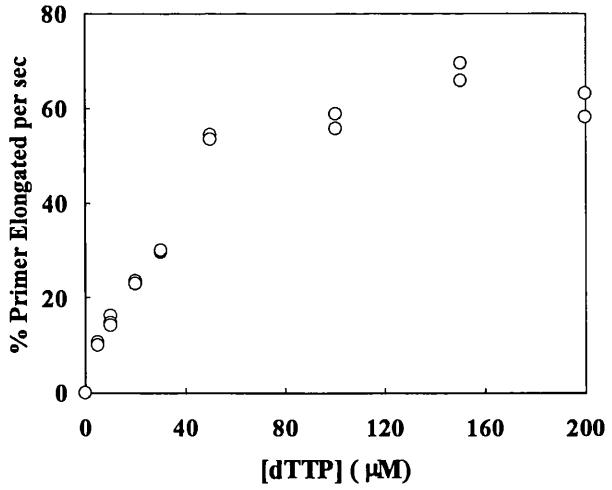


Figure 3.18. Graph showing the initial rates of incorporation of thymine opposite O^6-meG at different concentrations of dTTP. The experiments were carried out with 0.12 μM Klenow fragment, 0.1 μM DNA and 5-200 μM dTTP. Two or three data points are shown at each concentration of dTTP used, as each experiment was performed in duplicate or triplicate.

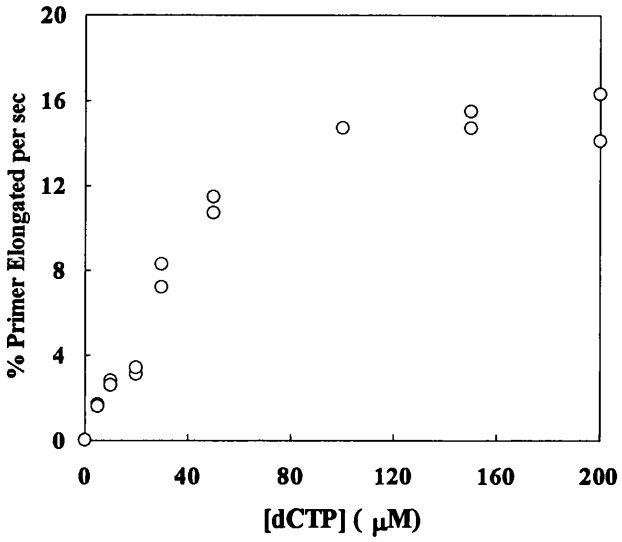


Figure 3.19. Graph showing the initial rates of incorporation of cytosine opposite O^6-meG at different concentrations of dCTP. The experiments were carried out with 0.12 μM Klenow fragment, 0.1 μM DNA and 5-200 μM dCTP. Two data points are shown at each concentration of dCTP used, as each experiment was performed in duplicates.

Besides the above Michaelis-Menten analysis, the data shown in Figures 3.16 and 3.17 were also used to calculate the values of K_m and V_{max} for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. Since the data points from Figures 3.16 and 3.17 were obtained from different experiments, this served as a check for the values of K_m and V_{max}. The initial rates of pre-steady state incorporation at different dNTP concentrations were calculated at 200 ms for dTTP and 1 s for dCTP. A curve-fitting program (Multifit 1.5 for the Macintosh, Day Computing, P.O. Box 327, Milton, Cambridge CB4 4WL, England) was then used to determine K_m and V_{max}. Figure 3.20 shows the Michaelis-Menten analysis of the initial rates of incorporation of thymine and cytosine opposite O^6-meG in the template DNA strand obtained from Figures 3.16 and 3.17. Note that the scales for the % incorporation of thymine (0-80%) is different from that of the % incorporation of cytosine (0-20%). The K_m for the incorporation of thymine and cytosine opposite O^6-meG in the template strand are 27.6 μM and 26.4 μM respectively, while the values of V_{max} for incorporation of thymine and cytosine differ by about 4-fold: 96.7 % s^{-1} and 25.2 % s^{-1} for thymine and cytosine respectively. These results are slightly different to those obtained from Figures 3.18 and 3.19, probably because the initial rates in Figure 3.20 have been underestimated. However, the values obtained are of similar magnitude to those from Figures 3.18 and 3.19, and it is reasonable to conclude that the K_m for the incorporation of thymine is the same as that for the incorporation of cytosine, and that the V_{max} for the incorporation of thymine is ~ 4-5 times that for the incorporation of cytosine.

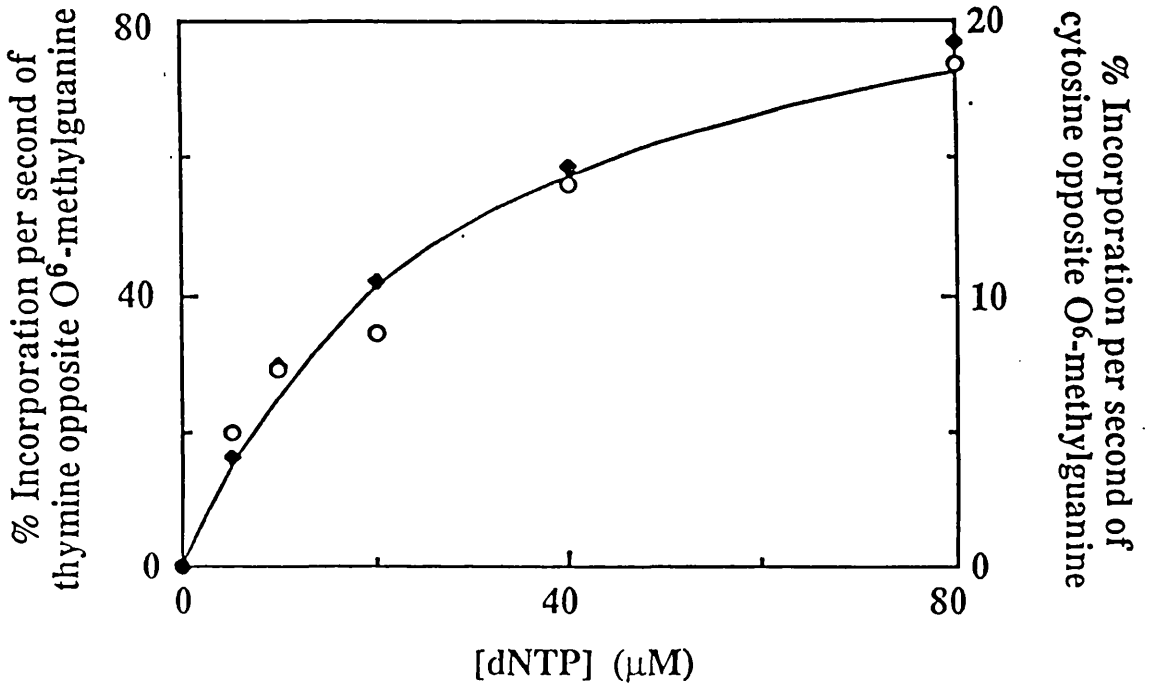


Figure 3.20. Initial rates of incorporation of thymine (o) or cytosine (♦) opposite O^6-meG. Klenow fragment (0.12 μM) and DNA duplex A (0.1 μM) were incubated with 5-80 μM dTTP or dCTP. The reaction was stopped after 200 ms for dTTP and 1 s for dCTP. Note that the scales for dTTP and dCTP differ. The line is a Michaelis-Menten fit to these results implying K_m (dTTP) = 27.6 μM, K_m (dCTP) = 26.4 μM; V_{max} (dTTP) = 96.7 nMs^{-1}, and V_{max} (dCTP) = 25.2 nMs^{-1}.

The pre-steady state experiments mentioned above were carried out with the rapid-quench apparatus that was first built. During the course of the Ph.D. project, a second apparatus, the quench-flow apparatus, was built in order to ease the of performance the experiments. The results of pre-steady state experiments carried out in the quench-flow apparatus were compared with those carried out in the rapid-quench apparatus in Figures 3.21 and 3.22. The results of the experiments carried out on the rapid-quench apparatus were generally very reproducible, even when the experiments were performed on different days. By contrast, the results with the quench-flow apparatus were highly unsatisfactory and totally unreproducible. Thus, all the pre-steady state experiments described in this thesis were carried out with the rapid-quench apparatus.

Figure 3.21. Results of experiments carried out on the quench-flow apparatus. The experiments were carried out with 0.12 μM Klenow, 0.1 μM DNA and 20 μM (o) or 100 μM dTTP (•) over time intervals of 500 ms- 4 s. These experiments were repeated on the rapid-quench apparatus for comparison (Figure 3.22).

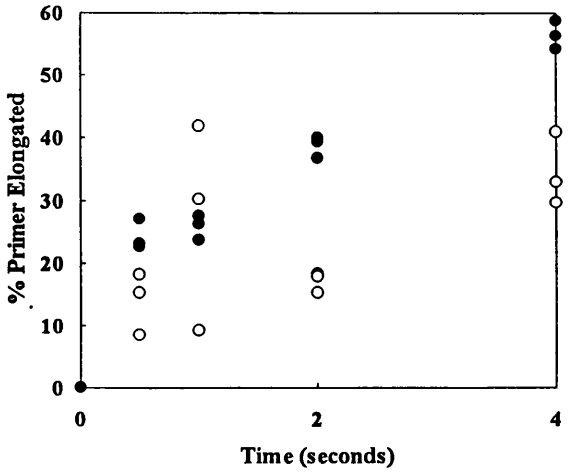
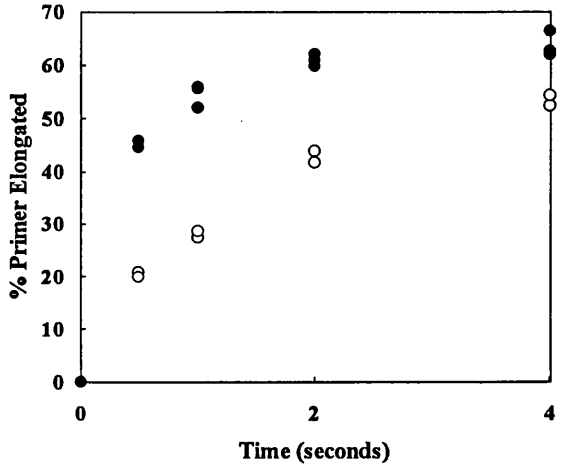


Figure 3.22. Results of experiments carried out on the rapid-quench apparatus. The experiments were carried out with 0.12 μM Klenow, 0.1 μM DNA and 20 μM (o) or 100 μM dTTP (•)over time intervals of 500 ms- 4 s. These results were compared with those of the same experiments performed on the quench-flow apparatus (Figure 3.21).



Pre-steady state polymerisation using phosphorothioate nucleotides. Pre-steady state kinetics of incorporation of the *S*-stereoisomer of the 2'-deoxynucleoside 5'-O-(1-thiotriphosphates), (S_p)-dTTPαS or (S_p)-dCTPαS, opposite O^6-meG in the template DNA strand are shown in Figure 3.23. The rate of incorporation using 40 μM (S_p)-dTTPαS (a) was linear with a very shallow gradient (approx. 0.9% elongated per sec). This rate was approximately 75 times slower than the initial rate of incorporation using the same concentration of the normal dTTP (approximately 70% elongated per sec). Like (S_p)-dTTPαS, the rate of incorporation of 40 μM (S_p)-dCTPαS opposite O^6-meG was also linear over the observation period (18 seconds) and approximately 75 times slower (roughly 0.2% elongated per sec) than the incorporation of the normal cytosine nucleotide opposite O^6-meG in the template strand (Figure 3.23b). Thus, although thymine and cytosine are incorporated opposite O^6-meG in the template strand with different rates, they are affected to the same degree when these normal thymine and cytosine nucleotides are replaced by their respective phosphorothioate nucleotide analogues.

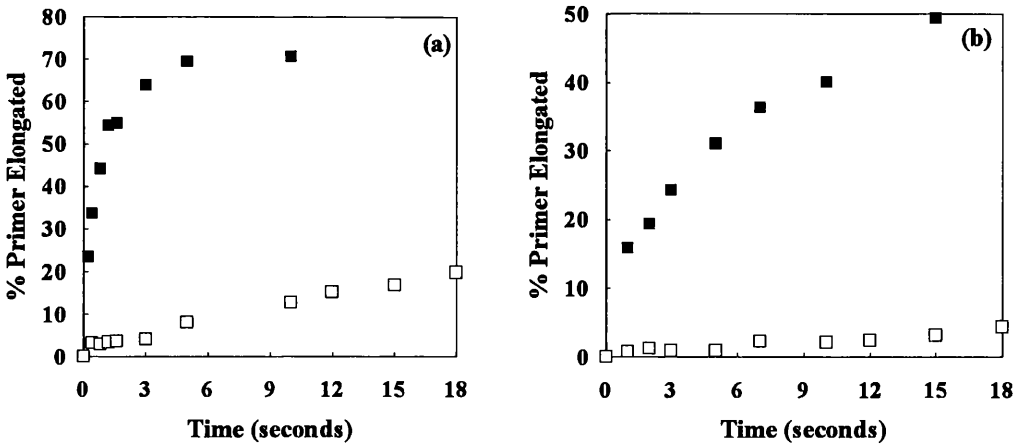


Figure 3.23. Comparison of the incorporation of normal (□) and phosphorothioate nucleotide analogues (▪) opposite O^6-meG in the template strand. Klenow fragment (0.12 μM) and DNA duplex A (0.1 μM) were incubated with 40 μM dTTP or (S_p)-dTTPαS (a) or 40 μM dCTP or (S_p)-dCTPαS (b).

Calculation of rate constants. The rate constants k_4 to k_{10} and k_{13} (Figure 3.5, page 93) for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand were calculated using the computer program FACSIMILE. Details of the calculation are given in the next section (section 3.4), and the results are summarized in Tables 12 and 13. Several sets of calculations were carried out. First, the model of polymerisation was tested to see if the two conformational steps proposed by Kuchta et al. (1988) were necessary for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. This was done by proposing four kinetic schemes:

- the first kinetic scheme did not involve any conformational change;
- the second kinetic scheme had only the first conformational change induced by dNTP binding and which occurs just before phosphodiester bond formation;
- the third kinetic scheme had only the second conformational change which occurs after phosphodiester bond formation; and
- the fourth model had both the first and second conformational changes which occur before and after phosphodiester bond formation respectively.

Each model was then used in computer simulation and parameter-fitting processes to see if the kinetic scheme proposed could be used to generate reaction progress curves that would fit the observed data shown in Figures 3.16 and 3.17. This is explained in greater detail in Section 3.4 of this chapter. Each of the four models was tested in turn, and the model that gave the best fit to

the experimental data was the one which included the two conformational changes before and after phosphodiester bond formation.

Having established that both conformational steps are needed in the kinetic scheme, the rate constants were then calculated from the pre-steady state data shown in Figures 3.16 and 3.17. Two constraints were applied to the solution. The first was to constrain the solution to the thermodynamics of the process, i.e. to constrain it to an overall Gibbs' free energy change of -4 kcal/mol. The justification for the assumed overall free energy change is given in the Section 3.4.

The second constraint was to fix the forward and backward rates for the formation of the phosphodiester bond (i.e. k_7 and k_8) at a value 150 times slower for the incorporation of the phosphorothioate nucleotide analogues than for the incorporation of the normal, oxygen-containing nucleotides. This constraint was based on the observed 75-fold reduction in rate of polymerisation when dNTPαS was used in place of dNTP for incorporation opposite O^6-meG in the template DNA strand (Fig. 3.23). This indicates that the formation of the phosphodiester bond is the rate-limiting step, i.e. the observed 75-fold lower rate of incorporation of dNTPαS than the regular nucleotides was entirely the result of a difference in the formation of the phosphodiester bond. This was deduced from work by Benkovic & Schray (1971) which showed that thio-substitution reduces the rate of reactions with phosphate triesters by 30-100-fold. It should be noted that although the literature suggests that effect of thio-substitution is 30-100, mathematical analysis of the data suggests that a value of 150 would be more appropriate for the effect of thio-substitution on the incorporation of thymine and of cytosine opposite O^6-meG in the template strand.

The results of the rate constants calculated from the application of the two constraints mentioned above are shown in Table 12, and the fit of the computer predicted curves to the observed data is shown in Figures 3.24 and 3.25. ***The only significant difference between the rate constants for incorporation of thymine and of cytosine is a five-fold difference in the rate of the formation of the phosphodiester link.***

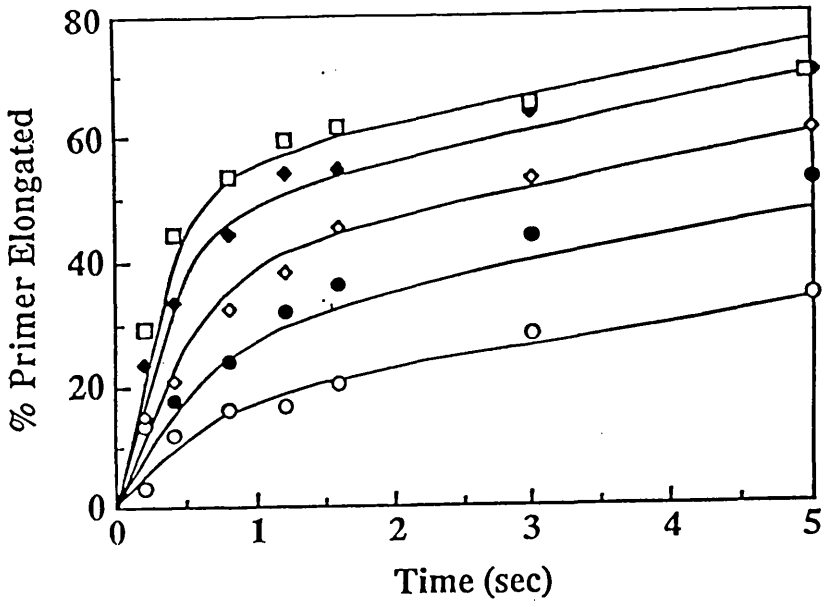


Figure 3.24. Fit of the computer predicted curves to pre-steady state incorporation of thymine opposite O^6-meG in the template strand. The points are the results from experiments performed by pre-incubating 0.12 μM Klenow fragment with 0.1 μM DNA duplex A in the presence of EDTA and initiating the polymerisation reaction by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◇), 40 μM (◆) and 80 μM (□) dTTP. The lines are the computer predicted outcome from the kinetic scheme in Figure 3.5 and the rate constants in Table 12.

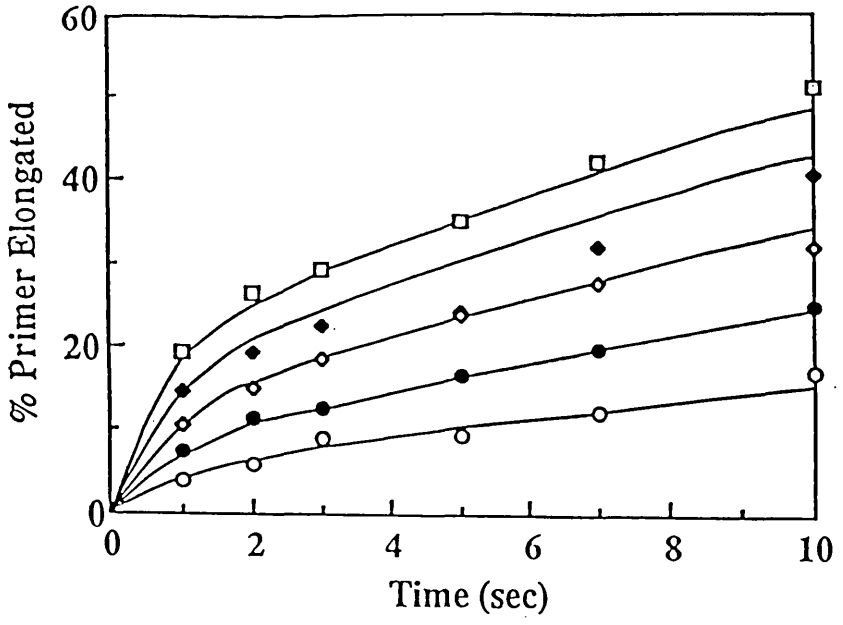


Figure 3.25. Fit of the computer predicted curves to pre-steady state incorporation of cytosine opposite O^6-meG in the template strand. The points are the results from experiments performed by pre-incubating 0.12 μM Klenow fragment with 0.1 μM DNA duplex A in the presence of EDTA and initiating the polymerisation reaction by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◇), 40 μM (◆) and 80 μM (□) dCTP. The lines are the computer predicted outcome from the kinetic scheme in Figure 3.5 and the rate constants in Table 12.

Table 12. Rate and equilibrium/dissociation constants for the incorporation of thymine (column 3) and of cytosine (column 4) opposite O^6-meG in the template strand by Klenow fragment. The values calculated from the pre-steady state data shown in Figures 3.16 and 3.17 were constrained by the assumption that the overall equilibrium constant for polymerisation is 1000, and that thio substitution reduces the forward and backward rate of phosphodiester bond formation by 150-fold. These constraints are discussed in the text. The values calculated by FACSIMILE are accompanied by their 5% and 95% confidence limits.

Reaction	Constants	Thymine (5%-95% limits)		Cytosine (5%-95% limits)		Units
$E + D \rightleftharpoons E.D$	[a] k_1	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	k_2	0.05		0.05		s^{-1}
	[b] k_2/k_1	5		5		nM
$E.D + N \rightleftharpoons E.D.N$	[a] k_3	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	[c] k_4	1000		1000		s^{-1}
	k_3/k_4	100		100		μM
$E.D.N \rightleftharpoons Ei.D.N$	[d] k_5	59.1	(33.1 - 105)	13.1	(3.6 - 47.1)	s^{-1}
	[d] k_6	22.2	(11.7 - 42.2)	3.82	(0.95 - 15.4)	s^{-1}
	[d] k_5/k_6	2.67	(2.26 - 3.15)	2.67	(2.67 - 4.38)	
$Ei.D.N \rightleftharpoons Ei.D1.PPi$	[d] k_7	3.79	(3.43 - 4.18)	0.65	(0.57 - 0.74)	s^{-1}
	[d] k_8	1.71	(1.54 - 1.90)	1.35	(1.11 - 1.65)	s^{-1}
	[d] k_7/k_8	2.21	(1.93 - 2.54)	0.48	(0.39 - 0.59)	
$Ei.D1.PPi \rightleftharpoons E.D1.PPi$	[d] k_9	0.25	(0.21 - 0.31)	0.23	(0.17 - 0.31)	s^{-1}
	[d] k_{10}	0.0015	(0.0013 - 0.0017)	0.00037	(0.00032 - 0.00044)	s^{-1}
	[d] k_9/k_{10}	170	(152 - 189)	610	(473 - 786)	
$E.D1.PPi \rightleftharpoons E.D1 + PPi$	k_{11}	1000		1000		s^{-1}
	[a] k_{12}	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	[b] k_{11}/k_{12}	100		100		μM
$E.D1 \rightleftharpoons E + D1$	[h] k_{13}	0.06	(0.055 - 0.070)	0.03	(0.023 - 0.039)	s^{-1}
	[a] k_{14}	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	[h] k_{13}/k_{14}	6.2	(5.5 - 7.0)	4.1	(3.6 - 5.0)	nM
overall incorporation of nucleotide	[e] K_m	30-35		30-35		μM
	[f] K_m	27.6		26.4		μM
	[g] K_m	30.7	(27.6 - 34.3)	25.5	(21.4 - 30.5)	μM
	[d] V_{max}	323	(307 - 340)	61	(42 - 117)	nMs^{-1}
3' to 5' exonuclease activity	[i] k_{exo}	0.0016		0.0028		s^{-1}

Continued on next page.

Table 12. Continued.

a	The diffusion-limited rate constants were assumed to be $1x10^7$ $M^{-1}s^{-1}$.
b	Data taken from Kuchta et al., 1987.
c	Assumed value, discussed in Section 3.4.
d	Determined by mathematical analysis (FACSIMILE) of the pre-steady state data shown in Figures 3.16 and 3.17.
e	Estimated from Figure 3.18 and 3.19.
f	Calculated by obtaining initial rates from Figures 3.16 and 3.17, and then analysing these initial rates by Mutlifit 1.5, a curve fitting program for the Macintosh.
g	Calculated by FACSIMILE by fitting the integrated form of the Michaelis-Menten equation to the pre-steady state results shown in Figures 3.16 and 3.17.
h	Determined by mathematical analysis (FACSIMILE) of the steady state data shown in Figure 3.11., and using rate constants determined in d and i.
i	Determined from Figure 3.10.

The rate constant for the dissociation of Klenow fragment from the product DNA, i.e. k_{13} in Figure 3.5 (page 93), was determined by FACSIMILE in a separate analysis. The rate constants shown in Table 12 that were determined by the analysis of the pre-steady state data were fixed at the values calculated, and the progress curve was fitted to the data from the steady state incorporation experiments (Figure 3.11) by allowing k_{13} to vary. In this way, the rate of dissociation of Klenow from the product DNA was found to be 0.06 s^{-1} for the product DNA containing a thymine opposite the O^6-meG, and 0.03 s^{-1} for the product DNA containing a cytosine opposite the O^6-meG. The fit of the computer predicted curves to the experimental data is shown in Figure 3.26.

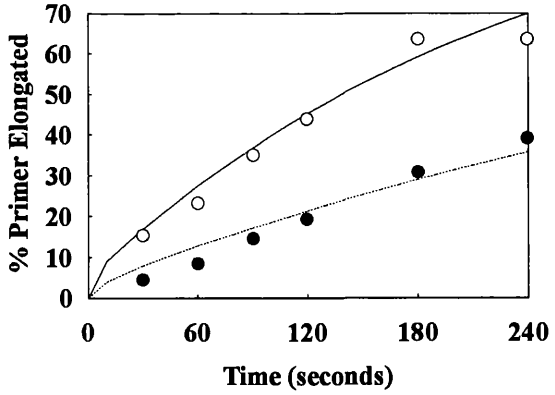


Figure 3.26. Graph showing the fit of the computer predicted curves to steady state incorporation of thymine (○) and of cytosine (●) opposite O^6-meG in the template strand. The points are results from experiments performed by pre-incubating 1 μM Klenow fragment with 0.109 μM DNA duplex and 20 μM dTTP or dCTP. The lines are the computer predicted outcome from the kinetic scheme in Figure 3.5 and the rate constants in Table 12.

Michaelis-Menten analysis of the initial rates of incorporation of thymine and cytosine opposite O^6-methylguanine in the template DNA strand by FACSIMILE. Two independent methods were used to calculate the values of K_m and V_{max} from the pre-steady state data. The first method was to obtain K_m and V_{max} from the measured initial rates of pre-steady state incorporation at different dNTP concentrations (Figure 3.18 and 3.19). The second method was to fit the experimental data directly to the integrated form of the Michaelis-Menten equation using FACSIMILE in order to calculate K_m and V_{max}. This was done by rewriting the Michaelis-Menten equations in terms of the internal rate constants of the polymerisation model, and then using computer analysis and parameter fitting to determine the values of K_m and V_{max}. This is discussed in greater detail in Section 3.4. The results of computer analysis also suggested values of K_m and V_{max} that are very similar to those obtained from Figures 3.18, 3.19 and 3.20 (see Table 12).

Competition assays with dTTP and (S_p)-dTTPαS. The assumption that a large (30-100-fold) reduction in the rate of incorporation with a phosphorothioate nucleotide analogue is indicative of a rate-limiting phosphodiester bond formation step has been previously used by Bryant et al. (1983), Mizrahi et al. (1985), Kuchta et al. (1987, 1988), Patel et al. (1991) and Wong et al. (1991) to solve the mechanistic kinetic schemes of Klenow fragment and of T7 DNA polymerase. However, during the calculations of the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand, a report by Herschlag et al. (1991) was published which challenged the use of this assumption by these authors. The essential aspect of the criticism of Herschlag et al. of the earlier work is that ***thio-substitution reduces the rates of reactions with phosphate diesters by only 4-11-fold***, whereas Bryant et al. (1983) and the other workers who came after them had assumed that thio-substitution reduces the rates of reactions with phosphate diesters by as much as 30-100-fold. Thus, the effect of the dNTPαS on polymerisation had to be re-examined to see if the phosphodiester bond formation is indeed the rate-limiting step during the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. The competition assays with both dTTP and (S_p)-dTTPαS were designed for this purpose, and the results of these experiments are shown in Figure 3.27. The results show that even when large amounts of (S_p)-dTTPαS was added, there was no inhibition of polymerisation, which confirms that the slow incorporation of the (S_p)-dTTPαS observed in the previous experiments could not be due to a slower rate of phosphodiester bond formation alone. Since (S_p)-dTTPαS did not act as a competitive inhibitor in these experiments, it is obvious that the assumption that thio-substitution would only affect the forward and backward steps of the phosphodiester bond formation during polymerisation is incorrect.

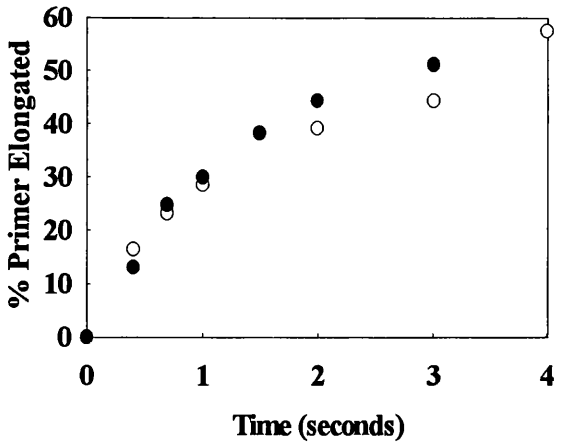


Figure 3.27. Results from competition experiments with dTTP and (S_p)-dTTPαS. 0.1 μM of duplex A was incubated with 0.12 μM Klenow fragment in the presence of EDTA, and the reaction was initiated by the addition of magesium ions and 20 μM dTTP (○) or 20 μM dTTP and 40 μM (S_p)-dTTPαS (●).

Competition assays with thymine and cytosine and recalculation of the rate constants. Figure 3.28 is an autoradiograph of the separation of the products of elongation from the unelongated primer by gel electrophoresis. As mentioned in the Materials and Methods section, the mobility of the 9+T-mer was slightly slower than that of the 9+C-mer in the gel. Although this small difference in the mobilities of the 9+T-mer and the 9+C-mer was insufficient to separate the two products of elongation, one can get a fairly good idea of the relative amounts of each product formed from the autoradiograph. For example, in Figure 3.28, in the lanes marked f and g, one can guess from the intensities of the bands that the amount of the amount of 9+T-mer was roughly five times that of the 9+C-mer, where as in lanes j and k, nearly equal amounts of each product was formed.

The results of the two competition assays with thymine and cytosine are shown as data points in Figures 3.29 and 3.30. These experiments were originally intended to serve as checks for the accuracy of the rate constants shown in Table 12, so the progress curves predicted by the computer are represented by lines in Figures 3.29 and 3.30. These figures clearly show that the experimental results were different from that predicted from the rate constants in Table 12. For the incorporation of cytosine, the predicted results were close to the observed data points, but for the incorporation of thymine, the predicted outcome did not fit the observed data points, particularly in the competition assay with 20 μM dTTP and 80 μM dCTP. The predicted amount of thymine incorporated appears to be underestimated in both competition assays.

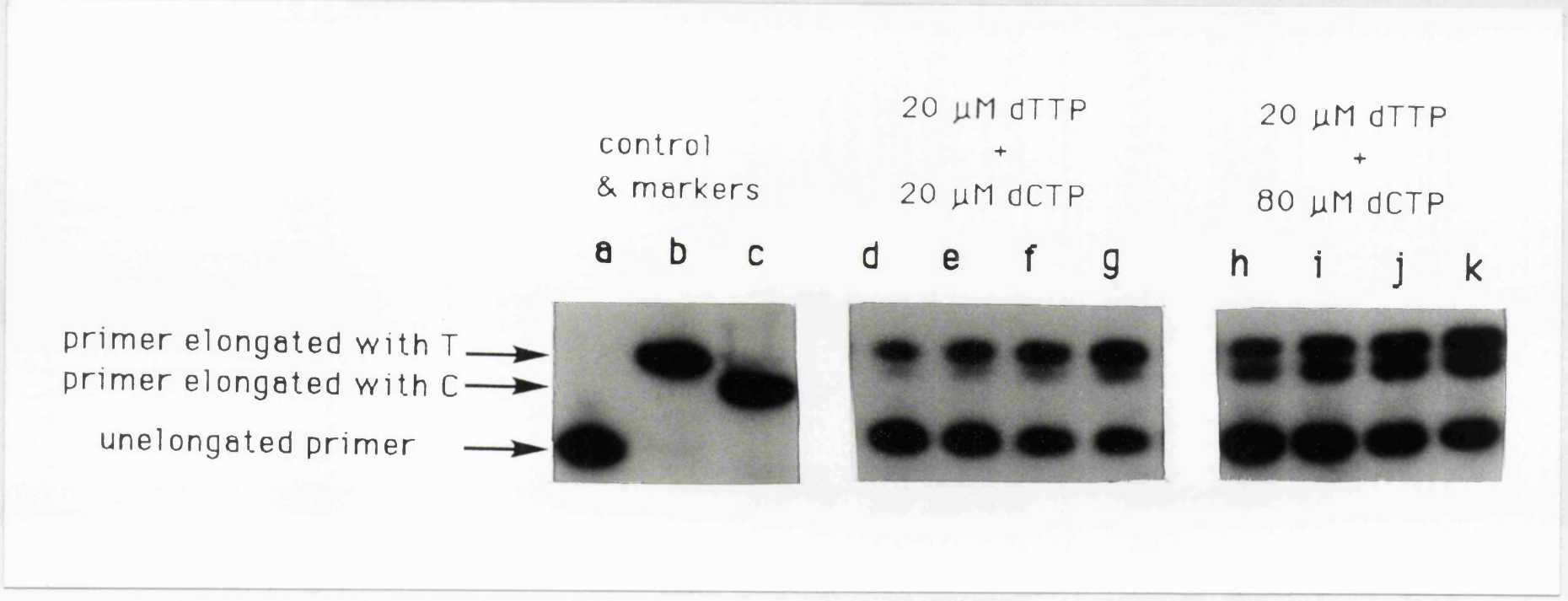


Figure 3.28. Autoradiograph showing the separation of the products of elongation from the unelongated primer by gel electrophoresis. Sections from three gels are shown.

- Starting from the left, lane a is the control (i.e. unelongated primer), while lanes b and c are chemically synthesized markers for the primer elongated with thymine (9+T-mer) and the primer elongated with cytosine (9+C-mer) respectively.
- The gel in the middle contains samples from the competition assay which used 20 µM dTTP and 20 µM dCTP. Lanes d, e, f and g are samples from this competition experiment at 0.5 s, 1 s, 2 s and 4 s respectively.
- The gel on the right contains samples from the competition assay which used 20 µM dTTP and 80 µM dCTP. Lanes h, i, j and k are samples this competition experiment at 0.5 s, 1 s, 2 s and 4 s respectively.

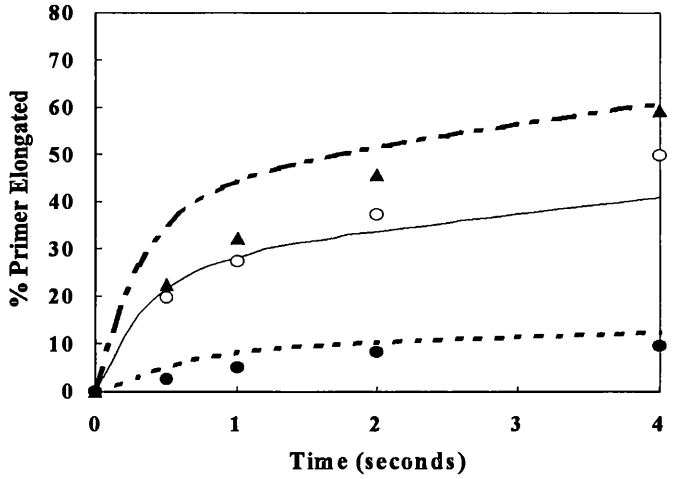


Figure 3.29. Comparison of the experimental results of the thymine/cytosine competition assay (20 μM T + 20 μM C) with that predicted by computer based on the rate constants in Table 12. The experiments were carried out with 0.12 μM Klenow fragment, 0.1 μM DNA, 20 μM dTTP and 20 μM dCTP. The total amount of primer elongated (•) was measured by gel electrophoresis (Figure 3.28). The the 9+T-mer (○) was measured by separation on a Pharmacia Mono-Q column. The amount 9+C-mer (●) was deduced from subtraction of the 9+T-mer from the total amount of primer elongated. The computer predicted results for the amount of 9+T-mer, 9+C-mer and the total amount of primer elongated are represented by —— , - - - and -.-.- respectively.

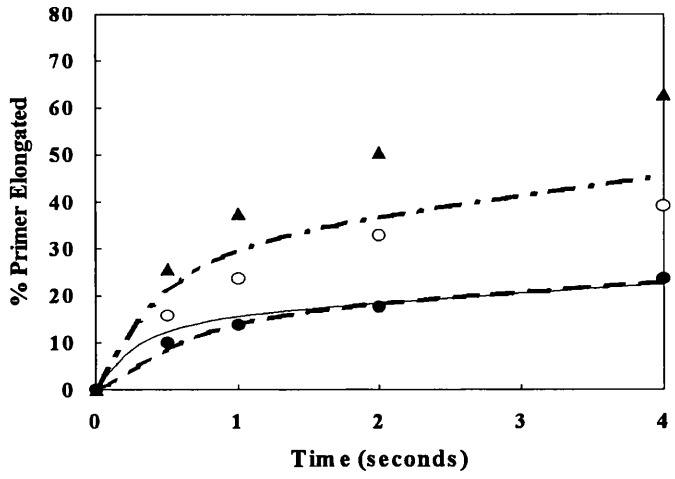


Figure 3.30. Comparison of the experimental results of the thymine/cytosine competition assay (20 μM T + 80 μM C) with that predicted by computer based on the rate constants in Table 12. The experiments were carried out with 0.12 μM Klenow fragment, 0.1 μM DNA, 20 μM dTTP and 80 μM dCTP. The total amount of primer elongated (•) was measured by gel electrophoresis (Figure 3.28). The the 9+T-mer (○) was measured by separation on a Pharmacia Mono-Q column. The amount 9+C-mer (●) was deduced from subtraction of the 9+T-mer from the total amount of primer elongated. The computer predicted results for the amount of 9+T-mer, 9+C-mer and the total amount of primer elongated are represented by —— , - - - and -.-.- respectively.

Since the assumption underlying the use of the phosphorothioate data, i.e. the assumption that thio-substitution affects only the forward and backward rates of phosphodiester bond formation during incorporation of thymine and of cytosine opposite O^6-meG in the template strand, as a constraint for the calculation of the rate constants was incorrect, the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand had to be recalculated. This was done by analysing the experimental data from the thymine/cytosine competition experiments shown Figures 3.29 and 3.30. Details of the analysis are described in Section 3.4. At this point, it was discovered that the reaction scheme shown in Figure 3.5, which produced a satisfactory fit to the observed data points in Figure 3.16 and 3.17 in all the previous calculations, could not fit the results of the competition data at all. Thus, the mechanistic kinetic scheme had to be re-examined, and it was eventually found that a good fit to the thymine/cytosine competition data was possible if a dNTP-exchange step that occurs in the enzyme.DNA.dNTP ternary complex before the first enzyme conformational change was included in the reaction scheme (Step 3, Figure 3.31). In the previous experiments in which only one kind of nucleotide was used (i.e. either T or C was incorporated), this nucleotide exchange step was kinetically invisible because the products of elongation with or without nucleotide-exchange would all be the same. This step could only be observed in the competition assays because the products of elongation were different if a nucleotide exchange had taken place. For example, when a dCTP that was bound to an enzyme-DNA complex was exchanged for a dTTP, this led to the formation of a 9+T-mer instead of a 9+C-mer.

Figure 3.32 shows the fit of the curves predicted according to the reaction scheme shown in Figure 3.5, which does not include a dNTP-exchange step, while Figure 3.33 shows the fit of the curves predicted according to the reaction scheme shown in Figure 3.31 which includes a dNTP-exchange step. The fit of the computer predicted curves for the incorporation of cytosine is good in both figures and is only slightly altered by the inclusion of the dNTP-exchange step in the reaction scheme. The most noticeable improvement in the fit of the computer-predicted curves after the inclusion of the dNTP-exchange step in the reaction scheme is that for the incorporation of thymine in the 20 μM dTTP and 80 μM dCTP competition experiment. The better fit of the computer predicted curves to the data after the inclusion of the dNTP-exchange step in the reaction scheme is seen more clearly in the residual sum of squares of the computer-predicted curves. For the curves predicted according to the reaction scheme without the dNTP-exchange step (Figure 3.32), the total RSQS was 44.9 compared with an expected range of 19.0 to 50.7, while for the curves predicted according to the reaction scheme with the dNTP-exchange step (Figure 3.33), the total RSQS was 33.5 compared with an expected range of 19.0 to 50.7.

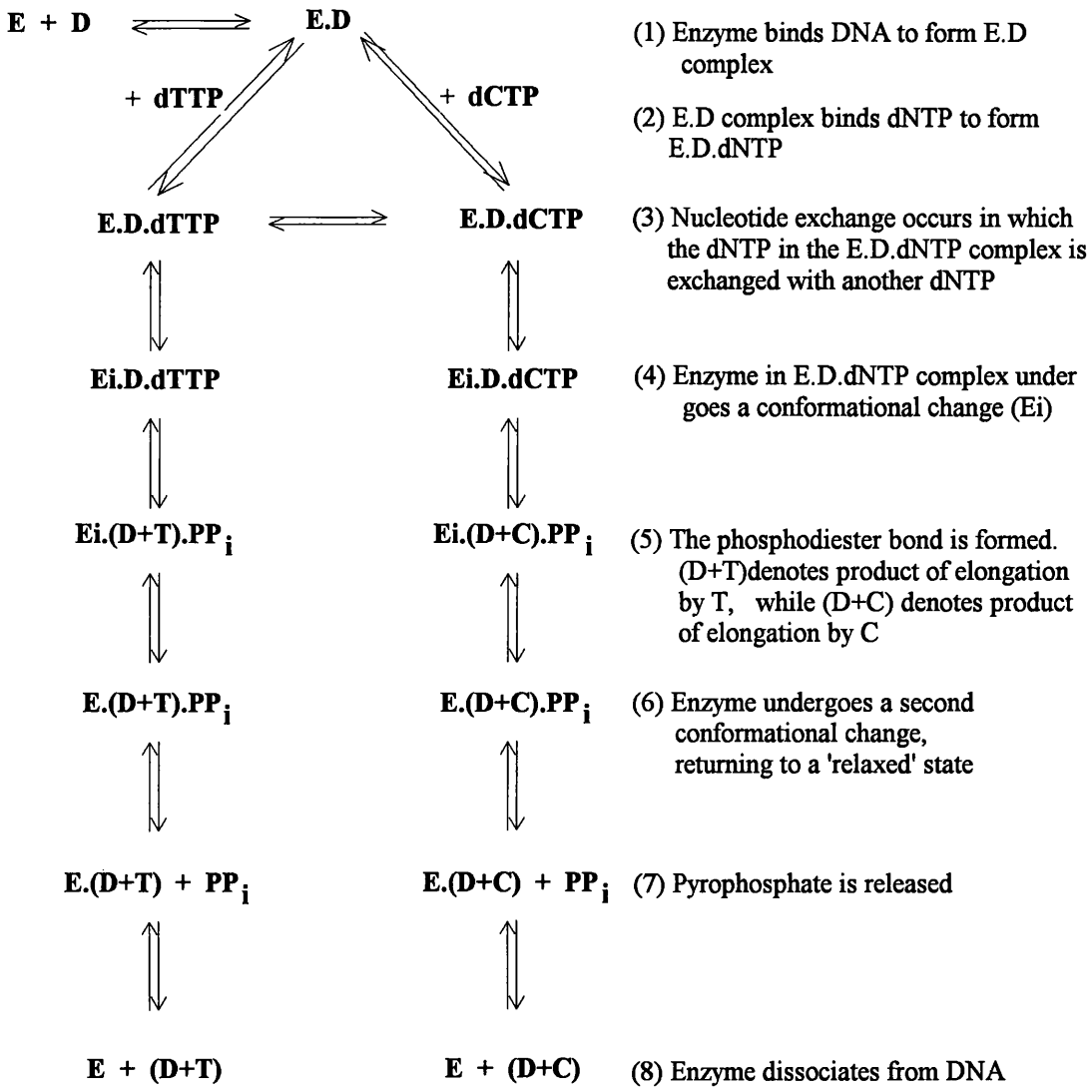


Figure 3.31. Kinetic scheme for the incorporation of thymine and cytosine proposed to fit the data from the thymine/cytosine competition assays. The only difference between this kinetic scheme and that shown in Figure 3.5 is the presence of a nucleotide exchange step in the present scheme (step 3 in the present scheme).

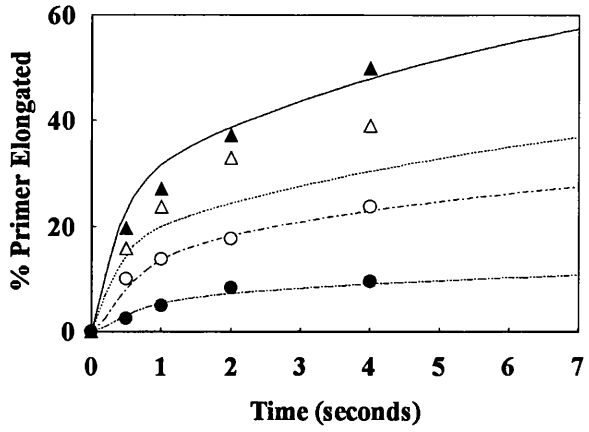


Figure 3.32. Graph showing the fit of computer predicted curves based on the reaction scheme shown in Figure 3.5, which does not include a dNTP-exchange step, to the data from the competition assays. The points are the expeimental results while the lines are the computer predicted curves for the incorporation of thymine and cytosine from two competition assays: —— and –.–.– are the predicted curves fitted to the results for the incorporation of thymine (•) and cytosine (•) from the competition assay with 20 μM dTTP + 20 μM dCTP. –.–.– and – – – are the predicted curves fitted to the results for the incorporation of thymine (Δ) and cytosine (○) from the competition assay with 20 μM dTTP + 80 μM dCTP.

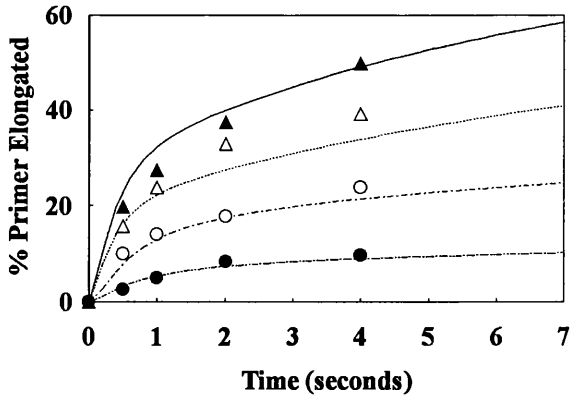


Figure 3.33. Graph showing the fit of computer predicted curves based on the reaction scheme shown in Figure 3.31, which includes a dNTP-exchange step, to the data from the competition assays. The points are the expeimental results while the lines are the computer predicted curves for the incorporation of thymine and cytosine from two competition assays: —— and –.–.– are the predicted curves fitted to the results for the incorporation of thymine (•) and cytosine (•) from the competition assay with 20 μM dTTP + 20 μM dCTP. –.–.– and – – – are the predicted curves fitted to the results for the incorporation of thymine (Δ) and cytosine (○) from the competition assay with 20 μM dTTP + 80 μM dCTP.

Using the new reaction scheme shown Figure 3.31, the rate constants for the incorporation of thymine and of cytosine were then recalculated. Data from the single nucleotide and thymine/cytosine competition experiments (Figures 3.16, 3.17 and 3.33) were analysed in a single program in order to recalculate the rate constants. Since there were many more data points from the single-nucleotide experiments than that from the competition assays, only some of the data points from the single-nucleotide experiments were included in the mathematical analysis. The rate of dissociation of Klenow from the product DNA was calculated by fixing the rate constants calculated from the analysis of the data from single nucleotide and thymine/cytosine competition experiments at the values calculated, and then fitting the progress curves to the steady state incorporation of dTTP or dCTP (Figure 3.11). The results are shown in Table 13, and the fit of the computer predicted curves to the experimental data is shown in Figures 3.33 (competition experiments) and 3.34 (single nucleotide experiments).

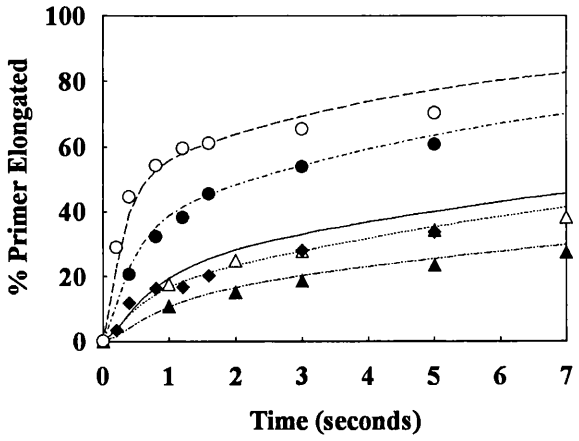


Figure 3.34. Graph showing the fit of computer predicted curves based on the reaction scheme shown in Figure 3.31 (with dNTP-exchange) to the data from single nucleotide incorporation experiments using 0.1 μM Duplex A and 0.12 μM Klenow. The points are the expeimental results for the incorporation of 80 μM dTTP (○), 20 μM dTTP (●), 5 μM dTTP (Δ), 80 μM dCTP (♦) and 20 μM dCTP (⋄). The lines are the computer predicted curves based on the reaction scheme in Figure 3.31 and the rate constants in Table 13: – – – for 80 μM dTTP, –.–.– for 20 μM dTTP, ····· for 5 μM dTTP, ___ for 80 μM dCTP and –..–..– for 20μM dCTP.

Table 13. Rate and equilibrium/dissociation constants for the incorporation of thymine (column 3) and of cytosine (column 4) opposite O^6-meG in the template strand by Klenow fragment. Data from the single-nucleotide experiments (Figures 3.16 and 3.17) and the thymine/cytosine competition experiments (Figure 3.33) were analysed as a single set. The values calculated by FACSIMILE were constrained by the assumption that the overall equilibrium constant for polymerisation is 1000, and that the K_m for the incorporation of thymine is the equal to that for cytosine. These constraints are discussed in Section 3.4. The values calculated by FACSIMILE are accompanied by their 5% and 95% confidence limits.

Reaction	Constants	Thymine (5%-95% limits)		Cytosine (5%-95% limits)		Units
$E + D \rightleftharpoons E.D$	[a] k_1	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	k_2	0.05		0.05		s^{-1}
	[b] k_2/k_1	5		5		nM
$E.D + N \rightleftharpoons E.D.N$	[a] k_3	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	[c] k_4	1000		1000		s^{-1}
	k_3/k_4	100		100		μM
$E.D.N \rightleftharpoons Ei.D.N$	[d] k_5	24.3	(5.06 - 116)	10.5	(1.45 - 76.1)	s^{-1}
	[d] k_6	8.21	(1.12 - 60.1)	4.55	(0.54 - 38.1)	s^{-1}
	k_5/k_6	2.96		2.31		
$Ei.D.N \rightleftharpoons Ei.D1.PPi$	[d] k_7	3.87	(2.08 - 7.21)	0.73	(0.35 - 1.54)	s^{-1}
	[d] k_8	1.89	(0.67 - 5.33)	0.92	(0.24 - 3.70)	s^{-1}
	k_7/k_8	2.05		0.79		
$Ei.D1.PPi \rightleftharpoons E.D1.PPi$	[d] k_9	0.35	(0.21 - 0.59)	0.24	(0.10 - 0.59)	s^{-1}
	[d] k_{10}	0.0021	(0.0016 - 0.0028)	0.00042	(0.00028 - 0.00066)	s^{-1}
	k_9/k_{10}	167		548		
$E.D1.PPi \rightleftharpoons E.D1 + PPi$	k_{11}	1000		1000		s^{-1}
	[a] k_{12}	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	[b] k_{11}/k_{12}	100		100		μM
$E.D1 \rightleftharpoons E + D1$	[h] k_{13}	0.06	(0.055 - 0.070)	0.03	(0.021 - 0.036)	s^{-1}
	[a] k_{14}	$1x10^7$		$1x10^7$		$M^{-1}s^{-1}$
	k_{13}/k_{14}	6.2	(5.5 - 7.0)	4.1	(3.6 - 5.0)	nM
$E.D.dCTP + dTTP \rightleftharpoons E.dTTP + dCTP$	[d] k_{15}	$1.9x10^7$	($1.3x10^7$-$2.8x10^7$)			s^{-1}
	k_{16}	$1x10^7$				s^{-1}
	[d] k_{15}/k_{16}	1.91	(1.28 - 2.84)			
overall incorporation of nucleotide	[e] K_m	30-35		30-35		μM
	[f] K_m	27.6		26.4		μM
	[g] K_m	33.5	(24.1 - 46.5)	33.5	(24.1 - 46.5)	μM
	[d] V_{max}	310	(191 - 503)	58	(30.5 - 113)	nMs^{-1}
3' to 5' exonuclease activity	[i] k_{exo}	0.0016		0.0028		s^{-1}

Continued next page.

Table 13. Continued.

a	The diffusion-limited rate constants were assumed to be $1x10^7$ $M^{-1}s^{-1}$.
b	Data taken from Kuchta et al., 1987.
c	Assumed value, discussed in Section 3.4.
d	Determined by mathematical analysis (FACSIMILE) of the data from single-nucleotide and thymine/cytosine competition experiments, shown in Figures 3.16, 3.17, and 3.33.
e	Estimated from Figure 3.18 and 3.19.
f	Calculated by obtaining initial rates from Figures 3.16 and 3.17, and then analysing these initial rates by Mutlifit 1.5, a curve fitting program for the Macintosh.
g	Calculated by FACSIMILE by fitting the integrated form of the Michaelis-Menten equation to the pre-steady state results shown in Figures 3.16 and 3.17.
h	Determined by mathematical analysis (FACSIMILE) of the steady state data shown in Figure 3.11., and using rate constants determined in d and i.
i	Determined from Figure 3.10.

Addition of the next correct nucleotide after a thymine or a cytosine base paired with O^6-methylguanine in the template strand. Figure 3.35 shows the steady state rates of adding the next correct nucleotide after a T:O^6-meG, a C:O^6-meG and a T:A base pair. The steady state rate of addition after a T:O^6-meG base pair is very similar to that of addition after a T:A base pair, whereas addition after a C:O^6-meG base pair is considerably slower.

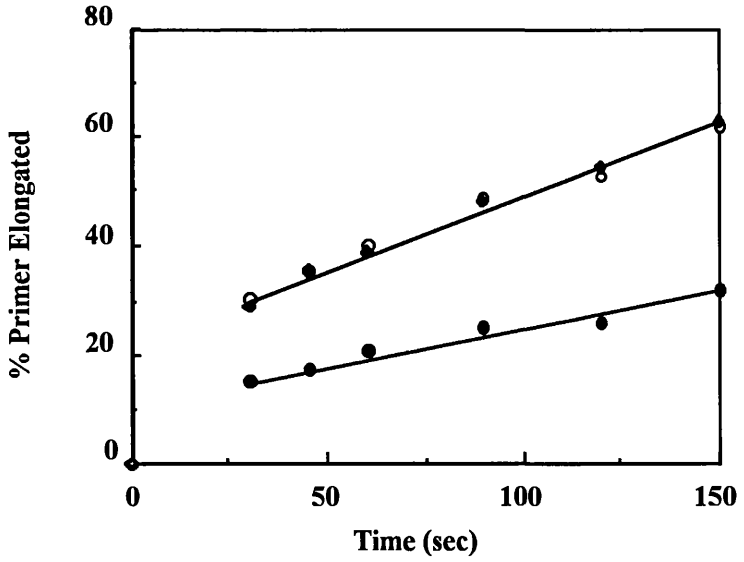


Figure 3.35: Addition of a correct nucleotide following the thymine in an O^6-meG:T pair (o), the cytosine in an O^6-meG:C pair (•), or the thymine in an A:T (♦) pair which was included as control. Klenow fragment (20 nM) was incubated with 1 μM DNA duplex B, C or D and either 50 μM dATP (duplexes B or C) or 50 μM dCTP (duplex D).

3.4 Mathematical analysis of the experimental results

Setting up FACSIMILE

The computer analysis of the experimental data was performed using a computer program called FACSIMILE, developed by the United Kingdom Atomic Energy Authority at Harwell. When the analysis was first begun, two versions of the program were available. One version ran on an IBM personal computer, and the other ran on a mainframe computer, and at that time it was adapted to the AMDHAL 5890 computer in Manchester, running the IBM Conversational Monitor System (CMS). The mainframe was accessed via the Joint Academic Network (JANET), which links the Universities in the UK so as to facilitate the exchange of data. Both of these versions were designed for computer experts and neither of them was user-friendly. The instruction manual is effectively incomprehensible to the non-mathematician and presented great problems, and a substantial amount of time was spent on writing and debugging programs to solve the kinetic problem.

The analysis was initially carried out on the mainframe computer. This had the advantage that the execution time was very short. However, access to the mainframe was very difficult, and the file editor used for writing/editing programs was very unsatisfactory. Thus, the mainframe version of FACSIMILE was replaced by the PC version. At that time, the PC version was written for use on an IBM XT machine which uses the Intel 8088 microprocessor. The Intel 8088 microprocessor is a second generation microprocessor which recognises 16-bit words and which has an internal clock running at 6 MHz. The PC used for the analysis of the data was an IBM PS2 model 30/286; this uses an Intel 80286 processor, which also recognises 16-bit words but has a faster internal clock running at 10 MHz.

At this stage, severe problems arose because of the limited memory capacity of the IBM PS2 30/286 model, with only secondary storage (i.e. storage for the hard-disk), 1 Mbyte RAM and no cache. The operating system (which is the main communications medium between the user and the hardware) used at the time was DOS 3, which can only address 640 kbytes of RAM, part of which is taken up by DOS 3 itself. This leaves very little space for FACSIMILE to run. FACSIMILE itself is a big program, but its storage requirement does not end here. FACSIMILE is written in FORTRAN 77, and so it requires a FORTRAN 77 compiler for execution.

Obviously, after accommodating DOS 3 and the FORTRAN 77 compiler in the RAM there is insufficient memory left to hold the entire FACSIMILE program. The capacity of RAM

had to be expanded, and this was done with the aid of a piece of software known as a memory manager, which is designed to obtain the best possible use of RAM (and cache, if installed). Some of the secondary memory is made to behave like RAM, so that large chunks of programs normally resident in RAM such as DOS 3 and compilers can be stored in such expanded memory, leaving more RAM memory available for the execution of application programs. However, despite this improved memory handling, there was still insufficient RAM to store the whole FACSIMILE program, and parts of it had to be stored in the hard disk, and the program then worked by cycling bits of the program and data to and from the hard disk.

The combination of the slowness of the Intel 80286 processor and the memory limitation had severe effects on the speed of calculations. A single calculation could take anything between an hour to more than five hours to complete. This did not appear a serious problem at first because it was thought that only a few runs would be required to analyse the data. However, this view soon changed as it was realised that writing a program to analyse the data is an evolutionary process, and at least a hundred runs were necessary to get the program to work. The processing and storage limitations of the computer therefore became a big handicap. The long calculation times proved to be an insurmountable difficulty in carrying out this evolutionary task because the slightest error in coding, for example, missed or incorrect punctuation, might not come to light until the calculation was complete or crashed several hours after it has been executed.

At this point, a return to the mainframe computer was considered. Unfortunately this became impossible because the mainframe at Manchester had been changed to a UNIX system, and FACSIMILE could not run on this system. A great deal of time and effort was spent (Dr. E. M. Chance) trying to adapt FACSIMILE to the UNIX system, but that adaptation was never completed. Eventually, one of the original authors of FACSIMILE, Dr. A. R. Curtis, revised it to run on the Intel 80386 or 80486 processor. In order to use this new PC version of FACSIMILE, the existing IBM PS2 30/286 model was upgraded with a 386 SX board (Kingston SX Now, from Kingston Technology, Corporation, Fountain Valley, CA 92708, USA). The Kingston SX 386 board has a 16-bit 25 MHz microprocessor chip that is equivalent to the Intel 80386 chip and a 16 kbyte memory cache. At the same time, a 25 MHz Kingston 80387 SX maths co-processor was installed, and the RAM was upgraded from 1 Mbyte to 4 Mbytes by replacing the four 256 kbytes memory chips originally present in the computer with four 1 Mbyte memory chips. With an improved RAM capacity, a faster processor (80/386 running at 25 MHz) working in conjunction with a 25 MHz maths co-processor, and a 16 kbyte cache, the calculation time was dramatically decreased. What used to take hours to run now requires only a few minutes. In practice, the cache makes a great difference to the run time of FACSIMILE. For instance, a fairly simple calculation which took 389 seconds of CPU time to

complete without the cache, took merely 143 seconds when the cache was installed. Another advantage that came with the upgrade to an IBM 386-equivalent was that the machine became compatible with DOS 5, an updated version of the DOS 3 operating system. DOS 5 has a file editor that is far superior to that was originally used. In practical terms, this means that writing and editing the programs became much easier and faster.

Although setting up a PC to run FACSIMILE is now relatively easy, writing a program to solve an experimental model using FACSIMILE still remains a difficult task for the non-computer expert. FACSIMILE is written in FORTRAN 77, so the investigator has to write the whole analysis program in FORTRAN 77, from declaration of the parameters, defining the model used, setting up arrays for the experimentally derived data, setting up work space for calculations, telling the computer when to execute the program and how to analyse the data (e.g. which data to include in the analysis, what error margin to allocate, and setting the maximum number of reiterations and etc.), to perform a simulation based on the calculated results and how to display the output of the analysis (in graphical presentation or in a table and etc.). This is illustrated in the final part of this section, where details of how the solution to the problem of incorporation by Klenow fragment of thymine and of cytosine opposite O^6-meG in the template DNA was arrived are described. But before this is presented, it is necessary to understand how FACSIMILE works, and this is described in the following paragraphs.

What FACSIMILE does

The aim of the mathematical analysis of the experimental data is to determine all or most of the rate constants for all the steps involved in the polymerisation of thymine and of cytosine opposite O^6-meG in the template strand. There are two ways of solving rate equations. The first is to solve the rate equations algebraically. For example, for a simple first-order reaction A $\rightarrow$ B, the differential rate equation is $d[A]/dt = -k[A]$ where k is the rate constant. Upon integration, this becomes $[A] = [A_0]e^{(-kt)}$, and a value for k can be solved by measuring [A] at different reaction times (t). However, this approach becomes very difficult with a multistep reaction. Consider a slightly more complex reaction scheme involving 2 steps,

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} AB \underset{k_4}{\overset{k_3}{\rightleftharpoons}} C$$

There are four rate constants, and a solution to all four rate constants would require solving the four simultaneous differential rate equations that govern this reaction:

$$d[A]/dt = k_2[A.B] - k_1[A].[B],$$
$$d[B]/dt = k_2[A.B] - k_1[A].[B],$$
$$d[A.B]/dt = k_1[A].[B] - (k_2 + k_3)[A.B] + k_4[C], \quad \text{and}$$
$$d[C]/dt = k_3[A.B] - k_4[C].$$

To solve these equations would require measurement of rates of change of substrate, intermediate and product, which may be very difficult especially if the transient intermediates are difficult to detect. A multistep reaction scheme involving more substrates and intermediates would be more difficult to solve by this approach. The sets of simultaneous rate equations can sometimes be simplified by making certain assumptions, but the assumptions have to be very carefully considered. For example, as pointed out before, Boosalis et al. (1987) made the assumption that the dissociation of polymerase from DNA during steady state polymerisation is reversible only for the substrate DNA, but not for product DNA in order to reduce the number of unknown rate constants in their calculations. This assumption was not substantiated and from the experiments on Klenow fragment and T7 DNA polymerase, it is clear that this assumption is not valid. In addition, one of the most common errors in the algebraic approach of solving rate equations is the occurrence of mistakes in the rate equation itself. As noted by Davis (1992), "kinetic equations are notorious for the number of typographical and algebraic errors that appear in published material", and one has to ensure that the rate equation used is correct before proceeding with further calculations.

An alternative approach to solutions of kinetic problems is to integrate the rate equations numerically. This is particularly useful in cases where the kinetic scheme is unknown and involves many substrates, and has become more popular with the availability of the appropriate computer software. The principle of numerical integration can be illustrated by considering the reversible reaction:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$$

The differential rate equation for this reaction is

$$\frac{d[A]}{dt} = k_2.[C] - k_1.[A][B].$$

For a very small time interval Δt, the following approximation is true:

$$d[A].dt \cong \Delta[A].\Delta t$$

Thus, we can write $\Delta[A] = k_2[C].\Delta t - k_1[A][B].\Delta t$

If this is repeated many times over different Δt periods, then a profile of the concentration of A over time t can be obtained by summing the $\Delta[A]$s:

$$[A] = [A_0] + \sum_{t=0}^{t} \Delta[A]$$

$$= [A_0] + \sum_{t=0}^{t} (k_2[C]\Delta t - k_1[A][B]\Delta t)$$

where $[A_0]$ is the initial concentration of A. This method of numerical integration is known as Euler's method. Using the law of conservation of mass, we can write the following equations for B and C :

$$[B] = [B_0] + \sum_{t=0}^{t} \Delta[A] \quad \text{and} \quad [C] = [C_0] - \sum_{t=0}^{t} \Delta[A]$$

By putting in values for k_1 and k_2, we can then work out a profile for the concentration of A, B and C with time, and if we then compare these calculated profiles with those experimentally determined, we are in a position to see how close the values of k_1 and k_2 chosen for that particular calculation are to the actual values. Thus, ***the principle of applying numerical integration to solve kinetic schemes is to simulate the profile of the reactants, intermediates and products over the experimental time course and compare this with the observed data. By changing the values of the rate constants, the profile of the simulated curve is changed. The simulated curves are optimized in repeated simulations by using least square fit analysis until the difference between the simulated and the experimental curves reaches a minimum.*** For a simple reaction scheme this can be done quite easily, but when a complex reaction scheme is encountered, then there will be many rate laws and their integration becomes exceedingly tedious. This is where the power of the computer comes in. With the help of a computer, even the most complex rate laws can be integrated numerically.

The accuracy of numerical integration depends on the integration interval Δt; the smaller the integration interval the better the approximation. This is particularly important in reactions where the reaction rates vary markedly over a small time interval. Such reactions are described as "stiff" systems, and in general they occur when the values of the rate constants are very different in magnitude. For example, the intermediate in a reaction could be formed very quickly but is converted into product slowly, so the system is very stiff initially because of the rapid formation of the intermediate (Barshop et al., 1983). To solve differential equations that are very stiff would mean that very small integration intervals have to be used in Euler's method, and this poses a great disadvantage because the large number of calculations with very small Δt would take up long computer execution time. FACSIMILE overcomes the problem of stiffness by using Gear's method where the integration interval is adjusted according to the stiffness: a small Δt is used for integration when the system is stiff, but as the stiffness decreases, the Δt is increased accordingly in order to save computing time (Chance & Curtis, 1970). The reader is referred to the paper by Chance & Curtis for details of this method.

Like many other programs that fit mathematical models to experimental data, FACSIMILE uses non-linear regression analysis to do this. Regression analysis uses the least

square criterion. The residual sum of squares of the errors (RSQS, i.e. difference between the calculated and observed value) is first calculated:

$$RSQS = \Sigma error^2 = \Sigma (calculated - observed)^2$$

The smaller the RSQS, the better the fit. The line of best fit or the regression line is the line for which the RSQS is minimum, and therefore ***the RSQS is minimised in order to find the equation that best fits the given data.*** How is the RSQS minimised ? For every rate equation integrated, the RSQS is calculated; then a change is made to the rate equation and the process repeated. In FACSIMILE, the change that is made is the natural logarithm of the specified parameter, which must therefore have a positive value. If a positive initial estimate is not given, a default value of 1 is assigned. If the RSQS is smaller than the previous one, than there is an improvement of the fit. If however the RSQS gets bigger, then the direction of the change is reversed to minimise the RSQS. The process is repeated until the RSQS reaches a minimum, which indicates that the corresponding rate equation is the one that best fits the data. As pointed out by Smith (1992), a helpful way of visualizing non-linear regression analysis is to consider a topographical analogy of a two-parameter model: if we plot the RSQS against all the parameter values, then the lowest point of the resulting response surface represents the minimum RSQS (Fig.3.36); the aim of non-linear regression analysis is to find this minimum, and the process is sometimes described as a convergence to the minimum.

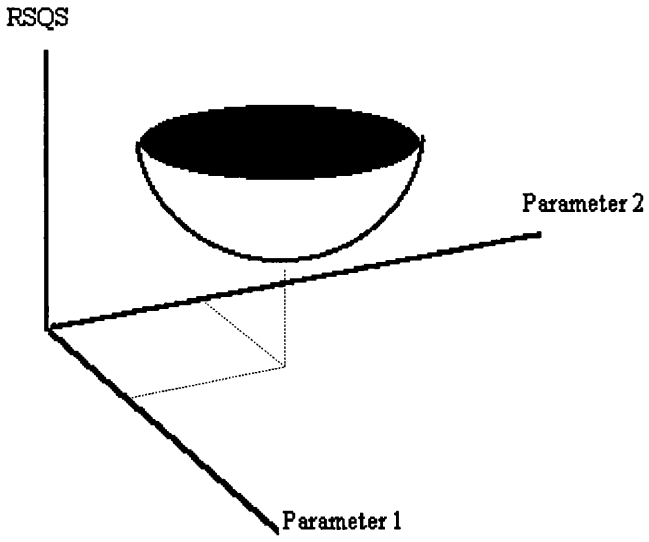


Figure 3.36. A topological analogy of non-linear regression showing convergence to the minimum (taken from Smith, K. in Laboratory Equipment Digest, February 1992).

The model in Figure 3.36 is a two-parameter model. A model involving many more parameters would be much more complicated, and there is always the possibility that more than one minimum exists in such a complex model. Under these circumstances, application of suitable

constraints will help in deciding which minimum is the true minimum for the system. This is discussed later.

How is the RSQS minimised in a complicated system where many rate constants need to be optimized? In FACSIMILE, the organisation of the parameter fitting process is divided into two stages. In the first stage, simulation runs are carried out with different parameter values to find out the minimum RSQS, but only one parameter is varied in each simulation. The first run uses the initial estimates of the parameters provided by the user. In the next p runs (where p = number of independent parameters to be optimized) each of the parameters is varied in turn, and a sensitivity matrix is constructed which relates the sensitivity of the RSQS to the parameter being varied. If a change in a particular parameter lowers the RSQS, then the initial estimate for that parameter is replaced by the parameter value which produces the smaller RSQS. Then iterative minimisation of the RSQS is carried out using and updating the sensitivity matrix (FACSIMILE User's Manual, 1988).

The second stage of FACSIMILE's parameter fitting process begins only after the non-linear regression analysis shows no further convergence to the minimum. In the first stage, the sensitivity matrix is built from simulations with initial estimates of the parameters. In the second stage, the aim is to recalculate the sensitivity matrix using the best parameter values obtained in the first stage. Thus, another (p +1) simulations are carried out but this time the best parameter values from stage 1 are used, and one parameter is varied at a time.

Statistical tests to determine the goodness of fit are also carried out in the second stage of parameter fitting. Although the aim in the analysis of the experimental data is to obtain values for the rate constants, it is equally important to know how accurate the determined values are. After performing the optimization, FACSIMILE carries out a series of statistical analyses to see how well the computed solution fits the experimental data. First, FACSIMILE can estimate a range within which the RSQS should fall if the proposed model is correct and if the residuals are due to random experimental errors. It also calculates the following indices:

- CORI: a correlation index of the residuals. One can tell if the predicted curve has a good fit to the observed data points by looking at the distribution of the data points relative to the predicted curve: a curve which fits the data points well should have the data points evenly distributed above and below the calculated values. The value of |CORI| gives an indication of the distribution of the data points, and |CORI| should be close to 1 if the residuals are unbiased. If |CORI| is much bigger than 1, then the data points are biased one way.
- AUCO: an auto-correlation index. This indicates whether the residuals are random; if they are, then AUCO will be similar to AUCR.
- AUCR: the expected value of AUCO if the residuals are random.

Details of the statistical analyses can be found in FACSIMILE/CHEKMAT User's Manual (1988).

If the RSQS falls within the expected range, and if the CORI, AUCO and AUCR are acceptable, then FACSIMILE varies all the parameter values by a small amount to see how this affects the RSQS. Another sensitivity matrix, called the dependence matrix, is calculated which relates the sensitivity of each RSQS to the parameter being varied, and the singular values (analogous to eigen values) in this dependence matrix indicate if the parameters have been well-determined. If all the singular values are large, then all the parameters have been well-determined. In this case, FACSIMILE then proceeds with the calculation of the confidence limits of each calculated parameter from the singular values of the dependence matrix and the error estimates for the data values. 5% and 95% confidence limits are reported, which means that there is a 5% probability that the true value lies below the lower limit, and a 95% probability that it lies below the upper limit. A parameter that is well-determined by the data would have 5% and 95% confidence limits that are very close to each other.

If one or more small singular values are found in the dependence matrix, this indicates that one or more of the parameters have not been well determined. Under these circumstances, FACSIMILE then identifies the parameters that are not well determined and fixes them at the values last used in the calculations. The sequence of the parameters is rearranged so that the most accurate parameter comes first and the most ill-determined parameter comes last, and another fitting process is carried out to see if any improvements can be made to the fitting by using a reduced number of parameters.

Mathematical analyses of the pre-steady state data

Having briefly described what FACSIMILE does, one can now describe how the mathematical analysis of the experimental data for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand was carried out. The fitting of the curves to the data for incorporation of thymine will be used to provide an illustration. Fitting of the curves to the data for incorporation of cytosine was done in a similar way as that for the thymine.

The first attempts to analyse the experimental results were made with the data for the incorporation of thymine opposite O^6-meG in the template strand. These were the results shown in Figures 3.16. Rate equations for the steps in the reaction were written. Initially, the most simple kinetic scheme was used (Figure 3.37) which included five steps. Each step was taken to be reversible and assigned separate rate constants, for example, the forward and backward rate

constants for step 1 were denoted by k_1 and k_2 respectively, the forward and backward rate constants for step 2 were denoted by k_3 and k_4 respectively and so on.

E + D	$\underset{k_2}{\overset{k_1}{\rightleftharpoons}}$	**E.D**	(1) Enzyme binds DNA to form E.D complex
E.D + dNTP	$\underset{k_4}{\overset{k_3}{\rightleftharpoons}}$	**E.D.dNTP**	(2) E.D complex binds dNTP to form E.D.dNTP
E.D.dNTP	$\underset{k_6}{\overset{k_5}{\rightleftharpoons}}$	**$E.D_1.PP_i$**	(3) The phosphodiester bond is formed.
$E.D_1.PP_i$	$\underset{k_8}{\overset{k_7}{\rightleftharpoons}}$	**$E.D_1 + PP_i$**	(4) Pyrophosphate is released
$E.D_1$	$\underset{k_{10}}{\overset{k_9}{\rightleftharpoons}}$	**$E + D_1$**	(5) Enzyme dissociates from DNA

Figure 3.37. The 5-step kinetic scheme that was first used to fit the data in Figure 3.16 and 3.17. Note that no enzyme conformational change is included in this scheme.

Using rate laws relating to the equations written above, the computer was programmed to fit them to the data observed (Figure 3.16). An output from one of the first parameter-fitting programs is given here as an illustration, shown in a different font and in bold. In order to explain the computer code, the program is shown in several parts, each part accompanied by an explanation. Only relevant parts of the program are shown, but a complete output of one of the calculations is included in the Appendix.

In this particular run, the pre-steady state incorporation of dTTP (5, 10, 20, 40 and 80μM) opposite O^6-meG at different time points (200 ms to 5 s) was analysed. Although measurements were made up to 10 s, the progress curves were fitted to data points up to 5 s only since there was very little change in the amount of product accumulated from 5 s to 10 s. The first step in the program was to declare all the variables and parameters. A variable is a quantity whose value is determined by the solution of a differential equation (e.g. [enzyme-DNA complex]), while a parameter is a quantity which may be constant through the time of integration or may be calculated as a function of time, variables or other parameters (e.g. the rate constants). In this particular run, the results of experiments conducted at five concentrations of dTTP were analysed, thus each variable has 5 different values and these values have to be entered as an array. The dimension of the array was set at 5 **(#1=5)** in the computer code below, and this was executed as soon as the computer reads this line by the "**compile instant;**" command. This routine

(i.e. the "**compile instant**" routine) was terminated by the "**;" command. The line reading "**varia<#1> ed edt ed1pp ed1 ttp pp;**" in the computer code declares ed, edt, ed1pp, ed1, ttp and pp as variables, each with 5 values. The enzyme is represented by e, the unelongated DNA primer as d, the elongated DNA primer as d1, the thymine nucleotide as t or ttp, and the pyrophosphate as pp. The initial concentration of dTTP was called s0, and the concentrations used in the five experiments are declared immediately after this. There is then a colon, followed by s5, s10, etc. These numbers (s5-s80) are labels for the five progress curves at the different dTTP concentrations:

```
compile instant;
#1=5;
**;
varia<#1> ed  edt ed1pp ed1 ttp pp;
param<#1> s0 5e-6 10e-6 20e-6 40e-6 80e-6 40e-6:s5 s10 s20 s40 s80;
```

Next, initial estimates for the rate constants were declared. Notice that these parameters were not entered as arrays. All diffusion-limited rate constants (k_3 and k_8) were assigned a value of 10^7 $M^{-1}s^{-1}$. The rate of pyrophosphate release (k_7) was fixed at 1000 s^{-1} based on calculations from a reported dissociation constant of 100 μM (Kuchta et al., 1987). Initial estimates for the parameters to be optimized were as follows: $k_4=10^5$ s^{-1}; $k_5=100$ s^{-1} and $k_6=100$ s^{-1}. The initial enzyme concentration (e0) was 0.12 μM.

```
param k3 1e7 k4 1e5 k5 100 k6 100 k7 1000 k8 1e7;
param e0 1.2e-7;
```

Next, the initial concentrations of the enzyme-DNA complex and the dTTP were assigned. In FACSIMILE, details of the kinetic problem is specified by the user in blocks of code, and each code block begins with a "compile" command and ends with a "**;" command. In the code block below, the "**compile initial;**" command instructs the computer to set the initial values of the variables and parameters according to the lines following this command.

```
compile initial;
ed0=d0;
ttp=s0;
**;
```

The equilibrium constants ks2, ks3 and ks4 were determined from the ratio of the forward and backward rate constants, and the algebraic equations expressing the relationship between the parameters were also included in the "compile initial" code block:

```
param ks2 ks3 ks4;
compile initial;
ks2=k3/k4;
ks3=k5/k6;
ks4=k7/k8;
```

Next, rate equations for the reactions were stated. The command "**compile equation;**" tells the computer that the lines in this code block specify the derivatives of the variables with respect to time. Memory space required by FACSIMILE for these specifications was allocated through the "**array <#1> ws;**" command. The derivatives of the variables were specified in the following way: for example, the line in the code block reading "**%k3%k4: ed+ttp=edt;**" shows that for the binding of dTTP to the enzyme-DNA complex, the forward and backward rate constants are k_3 and k_4 respectively:

```
compile equation;
array<#1> ws;
%k3%k4:   ed+ttp=edt;
%k5%k6:   edt=ed1pp;
%k7%k8:   ed1pp=ed1+pp;
**;
```

In the calculations, the step in which the polymerase binds to the primer was not included; instead the concentration of the ED complex at the beginning of the reaction was assumed to be equal to the initial concentration of DNA, D_0. This was because in the experiments, the polymerisation reactions were initiated by the addition of dNTP and Mg^{2+} to a solution containing a pre-formed Klenow-DNA complex, and the equilibrium for this can be worked out as a solution to a quadratic equation. The reaction to consider is

$$E + D \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E.D$$

For this reaction, we can write the following differential rate equations:

$$\frac{d[ED]}{dt} = k_1[E][D] \quad \text{and} \quad \frac{d[E]}{dt} = \frac{d[D]}{dt} = k_2[ED]$$

Equilibrium is achieved during the pre-incubation of the enzyme with the primer-template, and at equilibrium, the rate of the forward reaction equals the rate of the reverse reaction, so we can

$$k_1[E][D] = k_2[ED]$$

write

$$[E][D] = \frac{k_2}{k_1}.[ED]$$

$$[E][D] = K_d.[ED]$$

where K_d is the dissociation constant. If we now write [E] and [D] in terms of their initial concentrations E_0 and D_0, we can derive a qaudratic equation:

$$(E_0 - [ED])(D_0 - [ED]) = K_d.[ED]$$

$$[ED]^2 - (K_d + E_0 + D_0).[ED] + E_0.D_0 = 0.$$

c.f. $$ax^2 + bx + c = 0$$

By putting $a = 1$, $b = -(K_d + E_0 + D_0)$ and $c = E_0.D_0$, [ED] can be solved from

$$[ED] = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

K_d was assumed to be 5 nM based on the value reported by Kuchta et al. (1987) for the binding of Klenow to a 13/20-mer DNA. E_0 and D_0 were 0.12 μM and 0.1 μM respectively. Thus, the amount of [ED] at the start of the reaction was calculated at 0.999 μM, which is almost equal to D_0. Thus, [ED] at the start of reaction was assumed to equal D_0.

Since the gel assays measured only the amounts of unelongated and elongated primers, 9-mers and 10-mers respectively, the 9-mers had to be expressed as the sum of the ed and edt complexes, and the 10-mers as the sum of the ed1pp and ed1 complexes. This was done by first declaring the ninemers and tenmers, and then defining nm5-nm80 to represent the amount of 9-mer measured at the five concentrations of dTTP, and tm5-tm80 to represent the amount of 10-mer measured at the five concentrations of dTTP. Because the relative amounts of 9-mers and 10-mers were measured, the 9-mers and 10-mers had to be expressed as percentages of the total DNA concentration. The "**compile resid;**" command was used to tell the computer to set the measured amounts of ninemer and tenmer as functions of ed, edt, ed1pp and ed1 according to the instructions following this command. The "**array<#1> ws;**" command allocates work space (memory) for the arrays declared, while the "**array;**" command towards the end of the code block tells the computer to end the "array" instruction and return to scalar operations.

```
param<#1> ninemer:nm5 nm10 nm20 nm40 nm80;
param<#1> tenmer:tm5 tm10 tm20 tm40 tm80;
compile resid;
array<#1> ws;
ninemer=100*(ed+edt)/d0;
tenmer= 100*(ed1pp+ed1)/d0;
array;
**;
```

Finally, the data was entered as an array. The line reading "**data .035;**" tells the computer that the experimental error for the observed data is 3.5%, so that during parameter-fitting, an allowance should be made for an experimental error of 3.5%. This value is normally supplied by the user, but if this is omitted, then FACSIMILE will use a default value of 0.01. In the experiments carried out, the experimental error was unknown, but an estimated experimental error of 3.5% enabled FACSIMILE to fit the data and so this value was used. (If one assumes a large experimental error, then it is possible to fit quite a wide range of numbers to the same kinetic scheme, but if the experimental error is underestimated, then it becomes impossible to fit the data). The data for the five experiments performed at 5-80 μM dTTP over reaction time periods of 200 ms to 5 seconds were entered in an array. The first column in the array was the reaction time in seconds. The second, fourth, sixth, eighth and tenth columns were the experimental data for %9-mers at 5, 10, 20, 40 and 80 μM dTTP respectively, and the third, fifth, seventh, ninth and eleventh columns were the experimental data for %10-mers at 5, 10, 20, 40 and 80 μM dTTP respectively. The "**vary**" command specified that only 4 primary parameters k_3, k_4, k_5 and k_6 and 2 secondary parameters ks2 and ks3 were to be varied in this run. Parameter fitting was initiated by the "**begin**" command.

```
data .035;
time      nm5  tm5  nm10  tm10  nm20  tm20  nm40  tm40  nm80  tm80;
range     96.8  33.8  86.3   52.6   84.7    60.7   76.5   69.6    74.8    70.3;
 0.2  96.8   3.2  86.3   13.7   84.7    15.3   76.5   23.5    74.8    29.2;
 0.4  88.2  11.8  82.3   17.7   79.3    20.7   66.3   33.7    55.7    44.3;
 0.8  83.7  16.3  75.7   24.3   67.6    32.4   55.7   44.3    45.9    54.1;
 1.2  83.3  16.7  68.1   31.9   61.9    38.1   45.6   54.4    40.4    59.6;
 1.6  79.6  20.4  63.7   36.3   54.5    45.5   45.2   54.8    38.9    61.1;
 3.0  72.0  28.0  56.2   43.8   46.2    53.8   36.2   63.8    34.7    65.3;
 5.0  66.2  33.8  47.4   52.6   39.3    60.7   30.4   69.6    29.7    70.3;
**;
vary k3 k4 k5 k6:ks2 ks3;
begin;
```

The output from this run is shown in the following pages. The parameter-fitting starts once the "**vary**" and "**begin**" commands have been read. Included in the output is the number of progress curves and data points analysed in the run. In this particular run, the results of 5 experiments (at 5 [dTTP]) were analysed, and in each experiment, measurements of 9-mer, and 10-mer were made at 7 time points, generating two progress curves for each experiment. Thus, a total of 10 progress curves and 70 (2 x 7 x 5) data points were used in this analysis.

```
10 CURVES
     7 TIME      POINTS
     70 OBSERVATIONS
A TIME      POINT INCREMENT OF     1 AND OFFSET OF     0 SPECIFIED
      10 DEPENDENT VARIABLES SELECTED
      7      TIME      POINTS AND    70 OBSERVATIONS USED
```

The output generated reflects the progress of parameter-fitting. The first integration was performed by using the initial estimates of the parameters supplied by the user. Statistical calculations (RSQS, CORI, AUCO and AUCR) were then carried out after the integration. The values of RSQS, CORI, AUCO and AUCR were printed for each progress curve, followed by the total residual sum of squares, the mean absolute correlation index and the value of each parameter.

```
OPTIMISATION RUN WITH 70 OBSERVATIONS AND 4 VARIED PARAMETERS


 AFTER   0 ITERATIONS AND   1 RUNS, INDIVIDUAL CURVE STATISTICS ARE
       RSQS        CORI    AUCO    AUCR
    8.8487E+01  -2.4923   0.8605  0.3499
    7.2576E+02   2.4923   0.8605  0.3499
    2.6068E+02  -2.5919   0.8919  0.3499
    7.0171E+02   2.5919   0.8919  0.3499
    1.9042E+02  -2.5168   0.9054  0.3499
    3.7077E+02   2.5168   0.9054  0.3499
    2.2519E+02  -2.0621   0.8317  0.3499
    2.7205E+02   2.0621   0.8317  0.3499
    1.9107E+02  -0.0972   0.6613  0.3499
    2.3452E+02   0.1995   0.6644  0.3499
   TOTAL RESIDUAL SUM OF SQUARES IS      3.26066E+03
   MEAN ABSOLUTE CORRELATION INDEX IS   1.96230
   VARY( 1):K3         =  1.0000E+07
   VARY( 2):K4         =  1.0000E+05
   VARY( 3):K5         =  1.0000E+02
   VARY( 4):K6         =  1.0000E+02
   VARY( 5):KS2        =  1.0000E+02
   VARY( 6):KS3        =  1.0000E+00
```

FACSIMILE then started improving the parameter estimates (iteration) and the statistics of the best fit was reported after every 4 integration runs (one for every each primary parameter varied). Notice that the total residual sum of squares has been greatly reduced from the previous runs.

```
      5 SIMULATION RUNS COMPLETED
      6 SIMULATION RUNS COMPLETED
      7 SIMULATION RUNS COMPLETED
      8 SIMULATION RUNS COMPLETED

 AFTER  4 ITERATIONS AND  9 RUNS, INDIVIDUAL CURVE STATISTICS ARE
    RSQS       CORI    AUCO    AUCR
  3.4292E+01 -2.4313  0.8809  0.3499
  2.8126E+02  2.4313  0.8809  0.3499
  1.0072E+02 -2.4797  0.8908  0.3499
  2.7112E+02  2.4797  0.8908  0.3499
  5.8088E+01 -1.3030  0.6353  0.3499
  1.1310E+02  1.3030  0.6353  0.3499
  1.2286E+02  0.1995  0.6351  0.3499
  1.4842E+02 -0.1995  0.6351  0.3499
  3.0353E+02  1.8865  0.7909  0.3499
  3.5080E+02 -1.7804  0.7853  0.3499
 TOTAL RESIDUAL SUM OF SQUARES IS     1.78419E+03
 MEAN ABSOLUTE CORRELATION INDEX IS  1.64940
 VARY( 1):K3       =  1.0486E+07
 VARY( 2):K4       =  8.1545E+04
 VARY( 3):K5       =  1.2786E+02
 VARY( 4):K6       =  1.0556E+02
 VARY( 5):KS2      =  1.2859E+02
 VARY( 6):KS3      =  1.2113E+00
```

In the parameter-fitting process, 64 iterations were performed and the total residual sum of squares was finally reduced to 219, which is outside the expected range (see output below). Since the total residual sum of squares could not be reduced any further, FACSIMILE gave a warning that "EITHER (A) OBSERVATIONAL ERRORS WERE UNDERESTIMATED OR (B) THERE ARE SYSTEMATIC ERRORS IN MODEL, CONSTANTS OR DATA".

```
AFTER  64 ITERATIONS AND  69 RUNS, INDIVIDUAL CURVE STATISTICS ARE
    RSQS       CORI    AUCO    AUCR
  5.0422E+00  0.0960  0.4639  0.3499
  4.1356E+01 -0.0960  0.4639  0.3499
  1.5205E+01 -2.1444  0.4382  0.3499
  4.0930E+01  2.1444  0.4382  0.3499
  9.7739E+00  0.0867  0.2853  0.3499
  1.9031E+01 -0.0867  0.2853  0.3499
  1.8568E+01 -0.7103  0.2033  0.3499
  2.2432E+01  0.7103  0.2033  0.3499
  1.9116E+01  0.8953  0.5311  0.3499
  2.7551E+01 -0.4838  0.5452  0.3499


 TOTAL RESIDUAL SUM OF SQUARES IS     2.19005E+02
  MEAN ABSOLUTE CORRELATION INDEX IS  0.74538
```

```
VARY( 1):K3       = 3.7691E+09
  VARY( 2):K4     = 2.1730E+05
  VARY( 3):K7     = 1.2712E+01
  VARY( 4):K8     = 4.6408E+03
  VARY( 5):KS2    = 1.7345E+04
  VARY( 6):KS4    = 2.7392E-03
      69 SIMULATION RUNS COMPLETED
      70 SIMULATION RUNS COMPLETED
      70 SIMULATION RUNS COMPLETED

 RESIDUAL SUM OF SQUARES IS          218.5142873
 COMPARED WITH EXPECTED RANGE  45.4236  90.3524
 FOR  66 DEGREES OF FREEDOM
 EITHER (A) OBSERVATIONAL ERRORS WERE UNDERESTIMATED
 OR (B) THERE ARE SYSTEMATIC ERRORS IN MODEL, CONSTANTS OR DATA
 ANALYSIS OF VARIANCE PROCEEDS ON ASSUMPTION OF GOOD FIT

 BUT INSPECT RESIDUALS FOR EVIDENCE OF SYSTEMATIC ERROR
```

Next, each parameter was varied slightly so that a dependence matrix which relates the sensitivity of each RSQS to the parameter being varied, and the singular values of the values of the dependence matrix were reported. In this calculation, the singular value corresponding to k_4 (0.014) was much smaller than 1, which indicates that this parameter was not well-determined by the data.

```
 SINGULAR VALUES OF DEPENDENCE MATRIX

   VALUES   120.054  0.014   2.771  15.271
```

Since k_4 was not well-determined by the data, the parameter-fitting was repeated by varying only 3 primary parameters instead of 4 and the value of k_4 was fixed at the value last used in the calculations. The sequence of the parameters being varied was rearranged so that the most accurate parameter comes first (k_3) and the most ill-determined parameter (k_4)comes last.

```
PERMUTE SEQUENCE OF VARIED PARAMETERS FROM

    1       2       3       4
 TO
    1       4       3       2

 DATA DO NOT DETERMINE THE FOLLOWING  1 VARIED PARAMETERS
 SO DEFINE Q(I)=PARAM(I)/VALUE(I) AS INDEPENDENT NUMBERS (CURRENTLY ALL = 1.0)
  I PARAM(I)  VALUE(I)

  4 K4        9.3377E+04

 RE-ENTERING WITH REDUCED NUMBER OF VARIED PARAMETERS

 SIZE NEEDED FOR ARRAY WSFIT =      593

OPTIMISATION RUN WITH 70 OBSERVATIONS AND 3 VARIED PARAMETERS
```

The results of this parameter-fitting with a reduced number of varied parameters are as follows:

```
AFTER  0 ITERATIONS AND  1 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS       CORI    AUCO    AUCR
  5.1042E+00  0.1608  0.4667  0.3499
  4.1864E+01 -0.1608  0.4667  0.3499
  1.4875E+01 -2.1204  0.4267  0.3499
  4.0042E+01  2.1204  0.4267  0.3499
  9.9365E+00  0.1377  0.2948  0.3499
  1.9347E+01 -0.1377  0.2948  0.3499
  1.8532E+01 -0.6881  0.2017  0.3499
  2.2389E+01  0.6881  0.2017  0.3499
  1.8972E+01  0.8961  0.5292  0.3499
  2.7353E+01 -0.4832  0.5437  0.3499
 TOTAL RESIDUAL SUM OF SQUARES IS     2.18415E+02
 MEAN ABSOLUTE CORRELATION INDEX IS  0.75935

 VARY( 1):K3       =  1.6268E+09
 VARY( 2):K6       =  4.8436E+03
 VARY( 3):K5       =  1.3104E+01
      1 SIMULATION RUNS COMPLETED
      2 SIMULATION RUNS COMPLETED
      3 SIMULATION RUNS COMPLETED
      4 SIMULATION RUNS COMPLETED
      4 SIMULATION RUNS COMPLETED

 PARAMETER NUMBER, TYPE,    NAME,   VALUE
         1  PRI.   K3       1.6301E+09
         2  PRI.   K6       4.8533E+03
         3  PRI.   K5       1.3104E+01
```

Finally, the final parameter values were given together with their 5% and 95% confidence limits. The SDLN is the estimated standard deviation for the logarithm of each parameter value and is used in calculating the confidence limits. A correlation matrix was also given which shows how the different parameters are related to one another.

```
FITTED VALUES, ACCURACIES AND CONFIDENCE LIMITS
  NO.  NAME     VALUE        SDLN  5 PERCENT  95 PERCENT
   1 K3       1.6301E+09 0.1363   1.3026E+09    2.0399E+09
   2      K6       4.8533E+03 0.4839   2.1894E+03    1.0758E+04
   3      K5       1.3104E+01 0.4482   6.2697E+00    2.7389E+01

  CORRELATION MATRIX COMPONENTS

  COLUMN  1       2       3
  ROW  1 1.000  -0.093  -0.300
  ROW  2 -0.093  1.000  0.976
  ROW  3 -0.300  0.976  1.000
```

As the results show, this particular run with a three-step mechanism was not successful. The total residual sum of squares was very much bigger than the expected range (218 compared with an expected range of 46.3 to 91.5), which indicated a poor fit of the

predicted curve to the observed data points. A graphical representation of the fit is given in Figure 3.38. Subsequently, three other reaction schemes, based on those of Klenow and T7 DNA polymerase deduced by the laboratories of Benkovic and of Johnson, were tested:

- involving a conformational change ***before*** phosphodiester bond formation;
- involving a conformational change ***after*** phosphodiester bond formation, and
- involving conformational changes ***before and after*** phosphodiester bond formation.

The residual sum of squares of the curves predicted according to these models are given in Table 14 and the fit of these curves to the experimental data are shown below in Figures 3.38-3.41. It is quite difficult to tell how well the predicted curves fit the observed the data just by visual inspection of Figures 3.38-3.41. However, from the total residual sum of squares, it is obvious that the reaction scheme without any conformational change, and that involving only the conformational change induced by dNTP-binding, could not be fitted to the observed data. The curves predicted according to these two models (Figures 3.38 and 3.39) are all hyperbolic whereas the ***data points appear to suggest a biphasic reaction profile***. When a conformational step after phosphodiester bond formation was included in the reaction scheme, the fit of the predicted curve to the observed data points became much improved (Figure 3.40): the computer predicted progress curves were biphasic instead of hyperbolic. However, further analysis of the reaction scheme including two conformational changes, one occurring before and the other occurring immediately after phosphodiester bond formation, showed that this reaction scheme has a slightly better fit to the observed data (Figure 3.41 and Table 14).

Table 14. The total residual sum of squares of the computer predicted progress curves. Four reaction schemes were analysed: (1) with ***no*** conformational change, (2) with a conformational ***preceding*** the formation of phosphodiester bond, (3) with a conformational change ***after*** the formation of the phosphodiester bond, and (4) with ***two*** conformational changes, one preceding and one after the formation of the phosphodiester bond. The reaction scheme with two conformational changes produced the best fit to the observed data points (lowest RSQS).

Reaction scheme	total RSQS	expected range for total RSQS
(1) ***no*** conformational change	218	46.3 - 91.5
(2) ***one*** conformational change ***before*** phosphodiester bond formation	220	46.3 - 91.5
(3) ***one*** conformational change ***after*** phosphodiester bond formation	85	43.8 - 88.0
(4) ***two*** conformational changes, one before and one after phosphodiester bond formation	72	43.8 - 88.0

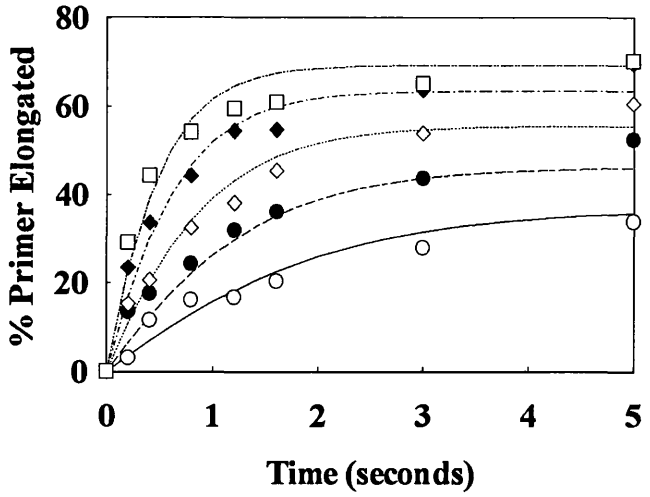


Figure 3.38. Graph showing the fit of the progress curves predicted by computer to the experimental data. The data points are the experimental results of pre-steady state incorporation of thymine opposite O^6-meG in the template strand. 0.12 μM Klenow fragment and 0.1 μM DNA duplex A were pre-incubated with EDTA and the reactions were initiated by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◊), 40 μM (♦) and 80 μM (□) dTTP. The lines are the predicted outcome from the kinetic scheme ***without any conformational change***.

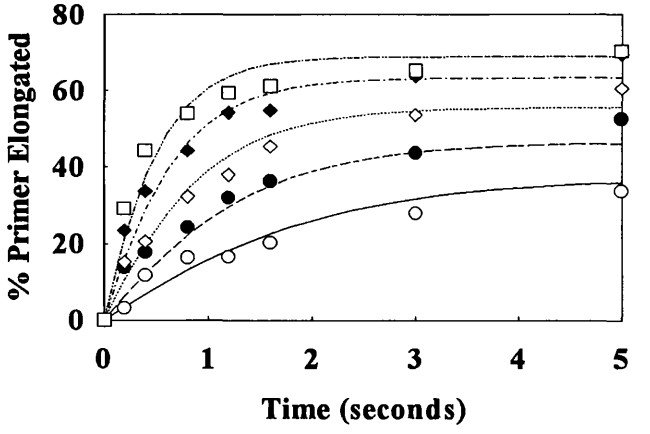


Figure 3.39. Graph showing the fit of the progress curves predicted by computer to the experimental data. The data points are the experimental results of pre-steady state incorporation of thymine opposite O^6-meG in the template strand. 0.12 μM Klenow fragment and 0.1 μM DNA duplex A were pre-incubated with EDTA and the reactions were initiated by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◊), 40 μM (♦) and 80 μM (□) dTTP. The lines are the predicted outcome from the kinetic scheme which has a conformational change occurring just ***before*** the phosphodiester bond is formed.

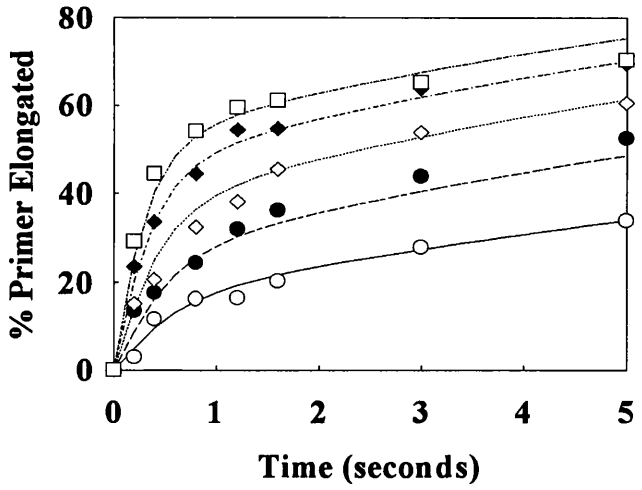


Figure 3.40. Graph showing the fit of the progress curves predicted by computer to the experimental data. The data points are the experimental results of pre-steady state incorporation of thymine opposite O^6-meG in the template strand. 0.12 μM Klenow fragment and 0.1 μM DNA duplex A were pre-incubated with EDTA and the reactions were initiated by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◊), 40 μM (♦) and 80 μM (□) dTTP. The lines are the predicted outcome from the kinetic scheme which has a conformational change occurring ***after*** the phosphodiester bond is formed.

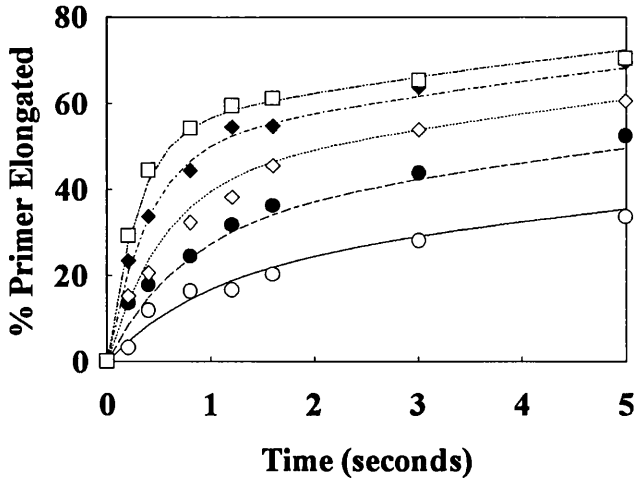


Figure 3.41. Graph showing the fit of the progress curves predicted by computer to the experimental data. The data points are the experimental results of pre-steady state incorporation of thymine opposite O^6-meG in the template strand. 0.12 μM Klenow fragment and 0.1 μM DNA duplex A were pre-incubated with EDTA and the reactions were initiated by the addition of magnesium and 5 μM (○), 10 μM (●), 20 μM (◊), 40 μM (♦) and 80 μM (□) dTTP. The lines are the predicted outcome from the kinetic scheme which has *two* conformational changes, occurring ***before and after*** the phosphodiester bond is formed.

Analysis of the data according to the reaction scheme with a conformational change after phosphodiester bond formation, and the reaction scheme which included two conformational changes, one before and one after phosphodiester bond formation, showed that ***the rate-limiting step was the conformational change which occurred after phosphodiester bond formation***. Thus, the intermediate that occurs immediately before this rate-limiting step, i.e. the $Ei.D_1.PP_i$ complex in Figure 3.5 (page 93), would accumulate during the initial phase of the reaction and therefore contribute most towards the observed 10-mers on the gels. This is illustrated by the biphasic progress curve for the $Ei.D_1.PP_i$ complex (Figure 3.42). Initially, the concentration of $Ei.D_1.PP_i$ rises very sharply with time because it is formed rapidly but converted to $E.D_1.PP_i$ very slowly. The progress curve then decreases until the gradient becomes very shallow and negative as the rate of $Ei.D_1.PP_i$ formation becomes slower than that for its conversion to the $E.D_1.PP_i$ complex. By contrast, after the second conformational change, the release of the pyrophosphate is very rapid, so rapid that the $E.D_1.PP_i$ complex does not accumulate, while the progress curve for the $E.D_1$ complex is a straight line with a shallow gradient, with the appearance of a first order reaction (Figure 3.42). Since each progress curve for the observed 10-mers is made up of the respective progress curves for the formation of $Ei.D_1.PP_i$, $E.D_1.PP_i$ and $E.D_1$, the only way to produce a biphasic curve for the observed 10-mers is to include a rate-limiting step after phosphodiester bond formation which would force one of the species that contribute to the observed 10-mers to accumulate. This explains why inclusion of the conformational change after phosphodiester bond formation in the reaction scheme made such a great improvement to the fitting of the computer predicted curves to the observed data.

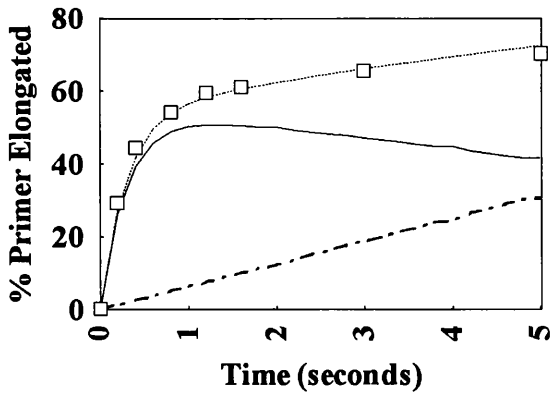


Figure 3.42. Progress curves of the $Ei.D_1.PP_i$ and $E.D_1$ complexes as predicted by computer analysis of data from the pre-steady state incorporation of 80 μM dTTP opposite O^6-meG in the template strand, based on the kinetic scheme which has two conformational changes occurring before and after the phosphodiester bond is formed. The points (□) are the experimental data points from Figure 3.16. The computer predicted results for the total amount of elongated products (i.e. $E.D_1 + Ei.D_1.PP_i + E.D_1.PP_i$), and for the amounts of $Ei.D_1.PP_i$ and $E.D_1$ complexes are represented by - - - -, ———, and —·—·— respectively.

Therefore, as shown by Kuchta et al (1988), two extra steps had to be included in the polymerisation reaction pathway. These were enzyme conformational changes immediately before and after phosphodiester bond formation (see Figure 3.5). To do so, additional parameters had to be included (see program below; ei represents the enzyme after it has undergone a conformational change). The values of k_3, k_{11} and k_{12} in this reaction scheme with two conformational changes (refer to Figure 3.5, page 93) were fixed at the same values as before, and k_4 to k_{10} were allowed to be varied. The definition of 10-mer was changed to "tenmer = (eid1pp + ed1pp + ed1 + d1) x 100/d0", and that for 9-mer was changed to "ninemer = (ed + edt + edt) x 100/d0". The exonuclease reactions were also included, although it was found that inclusion of the exonuclease reactions made negligible difference on the fitting because the exonuclease rate (measured at 1.6 x 10^{-3} s^{-1} for thymine paired with O^6-meG in the template strand) was so slow compared with the other rates in the polymerisation reaction. The modified program looked like this (only relevant parts of program are shown):

```
varia<#1> ed  edt eidt eid1pp ed1;
varia<#1> d1, ttp, pp;
param <#1> s0 5e-6 10e-6 20e-6 40e-6 80e-6:s5 s10 s20 s40 s80;
param k3 1e7 k4 1e3 k5 57.7 k6 22.2 k7 3.79 k8 1.67;
param k9 .245 k10 1e-3 k11 1e3 k12 1e7;
param ks2 ks3 ks4 ks5 ks6;
param kexo 0.0016 d0 1e-7 e0 1.2e-7;

compile equation;
array<#1> ws;
%k3%k4      :ed+ttp=edt;
%k5%k6      :edt=eidt;
%k7%k8      :eidt=eid1pp;
%k9%k10     :eid1pp=ed1pp;
%k11%k12    :ed1pp=ed1+pp;
%k13%k14    :ed1 = e + d1;
%kexo       :eid1pp = ed + pp;
%kexo       :ed1pp = ed + pp;
%kexo       :ed1 = ed + pp;
array;
**;
param<#1> ninemer:  nm5 nm10 nm20 nm40 nm80;

param<#1>  tenmer:  tm5 tm10 tm20 tm40 tm80;
```

```
compile resid;
array<#1> ws;
ninemer=100*(ed+edt+eidt)/d0;
tenmer= 100*(eid1pp+ed1pp+ed1+d1)/d0;
array;
**;
```

The possible existence of more than one minimum in a model with many parameters has been mentioned previously, and this problem was encountered in the calculations of the rate constants for incorporation of thymine and cytosine opposite O^6-meG. Although the simulated curves now fit the data, the parameters have not been well-determined. The main problem in assigning values to the individual rate constant was that the experimental design allowed one to measure elongated DNA, denoted by 10-mer (i.e. 10-mer = $Ei.D_1.PP_i + E.D_1.PP_i + Ei.D_1 + D_1$), but the amounts of the ($Ei.D_1.PP_i + E.D_1.PP_i + Ei.D_1 + D_1$) complexes could not be resolved. To a certain extent, FACSIMILE can get an equally good fit of the data using a number of values of the internal constants by reciprocally trading off one constant against another. These constants are shown as highly correlated in the correlation matrix calculated by FACSIMILE. If one constant can be traded off against another then the overall kinetics could be satisfied by a range of values for the rate constants and the results of the analysis will give no insight into the underlying process. To prevent this from happening, some constraints had to be applied to at least one pair of the intermediate rate constants.

Two constraints were used to limit the possible values of the rate constants. First the overall Gibbs' free energy change (ΔG) was assumed to be -4 kcal/mol (taken from Patel et al., 1991) thus fixing the ***overall equilibrium constant keq***, i.e. the ratio of the forward to the backward rates, as 1000. This Gibbs' free energy change is roughly what one would expect for the formation of a phosphodiester bond during DNA polymerization, because the ΔG for the hydrolysis of a nucleoside triphosphate to nucleoside monophosphate and pyrophosphate ions is -7.3 kcal/mol and the formation of a sugar phosphate (e.g. Glucose + Pi $\rightarrow$ Glucose-6-phosphate + H_2O) is +3.3 kcal/mol.

To set a value for the overall equilibrium constant keq as a constraint, another algebraic equation had to be added to the program:

```
k10=(k3*k5*k7*k9*k11)/(keq*k4*k6*k8*k12);
```

The effect of assuming different values of keq on the fit of the calculated progress curves to the experimental data and on the calculated values of the parameters were investigated by putting in different values of keq and observing the RSQS and the calculated values of the rate constants. Different values of keq were tried: 100 - 100 000. The fit of the progress curves (i.e. the RSQS) to the experimental data was not at all affected by the big changes in keq. All the calculated values were unaffected except the values of k_{10} and ks5; k_{10} was directly proportional and ks5 was inversely proportional to the value of keq. Since the large changes in keq produced no changes in the calculated values except k_{10} and ks5, it was considered justifiable to include keq in the calculations as a constraint.

The second constraint applied was to assume that the forward and backward rates of the bond-formation step, i.e. k_7 and k_8, were 150 times slower for the incorporation of the phosphorothioate nucleotide analogues ((S_p)-dNTPαS) than for the incorporation of the normal, oxygen-containing nucleotides (dNTP). An inspection of the rates of incorporation of normal and phosphorothioate nucleotides opposite O^6-meG in the template strand show that the rate of incorporation of thio-nucleotide was at least 75 times slower than the corresponding rate of incorporation of the regular nucleotide (Fig. 3.23). According to Bryant et al. (1983), Mizrahi et al. (1985) and Kuchta et al. (1987, 1988), this is indicative of a rate-limiting phosphodiester formation step in the enzyme mechanism. The rationale behind this was based on studies by Benkovic & Schray (1971) which showed that thio-substitution reduces the rates of reactions with phosphate triesters by 30-100-fold.

Thus, it should be possible to analyse the results of incorporation of both normal nucleotides and their phosphorothioate analogues as a single set of data. The following parameters were introduced into the FACSIMILE program: rks4, kcf and kcb; rks4 is the magnitude of the thio-effect, while kcf and kcb are the forward and backward rates of formation of the phosphodiester bond relative to the magnitude of the thio-effect. The line in the computer code below reading "**param<#1> kcf kcb rks4 1 1 1 1 1 150:rk1 rk2 rk3 rk4 rk5 rk6;**" was used to declare the parameters kcf, kcb and rks4. In this particular run, the results of six experiments were analysed: 5 μM dTTP, 10 μM dTTP, 20 μM dTTP, 40μM dTTP, 80 μM dTTP, and 40 μM dTTPαS. For the progress curves involving the normal nucleotides, rks4 is 1 but for the progress curve involving the phosphorothioate nucleotide, rks4 is 150; rks4 for the six progress curves were labelled rk1 to rk6, where rk1 is the magnitude of the thio-effect for the experiment with 5 μM dTTP, rk2 is the magnitude of the thio-effect for the experiment with 10 μM dTTP and so on. Thus, values for rks4 were assigned by the six numbers and labels following rks4 (**1 1 1 1 1 150:rk1 rk2 rk3 rk4 rk5 rk6;**). The algebraic relationships between kcf, k7 and rks4, and between kcb, k8 and rks4 were included in the "compile initial" code block where the initial values of the variables and parameters are defined (see page 149).

```
param<#1> kcf kcb rks4 1 1 1 1 1 150:rk1 rk2 rk3 rk4 rk5 rk6;
compile initial;
kcf=k7/rks4;
kcb=k8/rks4;
```

Although the literature suggests that the phosphorothioate-effect is 30-100, mathematical analysis suggests that a value of 150 would be more appropriate for the effect of thio-substitution for the incorporation of both thymine and cytosine opposite O^6-meG. Figure 3.23 shows that the rate of incorporation of thio-nucleotide was at least 75 times slower than the corresponding rate of incorporation of the regular nucleotide, thus the analysis was carried out with different values of rk6 ranging from 75 to 180. When rk6 was set at 75 or above 180, the residual sum of squares became very large. When rk6 was set at 100 or 120, the value of k_5 and k_6 were not well-determined, but when rk6 was fixed at 150, k_5 and k_6 became very well-defined, and for this reason, it was decided that the thio-effect was probably closer to 150 than 100. The results of the calculations for rk6 set at 100, 120 and 150 are shown in Table 15.

Table 15: Application of the phosphorothioate data as a constraint in the calculation of the rate constants for incorporation of thymine opposite O^6-meG in the template strand: effects of changing the magnitude of the thio-effect (rk6) on the calculated values of the parameters and on the residual sum of squares (RSQS). * denotes values which were not well-determined by the data (the 5% and 95% confidence limits differ by more than two or more orders of magnitude).

rk6	k4 (s^{-1})	k5 (s^{-1})	k6 (s^{-1})	k7 (s^{-1})	k8 (s^{-1})	k9 (s^{-1})	k10 (s^{-1})	ks2 μM	ks3	ks4	ks5	RSQS
100	n.d.	4850*	4850*	4.07	0.77	0.14	1.6e-3	n.d.	1.0	5.32	8.5	50.2
120	n.d.	53.8*	54.2*	4.76	0.97	0.06	1.6e-3	n.d.	1.0	4.94	91.8	46.6
150	n.d.	21.7	18.6	5.42	1.36	0.17	1.7e-3	n.d.	1.2	3.98	96.9	45.9

It proved impossible to determine k_4 within reasonable limits because the product of the reaction (E.D.dNTP, Figure 3.5) has a very low maximum concentration, and any change in k_4 could be compensated for by changes in k_5 or k_6. Trial calculations showed that the value of k_4 is almost certainly between 500 and 1500 s^{-1}, i.e. K_D is between 50 and 150 μM. A value of 1000 s^{-1} for k_4 was assumed (i.e. K_D = 100 μM), which allowed the internal rate constants shown in Table 12 to be calculated.

The FACSIMILE output of the mathematical analysis of the data for the pre-steady state incorporation of thymine opposite O^6-meG in the template strand using the two constraints mentioned above, i.e. (1) keq= 1000 and (2) thio effect for forward and backward rates of phosphodiester bond formation for a phosphorothioate nucleotide analogue is 150, is included in the appendix. The mathematical analysis of the data for the pre-steady state incorporation of cytosine opposite O^6-meG in the template strand (Figure 3.17) was also carried out in the same way.

Mathematical analysis of the steady state data

The rate of dissociation of Klenow from the elongated primer (10-mer), k_{13} in the reaction scheme in Figure 3.5, was not determined by the pre-steady state data. This is hardly surprising since the pre-steady state experiments were carried out with a slight excess of enzyme over DNA and a single catalytic cycle was followed, therefore the observed progress curves were not limited by the dissociation of enzyme from product DNA before a second round of polymerisation could begin, and the data obtained does not contain any information about the dissociation of enzyme from product DNA. To determine k_{13}, it was necessary to use the data from steady state experiments where there was an excess of DNA over enzyme and the observed progress curves therefore depended on the rate at which the enzyme dissociated from the product DNA. The steady state time course experiments were carried out at three enzyme concentrations. For the steady state incorporation of thymine, only the lowest enzyme concentration (0.109 μM) was used in mathematical analysis because at higher enzyme concentrations (0.327 and 0.654 μ M), most of the primer had already been elongated by the time the first observation was made.

The rate constants obtained from the previous analyses of the pre-steady state data were fixed in the analysis of k_{13}; only k_{13} and ks7 were allowed to vary. Since the reaction times were much longer in the steady state experiments, the exonuclease reactions became appreciable and were included in the analysis. Furthermore, since the polymerisation reaction was initiated by the addition of enzyme, the concentration of the enzyme.DNA complex was calculated from the solution of the quadratic equation derived from the initial concentrations of enzyme and DNA considered previously.

Michaelis-Menten analysis of the data

Once the reaction scheme for incorporation was known, it was possible to analyse the data according to Michaelis-Menten kinetics. The Michaelis-Menten equation for the rate of reaction of a single substrate with an enzyme is:

$$v = \frac{V_{max}.[S]}{K_m + [S]}$$

where v is the initial rate of formation of products, V_{max} is the maximal velocity that v tends towards at saturating substrate concentration ([S]), and K_m, the Michaelis constant, is the substrate concentration when $v = V_{max}/2$.

Two methods were used to calculate the values of K_m and V_{max} for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand from the pre-steady state data shown in Figures 3.6 and 3.17. The first method was to obtain K_m and V_{max} from the initial rates of pre-steady state incorporation at different dNTP concentrations (Figures 3.18 and 3.19). The second method was to fit the experimental data directly to the integrated form of the Michaelis-Menten equation using FACSIMILE in order to calculate K_m and V_{max}. Comparison of the values obtained by different methods served as a check to see if the calculated constants were of the right magnitude.

To fit the experimental data to the Michaelis-Menten equation by FACSIMILE, it was first necessary to rewrite the Michaelis-Menten equations in terms of the internal rate constants of the polymerisation model. The reactions to consider are:

$$E.D + dNTP \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E.D.dNTP \underset{k_6}{\overset{k_5}{\rightleftharpoons}} Ei.D.dNTP \underset{k_8}{\overset{k_7}{\rightleftharpoons}} Ei.D_1.PP$$

The reaction velocities, i.e. the net flux of each reaction, are

$$V1 = k_3.[dNTP].[E.D] - k_4[E.D.dNTP] \quad (1)$$

$$V2 = k_5.[E.D.dNTP] - k_6[Ei.D.dNTP] \quad (2)$$

$$V3 = k_7.[Ei.D.dNTP] - k_8[Ei.D_1.PPi] \quad (3).$$

Let us consider the reactions when the system is running at maximal velocity. V3 is greatest when $[Ei.D_1.PPi] = 0$. Thus, the maximum velocity of V3 is given by

$$V3 = k_7.[Ei.D.N]$$

$$[Ei.D.N] = V3/k_7 \quad (4).$$

Substituting (4) into (2), we get

$$V2 = k_5.[E.D.dNTP] - V3.\frac{k_6}{k_7}$$

$$[E.D.dNTP] = \frac{V2}{k_5} + \frac{V3.k_6}{k_5.k_7} \qquad (5).$$

Substituting (5) into (1), we get

$$[E.D] = \frac{V1}{[dNTP].k_3} + \frac{V2.k_4}{[dNTP].k_3.k_5} + \frac{V3.k_4.k_6}{[dNTP].k_3.k_5.k_7} \qquad (6).$$

If the total enzyme concentration is E_0, then

$$[E_0] = [E.D] + [E.D.dNTP] + [Ei.D.dNTP] \qquad (7).$$

When the system is at maximal velocity, [dNTP] $\rightarrow \infty$, [E.D] $\rightarrow 0$ and V2 = V3 = V_{max} where V_{max} is the maximum steady state velocity. Therefore, at V_{max},

$$[E_0] = [E.D.dNTP] + [Ei.D.dNTP],$$

and substituting (4) and (5) into this, we get

$$[E_0] = (\frac{V2}{k_5} + \frac{V3.k_6}{k_5.k_7}) + (\frac{V3}{k_7})$$

$$[E_0] = V_{max}(\frac{1}{k_5} + \frac{k_6}{k_5.k_7} + \frac{1}{k_7})$$

If we put $x = \frac{1}{k_7}$ and $y = \frac{1}{k_5} + \frac{k_6}{k_5.k_7}$, then $V_{max} = \frac{[E_0]}{(x+y)}$ (8).

Since $V_{max} = K_{cat}.[E_0]$, $K_{cat} = \frac{1}{(x+y)}$ (9).

Thus, V_{max} and K_{cat} are expressed in terms of the rate constants. To derive an expression for K_m in terms of the rate constants, one has to return to equation (7) again. Since $V_{max} = K_{cat}.[E_0]$, equation (7) becomes

$$\frac{V_{max}}{K_{cat}} = [E.D] + [E.D.dNTP] + [Ei.D.dNTP]$$

$$\frac{V_{max}}{K_{cat}} = (\frac{V1}{[dNTP].k_3} + \frac{V2.k_4}{[dNTP].k_3.k_5} + \frac{V3.k_4.k_6}{[dNTP].k_3.k_5.k_7}) + (\frac{V2}{k_5} + \frac{V3.k_6}{k_5.k_7}) + (\frac{V3}{k_7})$$

When [dNTP] = K_m, the reaction rate is half V_{max}, so

$$\frac{V_{max}}{K_{cat}} = \frac{V_{max}}{2K_m}\left(\frac{1}{k_3} + \frac{k_4}{k_3.k_5} + \frac{k_4.k_6}{k_3.k_5.k_7}\right) + \frac{V_{max}}{2}\left(\frac{1}{k_5} + \frac{k_6}{k_5.k_7} + \frac{1}{k_7}\right)$$

If we put $x = \frac{1}{k_7}$, $y = \frac{1}{k_5} + \frac{k_6}{k_5.k_7}$, $z = \frac{1}{k_3} + \frac{k_4}{k_3}(y)$, $K_{cat} = \frac{1}{(x+y)}$, and multiply both sides of the equation by $2/V_{max}$, the equation becomes

$$\frac{2}{K_{cat}} = \frac{z}{K_m} + (x + y)$$

$$K_m = \frac{z}{(x+y)} \qquad (10).$$

Once the parameters K_m, V_{max} and K_{cat} have been expressed in terms of the rate constants, (equations 8-10), these equations were then put into the FACSIMILE program (in the "compile initial" code block) and the parameters K_m and V_{max} allowed to be varied in the calculations. The program thus looked like this (only relevant parts are shown):

```
param k3 1e7 k4 1e3 k5 57.7 k6 22.2 k7 3.79 k8 1.67;
param k9 .245  k10 1e-3 k11 1e3 k12 1e7 keq 1e3 e0 1.2e-7 d0 1e-7;
param km 3.13e-5 vm 2.61e-7 kcat kmvm ks2 ks3 ks4 ks5 ks6 x y z;
compile initial;
x=1/k7;
y=(1/k5)+(k6/(k5*k7));
z=(1/k3)+(k4*y/k3);
vm=e0/(x+y);
km=z/(x+y);
kmvm=km/vm;
kcat=1/(x+y);
```

The values of K_m and V_{max} calculated by FACSIMILE were therefore fitted to the polymerisation model involving the two conformational change steps by FACSIMILE. The results of the Michaelis-Menten analyses by the two different approaches were very encouraging: the FACSIMILE-calculated values of K_m for thymine (30.7 μM (27.6 μM - 34.4 μM)(5% - 95%)) and for cytosine (25.5 μM (21.4 μM - 35 μM)) were statistically indistinguishable from the values obtained by analysis of the initial rates (27.6 μM and 26.4 μM for thymine and

cytosine respectively). A FACSIMILE output including the Michaelis-Menten analysis of the data is included in the Appendix.

Analysis of the data from the competition assays.

The results of the competition assay with dTTP and (S_p)-dTTPαS show that the assumption underlying the use of the phosphorothioate data, i.e. the assumption that the substitution by sulphur would affect only the forward and backward rates of phosphodiester bond formation during polymerisation (i.e. only k_7 and k_8), was incorrect. Thus, one would expect that the rate constants from the calculations constrained by the phosphorothioate data would also be incorrect. There was no reason to suppose that the kinetic mechanism was incorrect, and the data from the thymine/cytosine competition should allow us to calculate the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. This was duly performed but the model did not fit the data (Figures 3.29 and 3.30). The predicted outcome for the incorporation of cytosine was close to that observed, but the predicted outcome for the incorporation of thymine and of total incorporation, was significantly less than that observed, particularly in the competition experiment using 20 μM dTTP and 80 μM dCTP. The polymerisation reaction scheme was therefore re-examined again and it was subsequently discovered that if an extra step involving dNTP-exchange in the enzyme.DNA.dNTP complex was included into the reaction scheme, then a satisfactory fit to the observed data could be obtained (Figures 3.31, 3.32 and 3.33).

Thus, the reaction scheme shown in Figure 3.31 was eventually used to solve the mechanistic kinetic schemes for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. In order to recalculate the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand without using the phosphorothioate-effect as a constraint, data from the thymine/cytosine competition assays were used. This was done by analysing three sets of data in one FACSIMILE program: pre-steady state data for the incorporation of thymine alone and of cytosine alone opposite O^6-meG in the template strand (Figures 3.16 and 3.17) and pre-steady state data for the competition assays where both thymine and cytosine were incorporated opposite O^6-meG in the template strand.

In order to solve the internal rate constants, the values of K_m for the incorporation of thymine and cytosine were used as a constraint. The analyses of K_m and V_{max} all showed that the K_m for incorporation of thymine opposite O^6-meG in the template strand is very similar to that for the incorporation of cytosine, but the V_{max} for the incorporation of thymine is approximately 4-5 times that for the incorporation of cytosine (Figures 3.18, 3.19 and 3.20 and

Table 12). Thus, a constraint was applied by fixing the K_m for incorporation of thymine as equal to the K_m for incorporation of cytosine, and it was then possible to calculate the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand (Table 13).

The relevant parts of the output file from the mathematical analysis of the three sets of data including the thymine incorporation data, the cytosine incorporation data and the thymine/cytosine competition data is shown below. The code becomes quite complicated since the reaction schemes for thymine incorporation and cytosine incorporation were included. The enzyme is still denoted by **e** and after conformational change it is denoted by **ei**; the primer that is elongated by T is denoted by **d1** while that elongated by C is denoted by **d2**. A suffix t or c following the parameters $\mathbf{k_3}$-$\mathbf{k_{14}}$, **km** and **vm** was used to distinguish the corresponding parameters for thymine and cytosine incorporation respectively. Note that in this analysis,

- an extra step, that of dNTP-exchange with rate constants k_{13} and k_{14}, was included in the reaction scheme of thymine and of cytosine incorporation.
- K_m for the incorporation of thymine was set to equal that for the incorporation of cytosine, and this was used as a constraint to determine the internal rate constants, thus in the analysis k_5 was expressed in terms of K_m and the other rate constants.

Nine sets of data were analysed:

(1) T from (20 μM dTTP + 20 μM dCTP) experiment, denoted by T2020;
(2) C from (20 μM dTTP + 20 μM dCTP) experiment, denoted by C2020;
(3) T from (20 μM dTTP + 80 μM dCTP) experiment, denoted by T2080;
(4) C from (20 μM dTTP + 80 μM dCTP) experiment, denoted by C2080;
(5) 20 μM dTTP single nucleotide experiment, denoted by T2000;
(6) 80 μM dCTP single nucleotide experiment, denoted by C0080;
(7) 80 μM dTTP single nucleotide experiment, denoted by T8000;
(8) 20 μM dCTP single nucleotide experiment, denoted by C0020;
(9) 5 μM dTTP single nucleotide experiment, denoted by T0500;

A complete FACSIMILE output from this analysis is included in the Appendix.

```
param<#10> ws;
varia<#1> e ed edt eidt eid1pp ed1pp ed1;
varia<#1>  edc eidc eid2pp ed2pp ed2;
varia<#1> d0 1e-7 1e-7 1e-7 1e-7 1e-7 1e-7 1e-7;
param<#1> e0 1.2e-7  1.2e-7  1.2e-7  1.2e-7  1.2e-7  1.2e-7  1.2e-7;
varia<#1> d d1 d2 ttp ctp pp;
```

```
param<#1> t0 20e-6 20e-6 20e-6 00e-6 80e-6 00e-6 05e-6;
param<#1> c0 20e-6 80e-6 00e-6 80e-6 00e-6 20e-6 00e-6;
param  kmt 3.08e-5 vmt 2.63e-7 kmc 2.55e-5 vmc 4.84e-8;
param  k3t  1e7    k4t  1000;
param  k5t  59.2   k6t  22.2;
param  k7t  3.79   k8t  1.71;
param  k9t  .251   k10t 1.48e-3;
param  k11t 1000   k12t 1e7;
param  k13t 1e7    k14t 1e7 ;
param  k3c  1e7    k4c  1000;
param  k5c  13.1   k6c  3.82;
param  k7c  .649   k8c  1.35;
param  k9c  .227   k10c 3.74e-4;
param  k11c 1000   k12c 1e7;
param  k13c 1e7    k14c 1e7;
param  keq 1e3;
compile initial;
vmc=(e0c*k5c*k7c)/(k5c+k6c+k7c);
kmc=((k5c*k7c)+(k4c*k6c)+(k4c*k7c))/(k3c*(k5c+k6c+k7c));
k10c=(k3c*k5c*k7c*k9c*k11c)/(kequ*k4c*k6c*k8c*k12c);
k5t=((k6t+k7t)*(k4t-k3t*kmc))/(k3t*kmc-k7t);
vmt=(e0t*k5t*k7t)/(k5t+k6t+k7t);
vmtvmc=vmt/vmc;
kmt=((k5t*k7t)+(k4t*k6t)+(k4t*k7t))/(k3t*(k5t+k6t+k7t));
k10t=(k3t*k5t*k7t*k9t*k11t)/(kequ*k4t*k6t*k8t*k12t);
array<#1> ws;
ed0=d0;
ttp=t0;
ctp=c0;
array;
**;
compile equations;
array<#1> ws;
%k3t%k4t      :ed +ttp=edt;
%k5t%k6t      :edt=eidt;
%k7t%k8t      :eidt=eid1pp;
%k9t%k10t     :eid1pp=ed1pp;
%k11t%k12t    :ed1pp=ed1+pp;
```

```
%k3c%k4c      :ed+ctp=edc;

%k5c%k6c      :edc=eidc;

%k7c%k8c      :eidc=eid2pp;

%k9c%k10c     :eid2pp=ed2pp;

%k11c%k12c    :ed2pp=ed2+pp;

%k13%k14      :edc+ttp=edt+ctp;

%kexo:edt=ed0;

%kexo:edc=ed0;

array;

compile resid;

array<#1> ws;

xdt=100*(eid1pp+ed1pp+ed1)/d0;

xdc=100*(eid2pp+ed2pp+ed2)/d0;

array;

**;
data .075;
time    t2020   c2020   t2080   c2080   t2000   c0080   t8000   c0020   t0500;
range    49.8     9.6     39.2    23.7    60.7    42.0    70.3    31.9    33.8;
.40      0.00     0.0     0.00    0.00    20.7    0.00    44.3    0.00    11.8;
.50     19.8      2.6     15.8    9.95    0.00    0.00    0.00    0.00    0.00;
.80     0.00      0.0     0.00    0.00    32.4    0.00    54.1    0.00    16.3;
1.0     27.3      5.0     23.8    13.9    0.00    19.2    0.00    10.7    0.00;
1.2     0.00      0.0     0.00    0.00    38.1    0.00    59.6    0.00    16.7;
1.6     0.00      0.0     0.00    0.00    45.5    0.00    61.1    0.00    20.4;
2.0     37.3      8.4     33.0    17.6    0.00    27.5    0.00    0.00    0.00;
3.0     0.00      0.0     0.00    0.00    53.8    29.0    65.3    18.6    28.0;
4.0     49.8      9.6     39.2    23.7    0.00    0.00    0.00    0.00    0.00;
5.0     0.00      0.0     0.00    0.00    60.7    34.8    70.3    27.3    33.8;
7.0     0.00      0.0     0.00    0.00    0.00    42.0    0.00    31.9    0.00;

**;
vary k6t k7t k8t k9t k5c k6c k7c k8c k9c k13

:vmt k5t k10t ke3t ke4t ke5t kmc vmc vmtvmc k10c ke3c ke4c ke5c ke6c ke7;

**;

begin;
```

3.5 DISCUSSION

Steady state (Figures 3.11-3.13) and pre-steady state (Figures 3.14-3.19) studies on the incorporation of thymine and cytosine opposite O^6-meG in the template DNA strand have been carried out. The results obtained were analysed mathematically to determine the rate constants in the reaction scheme shown in Figures 3.5 and 3.31 and thus to show how *E. coli* Klenow fragment chooses to incorporate thymine rather than cytosine opposite the promutagenic base O^6-meG in the template DNA strand.

There are two great advantages in studying this particular misincorporation. First, the rate of incorporation is quite fast but not extremely fast and similar for both T and C. One problem with studies comparing normal incorporation and mismatch incorporation is that the former is so fast that there are technical difficulties in following its progress, and the latter is very slow. Second, the results of this kinetic analysis can be compared with structural studies on DNA containing such mismatches (Kalnik et al., 1988a, b).

The DNA polymerase chosen for the studies in this thesis was the Klenow fragment of DNA polymerase I from *E. coli*. Klenow fragment was chosen for several reasons. First of all, it is a simple peptide which does not require accessory proteins, it has a high fidelity, its rate of incorporation is not too fast, and the cloned gene has been expressed to yield very pure enzyme. Second, results with it could be compared with results of previous studies on the incorporation of mismatched nucleotides (Kuchta et al., 1988; Singer et al., 1989; Dosanjh et al., 1991 and Eger & Benkovic, 1992). Third, many detailed structural studies have been carried out on it, including X-ray crystallography (Ollis et al., 1985; Joyce & Steitz, 1987; Joyce, 1989 and Beese et al, 1993), allowing kinetic data to be related to structural data.

At the beginning of this introduction, it was mentioned that a year or so after this Ph.D. project had been started, Singer et al. (1989) published a paper which addressed the same question of how O^6-meG directs the misincorporation of thymine during DNA synthesis, and that shortly afterwards, Dosanjh et al. (1991) published another paper on the effects of neighbouring bases on the incorporation of thymine and of cytosine opposite a template O^6-meG. In the investigations carried out by Singer et al. (1989) and Dosanjh et al. (1991), the assays were carried out using the steady state method devised by Boosalis et al. (1987), which used a large excess of primer-template DNA over polymerase (Klenow fragment or pol a from Drosophila). The apparent K_m and V_{max} were calculated (summarized in Table 16), and the apparent K_m for cytosine was found to be 5.4 times higher than that for thymine, while the relative V_{max} for the incorporation of thymine and for cytosine were of similar magnitude. This led to the conclusion that the discrimination against cytosine was largely based on K_m and not V_{max}.

Singer et al. (1989) also studied the effect of the 3'-flanking base on the incorporation of thymine and cytosine opposite O^6-meG in the template strand, and concluded that when the O^6-meG in the template strand is flanked by a cytosine on its 3' side, there was a 6-7 fold preference for the formation of an O^6-meG:T base-pair compared with O^6-meG:C. If the 3'-flanking base next to the O^6-meG is changed to a thymine, both the K_m and V_{max} were markedly affected so that in this case formation of the O^6-meG:T base-pair was only slightly favoured over O^6-meG:C.

Table 16. Kinetics of incorporation of dTTP or dCTP opposite a template O^6-meG 3' to C in an M13 DNA fragment, reported by Singer et al. (1989). The reactions were carried out with Klenow fragment using an enzyme:DNA ratio of 1:30 over 90-second incubation periods. *f* is the ratio of each misincorporation efficiency compared to that for the incorporation of a correctly matched base-pair (C opposite template G).

Incorporation event	Apparent K_m (μM)	Relative V_{max}	*f*
C opposite G	0.16 ± 0.06	3.1 ± 0.7	1
T opposite G	264 ± 181	0.4 ± 0.1	0.8×10^{-4}
C opposite meG	233 ± 29	3.0 ± 1.4	6.7×10^{-4}
T opposite meG	43 ± 10	3.7 ± 0.8	44×10^{-4}

A subsequent study of the kinetics of adding a nucleotide following a thymine or a cytosine base-paired with O^6-meG in the template strand was carried out by the same laboratory (Dosanjh et al., 1991). Again, similar steady state kinetics gel assays were employed. The investigators also repeated their experiments with the incorporation of thymine and cytosine opposite O^6-meG in the template strand using even smaller amounts of polymerase (Klenow) and longer reaction times. The results of these experiments differed from those reported by the same laboratory in 1989 (Table 17), and Dosanjh et al. (1991) suggested that the differences might be a reflection of an effect of the 3' flanking base.

Table 17. Kinetics of incorporation of dTTP or dCTP opposite a template O^6-meG 3' with a cytosine on both the 3' and 5' flanking sides, reported by Dosanjh et al. (1991). The reactions were carried out with Klenow fragment using an enzyme:DNA ratio of 1:3000 over a period of 0-30 minutes. *f* is the ratio of each misincorporation efficiency compared to that for the incorporation of a correctly matched base-pair (C opposite template G).

Incorporation event	Apparent K_m (μM)	Relative V_{max}	*f*
C opposite G	0.04 ± 0.03	12 ± 4	1
T opposite G	56 ± 20	2.4 ± 1.5	1.3×10^{-4}
C opposite meG	84 ± 12	3.2 ± 1.8	1.3×10^{-4}
T opposite meG	26 ± 4	2.7 ± 1.7	1.3×10^{-4}

Table 18. Kinetics of incorporation of dTTP or dCTP opposite a template O^6-meG 3' with a thymine on both the 3' and 5' flanking sides, reported by Dosanjh et al. (1991). The reactions were carried out with Klenow fragment using an enzyme:DNA ratio of 1:3000 over a period of 0-30 minutes. *f* is the ratio of each misincorporation efficiency compared to that for the incorporation of a correctly matched base-pair (C opposite template G).

Incorporation event	Apparent K_m (μM)	Relative V_{max}	*f*
C opposite G	0.02 ± 0.008	3.8 ± 1.6	1
T opposite G	37 ± 22	2.2 ± 0.2	3.2×10^{-4}
C opposite meG	52 ± 18	2.9 ± 0.5	3.2×10^{-4}
T opposite meG	45 ± 12	2.9 ± 0.6	3.2×10^{-4}

These kinetic studies by Singer et al. (1989) and Dosanjh et al. (1991) of the incorporation of thymine or cytosine opposite O^6-meG in the template DNA strand have used steady state conditions in which there is substantially more DNA than the DNA polymerase. The steady state incorporation of thymine or cytosine opposite O^6-meG in the template DNA strand, shown in Figure 11, illustrate the problem with this approach. There is a pre-steady state burst of incorporation of thymine but not of cytosine. Under steady state conditions almost all the enzyme molecules are bound to DNA (K_d = 5 nM, Kuchta et al., 1987). The elongation during the pre-steady state burst is largely that of the DNA to which the enzyme is initially bound. Thus the occurrence of this pre-steady state burst indicates that the steady state incorporation of thymine is predominantly limited by a post-synthetic event such as the dissociation of the enzyme from the DNA and its binding to another substrate DNA duplex, rather than being limited by the

rate of incorporation itself. With cytosine however, there is no appreciable pre-steady state burst so the steady state rate of incorporation is limited predominantly by the rate of incorporation itself. Thus, a comparison of the steady state incorporation of thymine with that of cytosine would be a comparison of two different things. This can be seen very clearly from Figures 3.13, 3.18 and 3.19, which show the big difference between the Michaelis-Menten analyses of the ***steady state*** and ***initial rates*** of incorporation of thymine and of cytosine opposite O^6-meG in the template strand.

This is not the only problem with steady state kinetics. The establishment by Benkovic and by Johnson, and their respective colleagues, that the incorporation of a single nucleotide by Klenow fragment or T7 DNA polymerase involves five kinetically discernible steps (Fig. 3.5, steps 2-6; Kuchta et al., 1987, 1988; Patel et al., 1991; Wong et al., 1991; Dahlberg & Benkovic, 1991; Eger & Benkovic, 1992) has refined the question of misincorporation. ***One needs to ask how the ten rate constants differ so that thymine is preferentially incorporated.*** To answer this question one needs to carry out pre-steady state kinetic studies, i.e. start the reaction with all the DNA present as a enzyme-DNA complex and follow one catalytic cycle. The results are shown in Tables 12 and 13.

The analyses described in this thesis showed that the conformational changes play an important part in the reaction scheme for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. ***The second conformational change which occurs after the formation of the phosphodiester bond is particularly important because its slowness forces the products of elongation to accumulate, which explains the biphasic progress curves observed in following a catalytic cycle of polymerisation*** (Figure 3.36). The conformational change that occurs just before phosphodiester bond formation also improves the fitting of the predicted curves to the observed data, and the very recent report by Beese et al. (1993) which provided structural evidence for a conformational change upon binding of double-stranded DNA to Klenow fragment also corroborate the importance of the conformational changes in the polymerisation reaction scheme. One of the regions of Klenow fragment involved in the conformational change described by these authors is a flexible region that is partially disordered before binding duplex DNA, but this becomes ordered as duplex DNA is bound (Figure 1.18). The crystal of Klenow complexed with duplex DNA studied by these authors was formed by allowing Klenow to incorporate 2',3' riboepoxy adenosine triphosphate onto the 3' end of the primer strand, so the Klenow fragment in the resulting crystal complex had already catalysed an incorporation, and it is very tempting to say that this conformational change observed by Beese et al. corresponds to the conformational change induced by the binding of dNTP proposed for the mechanism of polymerisation by Klenow fragment. Furthermore, two residues from the flexible sub-domain that is involved in the conformational change proposed by Beese et al., Ser^{582} and

Asn^{579}, are either invariant (Ser^{582}) or highly conserved in the six DNA polymerases comprising the Pol I polymerase family (Beese et al., 1993), another indication of the importance of the conformational change. The similarity between the mechanism of polymerisation by Klenow fragment and that for T7 DNA polymerase in that both involve conformational changes (Kuchta et al., 1987, 1988, Patel et al., 1991, and Wong et al., 1991) could perhaps also be explained by the homology between the two polymerases in the flexible sub-domain that is proposed to be involved in a conformational change by Beese et al. (1993).

The initial rate of incorporation of thymine and of cytosine, i.e. that in the first 200 ms (T) or 1 s (C) followed Michaelis-Menten kinetics. Figure 3.20 shows that although the actual rates of elongation are different, thymine and cytosine have similar K_ms (27.6 μM and 26.4 μM respectively), and these K_ms are similar to that (20 μM) reported for the incorporation of thymine opposite adenine (Dahlberg & Benkovic, 1991). In this complex reaction scheme, the values of K_m do not represent the dissociation constant for the binding of dNTP. However they form a useful check to the calculations below.

The pre-steady state progress curves (Figs. 3.24 and 3.25) can be regarded as the integration of the differential equations implied by Figure 3.5. The objective of the analysis of the data is to assign values to the rate constants in Figure 3.5. This was carried out by using the FACSIMILE program which is a powerful tool in analysing systems dependent on differential equations (Chance & Curtis, 1970; Curtis & Chance, 1972; Chance et al., 1977; Curtis & Sweetenham, 1988). It works either to predict the outcome of mixing any initial concentrations of reactants in a system for which the rate constants are already known, or, otherwise, discover the rate constants if one already knows the time course of appearance of the product. In the second mode, the parameter-fitting mode, it is possible to enter the data as an array, in this case a table with columns showing the elongation of the primer at different times with each column being results with each concentration of dNTP or (S_p)-dNTPαS. The reaction scheme (Figure 3.5) was also entered. The computer then calculated the set of rate constants that gave the least residual sum of squares fit to the data. The adequacy of the fit was assessed in several ways. The total residual sum of squares, i.e. the fit of all the data points to their respective curves was compared with the expected range of the residual sum of squares calculated from the degrees of freedom and the estimated standard deviation of the data points. The residual sum of squares fit to each individual curve, in this case the curve corresponding to each dNTP concentration, was also calculated. A calculation showing whether the residuals to each curve were as likely to be positive as negative, and other calculations which also detect systematic errors, were made. The ratio of the forward and the backward rate constants was calculated to allow the free energy change for each step to be estimated, and the K_m and V_{max} were also calculated. Comparison of this K_m with that obtained with a Michaelis-Menten analysis of the initial rates (Figures 3.18 and

3.19) gave a check that the calculated rate constants were not seriously in error. The 5% and 95% confidence limits to all of these parameters were calculated. As will be discussed below, this showed that although the forward and backward rate constants for some steps could not be determined with satisfactory accuracy, their ratio, and hence the Gibbs' free energy change for that particular step, could be accurately determined. Furthermore, the computer calculates the correlation factor between the different calculated parameters. This was particularly useful in this case because to some extent it is possible to get an equally good fit by reciprocally changing two rate constants in the scheme - e.g. compensating an increase in a forward rate constant with an increase in a backward rate constant, or compensating a decrease in the forward rate to one step by increasing the forward rate to the following step, and in assessing any result it is important to know the extent to which this internal compensation is possible.

The main problem with analysing a system as complex as this is that one measures only the end-product (i.e. the sum of $Ei.D_1.PP_i$, $E.D_1.PP_i$, $E.D_1$ and free D_1). Thus, the overall kinetics can be satisfied by a range of values for the internal rate constants. This is because an increase in any forward rate constant can usually be offset by a decrease in another forward rate constant or an increase in a backward rate constant. Two constraints were initially applied to overcome this problem. The first was to constrain the overall kinetic solution to the thermodynamics of the process. The literature gives conflicting guidance on this point. Patel et al. (1991) measured by equilibrium pyrophosphorolysis the ΔG for the incorporation of dTMP into DNA by T7 DNA polymerase as -3.97 kcal/mol. This result is roughly what one would expect for the formation of a phosphodiester bond during DNA polymerization. However, other workers reported rather different results. Calculation from the data of Dahlberg & Benkovic (1991) (see Table 19) for normal incorporation of thymine opposite adenine suggests that the overall ΔG is about -4.75 kcal/mol. Calculation from the data of Eger & Benkovic (1992) (Table 19) for the misincorporation of adenine opposite adenine suggests a ΔG of about -5.51 kcal/mol. Although these values of ΔG are different, a large part of the difference is in the value given to equilibrium constant for the conformational change that occurs after phosphodiester bond formation ($Ei.D_1.PP_i \rightleftharpoons E.D_1.PP_i$), i.e. k_9/k_{10} in Figure 3.5 (see Table 19). The effects of assuming different values for the overall equilibrium constant were investigated, and it was found that of all the rate constants in the reaction scheme in Figure 3.5, only k_{10} is sensitive to the magnitude of the overall equilibrium constant. The calculated value of all the other rate constants remained unaffected by changes in the assumed overall equilibrium constant. Thus, it was considered justifiable to constrain the solution to an overall equilibrium constant of 1000 implying a ΔG of -4 kcal/mol.

The second constraint was to compare the rates of incorporation of regular dNTP and their phosphorothioate analogues (S_p)-dNTPαS into DNA. Because sulphur is less

electronegative than oxygen, incorporation of (S_p)-dNTPαS into DNA is slower. The thio-substitution alters the rate constants for the chemistry step (i.e. k_7 and k_8) and does not materially affect the other rate constants (Benkovic & Schray, 1971; Eger & Benkovic, 1992). The exact magnitude of this thio effect on rate constants is not known but it has been assumed that the formation of a phosphothiodiester bond using (S_p)-dNTPαS by DNA polymerase and its hydrolysis is approximately 30-100 times slower than the formation or hydrolysis of the normal phosphodiester bond (Bryant et al., 1983, Mizrahi et al., 1985, Kuchta et al. 1988, Patel et al. 1991, Wong et al., 1991).

Comparison of these rates in studies of DNA synthesis has proved to be an invaluable tool which, for example, confirmed the existence of the conformational change before polymerization (Figure 3.5, step 3) and showed that in normal incorporation this conformational change, rather than the chemical step, is rate limiting. If during polymerization the formation of a phosphodiester bond between the 3'-OH group of the deoxyribose sugar at the primer terminus and the α-phosphate of the incoming dNTP is rate-limiting, then the rate of polymerization using the phosphorothioate analogue of the regular deoxynucleotide triphosphate, (S_p)-dNTPαS, should decrease by approximately 30-100-fold. In normal nucleotide incorporation, the phosphorothioate nucleotide was incorporated only 3-7 fold slower than the normal (Kuchta et al., 1987; Patel et al., 1991). In the formation of mismatches, the ratio is higher but still much smaller than 100. For example, for the formation of A.A mismatches by Klenow, the ratio is about 65 (Kuchta et al., 1988), and with T7 DNA polymerase the ratios for the formation of A.A, C.A and G.A mismatches by incorporating adenine, cytosine or guanine opposite adenine in the template DNA strand are 19, 17 and 37 respectively (Wong et al., 1991).

The value initially assumed for the thio-effect taken from the literature was 100. This figure is somewhat arbitrary, and thus the effect of changing the assumed value for the thio-effect was studied. This suggested that the best fit to the data could be obtained for an assumed thio-effect of 150 for both the incorporation of thymine and cytosine opposite O^6-meG in the template strand. The constants given in Table 12 were calculated assuming a thio-effect of 150.

The rate of incorporation of the corresponding (S_p)-dTTPαS and of (S_p)-dCTPαS (20μM) opposite template O^6-meG were at least 75 times slower than the incorporation of the regular nucleotides (Figure 3.23). This indicates that the formation of the phosphodiester bond is the rate limiting step. The objective of the mathematical analysis was to fit a combined data set of incorporation of both regular nucleotides and their phosphorothioate analogues by a single set of rate constants differing only by the fact that k_7 and k_8 for the phosphorothioate nucleotide analogues are 150 times slower than k_7 and k_8 for the regular nucleotides. This constraint fixes the values of k_7 and k_8 to within very close limits.

The possibility of analysing the data for incorporation of the phosphorothioate nucleotide analogues as a single data set with the data for incorporation of normal nucleotides illustrates clearly the power of FACSIMILE. Previous workers have either used the phosphorothioate data just as a rule-of-thumb diagnostic tool to give an idea of the magnitude of the rate of the polymerization step relative to the others (Kuchta et al., 1988), or it has been applied to pre-steady state data using a mathematical formula derived from steady state kinetics (Wong et al., 1991). In the present study, FACSIMILE allows the results of incorporation of both normal nucleotides and their phosphorothioate analogues (e.g. the data in Figures 3.16, 3.17 and 3.23) to be analysed as a single set of data.

The principle behind the thio-substitution effect on the rate-limiting step during polymerisation is that substitution of one of the non-bridging phosphoryl oxygen atoms by a sulphur atom will slow down the rate of phosphodiester bond formation because thiophosphodiesters are chemically less reactive than phosphodiesters, and if the rate of phosphodiester bond is rate-limiting during polymerisation, then this thio-substitution will dramatically reduce the rate of polymerisation. However, this slower incorporation of the dNTPα S is deduced from studies with phosphate triesters (Benkovic & Schray, 1971) and a recent report by Herschlag et al. (1991) has shown that the thio-substitution has different effects on phosphate triesters and phosphate diesters. In the discussion of their paper they asserted that thio-substitution reduces the rates of reactions with phosphate triesters by 30-100-fold while their work has shown that with phosphate diesters, thio-substitution only slows the rate of reactions by 4-11-fold. In this light, the slow incorporation of the phosphorothioate analogues could be due to several possibilities: slow binding of the dNTPαS, or slow conformational change of the ternary complex induced by the dNTPαS, or a slow rate of phosphodiester bond formation, or a combination of any of the three. Competition experiments with dTTP and dTTPα S were therefore designed to see how thio-substitution affects polymerisation, and the results showed that the dTTPαS does not act as a competitive inhibitor of dTTP. This means that the slow incorporation of dTTPαS could not be due to the slower rate of formation of the thiophosphodiester bond alone, contrary to the assumption made by Mizrahi et al.(1985), Kuchta et al. (1987, 1988), Patel et al. (1991) and Wong et al. (1991).

Since the application of the phosphorothioate data as a second constraint was questionable, the rate constants had to be recalculated without using the phosphorothioate data. This was where the data from the thymine/cytosine competition assays came into use. These competition assays were originally carried out to see if the rate constants calculated using the phosphorothioate constraint could be used to predict the outcome of different experiments with thymine and cytosine. If the observed results are close to that predicted, then the calculated parameters could not be substantially in error. However, the experimental results of the

competition assays were rather different from the predicted progress curves, and were not compatible with the reaction scheme including the two conformational changes shown in Figure 3.5. After several attempts to fit the curves by altering the reaction scheme for the polymerisation pathway, it was found that the data could be fitted only if an extra step of dNTP-exchange which occurs before the first conformational change step was included. This dNTP-exchange step was kinetically invisible in the pre-steady state experiments using only a single type of nucleotide because under those circumstances, an exchange would result in the same type of product being formed. Only when two types of nucleotides were present could this dNTP-exchange step be identified.

The use of competition assay is not a new concept. As shown by Eger & Benkovic and by the analyses described in this thesis, data from competition assays could provide a wealth of information with the use of a powerful analytical method. Unfortunately, the usefulness of data from competition assays are frequently underestimated because of the lack of appropriate analyses.

The data from the thymine/cytosine competition assays were used with the data for incorporation of thymine and of cytosine to recalculate the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand. Since the phosphorothioate data could not be used as a constraint for determining the values of the rate constants, another constraint had to be applied in order to calculate the values of the internal rate constants. Thus, the values of K_m was used as a constraint. As mentioned earlier, the K_m for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand have been determined by three independent methods (Figures 3.18, 3.19, 3.20 and Table 12). The K_m was calculated from the rate constants and compared with the approximate obtained from the Michaelis-Menten analysis of initial rates (Figure 3.18, 3.19 and 3.20). The calculated K_m for thymine (30.7 μM (27.6 μM - 34.4 μM)(5% - 95%)) and for cytosine (25.5 μM (21.4 μM - 35 μM)) are statistically indistinguishable from the values obtained by Michaelis-Menten analysis of the initial rates (30-35 μM for thymine and cytosine from Figures 3.18 and 3.19, and 27.6 μM and 26.4 μM for thymine and cytosine respectively from Figure 3.20). The reasonable correspondence between the actual and the predicted K_ms leads to the conclusion that these values cannot cannot be substantially in error, and it is therefore reasonable to conclude from the results of the Michaelis-Menten analyses that the K_m for thymine and for cytosine are the same. Thus, in the analyses to recalculate the rate constants based on data from the single nucleotide and the competition experiments, the solution was constrained by setting the K_m for the incorporation of thymine opposite O^6-meG as equal to that for the incorporation of cytosine.

Since the data from the thymine/cytosine competition assays were too little to solve the rate constants for the incorporation of thymine and of cytosine opposite O^6-meG in the template strand, and since the data from the competition experiments must be compatible with that from the single nucleotide experiments, it was necessary to combine the data from the competition assays with the data from the single nucleotide experiments in order to recalculate the rate constants. Here, there are two methods of combining the three data sets. Because the three data sets were obtained from experiments performed on different days, there would be significant experimental errors between the three sets of data. One way of avoiding this is to analyse the data sets separately and see if the rate constants calculated could fit all the data sets. Data from the single-nucleotide experiments are analysed first to obtain as many rate constants as possible. Then, data from the thymine/cytosine competition experiments are analysed, using the constants that are well-defined in the initial calculations to calculate the other poorly defined constants. A circular strategy of computation is required to see if all the constants calculated could fit all three data sets, and this could present a problem if the rate constants do not fit some of the data.

Thus, in the analyses presented, the rate constants were calculated by a second method, which was to analyse the three sets of data in a single program. Since there were many more data points from the single nucleotide experiments than those from the thymine/cytosine competition experiments, some of the data points from the single nucleotide experiments were not included in the final analysis so that the data points form the single nucleotide experiments and the competition experiments were given equal weights in determining the rate constants. Again, this illustrates the strength of FACSIMILE in the analysis of kinetic problems because it enables the user to link different sets of data to one kinetic scheme and analyse different sets of data in one program.

Apart from the rate constants for the nucleotide exchange steps, the rate constants calculated by using the competition data and the single-nucleotide data constrained by the same K_m for both incorporation of thymine and cytosine opposite O^6-meG in the template strand, shown in Table 13, are not significantly different from those calculated from the single-nucleotide data constrained by the phosphorothioate-effect, shown in Table 12. In both sets of calculations, the V_{max} for the incorporation of thymine is approximately 5.3 times that for the incorporation of cytosine, which is similar to the value of 4.8 obtained from Michaelis-Menten analysis of the initial rates (Figures 3.18 and 3.19). The rate constants for the forward and backward steps of the formation of the phosphodiester bond, k_7 and k_8 in Figure 3.5, are well-determined and are of similar magnitude in both calculations. Similarly for k_9, the values determined from the two calculations are of similar magnitude, although k_9 for the incorporation of cytosine is less well-determined in the analysis using data from both single-nucleotide and competition experiments.

The biggest difference in the value of the rate constants obtained from the two calculations comes from k_5 and k_6, the forward and backward rate constants for the first conformational change induced by dNTP-binding. These rate constants are rather poorly determined in both calculations, as can be seen from the widely differing 5% and 95% confidence limits. However, the range of the confidence limits do show that the rate of the conformational change is faster than the rate of phosphodiester bond formation. The ratio of k_5:k_6 remains the same and is very well determined. This could perhaps be explained by the fact that progress curves are in general limited by the slow steps in a reaction pathway, and in this case, the rate of the first conformational change is fast relative to the formation of the phosphodiester bond, thus the data points provided information for determining the rate of phosphodiester bond formation but not for the rate of the conformational change except that it is faster than the rate of the phosphodiester bond formation. On the other hand, the ratio of k_5:k_6 is well-determined from both calculations because it is related to the flux of the reactants through the reaction pathway. Hence, it is possible to obtain a good fit to the observed data by putting in different values for k_5 and k_6 but keeping the ratio of k_5:k_6 constant.

From the mathematical analyses described, ***the discrimination against the incorporation of cytosine opposite O^6-meG in the template strand occurs primarily at the formation of the phosphodiester bond***. The K_m and V_{max} for the incorporation of thymine and of cytosine has shown that the ***K_m is similar for the incorporation of thymine and cytosine, but the V_{max} for the incorporation of thymine is approximately 5.3 times that for the incorporation of cytosine***. The biggest difference between the incorporation of thymine and cytosine opposite O^6-meG in the template strand lies in the respective rates of formation of the phosphodiester bond, k_7, which are 3.87 s^{-1} and 0.73 s^{-1} for the incorporation of thymine and cytosine respectively. The ratio of k_7T:k_7C is 5.3, the same as the ratio of $V_{max}T$:$V_{max}C$.

There are two other secondary factors that also favour the incorporation of thymine but discriminate against the incorporation of cytosine opposite O^6-meG in the template strand. The first is the nucleotide exchange reaction that favours the exchange of a dCTP that is bound to an enzyme.DNA complex with a dTTP (rate for this exchange, i.e. E.D.dCTP + dTTP $\rightarrow$ E.D.dTTP + dCTP was calculated at 1.9×10^7 $M^{-1}s^{-1}$, while the rate for E.D.dTTP + dCTP $\rightarrow$ E.D.dCTP + dTTP was calculated at 1×10^7 $M^{-1}s^{-1}$). The reason that the amount of thymine incorporated in the competition experiments were higher than expected from the reaction scheme in Figure 3.5 was not because the rate constants used were wrong, but because of the nucleotide exchange step was absent from the reaction scheme in Figure 3.5. The second discriminating mechanism comes from the 3' to 5' exonuclease activity of Klenow fragment, which is twice as fast in the removal of a cytosine base-paired with an O^6-meG in the template strand than the removal of a thymine base-paired with an O^6-meG in the template strand (Figure 3.10).

After it was realised that O^6-meG does not form a more stable base pair with thymine than with cytosine (Gaffney & Jones, 1989), the discussion of the misincorporation of thymine opposite O^6-meG focused on two possibilities:

1. discrimination through K_m possibly because of the structural similarities between O^6-meG and adenine (an analogous situation would be the recognition of O^6-meG by adenosine deaminase (Pegg & Swann, 1979);
2. that the discrimination occurred because the O^6-meG:T base pair retains the Watson-Crick configuration (where N1 of the purine is juxtaposed to N3 of the pyrimidine) while the O^6-meG:C base pair is a wobble pair (Patel et al., 1986a, b; reviewed by Swann, 1990).

These possibilities have been discussed in the wider context of miscoding by Echols and Goodman (1991). ***The results reported now showing that the discrimination occurs at the formation of the phosphodiester link support the view that the Watson-Crick conformation of the O^6-meG:T base-pair is the important factor in ensuring the incorporation of thymine opposite O^6-meG. The reason for this is probably the distortion of the phosphodiester link which occurs on the formation of the wobble O^6-meG:C base-pair***. The presence of an O^6-meG:C base pair perturbs the ^{31}P-NMR spectrum. The phosphorus resonances of dodecamer DNA duplexes containing either O^6-ethylG:T or O^6-ethylG:C base-pairs have been assigned (Kalnik et al., 1988a, b). This assignment showed that the greatest change from the spectrum of the control dodecamer with a G:C base pair, was in the chemical shift of the phosphodiester link 5' to the cytosine in the O^6-ethylG:C base pair (moved 0.59 ppm upfield) and in that of the phosphodiester link 3' to the cytosine (moved 0.5 ppm downfield; see Table 5, page 41). The possible relation between such chemical shifts and the conformation of the phosphodiester has been discussed by Gorenstein et al. (1988). It seems reasonable to ascribe the slow incorporation of cytosine opposite O^6-meG to the stereochemical problems of aligning the 3'-OH of the primer strand with the α-phosphorus of the nucleoside triphosphate.

Furthermore, recent X-ray crystallography studies of Klenow fragment complexed with double-stranded DNA has shown that the interaction between the polymerase and DNA is mainly with the phosphate backbone of the DNA, particularly with the primer strand. This is also in agreement with the results reported now which shows that the discrimination against the incorporation of cytosine opposite O^6-meG occurs at the formation of the phosphodiester link.

If a distortion of the phosphodiester backbone impairs the incorporation of cytosine opposite O^6-meG because it presents problems in aligning the 3'-OH of the primer with the α-phosphorus of the incoming dCTP, one might expect the addition of a correctly paired nucleotide after a cytosine in a C:O^6-meG base pair to be slower than the corresponding addition after thymine in an O^6-meG:T base pair. Figure 3.35 shows that in steady state polymerisation experiments, addition of the nucleotide after the cytosine in an O^6-meG:C base pair is slower

than that after a thymine in a T:O^6-meG base pair. The latter rate appears to be very similar to the addition after a thymine in a regular A:T base pair, but since steady state experiments do not necessarily reflect the true kinetics of polymerisation, one should not assume that the two polymerization rates are the same.

The observation that misincorporation can occur following the cytosine in an O^6-meG:C base-pair but not after the thymine in an O^6-meG:T base-pair during the pre-steady state experiments could perhaps be explained by the distortion of the phosphodiester backbone as well, as this distortion could favour the formation of a mismatched base-pair following the O^6-meG:C base-pair rather than a correctly matched base-pair.

The approximate Gibbs' free energy changes for the polymerization of thymine and cytosine opposite O^6-meG in the template DNA strand are shown in Table 19. The free energies were calculated from the respective rate constants for the steps involved in the incorporation of these nucleotides opposite template O^6-meG. The free energy changes for the incorporation of the normal nucleoside triphosphates by Klenow fragment (Dahlberg & Benkovic, 1991) and T7 DNA polymerase (Patel et al., 1991) are also given for comparison. In the incorporation of thymine/cytosine opposite O^6-meG in the template strand, ***the major part of the overall Gibbs' free energy change for the process comes from the relaxation step*** (Figure 3.5, step 5) where the enzyme in the $Ei.D_1.PP_i$ complex changes from an 'energized' state to a 'relaxed' state, after the phosphodiester bond has formed. ***The significance of this is that all the steps before this are freely reversible, but once this enzyme relaxation step has occurred, the incorporation is committed and will be reversed only by action of 3'→5' exonuclease or by mismatch repair***. Eger & Benkovic (1992) also found that for misincorporation by Klenow fragment, the greatest part of the energy change is in the conformational change after polymerisation. By contrast, Dahlberg and Benkovic (1991) reported that for correct incorporation by Klenow fragment this step is isoenergetic, and that the greatest part of energy change is in the conformational change prior to polymerisation. This might suggest a major difference between the kinetics of misincorporation and correct incorporation, but it should also be remembered that Patel et al. (1991) reported that for correct incorporation by T7 DNA polymerase, the greatest part of the energy change is in the conformational change after polymerisation has occurred. In the present report, and in the report by Eger & Benkovic (1991) in which misincorporaiton by Klenow was examined, this second conformational change is the slowest forward step in the kinetic pathway. One interesting aspect of the energetic profile of incorporation opposite O^6-meG is that the formation of phosphodiester bond during the incorporation of cytosine is endothermic (ΔG = + 0.14 kcal/mol). Patel et al. (1991) suggested that with T7 DNA polymerase this step is also endothermic (ΔG = + 0.40 kcal/mol) for the incorporation of thymine opposite adenine.

Table 19. Gibbs' free energy changes (ΔG) for the incorporation by Klenow fragment of different nucleotides opposite O^6-meG, compared with ΔG calculated from previously published rate constants for the incorporation by Klenow fragment or T7 DNA polymerase of different nucleotides.

Reaction	ΔG (kcal/ mol)	[a] T opp. O^6-meG (Klenow)	[a] C opp. O^6-meG (Klenow)	[b] T opp. A (T7 DNA pol)	[c] T opp. A (Klenow)	[d] A opp. A (Klenow)
E.D + N ⇌ E.D.N	ΔG1	-5.43	-5.43	-6.45	-7.20	-5.12
E.D.N ⇌ Ei.D.N	ΔG2	-0.64	-0.50	-0.64	-1.67	-1.20
Ei.D.N ⇌ Ei.D1.PPi	ΔG3	-0.43	+0.14	+0.40	-0.82	-1.30
Ei.D1.PPi ⇌ E.D1.PPi	ΔG4	-3.03	-3.73	-2.48	0.00	-2.10
E.D1.PPi ⇌ E.D1 + PPi	ΔG5	+5.43	+5.43	+3.67	+4.94	+4.21
		-4.10	-4.09	-5.50	-4.75	-5.51

All ΔG values were calculated according to the equation $\Delta G = -RT\ln K_{eq}$, where R is 1.987 x 10-3 kcal/mol (gas constant), T is 298 K, and K_{eq} is the equilibrium constant of the individual reaction. [a] See Table 20. [b] Rate constants were taken from Patel et al. (1991). [c] Rate and equilibrium constants were taken from Dahlberg & Benkovic (1991). [d] Rate constants were taken from Eger & Benkovic (1992).

The overall conclusion from this, and previous studies by others, is that the DNA polymerase has evolved to catalyse the formation of phosphodiesters only when the 3'-OH of the primer strand and the α-phosphorus of the incoming dNTP are in the alignment found in Watson-Crick pairs. Any deviation from this will have a very marked effect on the rate of synthesis. Because the greatest energy change occurs after the formation of the phosphodiester bond, the reverse rates of the early steps are relatively fast, allowing a mismatched base to be replaced without the need for the 3'→5' exonuclease. It has already been reported (Singer et al., 1989; Dosanjh et al., 1991) that the rate of incorporation of thymine or cytosine opposite O^6-meG depends to some extent on the flanking base pairs. Since it is known (Georgiadis et al., 1991) that the conformation of the O^6-meG containing base-pairs depends on the flanking base pairs, this would be expected.

APPENDIX

Appendix 1.

FACSIMILE analysis of the pre-steady state incorporation of T oppoiste O^6-meG in the template strand. The solution is constrained by (1) setting the overall equilibrium constant (keq) for polymerisation as 1000, and (2) the fixing the forward and backward rates for the formation of phosphodiester bond at a value 150 times slower for the incorporation of phosphorothioate nucleotide analogue than for the incorporation of the normal, oxygen-containing nucleotides. K_m and V_{max} for the incorporation of T are also analysed.

```
FACSIMILE OUTPUT ON  5/ 8/1993 AT 14:49:27
*          TTP580.fac;
punch 6;
compile instant;

 LOCN   INSTRUCTION

   74:#1=6;
   77:#2=5;
   80:#20=5*#1;
   89:**;

NO. OF VARIABLES =     0
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE     TOTAL
INTEGER SCALARS AND ARRAYS             67     19933     20000
REAL SCALARS AND ARRAYS                 6     19994     20000
INDEXES AND SECRETS                    99       400       499
CODE ARRAY                             70      9930     10000
NAME DICTIONARY                        29       971      1000
TEXT ARRAY                             80      4920      5000


varia<#1> ed    :<#2> edx;
varia<#1> edt   :<#2> edtx;
varia<#1> eidt  :<#2> eidtx;
varia<#1> eid1pp:<#2> eid1ppx;
varia<#1> ed1pp :<#2> ed1ppx;
varia<#1> ed1   :<#2> ed1x;
varia<#1> ttp pp;
param<#1> s0 5e-6 10e-6 20e-6 40e-6 80e-6 40e-6:s5 s10 s20 s40 s80 t40;
param<#1> kcf kcb rks4 1 1 1 1 1 150:rk1 rk2 rk3 rk4 rk5 rk6;
```

```
param<#20> ws;
param k3 1e7 k4 1e3 k5 57.7 k6 22.2 k7 3.79 k8 1.67 k9 .245;
param k10 1e-3 k11 1e3 k12 1e7 keq 1e3 d0 1e-7 e0 1.2e-7;
param km 3.13e-5 vm 2.61e-7 kcat kmvm ks2 ks3 ks4 ks5 ks6 x y z;
param g2 g3 g4 g5 g6 r 1.987e-3 t 300;
compile initial;

 LOCN   INSTRUCTION

   74:x=1/k7;
   83:y=(1/k5)+(k6/(k5*k7));
  103:z=(1/k3)+(k4*y/k3);
  122:vm=e0/(x+y);
  135:km=z/(x+y);
  148:kmvm=km/vm;
  157:kcat=1/(x+y);
  170:k10=(k3*k5*k7*k9*k11)/(keq*k4*k6*k8*k12);
  204:ks2=k3/k4;
  213:g2=-r*t*log(ks2);
  228:ks3=k5/k6;
  237:g3=-r*t*log(ks3);
  252:ks4=k7/k8;
  261:g4=-r*t*log(ks4);
  276:ks5=k9/k10;
  285:g5=-r*t*log(ks5);
  300:ks6=k11/k12;
  309:g6=-r*t*log(ks6);
  324:array<#1> ws;
```

```
  327:ed0=d0;
  334:ttp=s0;
  342:kcf=k7/rks4;
  353:kcb=k8/rks4;
  364:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           1
                          ALLOWED HENCEFORTH   20

  366:#10=0;
  369:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS             74    19926     20000
REAL SCALARS AND ARRAYS               141    19859     20000
INDEXES AND SECRETS                    99      400       499
CODE ARRAY                            370     9630     10000
NAME DICTIONARY                        94      906      1000
TEXT ARRAY                             80     4920      5000


compile equation;

 LOCN   INSTRUCTION


*****    EQUATION FIXES MAX. NO. OF VARIABLES AS    48

  374:array<#1> ws;
  377:%k3%k4:   ed+ttp=edt;
  399:%k5%k6:   edt=eidt;
  417:%kcf%kcb: eidt=eid1pp;
  437:%k9%k10:  eid1pp=ed1pp;
  455:%k11%k12: ed1pp=ed1+pp;
  477:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  479:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS            159    19841     20000
REAL SCALARS AND ARRAYS               767    19233     20000
INDEXES AND SECRETS                    99      400       499
CODE ARRAY                            480     9520     10000
NAME DICTIONARY                       103      897      1000
TEXT ARRAY                             80     4920      5000
```

```
param<#1> ninemer:nm5 nm10 nm20 nm40 nm80 nmth;
param<#1>  tenmer:tm5 tm10 tm20 tm40 tm80 tmth;
compile resid;

 LOCN   INSTRUCTION

  484:array<#1> ws;
  487:ninemer=100*(ed+edt+eidt)/d0;
  510:tenmer= 100*(eid1pp+ed1pp+ed1)/d0;
  533:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  535:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                              IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS                159     19841      20000
REAL SCALARS AND ARRAYS                   780     19220      20000
INDEXES AND SECRETS                        99       400        499
CODE ARRAY                                536      9464      10000
NAME DICTIONARY                           118       882       1000
TEXT ARRAY                                 80      4920       5000



data .035;
time   nm5  tm5 nm10 tm10 nm20 tm20 nm40 tm40 nm80 tm80 nmth tmth;
range 96.8 33.8 86.3 52.6 84.7 60.7 76.5 69.6 74.8 70.3 85.4 14.6;
 0.2  96.8  3.2 86.3 13.7 84.7 15.3 76.5 23.5 74.8 29.2  0.0  0.0;
 0.4  88.2 11.8 82.3 17.7 79.3 20.7 66.3 33.7 55.7 44.3  0.0  0.0;
 0.8  83.7 16.3 75.7 24.3 67.6 32.4 55.7 44.3 45.9 54.1 99.5  0.5;
 1.2  83.3 16.7 68.1 31.9 61.9 38.1 45.6 54.4 40.4 59.6 98.8  1.1;
 1.6  79.6 20.4 63.7 36.3 54.5 45.5 45.2 54.8 38.9 61.1  0.0  0.0;
 3.0  72.0 28.0 56.2 43.8 46.2 53.8 36.2 63.8 34.7 65.3 98.2  1.8;
 5.0  66.2 33.8 47.4 52.6 39.3 60.7 30.4 69.6 29.7 70.3 94.2  5.8;
10.0   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 89.8 10.2;
12.0   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 87.2 12.8;
15.0   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 85.4 14.6;
**;

   12 CURVES
   10 TIME       POINTS
   84 OBSERVATIONS


A TIME       POINT INCREMENT OF     1 AND OFFSET OF     0 SPECIFIED
    12 DEPENDENT VARIABLES SELECTED
    10 TIME       POINTS AND    84 OBSERVATIONS USED


vary k5 k6 k7 k8 k9:k10 ks3 ks4 ks5 km vm kmvm kcat;
begin;

SPACE FOR                              IN USE      FREE      TOTAL
REAL SCALARS AND ARRAYS                  2317     17683      20000
```

```
SIZE NEEDED FOR ARRAY WSFIT =      1083

OPTIMISATION RUN WITH    84 OBSERVATIONS AND   5 VARIED PARAMETERS

TIME  0.0000000E+00  TCPU      1.59      0.00
  STP      0  TRY      0  DER      0  JAC     0  LUD      0  RST     0

AFTER   0 ITERATIONS AND   1 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS        CORI      AUCO     AUCR
  1.5667E+00  -0.3843  -0.0837   0.3499
  1.2850E+01   0.3843  -0.0837   0.3499
  1.0607E+01  -2.1980   0.5693   0.3499
  2.8553E+01   2.1980   0.5693   0.3499
  5.7762E+00   0.9226   0.4827   0.3499
  1.1247E+01  -0.9226   0.4827   0.3499
  4.3900E+00  -1.4561  -0.0175   0.3499
  5.3036E+00   1.4561  -0.0175   0.3499
  5.6206E+00  -0.0528   0.2760   0.3499
  8.1891E+00   0.6146   0.4851   0.3499
  3.9705E-01  -0.2986  -0.0280   0.3499
  1.3658E+01   0.2448   0.0178   0.3499
TOTAL RESIDUAL SUM OF SQUARES IS      1.08159E+02
MEAN ABSOLUTE CORRELATION INDEX IS   0.92775
VARY(  1):K5          =  5.7700E+01
VARY(  2):K6          =  2.2200E+01
VARY(  3):K7          =  3.7900E+00
VARY(  4):K8          =  1.6700E+00
VARY(  5):K9          =  2.4500E-01
VARY(  6):K10         =  1.4451E-03
VARY(  7):KS3         =  2.5991E+00
VARY(  8):KS4         =  2.2695E+00
VARY(  9):KS5         =  1.6953E+02
VARY( 10):KM          =  3.1316E-05
VARY( 11):VM          =  3.1356E-07
VARY( 12):KMVM        =  9.9873E+01
VARY( 13):KCAT        =  2.6130E+00
        1 SIMULATION RUNS COMPLETED
        2 SIMULATION RUNS COMPLETED
        3 SIMULATION RUNS COMPLETED
        4 SIMULATION RUNS COMPLETED
        5 SIMULATION RUNS COMPLETED

AFTER   0 ITERATIONS AND   6 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS        CORI      AUCO     AUCR
  1.5466E+00   0.9079  -0.0452   0.3499
  1.2685E+01  -0.9079  -0.0452   0.3499
  6.2741E+00  -1.9127   0.4220   0.3499
  1.6889E+01   1.9127   0.4220   0.3499
  8.3223E+00   1.7028   0.6234   0.3499
  1.6204E+01  -1.7028   0.6234   0.3499
  3.2834E+00  -0.0476  -0.3469   0.3499
  3.9667E+00   0.0476  -0.3469   0.3499
  6.7427E+00   1.0553   0.2709   0.3499
  8.3541E+00  -0.4463   0.4260   0.3499
  3.2854E-01   0.7053  -0.2838   0.3499
  1.1353E+01  -0.7599  -0.2207   0.3499
TOTAL RESIDUAL SUM OF SQUARES IS      9.59498E+01
MEAN ABSOLUTE CORRELATION INDEX IS   1.00904
VARY(  1):K5          =  5.9161E+01
VARY(  2):K6          =  2.2200E+01
VARY(  3):K7          =  3.8859E+00
VARY(  4):K8          =  1.6700E+00
```

```
VARY(  5):K9          =  2.5120E-01
VARY(  6):K10         =  1.5577E-03
VARY(  7):KS3         =  2.6649E+00
VARY(  8):KS4         =  2.3269E+00
VARY(  9):KS5         =  1.6126E+02
VARY( 10):KM          =  3.0870E-05
VARY( 11):VM          =  3.2362E-07
VARY( 12):KMVM        =  9.5391E+01
VARY( 13):KCAT        =  2.6968E+00
        6 SIMULATION RUNS COMPLETED
        6 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS                95.9497512
COMPARED WITH EXPECTED RANGE   56.3019  105.4762
FOR   79 DEGREES OF FREEDOM
ANALYSIS OF VARIANCE PROCEEDS ON ASSUMPTION OF GOOD FIT
BUT INSPECT RESIDUALS FOR EVIDENCE OF SYSTEMATIC ERROR

AFTER   0 ITERATIONS AND   1 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  1.5466E+00   0.9079  -0.0452   0.3499
  1.2685E+01  -0.9079  -0.0452   0.3499
  6.2741E+00  -1.9127   0.4220   0.3499
  1.6889E+01   1.9127   0.4220   0.3499
  8.3223E+00   1.7028   0.6234   0.3499
  1.6204E+01  -1.7028   0.6234   0.3499
  3.2834E+00  -0.0476  -0.3469   0.3499
  3.9667E+00   0.0476  -0.3469   0.3499
  6.7427E+00   1.0553   0.2709   0.3499
  8.3541E+00  -0.4463   0.4260   0.3499
  3.2854E-01   0.7053  -0.2838   0.3499
  1.1353E+01  -0.7599  -0.2207   0.3499
TOTAL RESIDUAL SUM OF SQUARES IS      9.59498E+01
MEAN ABSOLUTE CORRELATION INDEX IS   1.00904
VARY(  1):K5          =  5.9161E+01
VARY(  2):K6          =  2.2200E+01
VARY(  3):K7          =  3.8859E+00
VARY(  4):K8          =  1.6700E+00
VARY(  5):K9          =  2.5120E-01
VARY(  6):K10         =  1.5577E-03
VARY(  7):KS3         =  2.6649E+00
VARY(  8):KS4         =  2.3269E+00
VARY(  9):KS5         =  1.6126E+02
VARY( 10):KM          =  3.0870E-05
VARY( 11):VM          =  3.2362E-07
VARY( 12):KMVM        =  9.5391E+01
VARY( 13):KCAT        =  2.6968E+00
        1 SIMULATION RUNS COMPLETED
        2 SIMULATION RUNS COMPLETED
        3 SIMULATION RUNS COMPLETED
        4 SIMULATION RUNS COMPLETED
        5 SIMULATION RUNS COMPLETED
        6 SIMULATION RUNS COMPLETED
        6 SIMULATION RUNS COMPLETED

 PARAMETER NUMBER, TYPE,     NAME,    VALUE
              1   PRI.     K5          5.9161E+01
              2   PRI.     K6          2.2200E+01
              3   PRI.     K7          3.8859E+00
              4   PRI.     K8          1.6700E+00
              5   PRI.     K9          2.5120E-01
              6   SEC.     K10         1.5577E-03
```

7	SEC.	KS3	2.6649E+00
8	SEC.	KS4	2.3269E+00
9	SEC.	KS5	1.6126E+02
10	SEC.	KM	3.0870E-05
11	SEC.	VM	3.2362E-07
12	SEC.	KMVM	9.5391E+01
13	SEC.	KCAT	2.6968E+00

DEPENDENCE MATRIX OF LN(P(I)) ON LN(P(J)), J =

I	1	2	3	4	5
6	1.000E+00	-1.000E+00	1.000E+00	-1.000E+00	1.000E+00
7	1.000E+00	-1.000E+00	3.849E-15	3.849E-15	3.849E-15
8	1.840E-14	1.840E-14	1.000E+00	-1.000E+00	1.840E-14
9	-1.000E+00	1.000E+00	-1.000E+00	1.000E+00	5.703E-14
10	-6.855E-01	5.831E-01	1.109E-01	1.175E-13	1.175E-13
11	3.058E-01	-2.606E-01	9.544E-01	1.891E-13	1.891E-13
12	-9.913E-01	8.437E-01	-8.435E-01	-5.443E-14	-5.443E-14
13	3.058E-01	-2.606E-01	9.544E-01	-2.307E-14	-2.307E-14

RESIDUAL SUM OF SQUARES = 9.5950E+01 FOR 79 DEGREES OF FREEDOM

SINGULAR VALUES OF DEPENDENCE MATRIX AND STRUCTURE OF COMPONENTS

	VALUES	137.157	30.387	16.171	11.289	3.420
K5		-0.47416	-0.16232	0.35807	-0.38557	-0.68698
K6		0.39291	0.26074	-0.63706	-0.11909	-0.59800
K7		-0.67020	-0.28639	-0.63137	0.25899	0.05580
K8		0.39320	-0.80608	0.07468	0.36089	-0.24455
K9		-0.13046	0.41699	0.24848	0.79990	-0.32791

VARIANCE ESTIMATE = 1.2145538

FITTED VALUES, ACCURACIES AND CONFIDENCE LIMITS

NO.	NAME	VALUE	SDLN	5 PERCENT	95 PERCENT
1	K5	5.9161E+01	0.2260	4.0795E+01	8.5795E+01
2	K6	2.2200E+01	0.1981	1.6026E+01	3.0752E+01
3	K7	3.8859E+00	0.0543	3.5538E+00	4.2492E+00
4	K8	1.6700E+00	0.0913	1.4370E+00	1.9407E+00
5	K9	2.5120E-01	0.1333	2.0173E-01	3.1281E-01
6	K10	1.5577E-03	0.0717	1.3843E-03	1.7528E-03
7	KS3	2.6649E+00	0.0799	2.3367E+00	3.0391E+00
8	KS4	2.3269E+00	0.1105	1.9402E+00	2.7907E+00
9	KS5	1.6126E+02	0.0811	1.4113E+02	1.8427E+02
10	KM	3.0870E-05	0.0668	2.7657E-05	3.4457E-05
11	VM	3.2362E-07	0.0315	3.0727E-07	3.4084E-07
12	KMVM	9.5391E+01	0.0548	8.7163E+01	1.0440E+02
13	KCAT	2.6968E+00	0.0315	2.5606E+00	2.8403E+00

CORRELATION MATRIX COMPONENTS

COLUMN	1	2	3	4	5	6	7	8	9	10	11	12	13
ROW 1	1.000	0.937	-0.481	0.795	0.690	0.467	0.504	-0.893	0.721	-0.741	-0.134	-0.826	-0.134
ROW 2	0.937	1.000	-0.186	0.790	0.714	0.371	0.172	-0.744	0.845	-0.461	0.110	-0.625	0.110
ROW 3	-0.481	-0.186	1.000	-0.092	-0.111	-0.331	-0.897	0.568	0.110	0.882	0.896	0.560	0.896
ROW 4	0.795	0.790	-0.092	1.000	0.880	0.615	0.290	-0.872	0.903	-0.485	0.297	-0.762	0.297
ROW 5	0.690	0.714	-0.111	0.880	1.000	0.855	0.180	-0.782	0.888	-0.374	0.160	-0.548	0.160
ROW 6	0.467	0.371	-0.331	0.615	0.855	1.000	0.400	-0.671	0.521	-0.471	-0.129	-0.499	-0.129
ROW 7	0.504	0.172	-0.897	0.290	0.180	0.400	1.000	-0.681	-0.058	-0.953	-0.652	-0.786	-0.652
ROW 8	-0.893	-0.744	0.568	-0.872	-0.782	-0.671	-0.681	1.000	-0.692	0.835	0.195	0.905	0.195
ROW 9	0.721	0.845	0.110	0.903	0.888	0.521	-0.058	-0.692	1.000	-0.199	0.377	-0.459	0.377
ROW 10	-0.741	-0.461	0.882	-0.485	-0.374	-0.471	-0.953	0.835	-0.199	1.000	0.582	0.884	0.582
ROW 11	-0.134	0.110	0.896	0.297	0.160	-0.129	-0.652	0.195	0.377	0.582	1.000	0.134	1.000
ROW 12	-0.826	-0.625	0.560	-0.762	-0.548	-0.499	-0.786	0.905	-0.459	0.884	0.134	1.000	0.134
ROW 13	-0.134	0.110	0.896	0.297	0.160	-0.129	-0.652	0.195	0.377	0.582	1.000	0.134	1.000

***** COMMAND INTERPRETER RE-ENTERED

```
nofit;
param<#1> otemmer:otm5 otm10 otm20 otm40 otm80;
param sc 80;
gstream 1 2;
tm5   sc 1;
tm20  sc 2;
tm80  sc 3;
otm5  sc a;
otm20 sc b;
otm80 sc c;
**;
compile graph;

 LOCN   INSTRUCTION

  540:call resid;
  542:gstream 1;
  548:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	869	19131	20000
REAL SCALARS AND ARRAYS	2567	17433	20000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	549	9451	10000
NAME DICTIONARY	136	864	1000
TEXT ARRAY	80	4920	5000

```
param sx sy 75;
compile instant;

 LOCN   INSTRUCTION

  553:sx=0.3*d0;
  562:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS            869    19131     20000
REAL SCALARS AND ARRAYS              2570    17430     20000
INDEXES AND SECRETS                    99      400       499
CODE ARRAY                            549     9451     10000
NAME DICTIONARY                       139      861      1000
TEXT ARRAY                             80     4920      5000



gstream 2 3;
edx     d0  1;
edtx    d0  2;
eidtx   d0  3;
nm80    100 s;
**;
gstream 3 4;
eid1ppx  sx  4;
ed1ppx   sx  5;
ed1      sx  6;
tm80     sy  p;
otm80    sy  o;
**;
pstream 2 7;
time tm5 tm10 tm20 tm40 tm80 tmth;
**;
compile plot;

 LOCN   INSTRUCTION

  553:call resid;
  555:gstream 2;
  561:gstream 3;
  567:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS            869    19131     20000
REAL SCALARS AND ARRAYS              2570    17430     20000
INDEXES AND SECRETS                    99      400       499
CODE ARRAY                            568     9432     10000
NAME DICTIONARY                       139      861      1000
TEXT ARRAY                             80     4920      5000
```

```
compile table;

 LOCN   INSTRUCTION

  572:call resid;
  574:pstream 2;
  576:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                        IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS          869    19131     20000
REAL SCALARS AND ARRAYS            2570    17430     20000
INDEXES AND SECRETS                  99      400       499
CODE ARRAY                          577     9423     10000
NAME DICTIONARY                     139      861      1000
TEXT ARRAY                           80     4920      5000



pstream 1 1;
k3 k4 k5 k6 k7 k8 k9 k10 k11 k12;
ks2 ks3 ks4 ks5 ks6 keq km vm d0;
g2 g3 g4 g5 g6;
s0;
rks4;
**;
compile ivalues;

 LOCN   INSTRUCTION

  581:pstream 1;
  583:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                        IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS          869    19131     20000
REAL SCALARS AND ARRAYS            2570    17430     20000
INDEXES AND SECRETS                  99      400       499
CODE ARRAY                          584     9416     10000
NAME DICTIONARY                     139      861      1000
TEXT ARRAY                           80     4920      5000



compile set;

 LOCN   INSTRUCTION

  588:for #11=#10;
  595:otm5= vobs<1,#11>;
  601:otm10=vobs<3,#11>;
  607:otm20=vobs<5,#11>;
  613:otm40=vobs<7,#11>;
  619:otm80=vobs<9,#11>;
```

```
  625:#10=#10+1;
  634:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS           869     19131     20000
REAL SCALARS AND ARRAYS             2570     17430     20000
INDEXES AND SECRETS                  100       399       499
CODE ARRAY                           645      9355     10000
NAME DICTIONARY                      139       861      1000
TEXT ARRAY                            80      4920      5000



compile clear;

 LOCN   INSTRUCTION

  649:array<#1> ws;
  652:otenmer=0;
  659:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           1
                          ALLOWED HENCEFORTH   20

  661:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS           869     19131     20000
REAL SCALARS AND ARRAYS             2570     17430     20000
INDEXES AND SECRETS                  100       399       499
CODE ARRAY                           662      9338     10000
NAME DICTIONARY                      139       861      1000
TEXT ARRAY                            80      4920      5000



param<#1> sv v1 v2 v3 vi #1*1e-12;
pstream 3 1;
v1;
v2;
v3;
vi;
sv;
**;
compile velocity;

 LOCN   INSTRUCTION

  666:array<#1> ws;
  669:'ed=0;
  676:'edt=0;
  683:'eidt=0;
  690:'eid1pp=0;
```

```
  697:'ed1pp=0;
  704:'ed1=0;
  711:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           1
                          ALLOWED HENCEFORTH   20

  713:call equations;
  715:array<#1> ws;
  718:v1='edt;
  726:v2='eidt;
  734:v3='eid1pp;
  742:vi=amax(v3,vi);
  755:sv=s0/vi;
  767:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  769:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS            869    19131     20000
REAL SCALARS AND ARRAYS              2600    17400     20000
INDEXES AND SECRETS                   100      399       499
CODE ARRAY                            770     9230     10000
NAME DICTIONARY                       144      856      1000
TEXT ARRAY                             80     4920      5000



whenever
time=0 call ivalues;
time=0+.001*200 call velocity;
time=0+.1*50 call graph;
time=0+.05*50 call plot;
time=0+.2*50% call table;
time=.19 .39 .79 1.19 1.59 2.99 4.99 call set;
time=.21 .41 .81 1.21 1.61 3.01 5.01 call clear;
**;
begin;

TIME  0.0000000E+00  TCPU     26.70      0.00
  STP   1457  TRY   1493  DER    2084  JAC    12  LUD    305  RST    12

 PRINT STREAM NO.  1
  K3          K4          K5          K6          K7          K8          K9          K10
  KS2         KS3         KS4         KS5         KS6         keq          KM          VM
  G2          G3          G4          G5          G6
  S0                <1>         <2>         <3>         <4>         <5>
  RKS4              <1>         <2>         <3>         <4>         <5>
  1.0000E+07  1.0000E+03  5.9161E+01  2.2200E+01  3.8859E+00  1.6700E+00  2.5120E-01  1.5577E-03
  1.0000E+04  2.6649E+00  2.3269E+00  1.6126E+02  1.0000E-04  1.0000E+03  3.0870E-05  3.2362E-07
 -5.4903E+00 -5.8428E-01 -5.0343E-01 -3.0300E+00  5.4903E+00
  5.0000E-06  1.0000E-05  2.0000E-05  4.0000E-05  8.0000E-05  4.0000E-05
  1.0000E+00  1.0000E+00  1.0000E+00  1.0000E+00  1.0000E+00  1.5000E+02
```

```
***** COMMAND INTERPRETER RE-ENTERED

compile instant;

 LOCN   INSTRUCTION

  774:pstream 3;
  776:**;

NO. OF VARIABLES =    48
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                        IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS          869    19131     20000
REAL SCALARS AND ARRAYS            2703    17297     20000
INDEXES AND SECRETS                 100      399       499
CODE ARRAY                          770     9230     10000
NAME DICTIONARY                     144      856      1000
TEXT ARRAY                           80     4920      5000


 PRINT STREAM NO.  3
  V1                <1>          <2>          <3>          <4>          <5>
  V2                <1>          <2>          <3>          <4>          <5>
  V3                <1>          <2>          <3>          <4>          <5>
  VI                <1>          <2>          <3>          <4>          <5>
  SV                <1>          <2>          <3>          <4>          <5>
 -1.2800E-09 -3.7727E-09 -8.9121E-09 -1.6223E-08 -2.3369E-08 -3.3192E-10
 -1.0763E-10 -4.6314E-09 -1.6269E-08 -3.4839E-08 -5.4021E-08  1.4384E-10
  2.6791E-08  4.4858E-08  6.6843E-08  8.6856E-08  1.0025E-07  1.0666E-09
  3.0541E-08  5.3654E-08  8.6622E-08  1.2567E-07  1.6299E-07  1.0796E-09
  1.6371E+02  1.8638E+02  2.3089E+02  3.1830E+02  4.9083E+02  3.7052E+04
stop;

TIME  1.0000000E+01  TCPU     32.25      0.00
  STP     118  TRY     121  DER      166  JAC      1  LUD      25  RST      1

 MAXIMUM SPACE REQUIREMENTS -
NAMES   144,  INTS     1413,  REALS     2703,  CODE    770,  VBLS     48

MEMORY PAGES        0     528        0     228     960     960        0
```

```
PLOTTING PARAMETERS FOR GRAPH STREAM  1
INDEPENDENT VARIABLE :TIME       (PLOTTED VERTICALLY)
DEPENDENT VARIABLES (PLOTTED HORIZONTALLY):
 TM5          SCALE  8.0000E+01   1
 TM20         SCALE  8.0000E+01   2
 TM80         SCALE  8.0000E+01   3
 OTM5         SCALE  8.0000E+01   a
 OTM20        SCALE  8.0000E+01   b
 OTM80        SCALE  8.0000E+01   c
          0                                        0.5                                        1
 0.000E+00 c---------:---------:---------:---------:---------:---------:---------:---------:---------:
 1.000E-01 c  1     2:      3  :         :         :         :         :         :         :         :
 2.000E-01 :   a  1  :        b:         :  3   c  :         :         :         :         :         :
 3.000E-01 c         1         :     2   :         :    3    :         :         :         :         :
 4.000E-01 :         :  1 a    :     b   :  2      :         :  3 c    :         :         :         :
 5.000E-01 c---------:----1----:---------:------2--:---------:-------3-:---------:---------:---------:
 6.000E-01 c         :      1  :         :         :2        :         : 3       :         :         :
 7.000E-01 c         :        1:         :         :   2     :         :    3    :         :         :
 8.000E-01 :         :         a         :         :b    2   :         :       c :         :         :
 9.000E-01 c         :         :1        :         :       2:         :        3  :         :         :
 1.000E+00 c---------:---------:-1-------:---------:---------2---------:---------:3--------:---------:
 1.100E+00 c         :         :  1      :         :         : 2       :         : 3       :         :
 1.200E+00 :         :         :a  1     :         :       b :  2      :         :  3 c    :         :
 1.300E+00 c         :         :    1    :         :         :   2     :         :   3     :         :
 1.400E+00 c         :         :     1   :         :         :    2    :         :    3    :         :
 1.500E+00 c---------:---------:------1--:---------:---------:-----2---:---------:-----3---:---------:
 1.600E+00 :         :         :     a1  :         :         :      b  :         :     c3  :         :
 1.700E+00 c         :         :       1 :         :         :       2 :         :      3  :         :
 1.800E+00 c         :         :       1 :         :         :        2:         :       3 :         :
 1.900E+00 c         :         :        1:         :         :        2:         :        3:         :
 2.000E+00 c---------:---------:---------1---------:---------:---------2---------:--------3:---------:
 2.100E+00 c         :         :         1         :         :         :2        :         3         :
 2.200E+00 c         :         :         :1        :         :         :2        :         :3        :
 2.300E+00 c         :         :         :1        :         :         : 2       :         :3        :
 2.400E+00 c         :         :         : 1       :         :         :  2      :         : 3       :
 2.500E+00 c---------:---------:---------:-1-------:---------:---------:--2------:---------:-3-------:
 2.600E+00 c         :         :         : 1       :         :         :   2     :         :  3      :
 2.700E+00 c         :         :         :  1      :         :         :    2    :         :   3     :
 2.800E+00 c         :         :         :  1      :         :         :    2    :         :   3     :
 2.900E+00 c         :         :         :   1     :         :         :     2   :         :    3    :
 3.000E+00 :---------:---------:---------:---1a----:---------:---------:-----2b--:---------:-c--3----:
 3.100E+00 c         :         :         :    1    :         :         :      2  :         :    3    :
 3.200E+00 c         :         :         :    1    :         :         :       2 :         :    3    :
 3.300E+00 c         :         :         :     1   :         :         :       2 :         :     3   :
 3.400E+00 c         :         :         :     1   :         :         :        2:         :     3   :
 3.500E+00 c---------:---------:---------:-----1---:---------:---------:--------2:---------:-------3-:
 3.600E+00 c         :         :         :      1  :         :         :         2         :      3  :
 3.700E+00 c         :         :         :      1  :         :         :         2         :       3:
 3.800E+00 c         :         :         :       1 :         :         :         :2        :       3:
 3.900E+00 c         :         :         :       1 :         :         :         :2        :        3
 4.000E+00 c---------:---------:---------:-------1-:---------:---------:---------:-2-------:---------3
 4.100E+00 c         :         :         :        1:         :         :         : 2       :         :3
 4.200E+00 c         :         :         :        1:         :         :         :  2      :         :3
 4.300E+00 c         :         :         :         1         :         :         :  2      :         : 3
 4.400E+00 c         :         :         :         1         :         :         :   2     :         : 3
 4.500E+00 c---------:---------:---------:---------1---------:---------:---------:---2-----:---------:--3------:
 4.600E+00 c         :         :         :         :1        :         :         :    2    :         :  3      :
 4.700E+00 c         :         :         :         :1        :         :         :    2    :         :  3      :
 4.800E+00 c         :         :         :         : 1       :         :         :     2   :         :   3     :
 4.900E+00 c         :         :         :         : 1       :         :         :     2   :         :   3     :
 5.000E+00 :---------:---------:---------:---------:-a-------:---------:---------:-----b2--:--------c-:----3----:
```

```
PLOTTING PARAMETERS FOR GRAPH STREAM  2
INDEPENDENT VARIABLE :TIME       (PLOTTED VERTICALLY)
DEPENDENT VARIABLES (PLOTTED HORIZONTALLY):
 EDX         SCALE  1.2000E-07   1
 EDTX        SCALE  1.2000E-07   2
 EIDTX       SCALE  1.2000E-07   3
 NM80        SCALE  1.0000E+02   s
          0                                    0.5                                    1
 0.000E+00 3---------:---------:---------:---------:---------:---------:---------:--1------:---------s
 5.000E-02 :   2 3   :         :         :         :         :         :         : 1       :         :  s      :
 1.000E-01 :  2   3  :         :         :         :         :         :         1         :     s   :         :
 1.500E-01 :  2    3 :         :         :         :         :         :        1:         s:        :         :
 2.000E-01 :  2    3 :         :         :         :         :         :      1 : s       :         :         :
 2.500E-01 :--2----3-:---------:---------:---------:---------:---------:------1s-:---------:---------:---------:
 3.000E-01 :  2    3 :         :         :         :         :         :    s 1  :         :         :         :
 3.500E-01 :  2    3 :         :         :         :         :        :s   1    :         :         :         :
 4.000E-01 :  2    3 :         :         :         :         :      s :   1     :         :         :         :
 4.500E-01 :  2    3 :         :         :         :         :    s    : 1      :         :         :         :
 5.000E-01 :--2----3-:---------:---------:---------:---------:---s-----:--1------:---------:---------:---------:
 5.500E-01 :  2    3 :         :         :         :         :  s       : 1      :         :         :         :
 6.000E-01 :  2    3 :         :         :         :         s          :1       :         :         :         :
 6.500E-01 :  2    3 :         :         :         :        s:          :1       :         :         :         :
 7.000E-01 :  2    3 :         :         :         :      s :           1        :         :         :         :
 7.500E-01 :--2----3-:---------:---------:---------:------s--:---------1--------:---------:---------:---------:
 8.000E-01 :  2   3  :         :         :         :    s    :          1        :         :         :         :
 8.500E-01 :  2   3  :         :         :         :   s     :         1:        :         :         :         :
 9.000E-01 :  2   3  :         :         :         :  s      :         1:        :         :         :         :
 9.500E-01 :  2   3  :         :         :         :  s      :        1 :        :         :         :         :
 1.000E+00 :--2---3--:---------:---------:---------:--s------:-------1-:---------:---------:---------:---------:
 1.050E+00 :  2   3  :         :         :         : s       :        1 :        :         :         :         :
 1.100E+00 :  2   3  :         :         :         : s       :        1 :        :         :         :         :
 1.150E+00 :  2   3  :         :         :         : s       :       1  :        :         :         :         :
 1.200E+00 :  2   3  :         :         :         :s        :       1  :        :         :         :         :
 1.250E+00 :--2---3--:---------:---------:---------:s--------:------1---:---------:---------:---------:---------:
 1.300E+00 :  2   3  :         :         :         :s        :     1    :        :         :         :         :
 1.350E+00 :  2   3  :         :         :         s         :     1    :        :         :         :         :
 1.400E+00 :  2   3  :         :         :         s         :     1    :        :         :         :         :
 1.450E+00 :  2   3  :         :         :         s         :     1    :        :         :         :         :
 1.500E+00 :--2---3--:---------:---------:--------s:---------:-----1----:---------:---------:---------:---------:
 1.550E+00 :  2   3  :         :         :        s:         :    1     :        :         :         :         :
 1.600E+00 :  2   3  :         :         :        s:         :    1     :        :         :         :         :
 1.650E+00 :  2   3  :         :         :       s :         :    1     :        :         :         :         :
 1.700E+00 :  2   3  :         :         :       s :         :    1     :        :         :         :         :
 1.750E+00 :--2---3--:---------:---------:------s--:---------:----1-----:---------:---------:---------:---------:
 1.800E+00 :  2   3  :         :         :      s  :         :    1     :        :         :         :         :
 1.850E+00 :  2   3  :         :         :     s   :         :   1      :        :         :         :         :
 1.900E+00 :  2   3  :         :         :     s   :         :   1      :        :         :         :         :
 1.950E+00 :  2   3  :         :         :     s   :         :   1      :        :         :         :         :
 2.000E+00 :--2---3--:---------:---------:-----s---:---------:---1------:---------:---------:---------:---------:
 2.050E+00 :  2   3  :         :         :    s    :         :  1       :        :         :         :         :
 2.100E+00 :  2   3  :         :         :    s    :         :  1       :        :         :         :         :
 2.150E+00 :  2   3  :         :         :    s    :         :  1       :        :         :         :         :
 2.200E+00 :  2   3  :         :         :   s     :         : 1        :        :         :         :         :
 2.250E+00 :--2---3--:---------:---------:----s----:---------:--1-------:---------:---------:---------:---------:
 2.300E+00 :  2   3  :         :         :   s     :         : 1        :        :         :         :         :
 2.350E+00 :  2   3  :         :         :   s     :         : 1        :        :         :         :         :
 2.400E+00 :  2   3  :         :         :   s     :         : 1        :        :         :         :         :
 2.450E+00 :  2   3  :         :         :  s      :         : 1        :        :         :         :         :
 2.500E+00 :--2---3--:---------:---------:---s-----:---------:--1-------:---------:---------:---------:---------:
```

```
PLOTTING PARAMETERS FOR GRAPH STREAM  3
INDEPENDENT VARIABLE :TIME       (PLOTTED VERTICALLY)
DEPENDENT VARIABLES (PLOTTED HORIZONTALLY):
 EID1PPX      SCALE  3.6000E-08   4
 ED1PPX       SCALE  3.6000E-08   5
 ED1          SCALE  3.6000E-08   6
 TM80         SCALE  7.5000E+01   p
 OTM80        SCALE  7.5000E+01   o
          0                                   0.5                                   1
0.000E+00 o---------:---------:---------:---------:---------:---------:---------:---------:---------:
5.000E-02 o 4     p :         :         :         :         :         :         :         :         :
1.000E-01 o     4   :       p :         :         :         :         :         :         :         :
1.500E-01 o         4         :       p :         :         :         :         :         :         :
2.000E-01 6         :   4     :         :     p   o:        :         :         :         :         :
2.500E-01 o6--------:-------4-:---------:---------:-p-------:---------:---------:---------:---------:
3.000E-01 o6        :         :4        :         :       p :         :         :         :         :
3.500E-01 o6        :         :   4     :         :         : p       :         :         :         :
4.000E-01 56        :         :      4  :         :         :     p   o:        :         :         :
4.500E-01 o 6       :         :        4:         :         :       p:        :         :         :
5.000E-01 o-6-------:---------:---------:4--------:---------:---------:-p-------:---------:---------:
5.500E-01 o 6       :         :         : 4       :         :         :   p     :         :         :
6.000E-01 o  6      :         :         :   4     :         :         :     p   :         :         :
6.500E-01 o  6      :         :         :     4   :         :         :       p :         :         :
7.000E-01 o   6     :         :         :       4 :         :         :         p         :         :
7.500E-01 o---6-----:---------:---------:---------4:---------:---------:---------:p--------:---------:
8.000E-01 5    6    :         :         :         4         :         :         : o       :         :
8.500E-01 o    6    :         :         :         :4        :         :         :  p      :         :
9.000E-01 o     6   :         :         :         : 4       :         :         :   p     :         :
9.500E-01 o     6   :         :         :         : 4       :         :         :    p    :         :
1.000E+00 o------6--:---------:---------:---------:--4------:---------:---------:-----p---:---------:
1.050E+00 o      6  :         :         :         :  4      :         :         :      p  :         :
1.100E+00 o       6 :         :         :         :  4      :         :         :      p  :         :
1.150E+00 o        6:         :         :         :   4     :         :         :       p :         :
1.200E+00 5        6:         :         :         :   4     :         :         :       po:         :
1.250E+00 o---------6---------:---------:---------:----4----:---------:---------:--------p:---------:
1.300E+00 o         6         :         :         :    4    :         :         :       p:         :
1.350E+00 o         :6        :         :         :    4    :         :         :        p         :
1.400E+00 o         :6        :         :         :    4    :         :         :        p         :
1.450E+00 o         : 6       :         :         :    4    :         :         :        :p        :
1.500E+00 o---------:--6------:---------:---------:------4---:---------:---------:--------:p--------:
1.550E+00 o         :  6      :         :         :     4    :         :         :        :p        :
1.600E+00 5         :   6     :         :         :     4    :         :         :        :op       :
1.650E+00 o         :   6     :         :         :     4    :         :         :        : p       :
1.700E+00 o         :    6    :         :         :     4    :         :         :        :  p      :
1.750E+00 o---------:-----6---:---------:---------:------4---:---------:---------:---------:--p-----:
1.800E+00 o         :     6   :         :         :     4    :         :         :        :  p      :
1.850E+00 o         :      6  :         :         :     4    :         :         :        :   p     :
1.900E+00 o         :      6  :         :         :     4    :         :         :        :   p     :
1.950E+00 o         :       6 :         :         :     4    :         :         :        :   p     :
2.000E+00 o---------:-------6-:---------:---------:------4---:---------:---------:---------:----p----:
2.050E+00 o         :        6:         :         :     4    :         :         :        :    p    :
2.100E+00 o         :         6         :         :     4    :         :         :        :    p    :
2.150E+00 o         :         6         :         :     4    :         :         :        :     p   :
2.200E+00 o         :         :6        :         :     4    :         :         :        :     p   :
2.250E+00 o---------:---------:6--------:---------:------4---:---------:---------:---------:-----p---:
2.300E+00 o         :         : 6       :         :     4    :         :         :        :      p  :
2.350E+00 o         :         :  6      :         :     4    :         :         :        :      p  :
2.400E+00 o         :         :  6      :         :     4    :         :         :        :      p  :
2.450E+00 o         :         :   6     :         :     4    :         :         :        :       p :
2.500E+00 o---------:---------:----6----:---------:------4---:---------:---------:---------:-------p-:
```

PRINT STREAM NO. 2

TIME	TM5	TM10	TM20	TM40	TM80	TMTH
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.0000E-01	5.2909E+00	9.2864E+00	1.4870E+01	2.1146E+01	2.6655E+01	1.9312E-01
4.0000E-01	1.0064E+01	1.7049E+01	2.5985E+01	3.4975E+01	4.2043E+01	4.1552E-01
6.0000E-01	1.3498E+01	2.2297E+01	3.2866E+01	4.2722E+01	4.9932E+01	6.3699E-01
8.0000E-01	1.6035E+01	2.5961E+01	3.7309E+01	4.7314E+01	5.4284E+01	8.5753E-01
1.0000E+00	1.7971E+01	2.8625E+01	4.0340E+01	5.0252E+01	5.6939E+01	1.0772E+00
1.2000E+00	1.9503E+01	3.0652E+01	4.2540E+01	5.2303E+01	5.8757E+01	1.2960E+00
1.4000E+00	2.0762E+01	3.2274E+01	4.4255E+01	5.3886E+01	6.0171E+01	1.5139E+00
1.6000E+00	2.1839E+01	3.3638E+01	4.5685E+01	5.5217E+01	6.1384E+01	1.7310E+00
1.8000E+00	2.2794E+01	3.4839E+01	4.6948E+01	5.6412E+01	6.2495E+01	1.9473E+00
2.0000E+00	2.3665E+01	3.5934E+01	4.8108E+01	5.7526E+01	6.3543E+01	2.1628E+00
2.2000E+00	2.4477E+01	3.6953E+01	4.9194E+01	5.8578E+01	6.4539E+01	2.3775E+00
2.4000E+00	2.5246E+01	3.7918E+01	5.0226E+01	5.9583E+01	6.5494E+01	2.5915E+00
2.6000E+00	2.5984E+01	3.8845E+01	5.1219E+01	6.0554E+01	6.6419E+01	2.8048E+00
2.8000E+00	2.6700E+01	3.9743E+01	5.2183E+01	6.1496E+01	6.7315E+01	3.0173E+00
3.0000E+00	2.7399E+01	4.0620E+01	5.3125E+01	6.2417E+01	6.8191E+01	3.2291E+00
3.2000E+00	2.8085E+01	4.1479E+01	5.4047E+01	6.3316E+01	6.9044E+01	3.4402E+00
3.4000E+00	2.8760E+01	4.2323E+01	5.4949E+01	6.4194E+01	6.9875E+01	3.6507E+00
3.6000E+00	2.9427E+01	4.3154E+01	5.5834E+01	6.5051E+01	7.0683E+01	3.8605E+00
3.8000E+00	3.0086E+01	4.3971E+01	5.6700E+01	6.5887E+01	7.1470E+01	4.0696E+00
4.0000E+00	3.0737E+01	4.4775E+01	5.7549E+01	6.6703E+01	7.2235E+01	4.2781E+00
4.2000E+00	3.1381E+01	4.5567E+01	5.8381E+01	6.7500E+01	7.2979E+01	4.4860E+00
4.4000E+00	3.2019E+01	4.6348E+01	5.9197E+01	6.8276E+01	7.3703E+01	4.6932E+00
4.6000E+00	3.2650E+01	4.7117E+01	5.9996E+01	6.9034E+01	7.4408E+01	4.8999E+00
4.8000E+00	3.3276E+01	4.7876E+01	6.0780E+01	6.9774E+01	7.5093E+01	5.1059E+00
5.0000E+00	3.3895E+01	4.8622E+01	6.1548E+01	7.0496E+01	7.5761E+01	5.3114E+00
5.2000E+00	3.4508E+01	4.9359E+01	6.2301E+01	7.1201E+01	7.6410E+01	5.5163E+00
5.4000E+00	3.5116E+01	5.0084E+01	6.3039E+01	7.1889E+01	7.7042E+01	5.7206E+00
5.6000E+00	3.5718E+01	5.0799E+01	6.3763E+01	7.2561E+01	7.7657E+01	5.9243E+00
5.8000E+00	3.6314E+01	5.1504E+01	6.4473E+01	7.3216E+01	7.8256E+01	6.1275E+00
6.0000E+00	3.6904E+01	5.2199E+01	6.5169E+01	7.3856E+01	7.8838E+01	6.3301E+00
6.2000E+00	3.7489E+01	5.2883E+01	6.5851E+01	7.4481E+01	7.9405E+01	6.5322E+00
6.4000E+00	3.8068E+01	5.3558E+01	6.6519E+01	7.5090E+01	7.9957E+01	6.7338E+00
6.6000E+00	3.8643E+01	5.4223E+01	6.7175E+01	7.5686E+01	8.0494E+01	6.9348E+00
6.8000E+00	3.9211E+01	5.4879E+01	6.7818E+01	7.6266E+01	8.1016E+01	7.1353E+00
7.0000E+00	3.9774E+01	5.5525E+01	6.8448E+01	7.6833E+01	8.1525E+01	7.3353E+00
7.2000E+00	4.0333E+01	5.6162E+01	6.9066E+01	7.7387E+01	8.2020E+01	7.5348E+00
7.4000E+00	4.0885E+01	5.6790E+01	6.9672E+01	7.7927E+01	8.2501E+01	7.7338E+00
7.6000E+00	4.1433E+01	5.7409E+01	7.0266E+01	7.8454E+01	8.2970E+01	7.9323E+00
7.8000E+00	4.1976E+01	5.8019E+01	7.0848E+01	7.8969E+01	8.3426E+01	8.1303E+00
8.0000E+00	4.2513E+01	5.8620E+01	7.1419E+01	7.9471E+01	8.3870E+01	8.3278E+00
8.2000E+00	4.3046E+01	5.9212E+01	7.1978E+01	7.9962E+01	8.4302E+01	8.5248E+00
8.4000E+00	4.3573E+01	5.9796E+01	7.2527E+01	8.0440E+01	8.4723E+01	8.7213E+00
8.6000E+00	4.4096E+01	6.0372E+01	7.3065E+01	8.0908E+01	8.5132E+01	8.9174E+00
8.8000E+00	4.4614E+01	6.0940E+01	7.3592E+01	8.1364E+01	8.5530E+01	9.1129E+00
9.0000E+00	4.5127E+01	6.1499E+01	7.4110E+01	8.1809E+01	8.5918E+01	9.3080E+00
9.2000E+00	4.5635E+01	6.2050E+01	7.4617E+01	8.2244E+01	8.6295E+01	9.5027E+00
9.4000E+00	4.6139E+01	6.2594E+01	7.5114E+01	8.2668E+01	8.6662E+01	9.6969E+00
9.6000E+00	4.6637E+01	6.3129E+01	7.5601E+01	8.3082E+01	8.7019E+01	9.8906E+00
9.8000E+00	4.7132E+01	6.3657E+01	7.6079E+01	8.3486E+01	8.7367E+01	1.0084E+01
1.0000E+01	4.7621E+01	6.4178E+01	7.6547E+01	8.3881E+01	8.7706E+01	1.0277E+01

Appendix 2.

Computer output from FACSIMILE analysis of the steady state incorporation of T opposite O^6-meG in the template strand.

```
*T opposite meG, time course. Determination of koff;
punch 6;
compile instant;


 LOCN   INSTRUCTION

   74:#1=1;
   77:#2=#1-1;
   86:#3=5*#1;
   95:**;

NO. OF VARIABLES =     0
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0


SPACE FOR                              IN USE      FREE      TOTAL
```

```
INTEGER SCALARS AND ARRAYS          41     23791     23832
REAL SCALARS AND ARRAYS              6     23826     23832
INDEXES AND SECRETS                 99       400       499
CODE ARRAY                          70     14930     15000
NAME DICTIONARY                     10       990      1000
TEXT ARRAY                         180     19820     20000


varia<#1> e ed edt eidt eid1pp ed1pp ed1 ed2;
varia<#1> d d1 d2 tmp ttp pp;
param t0 20e-6 d0 1e-6;
varia<#1> dnax:d109;
param<#1> e0 1.09e-7;
param k1  1e7    k2 0.05    ks1 ;
param k3  1e7    k4 1000    ks2 ;
param k5  59.1   k6 22.2    ks3 ;
param k7  3.79   k8 1.71    ks4 ;
param k9  0.25   k10 1.5e-3 ks5 ;
param k11 1e3    k12 1e7    ks6 ;
param k13 0.05   k14 1e7    ks7 ;
param k15 0.05   k16 1e7    ks8 ;
param keq 900    kexo 0.0016;
param<#1> a b c f x;
param<#3> ws;

compile initial;


 LOCN   INSTRUCTION

   74:#10=0;
   77:array<#1> ws;
   80:a=1;
   87:c=e0*d0;
   98:b=-(e0+d0+ks1);
  113:f=(b*b)-(4*a*c);
  137:x=(-b-sqrt(f))/(2*a);
  160:x=ed;
  168:e=e0-ed;
  180:d=d0-ed;
  191:ttp=t0;
  198:array;


MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  200:  k10=(k1*k3*k5*k7*k9*k11*k13)/(k2*k4*k6*k8*keq*k12*k14);
  246:  ks1=k2/k1;
  255:  ks2=k4/k3;
  264:  ks3=k5/k6;
  273:  ks4=k7/k8;
  282:  ks5=k9/k10;
  291:  ks6=k11/k12;
  300:  ks7=k13/k14;
  309:  ks8=k15/k16;
  318:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

```
SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS             48     23784     23832
REAL SCALARS AND ARRAYS               115     23717     23832
INDEXES AND SECRETS                    99       400       499
CODE ARRAY                            319     14681     15000
NAME DICTIONARY                        64       936      1000
TEXT ARRAY                            720     19280     20000



compile equations;


 LOCN   INSTRUCTION


*****    EQUATION FIXES MAX. NO. OF VARIABLES AS    45

  332:array<#1> ws;
  335:%k1%k2    :e   +d=ed;
  357:%k3%k4    :ed+ttp=edt;
  379:%k5%k6    :edt   =eidt;
  397:%k7%k8    :eidt  =eid1pp;
  415:%k9%k10   :eid1pp=ed1pp;
  433:%k11%k12  :ed1pp =ed1 + pp;
  455:%k13%k14  :ed1   =e + d1;
  477:%kexo     :ed1pp =ed + tmp + pp;
  500:%kexo     :ed1   =ed + tmp + pp;
  523:%kexo     :ed    =ed2 + tmp + pp;
  546:%k15%k16  :ed2   =e + d2;
  568:array;


MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  570:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0


SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS            133     23699     23832
REAL SCALARS AND ARRAYS               660     23172     23832
INDEXES AND SECRETS                    99       400       499
CODE ARRAY                            571     14429     15000
NAME DICTIONARY                        72       928      1000
TEXT ARRAY                            800     19200     20000



compile resid;


 LOCN   INSTRUCTION

  575:array<#1> ws;
  578:dnax=eid1pp + ed1pp + ed1 + d1;
```

```
 598:d=d0-ed-edt-eidt-eid1pp-ed1pp-ed1-d1-d2-ed2;
 641:e=e0-ed-edt-eidt-eid1pp-ed1pp-ed1-ed2;
 677:array;


MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           1
                          ALLOWED HENCEFORTH   20

 679:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0


SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS            133     23699      23832
REAL SCALARS AND ARRAYS               660     23172      23832
INDEXES AND SECRETS                    99       400        499
CODE ARRAY                            680     14320      15000
NAME DICTIONARY                        72       928       1000
TEXT ARRAY                            800     19200      20000



data 0.06;
time   d109;
range .692e-6;
30    .154e-6;
60    .233e-6;
90    .351e-6;
120   .440e-6;
180   .637e-6;
240   .692e-6;
**;


    1 CURVES
    6 TIME       POINTS
    6 OBSERVATIONS


A TIME       POINT INCREMENT OF     1 AND OFFSET OF     0 SPECIFIED
     1 DEPENDENT VARIABLES SELECTED
     6 TIME       POINTS AND     6 OBSERVATIONS USED

vary k13:ks7;
save;
compile instant;


 LOCN   INSTRUCTION

  684:begin;
  686:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

```
SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS            140     23692      23832
REAL SCALARS AND ARRAYS               775     23057      23832
INDEXES AND SECRETS                    99       400        499
CODE ARRAY                            680     14320      15000
NAME DICTIONARY                        81       919       1000
TEXT ARRAY                            890     19110      20000



SPACE FOR                          IN USE      FREE      TOTAL
REAL SCALARS AND ARRAYS               806     23026      23832


SIZE NEEDED FOR ARRAY WSFIT =         31


OPTIMISATION RUN WITH    6 OBSERVATIONS AND  1 VARIED PARAMETERS


TIME    .0000000E+00  TCPU      10.44        .00
  STP       0  TRY       0  DER        0  JAC      0  LUD       0  RST      0


AFTER   1 ITERATIONS AND   1 RUNS INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  7.2160E+00    1.3879     .5733     .3727
TOTAL RESIDUAL SUM OF SQUARES IS      7.21596E+00
MEAN ABSOLUTE CORRELATION INDEX IS    1.38791
VARY(  1):K13         =  5.0000E-02
VARY(  2):KS7         =  5.0000E-09


AFTER   2 ITERATIONS AND   1 RUNS INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  6.1660E+00    1.2705     .5346     .3727
TOTAL RESIDUAL SUM OF SQUARES IS      6.16605E+00
MEAN ABSOLUTE CORRELATION INDEX IS    1.27046
VARY(  1):K13         =  5.1266E-02
VARY(  2):KS7         =  5.1266E-09


AFTER   3 ITERATIONS AND   1 RUNS INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  2.5906E+00    -.6970     .1553     .3727
TOTAL RESIDUAL SUM OF SQUARES IS      2.59057E+00
MEAN ABSOLUTE CORRELATION INDEX IS     .69696
VARY(  1):K13         =  6.1598E-02
VARY(  2):KS7         =  6.1598E-09
        3 SIMULATION RUNS COMPLETED


RESIDUAL SUM OF SQUARES IS                 2.5905745
COMPARED WITH EXPECTED RANGE      .8312   12.8219
FOR    5 DEGREES OF FREEDOM
ANALYSIS OF VARIANCE PROCEEDS ON ASSUMPTION OF GOOD FIT
BUT INSPECT RESIDUALS FOR EVIDENCE OF SYSTEMATIC ERROR


AFTER   0 ITERATIONS AND   1 RUNS INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  2.5906E+00    -.6970     .1553     .3727
```

```
TOTAL RESIDUAL SUM OF SQUARES IS      2.59057E+00
MEAN ABSOLUTE CORRELATION INDEX IS    .69696
VARY(  1):K13         =  6.1598E-02
VARY(  2):KS7         =  6.1598E-09
       2 SIMULATION RUNS COMPLETED


PARAMETER NUMBER, TYPE,     NAME,    VALUE
             1   PRI.     K13         6.1598E-02
             2   SEC.     KS7         6.1598E-09


DEPENDENCE MATRIX OF LN(P(I)) ON LN(P(J)), J =
      1                              ..
  I


  2  1.000E+00


RESIDUAL SUM OF SQUARES =  2.5906E+00 FOR    5 DEGREES OF FREEDOM


SINGULAR VALUES OF DEPENDENCE MATRIX AND STRUCTURE OF COMPONENTS


   VALUES     10.527


K13          1.00000
VARIANCE ESTIMATE =        .5181149


FITTED VALUES, ACCURACIES AND CONFIDENCE LIMITS
NO.   NAME        VALUE        SDLN   5 PERCENT  95 PERCENT
 1 K13          6.1598E-02   .0684  5.5045E-02  6.8932E-02
 2 KS7          6.1598E-09   .0684  5.5045E-09  6.8932E-09


 CORRELATION MATRIX COMPONENTS


COLUMN   1      2


ROW  1  1.000  1.000


ROW  2  1.000  1.000


***** COMMAND INTERPRETER RE-ENTERED


nofit;
param xs;
compile instant;


 LOCN   INSTRUCTION

  684:xs=1e-6;
  690:**;
```

```
NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	864	22968	23832
REAL SCALARS AND ARRAYS	1150	22682	23832
INDEXES AND SECRETS	99	400	499
CODE ARRAY	680	14320	15000
NAME DICTIONARY	83	917	1000
TEXT ARRAY	910	19090	20000

```
param<#1> dnao #1*0:d109o d327o;
gstream 1 2 100;
d109   xs a;
d109o  xs 0;
**;
compile gra1;

 LOCN   INSTRUCTION

  684:call resid;
  686:gstream 1;
  692:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	864	22968	23832
REAL SCALARS AND ARRAYS	1153	22679	23832
INDEXES AND SECRETS	99	400	499
CODE ARRAY	693	14307	15000
NAME DICTIONARY	86	914	1000
TEXT ARRAY	940	19060	20000

```
pstream 1 1 10;
ks2 k3 k4 ks4 k5 k6 k7 ks5 k9 k10;
**;
compile ivalues;

 LOCN   INSTRUCTION

  697:pstream 1;
  699:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	864	22968	23832
REAL SCALARS AND ARRAYS	1153	22679	23832
INDEXES AND SECRETS	99	400	499

```
CODE ARRAY                          700     14300     15000
NAME DICTIONARY                      86       914      1000
TEXT ARRAY                          940     19060     20000



pstream 2 1 10;
time d109;
**;
compile table;


 LOCN   INSTRUCTION

  704:call resid;
  706:pstream 2;
  708:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0


SPACE FOR                       IN USE      FREE     TOTAL
INTEGER SCALARS AND ARRAYS         864     22968     23832
REAL SCALARS AND ARRAYS           1153     22679     23832
INDEXES AND SECRETS                 99       400       499
CODE ARRAY                         709     14291     15000
NAME DICTIONARY                     86       914      1000
TEXT ARRAY                         940     19060     20000



compile set;


 LOCN   INSTRUCTION

  713:for #11=#10;
  720:do 1 for #12=0(1)#2;
  728:dnao<#12>=vobs<#12,#11>;
  734:label 1;
  761:#10=#10+1;
  770:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0


SPACE FOR                       IN USE      FREE     TOTAL
INTEGER SCALARS AND ARRAYS         864     22968     23832
REAL SCALARS AND ARRAYS           1153     22679     23832
INDEXES AND SECRETS                101       398       499
CODE ARRAY                         771     14229     15000
NAME DICTIONARY                     86       914      1000
TEXT ARRAY                         940     19060     20000



compile clear;
```

```
LOCN   INSTRUCTION

 775:array<#1> ws;
 778:dnao=0;
 785:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR      5
                          ENCOUNTERED         1
                          ALLOWED HENCEFORTH  20

 787:**;

NO. OF VARIABLES =    45
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS           864    22968     23832
REAL SCALARS AND ARRAYS             1153    22679     23832
INDEXES AND SECRETS                  101      398       499
CODE ARRAY                           788    14212     15000
NAME DICTIONARY                       86      914      1000
TEXT ARRAY                           940    19060     20000


when
time= 0   call ivalues;
time= 9 19 29 39 49 59 69 79 89 99 109 119 129 139 149
      159 169 179 189 199 209 219 229 239 call set;
time= 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
      160 170 180 190 200 210 220 230 240 call table;
time= 11 21 31 41 51 61 71 81 91 101 111 121 131 141 151
      161 171 181 191 201 211 221 231 241% call clear;
**;
begin;


TIME   .0000000E+00  TCPU    102.44       .00
  STP    894  TRY    917  DER    1685  JAC    11  LUD    174  RST    5


 PRINT STREAM NO.  1
  KS2          K3           K4           KS4          K5           K6           K7           KS5          K9
  1.0000E-04   1.0000E+07   1.0000E+03   2.2164E+00   5.9100E+01   2.2200E+01   3.7900E+00   1.2381E+02   2.50

 PRINT STREAM NO.  2
  TIME         D109
  1.0000E+01   9.0286E-08
  2.0000E+01   1.3131E-07
  3.0000E+01   1.6920E-07
  4.0000E+01   2.0546E-07
  5.0000E+01   2.4040E-07
  6.0000E+01   2.7412E-07
  7.0000E+01   3.0663E-07
  8.0000E+01   3.3797E-07
  9.0000E+01   3.6815E-07
  1.0000E+02   3.9720E-07
  1.1000E+02   4.2513E-07
  1.2000E+02   4.5198E-07
```

1.3000E+02	4.7775E-07
1.4000E+02	5.0249E-07
1.5000E+02	5.2622E-07
1.6000E+02	5.4896E-07
1.7000E+02	5.7073E-07
1.8000E+02	5.9157E-07
1.9000E+02	6.1151E-07
2.0000E+02	6.3057E-07
2.1000E+02	6.4877E-07
2.2000E+02	6.6615E-07
2.3000E+02	6.8274E-07
2.4000E+02	6.9856E-07

Appendix 3.

Computer output from FACSIMILE analysis of the data from the thymine/cytosine competition experiments and the single nucleotide incorporation experiments. The solution is constrained by fixing the overall equilibrium constant (keq) at 1000, and setting the K_m for incorporation of T as equal to that for the incorporation of C.

```
FACSIMILE OUTPUT ON  9/ 8/1993 AT 17: 6:42
*                    ttpctpc;
punch 6;
text #t 1;
compile instant;

 LOCN   INSTRUCTION

   74:#t=!asci<9>;  sets #t to tab
   84:#1=7;
   87:#10=5*#1;
   96:**;

NO. OF VARIABLES =     0
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS             68    19932     20000
REAL SCALARS AND ARRAYS                 6    19994     20000
INDEXES AND SECRETS                    99      400       499
CODE ARRAY                             70     9930     10000
NAME DICTIONARY                        29      971      1000
TEXT ARRAY                             81     4919      5000


param<#10> ws;
varia<#1> e ed edt eidt eid1pp ed1pp ed1;
varia<#1>      edc eidc eid2pp ed2pp ed2;
varia<#1> d0 1e-7 1e-7 1e-7 1e-7 1e-7 1e-7 1e-7;
varia<#1> dt dc ttp ctp tmp cmp pp;
param<#1> t0 20e-6 20e-6 20e-6 00e-6 80e-6 00e-6 05e-6;
param<#1> c0 20e-6 80e-6 00e-6 80e-6 00e-6 20e-6 00e-6;
param<#1> e0 1.2e-7  1.2e-7  1.2e-7  1.2e-7  1.2e-7  1.2e-7  1.2e-7;
param kexo .03 kmt 3.08e-5 vmt 2.63e-7 kmc 2.55e-5 vmc 4.84e-8;
```

```
param e0tc 1.2e-7 e0t 1.2e-7 e0c 1.2e-7;
param  ke1  k1   1e7     k2   .05;
param  ke2t k3t  1e7     k4t  1000;
param  ke3t k5t  14.3    k6t  3.99;
param  ke4t k7t  6.85    k8t  4.28;
param  ke5t k9t  .368    k10t 2.12e-3;
param  ke6t k11t 1000    k12t 1e7;
param  ke2c k3c  1e7     k4c  1000;
param  ke3c k5c  5.06    k6c  2.82;
param  ke4c k7c  1.055   k8c  1.45;
param  ke5c k9c  .318    k10c 4.16e-4;
param  ke6c k11c 1000    k12c 1e7;
param  ke7  k13  1e7     k14  1e7;
param  keq 1e3 vmtvmc;
compile initial;

 LOCN   INSTRUCTION

   74:#21=0;
   77:ke1=k1/k2;
   86:ke2t=k3t/k4t;
   95:ke3t=k5t/k6t;
  104:ke4t=k7t/k8t;
  113:ke5t=k9t/k10t;
  122:ke6t=k11t/k12t;
  131:ke2c=k3c/k4c;
  140:ke3c=k5c/k6c;
```

```
 149:ke4c=k7c/k8c;
 158:ke5c=k9c/k10c;
 167:ke6c=k11c/k12c;
 176:ke7=k13/k14;
 185:vmc=(e0c*k5c*k7c)/(k5c+k6c+k7c);
 207:kmc=((k5c*k7c)+(k4c*k6c)+(k4c*k7c))/(k3c*(k5c+k6c+k7c));
 244:k10c=(k3c*k5c*k7c*k9c*k11c)/(keq*k4c*k6c*k8c*k12c);
 278:k5t=((k6t+k7t)*(k4t-k3t*kmc))/(k3t*kmc-k7t);
 308:vmt=(e0t*k5t*k7t)/(k5t+k6t+k7t);
 330:vmtvmc=vmt/vmc;
 339:kmt=((k5t*k7t)+(k4t*k6t)+(k4t*k7t))/(k3t*(k5t+k6t+k7t));
 376:k10t=(k3t*k5t*k7t*k9t*k11t)/(keq*k4t*k6t*k8t*k12t);
 410:e0<0>=e0tc;
 416:e0<1>=e0tc;
 422:e0<2>=e0t;
 428:e0<3>=e0c;
 434:e0<4>=e0t;
 440:e0<5>=e0c;
 446:e0<6>=e0t;
 452:array<#1> ws;
 455:ed=d0;
 463:ttp=t0;
 471:ctp=c0;
 479:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           1
                          ALLOWED HENCEFORTH   20

 481:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE     FREE     TOTAL
INTEGER SCALARS AND ARRAYS             75    19925     20000
REAL SCALARS AND ARRAYS               248    19752     20000
INDEXES AND SECRETS                    99      400       499
CODE ARRAY                            482     9518     10000
NAME DICTIONARY                       100      900      1000
TEXT ARRAY                             81     4919      5000


compile equations;

 LOCN   INSTRUCTION


*****    EQUATION FIXES MAX. NO. OF VARIABLES AS   140

 486:array<#1> ws;
 489:%k3t %k4t :   ed +ttp=edt;
 511:%k5t %k6t :   edt    =eidt;
 529:%k7t %k8t :   eidt   =eid1pp;
 547:%k9t %k10t:   eid1pp =ed1pp;
 565:%k11t%k12t:   ed1pp  =ed1+pp;
 587:%k3c %k4c :   ed +ctp=edc;
 609:%k5c %k6c :   edc    =eidc;
 627:%k7c %k8c :   eidc   =eid2pp;
```

```
  645:%k9c %k10c:   eid2pp =ed2pp;
  663:%k11c%k12c:   ed2pp  =ed2+pp;
  685:%k13 %k14 :   edc+ttp=edt+ctp;
  711:%kexo:ed1=ed;
  726:%kexo:ed2=ed;
  741:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  743:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS            160     19840     20000
REAL SCALARS AND ARRAYS              1978     18022     20000
INDEXES AND SECRETS                    99       400       499
CODE ARRAY                            744      9256     10000
NAME DICTIONARY                       109       891      1000
TEXT ARRAY                             81      4919      5000



param<#1> xdt:t2020 t2080 t2000 t0080 t8000 t0020 t0500;
param<#1> xdc:c2020 c2080 c2000 c0080 c8000 c0020 c4000;
compile resid;

 LOCN   INSTRUCTION

  748:array<#1> ws;
  751:xdt=100*(eid1pp+ed1pp+ed1)/d0;
  775:xdc=100*(eid2pp+ed2pp+ed2)/d0;
  799:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           2
                          ALLOWED HENCEFORTH   20

  801:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS            160     19840     20000
REAL SCALARS AND ARRAYS              1993     18007     20000
INDEXES AND SECRETS                    99       400       499
CODE ARRAY                            802      9198     10000
NAME DICTIONARY                       126       874      1000
TEXT ARRAY                             81      4919      5000



data .075;
time  t2020 c2020 t2080 c2080 t2000 c0080 t8000 c0020 t0500;
range  49.8   9.6  39.2  23.7  60.7  42.0  70.3  31.9  33.8;
.40    00.0   0.0  00.0  00.0  20.7  00.0  44.3  00.0  11.8;
```

```
.50    19.8   2.6   15.8   9.95   00.0   00.0   00.0   00.0   00.0;
.80    00.0   0.0   0.0    0.0    32.4   0.0    54.1   0.0    16.3;
1.0    27.3   5.0   23.8   13.9   00.0   19.2   00.0   10.7   00.0;
1.2    00.0   0.0   00.0   00.0   38.1   00.0   59.6   00.0   16.7;
1.6    00.0   0.0   0.0    0.0    45.5   00.0   61.1   00.0   20.4;
2.0    37.3   8.4   33.0   17.6   00.0   27.5   00.0   00.0   00.0;
3.0    00.0   0.0   00.0   00.0   53.8   29.0   65.3   18.6   28.0;
4.0    49.8   9.6   39.2   23.7   00.0   00.0   00.0   00.0   00.0;
5.0    00.0   0.0   00.0   00.0   60.7   34.8   70.3   27.3   33.8;
7.0    00.0   0.0   00.0   00.0   00.0   42.0   00.0   31.9   00.0;
**;

   9 CURVES
  11 TIME       POINTS
  43 OBSERVATIONS


A TIME       POINT INCREMENT OF     1 AND OFFSET OF    0 SPECIFIED
    9 DEPENDENT VARIABLES SELECTED
   11 TIME       POINTS AND    43 OBSERVATIONS USED

vary k6t k7t k8t k9t
k5c k6c k7c k8c k9c k13:vmt k5t k10t ke3t ke4t ke5t
kmc vmc vmtvmc k10c ke3c ke4c ke5c ke6c ke7;
begin;

SPACE FOR                              IN USE      FREE      TOTAL
REAL SCALARS AND ARRAYS                  3642     16358     20000

SIZE NEEDED FOR ARRAY WSFIT =      1196

OPTIMISATION RUN WITH    43 OBSERVATIONS AND  10 VARIED PARAMETERS

TIME  0.0000000E+00  TCPU      1.76      0.00
  STP      0  TRY      0  DER       0  JAC     0  LUD      0  RST     0

AFTER   0 ITERATIONS AND   1 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  2.0778E+00  -0.7635    0.3589    0.4330
  3.9723E+00   0.6952    0.2129    0.4330
  2.2333E+01   1.7398    0.6798    0.4330
  1.6905E+00   1.2135    0.0416    0.4330
  1.4038E+00  -1.5521    0.4272    0.3727
  5.1793E+00  -1.5545    0.6708    0.4000
  2.1281E+00   0.2751    0.4217    0.3727
  4.4540E+00   1.6915    0.5113    0.4330
  3.1555E+00   0.6005    0.2005    0.3727
TOTAL RESIDUAL SUM OF SQUARES IS      4.63942E+01
MEAN ABSOLUTE CORRELATION INDEX IS   1.12064
VARY(  1):K6T         =  3.9900E+00
VARY(  2):K7T         =  6.8500E+00
VARY(  3):K8T         =  4.2800E+00
VARY(  4):K9T         =  3.6800E-01
VARY(  5):K5C         =  5.0600E+00
VARY(  6):K6C         =  2.8200E+00
VARY(  7):K7C         =  1.0550E+00
VARY(  8):K8C         =  1.4500E+00
VARY(  9):K9C         =  3.1800E-01
VARY( 10):K13         =  1.0000E+07
VARY( 11):VMT         =  4.6822E-07
```

```
VARY( 12):K5T        =  1.4347E+01
VARY( 13):K10T       =  2.1178E-03
VARY( 14):KE3T       =  3.5840E+00
VARY( 15):KE4T       =  1.6005E+00
VARY( 16):KE5T       =  1.7358E+02
VARY( 17):KMC        =  4.3429E-05
VARY( 18):VMC        =  7.1695E-08
VARY( 19):VMTVMC     =  6.5308E+00
VARY( 20):K10C       =  4.1516E-04
VARY( 21):KE3C       =  1.7943E+00
VARY( 22):KE4C       =  7.2759E-01
VARY( 23):KE5C       =  7.6442E+02
VARY( 24):KE6C       =  1.0000E-04
VARY( 25):KE7        =  1.0000E+00
        1 SIMULATION RUNS COMPLETED
        2 SIMULATION RUNS COMPLETED
        3 SIMULATION RUNS COMPLETED
        4 SIMULATION RUNS COMPLETED
        5 SIMULATION RUNS COMPLETED
        6 SIMULATION RUNS COMPLETED
        7 SIMULATION RUNS COMPLETED
        8 SIMULATION RUNS COMPLETED
        9 SIMULATION RUNS COMPLETED
       10 SIMULATION RUNS COMPLETED

AFTER   0 ITERATIONS AND  11 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI      AUCO      AUCR
  2.5469E+00  -1.1760    0.4932    0.4330
  3.8542E+00   0.5789    0.1856    0.4330
  1.8680E+01   1.7051    0.6682    0.4330
  1.4892E+00   1.0238   -0.0346    0.4330
  2.0978E+00  -1.9379    0.5389    0.3727
  6.8345E+00  -1.6947    0.7019    0.4000
  2.0596E+00  -0.2894    0.3734    0.3727
  3.2505E+00   1.6176    0.4367    0.4330
  3.4615E+00  -0.0435    0.2197    0.3727
TOTAL RESIDUAL SUM OF SQUARES IS      4.42741E+01
MEAN ABSOLUTE CORRELATION INDEX IS   1.11854
VARY(  1):K6T        =  3.9900E+00
VARY(  2):K7T        =  7.0234E+00
VARY(  3):K8T        =  4.2800E+00
VARY(  4):K9T        =  3.6800E-01
VARY(  5):K5C        =  5.0600E+00
VARY(  6):K6C        =  2.8200E+00
VARY(  7):K7C        =  1.0817E+00
VARY(  8):K8C        =  1.4500E+00
VARY(  9):K9C        =  3.1800E-01
VARY( 10):K13        =  1.0253E+07
VARY( 11):VMT        =  4.7872E-07
VARY( 12):K5T        =  1.4481E+01
VARY( 13):K10T       =  2.1916E-03
VARY( 14):KE3T       =  3.6293E+00
VARY( 15):KE4T       =  1.6410E+00
VARY( 16):KE5T       =  1.6791E+02
VARY( 17):KMC        =  4.3599E-05
VARY( 18):VMC        =  7.3291E-08
VARY( 19):VMTVMC     =  6.5317E+00
VARY( 20):K10C       =  4.2567E-04
VARY( 21):KE3C       =  1.7943E+00
VARY( 22):KE4C       =  7.4601E-01
VARY( 23):KE5C       =  7.2862E+02
VARY( 24):KE6C       =  1.0000E-04
```

```
VARY( 25):KE7        =  1.0253E+00
       11 SIMULATION RUNS COMPLETED
       12 SIMULATION RUNS COMPLETED
       13 SIMULATION RUNS COMPLETED
       14 SIMULATION RUNS COMPLETED
       15 SIMULATION RUNS COMPLETED
       16 SIMULATION RUNS COMPLETED
       17 SIMULATION RUNS COMPLETED
       18 SIMULATION RUNS COMPLETED
       19 SIMULATION RUNS COMPLETED
       20 SIMULATION RUNS COMPLETED
       20 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS              33.5238820
COMPARED WITH EXPECTED RANGE   19.0349   50.7283
FOR   33 DEGREES OF FREEDOM
ANALYSIS OF VARIANCE PROCEEDS ON ASSUMPTION OF GOOD FIT
BUT INSPECT RESIDUALS FOR EVIDENCE OF SYSTEMATIC ERROR

AFTER   0 ITERATIONS AND   1 RUNS, INDIVIDUAL CURVE STATISTICS ARE
     RSQS         CORI     AUCO     AUCR
  2.9173E+00  -1.5319   0.6313   0.4330
  3.9771E+00   0.5879   0.2597   0.4330
  7.3652E+00   1.5504   0.6161   0.4330
  3.6198E+00   1.6875   0.2349   0.4330
  1.4726E+00  -1.9951   0.4943   0.3727
  5.5832E+00  -1.8185   0.7474   0.4000
  3.4262E+00  -0.3042   0.3935   0.3727
  1.8610E+00   0.8033   0.0909   0.4330
  3.3015E+00   0.3663   0.2057   0.3727
TOTAL RESIDUAL SUM OF SQUARES IS      3.35239E+01
MEAN ABSOLUTE CORRELATION INDEX IS   1.18277
VARY(  1):K6T        =  8.2118E+00
VARY(  2):K7T        =  3.8738E+00
VARY(  3):K8T        =  1.8881E+00
VARY(  4):K9T        =  3.5261E-01
VARY(  5):K5C        =  1.0509E+01
VARY(  6):K6C        =  4.5499E+00
VARY(  7):K7C        =  7.3233E-01
VARY(  8):K8C        =  9.5108E-01
VARY(  9):K9C        =  2.4070E-01
VARY( 10):K13        =  1.8824E+07
VARY( 11):VMT        =  3.1034E-07
VARY( 12):K5T        =  2.4272E+01
VARY( 13):K10T       =  2.1383E-03
VARY( 14):KE3T       =  2.9558E+00
VARY( 15):KE4T       =  2.0517E+00
VARY( 16):KE5T       =  1.6490E+02
VARY( 17):KMC        =  3.3500E-05
VARY( 18):VMC        =  5.8483E-08
VARY( 19):VMTVMC     =  5.3064E+00
VARY( 20):K10C       =  4.2807E-04
VARY( 21):KE3C       =  2.3097E+00
VARY( 22):KE4C       =  7.6999E-01
VARY( 23):KE5C       =  5.6229E+02
VARY( 24):KE6C       =  1.0000E-04
VARY( 25):KE7        =  1.8824E+00
        1 SIMULATION RUNS COMPLETED
        2 SIMULATION RUNS COMPLETED
        3 SIMULATION RUNS COMPLETED
        4 SIMULATION RUNS COMPLETED
        5 SIMULATION RUNS COMPLETED
```

```
       6 SIMULATION RUNS COMPLETED
       7 SIMULATION RUNS COMPLETED
       8 SIMULATION RUNS COMPLETED
       9 SIMULATION RUNS COMPLETED
      10 SIMULATION RUNS COMPLETED
      11 SIMULATION RUNS COMPLETED
      11 SIMULATION RUNS COMPLETED

PARAMETER NUMBER, TYPE,     NAME,    VALUE
             1   PRI.     K6T        8.2118E+00
             2   PRI.     K7T        3.8738E+00
             3   PRI.     K8T        1.8881E+00
             4   PRI.     K9T        3.5261E-01
             5   PRI.     K5C        1.0509E+01
             6   PRI.     K6C        4.5499E+00
             7   PRI.     K7C        7.3233E-01
             8   PRI.     K8C        9.5108E-01
             9   PRI.     K9C        2.4070E-01
            10   PRI.     K13        1.8824E+07
            11   SEC.     VMT        3.1034E-07
            12   SEC.     K5T        2.4272E+01
            13   SEC.     K10T       2.1383E-03
            14   SEC.     KE3T       2.9558E+00
            15   SEC.     KE4T       2.0517E+00
            16   SEC.     KE5T       1.6490E+02
            17   SEC.     KMC        3.3500E-05
            18   SEC.     VMC        5.8483E-08
            19   SEC.     VMTVMC     5.3064E+00
            20   SEC.     K10C       4.2807E-04
            21   SEC.     KE3C       2.3097E+00
            22   SEC.     KE4C       7.6999E-01
            23   SEC.     KE5C       5.6229E+02
            24   SEC.     KE6C       1.0000E-04
            25   SEC.     KE7        1.8824E+00

DEPENDENCE MATRIX OF LN(P(I)) ON LN(P(J)), J =
      1          2          3          4          5          6          7          8          9          10

  I

 11 -2.602E-13  1.004E+00 -1.748E-13 -1.748E-13  3.343E-01 -2.883E-01 -4.715E-02 -1.748E-13 -1.748E-13 -1.748E-13

 12  6.797E-01  3.325E-01 -1.811E-14 -1.811E-14  1.006E+00 -8.669E-01 -1.418E-01 -1.811E-14 -1.811E-14 -1.811E-14

 13 -3.203E-01  1.332E+00 -1.000E+00  1.000E+00  1.006E+00 -8.669E-01 -1.418E-01  8.088E-14  8.088E-14  8.088E-14

 14 -1.000E+00  6.797E-01  3.325E-01 -3.957E-14 -3.957E-14  1.006E+00 -8.669E-01 -1.418E-01 -3.957E-14 -3.957E-14

 15 -1.759E-14  1.000E+00 -1.000E+00 -1.759E-14 -1.759E-14 -1.759E-14 -1.759E-14 -1.759E-14 -1.759E-14 -1.759E-14

 16  1.743E-13  3.203E-01 -1.332E+00  2.000E+00 -1.000E+00 -1.006E+00  8.669E-01  1.418E-01  1.743E-13  1.743E-13

 17 -1.405E-13 -1.405E-13 -1.405E-13 -1.405E-13 -6.643E-01  5.719E-01  9.359E-02 -1.405E-13 -1.405E-13 -1.405E-13

 18 -3.383E-13 -3.383E-13 -3.383E-13 -3.383E-13  3.343E-01 -2.883E-01  9.536E-01 -3.383E-13 -3.383E-13 -3.383E-13
```

```
19 -1.162E-13  1.004E+00 -3.253E-14 -3.253E-14 -1.162E-13 -1.999E-13 -1.001E+00 -3.253E-14 -3.253E-14 -3.253E-14

20 -1.605E-13 -1.605E-13 -1.605E-13 -1.605E-13  1.000E+00 -1.000E+00  1.000E+00 -1.000E+00  1.000E+00 -1.605E-13

21  1.930E-14  1.930E-14  1.930E-14  1.930E-14  1.000E+00 -1.000E+00  1.930E-14  1.930E-14  1.930E-14  1.930E-14

22  5.529E-15  5.529E-15  5.529E-15  5.529E-15  5.529E-15  5.529E-15  1.000E+00 -1.000E+00  5.529E-15  5.529E-15

23  4.489E-14  4.489E-14  4.489E-14  4.489E-14  4.489E-14 -1.000E+00  1.000E+00 -1.000E+00  2.000E+00 -1.000E+00

24 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13 -4.302E-13

25  7.183E-15  7.183E-15  7.183E-15  7.183E-15  7.183E-15  7.183E-15  7.183E-15  7.183E-15  7.183E-15  1.000E+00

RESIDUAL SUM OF SQUARES =  3.3524E+01 FOR   33 DEGREES OF FREEDOM

SINGULAR VALUES OF DEPENDENCE MATRIX AND STRUCTURE OF COMPONENTS

   VALUES     52.978   45.525    8.310    7.598    5.016    4.220    2.201    1.765    0.880    0.476

K6T          0.08146  0.04020 -0.06468  0.17781 -0.03294  0.21134 -0.46154 -0.02680 -0.73350  0.39805

K7T         -0.69277 -0.24695  0.11708  0.12374  0.06282  0.49935 -0.28659 -0.20716  0.20674 -0.09474

K8T          0.38678  0.12617 -0.11273  0.51898  0.06969 -0.12433 -0.50397 -0.31734  0.39275 -0.15225

K9T         -0.15161 -0.04970  0.09136 -0.61416 -0.16556 -0.48599 -0.53381 -0.19871  0.03051 -0.00652

K5C         -0.43689  0.30234 -0.32417  0.19506  0.15323 -0.37395  0.19150 -0.22168  0.11960  0.55781

K6C          0.37788 -0.28183  0.19132 -0.28281  0.15228  0.29562  0.10021 -0.35025  0.28425  0.58061

K7C          0.02085  0.71107  0.12754 -0.30378  0.48483  0.26556 -0.02424 -0.18930 -0.07264 -0.19523

K8C          0.03179 -0.44437 -0.36538 -0.04669  0.45552 -0.14894  0.16883 -0.44531 -0.32695 -0.32008

K9C         -0.01692  0.16390  0.31251  0.13075 -0.55165 -0.01069  0.29991 -0.63554 -0.20495 -0.13340

K13         -0.07385 -0.13496  0.75626  0.27741  0.41128 -0.36646  0.01271  0.07946 -0.12155  0.05243
VARIANCE ESTIMATE =       1.0158752

FITTED VALUES, ACCURACIES AND CONFIDENCE LIMITS
NO.   NAME       VALUE        SDLN   5 PERCENT  95 PERCENT
 1 K6T         8.2118E+00  1.2104  1.1212E+00  6.0142E+01
 2 K7T         3.8738E+00  0.3777  2.0812E+00  7.2104E+00
 3 K8T         1.8881E+00  0.6314  6.6831E-01  5.3344E+00
 4 K9T         3.5261E-01  0.3088  2.1217E-01  5.8603E-01
 5 K5C         1.0509E+01  1.2039  1.4504E+00  7.6140E+01
 6 K6C         4.5499E+00  1.2916  5.4362E-01  3.8081E+01
 7 K7C         7.3233E-01  0.4531  3.4754E-01  1.5431E+00
 8 K8C         9.5108E-01  0.8260  2.4442E-01  3.7008E+00
 9 K9C         2.4070E-01  0.5473  9.7831E-02  5.9222E-01
10 K13         1.8824E+07  0.2410  1.2663E+07  2.7984E+07
```

11	VMT	3.1034E-07	0.2946	1.9115E-07	5.0383E-07
12	K5T	2.4272E+01	0.9497	5.0891E+00	1.1576E+02
13	K10T	2.1383E-03	0.1691	1.6190E-03	2.8242E-03
14	KE3T	2.9558E+00	1.7297	1.7176E-01	5.0866E+01
15	KE4T	2.0517E+00	0.3167	1.2185E+00	3.4546E+00
16	KE5T	1.6490E+02	2.7866	1.6843E+00	1.6145E+04
17	KMC	3.3500E-05	0.1990	2.4146E-05	4.6477E-05
18	VMC	5.8483E-08	0.3987	3.0350E-08	1.1269E-07
19	VMTVMC	5.3064E+00	0.4208	2.6556E+00	1.0603E+01
20	K10C	4.2807E-04	0.2666	2.7610E-04	6.6370E-04
21	KE3C	2.3097E+00	0.2810	1.4548E+00	3.6670E+00
22	KE4C	7.6999E-01	0.4464	3.6949E-01	1.6046E+00
23	KE5C	5.6229E+02	1.7744	3.0362E+01	1.0413E+04
24	KE6C	1.0000E-04	0.0000	1.0000E-04	1.0000E-04
25	KE7	1.8824E+00	0.2410	1.2663E+00	2.7984E+00

CORRELATION MATRIX COMPONENTS

COLUMN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	17	18	19	20	21	22	23	24	25							
ROW 1	1.000	-0.727	-0.783	0.013	0.590	0.485	-0.498	-0.271	-0.096	0.699	-0.707	0.985	-0.305	-0.407	0.693	-0.355
	-0.678	-0.397	-0.119	0.113	0.299	-0.003	-0.508	-0.669	0.699							
ROW 2	-0.727	1.000	0.925	0.353	-0.465	-0.294	0.501	0.203	0.121	-0.756	0.987	-0.778	0.262	0.421	-0.651	0.260
	0.883	0.348	0.361	-0.204	-0.640	0.132	0.425	0.217	-0.756							
ROW 3	-0.783	0.925	1.000	0.484	-0.409	-0.281	0.400	0.155	0.054	-0.677	0.956	-0.775	0.469	0.494	-0.891	0.216
	0.689	0.283	0.402	-0.177	-0.464	0.120	0.360	0.230	-0.677							
ROW 4	0.013	0.353	0.484	1.000	-0.032	0.000	0.076	0.029	0.037	-0.101	0.405	0.012	0.785	0.083	-0.544	0.117
	0.143	0.050	0.236	-0.026	-0.135	0.024	0.043	-0.359	-0.101							
ROW 5	0.590	-0.465	-0.409	-0.032	1.000	0.977	-0.901	-0.810	-0.474	0.395	-0.402	0.635	-0.230	0.462	0.262	-0.952
	-0.584	-0.879	0.552	-0.211	-0.207	0.585	-0.910	-0.690	0.395							
ROW 6	0.485	-0.294	-0.281	0.000	0.977	1.000	-0.863	-0.845	-0.493	0.255	-0.245	0.507	-0.234	0.587	0.209	-0.974
	-0.399	-0.883	0.665	-0.293	-0.410	0.688	-0.894	-0.678	0.255							
ROW 7	-0.498	0.501	0.400	0.076	-0.901	-0.863	1.000	0.920	0.657	-0.338	0.433	-0.564	0.159	-0.467	-0.201	0.889
	0.629	0.981	-0.626	0.315	0.109	-0.686	0.907	0.390	-0.338							
ROW 8	-0.271	0.203	0.155	0.029	-0.810	-0.845	0.920	1.000	0.820	-0.139	0.157	-0.308	0.143	-0.673	-0.066	0.884
	0.315	0.968	-0.808	0.583	0.413	-0.917	0.909	0.256	-0.139							
ROW 9	-0.096	0.121	0.054	0.037	-0.474	-0.493	0.657	0.820	1.000	-0.145	0.084	-0.135	0.066	-0.483	0.037	0.559
	0.215	0.694	-0.599	0.879	0.235	-0.849	0.782	-0.118	-0.145							
ROW 10	0.699	-0.756	-0.677	-0.101	0.395	0.255	-0.338	-0.139	-0.145	1.000	-0.732	0.731	-0.244	-0.406	0.448	-0.194
	-0.711	-0.206	-0.317	0.106	0.518	-0.085	-0.433	-0.369	1.000							
ROW 11	-0.707	0.987	0.956	0.405	-0.402	-0.245	0.433	0.157	0.084	-0.732	1.000	-0.735	0.306	0.464	-0.729	0.200
	0.798	0.292	0.423	-0.204	-0.596	0.150	0.368	0.166	-0.732							
ROW 12	0.985	-0.778	-0.775	0.012	0.635	0.507	-0.564	-0.308	-0.135	0.731	-0.735	1.000	-0.258	-0.369	0.616	-0.400
	-0.787	-0.445	-0.093	0.125	0.387	-0.004	-0.553	-0.658	0.731							
ROW 13	-0.305	0.262	0.469	0.785	-0.230	-0.234	0.159	0.143	0.066	-0.244	0.306	-0.258	1.000	0.088	-0.623	0.280
	0.090	0.158	0.065	0.055	0.089	-0.104	0.218	0.066	-0.244							
ROW 14	-0.407	0.421	0.494	0.083	0.462	0.587	-0.467	-0.673	-0.483	-0.406	0.464	-0.369	0.088	1.000	-0.483	-0.680
	0.225	-0.588	0.882	-0.462	-0.722	0.771	-0.477	-0.088	-0.406							

```
ROW 15  0.693 -0.651 -0.891 -0.544  0.262  0.209 -0.201 -0.066  0.037  0.448 -0.729  0.616 -0.623 -0.483  1.000 -0.121
       -0.320 -0.148 -0.370  0.110  0.162 -0.081 -0.210 -0.199  0.448

ROW 16 -0.355  0.260  0.216  0.117 -0.952 -0.974  0.889  0.884  0.559 -0.194  0.200 -0.400  0.280 -0.680 -0.121  1.000
        0.402  0.912 -0.724  0.336  0.396 -0.734  0.895  0.541 -0.194

ROW 17 -0.678  0.883  0.689  0.143 -0.584 -0.399  0.629  0.315  0.215 -0.711  0.798 -0.787  0.090  0.225 -0.320  0.402
        1.000  0.464  0.119 -0.172 -0.669  0.055  0.533  0.339 -0.711

ROW 18 -0.397  0.348  0.283  0.050 -0.879 -0.883  0.981  0.968  0.694 -0.206  0.292 -0.445  0.158 -0.588 -0.148  0.912
        0.464  1.000 -0.743  0.401  0.292 -0.796  0.899  0.360 -0.206

ROW 19 -0.119  0.361  0.402  0.236  0.552  0.665 -0.626 -0.808 -0.599 -0.317  0.423 -0.093  0.065  0.882 -0.370 -0.724
        0.119 -0.743  1.000 -0.523 -0.694  0.859 -0.595 -0.225 -0.317

ROW 20  0.113 -0.204 -0.177 -0.026 -0.211 -0.293  0.315  0.583  0.879  0.106 -0.204  0.125  0.055 -0.462  0.110  0.336
       -0.172  0.401 -0.523  1.000  0.442 -0.759  0.550 -0.202  0.106

ROW 21  0.299 -0.640 -0.464 -0.135 -0.207 -0.410  0.109  0.413  0.235  0.518 -0.596  0.387  0.089 -0.722  0.162  0.396
       -0.669  0.292 -0.694  0.442  1.000 -0.655  0.209  0.161  0.518

ROW 22 -0.003  0.132  0.120  0.024  0.585  0.688 -0.686 -0.917 -0.849 -0.085  0.150 -0.004 -0.104  0.771 -0.081 -0.734
        0.055 -0.796  0.859 -0.759 -0.655  1.000 -0.761 -0.077 -0.085

ROW 23 -0.508  0.425  0.360  0.043 -0.910 -0.894  0.907  0.909  0.782 -0.433  0.368 -0.553  0.218 -0.477 -0.210  0.895
        0.533  0.899 -0.595  0.550  0.209 -0.761  1.000  0.452 -0.433

ROW 24 -0.669  0.217  0.230 -0.359 -0.690 -0.678  0.390  0.256 -0.118 -0.369  0.166 -0.658  0.066 -0.088 -0.199  0.541
        0.339  0.360 -0.225 -0.202  0.161 -0.077  0.452  1.000 -0.369

ROW 25  0.699 -0.756 -0.677 -0.101  0.395  0.255 -0.338 -0.139 -0.145  1.000 -0.732  0.731 -0.244 -0.406  0.448 -0.194
       -0.711 -0.206 -0.317  0.106  0.518 -0.085 -0.433 -0.369  1.000


***** COMMAND INTERPRETER RE-ENTERED

nofit;
param sx1 50 sx2 100;
param<#1> xdto:t2020o t2080o t2000o t0080o t8000o t0500o;
param<#1> xdco:c2020o c2080o c2000o c0080o c0020o c4000o;
gstream 1 2 100;
t2020    sx1  a;
t2080    sx1  b;
c2080    sx1  c;
c2020    sx1  d;
t2020o   sx1  1;
t2080o   sx1  2;
c2080o   sx1  3;
c2020o   sx1  4;
**;
gstream 2 3 100;
t8000    sx2  a;
t2000    sx2  b;
c0080    sx2  c;
t0500    sx2  d;
c0020    sx2  e;
t8000o   sx2  1;
t2000o   sx2  2;
c0080o   sx2  3;
t0500o   sx2  4;
c0020o   sx2  5;
```

```
**;

pstream 2 10;
Time t2020 t2080 c2020 c2080 t8000 t2000 c0080 t0500 c0020;
**;
compile plot;

 LOCN   INSTRUCTION

  806:call resid;
  808:gstream 1;
  814:write 1=13:100, time
  814:#t t2020 #t t2020o #t t2080 #t t2080o
  814:#t c2080 #t c2080o #t c2020 #t c2020o%;
  873:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS           2282     17718      20000
REAL SCALARS AND ARRAYS              4483     15517      20000
INDEXES AND SECRETS                    99       400        499
CODE ARRAY                            874      9126      10000
NAME DICTIONARY                       154       846       1000
TEXT ARRAY                             81      4919       5000



compile graph;

 LOCN   INSTRUCTION

  878:call resid;
  880:gstream 2;
  886:write 1=14:200, time
  886:#t t8000 #t t8000o #t t2000 #t t2000o
  886:#t c0080 #t c0080o #t t0500 #t t0500o #t c0020 #t c0020o%;
  957:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                          IN USE      FREE      TOTAL
INTEGER SCALARS AND ARRAYS           2283     17717      20000
REAL SCALARS AND ARRAYS              4483     15517      20000
INDEXES AND SECRETS                    99       400        499
CODE ARRAY                            958      9042      10000
NAME DICTIONARY                       155       845       1000
TEXT ARRAY                             81      4919       5000



compile table;

 LOCN   INSTRUCTION

  962:call resid;
  964:pstream 2;
  966:**;
```

```
NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	2283	17717	20000
REAL SCALARS AND ARRAYS	4483	15517	20000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	967	9033	10000
NAME DICTIONARY	155	845	1000
TEXT ARRAY	81	4919	5000

```
compile format;

 LOCN   INSTRUCTION

  971:100(f7,2,n8(b1,f7,2));
  990:200(f7,2,n8(b1,f7,2));
 1009:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	2283	17717	20000
REAL SCALARS AND ARRAYS	4483	15517	20000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	1010	8990	10000
NAME DICTIONARY	155	845	1000
TEXT ARRAY	81	4919	5000

```
compile set;

 LOCN   INSTRUCTION

 1014:for #20=#21;
 1021:t2020o=vobs<0,#20>;
 1027:c2020o=vobs<1,#20>;
 1033:t2080o=vobs<2,#20>;
 1039:c2080o=vobs<3,#20>;
 1045:t2000o=vobs<4,#20>;
 1051:c0080o=vobs<5,#20>;
 1057:t8000o=vobs<6,#20>;
 1063:c0020o=vobs<7,#20>;
 1069:t0500o=vobs<8,#20>;
 1075:#21=#21+1;
 1084:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	2283	17717	20000
REAL SCALARS AND ARRAYS	4483	15517	20000
INDEXES AND SECRETS	100	399	499

```
CODE ARRAY                              1095      8905     10000
NAME DICTIONARY                          155       845      1000
TEXT ARRAY                                81      4919      5000



compile clear;

 LOCN   INSTRUCTION

 1099:array<#1> ws;
 1102:xdto=0;
 1109:xdco=0;
 1116:array;

MAXIMUM NESTING DEPTHS:   ALLOWED SO FAR        5
                          ENCOUNTERED           1
                          ALLOWED HENCEFORTH   20

 1118:**;

NO. OF VARIABLES =   140
NO. IMPLICIT     =     0
NO. OF ERRORS    =     0

SPACE FOR                           IN USE      FREE     TOTAL
INTEGER SCALARS AND ARRAYS            2283     17717     20000
REAL SCALARS AND ARRAYS               4483     15517     20000
INDEXES AND SECRETS                    100       399       499
CODE ARRAY                            1119      8881     10000
NAME DICTIONARY                        155       845      1000
TEXT ARRAY                              81      4919      5000



when
time=.39 .49 .79  .99 1.19 1.59 1.99 2.99 3.99 4.99 6.99 call set;
time=.41 .51 .81 1.01 1.21 1.61 2.01 3.01 4.01 5.01 7.01%call clear;
time=0+.1*40 call plot;
time=0+.2*35 call graph;
time=0+.2*50% call table;
**;
begin;

TIME  0.0000000E+00  TCPU    142.75      0.00
  STP   3629  TRY   3722  DER    5070  JAC    31  LUD    733  RST    31


***** COMMAND INTERPRETER RE-ENTERED

stop;

***** FORMATTED WRITE STRUCTURE NUMBER  1 CLEARED

TIME  7.0100000E+00  TCPU    150.05      0.00
  STP    116  TRY    119  DER     162  JAC     1  LUD     23  RST     1

 MAXIMUM SPACE REQUIREMENTS -
NAMES   155,  INTS     3875,  REALS     4573,  CODE   1119,  VBLS    140

MEMORY PAGES        0     492       0     192     960     960       0
```

```
PLOTTING PARAMETERS FOR GRAPH STREAM  1
INDEPENDENT VARIABLE :TIME       (PLOTTED VERTICALLY)
DEPENDENT VARIABLES (PLOTTED HORIZONTALLY):
 T2020        SCALE  5.0000E+01   a
 T2080        SCALE  5.0000E+01   b
 C2080        SCALE  5.0000E+01   c
 C2020        SCALE  5.0000E+01   d
 T20200       SCALE  5.0000E+01   1
 T20800       SCALE  5.0000E+01   2
 C20800       SCALE  5.0000E+01   3
 C20200       SCALE  5.0000E+01   4
          0                                    0.5                                    1
 0.000E+00 4---------:---------:---------:---------:---------:---------:---------:---------:---------:
 1.000E-01 4dc   b a :         :         :         :         :         :         :         :         :
 2.000E-01 4 d  c    :    b    a         :         :         :         :         :         :         :
 3.000E-01 4   d    c:         :  b      :a        :         :         :         :         :         :
 4.000E-01 4    d    : c       :       b :        a:         :         :         :         :         :
 5.000E-01 :----4d---:------c---3---------:-2-------1-----a---:---------:---------:---------:---------:
 6.000E-01 4      d  :       c  :         :    b    :         :a        :         :         :         :
 7.000E-01 4       d :         c         :      b   :         :   a     :         :         :         :
 8.000E-01 4        d:         : c       :         :b        :       a:         :         :         :
 9.000E-01 4         d         :   c     :         :  b      :         : a       :         :         :
 1.000E+00 :---------4d--------:-----c-3-:---------:---b---2-:----1----:---a-----:---------:---------:
 1.100E+00 4         :d        :      c  :         :    b    :         :     a   :         :         :
 1.200E+00 4         : d       :       c :         :     b   :         :      a: :         :         :
 1.300E+00 4         : d       :         c         :      b  :         :       a :         :         :
 1.400E+00 4         :  d      :         c         :       b:         :        : a       :         :
 1.500E+00 4---------:--d------:---------:c--------:---------b---------:---------:--a------:---------:
 1.600E+00 4         :   d     :         : c       :         :b        :         :   a     :         :
 1.700E+00 4         :   d     :         :  c      :         : b       :         :    a    :         :
 1.800E+00 4         :   d     :         :   c     :         :  b      :         :     a   :         :
 1.900E+00 4         :   d     :         :   c     :         :   b     :         :      a: :         :
 2.000E+00 :---------:----d-4--:---------:---3-----:---------:----b----:------2---:-----1----a---------:
 2.100E+00 4         :    d    :         :    c    :         :    b    :         :        :a       :         :
 2.200E+00 4         :    d    :         :     c   :         :     b   :         :        : a      :         :
 2.300E+00 4         :    d    :         :      c  :         :      b  :         :        :  a     :         :
 2.400E+00 4         :     d   :         :      c  :         :       b :         :        :   a    :         :
 2.500E+00 4---------:-----d---:---------:------c--:---------:-------b-:---------:---------:----a----:---------:
 2.600E+00 4         :     d   :         :       c :         :        b:         :         :    a   :         :
 2.700E+00 4         :     d   :         :       c :         :         b         :         :     a  :         :
 2.800E+00 4         :     d   :         :        c:         :         b         :         :      a :         :
 2.900E+00 4         :     d   :         :        c:         :         :b        :         :       a:         :
 3.000E+00 4---------:------d--:---------:--------c:---------:---------:-b-------:---------:---------a---------:
 3.100E+00 4         :      d  :         :         c         :         : b       :         :         :a        :
 3.200E+00 4         :      d  :         :         c         :         :  b      :         :         :a        :
 3.300E+00 4         :      d  :         :         :c        :         :   b     :         :         : a       :
 3.400E+00 4         :      d  :         :         :c        :         :   b     :         :         :  a      :
 3.500E+00 4---------:------d--:---------:---------:c--------:---------:----b----:---------:---------:---a-----:
 3.600E+00 4         :      d  :         :         :c        :         :    b    :         :         :    a    :
 3.700E+00 4         :      d  :         :         : c       :         :     b   :         :         :     a   :
 3:800E+00 4         :       d :         :         : c       :         :     b   :         :         :     a   :
 3.900E+00 4         :       d :         :         : c       :         :     b   :         :         :      a  :
 4.000E+00 :---------:-------d4:---------:---------:--c---3--:---------:-------b-:-------2-:---------:-------a-1
```

```
PLOTTING PARAMETERS FOR GRAPH STREAM  2
INDEPENDENT VARIABLE :TIME       (PLOTTED VERTICALLY)
DEPENDENT VARIABLES (PLOTTED HORIZONTALLY):
 T8000        SCALE  1.0000E+02   a
 T2000        SCALE  1.0000E+02   b
 C0080        SCALE  1.0000E+02   c
 T0500        SCALE  1.0000E+02   d
 C0020        SCALE  1.0000E+02   e
 T80000       SCALE  1.0000E+02   1
 T20000       SCALE  1.0000E+02   2
 C00800       SCALE  1.0000E+02   3
 T05000       SCALE  1.0000E+02   4
 C00200       SCALE  1.0000E+02   5
          0                                        0.5                                        1
0.000E+00 5---------:---------:---------:---------:---------:---------:---------:---------:---------:---------:
2.000E-01 5 e d     :b        : a       :         :         :         :         :         :         :         :
4.000E-01 5   e   dc: 4       :2 b      :         a:   1    :         :         :         :         :         :
6.000E-01 5      e  : dc      :         b         :        a:         :         :         :         :         :
8.000E-01 5        e:   d 4c  :         : 2  b    :         :  a1     :         :         :         :         :
1.000E+00 4---------:5-----d-3:---------:--------b:---------:-----a---:---------:---------:---------:---------:
1.200E+00 5         : e    4d : c       :       2 :b        :       a 1       :         :         :         :
1.400E+00 5         :  e     d   c      :         : b       :         a       :         :         :         :
1.600E+00 5         :    e   4d   c     :         :    b2   :         :1      :         :         :         :
1.800E+00 5         :     e  : d    c   :         :      b  :         : a     :         :         :         :
2.000E+00 5---------:------e--:--d----3-:---------:-------b-:---------:---a-----:---------:---------:---------:
2.200E+00 5         :       e :   d    c:         :         b         :    a    :         :         :         :
2.400E+00 5         :        e:    d    c         :         :b        :     a   :         :         :         :
2.600E+00 5         :         e:     d   :c       :         : b       :      a  :         :         :         :
2.800E+00 5         :          e      d  : c      :         :  b      :       a :         :         :         :
3.000E+00 :---------:---------5e-------43:--c------:---------:---2-----:-----1---a:---------:---------:---------:
3.200E+00 5         :          :e       d:   c    :         :    b    :         a         :         :         :
3.400E+00 5         :          :e       d:    c   :         :     b   :         :a        :         :         :
3.600E+00 5         :          : e       d    c   :         :      b  :         : a       :         :         :
3.800E+00 5         :          : e        :d    c  :         :       b :         :  a      :         :         :
4.000E+00 5---------:---------:--e-------:-d----c--:---------:--------b:---------:---a-----:---------:---------:
4.200E+00 5         :          : e       : d    c  :         :         b         :    a    :         :         :
4.400E+00 5         :          :  e      :  d    c :         :         :b        :    a    :         :         :
4.600E+00 5         :          :  e      :   d    c:         :         : b       :     a   :         :         :
4.800E+00 5         :          :  e      :    d   c:         :         :  b      :      a  :         :         :
5.000E+00 :---------:---------:----e-5---:---43----c---------:---------:2-b------1------a--:---------:---------:
5.200E+00 5         :          :     e   :     d  :c         :         :   b     :       a :         :         :
5.400E+00 5         :          :     e   :      d :c         :         :    b    :        a:         :         :
5.600E+00 5         :          :      e  :      d : c        :         :     b   :        a:         :         :
5.800E+00 5         :          :      e  :       d: c        :         :     b   :         a         :         :
6.000E+00 5---------:---------:-------e--:-------d-:--c------:---------:------b---:---------a---------:---------:
6.200E+00 5         :          :       e :        d:  c     :         :       b :         :a        :         :
6.400E+00 5         :          :        e:         d   c    :         :       b :         :a        :         :
6.600E+00 5         :          :        e:         d   c    :         :        b:         : a       :         :
6.800E+00 5         :          :        e:          :d   c  :         :         b         : a       :         :
7.000E+00 4---------:---------:---------e-5-------:d3---c---:---------:---------b---------:--a------:---------:
```

PRINT STREAM NO. 2

TIME	T2020	T2080	C2020	C2080	T8000	T2000	C0080	T0500	C0020
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.0000E-01	1.0214E+01	7.7098E+00	1.0183E+00	2.6881E+00	2.2711E+01	1.1405E+01	3.5778E+00	3.7608E+00	1.5563E+00
4.0000E-01	1.9613E+01	1.4001E+01	2.4860E+00	6.2346E+00	3.9271E+01	2.2770E+01	8.6831E+00	8.2667E+00	4.1586E+00
6.0000E-01	2.5520E+01	1.7790E+01	3.7147E+00	9.0805E+00	4.7889E+01	3.0228E+01	1.3047E+01	1.1764E+01	6.6331E+00
8.0000E-01	2.9432E+01	2.0321E+01	4.6592E+00	1.1246E+01	5.2716E+01	3.5209E+01	1.6576E+01	1.4462E+01	8.7696E+00
1.0000E+00	3.2203E+01	2.2149E+01	5.3804E+00	1.2904E+01	5.5747E+01	3.8717E+01	1.9436E+01	1.6601E+01	1.0582E+01
1.2000E+00	3.4300E+01	2.3553E+01	5.9410E+00	1.4203E+01	5.7915E+01	4.1347E+01	2.1785E+01	1.8353E+01	1.2125E+01
1.4000E+00	3.5988E+01	2.4694E+01	6.3881E+00	1.5245E+01	5.9651E+01	4.3447E+01	2.3746E+01	1.9835E+01	1.3452E+01
1.6000E+00	3.7419E+01	2.5669E+01	6.7543E+00	1.6103E+01	6.1154E+01	4.5218E+01	2.5412E+01	2.1127E+01	1.4609E+01
1.8000E+00	3.8687E+01	2.6536E+01	7.0618E+00	1.6827E+01	6.2518E+01	4.6780E+01	2.6855E+01	2.2285E+01	1.5631E+01

2.0000E+00	3.9848E+01	2.7335E+01	7.3258E+00	1.7452E+01	6.3793E+01	4.8204E+01	2.8126E+01	2.3346E+01	1.6547E+01
2.2000E+00	4.0934E+01	2.8086E+01	7.5570E+00	1.8001E+01	6.5001E+01	4.9531E+01	2.9266E+01	2.4336E+01	1.7378E+01
2.4000E+00	4.1965E+01	2.8803E+01	7.7624E+00	1.8492E+01	6.6149E+01	5.0785E+01	3.0303E+01	2.5272E+01	1.8142E+01
2.6000E+00	4.2953E+01	2.9494E+01	7.9476E+00	1.8938E+01	6.7246E+01	5.1981E+01	3.1260E+01	2.6167E+01	1.8852E+01
2.8000E+00	4.3904E+01	3.0163E+01	8.1164E+00	1.9348E+01	6.8296E+01	5.3127E+01	3.2154E+01	2.7029E+01	1.9517E+01
3.0000E+00	4.4825E+01	3.0815E+01	8.2721E+00	1.9729E+01	6.9301E+01	5.4233E+01	3.2999E+01	2.7864E+01	2.0148E+01
3.2000E+00	4.5717E+01	3.1451E+01	8.4164E+00	2.0085E+01	7.0265E+01	5.5298E+01	3.3802E+01	2.8676E+01	2.0749E+01
3.4000E+00	4.6583E+01	3.2071E+01	8.5515E+00	2.0422E+01	7.1188E+01	5.6329E+01	3.4572E+01	2.9469E+01	2.1327E+01
3.6000E+00	4.7422E+01	3.2675E+01	8.6783E+00	2.0741E+01	7.2073E+01	5.7324E+01	3.5313E+01	3.0243E+01	2.1884E+01
3.8000E+00	4.8237E+01	3.3264E+01	8.7981E+00	2.1045E+01	7.2921E+01	5.8287E+01	3.6030E+01	3.1000E+01	2.2423E+01
4.0000E+00	4.9028E+01	3.3838E+01	8.9117E+00	2.1335E+01	7.3733E+01	5.9219E+01	3.6726E+01	3.1742E+01	2.2948E+01
4.2000E+00	4.9796E+01	3.4398E+01	9.0197E+00	2.1613E+01	7.4511E+01	6.0120E+01	3.7402E+01	3.2468E+01	2.3460E+01
4.4000E+00	5.0541E+01	3.4944E+01	9.1229E+00	2.1881E+01	7.5257E+01	6.0992E+01	3.8063E+01	3.3181E+01	2.3960E+01
4.6000E+00	5.1264E+01	3.5475E+01	9.2217E+00	2.2139E+01	7.5972E+01	6.1836E+01	3.8707E+01	3.3880E+01	2.4450E+01
4.8000E+00	5.1966E+01	3.5993E+01	9.3165E+00	2.2388E+01	7.6657E+01	6.2654E+01	3.9339E+01	3.4566E+01	2.4931E+01
5.0000E+00	5.2647E+01	3.6497E+01	9.4076E+00	2.2629E+01	7.7313E+01	6.3445E+01	3.9957E+01	3.5240E+01	2.5403E+01
5.2000E+00	5.3308E+01	3.6988E+01	9.4953E+00	2.2862E+01	7.7942E+01	6.4211E+01	4.0562E+01	3.5901E+01	2.5868E+01
5.4000E+00	5.3949E+01	3.7466E+01	9.5800E+00	2.3088E+01	7.8545E+01	6.4953E+01	4.1156E+01	3.6550E+01	2.6325E+01
5.6000E+00	5.4571E+01	3.7930E+01	9.6614E+00	2.3307E+01	7.9123E+01	6.5670E+01	4.1738E+01	3.7187E+01	2.6774E+01
5.8000E+00	5.5174E+01	3.8383E+01	9.7401E+00	2.3519E+01	7.9676E+01	6.6365E+01	4.2308E+01	3.7812E+01	2.7217E+01
6.0000E+00	5.5759E+01	3.8823E+01	9.8162E+00	2.3725E+01	8.0207E+01	6.7038E+01	4.2869E+01	3.8426E+01	2.7654E+01
6.2000E+00	5.6326E+01	3.9250E+01	9.8896E+00	2.3925E+01	8.0715E+01	6.7688E+01	4.3418E+01	3.9028E+01	2.8084E+01
6.4000E+00	5.6876E+01	3.9666E+01	9.9606E+00	2.4119E+01	8.1202E+01	6.8318E+01	4.3958E+01	3.9620E+01	2.8508E+01
6.6000E+00	5.7410E+01	4.0071E+01	1.0029E+01	2.4307E+01	8.1669E+01	6.8928E+01	4.4487E+01	4.0201E+01	2.8926E+01
6.8000E+00	5.7927E+01	4.0465E+01	1.0096E+01	2.4490E+01	8.2116E+01	6.9518E+01	4.5007E+01	4.0771E+01	2.9338E+01
7.0000E+00	5.8429E+01	4.0848E+01	1.0160E+01	2.4668E+01	8.2545E+01	7.0090E+01	4.5517E+01	4.1330E+01	2.9745E+01

PRINT STREAM NO. 2

TIME	T2020	T2080	C2020	C2080	T8000	T2000	C0080	T0500	C0020
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.0000E-01	1.0214E+01	7.7098E+00	1.0183E+00	2.6881E+00	2.2711E+01	1.1405E+01	3.5778E+00	3.7608E+00	1.5563E+00
4.0000E-01	1.9613E+01	1.4001E+01	2.4860E+00	6.2346E+00	3.9271E+01	2.2770E+01	8.6831E+00	8.2667E+00	4.1586E+00
6.0000E-01	2.5520E+01	1.7790E+01	3.7147E+00	9.0805E+00	4.7889E+01	3.0228E+01	1.3047E+01	1.1764E+01	6.6331E+00
8.0000E-01	2.9432E+01	2.0321E+01	4.6592E+00	1.1246E+01	5.2716E+01	3.5209E+01	1.6576E+01	1.4462E+01	8.7696E+00
1.0000E+00	3.2203E+01	2.2149E+01	5.3804E+00	1.2904E+01	5.5747E+01	3.8717E+01	1.9436E+01	1.6601E+01	1.0582E+01
1.2000E+00	3.4300E+01	2.3553E+01	5.9410E+00	1.4203E+01	5.7915E+01	4.1347E+01	2.1785E+01	1.8353E+01	1.2125E+01
1.4000E+00	3.5988E+01	2.4694E+01	6.3881E+00	1.5245E+01	5.9651E+01	4.3447E+01	2.3746E+01	1.9835E+01	1.3452E+01
1.6000E+00	3.7419E+01	2.5669E+01	6.7543E+00	1.6103E+01	6.1154E+01	4.5218E+01	2.5412E+01	2.1127E+01	1.4609E+01
1.8000E+00	3.8687E+01	2.6536E+01	7.0618E+00	1.6827E+01	6.2518E+01	4.6780E+01	2.6855E+01	2.2285E+01	1.5631E+01
2.0000E+00	3.9848E+01	2.7335E+01	7.3258E+00	1.7452E+01	6.3793E+01	4.8204E+01	2.8126E+01	2.3346E+01	1.6547E+01
2.2000E+00	4.0934E+01	2.8086E+01	7.5570E+00	1.8001E+01	6.5001E+01	4.9531E+01	2.9266E+01	2.4336E+01	1.7378E+01
2.4000E+00	4.1965E+01	2.8803E+01	7.7624E+00	1.8492E+01	6.6149E+01	5.0785E+01	3.0303E+01	2.5272E+01	1.8142E+01
2.6000E+00	4.2953E+01	2.9494E+01	7.9476E+00	1.8938E+01	6.7246E+01	5.1981E+01	3.1260E+01	2.6167E+01	1.8852E+01
2.8000E+00	4.3904E+01	3.0163E+01	8.1164E+00	1.9348E+01	6.8296E+01	5.3127E+01	3.2154E+01	2.7029E+01	1.9517E+01
3.0000E+00	4.4825E+01	3.0815E+01	8.2721E+00	1.9729E+01	6.9301E+01	5.4233E+01	3.2999E+01	2.7864E+01	2.0148E+01
3.2000E+00	4.5717E+01	3.1451E+01	8.4164E+00	2.0085E+01	7.0265E+01	5.5298E+01	3.3802E+01	2.8676E+01	2.0749E+01
3.4000E+00	4.6583E+01	3.2071E+01	8.5515E+00	2.0422E+01	7.1188E+01	5.6329E+01	3.4572E+01	2.9469E+01	2.1327E+01
3.6000E+00	4.7422E+01	3.2675E+01	8.6783E+00	2.0741E+01	7.2073E+01	5.7324E+01	3.5313E+01	3.0243E+01	2.1884E+01
3.8000E+00	4.8237E+01	3.3264E+01	8.7981E+00	2.1045E+01	7.2921E+01	5.8287E+01	3.6030E+01	3.1000E+01	2.2423E+01
4.0000E+00	4.9028E+01	3.3838E+01	8.9117E+00	2.1335E+01	7.3733E+01	5.9219E+01	3.6726E+01	3.1742E+01	2.2948E+01
4.2000E+00	4.9796E+01	3.4398E+01	9.0197E+00	2.1613E+01	7.4511E+01	6.0120E+01	3.7402E+01	3.2468E+01	2.3460E+01
4.4000E+00	5.0541E+01	3.4944E+01	9.1229E+00	2.1881E+01	7.5257E+01	6.0992E+01	3.8063E+01	3.3181E+01	2.3960E+01
4.6000E+00	5.1264E+01	3.5475E+01	9.2217E+00	2.2139E+01	7.5972E+01	6.1836E+01	3.8707E+01	3.3880E+01	2.4450E+01
4.8000E+00	5.1966E+01	3.5993E+01	9.3165E+00	2.2388E+01	7.6657E+01	6.2654E+01	3.9339E+01	3.4566E+01	2.4931E+01
5.0000E+00	5.2647E+01	3.6497E+01	9.4076E+00	2.2629E+01	7.7313E+01	6.3445E+01	3.9957E+01	3.5240E+01	2.5403E+01
5.2000E+00	5.3308E+01	3.6988E+01	9.4953E+00	2.2862E+01	7.7942E+01	6.4211E+01	4.0562E+01	3.5901E+01	2.5868E+01
5.4000E+00	5.3949E+01	3.7466E+01	9.5800E+00	2.3088E+01	7.8545E+01	6.4953E+01	4.1156E+01	3.6550E+01	2.6325E+01
5.6000E+00	5.4571E+01	3.7930E+01	9.6614E+00	2.3307E+01	7.9123E+01	6.5670E+01	4.1738E+01	3.7187E+01	2.6774E+01
5.8000E+00	5.5174E+01	3.8383E+01	9.7401E+00	2.3519E+01	7.9676E+01	6.6365E+01	4.2308E+01	3.7812E+01	2.7217E+01
6.0000E+00	5.5759E+01	3.8823E+01	9.8162E+00	2.3725E+01	8.0207E+01	6.7038E+01	4.2869E+01	3.8426E+01	2.7654E+01
6.2000E+00	5.6326E+01	3.9250E+01	9.8896E+00	2.3925E+01	8.0715E+01	6.7688E+01	4.3418E+01	3.9028E+01	2.8084E+01
6.4000E+00	5.6876E+01	3.9666E+01	9.9606E+00	2.4119E+01	8.1202E+01	6.8318E+01	4.3958E+01	3.9620E+01	2.8508E+01
6.6000E+00	5.7410E+01	4.0071E+01	1.0029E+01	2.4307E+01	8.1669E+01	6.8928E+01	4.4487E+01	4.0201E+01	2.8926E+01
6.8000E+00	5.7927E+01	4.0465E+01	1.0096E+01	2.4490E+01	8.2116E+01	6.9518E+01	4.5007E+01	4.0771E+01	2.9338E+01

References

Allen, D. J., Darke, P. L. and Benkovic, S. J. (1989). Fluorescent oligonucleotides and deoxynucleotide triphosphates - preparation and their Interaction with the large (Klenow) fragment of *Escherichia coli* DNA-polymerase-I. *Biochemistry 28*: 4601-4607.

Aquilina, G., Biondo, R., Dogliotti, E., Meuth, M. and Bignami, M. (1992). Expression of the endogenous O^6-methylguanine-DNA-methyltransferase protects Chinese Hamster Ovary cells from spontaneous G:C to A:T transitions. *Cancer Res.* **52**: 6371-6475.

Archer, M. C. (1989). Mechanisms of action of *N*-Nitroso compounds. *Cancer Surveys* **8**: 229-239.

Balmain, A. and Brown, K. (1988). Oncogene activation in chemical carcinogenesis. *Adv. Cancer Res.* **51**: 147-182.

Barshop, B. A., Wrenn, R. F., Frieden, C. (1983). Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM-a flexible, portable system. *Anal. Biochem.* **130**: 134-145.

Bartsch, H., Ohshima, H., Nair, J., Pignatelli, B. and Calmels, S. (1986). Modifiers of endogenous nitrosamine synthesis and metabolism. *Basic Life Sci.* **39**: 87-101.

Bartsch, H., Oshima, H., Pignatelli, B., and Calmels, S. (1989). Human exposure to endogenous *N*-nitroso compounds: quantitative estimates in subjects at high risk for cancer of the oral cavity, oesophagus, stomach and urinary bladder. *Cancer Surveys*, **8** (2): 335-362.

Basu, S., Basu, A. and Modak, M. J. (1988). Pyridoxal 5'-phosphate mediated inactivation of *Escherichia coli* DNA polymerase I: identification of lysine-635 as an essential residue for the processive mode of DNA synthesis. *Biochemistry* **27**: 6710-6.

Beese, L. S. and Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**: 25-33.

Beese, L. S., Derbyshire, V. and Steitz, T. A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**: 352-355.

Belinsky, S. A., Devereux, T. R., Maronpot, R. R., Stoner, G. D. and Anderson, M. W. (1989). Relationship between the formation of promutagenic adducts and the activation of the K-*ras* protooncogene in lung tumors from A/J mice treated with nitrosamines. *Cancer Res.* **49**: 5305-11.

Benkovic, S. J. & Schray, K. J. (1971) Chemical basis of biological phosphoryl transfer, in *The Enzymes* (P. D. Boyer, Ed.) Academic Press, pp. 201-238.

Benzer, S. and Freese, E. (1958). Induction of specific mutations with 5-bromouracil. *Proc. Natl. Acad. Sci. U.S.A.* **44**: 112-119.

Bhanot, O. S., Grevatt, P. C., Donahue, J. M., Cabrielides, C. N. and Soloman, J. J. (1992). *In vitro* DNA replication implicates O^2-ethyldeoxythymidine in transversion mutagenesis by ethylating agents. *Nuc. Acids Res.* **20**: 587-594.

Boosalis, M. S., Petruska, J. and Goodman, M. F. (1987). DNA polymerase insertion fidelity. *J. Biol. Chem.* **262**: 14689-14696.

Borowy-Borowski, B. H. and Chambers, R. W. (1987). A study of side reactions occurring during synthesis of oligodeoxynucleotides containing O^6-alkyldeoxyguanosine residues at preselected sites. *Biochemistry* **26**: 2465-2471.

Bos, J. L. (1989). *Ras* oncogenes in human cancer: a review. *Cancer Res.* **49**: 4682-4689.

Branch, P., Aquilina, G., Bignami, M. and Karran, P. (1993). Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage. *Nature* **362**: 652-654.

Brennan, R. G., Pyzalska, D., Blonski, W. J., Hruska, F. E. and Sundaralingam, M. (1986). Crystal structure of the promutagen O^4-methylthymidine: importance of the *anti* conformation of the O^4-methoxy group and possible mispairing of O^4-methylthymidine with guanine. *Biochemistry* **25**: 1181-1185.

Brookes, P. and Lawley, P. D. (1964). Evidence for the binding of polynuclear aromatic hydrocarbons to the nucleic acids of mouse skin: relation between carcinogenic power of hydrocarbons and their binding to deoxyribonucleic acid. *Nature* **202**: 781-784.

Brown, T., Kneale, G., Hunter, W. N. and Kennard, O. (1986). Structural characterization of the bromouracil:guanine base pair mismatch in a Z-DNA fragment. *Nucleic Acids Res.* **14**: 1807-1809.

Bryant, F. R., Johnson, K. A., and Benkovic, S. J. (1983). Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry* **22**: 3537-3546.

Carroll, S. S., Cowart, M. and Benkovic, S. J. (1991). A mutant of DNA polymerase I (Klenow Fragment) with reduced fidelity. *Biochemistry* **30**: 804-813.

Catalano, C. E., Allen, D. J. and Benkovic, S. J. (1990). Interaction of *Escherichia coli* DNA polymerase I with azido-DNA and fluorescent DNA probes: identification of protein-DNA contacts. *Biochemistry* **29**: 3612-3621.

Chance, E. M. and Curtis, A. R. (1970). Fast numerical simulation of biochemical systems. *FEBS Letters* 7: 47-50.

Chance, E. M., Curtis, A. R., Jones, I. P. & Kirby, C. R. (1977). FACSIMILE: a computer program for flow and chemistry simulation, and general initial value problems, in *Report A.E.R.E. R 8775*, Atomic Energy Research Establishment, Harwell, Berkshire, England.

Coulondre, C. and Miller, J. H. (1977). Genetic studies of the lac repressor IV. Mutagenic specificity in the *lac* I gene of *Escherichia coli*. *J. Mol. Biol.* **117**: 577-606.

Cruzeiro-Hansson, L. Swann, P. F., Pearl, L. and Goodfellow, J. M. (1992). Molecular dynamics of alkylated DNA. *Carcinogenesis* **13**: 2067-2073.

Curtis, A. R. and Chance, E. M. (1972) Numerical methods for simulation and optimization, in *Eighth FEBS Meeting: Analysis and Simulation of Biochemical Systems*, (Hemker, H. C. & Hess, B. Eds.), North-Holland/American Elsevier, Amsterdam, pp 39-57.

Curtis, A. R. & Sweetenham, W. P. (1988) FACSIMILE/CHEKMAT User's Manual, in *Report A.E.R.E. R 12805*, United Kingdom Atomic Energy Authority, Harwell, Berkshire, England.

Dahlberg, M. E. and Benkovic, S. J. (1991). Kinetic mechanism of DNA polymerase I (klenow fragment): identification of a second conformational change and evaluation of the internal equilibrium constant. *Biochemistry* **30**: 4835-4843.

Davies, J. S. (1992). Practical aspects of kinetic analysis. *Methods on Enzymology* **210**: 374-390.

Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M. and Steitz, T. A. (1988). Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240**: 199-201.

Donlin, M. J., Patel, S. S. and Johnson, K. A. (1991). Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* **30**: 538-546.

Dosanjh, M. K., Galeros, G., Goodman, M. F. and Singer, B. (1991). Kinetics of extension of O^6-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. *Biochemistry* **30**: 11595-11599.

Drake, J. W. (1969). Comparative rates of spontaneous mutation. *Nature* **221**: 1132-1133.

Driggers, P. H. and Beattie, K. L. (1988). Effect of pH on the base-mispairing properties of 5-bromouracil during DNA synthesis. *Biochemistry* **27**: 1729-1735.

Druckrey, H., Preussmann, R., Ivankovic, S. and Schmähl, D. (1967). Organotropic carcinogenic effects of 65 different *N*-nitroso compounds in BD rats. *Z. Krebsforsch* **69**: 103-201.

Druckrey, H. and Steinhoff, D. (1962). Erzeugung von leberkrebs an meerschweinchen. *Naturwissenschaften* **49**: 497-498.

Echols, H. and Goodman, M. F. (1991). Fidelity mechanisms in DNA replication. *Ann. Rev. Biochem.* **60**: 477-511.

Eger, B. T. and Benkovic, S. J. (1992). Minimal kinetic mechanism for misincorporation by DNA polymerase I (klenow fragment). *Biochemistry* **31**: 9227-9236.

El-Deiry, W. S., So, A. G. and Downey, K. M. (1988). Mechanisms of error discrimination by *Escherichia coli* DNA polymerase I. *Biochemistry* **27**: 546-553.

Eritja, R., Horowitz, D. M., Walker, P. A., Zichler-Martin, J. P., Boosalis, M. S., Goodman, M. F., Itakuru, K. and Kaplan, B. E. (1986). Synthesis and properties of oligonucleotides containing 2'-deoxynebularine and 2'-deoxyxanthosine. *Nucleic Acids Res.* **14**: 8135-8153.

Fersht, A. (1985). *Enzyme structure and mechanism.* W. H. Freeman and Co.

Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. and Steitz, T. A. (1988). Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 8924-8.

Freese, E. (1959). The difference between spontaneous and base-analogue induced mutations of phage T4. *Proc. Natl. Acad. Sci. USA.* **45**: 622-633.

Gaffney, B. L. and Jones, R. A. (1989). Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O^6-methylguanine with reference to both Watson-Crick pairs and to mismatched pairs. *Biochemistry 28*: 5881-5889.

Gaffney, B. L., Marky, L. A. and Jones, R. A. (1984). Synthesis and characterization of a set of four dodecadeoxyribonucleoside undecaphosphates containing O^6-methylguanine opposite adenine, cytosine, guanine, and thymine. *Biochemistry* **23**: 5686-5691.

Georgiadis, P., Smith, C. A. and Swann, P. F. (1991). Nitrosamine-induced cancer: selective repair and conformational differences between O^6-methylguanine residues in different positions in and around codon 12 of rat H-*ras*. *Cancer Res.* **51**: 5843-5850.

Gerchman, L. L. and Ludlum, D. (1973). The properties of O^6-methylguanine in templates for RNA polymerase. *Biochim. Biophys. Acta* **308**: 310-316.

Ginell, S. L., Kuzmich, S., Jones, R. A. and Berman, H. M. (1990). Crystal and molecular structure of a DNA duplex containing the carcinogenic lesion O^6-methylguanine. *Biochemistry* **29**: 10461-10465.

Goodman, M. F., Hopkins, R. L., Watanabe, S. M., Clayton, L. K. and Guidotti, S. (1983). On the molecular basis of mutagenesis: enzymological and genetic studies with the bacteriophage T4 system, in *ICN-UCLA Symposia on Molecular and Cellular Biology, vol XIX: Mechanistic Studies of DNA Replication and Genetic Recombination* (Alberts, B. Eds., Fox, C. F. Series Ed.) pp 685- 705, Academic Press.

Gorenstein, D. G. (1992). ^{31}P NMR of DNA. *Methods in Enzymology* **211**: 254-286.

Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V. and Jones, C. R. (1988). Assignments of ^{31}P NMR resonances in oligodeoxynucleotides: origin of sequence-specific variations in the deoxyribose phosphate backbone conformation and the ^{31}P chemical shifts of double-helical nucleic acids. *Biochemistry* **27**: 7223-7237.

Goswami, B., Gaffney, B. L. and Jones, R. A. (1993). Nitrogen-15-labeled oligodeoxynucleotides. 5. Use of ^{15}N to probe H-bonding in an O^6-MeG.T base-pair. *J. Am. Chem. Soc.* **115**: 3832-3833.

Goth, R. and Rajewsky, M. F. (1974a). Persistence of O^6-ethylguanine in rat-brain DNA: correlation with nervous system-specific carcinogenesis by ethylnitrosourea. *Proc. Natl. Acad. Sci. U.S.A.* **71**: 639-643.

Goth, R. and Rajewsky, M. F. (1974b). Molecular and cellular mechanisms associated with pulse-carcinogenesis in the rat nervous system by ethylnitrosourea: ethylation of nucleic acids and elimination rates of ethylated bases from the DNA of different tissues. *Z. Krebsforsch* **82**: 37-64.

Graves, R. J., Li, B. F. and Swann, P. F. (1989). Repair of O^6-methylguanine, O^6-ethylguanine, O^6-isopropylguanine and O^4-methylthymine in synthetic oligodeoxynucleotides by *Escherichia coli* ada gene O^6-alkylguanine-DNA-alkyltransferase. *Carcinogenesis* **10**: 661-666.

Grevatt, P. C., Soloman, J. J. and Bhanot, O. S. (1992). *In vitro* mispairing specificity of O^2-ethylthymidine. *Biochemistry* **31**: 4181-4188.

Guest, C. R., Hochstrasser, R. A., Dupuy, C. G., Allen, D. J., Benkovic, S. J. and Millar, D. P. (1991). Interaction of DNA with the Klenow fragment of DNA polymerase I studied by time-resolved fluorescence spectroscopy. *Biochemistry* **30**: 8759-8770.

Heath, D. F. (1962). The decomposition and toxicity of dialkylnitrosamines in rats. *Biochem. J.* **85**: 72-91.

Hecht, S. S. and Hoffmann, D. (1989). The relavance of tobacco-specific nitrosamines to human cancer. *Cancer Surveys* **8** (2): 274-294.

Herschlag, D., Piccirilli, J. A. and Cech, T. R. (1991). Ribozyme-catalyzed and non-enzymatic reactions of phosphate diesters: rate effects upon substitution of sulfur for a nonbridging phosphoryl oxygen atom. *Biochemistry* **30**: 4833- 4854.

Johnson, K. A. (1986). Rapid kinetic analysis of mechanochemical adenosinetriphosphatases. *Methods in Enzymology* **134**: 677-705.

Joyce, C. M. (1989). How DNA travels between the separate polymerase and 3'→5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **264**: 10858-10866.

Joyce, C. M. and Steitz, T. A. (1987). DNA polymerase I: from crystal structure to function via genetics. *Trends Biochem. Sci.* **12**: 288-292.

Joyce, C. M., Friedman, J. M., Beese, L., Freemont, P. S. and Steitz, T. A. (1988). Structural model for the editing decision of *Escherichia coli* DNA polymerase I, in *DNA Replication and Mutagenesis* (Moses, R. E. and Summers, W. C. Eds.) Washington D. C., American Society for Microbiology.

Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F. and Patel, D. J. (1988a). Base pair mismatches and carcinogen-modified bases in DNA: an NMR study of A.C and A.O^4meT pairing in dodecanucleotide duplexes. *Biochemistry* **27**: 100-108.

Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F. and Patel, D. J. (1988b). Base pair mismatches and carcinogen-modified bases in DNA: an NMR study of G.T and G.O^4meT pairing in dodecanucleotide duplexes. *Biochemistry* **27**: 108-115.

Kalnik, M. W., Li, B. F. L., Swann, P. F. and Patel, D. J. (1989a). O^6-ethylguanine carcinogenic lesions in DNA: an NMR study of O^6-etG.C pairing in dodecanucleotide duplexes. *Biochemistry* **28**: 6182-6192.

Kalnik, M. W., Li, B. F. L., Swann, P. F. and Patel, D. J. (1989b). O^6-ethylguanine carcinogenic lesions in DNA: an NMR study of O^6-etG.T pairing in dodecanucleotide duplexes. *Biochemistry* **28**: 6170-6181.

Karran, P. and Bignami, M. (1992). Self-destruction and tolerance in resistance of mammalian cells to alkylation damage. *Nucleic Acids Res.* **20**: 2933-2940.

Katritzky, A. R. and Waring, A. J. (1962). Tautomeric azines. Part I. The tautomerism of 1-methyluracil and 5-bromo-1-methyluracil. *J. Chem. Soc.*: 1540-1544.

Kornberg, A. and Baker, T. A. (1992). *DNA Replication*. New York, W. H. Freeman & Co.

Krieg, D. R. (1963a). Ethyl methanesulfonate-induced reversion of bacteriophage T4*rII* mutants. *Genetics* **48**: 561-580.

Krieg, D. R. (1963b). Specificity of chemical mutagenesis. *Prog. Nucleic Acid Res. Mol. Biol.* **2**: 125-168.

Kuchta, R. D., Mizrahi, V., Benkovic, P. A., Johnson, K. A. and Benkovic, S. J. (1987). Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry* **26**: 8410-8417.

Kuchta, R. D., Benkovic, P. and Benkovic, S. J. (1988a). Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry* **27** : 6716-6725.

Kumar, R., Sukumar, S. and Barbacid, M. (1990). Activation of *ras* oncogenes preceding the onset of neoplasia. *Science* **248**: 1101-1104.

Laval, J., Boiteux, S. and O'Connor, T. R. (1990). Physiological properties and repair of apurinic/ apyrimidinic sites and imidazole ring-opened guanines in DNA. *Mutation Res.* **233**: 73-79.

Lawley, P. D. and Brookes, P. (1962). Ionization of DNA bases or base analogues as a possible explanation of mutagenesis, with special reference to 5-bromodeoxyuridine. *J. Mol. Biol.* **4**: 216-219.

Lawley, P. D. and Thatcher, C. J. (1970). Methylation of deoxyribunucleic acid in cultured mammalian cells by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. *Biochem. J.* **116**: 693-707.

Lawley, P. D., Orr, D. J., Shah, S. A., Farmer, P. B. and Jarman, M. (1973). Reaction products from *N*-methyl-*N*-nitrosourea and deoxyribonucleic acid containing thymine residues: synthesis and identification of a new methylation product O^4-methylthymine. *Biochem. J.* **135**: 193-201.

Lawley, P. D. and Wallick, C. A. (1957). The action of alkylating agents on deoxyribonucleic acid and guanylic acid. *Chem. Ind.* 633.

Leaf, C. D., Wishnok, J. S. and Tannenbaum, R. S. (1989). Mechanisms of endogenous nitrosation. *Cancer Surveys* **8**: 323-334.

Leonard, G. A., Thomson, J., Wayson, W. P. and Brown, T. (1990). High-resolution structure of a mutagenic lesion in DNA. *Proc. Natl. Acad. Sci. USA* **87**: 9573-9576.

Li, B. F. L., Swann, P. F., Kalnik, M. W., Kouchakdjian, M. and Patel, D. J. (1988). NMR solution studies of covalent carcinogenic lesions in DNA: O^6-methylguanosine,O^6-ethylguanosine, and O^4-methylthymidine, in *NMR Spectroscopy in Drug Research* (Jaroszewski, J. W., Schaumberg, K., and Kofod, H. , Eds.) Copenhagen, Munksgaard. pp 309-340.

Li, B. F. L. and Swann, P. F. (1989). Synthesis and characterisation of oligodeoxynucleotides containing O^6-methyl-, O^6-ethyl- and O^6-isopropylguanine. *Biochemistry* **28**: 5779-5786.

Loeb, L. A. and Reyland, M. E. (1987). Fidelity of DNA synthesis, in *Nucleic Acids and Molecular Biology* (F. Eckstein and D. M. J. Lilley, Eds.) Springer, Berlin, pp. 157-173.

Loechler, E. L., Green, C. L. and Essigmann, J. M. (1984). *In vivo* mutagenesis by O^6-methylguanine built into a unique site in a viral genome. *Proc. Natl. Acad. Sci. U.S.A.* **81**: 6271- 6275.

Loechler, E. L. (1991). Rotation about the C6-O^6 bond in O^6-methylguanosine: the *syn* and *anti* conformers can be of similar energies in duplex DNA as estimated by molecular modelling techniques. *Carcinogenesis* **12**:1693-1699.

Loveless, A. (1958). Increased rate of plaque-type and host-range mutation following treatment of bacteriophage *in vitro* with ethyl methanesulphonate. *Nature* (London) **181**: 1212-1213.

Loveless, A. (1959). The influence of radiomimetic substances on deoxyribonucleic acid synthesis and function studied in *Escherichia coli*/phage systems. III. mutation of T2 bacteriophage as a consequence of alkylation *in vitro*: the uniqueness of ethylation. *Proc. Roy. Soc. B.* **150:** 497- 508.

Loveless, A. (1969). Possible relevance of O^6-alkylation of deoxyguanosine to the mutagenecity and carcinogenicity of nitrosamines and nitrosamides. *Nature* **223**: 206-207.

Magee, P. N. and Barnes, J. M. (1956). The production of malignant primary hepatic tumours in the rat by feeding dimethylnitrosamine. *Br. J. Cancer* **10**: 114-122.

Magee, P. N. and Farber, E. (1962). Toxic liver injury and carcinogenesis: methylation of rat liver nucleic acids by dimethylnitrosamine. *Biochem. J.* **83**: 114-124.

Magee, P. N. and Vandekar, M. (1958). Toxic liver injury: the metabolism of dimethylnitrosamine *in vitro*. *Biochem. J.* **70**: 600-605.

Marletta, M. A. (1989). Nitric oxide: biosynthesis and biological significance. *Trends in Biochem. Sci.* **14**: 488-492.

McHenry, C. S. (1988). DNA polymerase III holoenzyme of *Escherichia coli*. *Ann. Rev. Biochem.* **57**: 519-550.

Mhaskar, D. N. and Goodman, M. F. (1984). On the molecular basis of transition mutations: frequency of forming 2-aminopurine.cytosine base mispairs in the G.C → A.T mutational pathway by T4 DNA polymerase *in vitro*. *J. Biol. Chem.* **259**: 11713-11717.

Miller, J. A. (1970). Carcinogenesis by chemicals: an overview--G. H. A. Clowes memorial Lecture. *Cancer Res.* **30**: 559-576.

Miyamoto, S., Sukumar, S., Guzman, C., Osborn, C. R. and Nandi, S. (1990). Transforming c-Ki-*ras* mutation is a preneoplastic event in mouse mammary carcinogenesis induced *in vitro* by *N*- methyl-*N*-nitrosourea. *Mol. Cell. Biol.* **10**: 1593-1599.

Mizrahi, V., Henrie, R. N., Marlier, J. F., Johnson, K. A. and Benkovic, S. J. (1985). Rate-limiting steps in the DNA polymerase I reaction pathway. *Biochemistry* **24**: 4010-4018.

Nagata, C., Takeda, E. and Aida, M. (1982). Why O^6-alkylguanine is especially promutagenic: Ab initio molecular orbital considerations. *Mutation Res.* **105**: 1-8.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. and Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**: 762-766.

Pandey and Modak (1987). *Biochemistry* **26**: 7744-7748.

Parthasarathay, R. and Fridey, S. M. (1986). Conformation of O^6-alkylguanosines: molecular mechanism of mutagenesis. *Carcinogenesis* 7: 221-227.

Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L. and Jones, R. A. (1986a). Structural studies of the O^6-meG.C interaction in the d(C-G-C-G-A-A-T-T-C-O^6meG-C-G) duplex. *Biochemistry* **25**: 1027-1036.

Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L. and Jones, R. A. (1986b). Structural studies of the O^6-meG.T interaction in the d(C-G-T-G-A-A-T-T-C-O^6meG-C-G) duplex. *Biochemistry* **25**: 1036-1042.

Patel, S. S., Wong, I. and Johnson, K. A. (1991). Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**: 511-525.

Pedersen, L. G., Darden, T. A., Deerfield, D. W., II, Anderson, M. W. and Hoel, D. G. (1988). A theoretical study of the effect of methylation or ethylation at O^6-guanine in the structure and energy of DNA double strands. *Carcinogenesis* **9**: 1553-62.

Pegg, A. E. and Byers, T. L. (1992). Repair of DNA containing O^6-alkylguanine. *FASEB J.* **6**: 2302-2310.

Pegg, A. E. and Swann, P. F. (1979). Metabolism of O^6-alkyldeoxyguanosines and their effect on the removal of O^6-methylguanine from rat liver DNA. *Biochim. Biophys. Acta.* **565**: 241-252.

Perrino, F. W. and Loeb, L. A. (1989). Proofreading by the ε subunit of *Escherichia coli* DNA polymerase III increases the fidelity of calf thymus DNA polymerase α. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 3085-3088.

Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C. and Tinoco, I. J. (1988). Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 6252-6256.

Pohorille, A. and Loew, G. H. (1987). Base-pairing properties of *O*-methylated bases of nucleic acids: energetics and steric considerations. *Biophys. Chem.* **22**: 37-51.

Polesky, A. H., Steitz, T. A., Grindley, N. D. F. and Joyce, C. M. (1990). Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **265**: 14579-14591.

Preston, B. D., Zakour, R. A., Singer, B. and Loeb, L. A. (1988). Fidelity of base selection by DNA polymerases, in *DNA Replication and Mutagenesis* (Moses, R. E. and Summers, W. C. Eds.) American Society for Microbiology, Washington, D.C.

Psoda, A., Kierdaszuk, B., Pohorille, A., Geller, M., Kusmierek, J. T. and Shugar, D. (1981). Interaction of the mutagenic base analogs O^6-methylguanine and N^4-hydroxycytosine with potentially complementary bases. *Int. J. Quantum Chem.* **20**: 543-554.

Radman, M. and Wagner, R. (1988). The high fidelity of DNA duplication. *Sci. Am.* **259**: 24-30.

Reddy, E. P., Reynolds, R. K., Santos, E. and Barbacid, M. (1982). A point mutation is resposible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300**: 149-152.

Rush, J. and Konigsberg, W. H. (1990). Photoaffinity labeling of the Klenow fragment with 8-azido-dATP. *J. Biol. Chem.* **265**: 4821-4827.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*. New York, Springer-Verlag.

Saffhill, R., Margison, G. P. and O' Connor, P. J. (1985). Mechanisms of carcinogenesis induced by alkylating agents. *Biochim. Biophys. Acta* **823**: 111-145.

Sibghat-Ullah and Day, R. S., III. (1992). Incision at O^6-methylguanine:thymine mispairs in DNA by extracts of human cells. *Biochemistry* **31**: 7998-8008.

Sibghat-Ullah and Day, R. S., III. (1993). DNA-substrate sequence specificity of human G:T mismatch repair activity. *Nucleic Acids Res.* **21**: 1281-1287.

Singer, B. (1979). *N*-nitroso alkylating agents: formation and persistence of alkyl derivitives in mammalian nucleic acids as contributing factors in carcinogenesis. *JNCI* **62**: 1329- 1339.

Singer, B. and Grunberger, D. (1983). *Molecular Biology of Mutagens and Carcinogens*, Plenum, New York, pp347.

Singer, B., Chavez, F., Goodman, M. F., Essigmann, J. M. and Dosanjh, M. K. (1989). Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O^6-methylguanine in a defined primed oligonucleotide when *Escherichia coli* DNA polymerase I is used. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 8271-8274.

Smith, C. A., Xu, Y.-Z. and Swann, P. F. (1990). Solid-phase synthesis of oligodeoxynucleotides containing O^6-alkylguanine. *Carcinogenesis* **11**: 811-816.

Smith, K. (1992). Non-linear regression: cut errors by making the curve fit the data. *Laboratory Equipment Digest* **February Issue**: 31- 32.

Snow, E. T., Foote, R. S. and Mitra, S. (1984). Base-pairing properties of O^6-methylguanine in template DNA during *in vitro* DNA replication. *J. Biol. Chem.* **259**: 8095-8100.

Sriram, M., van der Marel, G. A., Roelen, H. L. P. F., van Boom, J. H. and Wang, A. H.-J. (1992). Conformation of B-DNA containing O^6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[6eG]AATTCGCG) complexed with Hoechst 33258 or Hoechst 33342. *EMBO J.* **11**: 225-232.

Sternglatz, H. and Bugg, C. E. (1975). *Biochim. Biophys. Acta* **378**: 1-11.

Sukumar, S. (1990). An experimental analysis of cancer: role of *ras* oncogenes in multistep carcinogenesis. *Cancer Cells* **2**: 199-204.

Swann, P. F. (1990). Why do O^6-alkylguanine and O^4-alkylthymine miscode? The relationship between the structure of DNA containing O^6-alkylguanine and O^4-alkylthymine and the mutagenic properties of these bases. *Mutation Res.* **233**: 81-94.

Swann, P. F. and Magee, P. N. (1968). Nitrosamine induced carcinogensis. The alkylation of nucleic acids of the rat by *N*-methyl-*N*-nitrosourea, dimethylnitrosamine, dimethyl sulphate and methyl methanesulphonate. *Biochem. J.* 110: 39-47.

Swann, P. F. and Magee, P. N. (1971). Nitrosamine induced carcinogensis. The alkylation of N-7 of guanine of nucleic acids of the rat by diethylnitrosamine, *N*-ethyl-*N*-nitrosourea and ethyl methanesulphonate. *Biochem. J.* **110**: 39-47.

Tabin, C. J., Bradley, S. M., Bargmann, C. I., Weinberg, R. A., Papageorge, A. G., Scolnick, E. M., Dhar, R., R., L. D. and Chang, E. S. (1982). Mechanism of activation of a human oncogene. *Nature* **300**: 143-149.

Topal, M. D. (1988). DNA repair, oncogenes and carcinogenesis. *Carcinogenesis* **9**: 691-696.