THEORETICAL INVESTIGATIONS ON THE PROPERTIES OF MODIFIED DNA

NORTH EASTERN HILL UNIVERSITY LIBRARY
No............
Date..........

BY

Divi Venkateswarlu

DEPARTMENT OF CHEMISTRY

SCHOOL OF PHYSICAL SCIENCES

A THESIS

SUBMITTED IN FULFILMENT OF THE REQUIREMENT

FOR THE DEGREE OF

Doctor of Philosophy

NEHU LIBRARY
Acc. No........ 103235
Acc. by ....
Date ....
Class by..............................
Sub Heading by ....
Enter by..............................
Transcribed by......................

To

NORTH-EASTERN HILL UNIVERSITY

SHILLONG

INDIA

1995

DS
574.9732
VEN

In memory of

my **Parents**

पूर्वोत्तर पर्वतीय विश्वविद्यालय

बिजनी परिसर, शिलांग - 793003 (मेघालय)

North-Eastern Hill University

Bijni Complex, Shillong - 793003 (Meghalaya)

Phone :
Gram : NEHU

Dr. R.H. Duncan Lyngdoh
Department of Chemistry

CERTIFICATE

I hereby certify that the entire work embodied in this thesis has been carried out by Mr. Divi Venkateswarlu under my guidance in the Department of Chemistry, North-Eastern Hill Univerisity, Shillong, and it has not been submitted elsewhere for any degree or diploma.

R. H. Duncan Lyngdoh

Research Supervisor

ACKNOWLEDGEMENTS

It is with great pleasure, I express my heart-felt gratitude to my thesis mentor and Guru, **Dr. R.H.Duncan Lyngdoh**, who introduced me to the field of Quantum biology in particular and computational chemistry, in general. This thesis would have not come out of this shape without his kind help. His great encouragement and untiring support in both academic and personal life, particularly during times when my career and personal life were going through very traumatic transition, cannot be simply forgotten throughout my life.

I wish to express my sincere thanks to:

* University Grants Commission, New Delhi and North-Eastern Hill Univerisity, for providing the necessary financial help through research fellowship

* Dr S N Rai, Director, Computer Centre, NEHU, for providing me the free access to the computer facilities of his centre.

* Professor H Junjappa and Professor (Mrs.) H.Ila, with whom I initially started my research career at NEHU, for their moral encouragement through my course of research work

* Professor Mihir K Choudhury, Head of the Chemistry Department, NEHU, for providing facilities and encouragement

* Mr P P Dey, Ms B Rynjah and other staff of Computer centre, who provided me the constant and untiring help, without whose help I would have not completed my research work in this time.

* All the faculty members of Chemistry Department, NEHU, for their support during my study

* Mr M V Basaveswara Rao, a close companion in my personal life at NEHU, for fraternal association throughout my stay at NEHU

* Director of NEC and Mr Stephen, Mr Sarkar, Mr Dinesh and Mr Rajasekhar, the staff of NIC unit for providing generous computer facilities

* Dr C. Satyajit Rao and other country-cousins, for their personal help and cultural interaction during my stay at Shillong

* Dr. Ch.V.K.MSarma and Dr. A.S.Bhanuprasad, of University of Hyderabad, for providing me the necessary software and other help for my research work

* Professor Donold G. Truhlar, Department of Chemistry, University of Minnesota, Minnesota, USA, for providing a free copy of AMSOL 4.5.1 software

* Ms Jayati and other colleagues of my research group, for thier support in research work

* Mr A Arunachalam and all my friends, who helped me in many ways to accomplish my goal at NEHU smoothly

and

* Ms. Sumana, my beloved friend and life-companian, for her great association

Divi Venkateswarlu

CONTENTS

		Page
CHAPTER I	INTRODUCTION	1
CHAPTER II	APPROACHES TO THE RESEARCH PROBLEM	18
CHAPTER III	PROTIC CHANGES IN NUCLEIC ACID BASES AND NUCLEOSIDES	28
CHAPTER IV	PAIRING PROPERTIES OF MODIFIED DNA BASES	85
CHAPTER V	STABILITY AND REPAIR OF MODIFIED DNA COMPONENTS	137
	REFERENCES	160

CHAPTER I

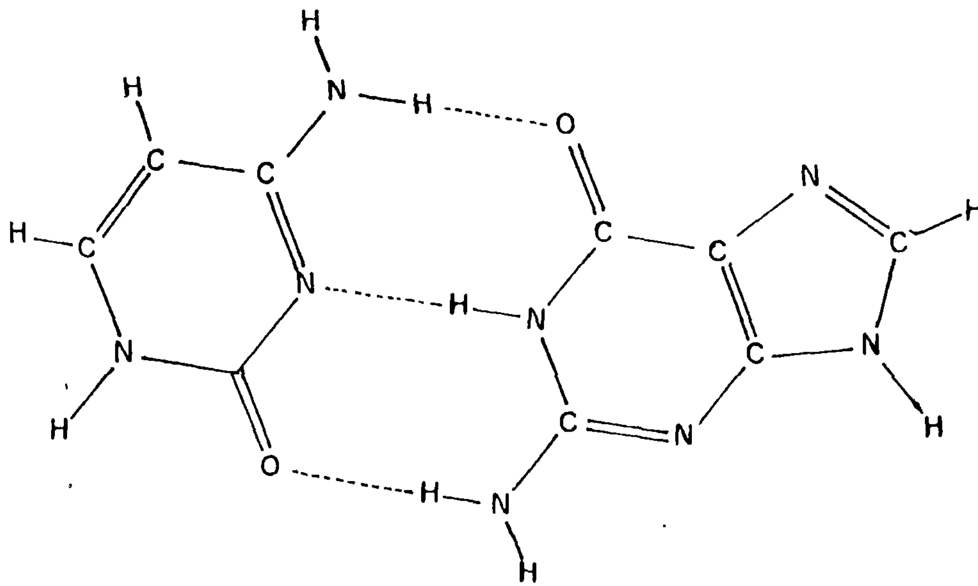
INTRODUCTION

I.1 Opening Remarks

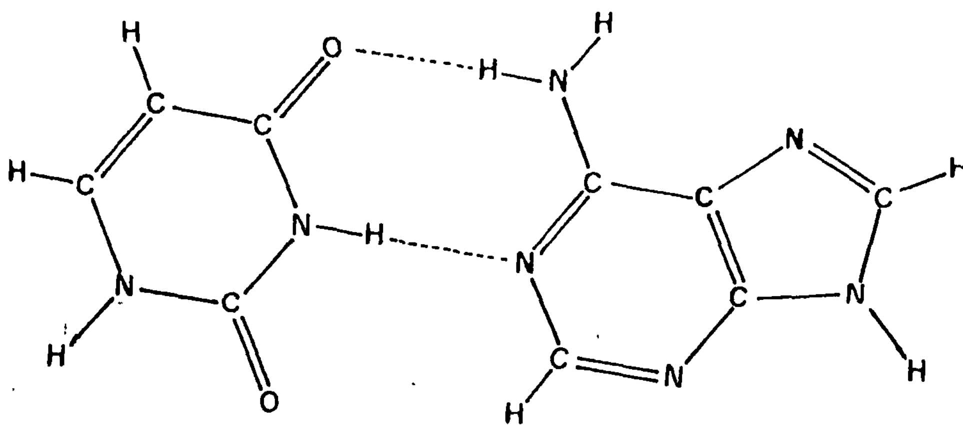
Among the outstanding achievements of the 20th century has been the establishment of the structures of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), and of the primary role of these macromolecules as hereditary determinants in biological reproduction, growth and development. Ever since the seminal contribution to our understanding of DNA structure by Watson and Crick (1953a, 1953b), there has been an exponential growth in the quantum of research study, both experimental and theoretical, which addresses the questions concerning the details of three-dimensional structure, dynamical behaviour and functioning of both DNA and RNA.

The DNA duplex consists of two complementary strands interwound to form a right-handed helix held together by hydrogen-bonding between heterocyclic bases. Each unit of a strand consists of a heterocyclic base (purine or pyrimidine), a deoxyribose sugar moiety and a phosphate group. The common purine bases are guanine and adenine, the pyrimidine bases being cytosine and thymine. Fig. I.1 portrays the two Watson-Crick base-pairs G:C and A:T. The integrity of the genetic code lies to a great extent in the high specificity of hydrogen-bonding linked to the complementary base-pairing of guanine to cytosine and thymine to adenine (Loeb & Kunkel 1982).

During the course of cell division, the double-stranded DNA splits into two single



Guanine: Cytosine pair



Adenine: Thymine pair

Fig. I.1.:Schematic representation of normal Watson-Crick base pairs

strands, each of which acts as a template for a new strand of DNA. As a result, each new cell acquires its own genome in the form of DNA duplexes. This process, known as replication, is not always perfect and can allow for errors at the molecular level leading to mutations that are perhaps essential for human evolution. There may be a competition between the correct GC and AT base pairs and eight possible base mispairs - AA, GG, AG, CC, TT, CT, AC and GT. The first three are purine-purine mismatches, the next three are pyrimidine- pyrimidine mismatches and the last two are purine-pyrimidine mismatches. After a cycle of replication, purine-purine mismatches and pyrimidine-pyrimidine mismatches yield **transversion** mutations, while purine-pyrimidine mismatches produce **transition** mutations. Fig. 1.2 shows the mutagenic pathways for both transversional and transitional mutations. Each single mistake can alter one codon triplet of information and may lead to a change in the amino acid sequence of the proteins synthesised therefrom. As a consequence the protein may be ineffective, its malfunction even resulting in a catastrophic event such as the death of the cell, a genetic desease, or a carcinogenic response.

The reasons for these molecular mistakes are not always very clear even today. Most endogenous mutations are speculated to be due to tautomerism in the DNA bases, or due to spontaneous deglycosylation or hydrolysis leading to abasic site formation, or to deamination or base ionisation or errors in replication which may lead to base mispairing (Lindahl 1993). Exogenous (induced) mutations may be due to environmental chemicals and mutagens, or other types of damages due to ionising radiation or free radicals (Singer & Grunberger 1983). All of these modifications can usually be subject to the efficient DNA repair process by proof-reading and repair enzymes that patrol the new DNA duplex. However, it has been long speculated that in some very specific cases of base mismatches, the repair enzymes fail to recognize the mismatch, which event may lead to permanent and inheritable damage (Modrich 1994; Sancar 1994).

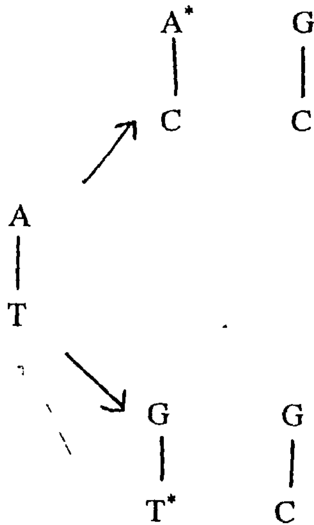
Figure I.2 Mutagenic pathways to : (a) Transitional mutations; and (b) Transversional mutations from AT and GC base pairs

Transition mutations

Tranversion mutations

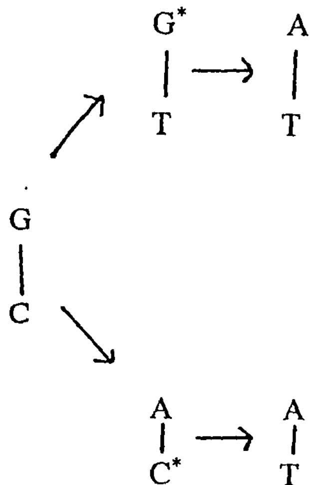
1) A-T type

Basepair Mispair Mutation



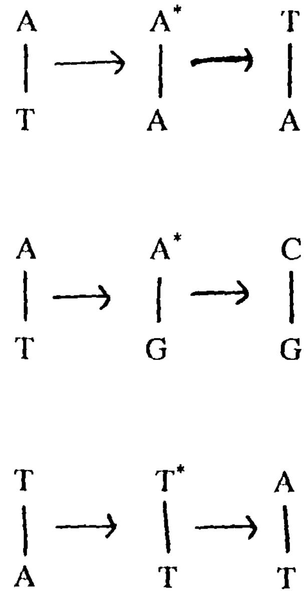
2) G-C type

Basepair Mispair Mutation



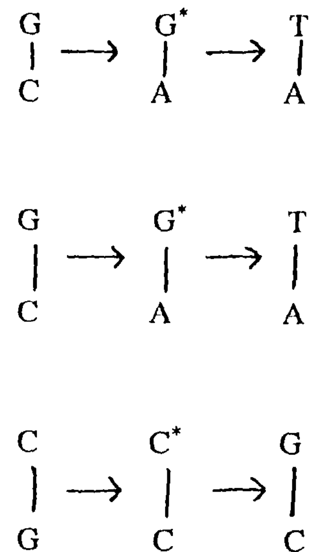
3) A-T type

Basepair Mispair Mutation



4) G-C type

Basepair Mispair Mutation



I.2 DNA-Related Xenobiotic Phenomena

Xenobiotic phenomena are those in which the structure and functioning of the living organism or its components are disturbed or altered by the action of agents foreign to the living system, which may be physical, chemical or biological in nature. In this, they stand in sharp contrast to phenomena of endogenous or spontaneous origin within the living system. Xenobiotic phenomena which involve the genome of the organism include gene mutations, carcinogenesis, teratogenesis and induced cell death, besides others. All of these may be described within the framework of a commonality of molecular basis involving the alteration of the genome in some way or the other.

I.2.1 Mutagenesis

Mutation of the DNA as the critical macromolecular target for the xenobiotic agent is a widely studied biological phenomenon. Every human genetic disease, as defined at the clinical level, is caused by a variety of mutations. Many types of changes in the sequence of DNA nucleotide can affect gene function. These alterations, or DNA mutations, fall under the following classes :

(a) Point mutations - Base substitutions or frameshifts involving just one base;

(b) Gene or chromosomal reallocations; and

(c) Gene or chromosomal amplifications.

Single nucleotide changes can substitute one amino acid for another (called **missense**) or cause premature termination of polypeptide synthesis (called **nonsense**). These may affect the level of transcription of a gene, or they may alter splice sites for RNA processing, often with drastic effects on gene expression. When the nucleotide change results in no amino acid substitution, it is called a silent mutation.

Missense mutations may affect any aspect of a protein's function, including catalytic activity, binding of cofactors, interactions with other macromolecules, and

stability.

However, some single amino acid changes may be truly neutral. A special case of missense mutations are "anti-nonsense" mutations, where the termination codon of an mRNA is converted to a normal amino acid codon. There are more than 100 known single amino acid substitutional muations documentated for beta-globin, and dozens more in alpha-globin. For example, replacement of histidine 58 in alpha-globin causes methemoglobinemia, a drastic reduction in the ability of hemoglobin to bind oxygen.

Nonsense mutations, which normally produce null alleles, may arise directly by nucleotide substitution or indirectly as a result of frameshift. In the later case, deletion or addition of 1 or 2 nucleotides (or a multiple thereof) changes the reading frame, so that the amino acid sequence becomes totally different from the normal sequence downstream of the mutation. Two examples of direct nonsense mutations involving beta-globin occur at codon 17 in Chinese populations and at codon 39 in Mediterranean populations. These lead to beta-zero thalassemia in homozygotes (Orkin 1983).

Deletions and **insertions** are another important source of genetic variation. Both may occur as a result of **unequal crossing-over**, but other mechanisms, not yet clearly understood, are apparently also involved. Small deletions and/or insertions may produce frameshifts. It is apparent that deletions account for at least 50% of Duchenne muscular dystrophy patients. Another known example is steroid sulphatase deficiency, which leads to one form of the disease called ichthyosis (scaly skin); in one study, 14 out of 15 apparently urelated patients had deletions (Bonifas *et al* 1987). Probably the most famous deletions are those that cause the large majority of alpha-thalassemia cases.

Chromosomal mutations represent gross changes that can be detected by microscopic examination of metaphase or late prophase chromosomes. Genomic accidents that alter the number of chromosomes or that produce cytologically detectable variations in one or more chromosomes are called chromosomal mutations. These

include polyploidy (extra sets of chromosomes), aneuploidy (any number of chromosomes other than a multiple of the haploid number), and rearrangements. There are two broad categories of chromosomal rearrangements: deletions and translocations. Deletions can be either "terminal" (referring to genetic material deleted from one end of the chromosome) or "interstitial" (implying loss of an internal section of a chromosome). Balanced translocations arise when there have been breaks in two non-homologous chromosomes, followed by joining of the broken ends in new combinations. Balanced translocations often have no deleterious effects on persons who carry them, but they are responsible for abnormal gamete formation, which may lead to abnormal offspring or multiple spontaneous abortions.

1.2.2 Induction of point mutations by DNA adducts

DNA adducts are the products of chemical interaction between a xenobiotic agent and DNA. They are thought to be the species directly responsible for the mutations induced by mutagens/carcinogens (Osborne 1984). However, little is known about the kind of mutations induced by specific DNA adducts (or "lesions") derived from most mutagens or carcinogens (Basu and Essigmann 1988), and even less about the corresponding mechanisms for induction of mutations (Loechler *et al* 1990). Happily, this is not the case for N-nitroso carcinogens and alkylating carcinogens, for which the molecular basis for DNA modification, mutagenesis and even carcinogenesis is relatively clear. The mechanisms of mutagenesis involving DNA lesions can be classified into three categories:

i) Misinformational Mechanisms

a) Chemical perturbations

Adduct-induced base tautomerisation and Adduct-induced base ionisation

b) Structural perturbations

Adduct-induced base-rotation and Adduct-induced base-wobble

A misinformational mechanism is operative when a DNA polymerase attempts to "read" the base moiety in a DNA lesion but misinterprets it. O^6-Methylguanine (Loechler *et al* 1984; Basu & Essigmann 1988) and O^4-alkylthymine (Preston *et al* 1986), formed from carcinogenic methylating agents, appear to belong in this class, and are frequently read as adenine and cytosine respectively during DNA replication.

ii) Noninformational Mechanisms

A noninformational mechanism is operative when a DNA polymerase chooses to incorporate a particular deoxynucleoside triphosphate (dNTP) opposite a DNA lesion for reasons not involving an attempt to "read" the lesions (Strauss *et al* 1982). Apurinic/apyrimidinicsites (AP-sites), which are chemically induced and formed spontaneously by various means, appear to belong in this class (Loeb & Preston 1986), and it is likely that the adducts derived from bulky mutagens/carcinogens also belong in this class (McCann *et al* 1975; Strauss *et al* 1982).

iii) Other mechanisms

When DNA polymerase is not involved in the fixation of a mutation, the mechanism will be governed by other types falling in this category. Mechanisms involving topoisomerases have been suggested in several cases (Burnouf *et al* 1989; Ripley *et al* 1988).

These modifications of DNA are what go to comprise the subject matter of this thesis, and may be held responsible for a wide variety of xenobiotic phenomena like carcinogenesis, teratogenesis, some cases of cytotoxicity and cell death etc. A few of these are dealt with below. For the most part, this thesis concerns only those modifications of DNA which lead to point mutational changes.

1.2.3 Carcinogenesis

Cancer may be defined as a cellular disorder characterised by the rapid and uncontrolled proliferation of cells which have been genetically altered in a specific manner. This gives rise to the growth and spread (metastasis) of tumours, which in principle may occur at any organ of the body. At the cytological level, apart from the uncontrolled cell division, other key features include the breakdown of normal mechanisms for growth, regulation and differentiation, marked changes in cell shape and motility, and a tendency towards the anaerobic mode of cellular respiration. Cancer, if unchecked, can lead to the death of the organism. Indeed around 20 to 25% of human deaths are attributed to cancer-related causes. A wide variety of xenobiotic agents may be responsible in inducing the process of carcinogenesis. The environmental causes of cancer may be attributed to:

(a) Physical factors, eg. electro magnetic or particulate radiation

(b) Chemical factors

(c) Biological factors, viz. tumour viruses

Upto 90% of human cancers today may be attributed to envir- onmental chemical carcinogens. Epidemiological studies have also earmarked gamma radiation and nuclear fission products as important causes of cancer, eg. among the victims of the atomic bomb holocausts of World War II. Apart from the Epstein-Barr virus (concomitant with endemic malaria among African children), and the human papilloma virus, tumour viruses have not in general been found to be operative in man, although well-known in various animal species ever since the discovery of the chicken sarcoma virus. The breakdown in immunological defences accompanying HIV infection is a secondary factor in the widespread occurrence of sarcomas among AIDS patients.

1.2.4 Teratogenesis

This refers to mutations in the foetus of a vertebrate species generally, resulting in the birth of deformed or malformed offspring. Ontologically, these may owe their origin to mutations in the sex cells (gametes) themselves prior to fertilisation, or to mutations occurring after the zygote or embryo is formed. In the latter case, for the malformation to be marked and well-characterised, the harmful mutations should occur in cells during the early stages of differentiation. Malformations include physical peculiarities like multiplicity or absence of limbs, organs or body parts, as well as inherent abnormalities in metabolism or vital organ functioning. The well-studied thalidomide tragedy is one instance in which intake of a xenobiotic drug administered to pregnant women resulted in a widespread occurence of births of human babies with gross bodily malformations. Teratogens may thus be classified as falling under the general class of mutagens and act through targetting DNA of the appropriate cells at their critical sites.

1.3 Role of Modified DNA for Xenobiotic Phenomena

The role of modified DNA for the underlying phenomenon of induced mutagenesis is implicit from its very definition itself. This role for cancer, teratogenesis and cytological changes is less obvious, and is discussed below for the case of cancer only :

1.3.1 Somatic mutation theory of cancer

This theory, now widely accepted as fact, acknowledges that cancer ultimately owes its origin to the action of carcinogens on the nuclear DNA of somatic cells (Straus 1981; Miller & Miller 1981). A survey of the various families of chemical carcinogens has demonstrated that, to a large extent, there exists a good deal of broad mechanistic commonality in the molecular basis of their carcinogenic action. It is convenient to distinguish two chief phases of the process.

(1) **Phase 1**, where the carcinogen operates to attack and modify the critical targets in cellular macromolecules.

(2) **Phase 2**, where the modified macromolecules operate so as to induce the cellular events associated with cancer induction.

Much of the groundwork concering elucidation of **Phase 1** was laid by Miller and Miller with co-workers, who forwarded the **Ultimate Carcinogen theory** as the possible mode of explanation for the initial formation of the reactive ultimate carcinogen species which then attacks the critical macromolecular targets. This theory holds that most chemical carcinogens are in themselves, or are transformed *in vivo* to, reactive electrophiles by spontaneous decomposition or enzyme-mediated metabolism after administration to the living system. These reactive electro- philes being the actual agents that critically modify the target macromolecules are termed the ultimate carcinogen(s). The chemical as administered to the body is termed as the parent carcinogen, while stable or metastable intermediates formed during the process of transformation *in vivo* are termed as the proximate carcinogens.

Phase 2, which is formulated as the somatic mutation theory of cancer, identifies the nuclear DNA of somatic cells as the critical macromolecular target for the ultimate carcinogen formed in Phase 1. This theory explains the major macromolecular events involved in the process of inducing cancer. The somatic mutation theory of cancer, which was initially put forward in germinal form by Boveri (1914) and later worked upon or refined further by Lowdin (1977), Ts'O (1980) and Skipper (1983), proposes a genetic origin for cancer. The DNA being the prime genetic material, this theory postulates a key role played by the critically altered nuclear DNA of somatic cells for the induction and maintenance of the cancerous state (Strauss 1981; Bishop 1985). DNA is known to be definitely altered and modified by chemical or physical carci- nogens, and when this alteration occurs in the critical manner, cancer may result. This

critical alteration is specifically known to involve certain sections of the genome which are termed as oncogenes. Oncogenes are universally present in the genome of all living species in a latent form known as the proto-oncogene. DNA mutation of the critical kind serves its carcinogenic function by transforming the latent proto-oncogene to its carcinogenically active oncogene counterpart.

The discovery of the existence and functioning of oncogenes, reviewed by Franks and Teich (1986), lends direct confirmation to this concept of somatic mutation, suggesting that modifications only at the critical sites on the proto-oncogene would be of carcinogenic significance (Varmus 1985). Thus, the mutation resulting from DNA modification may be simply a point mutation like base-substitution or frame-shifts, or involve actual breakage or translocation of a chromosome. Gene amplification serves as another mechanism for involvement of DNA in carcinogenesis. Such a transformed cell becomes capable of indefinite cell division independent of normal control mechanisms, thereby leading to tumour growth and development.

1.4 Structural Basis for DNA Modifications

The types of DNA modifications along with their properties which are studied in this thesis include the following :

(1) Prototropic changes in DNA bases : These include tautomerism (proton transfer from one nucleophilic site to another), proton addition to nucleophilic sites, and proton abstraction from nucleophilic sites on DNA bases.

(2) Alkylation of DNA : Induced by N-nitroso compounds and alkylating agents, it is of relevance for point mutagenesis and carcinogenesis, involves a variety of nucleophilic sites attacked, and can be linked to base-mismatches.

(3) C^8-oxidation of purines : The products could be of relevance for purine-purine mismatches as discussed below.

(4) Conformational changes of DNA : These could result in a purine nucleoside turning to the syn rotamer, of possible relevance for mismatching (see below).

(5) Dealkylation of alkylated DNA : The *in vivo* stability and loss of alkyl groups in DNA and their pH dependence is of likely importance for specificity in choice of repair mechanisms.

(6) Deglycosylation : The loss of purine or pyrimidine bases by cleavage of the glycosyl bond creates abasic sites, of significance for induction of noninformational mutations. It is also the mechanistic base for N-glycosylase excision repair.

1.5 Proton Transfer Reactions in DNA

The key role of point mutations (base substitutions, additions or deletions) in the activation process of oncogenes has been well-documented. Since the information in DNA is intimately connected with the pattern of proton distribution along the DNA chain, proton transfer reactions emerge as having a singular import for the disturbance in transmission of this information. Proton transfer can occur via tautomerism, protonation of nucleophilic sites, and deprotonation from a basic site.

1.5.1 Tautomerism of DNA bases

The structures of the commonly conceived tautomers of the 4 DNA bases are such as to facilitiate induction of base substitutions (Pullman & Pullmann 1964). The possible involvement of the minor tautomers in DNA mis-pairing has led to many experimental and theoretical studies on their relative thermodynamic stability. Lowdin (1963; 1977) suggested that unusual base tautomers might arise within the DNA helix through a double-proton transfer between adjacent guanine and cytosine moieties, converting the normal GC pair to another complementary base pair involving an enol tautomer of guanine and an imino tautomer of cytosine, which upon replication leads to mutation.

1.5.2 Protonation of nucleosides

Purine and pyrimidine heterocycles are very reactive and possess many nucleophilic centres. The study of protonation and the preferred site of protonation on the nucleosides and nucleotides is of wider interest for structural chemists as well as to bioorganic chemists as they formulate viable reaction schemes and mechanisms. The existence of a large number of basic sites on base, sugar and phosphate units led to several possibilities of protonation. Even in the relatively simple molecule adenine, there are five possible sites for the two observed ionisations from the protonated species, viz., N^1,N^3,N^7,N^9 and C_6NH_2. For the other nucleic acid bases and their respective nucleosides and nucleotides, the possible sites of protonation or ionisation are also numerous. A considerable amount of work has been done during the past three decades in identifying the sites of proton ionisation from these substances and several are now well established.

1.5.3 Deprotonation of nucleosides

The anionic species generated following deprotonation of nucleic acid bases may posses the non-transitory base-pair potential provided the associated basic pK_a values are amenable at the biological pH. A survey of the literature shows that deproto- nation of normal nucleic acid bases occurs at high basic pH. The knowledge of pK_a values associated with the deprotonation of the bases is of much relevance to mutagenicity as some of these deprotonated bases (viz. the alkylated ones) may possess sufficient stability at biological pH to mispair with non-complementary bases leading to spontaneous mutations.

1.5.4 Pairing properties of protically modified DNA bases

The stability range of protically modified DNA base systems is of intimate relevance to

the biological situation as some of these modified systems might well participate in base-pairing schemes leading to misincorporation of wrong bases during replication. For example, the N^1-protonated adenine tautomer may behave as guanine, mispairing with cytosine. Similarly, N^7-protonated guanine may mispair in the syn conformation with both purines and pyrimidines, leading to a mutagenic situation. In fact, many of these possibilities were studied extensively by NMR and other physical methods to understand the base-pairing properties of the mismatches.

I.4 Alkylation of DNA

The somatic mutation theory of cancer invokes the role of critically altered DNA of somatic cells in the initiation and maintenance of the cancerous state of the cell. It is well-established that alkylation of DNA can lead to proto-oncogene activation via point mutations, which thus demonstrates the link between DNA alkylation and carcinogenesis.

I.4.1 Alkylating agents as DNA modifiers

Alkylating agents encompass a very broad range of compounds, including the direct-acting alkyl esters, ethylenimines, epoxides and alkylnitrosoureas, as well as the indirect-acting nitrosamines and aflatoxins. A total of about 16 nucleophilic sites on DNA have been identified as targets for alkylation by carcinogens and mutagens, many of which predominantly react with DNA at the N^7-guanine position. The first known alkylating agent whose DNA adduct yield correlated with its mutagenicity is nitrogen mustard, extensively used during World War II. Alkylating agents can be broadly classified into mono and bifunctional types.

I.4.2 Features of DNA alkylation

Alkylating agents may form covalent adducts with DNA components either directly by virtue of their own inherent electrophilic reactivity, or indirectly through prior metabolic activation to reactive ultimate carcinogens. Nucleic acids and proteins are alkylated fairly evenly by alkylating agents *in vivo* and *in vitro*, both conditions producing similar trends. Nitrogen, oxygen, sulphur and nucleophilic carbon sites are alkylated in proteins, while the nucleic acid targets include both nitrogen and oxygen nucleophilic centres on the base residues and sugar moiety. Alkylation has been found to occur at the somatic DNA (nuclear), the mitochondrial, DNA, t-RNA and r-RNA, of which only somatic cell nuclear DNA is implicated for the role in carcinogenesis. Both *in vitro* and *in vivo* studies show similar patterns of alkylation at nucleophilic sites.

Direct-acting alkylating agents such as alkyl esters tend to attack DNA by a more or less SN^2 mechanism. The alkanediazonium ion product of N-nitroso compound decomposition or metabolism modifies DNA similarly, although with an appreciable amount of SN^1 character in the transition state, which is linked to O-selectivity of alkylation, itself a factor of relevance for carcinogenesis and mutagenesis.

Sites which have been isolated as alkylation targets on DNA *in vivo* and *in vitro* include the phosphate group oxygens (a major target), the $O^{2'}$- atom on the deoxyribose ring, the N^7-, O^6-, N^3- and C^8-atoms on guanine, the N^7-, N^3-, and N^1-atoms on adenine, the O^4-, O^2- and N^3- atoms of thymine and the O^2- and N^3 atoms of cytosine. One need not pay much attention to the alkylation of ribose-phosphate backbone owing to the evidence that it is only the purine or pyrimidine base sites that may be directly linked with the mechanisms for mutagenesis and carcinogenesis.

I.4.3 Properties of alkylated DNA components

Alkylated DNA components (notably the base residues) may be described with respect

103235

to the following properties dealt with in this thesis : (1) Watson-Crick proton acidity, related to the bestowal of mutagenic character, (2) Stability *in vitro* with regard to half-life and loss, (3) Specificity in choice of repair mechanisms.

While many alkylated DNA nucleosides (like N^7-G alkyl adducts) have no significance for base mispairing or mutagenicity, certain O-alkylated adducts (like the O^6-G alkyl adducts) exhibit a pronounced tendency to mispair with noncomplementary bases, which has been widely implicated in mutagenesis and carcinogenesis. This presence or absence of promutagenicity is here linked to Watson-Crick proton acidity, where the mutagenic adducts lose these protons at biological pH, while the non-mutagenic ones retain them.

Alkylated DNA components display varying stabilities *in vitro* and various mechanisms for loss from the nucleic acid system. In general, the N-alkylated base residues are easily lost through deglycosylation, while O-alkylated residues possess a stabler glycosyl bond. The alkyl groups in the base residues themselves possess varying labilities, where the O-alkyl groups are more labile than the N-alkyl groups.

These varying *in vitro* stabilities and mechanisms for loss of alkylated DNA components bear a convincing correlation to the highly specific choice of repair mechanism *in vivo*. In general, the N-alkylated lesions are repaired by glycosylase enzymes, while the O-alkylated ones are repaired by alkyltransferases.

The effects of DNA alkylation include mutagenesis, carcinogenesis, cell toxicity, teratogenesis and acute toxicity. It is thus seen that DNA alkylation be xenobiotic alkylating agents generally has a deleterious effect on the genome.

I.5 C8-Oxidations of Purines

C^8-Oxopurines, including 8-oxoguanine and 8-oxoadenine, have been much studied in recent years with regard to their formation and significance for mutagenesis. Studies

on the mutagenic consequences of oxidative DNA damage involving the formation of 8-oxopurine derivatives *in vitro* and *in vivo* have been reported (Kuchino *et al* 1987; Fraga *et al* 1990; Shibutani *et al* 1991; Wood *et al* 1992). Recently, the formation of 7,8-dihydro-8- oxoguanosine or deoxyguanosine by hydration of the guanine radical cation within DNA (Kasai *et al* 1992) or by singlet oxygen attack on the corresponding gaunosine has also been reported (Kasai *et al* 1992; Devasagayam *et al* 1991).

The reaction of purine bases of DNA with hydroxyl radicals can result in the formation of 8-oxoadenine and 8-oxoguanine, whose significance for mutagenesis is currently the subject of a great deal of interest. Guanine residues in DNA are hydroxylated by oxygen radicals at the C^8 position to form 7-hydro-8-oxo-2' deoxy-guanosine (Kasai & Nishimura 1984a-c; Dizdaroglu 1985). Introducing an oxygen atom at the C^8 position of dG alters the stereoelectronic properties of dG, forcing it to adopt the syn conformation, thus rendering it capable of pairing with nucleosides other than dC. Studies have indicated that 8-oxo-7H- dG(syn) also forms a stable structure in duplex DNA when paired with dA(anti). The prime mechanism for mutagenesis by these modified purines is purine-purine mismatching leading to base-pair transversions after rotation to the *syn* conformer. It has been noted that while 8-oxoguanine has the potential for base-mismatching, 8-oxoadenine is not so, maintaining a preference for normal non-mutagenic pairing with thymine.

I.9 Conformational Changes in DNA Nucleosides

I.9.1 Conformational changes in O-alkylated nucleosides

Miscoding alkylated nucelosides, like the exocyclic O-alkyated products, have been widely implicated for mutagenesis and carcinogenesis. The exocyclic groups can adopt either a syn or anti confirmation relative to the hydrogen bonding side, of which only the latter is meaningful in the context of mutagenesis. The syn to anti rotational barrier

thus may even block mutagenesis.

1.9.2 Conformational changes in purine nucleosides

The other significant conformationally variant feature of nucleosides is the glycosidic syn to anti rotation in purine nucleosides, which is of significance for mutagenic purine:purine mismatches. This feature may be seen in more pronounced form in C^8-oxidised purines. Recent studies show that syn oriented C^8- oxidised guanines mispair with adenine to lead to trasversional mutations. It is believed that the bulky oxo group at C^8 position in guanine forces it to adopt syn form, thereby allowing it to mispair with adenine and subsequent mutational events. On the contrary, C^8-oxidised adenine is not particularly mutagenic, from which it may be understood that the anti to syn rotation barrier may not be feasible energetically.

1.10 Concluding Remarks

The thrust of this thesis is to investigate the above-reviewed modifications of DNA at the molecular level by using the methods of theoretical chemistry, viz. quantum chemical models for molecular structure and interaction. Throughout the course of this investigation, it is deemed of paramount importance to be able to furnish convincing correlations between theoretically calculated parameters and experimentally determined properties. Going further beyond the scope of mere rationalisation, attempts are also made to generate reliable predictions for particular properties of some untested cases. Furthermore, some of these molecular properties are in turn related to events of macromolecular significance and biological importance, the connection being supplied by the microscopic molecular basis that these properties of modified DNA provide for phenomena at the macroscopic level of observation.

CHAPTER II

APPROACHES TO THE RESEARCH PROBLEM

II.1 Theoretical Calculations and Experimental Observations

It has always been a challenge for the theoretician to test out the validity of his model by comparing the results of theoretical calculation with undeniable experimental findings. To this is also added the further dimension of being able to predict features in areas where the experimentalist has not yet ventured. The relative low cost of theoretical calculations in contrast to the heavy expenditure often involved in experiment is another reason why theoretical calculations are often so popular. The challenge of successful model building, coupled with the features of rationalisations, generalisations and predictions, are a common feature of applied science in almost every field today.

II.1.1 Scope of the molecular orbital model - an overview

The theoretical model adhered to in this thesis is the molecular orbital approach to molecular structure and interaction, based upon the foundations of molecular quantum mechanics. The methods of molecular quantum chemistry are used here to tackle the problems envisaged in this research problem. Two broadly different conceptual approaches to the approximate solution of the Schrodinger equation for molecular systems are possible within the molecular orbital framework. The first type, the so-called ab initio methodo

logy, uses very rigorous non-empirical quantum chemical concepts in solving the wave function without recourse to any approximations in evaluating the various integrals, including the many-centre electron repulsion integrals. On the other hand, the second type, known as semi-empirical methods, are based on the ZDO (Zero Differential Overlap) approximation applied in varying degrees of rigour, used in conjunction with the use of calibrated parameters for evaluating some of the various integrals involved, including the effect of electron repulsion integrals and some core potentials.

Either of these theoretical approaches may ultimately be tested by systematic comparison of its findings with known experimental results. If comparisons prove favorable, the model acquires some predictive value in situations where experimental data are unavailable. These various theoretical models have been subjected to intense testing in this manner with respect to their range of applicability, their successes, and their failures, all of which are thoroughly documented, covering the whole range of models from the frankly empirical to the most rigorous, including the configuration interaction and Moller-Plesset approaches to electron correlation (Hehre *et al* 1985).

Molecular orbital models based on the Linear Combination of Atomic Orbitals (LCAO) concept, within the framework of the Hartree-Fock Self-Consistent Field (HF-SCF) approximation, have gained wide acceptance for their flexibility in constructing the model at the level of both ab initio and semi-empirical theories. Theoretical models of molecular structure and interaction at the Hartree-Fock limit have become fairly well characterized, and many of its successes and limitations are now well documented (Hehre *et al* 1986). These approaches are augmented by the inclusion of electron correlation explicitly through configuration interaction or Moller-Plesset perturbation theory at various levels. Another area with scope for further development is the extrapolation of the

gas-phase model to the solvent phase through application of a wide range of models for the solvent phase.

II.1.2 Quantum chemistry and biological phenomena

While *ab initio* methods at high level theory may be regarded as intrinsically superior to semi-empirical methods, the large size of many important biological molecules proves prohibitive for treatment by the use of these methodologies. Moreover, ab initio methods employing minimal basis sets are reported to present no marked improvement over most semi-empirical methods, and in some cases are even inferior to the newer well-parametrized semi-empirical methods, like the AM1 SCF-MO method (Dewar and Storch 1985).

With increasing sophistication in numerical methodology and computer architecture, it might be well possible in the near future to apply high level ab initio methods to large biomolecules. However, in view of the limitations imposed by the available computational resources here, it is unfortunately not yet feasible for this author to make free recourse to these methods, and resort is made to the use of widely appreciated and well documented semi-empirical methods to tackle the present research problem.

II.2 Theoretical Methodologies Employed

The work embodied in this thesis exploits the application of semi-empirical Selfconsisistent Molecular Orbital theory (SCF-MO) to problems related to some fundamental chemical processes that occur in nucleic acids. Semi-empirical SCF-MO theory has taken on a gradual evolution over the past three decades from earlier models such as the EHT and CNDO/2 methodologies. These old models are now virtually obsolete, with the advent of the more recent semi-empirical models, viz. the AM1 and PM3 SCF-MO models, which

have found a signficant place in the application of quantitative molecular orbital theory. The following sections give a brief account of the theory underlying the MNDO, AM1 and PM3 models.

II.2.1 MNDO, AM1 and PM3 semiempirical MO methods

Current semi-empirical self-consistent field molecular orbital (SCF-MO) methods, as embodied in the MOPAC package, are based on application of the ZDO approximation at the NDDO level (Neglect of Diatomic Differential Overlap) to the existing HF-SCF LCAO framework, with the added feature of parametrisation. In these methods, the wave function is restricted to an all-valence electron population, the inner shell being represented by a core model. Empirically determined parameters are utilised in order to evaluate the various types of potentials and interactions in the multielectron system. The basis for this parametrisation is with reference to certain known results of experimental science, so that the accuracy of a semi-empirical method is limited to the accuracy of experimental data used in obtaining these parameters. Moreover, the universality of application of the parametrisation schemes employed is often a moot question. The current MOPAC package includes the MNDO (Dewar & Thiel 1977), AM1 (Dewar *et al* 1985) and PM3 models (Stewart 1989), which differ from each other chiefly in the parametrisation schemes employed.

The power of semiempirical methods lies chiefly in their low computational expense of utilisation. The adjustable parameters within the parametrisation scheme may be optimized to reproduce important chemical properties like molecular heats of formation, molecular geometries and dipole moments. Often, comparison of ab initio methods with semiempirical methods has shown that these salient properties, such as the heat of

formation, lend themselves to an accuracy of reproduction of experiment where the results of semiempirical methods are quite comparable to or fare even better than those of ab initio calculations employing fairly extended basis sets. Ideally, these parameters shoud be fully optimized using as complete a set of predetermined reference molecules as is ideally possible, but to date this has not proved completely possible for obvious reasons. As a consequence, the performance of any semiempirical method cannot be regarded as universally appropriate for all properties of all molecules. Care should thus be taken that the validity of each model used is in consonance with the particular molecular property of interest and with the chemical and structural framework of the molecules being subjected to calculation, so as to furnish reasonably good qualitative or quantitative predictions.

II.2.2 Solvent-phase simulation in SCF-MO theory

One disadvantage of the usual SCF-MO theory is that the wave-function corresponds to the situation in a hypothetical motionless vacuum state or in the gas phase, while most chemical reactions of usual interest in fact occur in solution. The importance of simulating the solvent phase may be seen in the extensive work expended towards implementation of solvent effects through incorporating perturbation of the gas-phase molecular wavefunction by the solvent molecule. This may be done within the scope of quantum mechanics itself or through theoretical models leaning towards a classical approach. To date, this body of work has resulted in essentially three distinct approaches of the Self-Consistent Reaction Field (SCRF) theory (Tapia *et al* 1980; Cramer *et al* 1995) used to tackle solute-solvent interactions.

In the so-called **extended Born approach**, the solute is treated as an array of atom-centered spheres, each of which contributes to the overall hydration energy in a manner

related to Born's original treatment of monoatomic ions. A number of applications of this Generalized Born Model (GBM) approach have appeared, including highly parametrized but promising versions of the AM1 and PM3 models, as developed by Cramer and Truhlar for both aqueous and non-aqueous environments (Cramer & Truhlar 1992a, 1992b; Cramer *et al* 1995).

In the second method, the solute is placed in an ellipsoidal or often simply a spherical cavity within the solvent with which it interacts via its dipole or in some cases higher-order multipole moments. Although this method has enjoyed considerable popularity, there are obvious limitations associated with the need to constrain an arbitrarily shaped solute to a cavity of specified shape. For similar reasons, its application to reaction processes is much limited (Ford & Wang 1992a, 1992b).

In a third approach, pioneered by Miertus, Scrocco, and Tomasi (the MST method), the solute is placed in a cavity of arbitrary shape composed of finite surface elements (Miertus *et al* 1981; Miertus & Tomasi 1982). The reaction field is modelled by an induced charge on the cavity surface that is computed explicitly from the normal component of the electric field obtainable from the molecular wave function. Various applications and refinements of this general procedure have been presented by the original workers and by others. The MST theory has recently been adapted to semiempirical Hamiltonia in the laboratories of Orozco and Luque (Negre *et al* 1992; Luque *et al* 1993; Luque *et al* 1994; Orozco *et al* 1994).

The development and performance of these methodologies based on the polarized continuum methods (PCM) was reviewed recently by Tomasi (Tomasi & Persisco 1994). While the performance of some of these models in simulating the aqueous phase is not well understood owing to their quite recent evolution, some of these methodologies have

been incorporated in commercially available molecular orbital packages within both *ab initio* and semiempirical formalisms.

II.2.3 Class IV atomic charges

The widely used zero-differential overlap Mulliken's population analysis (Mulliken 1955) is derived from the density matrix elements by appropriate summation. This method has been much criticised in the literature, especially in connection with extended basis sets. For the minimal basis set employed in AM1 and PM3, it yields atomic charges that follow physically reasonable trends but that are not quantitatively accurate. In an effort to obtain more accurate and meaningful partial charges, Truhlar's research group have developed the Class IV charge model (Storer *et al* 1995) in attempt to obtain charge distribution quantitatively comparable with high level ab initio atomic charges.

Class I charges are those that are obtained directly from experiment by simple models. Class II charges are obtained directly from a quantum mechanical wave function, while class III charges are obtained from a physical observable predicted from the wave function, e.g. charges derived from a best fit to the molecular electrostatic potential. Finally, Class IV charge models transform Class II or Class III charges to more accurately predict the desired experimental property. The parameters for Class IV charge models are determined in order to convert Mulliken charges (obtained from the AM1 or PM3 wave-functions) into charges that more accurately reproduce experimental gas- phase dipole moments. Some current Class IV charge models include the CM1A and CM1P corresponding to the AM1 and PM3 wave-functions respectively. The advantage of these models is that they can yield atomic charges similar to high level ab initio charges, but with a minimal computational effort as is appropriate to the inexpensive AM1 or PM3

methodologies. Use will be made of these Class IV charge models, (the so-called CM1 charge distribution) in this thesis when the traditional Mulliken charge distribution does not provide realistic representation.

II.3 Foundation Material for Research Problem

II.3.1 Reactions and properties of modified DNA

As the title of the thesis indicates, the chief aim of this work is to utilise the theoretical models of Sec. II.2 above in order to furnish rationalisations, generalisations and predictions with respect to certain key features and properties of chemically modified DNA. Since the macromolecular structure of DNA in its entirety precludes any reasonable theoretical approach, the problem is envisaged in terms of the constituents of DNA that are subjected to chemical modification. The following give a clue to the different chemical modifications of DNA considered here, as well as the properties incorporated for study here :

1. Proton transfer reactions in normal nucleic acid bases and their nucleosides, which reactions include tautomerism, protonation and proton abstraction. These reactions are directly connected to the tautomeric equilibrium constants as well as the acidic and basic pKa values for the protic acid-base equilibria.

2. Proton transfer reactions (including tautomerism and proton abstraction) in alkylated DNA bases and nucleosides. These reactions are again linked to the acidic and basic pKa values of the alkylated systems, and also to the presence or absence of mutagenic properties as manifested by them.

3. Tautomerism and glycoside bond rotation in C^8-oxidised purines, with their relevance for mutagenesis.

4. Base-pairing properties for alkylated DNA bases and C^8-oxo purines, with emphasis upon the steric, structural and energetic requirements for inducement of point mutations.

5. *In vitro* and *in vivo* mechanisms for the loss and repair of modified nucleosides, here restricted mainly to alkylated systems.

II.3.2 Theoretical concepts applied

Each of the above features of chemically modified DNA needs to be expressed in terms of a specific model and specific determinants for structure and reactivity so as to permit the phenomena to be accessible to theoretical analysis. The following general tenets and principles prove useful in defining this approach :

1. Proton transfer equilibria are expressed in terms of thermodynamic quantities, such as free energy and enthalpy changes, from which equilibrium constants can be calculated.

2. Enthalpy change in general proves to be as useful, and much less compute-intensive, than free energy change, owing to the cancelling out of the entropy term differences in approximately equal and parallel manner.

3. Reaction equilibria in aqueous phase are studied by extrapolating the gas-phase study to incoroporate the solvent phase models mentioned above, particularly the AM1-SM2 model.

4. The charge distribution, in terms of Mulliken or Class IV charge models, can provide a systematic scaling of reactive atoms with respect to nucleophilicity or electrophilicity (provided the electrostatic contribution to their reactivity is appreciable).

5. Rotational barriers are estimated in terms of enthalpy changes along the rotation coordinate.

6. Facility of hydrogen-bonded pairing between bases is estimated in terms of the net

lowering of enthalpy for the paired system.

7. In many cases, appeal is made to the Wiberg bond index as a useful estimate of intrinsic bond strength. Bond strength, eg. for the glycoside linkage, is also expressed in terms of the enthalpy of gas-phase bond dissociation.

8. Labilities of alkyl groups attached to nucleophilic sites are estimated in terms of a combination of the above type of indices, focussing upon enthalpy changes for formal reactions. By laying recourse to these models and indices, the chief features of the research problem here find expression in numerical and quantitative terms, which is useful for making comparisons and for analysis of trends. The actual numbers obtained, however, are purely subject to the limitations and drawbacks of the theoretical methodologies used.

CHAPTER III

PROTIC CHANGES IN DNA BASES AND NUCLEOSIDES

III.1 Opening Remarks

The primary role of nucleic acids and their derivatives as the hereditary determinants in biological reproduction and growth has led many researchers to delve into a proper understanding of the detailed three-dimensional structure and functioning as well as the various physical properties and chemical reactions of nucleic acids and their constituents. Central to the concept of nucleic acid structure and functioning are three main factors : (a) the hydrogen-bonding between bases of the two strands (in the plane determined by a base pair), (b) stacking interactions between bases within each strand, and (c) interactions among phosphate ions (one in each of the two strands), counterions, and water molecules. The relative importance of these contributions is still not definitively known, but hydrogen bonds between base pairs in DNA are arguably the most important ones in all of biological chemistry (Watson & Crick 1953). While in DNA one finds generally only two base pairs i.e. guanine-cytosine (GC) and adenine-thymine (AT), both in the Watson-Crick configuration, the presence and significance in DNA of other possible base-pair motifs (arising chiefly out of purine-purine, purine-pyrimidine mismatches via Watson-Crick and Hoogstein and reverse Hoogstein type structures) is now being increasingly appreciated (Cruse *et al* 1983). Similarly, alternative base-pairs arising from protonation (Gehring *et*

al 1993; Leroy *et al* 1993; Ahmed *et al* 1994; Leroy *et al* 1994) or tautomeric changes (Lowdin 1965; Piccirilli *et al* 1990) are receiving attention. In most of these cases, protonation of certain nucleophilic sites in DNA bases plays a critical role in stabilisation of the mismatched base-pairs. The *in vivo* stability of these protonated base-pairs depends chiefly on the associated pK_a values and the site of protonation. Thus the position and sites of the relevant protons and of the basic sites involved are of vital significance in answering questions related to nucleic acid structure, stability and function. Protic changes in nucleic acid components may thus drastically affect some of the salient features of nucleic acid structure and functioning.

III.2 Proton Transfer Reactions in Nucleic Acids

Protic changes are of three types : tautomerism, protonation and proton abstraction. The first refers to an intramolecular proton shift leading to tautomer formation. The second involves protonation of basic sites, while the third concerns the abstraction of a proton from a basic site. The question regarding tautomeric shift concerns the identity of the proton involved and the site to which it shifts. For protonation, the question concerns the position of the basic site to be protonated, while for deprotonation the inquiry pertains to the identity of the proton abstracted. For all these possibilities, the question of relative feasibility arises too.

For a molecule to be capable of undergoing all the three different possibilities given above, it should possess labile (acidic) protons to be transferred or abstracted, as well as basic (nucleophilic) sites for protons to attach themselves. Nucleic acid bases furnish the ground for all three possibilities, since they possess the required protons as well as a number of basic sites. A large number of potentially reactive basic sites are present in

nucleic acids and this understandably led many workers to assign a given ionization reaction (protonation or deprotonation) to different sites as discussed below. While in many cases the controversy over the particular site involved has been settled, in some cases the problem still remain unresolved. Note that a given protonation or deprotonation reaction is identified by the basic site involved in the acceptance or loss of the concerned proton. A tautomeric reaction, however, is identified by two sites - the one from which the proton leaves, and the one to which it is subsequently attached. Fig. III.1 gives the numbering systems for the 5 nucleic acid bases, while Table.III.1 lists all the basic sites and all the protons present in each nucleic acid base in its normal form predominant at biological pH.

The sites involved in the tautomeric shifts and the corresponding energies and tautomeric equilibrium constants reported (Singer & Grunberger 1983) are presented for the 5 nucleic acid bases in Table III.2, which also gives the protonation sites and reported acidic pK_a values for the 5 bases, as well as the deprotonation sites and reported basic pK_a values, as obtained from experiment and from assignment following the results of experiment.

This chapter attempts to examine theoretically, within the AM1 SCF-MO semiempirical gas-phase and solution-phase formalisms, the energetics involved in the protic shifts and transfers in nucleic acid bases as well as alkylated bases and their nucleosides. This information is then used to relate qualitatively and quantitatively the theoretically examined protic acid-base equilibria to the large body of experimental data on tautomerism, proton acidities, and the acidic and basic pK_a values noted for these systems. These then in turn may be related to certain phenomena of biological relevance (notably point mutagenesis), which connection exists through the molecular basis that these protic reactions furnish for the phenomena.

III.3 Tautomerism in Nucleic Acid Base Systems

Intramolecular proton shifts in nucleic acid bases and nucleosides are regarded as a possible basis for spontaneous mutations of DNA (Topal & Fresco 1976a, 1976b). A finite possibility exists for tautomer formation (Singer & Grunberger 1983) at biological pH, the tautomers G^*, A^*, C^* and T^* of guanine, adenine, cytosine and thymine respectively involving shifts in the hydrogen-bonding protons.

The possible involvement of the rare tautomers of the nucleic acid bases for low levels of spontaneous mutations common to all cells was suggested by Watson and Crick from the time of their classic papers on DNA structure. It is commonly beleived that the different hydrogen-bonding characteristics of the tautomer would lead to a thermodynamically favourable pairing with the wrong base and, hence, to its misincorporation during replication. This view was amplified by Topal and Fresco (1976a, 1976b) who attempted to relate average mutation frequencies to the tautomeric equilibria of free nucleosides. The fidelity of the replication process *in vivo* per one base pair was found to be 10^8-10^{10}. The more recent study of Ferscht and Knill-Jones (1994) has shown the error rate in the procaryotic cell to be higher and close to the rate *in vitro* i.e., one mistake in 10^6 to 10^7 base pairs. Consequently, the question of whether point mutations could be formed spontaneously via the tautomerism of free nucelotides in solution remains unanswered. The possible base-pairing properties of rare tautomer (enol and imino forms) with wrong bases was studied at various levels of theoretical methods and the Lowdin's hypothesis of double proton transfer in base-pairs was studied quite recently by both semi-empirical and ab inito methods (Lowdin 1977).

However, all the above mentioned studies considered only the keto and imino forms as the minor tautomers leaving behind the possibility of other intramolecular proton shifts

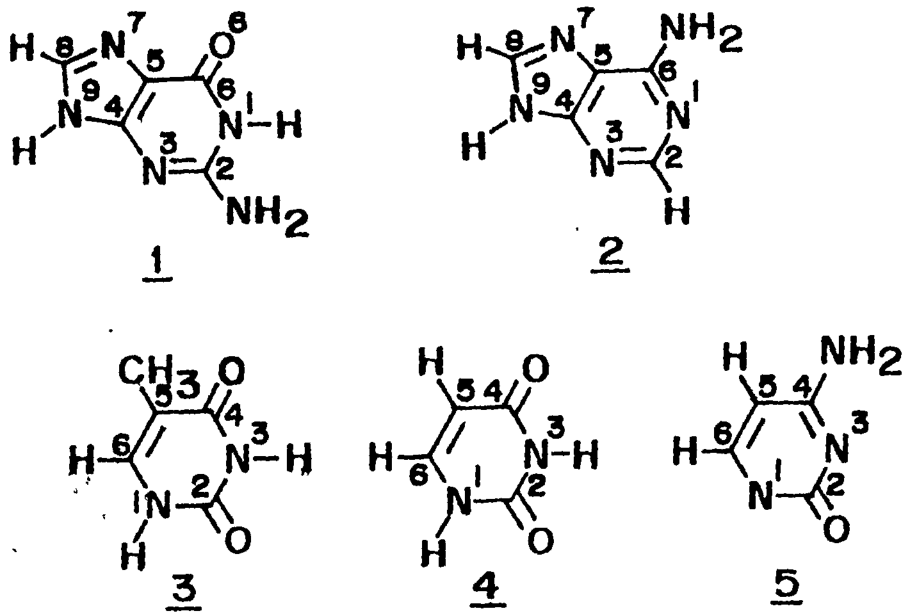


Fig III.1 : Numbering system for five nucleic acid bases; 1. Guanine 2. Adenine 3. Thymine 4. Uracil and 5. Cytosine

Table III.1 The basic sites and protons present in nucleic acid bases in their normal form predominant at biological *p*H.

Base system	basic sites	acidic protons present
Guanine	N^7,N^3,N^2,O^6	N^1,N^2,N^9
Adenine	N^7,N^3,N^1	N^6,N^9
Thymine	O^2,O^4	N^1,N^3
Cytosine	N^3,O^2,N^4	N^1,N^4
Uracil	O^2,O^4	N^1,N^3

Table III.2 Experimental data reported concerning protic shifts in nucleic acid bases, tautomeric transitions and their energies (E_t in kcal/mol), Calculated experimental equilibrium constants (K_t in $mol^{-1}K^{-1}$ at $298^{o}K$), protonation sites assigned and corresponding pK_a values, proton abstraction sites assigned and corresponding pK_a values (references as in text)

Base system	Tautomeric transition	E_t	K_t*	Prtn site	Acidic pK_a	Abstd proton	Basic pK_a
Guanine	N^1-G --> O^6-G	4.27	8.40×10^{-3}	N^7-G	3.3	N^1-G	9.42
Adenine	N^6-A --> N^1-A	14.06	1.46×10^{-7}	N^1-A	3.84	N^9-A	9.26
Cytosine	N^4-C --> N^3-C	12.68	6.86×10^{-7}	N^3-C	4.58	N^1/O^2	12.16
Thymine	N^3-T --> O^2-T /O^4-T	7.96	1.35×10^{-4}	?	< 0.5	N^3-T	9.90
Uracil	N^3-U --> O^2-U /O^4-U	7.59	2.04×10^{-4}	?	-3.38	N^3-U	9.51

* Equilibrium constants (K_t) are computed from the equation
$K_t = \exp(-E_t/RT)$ at 298 C where R is gas constant and T is temperature

unconsidered. No thorough study was reported on the possible significance of protic shifts to other nucleophilic sites on the five nucleic acid bases from the viewpoint of mutagenesis and carcinogenesis. The present theoretical study considers a variety of avenues for tautomer formation in order to establish which tautomeric routes might be predicted as the most feasible. Table III.2 also furnishes data on the experimentally established tautomeric transitions and energies.

The 5 nucleic acid bases guanine, adenine, cytosine, thymine and uracil were studied with respect to all the possible tautomeric shifts involving the Watson-Crick protons, viz the N^1- G, N^6-A, N^4-C, N^3-T and N^3-U protons. These are the protons believed to participate in the tautomeric shifts occuring spontaneously *in vivo*. The following shifts were studied : The N^1 proton of guanine to the N^2, N^3, N^7 and O^6 sites; the N^6 proton of adenine to the N^1, N^7 and N^3 sites; the N^4 proton of cytosine to the N^3 and O^2 sites, and the N^3 protons of thymine and uracil to the O^2 and O^4 sites. This study does not exhaust all tautomeric possibilities, since only shifts involving the Watson- Crick protons are included here. No note is taken of the proton shifts to or from the nitrogen atom bonded to the sugar moiety in the nucleosides (in an attempt to gauge the situation for the corresponding nucleosides). The tautomeric equilibria were studied using the AM1 SCF-MO method in gas-phase as well as simulated aqueous phase (Secs. III.2.1 and III.2.2 below).

III.2.1 AM1 gas-phase study of tautomerism

The AM1 gas-phase results are presented in Table III.3, giving the enthalpies ΔH_f and free energies ΔG_f of formation for all the tautomers studied, each being denoted in terms of the site on which the shifting proton is situated. The stabilities ΔH_s and ΔG_s of each tautomer with respect to the most stable one are also given, this being done to determinine

Table III.3. AM1 values for enthalpies ΔH_f and free energies ΔG_f of formation of nucleic acid bases and their tautomers, including their stabilities ΔH_s and ΔG_s with respect to the most stable tautomer in each case* (all values in Kcal/mol), and values of the equilibrium constants K_t for the tautomeric transitions based on calculated free energy differences at 25 C.

Tautomer	ΔH_f	ΔH_s	ΔG_f	ΔG_s	K_t
Guanine					
N^1-G	48.82	0.00	22.16	0.00	-
N^2-G	91.48	42.65	64.21	42.05	1.44×10^{-31}
N^3-G	62.98	14.15	35.88	13.72	8.65×10^{-11}
N^7-G	73.15	24.33	46.61	24.25	1.64×10^{-18}
O^6-G	53.47	4.64	26.75	4.59	4.30×10^{-04}
Adenine					
N^6-A	86.79	0.00	61.22	0.00	-
N^1-A	100.37	13.58	75.44	14.22	3.72×10^{-11}
N^7-A	129.35	42.56	104.40	43.18	2.14×10^{-32}
N^3-A	110.77	23.98	85.36	24.14	1.97×10^{-18}
Cytosine					
N^4-C	2.72	0.00	-21.59	0.00	-
N^3-C	6.19	3.47	-18.09	3.50	2.71×10^{-03}
O^2-C	24.41	21.69	0.13	21.46	1.82×10^{-16}
Thymine					
N^3-T	-61.01	0.00	-87.52	0.00	-
O^2-T	-40.79	20.22	-67.39	20.13	1.72×10^{-15}
O^4-T	-43.34	17.67	-69.30	18.22	4.33×10^{-14}
Uracil					
N^3-U	-53.86	0.00	-77.52	0.00	-
O^2-U	-33.29	20.57	-57.12	20.40	1.09×10^{-15}
O^4-U	-35.76	18.10	-59.42	18.10	5.30×10^{-14}

*Each tautomer represented by site of attachment of proton

how the more easily calculated enthalpy differences compare with the more rigorously obtained free energy differences.

Formally, the equilibrium constant K_t for a tautomeric transition involving a stabler tautomer **B** and another tautomer **B'** may be expressed in terms of the partition functions and the free energy difference (for the transition) as

$$K_t = Q/Q' \times \exp(-\Delta G/RT) \qquad \text{.........(III.1)}$$

where Q and Q'are the partition functions for the tautomers **B** and **B'** respectively, ΔG being the free energy difference for the tautomeric transition. The AM1 partition functions were calculated for a number of representative cases using the AMPAC package Version 2.1. It was found that, although the individual contributions to the partition functions varied from tautomer to tautomer, the net partition functions had identical values for any series of tautomers, so that Eqn. (III.1) above reduces to

$$K_t = \exp(-\Delta G/RT) \qquad \text{..........(III.2)}$$

Values of the equilibrium constants calculated this way are given in the last column of Table III.3, the temperature being 25° C (298° K). Using the relative values of the enthalpies and free energies of formation for the various tautomers, as well as the values of the equilibrium constants, the relative stability of the tautomers can be obtained.

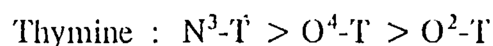
The most stable tautomers predicted for each case are the ones conventionally depicted as naturally occuring under normal conditions of temperature, pressure and pH.

These are the N^1-G, N^6-A, N^4-C, N^3-T and N^3-U tautomers of guanine, adenine, cytosine, thymine and uracil respectively, each tautomer being designated in terms of the basic site to which the concerned proton is attached. The order of stabilities of the different tautomers for each case are as depicted below, where the results obtained are the same whether using the ΔH_s or the ΔG_s index :

Guanine : N^1-G > O^6-G > N^3-G > N^7-G > N^2-G

Adenine : N^6-A > N^1-A > N^3-A > N^7-A

Cytosine: N^4-C > N^3-C > O^2-C


Thymine : N^3-T > O^4-T > O^2-T

Uracil : N^3-U > O^4-U > O^2-U

The most feasible tautomeric shifts predicted are the N^1-->O^6 shift for guanine, the N^6-->N^1 shift for adenine, the N^4-->N^3 shift for cytosine, and the N^3-->O^4 shifts for thymine and uracil. The most favourable shifts predicted thus all involve shift of the concerned proton to the most adjacent basic site, which in fact is what is generally held to be the situation in reality.

The stability ordering for the tautomers of both the purines follows the order N^1 > N^3 > N^7, where each tautomer is represented by the site of location of the concerned proton. The stability ordering for the tautomers of thymine and uracil both follow the order N^3 > O^4 > O^2, indicating that the C^5-methyl group of thymine would have no drastic effect on the relative stabilities of the tautomers. The charge separated tautomers N^7- G and N^2-G of guanine and the N^7-A tautomer of adenine are the least stable for each case, as would be expected from elementary considerations.

The relative feasibility of the most probable tautomeric shifts with respect to the base species is as follows : N^4-->N^3 for C > N^1-->O^6 for G > N^6-->N^1 for A > N^3-->O^4 for T$\simeq$$N^3$-->$O^4$ for U. The corresponding AM1 calculated free energies of tautomeric transition are 3.50, 4.59, 14.22, 18.22 and 18.10 kcal/mol for cytosine, guanine, adenine, thymine and uracil respectively. These values compare only broadly with those obtained experimentally, which are 12.68, 4.27, 14.06, 7.96 and 7.59 kcal/mol for cytosine, guanine, adenine, thymine and uracil respectively (Singer & Grunberger 1983). So, while these gas-phase AM1 calculations do not accurately reproduce the experimental values of the tautomeric transitions, correct predictions are made with regard to identification of the most stable tautomer as well as the most feasible tautomeric transition for each base system.

It may be noted that the relative stabilities from the view- point of free energy compare very well with those from the view-point of enthalpy, even though the individual free energies and heats of formation differ greatly. This is presumably because the entropy terms cancel out in approximately equal manner. This observation justifies the convenience introduced by substituting free energy change by enthalpy change for these cases as well as for the rest of the cases covered in this study of protic transfer equilibria.

III.2.2 Comparison of gas-phase and aqueous phase results

Since the above results of Sec. III.2.1. pertain to the gas phase, study of the tautomeric equilibria in aqueous medium appears to be called for here, which is in line with the biological situation in the cell. In order to assess the possible differentiating effect of aqueous solvation on the tautomeric energies and equilibria, the recent AM1-SM2 solvation model has been used to evaluate the aqueous free energies. The AM1-SM2 calculated

enthalpies, which implicitly considers the effect of aqueous environment within the extended Born approach (discussed in Chapter II), are presented in Table III.4 along with the AM1 gas-phase values for comparison. Under the AM1-SM2 data heading, the ENP (electronic, nuclear and polarisation) contributions and CDS (cavitation, dispersion and solvent rearrangement) contributions are also presented, along with their total contribution to the net enthalpy of formation.

Close comparison of the gas-phase and aqueous-phase data presented in Table III.4 reveals that, except for the case of guanine (where the N^7-protonated tautomer is predicted to be stabler than the O^6-protonated one), the energy differences between the various tautomers of a base are systematically similar to the gas-phase values, though the individual values vary (as expected from the energy-lowering effect of solvation). Fig. III.2a to 2e provides a schematic representation of the relative energies of various possible tautomers computed at the gas-phase AM1 and the aqueous phase AM1-SM2 levels of theory. This observed qualitative similarity between gas-phase and aqueous phase data sets predicts that the presence of a hydrated environment does not significantly alter the overall picture with regard to relative energetics. Table III.5 compares the reported experimental results with these AM1 gas-phase and AM1-SM2 data, which reveals that while the solvation AM1-SM2 model largely mimics the AM1 gas-phase results, neither do adequate justice to the situation as borne out by experiment, insofar as reproduction of tautomeric transition energies are concerned.

III.3 Protonation of Nucleic Acid Bases

Protonation of nucleophilic sites on the DNA base moieties is of relevance for the body of data on experimentally obtained acidic pK_a values for these systems. Assignment

Table III.4 AM1 and AM1-SM2 calculated Heats of formation of various nucleic acid tautomers (All values in Kcal/mol)

Tautomer	AM1	AM1-SM2				Relative Energies	
		ENP	CDS	Total	Hf+ Gs	AM1	AM1-SM2
Guanine							
-N1	48.32	-28.83	-11.47	-40.29	20.81	0.00	0.00
-O6(syn)	52.15	-18.65	-13.09	-31.74	28.55	3.83	5.64
-O6(anti)	53.47	-18.07	-13.13	-31.19	27.99	5.15	7.08
-N3	62.98	-48.31	-11.49	-59.83	29.83	14.66	8.92
-N7	73.15	-72.35	-11.58	-83.94	24.50	24.83	3.59
Adenine							
-N6	86.79	-20.29	-9.67	-29.96	63.06	0.00	0.00
-N1	100.37	-22.71	-10.12	-32.83	75.16	13.58	12.10
-N7	120.18	-57.35	-10.06	-67.41	78.85	33.39	15.79
-N3	109.24	-32.18	-10.19	-42.37	81.43	22.45	18.37
Thymine							
-N3	-61.01	-15.99	-6.15	-22.14	-77.87	0.00	0.00
-O4(syn)	-48.39	-15.99	-7.72	-23.71	-66.29	12.62	11.38
-O4(anti)	-43.34	-16.95	-7.75	-24.70	-61.56	17.67	16.11
-O2(syn)	-40.79	-23.01	-7.87	-30.64	-61.49	20.22	16.18
Cytosine							
-N4	2.72	-22.03	-8.35	-30.38	-19.03	0.00	0.00
-N3	6.19	-12.08	-8.59	-20.67	-10.92	3.47	8.11
-O2(syn)	24.41	-14.42	-10.36	-24.65	5.03	21.69	24.06
Uracil							
-N3	-53.86	-18.98	-6.59	-25.57	-72.42	0.00	0.00
-O4(syn)	-41.31	-17.10	-8.48	-25.58	-60.53	12.55	11.89
-O4(anti)	-35.76	-18.20	-8.45	-26.66	-55.31	18.10	17.11
-O2(syn)	-33.29	-27.25	-8.31	-35.56	-56.11	20.57	16.31

ENP : Contribution from electronic, nuclear and polarization terms

CDS : Contribution from cavitation, dispersion, and solvent-structural rearrangement arising from first-hydration shell effects

Note: All O^2 protonated tautomers of pyrimidine bases are stable only in syn conformation to N3 proton at both AM1 and AM1-SM2 level

Table III.5 Comparison of theoritically computed equilibirum constants for the most feasible tautomeric transition at AM1 and AM1-SM2 level with experimental equilibirum constants (E_t in Kcal/mol)

Base	Experimental		AM1		AM1-SM2	
	Transition	E_t	Transtition	E_t	Transition	E_t
G	$N^1 \to O^6$	4.27	$N^1 \to O^6$	3.83	$N^1 \to N^7$	3.59
A	$N^6 \to N^1$	14.06	$N^6 \to N^1$	13.58	$N^6 \to N^1$	12.10
C	$N^4 \to N^3$	12.68	$N^4 \to N^3$	3.47	$N^4 \to N^3$	8.11
T	$N^3 \to O^4$	7.96	$N^3 \to O^4$	12.62	$N^3 \to O^4$	11.38
U	$N^3 \to O^4$	7.59	$N^3 \to O^4$	12.55	$N^3 \to O^4$	11.89

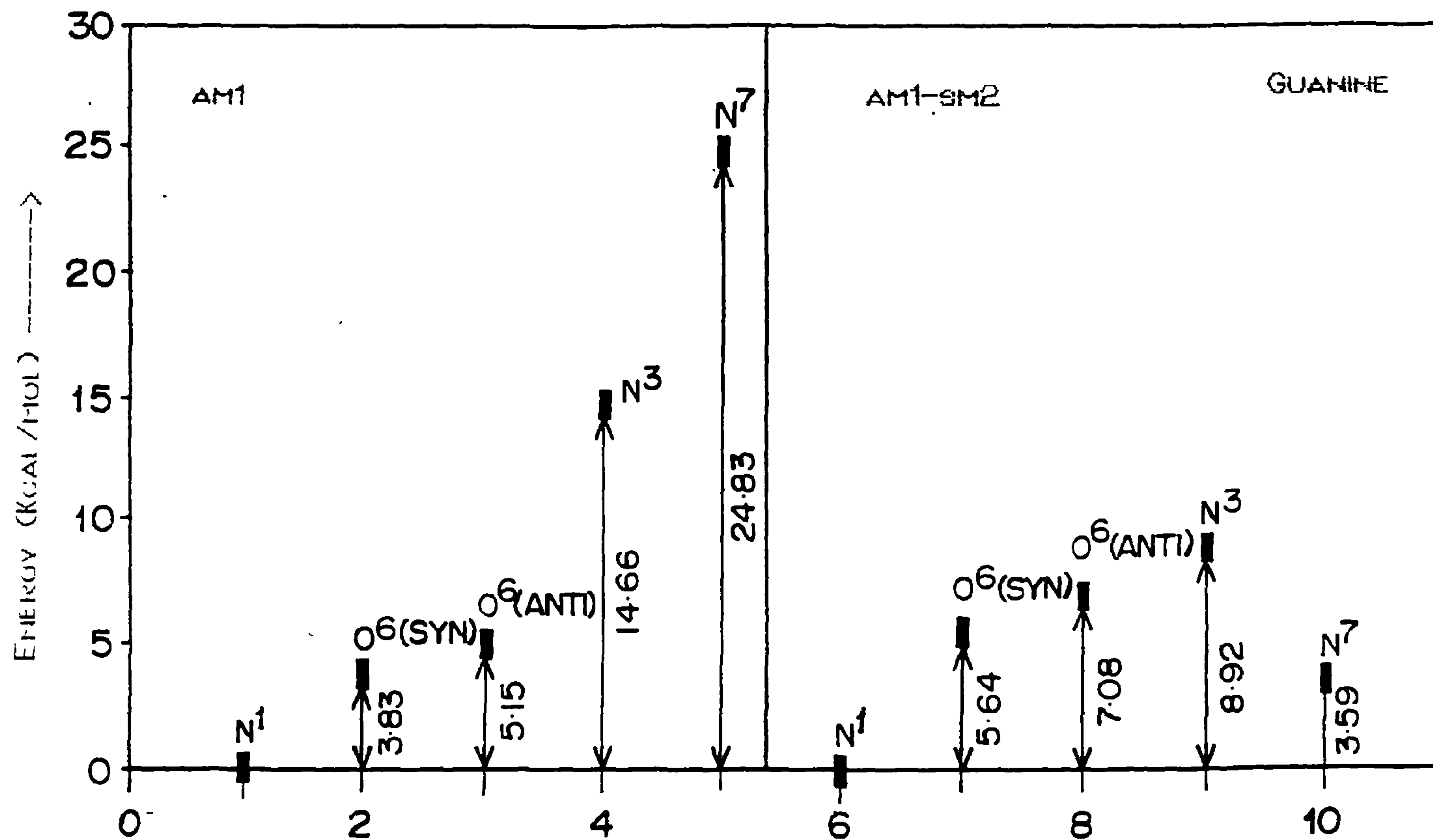
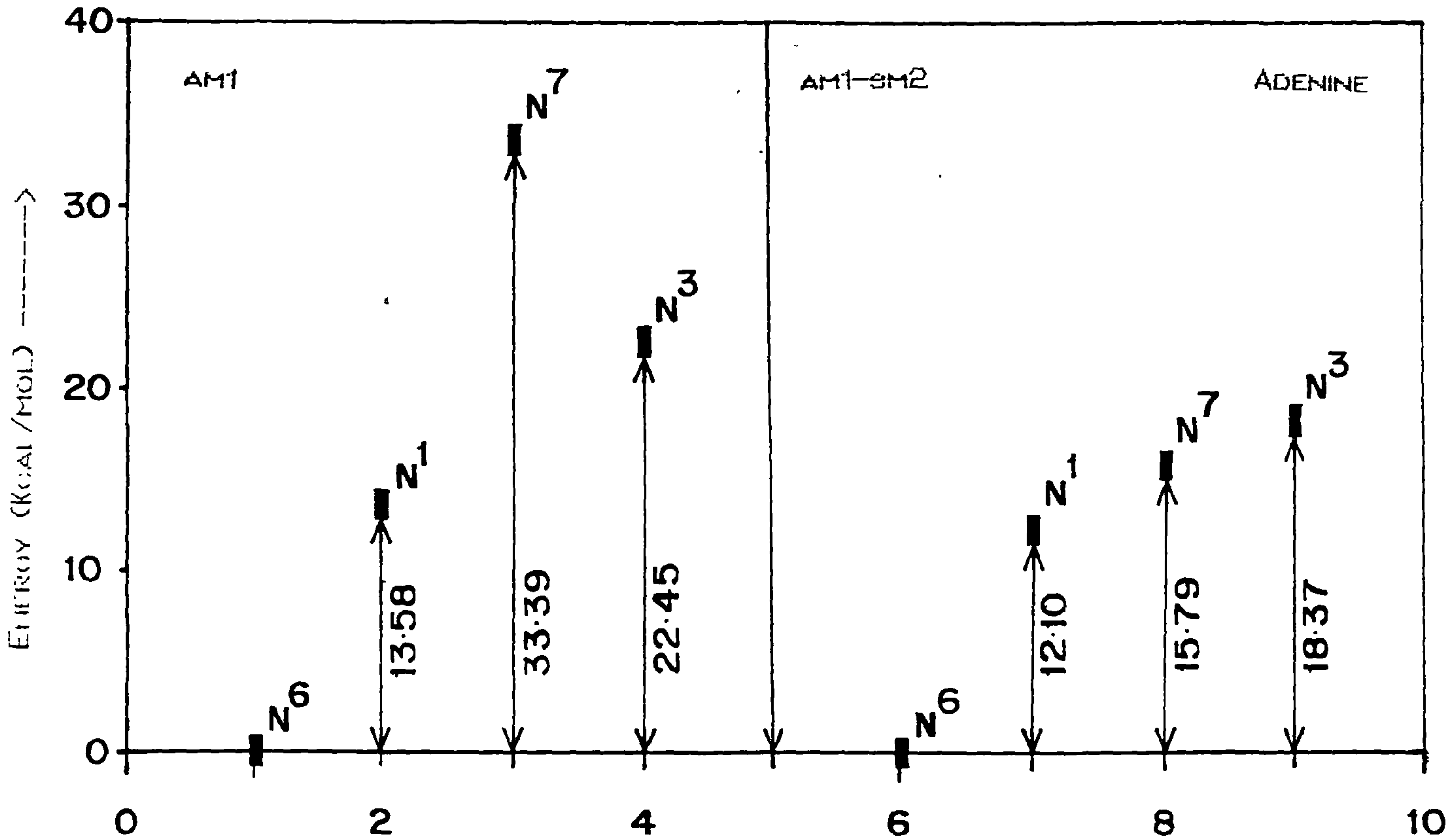
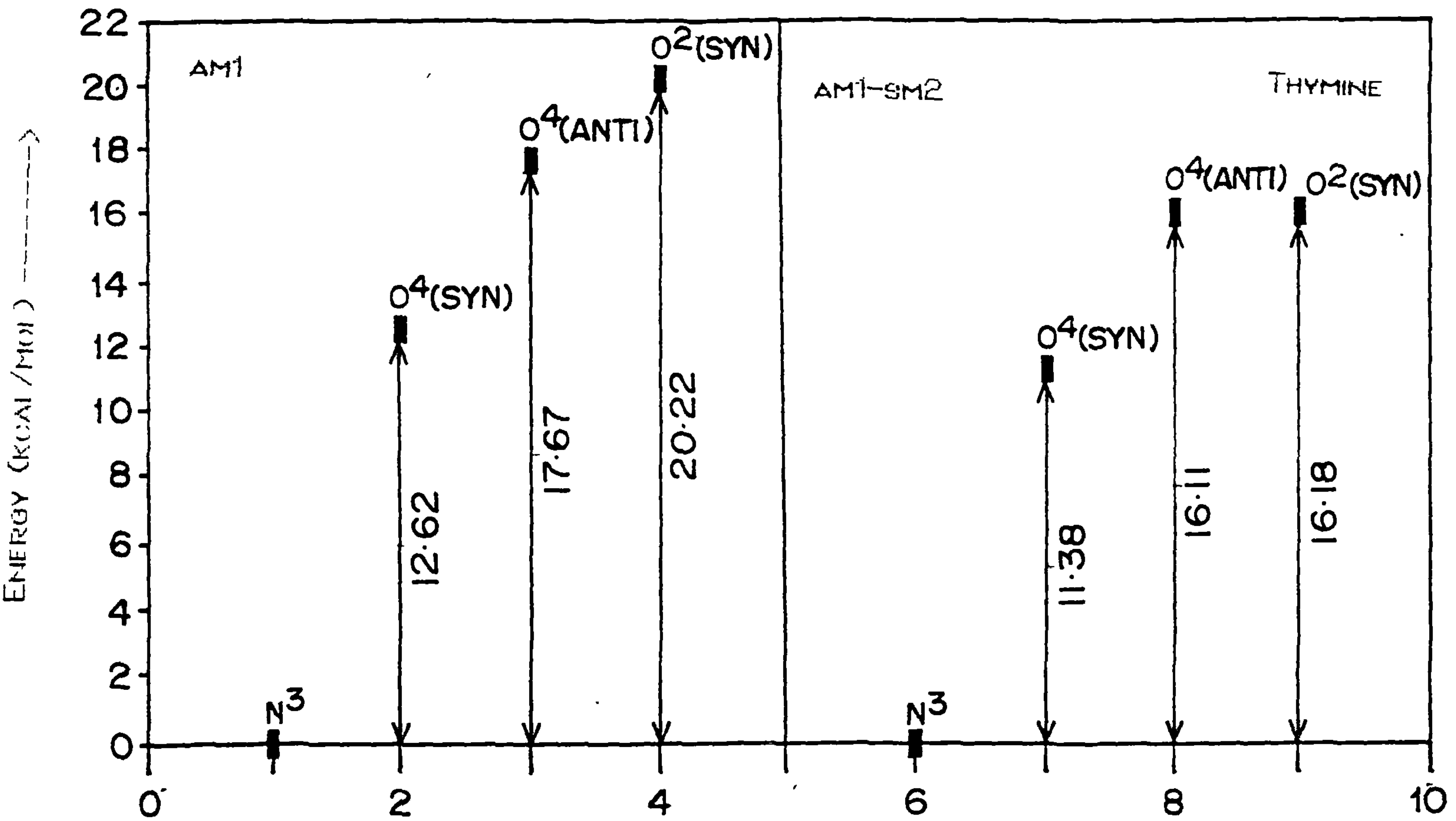
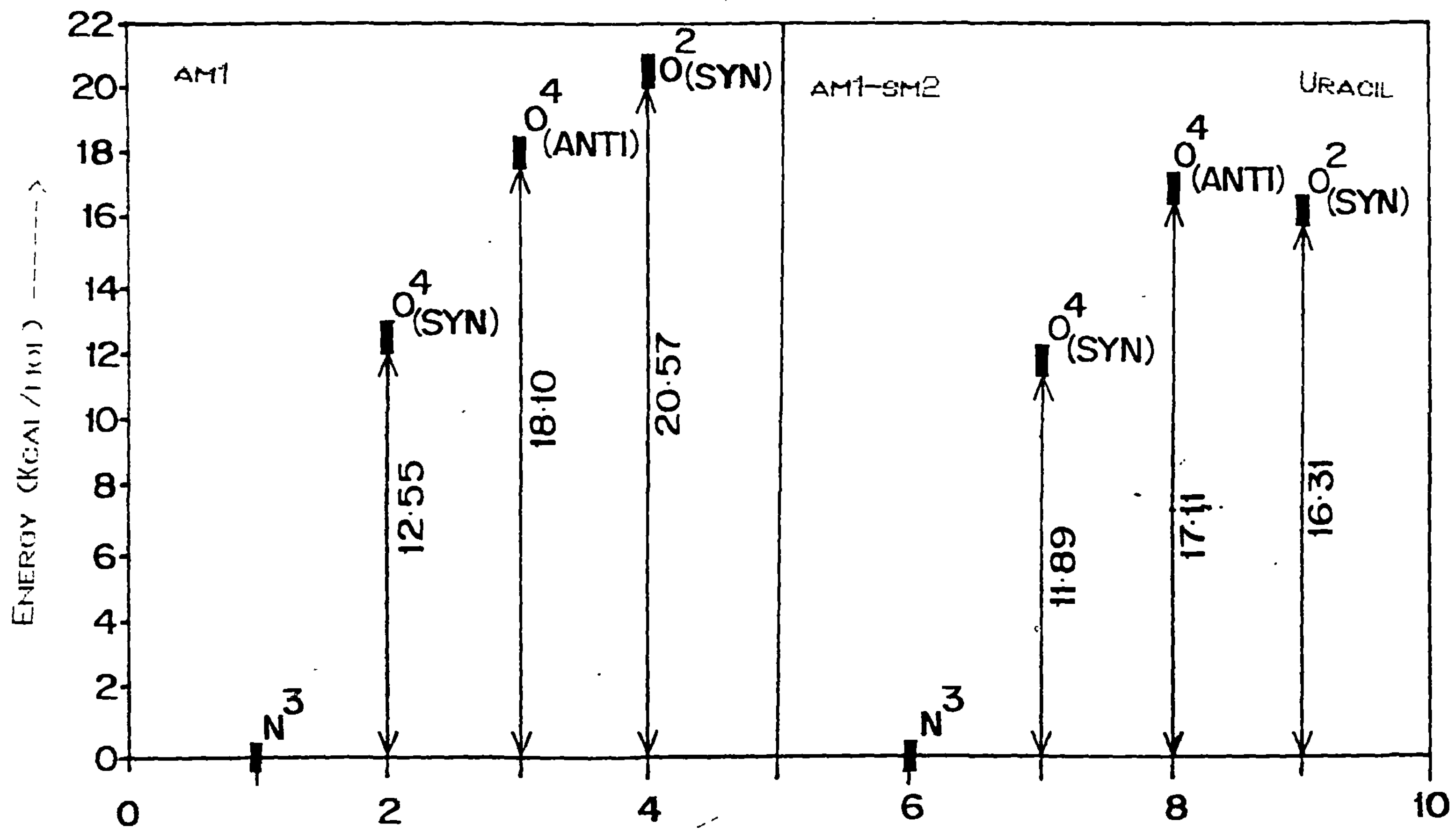


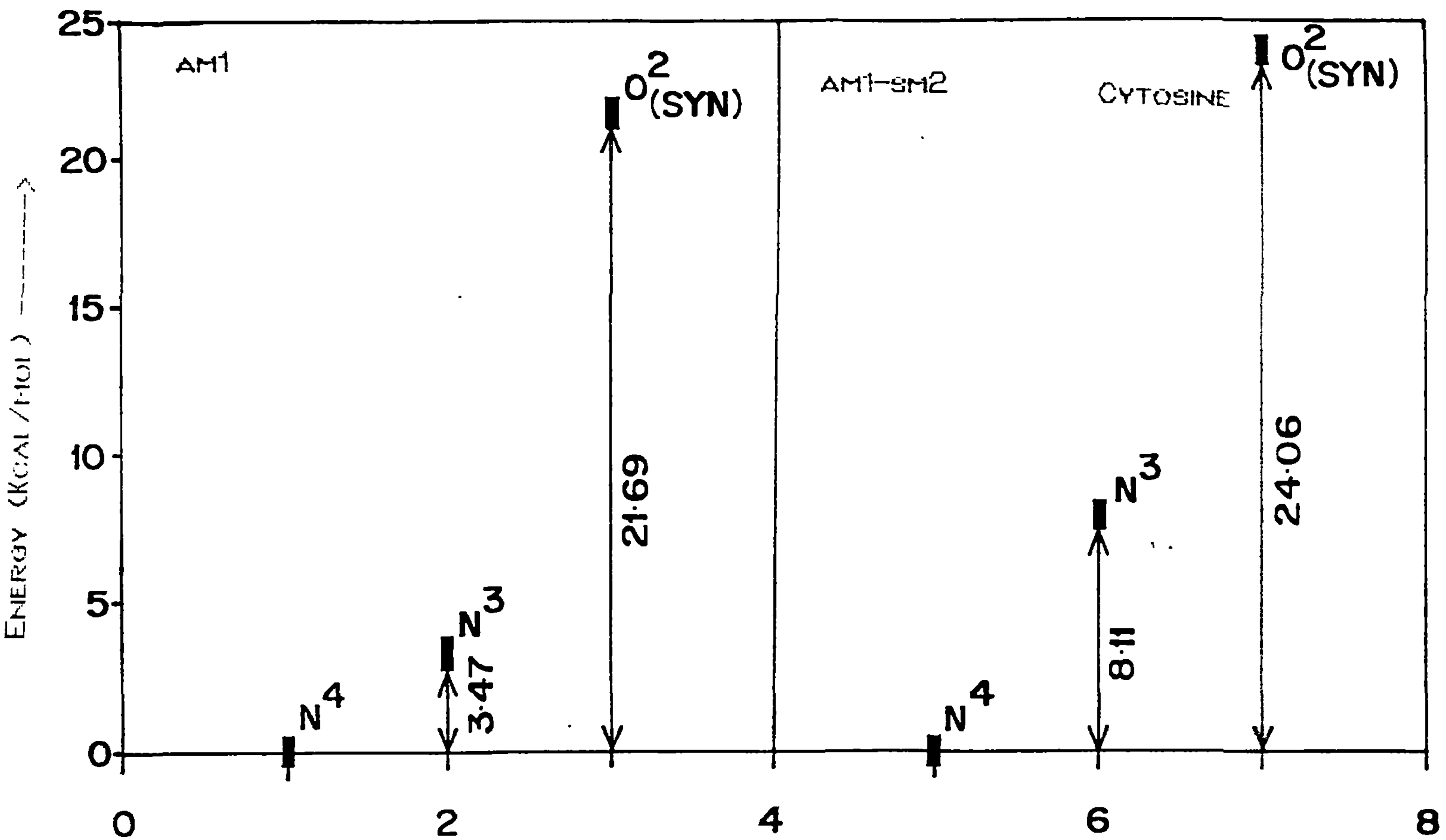
Fig. III.2a-2e : Schem atic representation of the relative energies ofvarious tautomers of five nucleic acid bases computed at AM1 and AM1-SM2 levels


AM1
AM1-SM2
ADENINE
ENERGY (KCAL/MOL) ------->
N6
N1
N7
N3
13·58
33·39
22·45
12·10
15·79
18·37
0
10
20
30
40
2
4
6
8


ENERGY (KCAL/MOL) ------->
22
20
18
16
14
12
10
8
6
4
2
0
AM1
N3
O4(SYN)
12·62
O4(ANTI)
17·67
O2(SYN)
20·22
AM1-SM2
N3
O4(SYN)
11·38
O4(ANTI)
16·11
O2(SYN)
16·18
THYMINE
0
2
4
6
8
10



AM1
AM1-SM2
URACIL
ENERGY (KCAL/MOL)
N^3
O^4(SYN)
O^4(ANTI)
O^2(SYN)
12.55
18.10
20.57
11.89
17.11
16.31
0
2
4
6
8
10
12
14
16
18
20
22



AM1
AM1-SM2
CYTOSINE
N4
N3
O2 (SYN)
3·47
21·69
8·11
24·06
ENERGY (KCAL/MOL) ------>
0
5
10
15
20
25
0
2
4
6
8

of protonation sites has largely followed experimental results, taking cues from theoretical and mechanistic considerations as well. For the sake of making comparisons, the pK_a values quoted here were all taken in aqueous solvent at 25°C.

III.3.1 Experimental data on nucleic acid base protonation

Experimental data here pertains to the assignment of protonation sites and corresponding acidic pKa values. The 5 nucleic acid bases are dealt with as follows :

Shapiro has pointed to the evidence for the N^7 position as the most likely protonation site for 9-substituted guanine nucleosides (Shapiro 1968) . He appealed to the results of X-ray diffraction, ir and nmr studies (Jardetzky & Jardetzky 1960; Miles *et al* 1963a, 1963b; Iball & Wilson 1965). Earlier studies, reviewed by Shapiro (1968), had erred in citing the 2-amino nitrogen as the protonation site. The acidic pK_a values reported for guanine, guanosine, deoxyguanosine and for guanine residues in DNA are 3.3, 1.9, 2.27 and 3.45 respectively (Reinert & Weiss 1969; Christensen *et al* 1970; Benoit & Frechette 1985).

The N^6-position of adenine was earlier wrongly assigned as the protonation site for adenine and its N^9-nucleosides (Taylor 1948; Alberty *et al* 1951; Pullman *et al* 1956; Beers & Steiner 1957). Later transition state calculations led Pullman to identify the N^1 position as the most likely protonation site in adenine (Pullman 1959). X-ray crystallographic studies (Cochran 1951), calorimetric studies (Christensen *et al* 1962, 1970a, 1970b), the arguments of Zubay (1958) and the ^1H and ^{13}C NMR studies of Benoit *et al* (1985) by excess acidity methods have led to the general opinion now that the N^1 site is the most likely protonation site for adenine and its derivatives. Acidic pK_a values reported for adenine, adenosine, deoxyadenosine and adenine residues in DNA are respectively 3.835, 3.50, 3.77 and 4.25 (Reinert & Weiss 1969; Christensen & Izatt 1962; Jordan *et al* 1956).

For cytosine and its derivatives, the N^1 atom had been cited earlier as the most likely protonation site (Lewin & Humphreys 1966). However, later calorimetric, nmr and absorption spectroscopy studies on cytosine protonation showed ionisation as occuring from the N^1 site (Miles *et al* 1963a, 1963b; Katritzky & Waring 1963; Ueda & Fox 1963; Christensen *et al* 1970a, 1970b). Though a tautomeric equilibrium involving microspecies had been proposed (Lewin & Humphreys 1966), proton nmr studies of the Cu^{2+}-cytidine system (Berger & Eichhorn 1971) eliminate binding to the N^4 site, inferring the N^3-C position as the most possible protonation site. The acidic pK_a values for cytosine, cytidine, 5'-dCMP and cytosine residues in DNA are reported as 4.58,4.22,4.44 and 4.50 respectively (Christensen *et al* 1967; Fox *et al* 1953; Cox 1966).

Protonation of cytosine plays an important role in modified DNA quadriplexes

For some time it has been known that nucleic acids containing stretches of cytidine residues can form parallel strands held together by cytosine-: protonated cytosine base pairs $C.C^+$ (Langridge & Rich 1963; Inman 1964). The structures of four-stranded molecules formed by two paralle-stranded duplexes, held together by $C:C^+$ base pairs are reported by NMR structural studies on d(TC_5) (Gehring *et al* 1993). It is beleived that the most important component of the stabilization of these four-stranded intercalative cytosine structures is the hemi-protonation of the cytosines giving rise to three hydrogen bonds.

For uracil, thymine and their ribonucleosides, protonation in strongly acidic pH ($<$ 0.5) was earlier reported using spectrophotometric data (Cohn 1955). No clearcut assignment of the ionisation sites has been reported so far. For uracil, Shapiro (1972) has confirmed the pK_a to be -3.38.

Protonation at basic sites in the 5 nucleic acid bases was studied here for the

following sites in each base : N^2, N^3, N^7 and O^6 of guanine; N^1, N^3, N^6 and N^7 of adenine; O^2 and N^4 of cytosine; and O^2 and O^4 of thymine and uracil. No consideration was made of the atoms attached to the sugar in the corresponding nucleosides, in order to assess the probable situation prevailing for the nucleosides.

III.3.2 AM1 approach to protonation facility

Based upon the AM1 SCF-MO obtained gas-phase equilibrium geometries and wave functions, the following AM1 theoretical indices for protonation facility were employed :

The enthalpy of protonation was defined in two ways : (a) From the reaction corresponding to addition of a bare proton to the nucleophilic site, described by the enthalpy term $\Delta H_p(p)$; and (b) from the water assisted protonation given in Eqn. (III.3) below, described by the enthalpy term $\Delta H_p(w)$:

$$B + H_2O \longrightarrow (BH)^+ + OH^- + \Delta H_p(w) \quad .... \text{(III.3)}$$

where **B** and **(BH)**$^+$ represent the normal and protonated forms of the base respectively.

In addition, protonation facility was estimated from the basicity of the nucleophilic site **X** protonated in the base, given by the Mulliken charge Q_x, and also by the Wiberg bond index W_{hx} for strength of the **H-X** bond formed on protonation. The results of other workers on the molecular electrostatic potential minima (MEP) associated with the sites **X** were also incorporated for comparison.

III.3.3 AM1 results on nucleic acid base protonation

Table III.6 furnishes the AM1 calculated values of the various theoretical indices for

Table III.6 AM1 values for indices of facility of protonation along with *ab initio* MEP values, of nucleic acid bases (ΔH_p(p) and ΔH_p(w) in Kcal/mol)

Species	$\Delta H_p(p)$	$\Delta H_p(w)$	W_{hx}	Q_{xh}	MEP
Guanine					
N^2-G	-189.53	-26.53	0.8996	0.2825	-2.0
N^3-G	-211.29	-48.30	0.8663	0.2920	-64.0
N^7-G	-222.08	-59.08	0.8513	0.3227	-91.0
O^6-G	-210.35	-47.35	0.9018	0.2642	-75.0
Thymine					
O^2-T	-196.24	-33 24	0.8896	0.2947	-50.2
O^4-T	-203.59	-40.59	0.8839	0.3023	-54.6
Adenine					
N^6-A	-203.65	-40.65	0.9034	0.2713	-24.8
N^1-A	-220.70	-57.70	0.8711	0.2858	-77.0
N^7-A	-212.78	-49.78	0.8604	0.3056	-69.4
N^3-A	-221.31	-58.31	0.8666	0.2914	-78.3
Cytosine					
N^4-C	-195.63	-32.63	0.9008	0.2773	-13.7
O^2-C	-214.46	-51.46	0.9044	0.2682	-84.8
N^3-C	-225.67	-62.67	0.8607	0.2977	-93.5
Uracil					
O^2-U	-195.32	-32.32	0.8855	0.3026	-
O^4-U	-201.97	-38.97	0.8869	0.3021	-

protonation facility, viz. the $\Delta H_p(p)$, $\Delta H_p(w)$, W_{hx} and Q_x indices for protonation at the various sites on the 5 nucleic acid bases. Table III.5 also compares the enthalpy indices with calculated values of the molecular electrostatic minima (MEP) for each site in the solitary base as obtained by Bonnaccorsi *et al* (1972, 1975) using *ab initio* level theory.

It is seen that the four types of indices serve to furnish the same broad trends for prediction of relative basicity in each case. The density matrix-based indices W_{hx} and Q_x predict more or less the same trends for protonation facility as the enthalpy indices, both for variation of site within a base as well as for comparison between different bases. Basicity being primarily a thermodynamic concept, chief recourse was made to the enthalpy index $\Delta H_p(w)$. For each base, the predicted ordering for basicity of the sites as predicted by this index is given as follows :

Guanine : N^7-G > N^3-G O^6-G > N^2-G

Adenine : N^3-A $\simeq$ N^1-A > N^7-A > N^6-A

Cytosine: N^3-C > O^2-C > N^4-C

Thymine : O^4-T > O^2-T

Uracil : O^4-U > O^2-U

The most basic sites in each base as predicted by these gas- phase AM1 calculations are the N^7-G, N^3-A (close to N^1-A), N^3-C, O^4-T and O^4-U sites of guanine, adenine, cytosine, thymine and uracil respectively. These semiempirical AM1 gas-phase predictions are thus in complete consensus with the *ab initio* molecular electrostatic potential minima indentifications of Bonnacorsi *et al* (1972, 1975). Both these sets of theoretical results are also seen to follow the general consensus of opinion as gathered from experiment for

guanine and cytosine, as discussed above, although not quite so for the case of adenine. However, the closeness in values here (less than 1 kcal/mol difference) between the $\Delta H_p(w)$ values for the N^1-A and N^3-A sites, if not reflecting on the limitations of these calculations, could suggest the possibility of competition. For thymine and uracil, with no experimental confirmation of protonation site, the results predict the O^4 sites as the most likely.

The ring nitrogens are predicted to be more basic than the exocyclic amino nitrogens, as would be expected from the presence of two already-bonded hydrogens in the latter. Protonation at an amino nitrogen is predicted as being least favourable among all sites for guanine, adenine and cytosine. The environment provided by the O^6 atom in guanine makes the N^7-G site particularly basic as compared to the N^7-A site in adenine. Likewise, the O^6-G site is rendered more basic by the imidazole nitrogen than the other oxygen sites on cytosine, thymine and uracil, as seen from a comparison of the enthalpy indices.

Table III.7 summarises these findings by presenting the most basic site in each of the 5 bases as predicted here, coupled with the values of the $\Delta H_p(w)$ index for protonation facility, the MEP values of Bonnaccorsi *et al*, the experimentally obtained pK_a values, and the experimental assignment of ionisation site. The predicted order of basicity is thus seen to be U $\approx$ T > C > A $\approx$ G, the reverse of which would point to the predicted ordering for the acidic pK_a values of the bases. The experimental ordering for the pK_a values is, however, T (U) > A > G > C, indicating not a very good degree of positive correlation. Here again, the predictions of the AM1 method are fairly reliable so far as identification of protonation sites is concerned, while quantitative correlation between gas-phase proton affinities and solvent-phase pK_a values is poor. The correlation might be better if the

Table III.7 Theoretical data for relative basicity of basic sites *in nucleic acid bases compared with pKa (acidic) values and their* assignments ($\Delta H_p(w)$ values in Kcal/mol)

Base	Site	$\Delta H_p(w)$	MEP	Acidic pK_a	Assignment[a]
G	N^7-G	-59.08	-91.0	3.30	N^7
A	N^3-A	-58.31	-78.3	3.84	N^{7}*
C	N^3-C	-62.67	-93.5	4.58	N^3
T	O^4-T	-40.59	-50.2	< 0.50	?
U	O^4-U	-38.97	-	-3.38	?

a Experimental results (references in text)
* close to N^3-A

gas-phase proton affinities were considered instead of the pK_a values, but these are not available in the literature. Recourse to calculations incorporating models for the solvent phase might also improve the picture here, as has been done for deoxyguanosine (Ford and Wang 1993).

III.4 Proton Abstraction in Nucleic Acid Bases

Deprotonation from nucleic acid bases and their nucleosides pertains to data on basic pK_a values for these systems. For alkylated DNA bases, it is connected to mechanisms for carcinogenesis and mutagenesis (Saffhill *et al* 1985; Lindahl *et al* 1988; Duncan & Davies 1989; Lyngdoh 1994). This Section focusses on deprotonation reactions for the free nucleic acid bases G, A, C, T and U. The precise identification of the protons involved here is uncertain in some cases. The basic pK_a values reported here all pertain to aqueous solvent at 25^0 C, so as to allow for comparisons on an equal basis.

III.4.1 Experimental data on nucleic acid base deprotonation

Basic pK_a values have been reported as 9.42, 9.25 and 9.26 for guanine, guanosine and deoxyguanosine respectively (Zubay 1958; Katritzky and Waring 1963), where the first and perhaps the last would involve the N^1 site, not the ribose moiety. Proton ionisation from the ribose group of guanosine has been shown without identification of site (Zubay 1958). The basic pK_a value for the guanine residue in DNA is 11.6 (Suchorukow *et al* 1964), perhaps corresponding to ionisation from the N^1-G site.

Deprotonation for free neutral adenine has been assigned to the N^9 proton (Lewin 1964). For the 9-substituted ribonucleoside derivative, the ionisation site has been shown to be located at the ribose moiety, both the 2'- and 3'-hydroxy groups being required for

this ionisation (Izatt *et al* 1966). On this basis, proton loss from deoxyadenosine, although not reported, would have to be from the N^6-amino proton. The pK_a values reported for adenine and adenosine are 9.259 and 12.35 respectively (Makar and Williams 1974; Izatt *et al* 1966), corresponding to proton loss from the N^9 site and sugar moiety respectively.

The basic pK_a value of 12.15 (Christensen *et al* 1967) for cytosine has been attributed to the N^1HC^2O grouping, where proton loss could be either from a hydroxyl group or an acidic imino group, depending upon which tautomer is involved. For cytidine, the most facile proton loss has been inferred as occuring from the ribose moiety (Christensen *et al* 1970a,1970b), for which a value of 12.5 has been reported.

The proton ionisation occuring with a pK_a value of around 9.6 for uracil, uridine, thymine and thymidine has had various sites assigned to it, including the 2-hydroxyl group for neutral uracil and thymine (Shugar and Fox 1952). The possibility of existence together in basic pH of both diketo and hydroxy forms has been invoked to seek rationale for the varying assignments. Both the N^1 and N^3 protons of uracil have been concluded to be involved in a simultaneous ionisation from both sites (Lewin 1964). N^1- deprotonation would be ruled out, of course, for the nucleosides, for which calorimetric results point to the first ionisation as occuring from the N^3 position. The basic pK_a values for thymine and thymidine are 9.90 and 9.79 respectively (Christensen *et al* 1967), while those for uracil, uridine and deoxyuridine are 9.28, 9.17 and 9.3 respectively (Christensen *et al* 1970a,1970b)

III.4.2 AM1 approach to deprotonation facility

For each nucleic acid base, a variety of possible protons were considered for the abstraction reaction, which included the following : guanine (N^1, N^2 and N^9), adenine (N^6 and N^9), cytosine (N^1, O^2 and N^4), and thymine and uracil (N^1, O^2 and N^3). The pyrimidine systems

incorporate the O^2-C, O^2-T and O^2-U tautomers for study as well, since there is some question as to involvement of these species as microspecies in the equilibria.

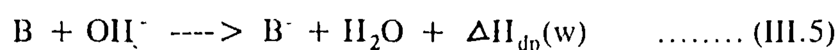
For each of these alternatives, AM1 SCF-MO based theoretical indices for deprotonation facility were framed as follows :

(a) Q_h, the Mulliken positive charge on the proton concerned,

(b) W_{xh} (the Wiberg index for bond strength between the proton **H** and the nucleophilic atom **X** to which it is attached,

(c) the enthalpy of abstraction of the proton, calculated in terms of the gas-phase proton abstraction enthalpy $\Delta H_{dp}(p)$ as given in Eqn. (III.4), and in terms of the enthalpy $\Delta H_{dp}(w)$ for deprotonation by a hydroxide anion as in Eqn. (III.5) :

$$B \longrightarrow B^- + H^+ + \Delta H_{dp}(p) \qquad \text{........ (III.4)}$$



$$B + OH^- \longrightarrow B^- + H_2O + \Delta H_{dp}(w) \qquad \text{........ (III.5)}$$

where **B** and **B**$^-$ represent the normal (neutral) and deprotonated (anionic) forms of the bases.

III.4.3 AM1 results on nucleic acid base deprotonation

Table III.8 presents the AM1 calculated indices for the feasibility of abstraction of these protons in the base systems, viz. the $H_{dp}(p)$, $H_{dp}(w)$, W_{xh} and Q_h indices. It is seen that the four sets of indices provide more or less the same basic trends of prediction. The density matrix-based indices (W_{xh} and Q_h) follow trends roughly similar to those furnished by the enthalpy indices with regard to the variety of protons in any base, as well as with regard to comparisons between different base systems. Since proton acidity is primarily

Table III.8 AM1 calculated indices for facility of deprotonation of nucleic acid bases($\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ values in Kcal/mol)

Proton	$\Delta H_{dp}(p)$	$\Delta H_{dp}(w)$	W_{xh}	Q_h
Guanine				
N^1-G	335.84	- 74.98	0.8834	0.2484
N^2-G	337.28	-72.54	0.8982	0.2066
N^9-G	346.03	-64.79	0.8776	0.2669
Adenine				
N^6-A	350.00	-60.82	0.9028	0.2407
N^9-A	332.89	-77.93	0.8765	0.2688
Cytosine				
N^1-C	338.87	-71.95	0.8827	0.2520
O^2-C	337.14	-73.68	0.9194	0.2446
N^4-C	351.49	-59.33	0.9033	0.2413
Thymine				
N^1-T*	327.94	-82.86	0.8784	0.2683
O^2-T	312.62	-98.20	0.9241	0.2317
N^3-T	344.09	-66.73	0.8742	0.2677
Uracil				
N^1-U*	328.39	-82.43	0.8784	0.2683
O^2-U	319.07	-91.75	0.9074	0.2675
N^3-U	344.48	-66.34	0.8743	0.2675

*The N^1-protons here for thymine and uracil pertain to the less stable tautomer in each case

as thermodynamic concept, the focus here is upon the enthalpy indices $\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$. The order of proton acidities for each base is predicted as follows by the $\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ indices :

Guanine : $N^1 > N^9 > N^2$

Adenine : $N^9 > N^6$

Cytosine : $N^1 > N^4 > O^2$

Thymine : $N^3 > N^1 > O^2$

Uracil : $N^3 > N^1 > O^2$

The above order of acidities could be interpreted as furnishing predictions for the successive proton ionisations in each case, where the protons attached to ring nitrogens are predicted to be more acidic than those attached to the exocyclic amino groups. The most acidic protons for each system are thus predicted to be as follows : N^1 (guanine), N^9 (adenine), N^1 (cytosine), and N^3 (thymine and uracil). This compares well with the assignments made following the observations of experiment in the cases of guanine, adenine, thymine and uracil, as discussed above, and lends further theoretical confirmation to these assignments. For the case of cytosine, the N^1 proton of the normal tautomer is predicted to be more acidic than the O^2 atom of the less stable tautomer. Similar arguments hold for the cases of uracil and thymine, where the N^3 proton of the normal tautomer is more acidic than the O^2 proton of the less stable tautomer.

Table III.9 sums up these findings by presenting the protons predicted here as most acidic (identified with respect to the basic site to which they are attached), their corresponding AM1 enthalpies of deprotonation $\Delta H_{dp}(w)$, the reported basic pK_a values,

Table III.9 Theoretical data for relative acidity of acidic sites in nucleic acid bases compared with pK_a (basic) values and their experimental assignments (ΔH_{dp} values in Kcal/mol)

Base	Site	$\Delta H_{dp}(w)$	Basic pK_a	Assignment[a]
G	N^1	-74.98	9.42	N^1
A	N^9	-77.93	9.26	N^9
T	N^3	-66.73	9.90	N^3C^4O
U	N^3	-66.34	9.51	N^3
C	N^4	-59.33	12.16	N^1/C^2O

[a] Experimental results (references in text)

and the reported assignments of deprotonation sites. With respect to the base systems, the predicted order of proton acidities is : $A > G > T \; U > C$. If these gas-phase calculations may be assumed to be applicable to the solvent phase, then the order of basic pK_a values expected from these results would be : $A < G < T \; U < C$. The order of aqueous phase pK_a values is, however, given by $A < G < T < U < C$, so that there is poor agreement between these gas-phase calculations and experimental results. Once again, use of gas-phase proton affinities for the experimental data set might lead to better correlations, but there is no data available. Calculations incorporating models for the solvent phase might also lead to to better correlations with the pK_a data set. So it once again emerges that these gas-phase AM1 calculations fare well for identification of proton ionisation sites, but faring less well in predicting quantitatively the ordering for solvent phase pK_a values.

III.5 Protic Reactions in Alkylated DNA Bases and Nucleosides

The somatic mutation theory of cancer points to the vital role played by critical alterations of the genome of somatic cells in the initiation and maintenance of the cancerous state of the cell (Ts'O 1980; Lowdin 1977 and Miller and Miller 1981). It is through this theory that a link may be established between DNA alkylation and carcinogenesis. Point mutation of the base-substitutional type has been well-documentated as one of the mechanisms by which latent proto-oncogenes of the ras and neu families are transformed to their carcinogenically active forms (Varmus 1984 and Bargmann *et al* 1986). It has been shown, as will be seen in the next section, that the abstraction or absence of certain Watson-Crick protons from their usual positions in modified guanines and thymines could furnish as plausible molecular basis for the adoption of mutagenic and pro-carcinogenic properties.

III.5.1 Role of proton abstraction for genotoxicity

About 16 sites for DNA modification by alkylating carcinogens have been discovered (Singer 1975; Pegg 1977), occurring at nucleophilic sites on the DNA base moiety, the deoxyribose ring and the phosphate backbone. Of these, the products of alkylation at the O^6-G site and the O^4-T site have been fairly conclusively shown by *in vitro* and *in vivo* studies to have the potential for inducing base misincorporation (Abbott & Saffhill 1979; Bhanot & Ray 1986; Singer *et al* 1978; Singer *et al* 1986; Preston *et al* 1986). The carcinogenic role of these products is also inferred from studies pointing to correlations between tumour incidence and the persistence of these alkylation products in the tissues of rodents treated with alkylating carcinogens (Goth & Rajewsky 1974; Margison & Kleihues 1975; Kleihues *et al* 1979; Cox & Irving 1977). This is in contrast to the evidence pointing to no significant promutagenic or procarcinogenic role for N^7- alkylguanines (Schoental 1969; Ludlum 1970).

Since not all alkylated DNA constituents are relevant to mutagenesis or cancer, some key differentiating element would be expected to play the role of demarcating the promutagenically alkylated DNA constituents from their nonmutagenic counterparts. At first sight, it would seem most probable that the adoption of mutagenicity by a modified DNA nucleoside should be attributable to the direct steric effect of the alkyl group which would block normal H-bonding. However, it may be argued that this would also block even unusual base-pairing in the double-helical configuration, as indicated by the theoretical study of Pohorille and Loew (1985), if the alkyl group is syn to the H-bonding side. If the alkyl group is in the anti position though, unusual base- pairing may be anticipated in the double-helical configuration provided the Watson-Crick protons are

displaced in such a manner as to simulate the tautomeric situation. The theoretical study of Duncan and Davies (Duncan & Davies 1989) indicated that the actual key molecular event which bestows the mutation-inducing property would be the loss of the Watson-Crick protons (the N^1-G proton for guanines and the N^3-T proton for thymines) from the modified DNA base moiety.

However, further study is due here since, in principle, the mutation-inducing process might well involve protons other than the ones studied by Duncan and Davies. Furthermore, the protic change could well be simple tautomerism apart from proton abstraction. Lastly, the gas-phase results of Duncan and Davies need to be supplemented by solvent-phase simulations in order to properly represent the situation in biological media.

These Watson-Crick protons may become absent from their usual positions in two ways, viz. by tautomeric shifts or by proton abstraction. The aims of the present study may thus be delineated as follows: (a) to examine for the possibility of proton shifts occurring in modified DNA bases leading to formation of tautomers, (b) to examine for the possibility of protons being abstracted from the alkylated DNA bases and nucleosides, and (c) to examine what situation might be likely to prevail at biological pH, particularly in the context of mutagenicity.

Fig. III.3 portrays the numbering system for methylated guanine and methylated thymine, where the Watson-Crick protons are the N^1-G and N^2-G protons for guanines, while for thymines it is the N^3-T proton. As shown in Fig. III.4, loss of these protons would not only defacilitiate normal base-pairing, but also pave the way for unusual base-mispairing schemes, where modified guanine mispairs with thymine, and modified thymine with guanine (shown in Fig. III.5). Calculated differences in Watson-Crick proton acidity had earlier been found to demarcate well between the promutagenic and the

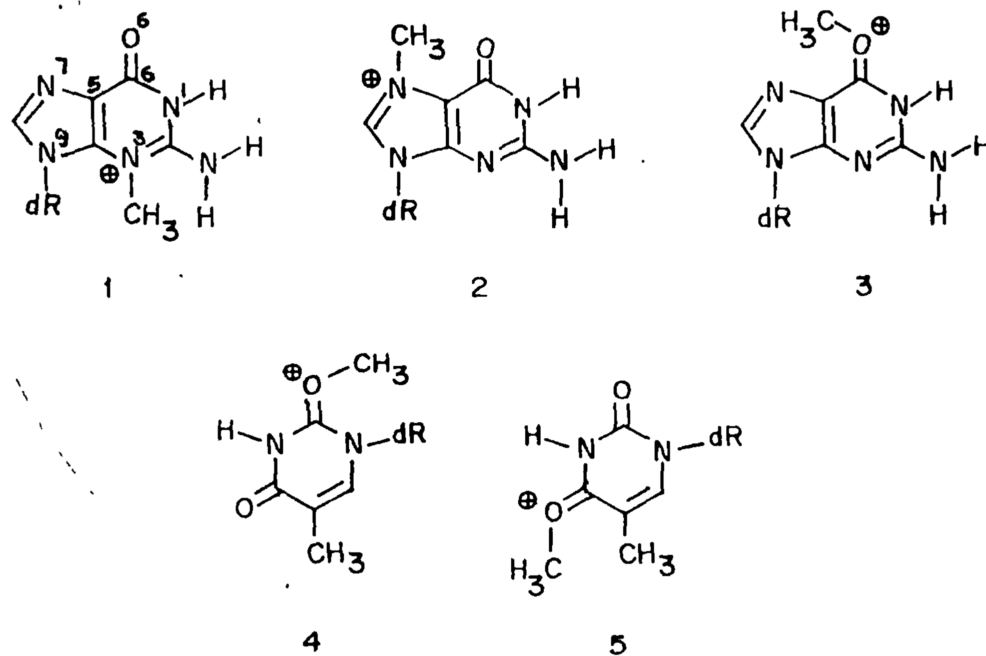
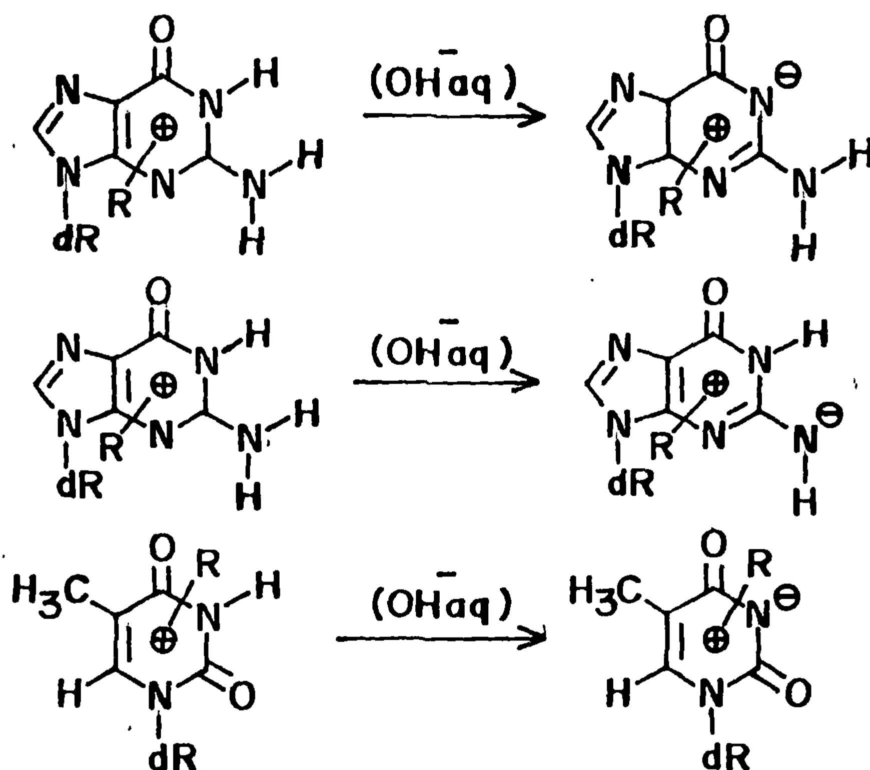


Fig. III.3 : Numbering system for the five methylated bases



dR = deoxyribose
R = Substitution at N^3, N^7, O^6 or O^2, O^4

Fig. III.4 : Reaction schemes showing the loss of Watson-Crick protons

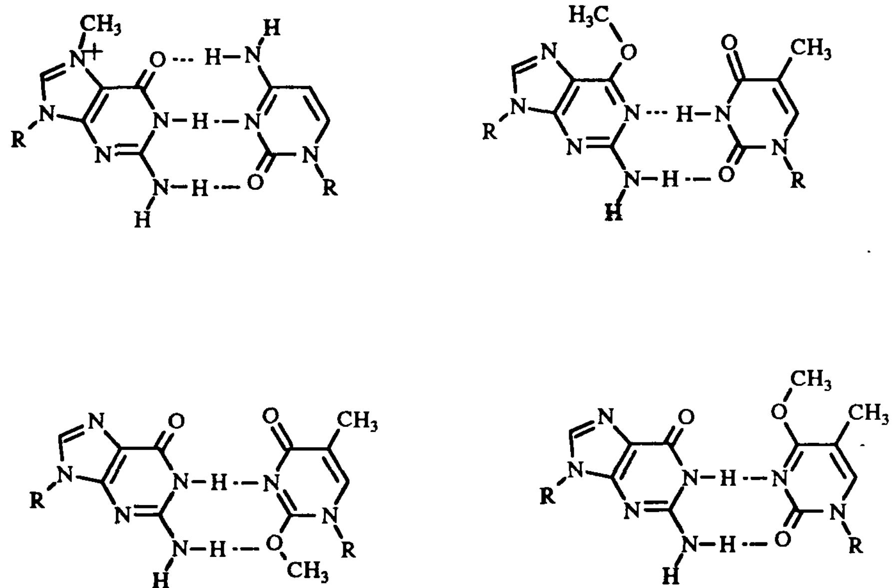


Fig. III.5 : Representation of some experimentally observed mismatches

nonmutagenic types of modified DNA bases (Duncan & Davies 1989).

III.5.2 Chemical changes following DNA alkylation

Formally, the alkylation of DNA nucleosides would involve the initial formation of a cationic species, representable as $(\mathbf{RdB})^+$ where **R** is an alkyl group and **dB** the deoxyribonucleoside moiety. What follows next could involve a variety of possibilities, including loss of alkyl group, loss of the base moiety (depurination or depyrimidination), shift or loss of protons, opening of the imidazole ring for N-alkylated purines, or simply no change at all.

This study assumes that a key factor for the manifestation of miscoding properties in the DNA bases guanine and thymine or their derivatives is the absence of the Watson-Crick protons from their usual positions (Duncan & Davies 1989). This could conceivably occur in two ways, viz. by intramolecular shift of the proton to some other site on the base, or by loss of the proton from the system. Hence, three alternatives are chosen for study : (1) the Watson-Crick proton shifts to some other site on the DNA base leading to formation of a tautomer, which may be expected to possess promutagenic properties, (2) the Watson-Crick proton is abstracted from the system, giving a species capable of inducing mutations, and (3) the system stays as it is. These processes are dependent, besides other factors, upon the pH of the surroundings, which factor is considered empirically, both at gas-phase and solvent-phase, in the course of this study.

The following alkylated DNA bases with their deoxyribonucleosides are incorporated initially for study here (see Fig. III.3) : N^3-methylguanine (N^3-MeG), N^7-methylguanine (N^7- MeG), O^6-methylguanine (O^6-MeG), O^2-methylthymine (O^2-MeT) and O^4-methylthymine (O^4-MeT). The deoxyribonucleosides are abbreviated here as for the

bases, substituting dG and dT for G and T respectively. N^3-MeG has been isolated as a minor product of *in vivo* DNA alkylation by carcinogens, but its promutagenic potential has not been experimentally elucidated. N^7-MeG, the chief alkylation product in most cases (Singer 1975; Pegg 1977), is characterised by non-miscoding properties (Schoental 1969; Ludlum 1970). O^6-MeG is evidently promutagenic (Abbott & Saffhill 1979; Bhanot & Ray 1986) and so are O^2-MeT and O^4-MeT (Singer *et al* 1986; Preston *et al* 1986).

III.5.3 Theoretical indices for proton transfer possibilities .

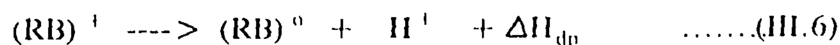
(a) Indices for tautomer formation

For tautomer formation, the feasibility of proton shift is equated to the thermodynamic stabilities ΔH_s and ΔG_s of the tautomer resulting from the shift, respectively derived from the AM1 enthalpies and free energies of formation ΔH_f and ΔG_f of the various tautomers, being calculated relative to the most stable tautomer in each case. Since the proton shift process studied here is intramolecular, external effects such as solvation may be expected to play a secondary role here (which is not true for the proton abstraction process). This means that a greater degree of confidence may be placed on the gas-phase values of the ΔH_s and ΔG_s indices to furnish predictions of the feasibility of the tautomer formation processes.

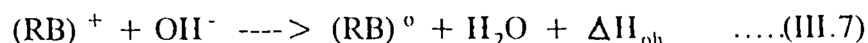
(b) Indices for proton abstraction

The proneness of a proton to abstraction is represented by the following indices : Q_h (the Mulliken positive charge on the proton of interest), W_{xh} (the Wiberg index for bond strength between the proton **H** and the nucleophilic atom **X** to which it is attached), and the enthalpy of abstraction of the Watson-Crick proton, calculated in terms of the gas-phase

proton abstraction enthalpy ΔH_{dp} as given in Eqn. (III.6), and in terms of the enthalpy ΔH_{oh} for deprotonation by a hydroxide anion as in Eqn. (III.7) :



$$(RB)^{+} \text{----}> (RB)^{o} + H^{+} + \Delta H_{dp} \quad \text{......(III.6)}$$



$$(RB)^{+} + OH^{-} \text{----}> (RB)^{o} + H_2O + \Delta H_{oh} \quad \text{....(III.7)}$$

where $(RB)^{+}$ and $(RB)^{o}$ respectively represent the cationic and deprotonated forms of the alkylated bases. For the deoxynucleoside systems, the same indices apply, only replacing (RB) by (RdB), the symbol for the alkylated deoxynucleoside.

III.5.4 Previous theoretical studies on DNA alkylation

Theoretical gas-phase studies on DNA base alkylation include the MNDO MO work of Ford and Scribner (1983; 1990) to explain O-selectivity trends, the *ab initio* MODPOT-/VRDDO/MERGE study of Hariharan *et al* (1979) on DNA base methylation, the MNDO MO work of Buda and Sygula (1983) on alkylated DNA bases, the INDO SCF-MO studies of Lyngdoh (1992, 1994a, 1994b) and of Duncan and Davies (1989a, 1989b, 1989c), the MINDO/3 MO study of Mohammed and Hopfinger (1980) on attack of DNA bases by the methanediazonium ion, the perturbational base-pairing studies of Psoda *et al* (1981) and of Pohorille and Loew (1985), the AM1 SCF-MO study of Ford and Wang (1992a, 1992b) on the possible mutagenic effects of protonation and alkylation upon G-C pairs, and the base-pairing studies of Venkateswarlu and Lyngdoh (1995). While trends concerning the product distribution of DNA alkylation and the general mutagenic effects of DNA alkylation have been worked on, it appears that further study needs to be devoted to the key molecular basis for the adoption of mutagenicity by some alkylated DNA

components and not by others. The two alternatives of tautomer formation and proton abstraction for alkylated DNA bases are dealt with as below :

III.5.5 Possibilities for tautomerism

The possibility of tautomerism involves shift of the Watson-Crick proton to another position on the DNA base moiety. This could result in a tautomer of possible mutagenic significance, since the Watson-Crick protons are not present in their usual positions and are unavailable for normal base-pairing. For the guanines, the proton could conceivably be variously attached to the N^1-G site (the normal site in free guanine) or to the N^2-G, N^3-G, N^7-G or O^6-G sites. For thymines, the alternatives open are at the N^3-T site (normal in free thymine) or the O^2-T and O^4-T sites.

Table III.10 gives the AM1 and MNDO SCF-MO enthalpies of formation of each tautomer for the various alkylated DNA bases, along with the relative stability of each species with respect to the most stable tautomer in each case as indicated by ΔH_s. Both sets indices furnish virtually identical predictions. However, whenever any variation is noted, greater weightage is laid upon the more recent AM1 method.

These gas-phase data for N^7-MeG and O^6-MeG predict that the initially formed cationic alkylated product (with the Watson- Crick protons in their usual position) would be the stablest among the various tautomers considered. These results would exclude the significant role of intramolecular proton shift for the N^7-MeG, O^6-MeG, O^2-MeT and O^4-MeT systems. However, N^3-MeG exhibits differences here in that the tautomer with the proton shifted to the O^6-G site is by far the stablest. The stability orderings for the various alkylated bases are as follows, indicating each tautomer by the position of the proton under consideration :

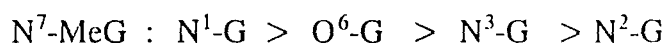
Table III.10 AM1 values for enthalpies ΔH_f of formation of alkylated guanines, thymines and tautomers, including their stabilities ΔH_s with respect to the most stable tautomer in each case (all values in kcal/mol)

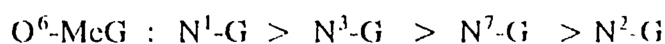
Species	AM1		MNDO	
	ΔH_f	ΔH_s	ΔH_f	ΔH_s
N^7-MeG				
N^1-G	194.02	0.00	161.48	0.00
N^2-G	224.71	30.69	190.83	29.33
N^3-G	210.96	16.94	179.52	18.04
O^6-G	208.38	14.37	165.18	3.70
N^3-MeG				
N^1-G	206.96	3.39	179.08	13.35
N^2-G	242.44	38.87	211.06	46.33
N^7-G	214.29	10.72	185.97	21.24
O^6-G	203.57	0.00	164.73	0.00
O^6-MeG				
N^1-G	202.17	0.00	165.69	0.00
N^2-G	220.51	18.34	181.39	15.70
N^3-G	206.93	4.76	169.65	4.26
N^7-G	211.57	9.41	174.19	8.50
O^2-MeT				
N^3-T	108.36	0.00	100.38	10.90
O^4-T	111.16	2.81	89.48	0.00
O^4-MeT				
N^3-T	104.64	0.00	98.28	6.80
O^2-T	114.39	10.76	91.48	0.00

Table III.11 Proton acidities in alkylated guanines and thymines, considering the stablest tautomers in each case (enthalpies in kcal/mol)[a]

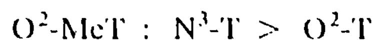
Proton Site	AM1			
	$\Delta H_{dp}(p)$	$\Delta H_{dp}(w)$	Q_h	W_{xh}
N^7-MeG				
N^1-G	247.72	-163.10	0.2767	0.8707
N^2-G	253.39	-157.43	0.2372	0.8876
O^6-MeG				
N^1-G	225.03	-185.79	0.2927	0.8648
N^2-G	242.97	-167.85	0.2485	0.8832
N^3-MeG				
O^6-G	229.14	-181.68	0.2880	0.8970
N^2-G	235.43	-175.39	0.2981	0.8759
O^2-MeT				
N^3-T	222.25	-188.57	0.2917	0.8602
O^4-MeT				
N^3-T	218.42	-192.40	0.3004	0.8576

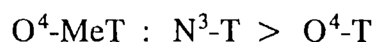
[a] Heat of formation of H^+ is taken from experimental value H^+ = 365.7 Kcal/mol


N^7-MeG : N^1-G > O^6-G > N^3-G > N^2-G


O^6-MeG : N^1-G > N^3-G > N^7-G > N^2-G

N^3-MeG : O^6-G > N^1-G > N^7-G > N^2-G


O^2-MeT : N^3-T > O^2-T


O^4-MeT : N^3-T > O^4-T

These results predict that, apart from N^3-MeG, the shift of the Watson-Crick protons to some other nucleophilic site is not likely. So for 4 of the above 5 systems, intra-molecular proton shift is not likely to be instrumental for the transformation of the modified DNA base to a promutagenic species with the Watson- Crick protons absent from their usual positions. For N^3-MeG, the stable O^6-protonated tautomer predicted could in itself be capable of inducing mutations because of the absence of the N^1-G proton, but it remains to be seen whether it remains in its cationic form or not, owing to to possibility of proton loss.

III.5.6 Possibilities for proton loss

The second possibility is that a proton may be abstracted from the cationic alkylated species initially formed to yield a neutral deprotonated species. The only systems considered for this are the most stable tautomers established earlier, since they would be expected to predominate before the deprotonation takes place by virtue of their greater stability. For the thymine systems, only one proton (the N^3-T proton) is available for deprotonation. For the guanine systems, the alternatives open are the N^1-G and N^2-G protons (for the N^7-MeG and O^6-MeG systems) and the O^6-G and N^2-G protons (for the N^3-MeG system), giving rise to the question as to which is most likely to be abstracted. For all the systems

considered, another question is at which pH this proton loss is likely to be effected, which is dealt with empirically in the next section

Table III 11 presents the AM1 values of the indices for proton acidity, viz. the Q_h, W_{xh},$\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ indices, each calculated for the stablest tautomer of each system as established from the calculations described earlier This data shows that for all the guanine systems, the N^2-G proton is invariably the less acidic one, marked by a higher enthalpy of deprotonation ΔH_{dp}, a lower Mulliken positive charge Q_h, and a lower bond strength index W_{xh} For N^7-MeG, and O^6-MeG, the N^1-G proton is the more acidic one, while for the stable tautomer of N^3-MeG, the O^6-G proton is the more acidic one Remembering however that the O^6-G proton of N^3-MeG is ultimately derived from the N^1-G proton, the overall loss of proton in all cases is thus predicted to involve the N^1-G proton, and not the N^2-G proton. For the thymine systems, there is no ambiguity, since the only proton considered is the N^3-T proton

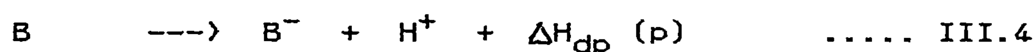
Determination of the relative ordering for gas-phase proton acidities with respect to the system may be done from the calculated values of $\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ The values give the following order for deprotonation facility of the base systems, considering only the more acidic protons for the guanine systems O^4 MeT > O^2 MeT > O^6 MeG > N^3-MeG > N^7-MeG. As this data pertains only to the solitary base systems, and do not give an adequate simulation of DNA environment, further work was done by taking recourse to the deoxynucleoside systems, as will be seen in the following Section

III.5.7 Proton acidities for nucleoside systems

The data of Table III.12 corresponds to the data of Table III 11, giving the complete deoxyribonucleoside instead of the solitary base Here, for the deoxyguanosine systems,

Table III.12 AM1 values of indices for Watson-Crick proton acidities of alkylated ribo- and deoxyribonucleosides (enthalpies in kcal/mole)

Species	pK_a	ΔH_{dp}(p)	ΔH_{oh}(w)	Q_h	W_{xh}	Mut/Non
O^4-MedT	0.30	216.07	-194.75	0.3113	0.8523	Mut
O^4-EtdU	0.66	220.11	-190.71	0.3111	0.8526	Mut
O^4-MedU	0.70	218.61	-192.21	0.3126	0.8518	Mut
O^4-MerU	0.80	223.40	-187.42	0.2947	0.8606	Mut
O^2-EtdU	0.92	222.58	-188.24	0.2865	0.8623	Mut
O^2-MedT	1.00	222.39	-188.43	0.2875	0.8615	Mut
O^6-MedG	2.40	225.10	-185.72	0.2919	0.8651	Mut
O^6-EtrG	2.50	228.00	-182.82	0.2895	0.8662	Mut
N^3-MedG	-	239.17	-171.65	0.2846	0.8656	Mut
N^7-MerG	7.00	251.97	-158.85	0.2741	0.8718	Non
N^7-MedG	7.20	252.73	-158.09	0.2740	0.8719	Non
N^7-EtrG	7.30	253.30	-157.52	0.2730	0.8723	Non



$$B \longrightarrow B^- + H^+ + \Delta H_{dp}\ (p) \qquad \text{..... III.4}$$

$$B + OH^- \longrightarrow B^- + H_2O + \Delta H_{dp}\ (w) \qquad \text{..... III.5}$$

where B represents the nucleoside and B^- represents the corresponding deprotonated counterpart.

Experimental gas-phase proton enthalpy value of 365.7 Kcal/mol is used for H^+ in III.4.

focus is laid only on the more acidic proton (as predicted by the data of **Table III.11**), and not on the N^2-G proton. The possibility for a good positive correlation between theoretical gas-phase and experimental solvent-phase proton acidities was further explored here by calculating the gas-phase deprotonation enthalpies for a fair number of alkylated deoxynucleoside systems whose pK_a values are known. Apart from the five systems already considered, the list of 12 alkylated nucleosides considered here includes the following : O^4-EtdU , O^4-MedU , O^4-MeU , O^2-EtdU , O^6-EtG , N^7- MeG, and N^7-EtG.

Fig. III.6 portrays the linear relation between experimental basic pK_a and the gas-phase AM1 enthalpies of deprotonation $\Delta H_{dp}(w)$ for these 12 alkylated nucleosides. Since reproduction of solution phase enthalpies is impossible by gas-phase calculations even at the *ab initio* level, what is sought for here is only the basic trends of correlation. Considering that the relation between the free energies of deprotonation and the corresponding pK_a values is in principle a linear one, it is remarkable that there is a good linear trend observed between the theoretically calculated AM1 values of deprotonation enthalpies and the experimental pK_a values, which may be described by the linear best fit

$$pK_a = M . \Delta H_{dp}(w) + C \qquad \text{......... (III.8)}$$

where M = 0.1959 and C = -42.33, the coefficient of linear regression having a convincing value of 0.9861. This fairly good fit between gas-phase and solvent-phase data is remarkable, and may be attributable to the similarity of type for the systems considered, all being alkylated nucleosides, for which solvation may be expected to exert parallel effects upon the various members. From this linear plot, a pK_a value of 2.25 is predicted for N^3-MedG, for which no experimental information is yet available.

The above linear correlation between AM1 calculated gas-phase deprotonation

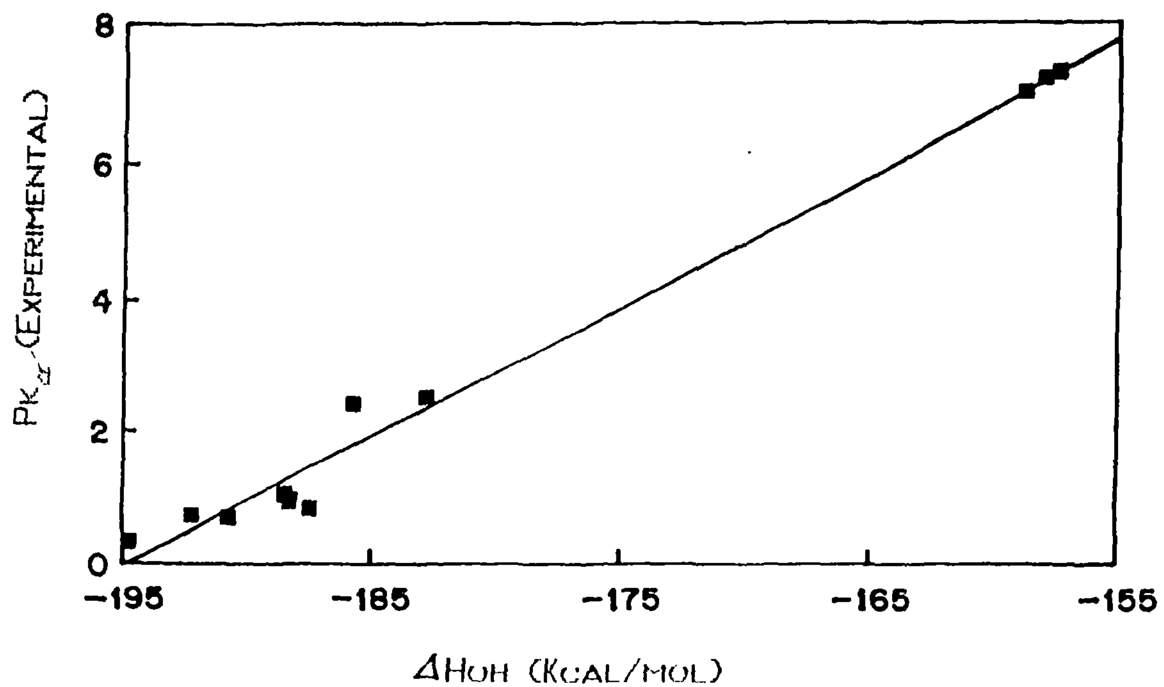


Fig. III.6 : Linear correlation plot between experimental pK_a values and AM1 calculated deprotonation enthalpies for various deoxyribonucleic acids .

enthalpies and experimental basic pK_a values is remarkably good for structurally similar alkylated nucleosides. A further effort was made to extend the study to reproduce acidic pK_a values as well, apart from the basic pK_a values just successfully studied, where the acidic data set refers formally to protonation reactions and the basic data set to deprotonation reactions. Both were incorporated into one framewrok, this being done by studying the process of protonation in terms of deprotonation in the opposite direction, viz. by simply reversing the direction of the protic acid-base equilibrium. The results were not encouraging when AM1 gas-phase enthalpies of deprotonation were plotted against the entire gamut of both acidic and basic aqueous phase pK_a values. This may be chiefly attributed to differences in the aqueous solvation effects arising from different chemical environments, one being acidic and one basic.

III.5.8 Solvent-phase proton acidities

While the above correlation of AM1 calculated gas-phase deprotonation enthalpies with experimental pK_a is remarkably good for structurally similar alkylated nucleosides, it is poor for structurally different systems and chemical environments. In order to evaluate the free energies associated with the full range of deprotonation reactions of various nucleic acid base systems (modified and unmodified), the aqueous solvation model **AM1-SM2** was employed. The performance of **AM1-SM2** model along with other **SMx** models has been reviewed recently (Cramer & Truhlar 1992; Tomasi & Persico, 1994) and was shown to furnish good linear correlations with experimental free energies of solvation. Full geometry optimisation using the AM1-SM2 solvation methodology is quite compute-intensive, despite the improvement of convergence strategies as reported by Cramer and Truhlar. The computation of free energies of the various deprotonation reactions for

nucleic acid systems was therefore restricted to the base level only.

Table III.13 presents the protic acid base equilibria, the experimental pK_a values, the corresponding deprotonation enthalpies $\Delta H_{dp}(w)$ (see Eqn III.8) computed at the gas-phase AM1 level, and the enthalpies of deprotonation $\Delta H_{dp}(aq)$ calculated at the simulated solvent-phase **AM1-SM12** level of theory for 10 nucleic acid base systems. This set includes the deprotonation reactions for the four unmodified DNA base systems and six alkylated base systems, spanning the whole range of pK_a values (acidic and basic). Figure III.7 illustrates the linear correlation plot of AM1-SM12 Predicted pK_a-values against experimental pK_a values. This plot reveals a good linear trend for the full spectrum of protic acid-base reactions in nucleic acid base systems. Fig. III.8 illustrates the linear correlation between the experimental pK_a values and AM1 gas-phase values, showing the poor regression for the same data set represented in Fig. III.7. The best linear fit for the plot shown in Fig.III.7 may be described as follows :

$$pK_a = M.\ H_{dp}(aq) + C \qquad \ldots\ (III\ 9)$$

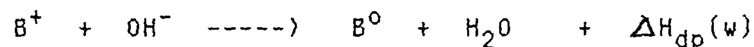
which gave a correlation coefficient $R_c = 0.9846$, M being -0.3131 and C = -46.65. This plot also predicts a pK_a value of 4.30 for the experimentally unknown case of N^3-MeG.

Finally, **Table III.14** sums up the findings of this Chapter concerning the protic acid-base equilibria of the five protonated nucleic acid bases, three alkylated DNA bases, and the five normal nucleic acid bases, giving the AM1 values of $\Delta H_{dp}(w)$, the experimental pK_a values, the experimental assignment of ionisation sites, and the theoretically predicted assignments obtained here. There is some linear progression in both the experimental and theoretical data sets pertaining to proton acidity, which speaks for the fair ability of the theoretical methods used here to reproduce trends at large. The use of the solvent-phase

Table III.13. Gas-phase and solvent-phase enthalpies of deprotonation of nucleic acid systems and the corresponding experimental pK_a values. along with theoretically predicted pK_a values (enthalphies in Kcal/mol)

Species	proton involved	pK_a (Exptl.)	$\Delta H_{dp(w)}$ [a]		pK_a (predicted)	
			AM1	AM1-SM2	AM1	AM1-SM2
O^4MeT	N^3	0.80	192.39	46.25	3.10	2.08
O^2MeT	N^3	1.40	196.22	49.33	3.33	1.13
O^6MeG	N^1	2.80	199.00	43.47	3.50	2.94
N^1HA	N^1	4.20	194.67	40.68	3.24	3.81
N^3MG	N^1	?	201.40	38.52	3.64	4.48
N^3HC	N^3	4.25	188.43	35.57	2.87	5.39
N^7MeG	N^1	7.20	221.69	30.76	4.86	6.88
N^3U	N^3	9.51	318.45	22.48	10.67	9.45
N^1MeU	N^3	9.70	318.38	24.24	10.66	8.91
N^3T	N^3	9.90	318.06	22.16	10.64	9.55
N^4C	N^4	12.16	312.86	10.73	10.33	13.09

[a] enthalpies of deprotonation (ΔH_{dp}) are computed according to the following equation



$$B^+ + OH^- \longrightarrow B^0 + H_2O + \Delta H_{dp}(w)$$

In AM1 gas-phase calculations, experimental proton (H^+) value 365.7 kcal/mol is considered.

Table III.13 Gas-phase and solvent-phase enthalpies of protonation of nucleic acid base systems and the corresponding experimental pK_a values along with theoretically predicted pK_a values (Enthalphies in Kcal/mol)

Reaction	pK_a(exptl.)	AM1*	pK_a(SM2)	AM1-SM2**
$O^{4}MeT^{0}$ + H^{+} --> $O^{4}MeT^{+}$	0.80	-192.40	1.88	-155.00
$O^{2}MeT^{0}$ + H^{+} --> $O^{2}MeT^{+}$	1.40	-188.57	0.92	-151.91
$O^{6}MeG^{0}$ + H^{+} --> $O^{6}MeG^{+}$	2.80	-185.79	2.76	-157.80
A + H^{+} --> $N^{1}HA^{+}$	4.20	-190.12	3.64	-160.59
$N^{3}MG^{0}$ + H^{+} --> $N^{3}MG^{+}$	?	-183.39	4.30	-162.32
C + H^{+} --> $N^{3}HC^{+}$	4.25	-196.36	5.23	-165.67
$N^{7}MeG^{0}$ + H^{+} --> $N^{7}MeG^{+}$	7.20	-163.10	6.74	-170.49
$N^{3}U^{-}$ + H^{+} --> U	9.51	-66.34	9.39	-178.95
$N^{1}MeU^{-}$ + H^{+} --> U	9.70	-66.41	8.89	-177.40
$N^{3}T^{-}$ + H^{+} --> T	9.90	-66.73	9:47	-179.21
$N^{4}C^{-}$ + H^{+} --> C	12.16	-71.93	13.00	-190.51

* In AM1 gas-phase calculations H^{+}(exptl.) = 365.7 Kcal/mol is used

** In AM1-SM2 calculations, H^{+} (solv.) = 258.0 kcal/mol is used

Table III.14 Theoretical Data for relative acidity of most acidic protons of nucleic acid bases compared with experimental pK_a values and their assignments ($\Delta H_{dp}(w)$ values in Kcal/mol)

System	$\Delta H_{dp}(w)$	pK_a	Exptl. assignment	Theoretically predicted
HT^+	201.97	-	?	O^4
HU^+	203.59	-3.38	?	O^4
HC^+	214.46	4.25	N^3	N^3
HA^+	220.70	3.77	N^3	N^3
HG^+	222.08	2.27	N^7	N^7
O^4-MeT^+	218.42	0.8	N^3	N^3
O^2-MeT^+	222.25	1.4	N^3	N^3
O^6-MeG^+	225.03	2.8	N^1	N^1
N^7-MeG^+	247.72	7.2	N^1	N^1
G	297.91	9.42	N^1	N^1
A	332.88	9.26	N^9	N^9
T	344.09	9.90	N^3	N^3
U	344.48	9.51	N^3	N^3
C	338.89	12.16	N^1/C^2O	N^4

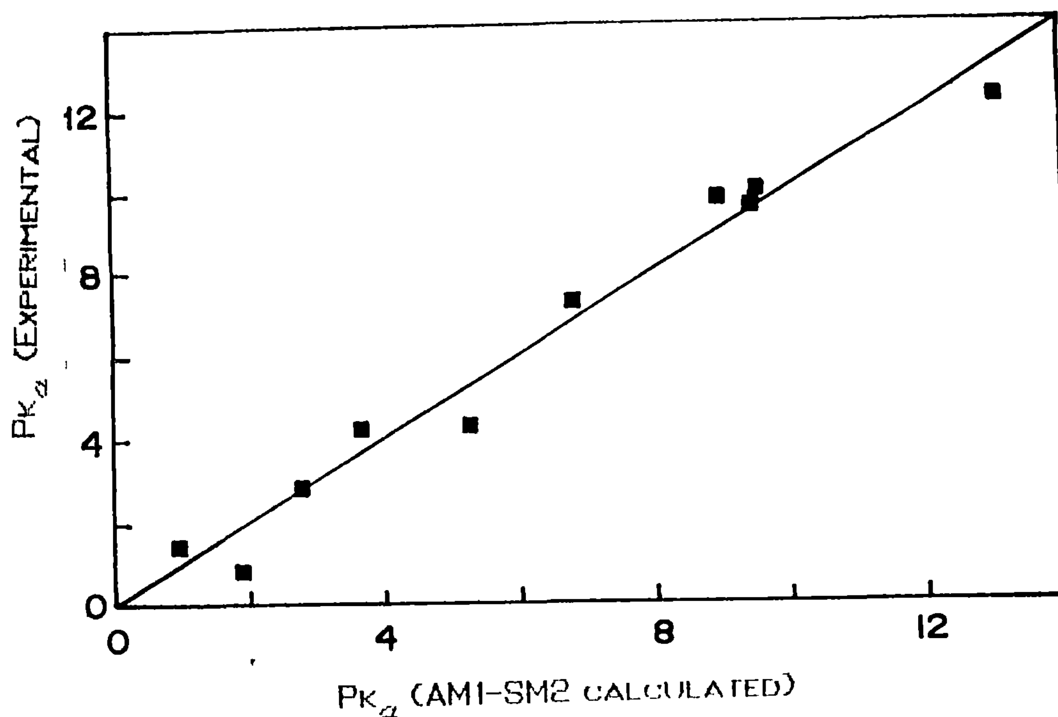


Fig. III.7 : Linear correlation plot between experimental pK_a values and AM1-SM2 computed pKa values

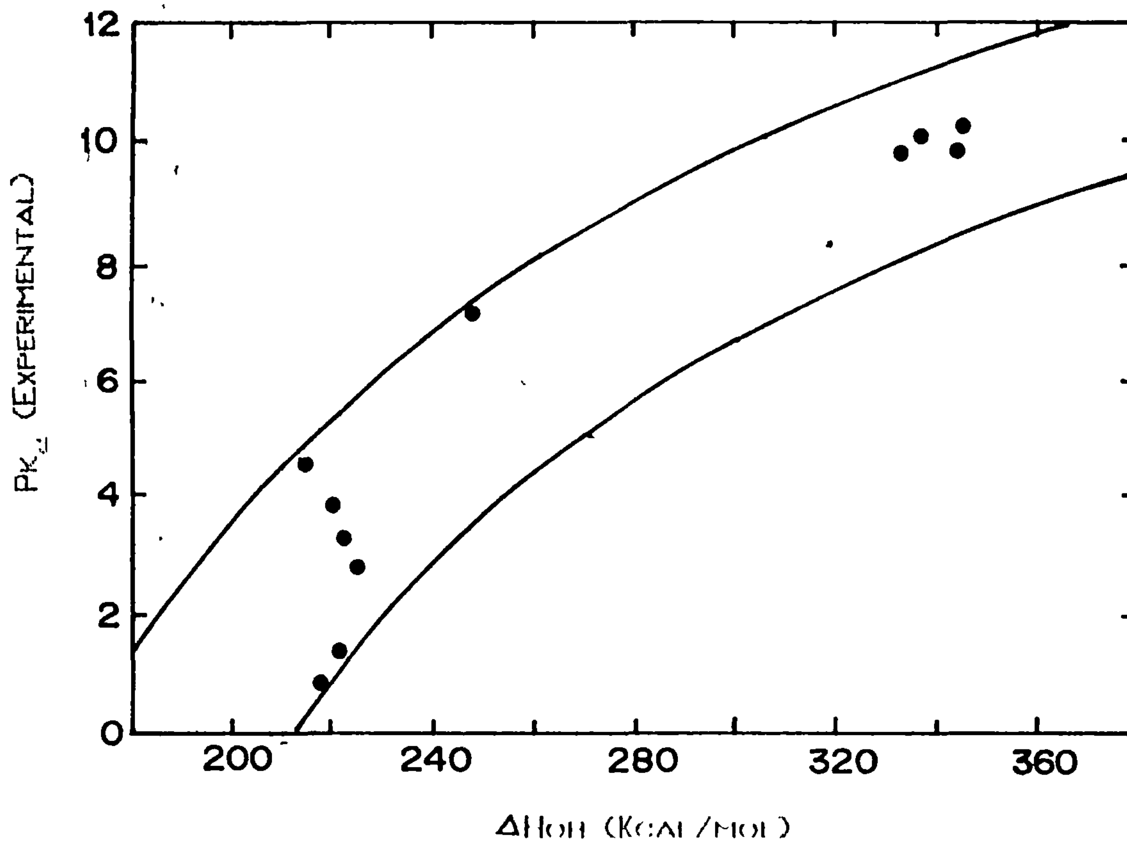


Fig. III.8 : Observed poor linear correlation between experimental pK_a values and AM1 gas-phase values for the same data set, shown in Fig. III.7

model undoubtedly improves the situation here. There is also a good correspondence between theoretical and experimental assignment of proton ionisation sites, which again speaks in favour of the theoretical method used.

III.5.8 Predictionof mutagenicity

Since for alkylated guanine and thymine nucleosides, the most acidic protons are here predicted to be the Watson-Crick ones, their acidity would be a reflection on the mutagenicity or non- mutagenicity of a particular alkylated system. Thus, predictions may be made here on the basis of pK_a value as to whether a given system would be mutagenic or non-mutagenic. For the ribonucleoside systems, the definition of mutagenic would have to refer to the ability to pair with the wrong ribonucleoside, causing errors in transcription (and not in DNA synthesis as would be the case for the deoxyribonucleoside systems). Assuming a biological pH value of about 7.0, it may be inferred that the systems with pK_a values above this value would exist predominantly in the undeprotonated form with the Watson-Crick protons present, and thus not be mutagenic, while those systems with pK_a values below 7.0 would be expected to exist in the deprotonated form predominantly, and thus be mutagenic. The classification into mutagenic (Mut) and nonmutagenic (Non) species is given in Table III.12 too, indicating that N^7-methyldeoxyguanosine would be non-mutagenic, together with N^7-methyl- and ethylguanosines. All the other alkylated nucleosides are predicted to be mutagenic, which is indeed the case. The yet untested case of N^3-methyldeoxyguanosine is also predicted to be mutagenic as indicated by its predicted pK_a value of 2.25 on the basis of gas-phase data, while the base is predicted to have a pK_a value of 4.30 on the basis of solvent-phase data. These predictions await experimental verification.

III.6 Concluding Remarks

The findings of the above studies may be summarised as follows :

(1) The most stable form in which the 5 nucleic acid bases are predicted to exist are the ones commonly known, while their most stable tautomeric counterparts involve shift of the hydrogen- bonding protons to adjacent basic sites.

(2) The sites predicted to be most favourable for protonation of the neutral species are the N^7-guanine, N^3-adenine(close to N^1), N^3-cytosine and the O^4 sites of thymine and uracil, which is corroborated to an extent by the findings of experiment. The sites predicted as most favourable for deprotonation in the neutral species are the N^1-guanine, N^9-adenine, N^1-cytosine and the N^3 sites of thymine and uracil, which compares well with the assignments of experiment.

(3) While these gas-phase AM1 calculations perform well insofar as identification of site for tautomerism, protonation and deprotonation goes, they do not fare as well in furnishing quantitative predictions for the relative energetics of the processes in solvent phase.

While the above findings are related to nucleic acid bases, the following inferences may be drawn for the protic changes expected in alkylated DNA bases and nucleosides, including the significance of these for mutagenicity :

(4) Formation of stable tautomers in significant amounts is not predicted for any of the systems except for N3-methylguanine, so that tautomer formation in itself may be largely excluded as a likely route for the formation of promutagenic and potentially miscoding alkylated DNA bases.

(5) Formation of potentially mutagenic species is predicted to involve loss of the N^1-G proton for the deoxyguanosine systems and the N^3-proton for the deoxythymidine systems, which protons have acidities predicted in the order O^4-MedT > O^2-MedT > O^6- MedT > N^3-MedG > N^7-MedG. This ordering is well-substantiated by the experimentally obtained pK_a values of the nucleosides.

(6) The theoretically calculated gas-phase AM1 values for enthalpy of proton loss correlate linearly with the experimental basic pK_a values for a large number of systems, from which plot a pK_a value of 2.25 is predicted for the untested case of N^3- methyl-deoxyguanosine.

(7) Theoretically calculated AM1-SM2 solvent-phase values for enthalpy of proton loss also correlates linearly with both acidic and basic experimental pK_a values for a large number of base systems, from which a pK_a value of 4.30 is predicted for the untested case of N^3-methylguanine.

(8) These gas-phase and solvent-phase calculated proton acidities corroborate the experimentally observed mutagenicities of all the O-alkylated base systems studied here, and the non-mutagenicity of N^7-alkylguanine systems.

(9) The N^3-methylguanine systems are predicted to be mutagenic on the basis of the pK_a values of 4.30 and 2.25 predicted for the base and the deoxynucleoside respectively.

(10) This study thus serves to substantiate the role of Watson- Crick proton shift as furnishing the key molecular basis for conferring of mutagenic properties; this through tautomerism for the free nucleic acid bases, and through deprotonation for the alkylated guanine and thymine systems.

CHAPTER IV

PAIRING PROPERTIES OF MODIFIED DNA BASES

IV.1 Biological Significance of Base Misincorporations

Molecular mistakes during DNA replication are inherent and may be essential to the human genome. The human genome has about 3 x 10^9 base pairs, so inevitably a few mistakes will be made during copying, most of which are potentially deleterous. Themolecular mechanism of copying parent DNA into daughter DNA, elucidated by Watson and Crick some 42 years ago, consists of very specific hydrogen bonding (characteristic for complementary strands) between purine and pyrimidine bases, viz. guanine to cytosine and thymine to adenine (Watson and Crick 1953). However, the degree of fidelity achieved by the association of Watson-Crick base pairs is not sufficient to maintain genetic integrity. There is a competition between the correct GC/AT base pairs and eight possible base mispairs - AA, GG, AG, CC, TT, CT, AC and GT. The first three are purine-purine mismatches, the second three are pyrimidine-pyrimidine mismatches and the last two are purine- pyrimidine mismatches. Fig. IV.1 gives various possible mispairs among DNA bases. After a cycle of replication, purine-purine and pyrimidine-pyrimidine mismatches give rise to transversion mutations, while purine pyrimidine mismatches produce transition mutations. A complex array of proofreading and repair enzymes patrols the new DNA duplex as it formed, seeking out mismatches and correcting them. These and

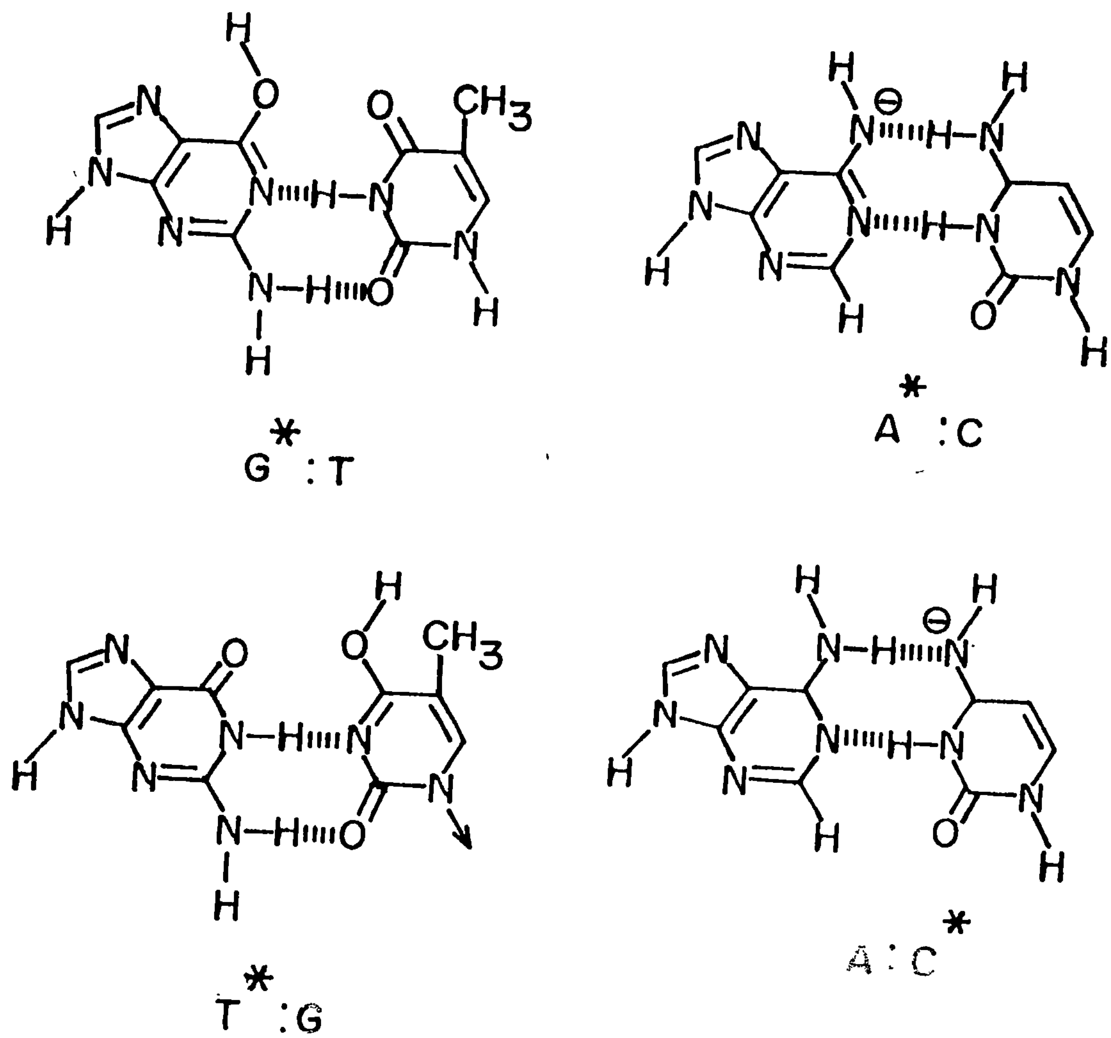


Fig. IV.1 : Various possible mispairs among DNA bases

related enzymes also repair chemical damage caused by cellular and environmental mutagens, UV and ionizing radiation. Repair enzyme systems, which are still not well understood, are remarkably efficient. They reduce the error rate during replication from 1 in 10^2 to 1 in 10^9.

In certain special contexts non-Watson-Crick base associations such as Wobble base pairs, Hoogstein and reverse Hoogstein type base pairs are possible and could play important role within the cell. Such structures occur in tRNAs (Saenger 1984), DNA triplexes (Hoogsteen 1963; Maher *et al* 1989), and in telomeres - the structures that terminate the wing of a chromosome (Kang et al 1992).

Besides these self-associating mismatches, DNA bases modified by xenobiotic agents such as alkylating agents cause certain specific types of mismatches with non-complementary bases, leading to mutational events (Singer & Grunberger 1983). While most of these mismatches are repaired by appropriate enzymes, some of them manage to escape the vigil. Some types of these mismatches closely resemble the typical Watson-Crick base pairs in their configurational aspects, thus escaping the proofreading enzymes. This can cause permanent mutations which may lead to catastrophic events such as the death of the cell, a genetic disease, or a carcinogenic lesion.

The specific choice of DNA repair by enzymes may possess some very specific structural origin, which speculation has prompted many experimental and structural research groups to attempt understanding the structure and thermodynamic stability of a variety of Watson-Crick and non-Watson-Crick base pairs both at the monomer and oligomer levels.

In this chapter the semi-empirical PM3 SCF-MO method is used to study and probe the base-pairing properties of altered DNA bases through examining the characteristics and

stability of these mismatches with respect to normal Watson-Crick base pairs, and thus possibly to get some insight into the physical basis of their ability to induce mutations in DNA. All could in turn lead one to arrive at some conclusions about their significance in terms of the integrity of the genetic code.

IV.2. Theoretical Methodology

Each putative base-pairing situation was studied by incorporating the two concerned base species together as a hydrogen-bonded complex, whose configuration generally initially approximated to the Watson-Crick double-helical one found in B-DNA. The entire complex was subjected to a single SCF-MO calculation strategy, which is described as the supermolecule approach. The initial configuration was then submitted to full geometry optimisation in order to arrive at the equilibrium situation appropriate to the method used. The PM3 SCF-MO method has been substantiated to be the only semiempirical methodology (using the NDDO scheme) with any ability to properly reproduce experimentally observed hydrogen bonding between nucleotide base pairs (Lively *et al* 1994). This observation is very important in the context of this work, since a large variety of semiempirical methodologies have been in current use.

Two strategies were adopted for optimisation of the base pairs: one retaining C_s (mirror-plane) symmetry for the pair, and the other without any symmetry constraints. The use of mirror-plane symmetry arises out of the general expectation that the normal Watson-Crick base pairs in the double-helical situation usually are observed to adopt an approximately planar configuration embracing both the bases. The use of the other alternative of free optimisation without any constraints was done with a view in mind to check if any possibilities exist for locating an energy miminum for the base-pair in a

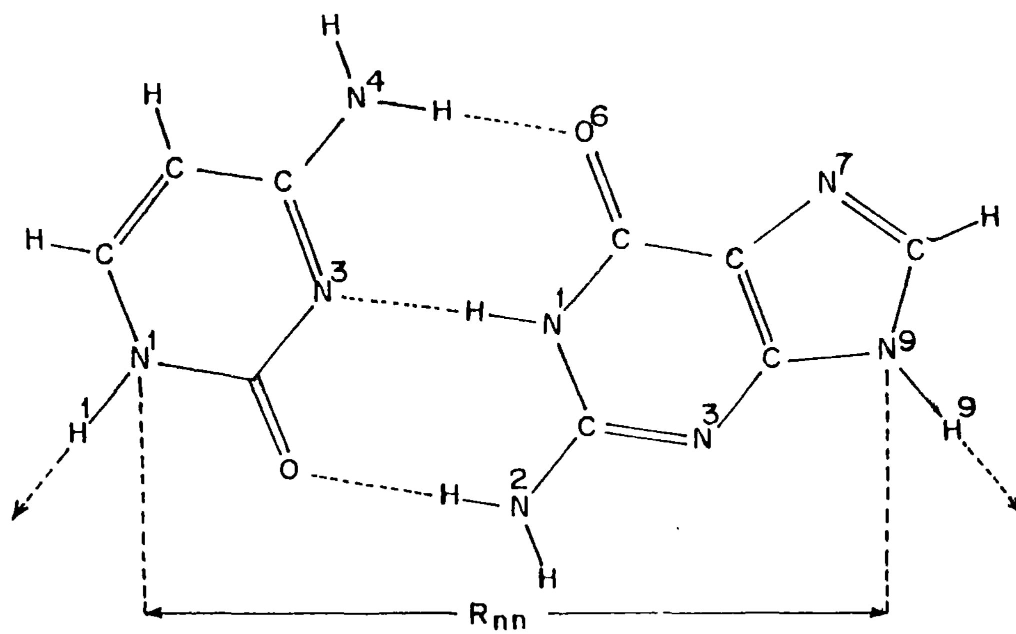
configuration other than a planar or Cs symmetry-constrained one. This use of two options also permitted the opportunity for seeing if two different optimisation strategies would lead to basically the same minimum, at least for some of the cases. Use of the unconstrained strategy was also important for examining cases where the equilibrium configuration departs far from the double-helical Watson-Crick one.

The thermodynamics of base-pair formation was studied from the enthalpy of base-pairing E_p, obtained from the heats of formation of the pair and of the individual bases. Note was taken of the lengths R_{hh} of the hydrogen bonds formed, and their number. This could allow for the possibilty of correlating the magnitude of the pairing energy Ep with the number of hydrogen bonds observed, and also with their length. In this Chapter, a hydrogen bond **H...X**is identified by the numberings on the hydrogen atom **H** involved (as attached to a basic site) and the electronegative atom **X** involved directly with the bond.

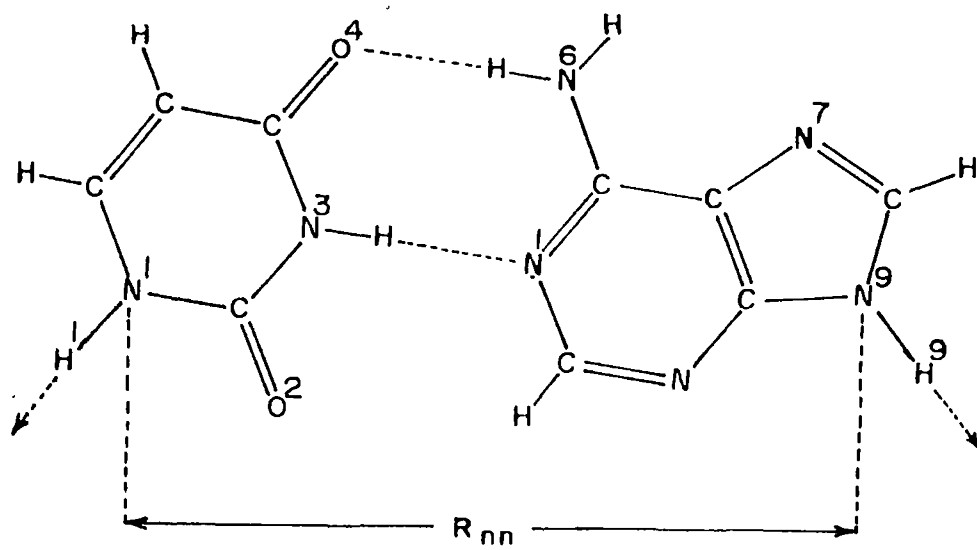
The intermolecular configuration of a base-pair was gauged by 3 markers - the distance R_{nn} between the sugar-bonding nitrogens of the 2 bases, the angle θ_{nh} between the N-H bonds at these nitrogens, and the dihedral angle ϕ_{nh} between these two N-H bonds. Fig. IV.2 gives a schematic representation of the configurational markers R_{nn}, θ_{nh} and ϕ_{nh} used in defining the alignment between base pairs consisting of a purine and a pyrimidine. Comparison of the values of these configurational markers for a particular base-pair with those for the standard Watson-Crick base-pairs could lead to a proper discussion of the degree with which the given base-pair resembled or departed from the standard double-helical configuration.

IV.3 Ionic and Tautomeric DNA Base Mismatches

As was discussed above, the integrity of the genetic code lies in the stability of the



Guanine: Cytosine pair



Adenine: Thymine pair

Fig. IV.2 : Schematic representation of the configurational markers observed between Watson-Crick base pairs

complementary base-pairing present in the DNA duplex. Nevertheless, the possibility that the presence of rare tautomeric forms of the bases might be responsible for the low level of spontaneous mutations believed common to all organisms was suggested by Topal and Fresco, as discussed in Chapter III. The different hydrogen-bonding characteristics of the tautomer may lead to thermodynamically favorable base-mispairing schemes, thus giving ground to the speculation that base misincorporations involving these tautomers could lead to spontaneous or endogenous mutations. Note that the tautomer concerned might be present either in the nucleic acid strand itself, or in the incoming species from the nucleotide pool. Either way, the possibility of mispairing could exist.

The tempting possibility of the significance of existence of rare tautomeric species in nature has led many experimental and theoretical groups to investigate the energetics of the proposed tautomeric forms. At the most, these studies go to show that the concentration of these tautomeric species is very low, perhaps not sufficient enough to induce point mutations significantly (Strazewski & Tamm 1990; Kwiatkowski *et al* 1986). In fact, while base tautomerization may well be responsible for replication- dependent misincorporation, this may not be due to the specific misincorporation of an incorrect base (Strazewski & Tamm 1990; Carroll & Benkovic 1990). Also, the base-pair induced shift in the tautomeric equilibrium of the modified DNA base N^4-methoxy cytosine, observed by NMR studies to exist predominantly in imino configuration, and its base-pair formation with adenine in Watson-Crick fashion, supports the hypothesis that rare tautomeric forms may be involved in mutagenesis (Fazakerley *et al* 1993). Nevertheless, the matter is still intriguing and could merit further investigation, particularly in the context of the belief that spontaneous mutations could be a common cause of endogenous cancers and other genetic disorders.

The precise manner in which the DNA bases pair in Watson-Crick fashion forms a primary structural basis for the DNA double- helix. It is noteworthy that both the purine-pyrimidine base- pairs normally occurring, viz. the guanine-cytosine (GC) and the adenine-thymine (AT) base pairs, are characterised by approx- imately the same values of the standard configurational markers, pointing to almost the same overall configuration for both. In the context of nature as a whole, however, the pairing patterns need not be confined to the Watson-Crick configuration, since the structures of the bases allow for other schemes, like the Hogsteen, reverse Watson-Crick and reverse Hogsteen pairing schemes. In fact, if external configurational constraints are completely lifted, the number of possible base-pairs increases to a total of 24, since any situation allowing stabilisation through hydrogen-bonding would be feasible.

In this dissertation, the effect of protic changes upon the base-pairing possibilities open to the naturally occuring DNA bases is studied by invoking the three possibilities mentioned in Chapter III - viz. those of tautomerism, of protonation, and of deprotonation. Sections IV.3.1 to IV.3.3 below delineate each possibility briefly in turn.

IV.3.1 Tautomeric base-pairing schemes

The normally stable forms of DNA bases are able, in the B-DNA double-helical configuration, to furnish the two Watson-Crick base pairs G:C and A:T. The next stablest tautomers, depicted as G^*, A^*, C^* and T^*, correspond to transfer of the hydrogen-bonding protons to the O^6-G, N^1-A, N^3-C and O^2-T positions for G, A, C and T respectively. These tautomers would be also able to pair in the double-helical configuration, as proposed by Pullman and Pullman (1962; 1964) and indicated by theoretical calculations (Kong *et al* 1987; Lipinski 1988; Kwiatkowski *et al* 1986; Czerminski *et al* 1990; Kwiatkowski & Person

1990).

The expected outcome of these unusual pairs in the biological context is the incorporation of a wrong base, leading to the transitional mutations G^{*} --> A, A^{*} --> G, C^{*} --> T and T^{*} --> C. It has been suggested that these kinds of pairs involving tautomers could constitute a possible basis for the induction of spontaneous mutations *in vivo.* This tautomer model for spontaneous mutagenesis complements the well-known proton tunnelling model of Lowdin for spontaneous mutagenesis (Lowdin 1955; 1977). The proton-tunnelling model invoked the possibility of quantum mechanical tunnelling of a proton through the hydrogen-bond potential well as the physical basis for base-misincorporation (Ford & Wang 1992a, 1992b). The probability of this occurring has been calculated as being exceedingly low. In fact, the chief objection to the tautomer model for spontaneous mutations has been this low probability and the transient existence of the tautomers.

IV.3.2 Protonated base-pairing schemes

Permanent (non-transient) base-pairs involving a cationic (protonated) species would have to occur at a pH below the acidic pK_a of the base (corresponding to protonation). Transient or very low concentrations of the protonated species could of course occur at higher pH values. Studies have been made *in vitro* of the base-pairing properties of protonated bases at a low pH environment in solvent phase. The base-pairing potential of N^3- protonated adenine present in a dodecameric nucleotide sequence at pH below 5.5 has been shown by nmr spectroscopy (Gao & Patel 1988) to involve a mismatch between protonated adenine in the *anti* conformation and guanine in the *syn* conformation. The pairing configuration is evidently not quite double-helical though. In a (CA) dodecamer,

N^3-protonated adenine has also been indicated by melting point studies (Brown *et al* 1990) to mismatch well with cytosine at pH 5.5, both being in the normal *anti* conformations.

IV.3.3 Deprotonated Base-Pairing Schemes

Non-transitory base-pairs incorporating an anionic (deprotonated) species would of necessity require a high pH environment well above the basic pK_a value of the base (corresponding to proton abstraction). A survey of the literature shows that deprotonation of normal nucleic acid bases occurs at high basic pH (above 9.0). This is not so for bases modified by alkylation, where many of them exist in deprotonated form at neutral pH or even lower pH. The base-pairing potential of deprotonated nucleic acid bases in their alkylated forms has been well- documentated (Voigt & Topal 1990; Gaffney & Jones 1989; Dosanjh *et al* 1990; Georgiadis *et al* 1991), being invoked as a basis for the mutagenic and carcinogenic properties of alkylating agents. One can surmise that the corresponding free unalkylated bases in their deprotonated forms would also base-pair in analogous fashion, the chief difference being the pH at which this would take place.

IV.3.4 Pairing properties of ions and tautomers of bases

Table IV.1 gives the PM3 pairing energies and configurational data for the normal G:C and A:T pairs as well as for the pairs expected for the tautomeric forms of the 4 DNA bases viz. G^*, A^*, C^* and T^*, which pair with T, C, A and G respectively, giving rise to base-pair transitions. For some species, the C_s symmetry constraint is adopted, and free optimisation for the others. There does not appear to be a marked difference introduced by either option.

The pairing energies for the G:C and A:T pairs are 11.79 and 5.61 kcal/mol respectively,

Table IV.1 PM3 calculated data for base pairs involving the normal and tautomeric forms of nucleic acid bases

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\varnothing_{nh}$ (deg)	θ_{nh} (deg)
Normal						
G:C	-11.79	$O^6G:H^4C$	1.804	9.044	-7.0	57.68
		$H^1G:N^3C$	1.780			
		$H^2G:O^2C$	1.840			
A:T	-6.10	$O^4T:H^6A$	1.823	9.125	6.64	42.46
		$H^3T:N^1A$	1.778			
Tautomers						
G^*:T	-4.33	$N^1G:H^3T$	1.809	9.084	-15.12	54.32
		$H^2G:O^2T$	1.796			
T^*:G	-4.49	$H^1G:N^3T$	1.839	9.016	-19.13	57.16
		$H^2G:O^4I$	1.849			
A^*:C	-10.03	$H^1A:N^3C$	1.781	9.107	-7.25	52.49
		$N^4A:H^4C$	1.782			
C^*:A	-5.30	$N^1A:H^3C$	1.781	9.131	-10.60	54.25
		$H^4A:N^4C$	1.813			

reflecting upon the existence of 3 and 2 hydrogen bonds respectively for these normal pairs. The normal double-helical configuration is taken here to be typified by these two pairs, with R_{nn} values of 9.044 and 9.134 A, ϕ_{nh} values of -7.0 and 0.0 degrees, and θ_{nh} values of 57.68 and 52.60 degrees for the G:C and A:T respectively, resorting to an unconstrained optimisation strategy. The tautomeric base pairs (G^*:T, T^*:G, A^*:C and C^*) are all seen to be favoured energetically in a configuration approximating well to the double-helical configuration as typified by the two normal pairs. The range of pairing energy values (-4.33, -9.98, -5.30 and -4.49 kcal/mol for the G^*:T, A^*:C, C^*:A and T^*:G pairs) compares well with the range displayed by the normal pairs. This data could point to the feasibility of these pairs occuring in nature in response to the transient or minute existence of the tautomeric forms. The results of these calculations may be taken to serve as a model for the spontaneous induction of base-pair transitions in nature.

Table IV.2 presents PM3 values of the pairing energies and configurational markers for base pairs involving the deprotonated forms of the DNA bases guanine and thymine, free optimisation and C_s symmetry imposition being both considered. The data of Table IV.2 would pertains only to a basic pH above the pK_a values of the corresponding nucleoside systems (about 13 for dG and dT), when the anionic forms may be expected to exist in appreciable concentrations.

The guanine anion is seen to pair most favourably with thymine when freely optimised, although the values of the configurational markers indicate some departure from the typical double-helical configuration by introduction of a twist. The corresponding C_s symmetry-constrained pair, although less favoured energetically, has a configuration closer to the double-helical one. The G^- anion can also pair favourably with cytosine, although the configuration departs from double-helical. The thymine anion pairs very favourably

Table IV.2 PM3 calculated data for base pairs involving the deprotonated forms of nucleic acid bases

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\varnothing_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
G^-:T	-7.62	$N^1G:H^3T$	1.713	9.017	-22.72	64.20
		$H^2G:O^2T$				
G^-:C	-5.61	$O^6G:H^4C$	1.801	9.972	-2.58	66.00
		$H^2G:O^2C$	3.728			
T^-:G	-19.33	$N^3T:H^1G$	1.777	9.107	-0.46	47.60
		$O^2T:H^2G$	1.754			
Geometries with C_s symmetry						
G^-:T	-3.74	$N^1G:H^3T$	1.744	9.017	0.00	47.59
		$H^2G:O^2T$	1.863			
G^-:C	-5.08	$O^6G:H^4C$	1.802	10.232	0.00	66.79
		$H^2G:O^2C$				
T^-:G	-19.93	$N^3T:H^1G$	1.777	9.105	0.00	47.93
		$O^2T:H^2G$	1.756			
T^-:A	-8.55	$O^4T:H^6A$	1.790	10.704	0.00	81.56

with guanine under both optimisation regimes, the large value of pairing energy being more than expected from the two hydrogen bonds involved. The T^-:G pair is predicted to have a configuration not far from double-helical. Although the T^-:A pair is also energetically possible, it would not be easily accomodated within the double-helical configuration, as seen from the values of the configurational markers. The consensus from these calculations is that the anions of the bases guanine and thymine would pair favourably only with thymine and guanine respectively in the double-helical configuration at the appropriate conditions of high pH, leading to base-pair transitions. However, the anionic forms of the normal base-pairs (viz. G^-:C and T^-:A) pair with departure from double-helical configuration. This may be interpreted as indicating that highly basic pH would have a destabilising effect upon the double-helix.

Table IV.3 furnishes PM3 pairing data for base pairs involving the protonated forms of the bases adenine, guanine and cytosine, where the protonation sites are as predicted from the considerations arrived at in Chapter Three above, viz. N^3-A, N^7-G and O^2-C for A^+, G^+ and C^+ respectively. These data would naturally pertain only to acidic pH values below the acidic pK_a values for the various nucleoside systems. The pairing energies are seen to be larger than usual, reflecting upon the attraction between the cationic protonated system and the nucleophilic pi cloud of the neutral pairing base. This is especially noticed in the rather small values of the configurational marker R_{nn}, indicating a closer (tighter) association of the bases in the pair. All in all, this data would lead to the general prediction that acidic pH would be a factor leading to stabilisation and higher melting point for a DNA duplex.

The A^+:C pair departs somewhat from the Watson-Crick configuration, as indicated by the large value of θ_{nh} (89.5°). The A^+:G pair involves a purine-purine mismatch, where

Table IV.3 PM3 calculated data for base pairs involving the protonated forms of nucleic acid bases

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\varnothing_{nh}$ (deg)	θ_{nh} (deg)
$A^+:C$	-28.65	$H^1A:O^2C$ $H^6A:N^3C$	1.757 1.784	8.995	-2.20	89.47
$A^+:G$	-28.55	$H^1A:N^7G$ $H^6A:O^6G$	1.739 1.777	8.971	-1.75	58.86
$C^+:C$	-27.15	$H^3C:O^2C$ $H^4C:N^3C$	1.758 1.773	7.396	0.31	40.09
$G^+:A$	-15.73	$N^7G:N^1A$ $O^6G:H^6A$	1.704 1.850	8.920	-10.31	58.96
$G^+:C$	-24.67	$N^7G:N^3C$ $O^6G:H^4C$	1.718 1.841	6.758	-2.94	53.69

the guanine is in the *syn* conformer with respect to the sugar moiety; this pair is close to the normal pairs in its configuration. The C^+:C mismatch departs far from double-helical configuration, and has been invoked in the formation of quadruplex strands in DNA, as discussed in Sec. III.5. The G^+:A complex is not too far from double-helical, again involving guanine in the *syn* conformer.

Most of the pairs studied were in basically Watson-Crick alignment, so as to permit interpretation in terms of stability of the double-helix.

IV.4 Base-Pairs with Alkylated DNA Bases

DNA base alkylation by xenobiotic agents has been linked to point mutation and oncogene activation, and is believed to exert its genotoxic effect via the unusual base-pairing schemes adopted by the alkylated base residues, which include the products of alkylation at the N^7-guanine (N^7-G), O^6-guanine (O^6-G), O^4- thymine (O^2-T), O^2-thymine (O^4-T) and N^3-guanine (N^3-G) sites (Singer *et al* 1978; Singer 1976).

IV.4.1 Base incorporation by alkylated nucleosides

N^7-Methylguanine does not lead to misincorporation of a non- complementary base when present in templates for nucleic acid polymerases (Schoental 1969; Ludlum 1970). O^6-Methylguanine has been shown to possess the ability to misincorporate thymine residues when present in *in vitro* and *in vivo* templates for DNA and RNA synthesis (Gerchman & Ludlum 1973; Abbott & Saffhill 1979; Mehta & Ludlum 1978; Bhanot & Ray 1986).

103235

O^4-Alkylthymines can similarly misincorporate guanine residues during nucleic acid synthesis, as demonstrated for O^4-methyl-, ethyl- and isopropylthymines (Singer *et al* 1983, 1986; Preston et al 1986, 1987). The evidence for O^2-methylthymine is less conclusive (Saffhill & Abbott 1978).

These findings may be summed up by proposing that N^7- methylguanine is nonmutagenic while O^6-methylguanine and O^4- methylthymine are pro-mutagenic, the position of O^2-methylthymine being not too clear and that of N^3-methylguanine being unknown.

IV.4.2 Previous experimental and theoretical studies

Much *in vitro* and theoretical work has been done to corroborate and add finer detail to the above basic findings, which may be summarised as in the following three paragraphs:

X-ray crystallography and NMR studies indicate that substitution of O^6-methylguanine (O^6-MeG) for guanine in synthetic DNA duplexes destabilises the helical structure (Voigt & Topal 1990). This has the effect of decreasing the T_m of DNA significantly. That these disruptions take on the nature of only local perturbations is indicated by x-ray studies (Brown *et al* 1985; Hunter *et al* 1986) and NMR studies (Patel *et al* 1986a, 1986b; Kalnik *et al* 1988a, 1988b). The O^6-MeG:C and O^6-MeG:T pairs are both recognised by the repair enzyme ABC excinuclease *in vitro*, which apparently recognises the local helical perturbation rather than the lesion itself (Voigt *et al* 1989). Although the conformation adopted by the O-methyl group is *syn* to the hydrogen bonding side in the solitary deoxynucleoside and in the single-stranded oligonucleotide, in the double-stranded duplex it goes out of plane, protruding into the major groove for the O^6-MeG:C pair (Parthasarathy & Fridey 1986; Ginell *et al* 1990). Proton NMR studies on a right-handed helical

dodecamer indicate an *anti* conformation for the methyl group in an O^6-MeG:C pair (Patel *et al* 1986). The crystal structure of a B-DNA oligonucleotide containing O^6-MeG also revealed an *anti* conformation for the methyl group, where the mismatched O^6-MeG:T pair is basically in Watson-Crick configuration (Leonard *et al* 1990). NMR solution studies have found that the O^6- alkylguanine:T base-pair most likely retains the Watson-Crick alignment, while the O^6-MeG:C pair adopts a wobble configuration (Li *et al* 1988). UV melting studies point to a Watson-Crick configuration for the O^6-MeG:C pair provided the modified base is N^1-protonated (Lionard *et al* 1990).

O^4-Alkylthymines are considered to be of greater relevance for carcinogenesis and mutagenesis than O^6-alkylguanines because of their greater resistance to repair (Swenberg *et al* 1984; Richardson *et al* 1985; Brent *et al* 1988). A Watson-Crick type of alignment was proposed by Singer for the pair between O^4-methyl thymine (O^4-MeT) and thymine, linked by two H-bonds, where the methyl group would not be *syn* to the hydrogen bonding side, but possibly in a conformation intermediate between *syn* and *anti* (Singer 1980). This is in contrast to a 2D NMR solution structure obtained which proposed only one hydrogen bond (Kalnik *et al* 1988). Both O^4-MeT:G and O^4-MeT:A base pairs were formed by bacterial and fruitfly DNA polymerase fragments acting on a 25-meric oligonucleotide template containing O^4-MeT at a unique site (Dosanjh *et al* 1990). The efficiency of forming the O^4- MeT:G pair was about 10 times greater than that of forming the O^4-MeT:A or T:G pairs, and bacterial T4 DNA polymerase allowed for stable incorporation of G opposite O^4-MeT in contrast to incorporation of G opposite T. The solitary nucleoside O^4-MedT has the methyl group *syn* to the hydrogen bonding zone , but when O^4-MedTTP is placed opposite a poly(dA) template, the methyl group reorients to the *anti*

conformation which is more favourable for base-pair formation (Engel & von Hippel 1987). Pohorille and Loew (1985) used a perturbation theory treatment to study the base-pairing properties of some O-methylated bases, all of which were without the relevant Watson--Crick protons. This study predicted the role of conformation of the O-methyl groups for favourable base-mismatching in the double-helical configuration, where only the *anti* conformers were conducive to base-pairing of any kind. Ford and Wang (1992a, 1992b) examined the possibilities of interstrand proton transfer serving as a basis for mutagenesis following protonation and methylation at the N^7- and O^6-guanine sites using the AM1 SCF-MO method. Quantum chemical calculations have indicated the *syn* conformer of O^6-MeG as being the most stable, while energy-minimisation studies using the AMBER methodology show that within a pentameric oligonucleotide, the *anti* conformer is more stable (Pederson et al 1988). The conformation and dynamics of oligonucleotides containing modified thymines have been studied using molecular dynamics simulations (Goodfellow & Williams 1992; Cruzeiro-Hannson *et al* 1992). Parker *et al* (1993) used molecular dynamics simulation studies on oligo-nucleotides containing O^6-MeG to indicate that the O^6-MeG:T pair is of the Watson-Crick type structure, so that the sequence containing this pair is closer to the normal unmodified sequence than that containing the O^6-MeG:C pair (Parker *et al* 1993).

IV.4.2 Physical basis for induction of point mutations

For the alkylated DNA bases considered here, this physicochemical basis is proposed to include two components, viz. (a) abstraction of the Watson-Crick proton (the N^1-proton for methylated guanines and the N^3-proton for methylated thymines) and (b) retention of the alkylating group in a conformation sterically conducive to base-mispairing in the

double-helical configuration. The study of these two structural criteria form the objective of this study.

Criterion (a) is dealt with by studying each methylated base with and without the concerned proton. Criterion (b) is incorporated for those cases with exocyclic methyl groups, by including two conformers for study - one with the methyl group *syn* to the hydrogen-bonding region, and the other *anti*. The various possibilities arising from joint consideration of these two criteria are examined for their base-pairing properties to see which situations are predicted as feasible.

5 methylated bases were incorporated for study of their base-pairing properties : N^7-methylguanine (N^7-MeG), O^6-methylguanine (O^6-MeG), N^3-methyl-guanine (N^3-MeG), O^2-methylthymine (O^2-MeT) and O^4-methylthymine (O^4-MeT) whose mutagenic properties were summarised above. Each system was studied in two forms - with and without the Watson-Crick proton, respectively designated by a plus and zero sign, eg. N^7-MeG$^+$ and N^7-MeG$^\circ$ for N^7-methylguanine. The O-methylated bases have *syn* and *anti* conformers for the methyl group (with respect to the hydrogen-bonding zone), and these are respectively designated by the symbols s and a in brackets, eg. O^6-MeG$^+$(a) and O^6-MeG$^+$(s) for cationic O^6- methylguanine. Fig.IV.3 depicts the 5 bases in their protonated form, with the O-methylated bases in their *anti* conformers. Fig. IV.4 gives the corresponding representations for the deprotonated bases, while Fig.IV.5 presents the *syn* conformers for the O-methylated bases.

IV.4.3 Pairing properties of methylated guanines

The 3 methylated guanines in their protonated and deprotonated forms were paired with cytosine and thymine, with O^6-MeG in the *anti* conformation. Table IV.4 presents results

for pairing of the protonated systems with cytosine, giving values of the pairing energy E_p, the 3 configurational markers R_{nn}, θ_{nh} and ϕ_{nh}, and the lengths R_{hb} of the hydrogen bonds formed, both optimisation options being incorporated, viz. free optimisation and imposition of C_s symmetry. Table IV.5 presents the same data for pairing of the deprotonated systems with cytosine. Table IV.6 presents the data for pairing of the deprotonated guanine systems with thymine.

The data of Table IV.4 correspond to pH values of the surrounding medium where the base systems exist in a form with the N^1-proton present. From the negative values of the pairing energies E_p, it is immediately obvious that this presence of the N^1-proton allows for the modified guanine systems to pair favourably with cytosine as the free base does. In fact, the net positive charge on the methylated guanines leads to pairing energies even lower than that for the G:C base pair. This may be taken to indicate that, at the appropriately low pH, substitution of these methylated guanines for guanine would stabilise the double-helix, lowering the T_m value for the duplex. These pairing schemes are non-mutagenic, not leading to a base substitution. This goes to indicate that the mere steric presence of the methyl group does not in itself constitute a molecular determinant for successful base-mismatching so long as the Watson-Crick proton is still present. The geometries of all the base-pair systems more or less conform to the double-helical configuration typified by the G:C pair. This is in line with the experimental evidence for the Watson-Crick configuration of the O^6-MeG$^+$:C pair, where the O^6-methylguanine is N^1-protonated. The imposition of the C_s symmetry constraint does not lead to results or conclusions markedly differing from those obtained without any constraints.

Table IV.5 furnishes the data for the pairing between cytosine and the guanine systems minus the N^1-protons, which would in each case correspond to a pH value for the

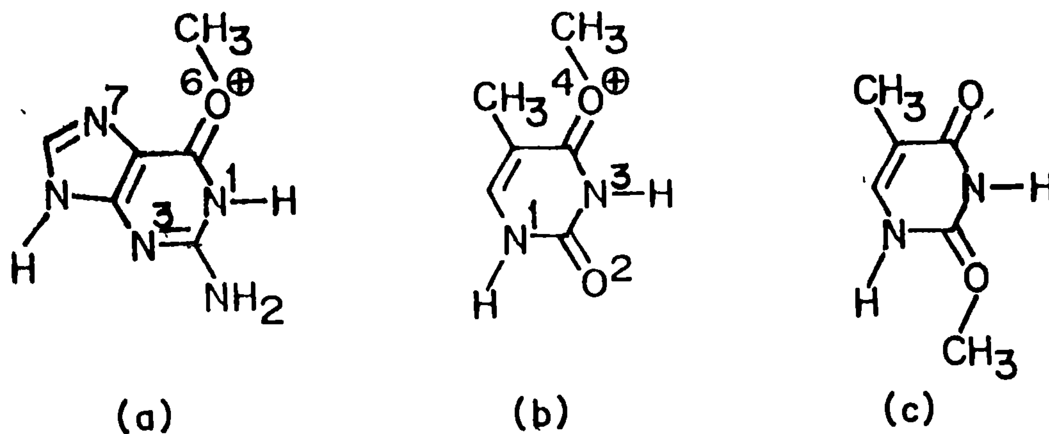


Fig. IV.3 : .methylated nucleic acid bases in their protonated form

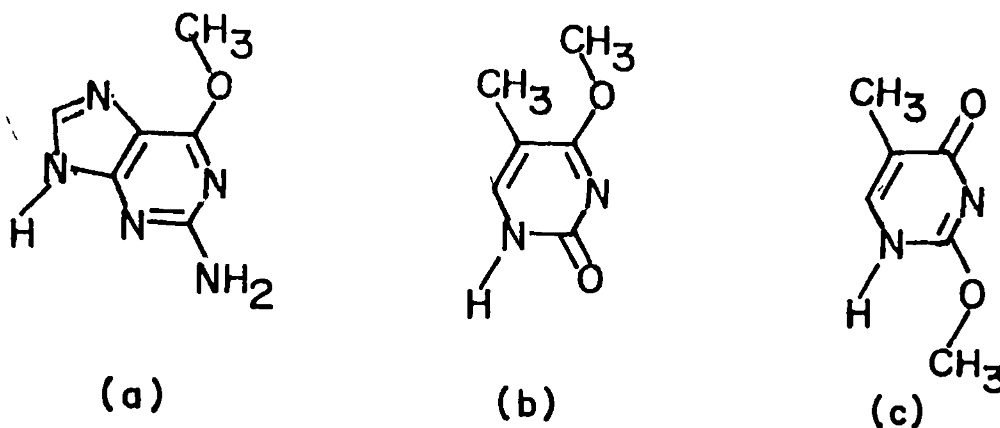


Fig. IV.4 : methylated nucleic acid bases in their deprotonated form

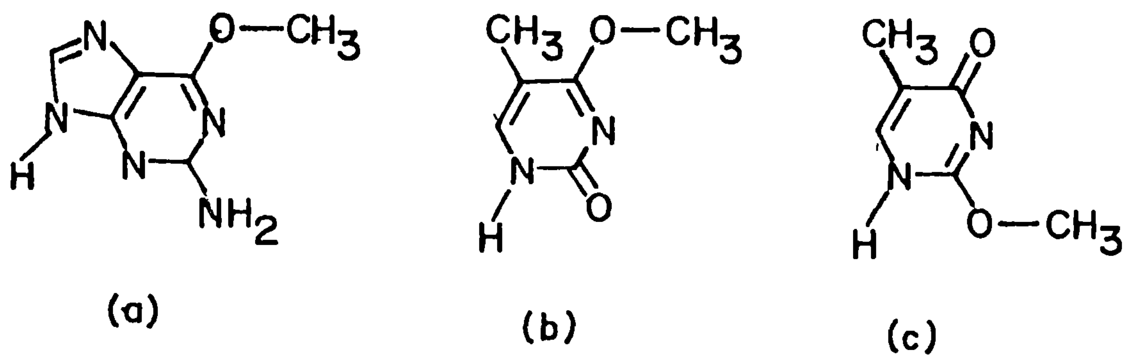


Fig. IV.5 : Syn conformers for O-methylated bases

Table IV.4 PM3 calculated data for pairing between cytosine and methylguanine systems which retain the N^1-proton

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\varnothing_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
G:C	-11.79	$O^6G:H^4C$	1.804	9.044	-7.00	57.68
		$H^1G:N^3C$	1.780			
		$H^2G:O^2C$	1.840			
N^3MeG^+:C	-30.67	$O^6G:H^4C$	1.832	9.152	2.12	55.00
		$H^1G:N^3C$	1.737			
		$H^2G:O^2C$	1.740			
N^7MeG^+:C	-23.41	$O^6G:H^4C$	1.851	9.081	3.05	52.85
		$H^1G:N^3C$	1.733			
		$H^2G:O^2C$	1.782			
O^6MeG^+(a):C	-23.33	$O^6G:H^4C$	1.929	9.053	-9.91	54.53
		$H^1G:N^3C$	1.745			
		$H^2G:O^2C$	1.778			
Geometries with C_s symmetry						
G:C	-11.09	$O^6G:H^4C$	1.794	9.115	0.00	53.39
		$H^1G:N^3C$	1.781			
		$H^2G:O^2C$	1.836			
N^3MeG^+:C	-30.68	$O^6G:H^4C$	1.828	9.147	0.00	53.71
		$H^1G:N^3C$	1.738			
		$H^2G:O^2C$	1.742			
N^7MeG^+:C	-23.80	$O^6G:H^4C$	1.843	9.076	0.00	51.63
		$H^1G:N^3C$	1.730			
		$H^2G:O^2C$	1.782			
O^6MeG^+(a):C	-23.54	$O^6G:H^4C$	1.924	9.082	0.00	50.34
		$H^1G:N^3C$	1.755			
		$H^2G:O^2C$	1.762			

Table IV.5 PM3 calculated data for pairing between cytosine and N^1-deprotonated methylguanine systems

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$Ø_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
N^3MeG^0:C	-3.33	$O^6G:H^4C$	1.840	9.250	16.41	54.64
		$H^2G:O^2C$	1.848			
N^7MeG^0:C	-0.17	$O^6G:H^4C$	1.839	9.121	10.28	56.30
		$H^2G:O^2C$	1.875			
O^6MeG^0(a):C	-1.76	$O^6G:H^4C$	1.945	9.233	-11.95	51.63
		$H^2G:O^2C$	1.861			
Geometries with C_s symmetry						
N^3MeG^0:C	-2.37	$O^6G:H^4C$	1.833	9.334	0.00	52.89
		$H^2G:O^2C$	1.869			
N^7MeG^0:C	0.44	$O^6G:H^4C$	1.831	9.255	0.00	51.69
		$H^2G:O^2C$	1.879			
O^6MeG^0(a):C	-0.58	$O^6G:H^4C$	3.573	9.619	0.00	38.11
		$H^2G:O^2C$	1.835			

medium which is higher than those corresponding to the systems considered in **Table IV.4**. Since in each case, the base pairing with these guanine systems is cytosine, none of these pairing schemes may be described as mutagenic. The first observation is that the pairing energies in each case are appreciably less negative than those of Table IV.4, indicating lesser favourability of base-pairing. This is because of the loss of the N^1-proton resulting in the disappearance of one hydrogen bond. These neutral methylated systems are characterised by low pairing energies and only two hydrogen bonds. As such, these pairs would be expected to destabilise the DNA duplex at the appropriate pH, as indicated by the destabilising effect noted experimentally by substitution of O^6-MeG for guanine. The two types of optimisation strategies lead to different pairing configurations, particularly for the O^6-MeG^0:C pair. Even if the fairly small degree of torsional twist ϕ_{nh} in the freely optimised geometries be regarded as accomodatable in the double- stranded situation, the values of the R_{nn} and θ_{nh} markers show some departure from the Watson-Crick alignment typified here by the G:C pair of Table IV.4. However, if the freely optimised geometries be taken as more representative of the real situation than the artificially symmetry--constrained geometries, these results may be taken as predictive of a wobble type of pairing configuration for these neutral alkyl guanines with cytosine, which has been indicated experimentally by NMR solution studies for the O^6-MeG^0:C pair (see above).

Table IV.6 gives the data for pairing between thymine and the deprotonated guanine systems, which pairings (like the pairings of Table IV.5) may be expected to occur at pH values higher than those for the systems in Table IV.4. Pairing between a guanine system and thymine would be mutagenic if successful, leading to a base transition (G --> A). Here again, pairing occurs via two hydrogen bonds, and the pairing energies are smaller for each case in comparison with the pairing between the protonated guanines and

Table IV.6 PM3 calculated data for pairing between thymine and N^1-deprotonated methylguanine systems

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\varnothing_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
N^3MeG^0:T	-4.92	N^1G:H^3T	1.820	9.069	-38.60	60.30
		H^2G:O^2T	1.805			
N^7MeG^0:T	-5.84	N^1G:H^3T	1.782	9.026	-41.25	59.40
		H^2G:O^2T	1.821			
O^6MeG^0(a):T	-5.68	N^1G:H^3T	1.830	9.610	-20.41	54.38
		H^2G:O^2T	1.816			
Geometries with C_s symmetry						
N^3MeG^0:T	-2.63	N^1G:H^3T	1.857	9.195	0.00	53.83
		H^2G:O^2T	1.785			
N^7MeG^0:T	-4.51	N^1G:H^3T	1.802	9.118	0.00	51.40
		H^2G:O^2T	1.792			
O^6MeG^0(a):T	-3.87	N^1G:H^3T	1.837	9.152	0.00	49.11
		H^2G:O^2T	1.793			

cytosine. But the pairing in each case is predicted to be favourable, as attested to by the negative values of E_p. The results obtained from free optimisation and from the C_s symmetry constraint differ to some extent here. While the geometries derived from C_s symmetry imposition could be possibly accomodated into the Watson-Crick configuration, there is a large departure from this configuration for the freely optimised geometries. This raises the question of whether the latter could fit in at all into the normal Watson-Crick double helix. If not, then one would have to depend upon the results obtained from C_s symmetry imposition to get a picture of what might be the actual situation prevailing in the normal double-helix. In this case, the small variations from Watson-Crick alingment as typified by the G:C pair (Table IV.4) may be interpreted in terms of the minor local perturbations in helical structure which have been indicated by experimental studies pertaining to this situation. That the configuration for the O^6-MeG^0(a):T pair is close to the Watson-Crick alignment is in line with the results of x-ray structure elucidation. It is also in good accord with the predictions of molecular dynamics simulation studies which indicate that the oligonucleotide sequence containing the O^6- MeG^0:T pair is closer to the normal one than that containing the O^6-MeG^0:C pair (see above).

From the data of Tables IV.5 and IV.6, it is predicted that O^6-MeG^0 can pair with both cytosine and thymine, although the latter is definitely preferred energetically and sterically, being closer to the normal Watson-Crick configuration. This points to the competition observed between incorporation of thymine and incorporation of cytosine when both are present for incorporation into a DNA strand by action of a bacterial DNA polymerase (Abbott & Saffhill 1979). The mutation efficiency has also been noted *in vivo* to be about 0.75 (Bhanot & Ray 1986), indicating a definite preference for the mutagenic O^6-MeG^0:T pair over the nonmutagenic O^6-MeG^0:C pair.

IV.4.4 Pairing properties of methylated thymines

The 2 methylated thymines in their protonated and deprotonated forms were paired with guanine and adenine, the methylated bases each in their *anti* conformations. Table IV.7 presents results for pairing of both the protonated and deprotonated systems with adenine, giving values of the pairing energy E_p, the 3 config- urational markers R_{nn}, θ_{nh} and ϕ_{nh}, and the lengths R_{hb} of the hydrogen bonds formed. Table IV.8 presents the same data for pairing of the deprotonated systems with guanine.

The data of Table IV.7 for protonated O-methylthymines pertains to pH values corresponding to retention of the N^3-T proton, where pairing proceeds via two hydrogen bonds. It is predicted that these protonated bases, while pairing favourably with adenine, adopt pairing configurations that deviate only somewhat from the double-helical one as typified by the A:T pair. None of these pairs are mutagenic, since the modified thymines pair with adenine, indicating that retention of the N^3 proton would allow for non-mutagenic pairing schemes. The values of E_p are larger in magnitude for modified thymines than for normal thymine, leading to the prediction that, at the appropriately acidic pH, substitution of O^2- or O^4-methylthymine for thymine would result in stabilisation of the double-helix. Imposition of C_s symmetry and free optimisation both lead to essentially the same conclusions.

The data of Table IV.8 for the N^3-deprotonated bases pertain to higher pH values where the proton is abstracted, and pairing with adenine has to occur through just one hydrogen bond. None of these pairs are predicted as energetically favoured, with C_s symmetry imposition and free optimisation leading to the same result. The pairing configurations for these pairs appear to deviate even more from double-helical than those of Table IV.7. This all leads to the inference that O^{2-} and O^4-methylthymines in their N

Table IV.7 PM3 calculated data for pairing between adenine and N^3-protonated thymine systems

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$Ø_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
A:T	-6.10	$O^4T:H^6A$	1.823	9.125	6.64	42.46
		$H^3T:N^1A$	1.778			
$O^4MeT^+(a):A$	-12.75	$O^4T:H^6A$	2.523	8.843	-4.90	51.23
		$H^3T:N^1A$	1.721			
$O^2MeT^+(a):A$	-14.69	$O^4T:H^6A$	1.864	9.045	6.61	51.23
		$H^3T:N^1A$	1.683			
Geometries with C_s symmetry						
A:T	-5.61	$O^4T:H^6A$	1.811	9.134	0.00	52.60
		$H^3T:N^1A$	1.766			
$O^4MeT^+(a):A$	-10.93	$O^4T:H^6A$	2.432	8.888	0.00	50.63
		$H^3T:N^1A$	1.725			
$O^2MeT^+(a):A$	-14.90	$O^4T:H^6A$	1.860	9.044	0.00	45.53
		$H^3T:N^1A$	1.684			

Table IV.8 PM3 calculated data for pairing between adenine and N^3-deprotonated thymine systems

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$Ø_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
$O^2MeT^0(a)$:A	0.28	O^4T:H^6A	1.851	9.086	39.09	62.09
$O^4MeT^0(a)$:A	1.85	O^4T:H^6A	2.848	9.186	11.63	45.99
Geometries with C_s symmetry						
$O^2MeT^0(a)$:A	5.28	O^4T:H^6A	1.851	10.768	0.00	76.86
$O^4MeT^0(a)$:A	3.11	O^4T:H^6A	3.117	9.936	0.00	47.39

^{N3}-deprotonated form would not form stable pairs with adenine in the double-helical configuration. The substitution of thymine by these O-methylated thymined in a DNA duplex would thus be expected to markedly destabilise the helix locally.

The data of Table IV.9 constitute the possibilities for pairing between deprotonated O-methylthymines and guanine, both optimisation strategies being used. These data pertain to a pH corresponding to abstraction of the N^3-protons, and indicate potentially mutagenic base-pairing situations. The results do not depend much upon the optimisation strategy employed. Both O^2- and O^4-methylthymines are seen to base-pair favourably with guanine, although the pairing energy for the former is appreciably lower than that for the latter, indicating that O^4- methylthymine might be a more efficient mutation-inducing agent than O^2-methylthymine. The base-pairing configurations for these systems are not far from double-helical, especially for the O^4- MeT^0(a):G pair, indicating that these pairs could be accomodated into this configuration. These findings are basically in accord with the mutagenic base-pair between O^4-MeT and G as proposed by Singer (Singer 1980), and do not agree with the single hydrogen- bonded base-pair proposed by Kalnik *et al* (1988).

The *anti* conformation for the O-methyl group is predicted to be favourable to these mutagenic pairing schemes, as indicated by experimental studies (Engel & von Hippel 1987). That the O^4-MeT:G pair is clearly more favourable than the O^4-MeT:A pair (compare Tables IV.8 and IV.9) is also predicted here, being in line with the relative efficiencies of formation of these pairs by action of polymerase fragments upon a 25-meric oligonucleotide template containing O^4-MeT (Dosanjh *et al* 1990).

IV.4.5 Conformational role of O-methyl groups

The exocyclic O-methyl group of O^6-methylguanine, O^2- and O^4- methylthymines may exist

Table IV.9 PM3 calculated data for pairing between guanine and N^3-deprotonated thymine systems

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\emptyset_{nh}$ (deg)	θ_{nh} (deg)
Freely optimised geometries						
$O^2MeT^0(a):G$	-1.62	$N^3T:N^1G$	1.973	9.297	21.16	50.82
		$O^2T:N^2G$	1.887			
$O^4MeT^0(a):G$	-6.13	$N^3T:N^1G$	1.868	9.014	-14.44	52.42
		$O^2T:N^2G$	1.833			
Geometries with C_s symmetry						
$O^2MeT^0(a):G$	-0.60	$N^3T:N^1G$	1.968	9.302	0.00	48.94
		$O^2T:N^2G$	1.906			
$O^4MeT^0(a):G$	-4.47	$N^3T:N^1G$	1.834	9.122	0.00	48.99
		$O^2T:N^2G$	1.858			

in the *syn* or *anti* conformation, defined with respect to the hydrogen-bonding zone. The conformation adopted is of import for the feasibility of base-pairing in the double-helical configuration, as is shown by the data of Table IV.10 on the pairs formed between the *syn* conformers and normal DNA bases. Data on the *anti* conformers is represented by the preceding Tables.

When the *syn* conformers are considered, the pairs between O^6- methylguanine and thymine, O^2-methylthymine and guanine, and O^4- methylthymine and guanine are predicted as energetically favourable, as seen from the negative values of E_p. The resultant geometries, however, indicate configurations far from double- helical. These results compare well with those of Pohorille and Loew (1985) who predicted through perturbational calculations that the *syn* conformers of O-methylated bases would furnish strongly repulsive interactions with normal DNA bases in the double-helical configuration. This is not the case for the *anti* conformers, as the preceding Tables show. It is interesting to observe the general agreement between our results and previously reported structural studies, including crystallographic, proton NMR and NMR solution studies, which all indicate that the methyl group of O^6-methylguanine adopts the *anti* conformation when faced with a pairing situation (Patel *et al* 1986; Ginell *et al* 1990; Leonard *et al* 1990).

There are conflicting reports on the possible orientation of the methyl group in the base-pair between O^4-methylthymine and guanine. Singer (1980) postulated a base-pairing scheme where the O^4-methyl group lies intermediate between *syn* and *anti* positions. The 2D NMR studies of Kalnik *et al* (1988) proposed a different scheme in which the O^4-methyl group adopts the *syn* orientation. Our studies suggest that though the O^4-methyl group may adopt both orientations when pairing, the base-pair with the *syn* conformer clearly deviates

from the double helical conformation (Table IV.10), while the one with the *anti* conformer closely approximates to the Watson-Crick alignment (Table IV.9).

The inference is that the *syn* conformers of these three O- methylated bases would not undergo mutagenic base pairing with the appropriate base under normal conditions (the *usual Watson- Crick double-helical configuration). This means that they would not be of* consequence for induction of base-substitutional mutations. A corollary of this is that the *syn* to *anti* rotational barrier would be a factor of relevance for the feasibility of an O-methylated base to induce point mutations.

IV.4.6 Summary of favoured pairing schemes

In the data of Tables IV.4 to IV.9, the feasibility of the base-pairs predicted favourable in the double-helical configuration is decided on energetic grounds, eliminating those pairs which occur out of this configuration, so that the favoured pairs include the following :

(a) For protonated alkylguanines, pairing with cytosine (viz. the N^3-MeG $^+$:C, N^7-MeG $^+$:C, O^6-MeG $^+$(a):C pairs) is favoured, all of which would be innocuous for mutation since they simulate the scheme present in the normal nonmutagenic G:C pair.

(b) For deprotonated alkylguanines, mutagenic pairing with thymine (viz. the N^3-MeG ^{0}:T, N^7-MeG ^{0}:T and O^6-MeG(a):T pairs) is energetically favoured over nonmutagenic pairing with cytosine (viz. the N^3-MeG ^{0}:C, N^7-MeG ^{0}:C and O^6-MeG ^{0}(a):C pairs). That both are nevertheless energetically allowed points to the possibility of competition between mutagenic and nonmutagenic pairing possibilities, the former being preferred (Abbott & Saffhill 1979).

(c) For protonated alkylthymines, the *anti* conformers only being considered, pairing with adenine *(viz. the O^2-MeT $^+$(a):A and O^4- MeT $^+$(a):A pairs) are favoured, both of which are*

Table IV.10 PM3 calculated data for pairing between normal DNA bases and the *syn* conformers of *O*-methylated bases (Cs symmetry imposed in all cases)

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$Ø_{nh}$ (deg)	θ_{nh} (deg)
O^6-MeG^0(s):T	-4.25	$N^1G:H^3T$	2.840	9.610	0.00	35.36
		$H^2G:O^2T$	1.836			
O^2-MeT^0(s):G	-6.40	$N^1G:H^3T$	2.428	9.958	0.00	60.91
		$H^2G:O^2T$	3.705			
O^4-MeT^0(s):G	-4.63	$N^1G:H^3T$	2.693	9.555	0.00	36.56
		$H^2G:O^2T$	1.813			

of no consequence for point mutations since they mimic the scheme present in the normal nonmutagenic T:A pair.

(d) For deprotonated alkylthymines (only *anti* conformers) pairing with guanine is energetically favoured over adenine. Though the pairing of O^4-MeT^0 with guanine is energetically more favorable than O^2-MeT^0, both of them are significant for inducing point mutations.

IV.4.7 Situation at biological pH

Out of the various favoured pairs mentioned in the last section, the question of which actually occur at biological pH depends upon whether the Watson-Crick proton in question (the N^1-proton for alkylguanines and the N^3-proton for alkylthymines) would be retained at this pH. The acidities of these protons have been linked to the pK_a of the base or nucleoside in question. This aspect has been treated in Chapter III, and the outcome of these theoretical predictions coupled with experimental pK_a values is that at biological pH (close to neutral), the methylated base systems would exist as their deoxynucleosides in the forms summarised as follows :

N^3-methylguanine : N^1-deprotonated (neutral) - N^3-MeG^0

O^6-methylguanine : N^1-deprotonated (neutral) - O^6-MeG^0

N^7-methylguanine : N^1-protonated (cationic) - N^7-MeG^+

O^2-methylthymine : N^3-deprotonated (neutral) - O^2-MeT^0

O^4-methylthymine : N^3-deprotonated (neutral) - O^4-MeT^0

The pairs that they would form at biological pH in the double- helical configuration thus include the following :

N^3-methylguanine : N^3-MeG^0:C (nonmutagenic) and N^3-MeG^0:T (mutagenic)

O^6-methylguanine : O^6-MeG^0(a):C (nonmutagenic) and O^6-MeG^0(a):T (mutagenic)

N^7-methylguanine : N^7-MeG$^+$:C (nonmutagenic)

O^2-methylthymine : O^2-MeT^0:G (mutagenic)

O^4-methylthymine : O^4-MeT^0:G (mutagenic)

The inference from the above predictions is that N^3- and O^6- methylguanines could pair in mutagenic fashion (although competing with nonmutagenic pairing), while N^7-methylguanine would be nonmutagenic. The pairing properties of cationic N^7-methylguanine with thymine have been examined here using the same methodology, and although the pairing energy is negative (-9.29 kcal/mol), the configurational parameters are such as to indicate that the potentially mutagenic pairing here would not be possible in the double-helical configuration chiefly owing to the presence of the N^1-proton. The values of the configurational markers are as follows :

R_{hb}: $N^1G:N^3T$ = 2.874 $\quad$ R_{nn}: 9.28 A^0

$N^2G:O^2T$ = 1.805 $\quad$ θ_{nh}: 48.53^0

O^2- and O^4-methylthymines also possess the ability to induce mutagenic pairing with guanine, while the possibility of non- mutagenic pairing with adenine is energetically unfavoured. The former is predicted to give a rather small pairing energy with guanine, while the latter pairs quite favourably with guanine, so that we may conclude that

O^4-methylthymine would be more efficient in inducing base-substitutions than O^2-methyl-thymine. These two methylated bases differ from N^3- and O^6-methylguanines in that here the chances for mutagenic pairing competing with non-mutagenic pairing weighs heavily in favour of the former, so that the alkylthymines (especially O^4-methylthymine) may be expected to be much more specific in their inducement of base- substitutions than the alkylguanines.

Fig. IV.6 portrays the base-pairs which be expected to predominate for each species at biological pH : (a) N^7- methylguanine (cationic) with cytosine, (b) N^3-methylguanine (deprotonated) with thymine, (c) O^6-methylguanine (deprotonated) with thymine, (d) O^2-methylthymine (deprotonated) with guanine (weak pairing), and (e) O^4-methylthymine (deprotonated) with guanine.

IV.6 Mutagenic Potential of C^8-oxidised Purines

The attack of reactive oxygen species on DNA represents an important mechanism by which exogenous and endogenous agents may inflict damage on cells. Oxygen-derived free radicals, produced by ionising radiation (Ward 1988; Huttterman *et al* 1978) or cellular metabolism (Cerutti 1985; Diguiseppi & Fridovich 1984), generate apurinic/apyrimidinic sites or base modification in DNA; both types of lesions may have mutagenic and carcinogenic effects (Breimer 1988).

Purine residues, both deoxyguanosine and deoxyadenosine, in DNA are hydroxylated by hydroxyl radicals, resulting in the formation of 8-oxoadenine (8OA) and 8-oxoguanine (8OG) (Kasai & Nishimura 1984a-c; Dizdaroglu 1985). . The contribution of these species to mutagenesis is currently the subject of a great deal of interest (Kouchakdjian *et al* 1991; Leonard *et al* 1992).

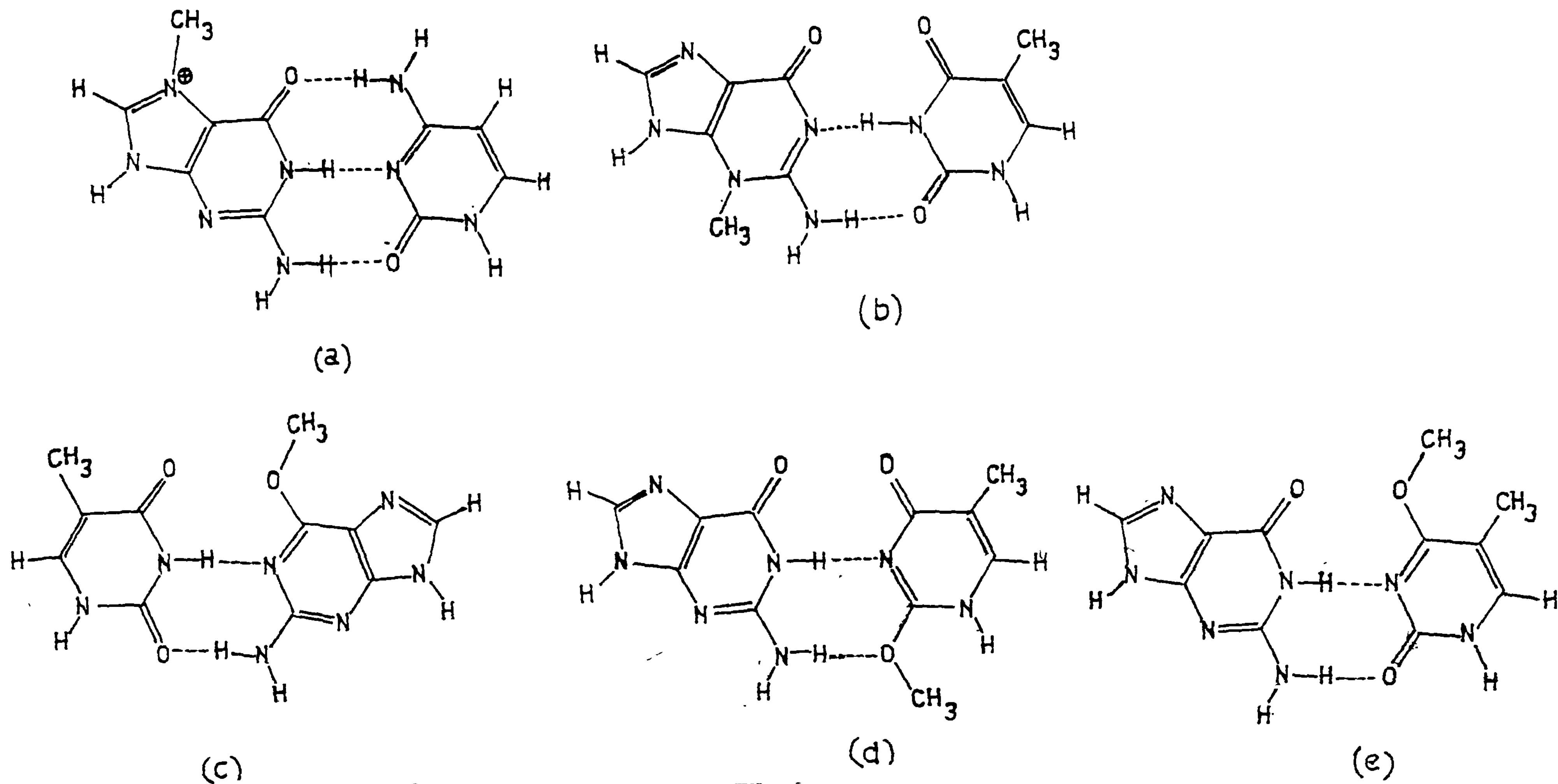


Fig. IV.6 : Base-pairs proposed to be predominant at biological pH

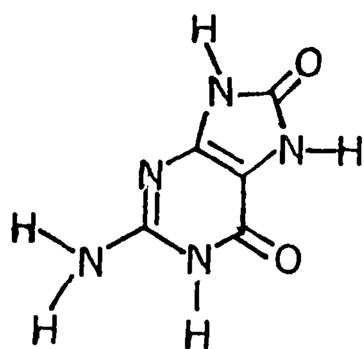
IV.6.1 Tautomerism in 8-oxidised purines

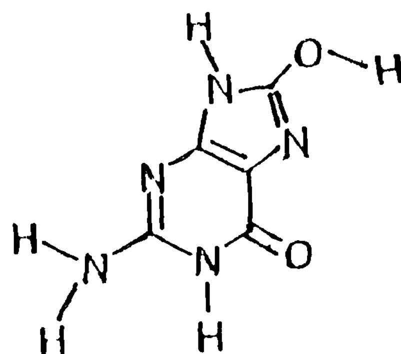
In their monomeric form, these modified bases can adopt several tautomeric structures (Aida & Nishimura 1987). Fig.IV.7 shows various possible tautomers of 8OA and 8OG. It isreported that under physiological conditions, the 6,8-dioxo species (designated as 8-oxo-7H-dG) predominates (Culp *et al* 1989) over other possible tautomers. NMR spectral evidence by Kouchakdjian *et al* (1991) indicates the existence of about 15% of minor tautomers. Therefore, the significance of minor tautomers, particularly the enol forms, in base-mispairing and induced mutagenesis cannot be ruled out. In view of this, it would be of interest to compare the energetics of various tautomers and the possible existence of the rarer tautomeric species.

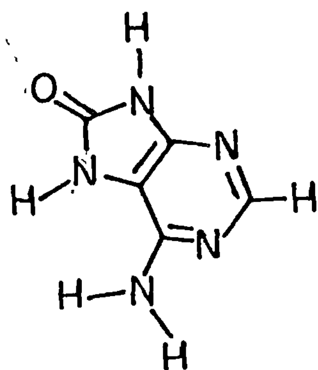
The PM3 SCF-MO fully optimised heats of formation of 8- oxodeoxyguanosine and 8-oxodeoxyadenosine and their enol counterparts are presented in Table IV.11. The tautomeric transition energies (E_{tt}) from the native keto to the enol form have been presented for both 8-oxo-deoxyguanosine and 8-oxo-deoxyadenosine in Table IV.12. The high E_{tt} values, 12.74 kcal/mol for 8-oxo-dG and 11.86 Kcal/mol for 8-oxo-dA, indicate that the tautomeric transition is a very unlikely route for both purines. Moreover, initial molecular modelling studies here indicate that for both bases, the enol form is not found to form any base-pairs with other DNA bases within the reasonable Watson- Crick configuration. On both counts, it is the keto (oxo) form which emerges as the species of relevance for consideration here.

IV.6.2 Rotational barriers in 8-oxidised purines

These 8-oxidised purines do not exert their mispairing properties in Watson-Crick fashion. Both 8-oxo-deoxyguanosine (8OdA) and 8- oxo-deoxyadenosine (8OdG) must adopt *syn*







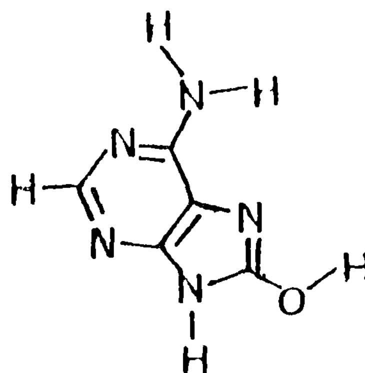


Fig. IV.7 : Various possible tautomers of 8-oxo-Adenine and 8-oxo-Guanine

Table IV.11 PM3 calculated Heats of formation of keto and enol forms of deoxyguanosine and deoxyadenosine (in Kcal/mol)

system	ΔH_f
Deoxynucleosides	
dG-8-oxo	-179.67
dG-8-ol	-166.93
dA-8-one	-131.00
dA-8-ol	-119.14

Table IV.12 PM3 data of Keto-enol tautomeric energetics of 8-oxo modified deoxy guanosine and 8-oxo deoxyadenosine (in Kcal/mol)

transition	E_{tt}
dG-8-one ---> dG-8-ol	12.74
dA-8-one ---> dA-8-ol	11.86

E_{tt} = Energy of tautomeric transition between keto and enol forms

Table IV.13 PM3 calculated data for base pairs involving the normal and tautomeric forms of nucleic acid bases

System	E_p (kcal/mol)	H-bond	R_{hb} (A)	R_{nn} (A)	$\emptyset_{nh}$ (deg)	θ_{nh} (deg)
G:C	-11.79	$O^6G:H^4C$	1.804	9.044	-7.0	57.68
		$H^1G:N^3C$	1.780			
		$H^2G:O^2C$	1.840			
A:T	-6.10	$O^4T:H^6A$	1.823	9.125	6.64	42.46
		$H^3T:N^1A$	1.778			
8OG(s):A(a)	-4.82	$O^6G:N^4A$	1.824	9.071	4.75	61.94
		$N^7G:N^1A$	1.773			
8OA(s):G(a)	-8.57	$O^8A:N^1G$	1.799	8.963	-38.78	62.73
		$N^7A:O^6G$	1.799			

conformations (with respect to the sugar moiety around the glycoside bond) in order to facilitiate transversional mutagenic events via purine-purine mismatches. The bulky oxygen atom at the C^8 position is expected to force the base to adopt the *syn* confomation from its normal *anti* form. Since these 8-oxidised bases in the *syn* rotamer are believed to be of particular mutagenic significance, the rotational barrier to adopt *syn* from *anti* position would be expected to play a key role in conferring mutagenic properties to these bases, besides the necessity of being structurally polymorphic with normal Watson-Crick base pairs so as to escape detection by the proof-reading enzymes. In order to evaluate and understand the underlying energies and the feasibility of 8- oxo-purines to adopt *syn* conformers, the *anti* to *syn* rotational barriers are estimated for both 8OdG and 8OdA.

Figs. IV.8 and IV.9 represent the PM3 computed rotational barriers for 8OdG and 8OdA respectively, with heat of formation plotted against the O1'-C1'-N9-C4 torsional angle ϕ_r. In both cases, two main minima are located, the first at ϕ_r about 67-70° and corresponding to the *anti* conformer. The second minimum is about 2.3 kcal/mol higher than the first for 8OdA, and about 1 kcal/mol for 8OdG; this minimum is located at ϕ_r equal to 260° for 8OdG and 270° for 8OdA. For both cases, the *anti* to syn rotational barrier is estimated as approximately 5 kcal/mol. This barrier is not insurmountable under favourable conditions of base-pairing (which can supply the energy required), thereby allowing for the possibility that these syn conformers of the two 8-oxopurines could be involved in purine:purine base-mismatches.

IV.6.3 Base pairing properties of 8-oxidised purines

Recent studies on the template properties of 8-oxo purines have established that translesional synthesis can proceed past 8-oxo- 7H-dG in primed template reactions

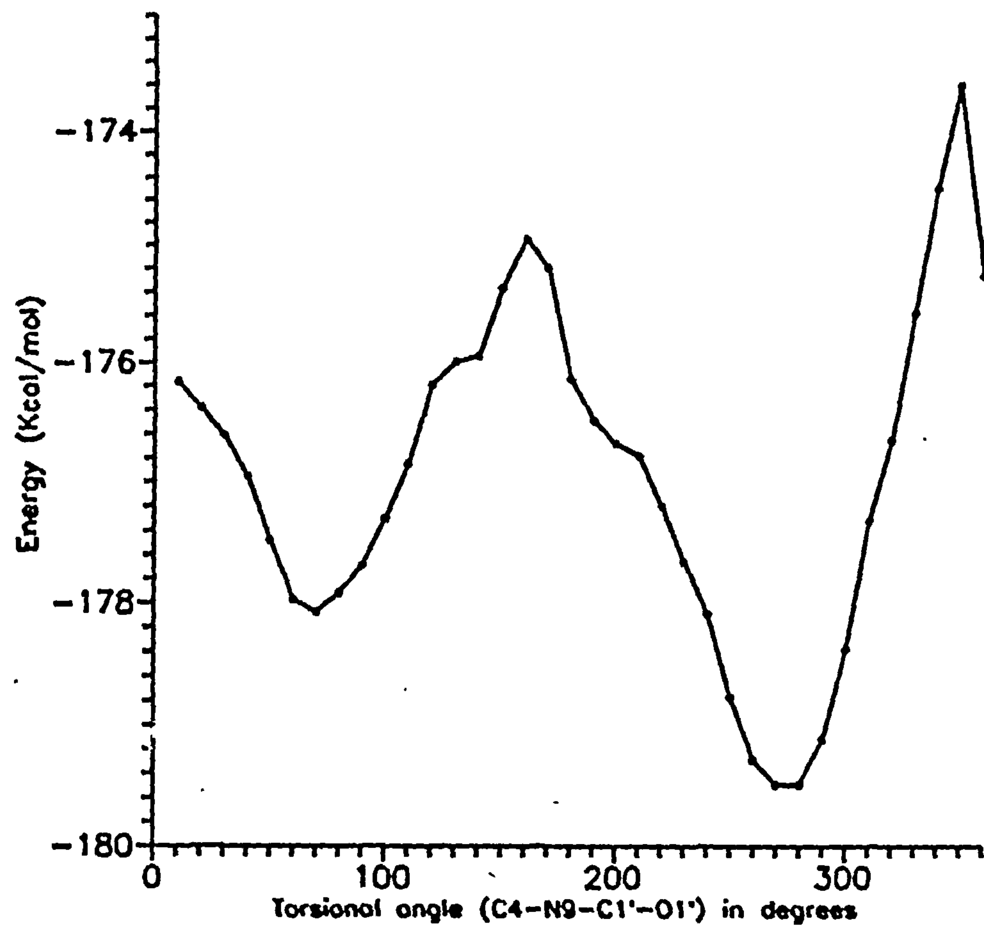


Fig. IV.8 : PM3 computed rotational barrier from anti to syn configuration for 8-oxo-Deoxyguanosine

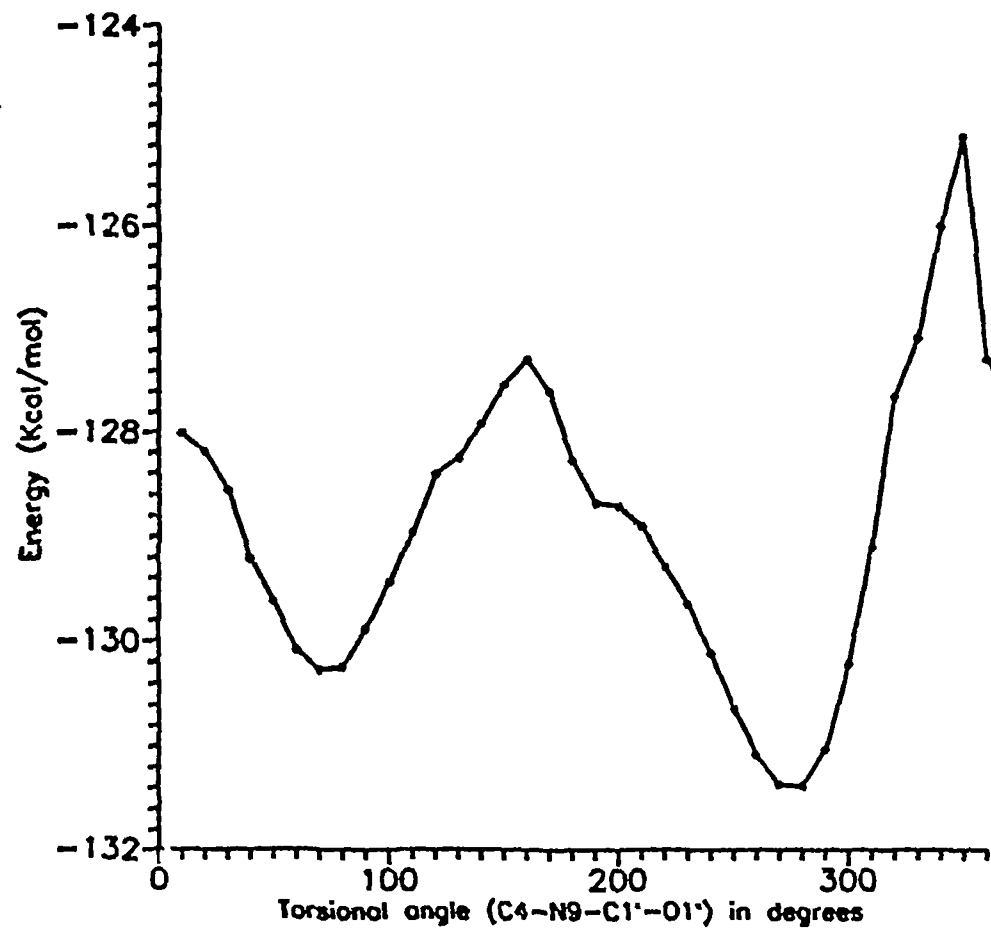
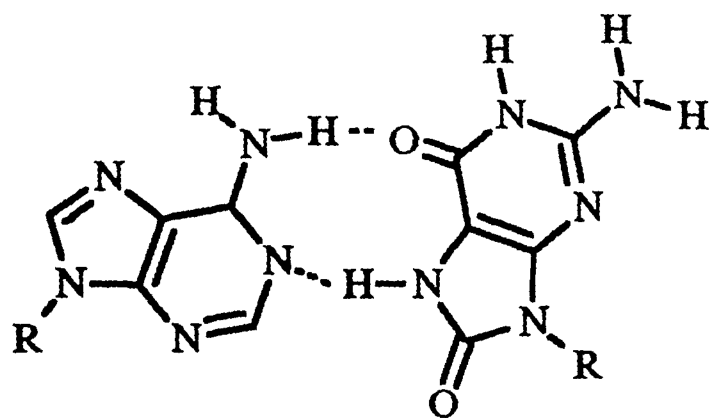


Fig. IV.9 : PM3 computed rotational barrier for anti to syn conformer for 8-oxo-Deoxyadenosine

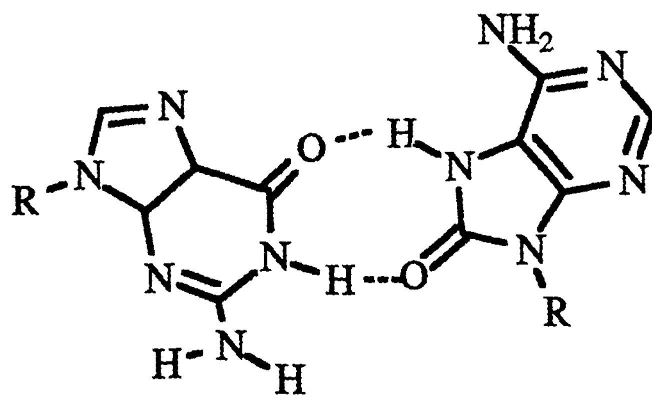
catalysed by DNA polymerase, in which case dA and/or dC is inserted opposite the lesion. The 8-oxo-7H-dG:dA pair is readily extended by DNA polymerase and does not appear to be subject to the editing function of this enzyme (Shibutani *et al* 1991). The presence of 8OG base in genomic DNA can lead to a G to T transversion mutation via an intermediate O8G:A base-pair.

NMR and X-ray studies show that the 8OG:A base-pair assumes a *syn:anti* configuration stabilised by two interbase hydrogen bonds (Kouchakdjian *et al* 1991; Leonard *et al* 1992). In addition to having a reasonable thermodynamic stability, the O8G:A base-pair is pseudosymmetric about the glycosidic bonds and is therefore structurally similar to a Watson-Crick base pair. The similarity is observed to be particularly striking in the minor groove where the 8-oxygen atom of O8G lies in the position that would be occupied by the 2-oxygen atom of the thymine base in an AT base pair. Thus, it is proposed that the O8G:A base pair is not readily recognised by proof-reading enzymes.

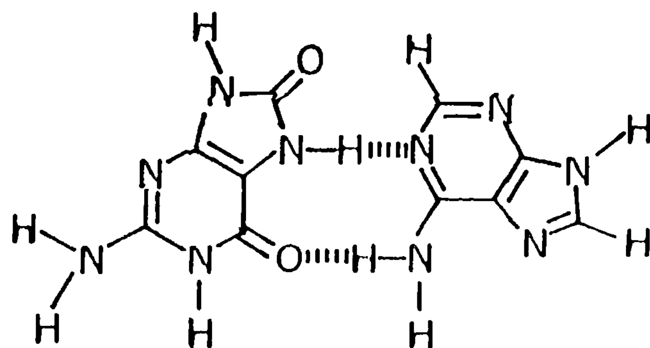
In contrast to O8G, 8-oxoadenine (O8A) is not particularly mutagenic. X-ray studies on a dodecanucleotide duplex have shown that the most likely alternative base-pair G:O8A is asymmetric, and is similar to a purine-pyrimidine mismatch (Leonard *et al* 1992). It is therefore likely to be an easy target for repair enzymes. As it was seen in Secs. IV.6.1 and IV.6.2,both 8OdG and 8odA can adopt the *syn* conformers through thermodynamically feasible rotational barriers. It is interesting to understand the base- pairing characteristics of these two modified bases to understand why 8OdG is mutagenic whereas 8OdA is not. Table IV.13 presents the PM3 calculated data for these two base-pairs. It is immediately obvious that 8-oxo-guanine pairs in *syn* fashion with adenine(anti) within a reasonably Watson-Crick configuration. On the other hand, 8-oxo-adenine in *syn* form does not base-pair with guanine(anti) within the permissible Watson-Crick configuration. The



$8OG_s:A_a$



$8OA_s:G_a$



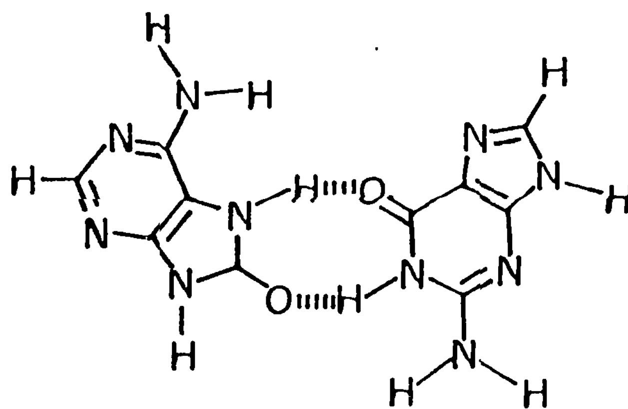
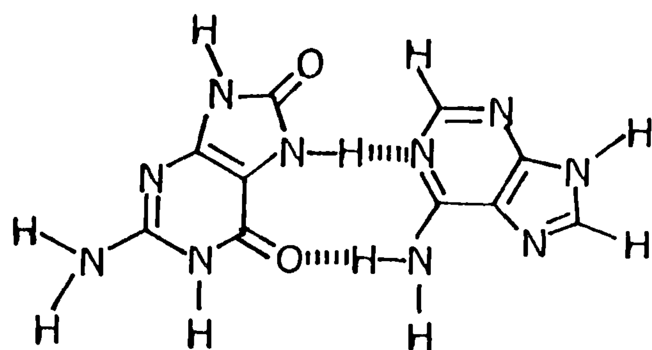


Fig. IV.10 : Possible base-pairing schemes between 8-oxo-Guanine (syn): Adenine (anti) and 8-oxo-Adenine (syn) : Guanine (anti)



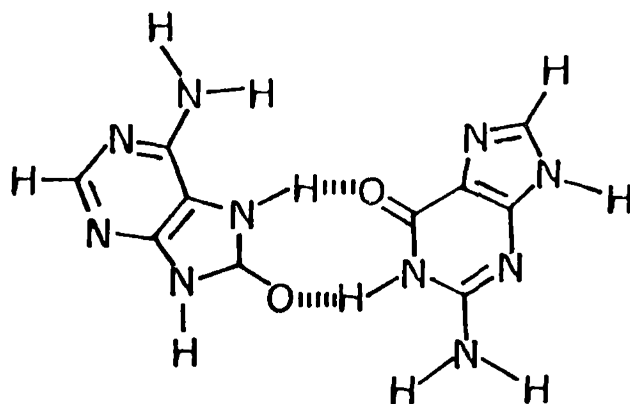


Fig. IV.10 : Possible base-pairing schemes between 8-oxo-Guanine (syn): Adenine (anti) and 8-oxo-Adenine (syn) : Guanine (anti)

strong deviation from double-helical nature here (having a marked propellor twist with ϕ_{nh} equal to -38.78°) may be one factor why 8-oxo-dA is apparently not mutagenic. **Fig. IV.10** portrays the base-pairs O8G(*syn*):A(*anti*) and O8A(*syn*):G(*anti*).

IV.7 Concluding Remarks

The above findings of this Chapter may be summarised as follows:

1. Studies pertaining to the base-pairing potential of tautomeric speices with normal bases points to the feasibility of these pairs occuring in nature in response to the transient or minute existence of the tautomeric forms. The results of these calculations may be taken to serve as a model for the spontaneous induction of base-pair transitions in nature.

2. The consensus from the calculations on the base-pairing properties of anions (deprotonated species) is that the anions of the bases guanine and thymine would pair favourably only with thymine and guanine respectively in the double-helical configuration at the appropriate conditions of high pH. The normal G:C and A:T base-pairs would be stabilised at high pH.

3. The high pairing energies resulting from pairing of protonated species indicate that the existence of these base-pair motifs at low acidic pH would stabilise the double-helix.

4. The criteria proposed for successful induction of base- substitutional mutations, viz. absence of the Watson-Crick proton and appropriate conformation of the methyl group in O-methylated bases, are seen to be borne out by the results of these base- pairing calculations.

5. Application of theoretical and experimentally-based consi- derations on the acidities of the Watson-Crick protons in question lead to a correct assignment of mutagenic potential for N^7- and O^6-alkylguanines and the O-methylated thymines, furnishing also a prediction

that N^3-methylguanine would be mutagenic.

6. PM3 studies on 8-oxoguanine and 8-oxoadenine reveal that they can pair in the easily adoptable *syn* conformers through purine- purine mismatches, where 8OG would be mutagenic while the corresponding pair with 8OA, would not be allowed in the usual double-helical configuration, and thus be non-mutagenic.

to serve as a model for the spontaneous induction of base-pair transitions in nature.

2. The consensus from the calculations on the base-pairing properties of anions (deprotonated species) is that the anions of the bases guanine and thymine would pair favourably only with thymine and guanine respectively in the double-helical configuration at the appropriate conditions of high pH. The normal G:C and A:T base-pairs would be stabilised at high pH.

3. The high pairing energies resulting from pairing of protonated species indicate that the existence of these base-pair motifs at low acidic pH would stabilise the double-helix.

4. The criteria proposed for successful induction of base- substitutional mutations, viz. absence of the Watson-Crick proton and appropriate conformation of the methyl group in O-methylated bases, are seen to be borne out by the results of these base- pairing calculations.

5. Application of theoretical and experimentally-based consi- derations on the acidities of the Watson-Crick protons in question lead to a correct assignment of mutagenic potential for N^7- and O^6-alkylguanines and the O-methylated thymines, furnishing also a prediction that N^3-methylguanine would be mutagenic

6. PM3 studies on 8-oxoguanine and 8-oxoadenine reveal that they can pair in the easily adoptable *syn* conformers through purine- purine mismatches, where 8OG would be mutagenic while the corresponding pair with 8OA would not be allowed in the usual double-helical configuration, and thus be non-mutagenic.

CHAPTER V

STABILITYAND REPAIR OF MODIFIED DNA COMPONENTS

V.1 Biochemical Scope of DNA Nucleoside Repair and Stability

Chemical carcinogenesis by alkylating agents being attributed to mutations induced by alkylation of DNA, it may be expected that an error-free repair of DNA damaged by these compounds would serve an important role as a deactivating path which detracts from the carcinogenic effect, or may even act to totally prevent the induction of tumours. Conversely, an error-prone repair process might contribute towards the process of initiation after exposure to the chemical carcinogen (Modrich 1994; Sancar 1994).

Lesions chemically induced in DNA are generally subject to removal or repair processes *in vivo*. Genetic alterations induced by endogenous mutations or environmental carcinogens are usually deleterious, and need to be removed to ensure health and survival of the organism. The harmful nature of most gene mutations is a problem for the biologist who seeks to reconcile the theory of evolution with genetics by attributing phylogeny and the origin of new species to genetic alterations. Most of the endogenous events include depurination reactions and the deamination of bases, as well as replication errors resulting from repair infidelity in DNA polymerases and their associated proteins. The possibility of base-substitutional mutation following mispairing with rare (transient) tautomeric forms of the DNA bases has not yet been substantiated satisfactorily, so that the repair process

associated with it is also correspondingly not clear.

Clear evidence exists for the processes associated with repair of various types of chemically induced lesions. Nucleotide or base excision is performed enzymatically for a number of lesions, including the thymine dimer formed through action of UV light. Base excision has been noted for many products of alkylation at ring nitrogens of DNA base residues, while alkyltransferase activity serves as a repair mode for products of DNA base alkylation at exocyclic oxygens. The possible effect of repair on carcinogenesis (serving as a deactivating pathway) has been suggested by a number of studies which demonstrated the connection between tumour incidence and the persistence of certain O-alkylated lesions in tissues or organs treated with alkylating agents and N-nitroso compounds. These studies were conducted through single dose, multiple dose and chronic exposure regimes on rodents (Goth & Rajewsky 1974; Kleihues & Margison 1974; Nicoll *et al* 1975; Margison *et al* 1976; Hodgson *et al* 1980; Margison & Kleihues 1975; Cox & Irving 1977). It is now well established that defects in DNA repair do contribute directly to carcinogenesis (Modrich 1994; Sancar 1994).

Apart from enzymatic removal of chemically induced lesions, there also exists the possibility of removal through spontaneous mechanisms like hydrolysis. Here, the *in vivo* hydrolysis has been subjected to much study using in vitro models. Glycosidic hydrolysis (also called glycolysis or deglycosylation) of purine nucleosides (depurination) is a fundamental process that has been studied both *in vivo* and *in vitro* (Singer *et al* 1978). Glycosidic hydrolyis of interstrand nucleotides yields apurinic sites (AP). The principal biological effect of such AP site-forming mutations is error-prone DNA synthesis.

Due to the relative instability of the N-glycosidic bond, such abasic sites in DNA many occur spontaneously with a relatively high frequency (estimated to be 10^4 lesions per

cellular cycle). It is believed that the deglycosylation reaction is much easier for purines than for pyrimidines since protonation of the N^7 atom of guanine and the N^3 atom of adenine labilises the glycosidic bond to apparently give a highly unstable cyclic carboxonium ion which leads to loss of 2-deoxyribose. This hydrolytic process is markedly accelerated by chemical modifications of the nucleic bases such as alkylating agents and other *carcinogens, and by physical agents such as UV and r radiation. In cells,* abasic sites are also produced enzymatically during repair of abnormal or modified bases. Here, the alterations in the base are recognized and removed by highly specific N-glycosylases, generating abasic sites as intermediates in the repair system. If not repaired, the abasic site, which is a noninformative lesion, may promote more or less random misincorporation of nucleotides *in vitro,* which can be mutagenic or even cytotoxic (Friedberg 1985).

The significance of these spontaneous mutations and the formation of abasic sites for mutagenesis, and possibly carcinogenesis, has led many experimental groups to be actively involved in detailed kinetic studies of the relative stabilities of the DNA nucleosides and alkylated nucleosides. These have also been accompanied by the more biologically relevant studies on the enzymatic repair processes involved in removal of the concerned DNA lesions.

Regarding spontaneous loss of alkyl groups from alkylated lesions, there does not appear to be any evidence that this could occur under biological conditions of temperature and pressure. Under *in vitro* conditions of elevated temperature and pH ranges far removed from the biological one, it is clear that dealkylation does occur, as will be dealt with in detail in Sec. V.2.1 below.

V.2 Experimental Aspects of DNA Nucleoside Stabilities

Much labour has been expended towards the study of loss, removal, repair and stability of free and modified DNA nucleosides, which is briefly described below :

V.2.1 In vitro stability of alkylated nucleosides

Alkylated DNA and RNA nucleosides possess varying degrees of in vitro stability, depending upon the alkylation site, the pH and the temperature. Stability here refers to the ability of an alkylated nucleoside to retain its original chemical structure, which includes the modified base residue and the sugar moiety together. The modified nucleoside may be altered *in vitro* by two chief ways - loss of the alkyl group (dealkylation) or loss of the intact base residue by breaking of the N-glycosidic linkage (deglycosylation). In practice, either one may follow the other, resulting in a net operation of both processes leading to generation of the free base and sugar. The liability of alkylated nucleosides to *in vitro* alteration in this manner has been studied primarily in the context of nucleic acid hydrolysis under varying conditions of pH and temperature.

There are noticeable differences in the *in vitro* lability of the alkyl group among alkylated nucleosides, depending upon the alkylation site and pH. It has long been known that N-alkylated purine nucleosides (alkylated at the N^3-G, N^7-G, N^3-A and N^7-A positions) retain their alkyl groups intact in the base residues following depurination under strongly acidic conditions of DNA hydrolysis (Lawley & Thatcher 1970; Frei & Lawley 1975; Magee & Farber 1962; Swann & Magee 1971). However, O-alkylated base residues lose their alkyl groups under strongly acidic conditions (Frei *et al* 1978; Singer *et al* 1978), which was one factor held responsible for the inability of earlier workers to detect these products of DNA alkylation by alkylating agents and N-nitroso compounds. Milder conditions of DNA

hydrolysis permitted retention of these O-alkyl groups (Kleihues & Margison 1971; O'Connor *et al* 1972, 1973), while enzymic hydrolysis at neutral pH enabled substantially complete retention of the O-alkyl group in the liberated base residue (Singer 1976; Beranek *et al* 1980). All this may be interpreted as meaning that N-alkyl groups are less labile than the corresponding O-alkyl groups, while raising of pH serves to strengthen the binding between the O-alkyl groups and their alkylation sites.

Loss of the alkylated base residue intact following hydrolytic cleavage of the glycosidic linkage has been a subject of much study. Alkylation of DNA purine nucleosides at their ring nitrogens facilitiates depurination even at neutral pH (Lawley, 1973). The rate of loss of methylated purine deoxynucleosides (i.e. liberation of the intact methylated base residue following deglycosylation) has been shown to follow the order : N^7-MeA > N^3-MeA > N^7-MeG > N^3-MeG. In contrast, O^4-alkylated DNA and RNA pyrimidine nucleosides resemble the unmodified nucleosides in possessing a stable glycosyl bond at neutral pH, which has also been found to be true for O^6-alkyldeoxyguanosines (Singer *et al* 1978; Margison & O'Connor 1979). Alkylation at the O^2-site of pyrimidine deoxynucleosides, however, has been found to have a destabilising effect on the glycoside linkage, releasing the alkylpyrimidine intact at neutral pH (Singer *et al* 1978). The unmodified purine and pyrimidine nucleosides, of course, are associated with a minimal tendency towards deglycosylation at neutral pH. Thus, it may be summarised that N-alkylated purine nucleosides with O^2-alkyl-pyrimidine nucleosides are noted for the greater lability of the glycosidic linkage at neutral pH as compared to the unmodified nucleosides, while O^4-alkylpyrimidine nucleosides and O^6-alkyl-deoxyguanosines are characterised by a relative stability of the glycoside bond at neutral pH. Higher temperature and lower pH (acid hydrolysis) does, however, permit for cleavage of the N-glycoside bond of these

O-alkylated nucleosides.

An important observation is that under the conditions of pH and temperature used in the studies of Singer *et al* (1978), the deglycosylation reaction precedes dealkylation, which means that hydrolysis of an alkylated DNA nucleoside results firstly in cleavage of the N-glycoside bond in all cases, releasing the alkylated base and sugar. Further reaction can then take place, depending upon the conditions, leading to loss of the alkyl group and generation of the free base.

V.2.2 In vivo repair mechanisms for alkylated nucleosides

Alkylated DNA nucleosides have been noted for the remarkable selectivity of choice in their mechanisms for repair *in vivo*. The repair enzymes and their mode of action are quite unique for each separate class of alkylated nucleoside. We may distinguish **two** chief modes for repair of alkylated lesions in DNA - one resorting to removal (excision) of the intact base residue by cleavage of the glycoside bond; and the other acting by simply removing only the alkyl group, leaving behind the unmodified DNA nucleoside. It has been noted that nucleosides alkylated at base ring nitrogens are subject to repair by a glycosylase enzyme which removes the excised base moiety with alkyl group intact, and leaves behind an apurinic site. On the other hand, nucleosides alkylated at the exocyclic oxygen sites are generally subject to repair by through removal of the O-alkyl group alone, leaving behind the repaired and unmodified nucleoside.

Evidence exists for the role of mammalian DNA glycosylases in the repair of nucleosides alkylated at the N^3-A, N^7-A, N^3-G and N^7-G positions (Laval 1977; Chetsanga & Lindahl 1979; Laval *et al* 1981; Pierre & Laval 1980; Ishiwata & Oikawa 1979; Brent 1979; Cathcart & Goldthwait 1981). The glycosylase enzyme 3-methyladenine DNA

glycosylase II (AlkA) has been found to catalyze the removal of 3-methyladenine and 7-methylguanine, as well as several other minor DNA lesions, through glycosidic bond hydrolysis (Nakabeppu *et al* 1984; Lindahl *et al* 1988). These enzymes act by cleaving the destabilised glycosidic linkage yielding the free alkylated purine and the sugar (Dodson *et al* 1994). Glycosylase base excision was, however, ruled out as a repair mechanism for various O-alkylated DNA nucleosides (Karran *et al* 1979; Goth-Goldstein 1980; Cleaver & Kaufman 1980).

Repair of O^6-alkyldeoxyguanosines, reviewed by Yarosh (1985) is known to be mediated by an alkyltransferase enzyme which removes only the O-alkyl group by transferring in on to a cysteine residue in the enzyme system. This mode of repair has been characterised for bacterial (*E.coli*) and various mammalian cell lines (Schendel & Robins 1978; Olsson & Lindahl 1980). The scope of this mechanism extends beyond the methyl case to include the ethyl, n-propyl, n-butyl, isopropyl, 2-hydroxyethyl and chloroethyl groups as well (Pegg *et al* 1983; Sedgewich & Lindahl 1982; Robins *et al* 1983; Kohn 1972; Pegg *et al* 1984). The repair reaction irreversibly produces the S-methylated cysteine residue on the enzyme in stoichiometric yield, and cannot be labelled as truly catalytic in nature. For O-alkylpyrimidines, the repair process is markedly slow, which is why this class of alkylated lesions has been earmarked as being of special significance for mutagenesis and cancer initiation (Singer 1986). A human repair capacity specific for O^4-ethyldeoxythymidine has been identified and partially characterised (Wani *et al* 1990).

V.2.3 Relevance for Repair Mechanisms

It may be noted here that a remarkable correlation exists between *in vitro* stabilities of alkylated nucleosides and their *in vivo* repair mechanisms. In general, those nucleosides

which are prone to deglycosylation *in vitro* are the ones which are also repaired by glycosylase excision of the base moiety. Likewise, those alkylated bases which are more susceptible to hydrolytic loss of alkyl group *in vitro* are the ones which in their nucleoside form lend themselves to repair via the alkyltransferase enzyme system. This correlation served as the basis for the theoretical study of Lyngdoh (1992) on the chemical factors influencing the precise choice of repair mechanism among the different classes of alkylated nucleosides. It was found that N^7-alkylguanine systems (repaired by N-glycosylase) were predicted as chemically more prone towards depurination than the O^6-methylguanine and O^4-thymine systems, while the O-alkylated bases were more chemically susceptible to loss of alkyl group than the N^7-alkylguanines, reflecting on the choice of alkyltransferase repair for the O-alkylated base systems. It may be noted here that perhaps a steric factor operates too, since the systems prone to deglycosylation rather than dealkylation have the alkyl group attached to a ring nitrogen, while the systems repaired through alkyltransferase enzymes have the alkyl group attached to an exocyclic oxygen atom.

V.4 Deglycosylation in Purine and Pyrimidine Nucleosides

V.4.1 Approach to study of deglycosylation

The reaction of deglycosylation was dealt with in this study by considering the following :

(1) The effect of protonation, ie. the pH factor

(2) The effect of site of alkyl group attachment

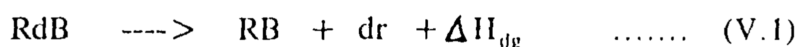
(3) The effect of N^7-protonation upon depurination

The reaction was studied by considering the following alkylated nucleosides in their protonated and deprotonated forms - N^7- methyldeoxyguanosine (N^7-MedG), N^3-methyl

deoxyguanosine (N^3- MedG), O^6-methyldeoxyguanosine (O^6-MedG), O^2-methyldeoxythymidine (O^2-MedT) and O^4-methyldeoxythymidine (O^4-MedT). The proton- ated and deprotonated forms are denoted by suffixing a plus (+) sign and a zero (0) sign respectively to the symbols above, eg. as for N^7-methyldeoxyguanosine, the protonated form is represented as N^7-MedG $^+$, and the deprotonated form as N^7-MedG ^{0}. Besides these alkylated nucleosides, the free nucleosides deoxyguanosine (dG) and deoxythymidine (dT) were also studied. The effects of protonation at the N^7-purine sites was also studied by considering the N^7-protonated deoxyguanosine (N^7-HdG) and the N7-protonated deoxyadenosine (N^7-HdA) systems.

Facility and susceptibility towards deglycosylation was gauged from a number of theoretical indices calculated using the AM1 SCF-MO method on fully optimised geometries for the various nucleoside, base and sugar systems. Figs. V.1 to Fig V.4 illustrate various deglycosylation and demethylation reaction schemes studied for both protonated and deprotonated alkylated deoxynucleosides. The indices are as follows:

Susceptibility towards cleavage of the glycosyl bond was estimated firstly from the intrinsic strength of the bond as obtained from the gas-phase heat of dissociation H_{dg} as given in Eqn (V.1) below:



RdB ----> RB + dr + ΔH_{dg} .... (V.1)

where **RdB** represents the alkylated deoxynucleoside, **RB** the alkylated base as immediately formed after dissociation and **dr** the sugar moiety in the form immediately obtained following the bond cleavage. Note that **RB** and **dr** are not the stable products of the bond dissociation.

Secondly, proneness towards deglycosylation was also estimated from the heat of cleavage ΔH_{ig} of the glycosyl bond when assisted by a water molecule attacking the C1'-atom of the

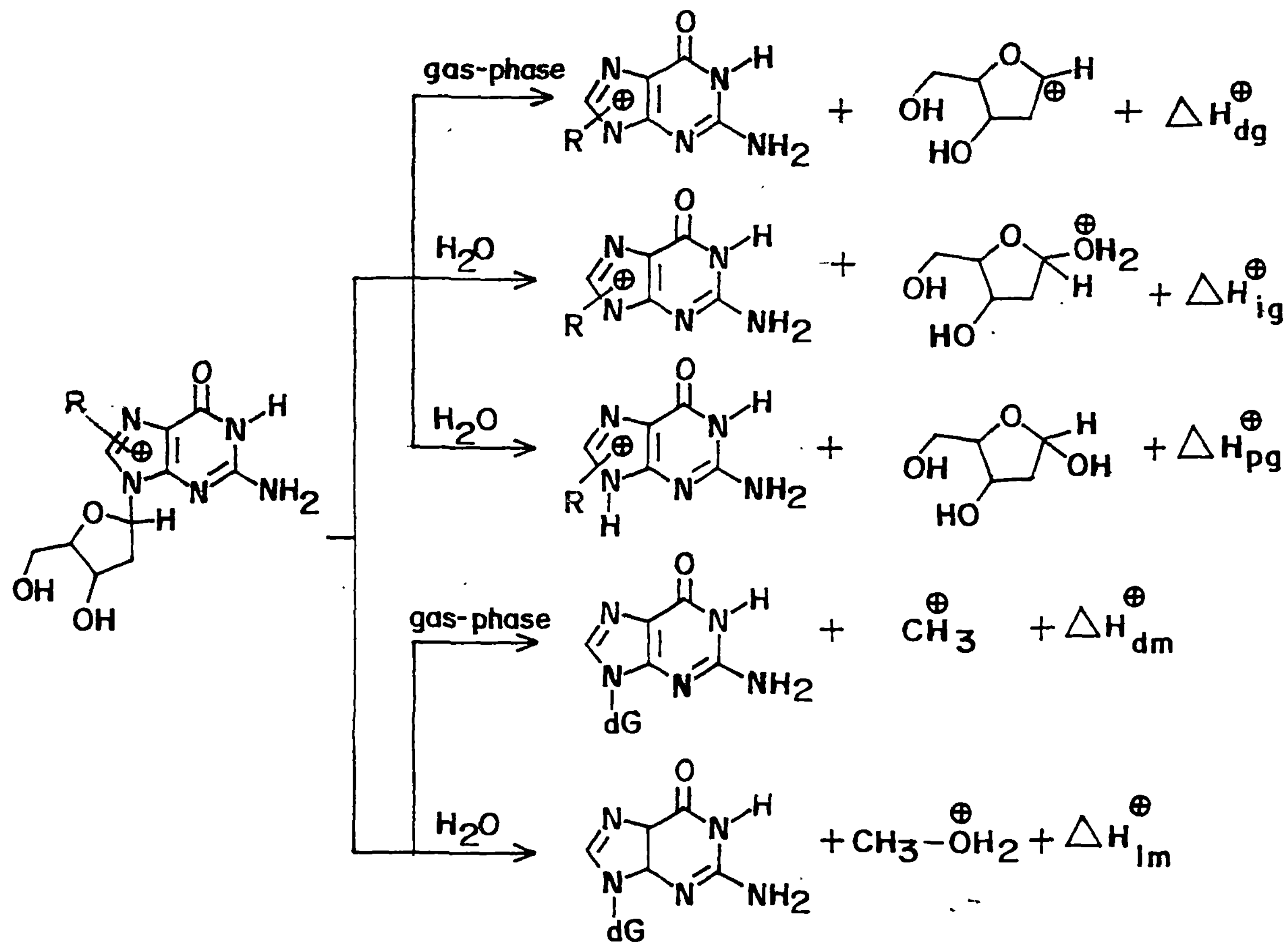


Fig. V.1: Deglycosylation and demethylation pathways for protonated alkyalted deoxyguanosines studied

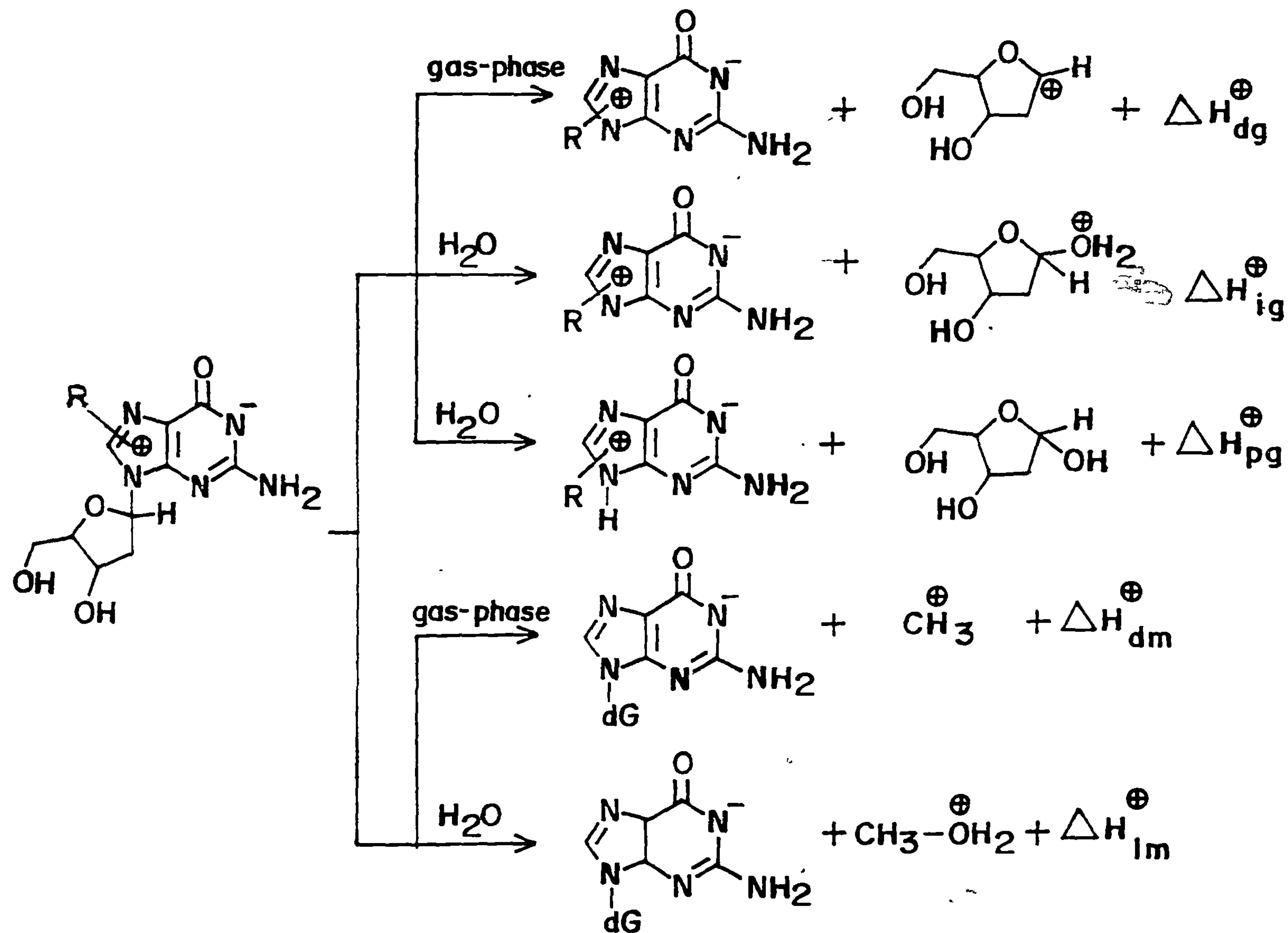


Fig. V.2: Deglycosylation and demethylation pathways for deprotonated alkyalted dexoyguanosines studied

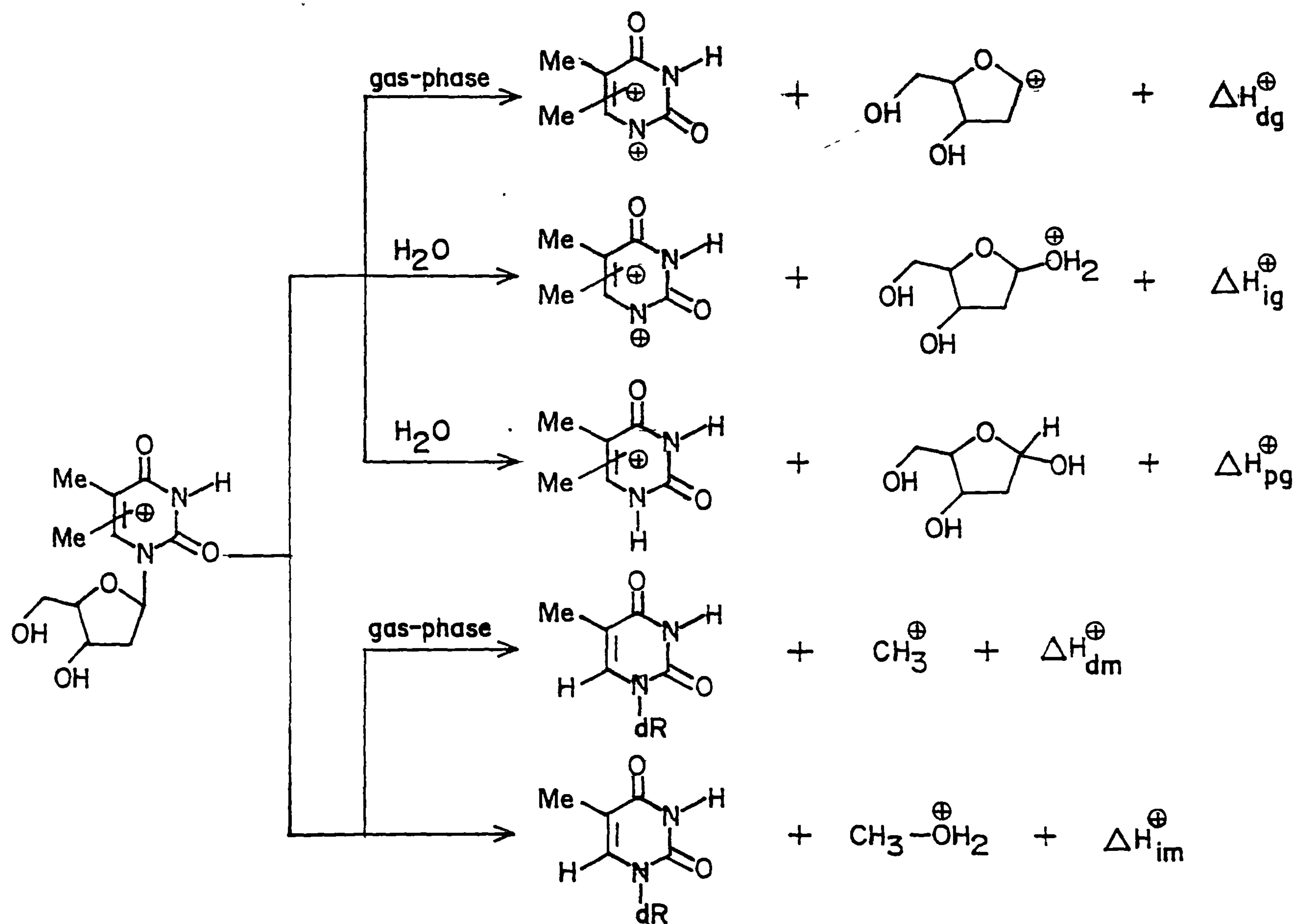


Fig. V.3: Deglycosylation and demethylation pathways for protonated alkylated deoxythymidines

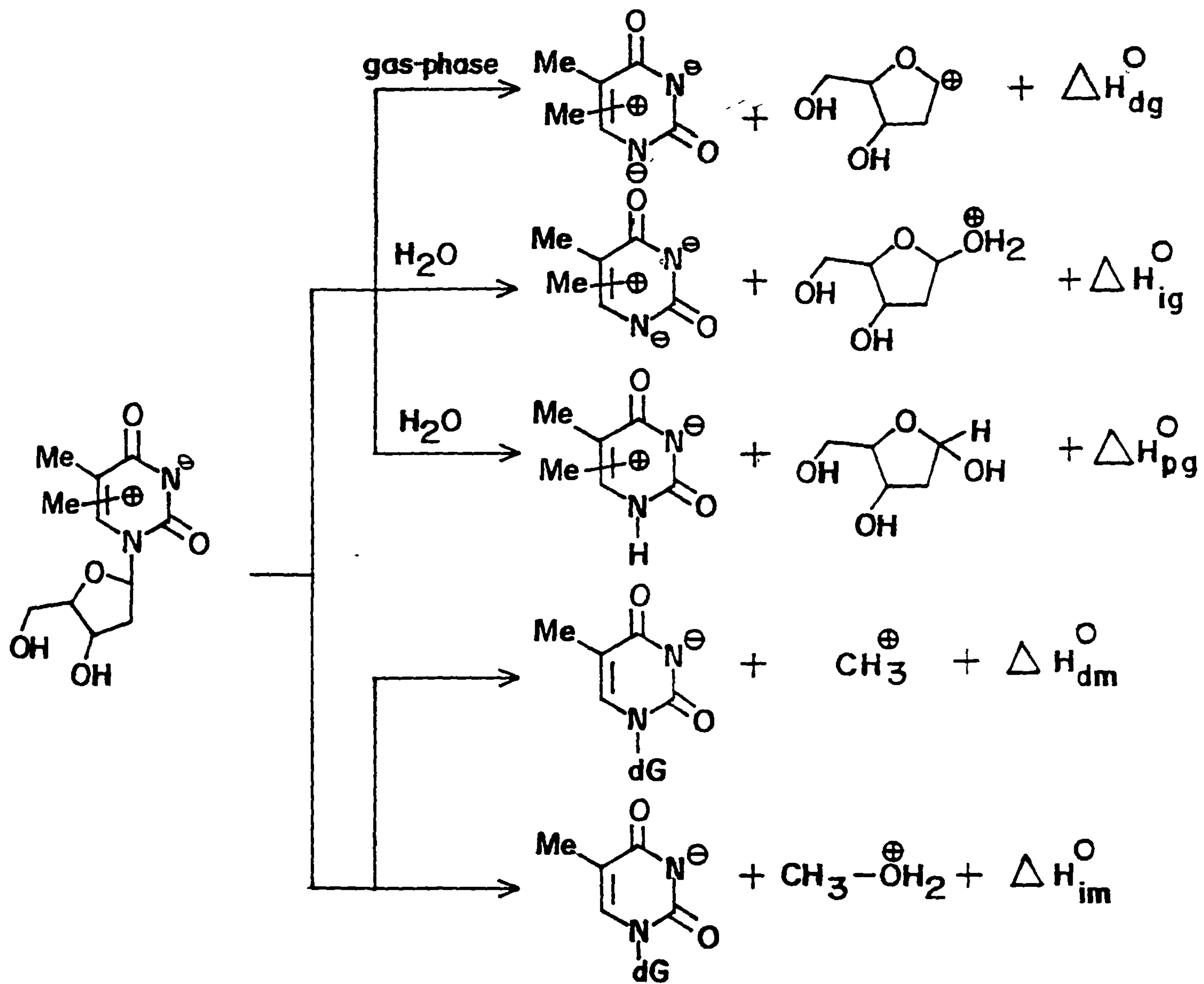


Fig. V.4: Deglycosylation and demethylation pathways for deprotonated alkylated deoxythymidines

sugar moiety. Here again, the reference is to the immediate products of the bond cleavage, not the stable forms, so that ΔH_{ig} may be given by Eqn. (V.2) below :

$$RdB \text{ -----> } RB + dR + \Delta H_{ig} \qquad \ldots\ldots \quad (V.2)$$

where **dR** is the form of the sugar molecule obtained immediately following the bond cleavage. **RB** and **dR** could represent possible intermediates formed during the deglycosylation reaction. The ΔH_{dg} and ΔH_{ig} indices may be interpreted as having some significance for the kinetics of the reaction where the true activation energies are not feasible to compute.

Thirdly, the thermodynamic facility of deglycosylation was expressed simply in terms of the enthalpy ΔH_{pg} of net conversion to the stable products for the reaction as assisted by a water molecule. These stable products are the alkylated base in the stable tautomeric form and the sugar system.

In addition, proneness towards deglycosylation was also expressed in terms of (1) the susceptibility of the C1'-atom of the sugar moiety towards nucleophilic attack by water as estimated through the CM1 (Class 4 charge model, discussed in Sec. II.3) charge Q_c on this atom, and (2) the strength of the glycosyl bond as given by the Wiberg index (or bond order) W_{cn} for this bond.

All these indices were employed for both the protonated and the deprotonated nucleoside systems, the former being distinguished by the suffix "+",and the latter by the suffix "0". This system applies even to the free nucleosides, even though the protonated form is neutral, and the deprotonated form is anionic.

Table V.1 Deglycosylation of watson-crick protonated alkyl nucleosides

Species	ΔH_{dg}^{+}	ΔH_{ig}^{+}	ΔH_{pg}^{+}	W_{cn}^{+}	$Q_{c1'}^{+}$
N^{7}- HdG^{+}	44.13	30.54	-12.78	0.8140	0.2257
N^{7}- HdA^{+}	40.62	27.03	-11.88	0.8071	0.2317
N^{7}-$MedG^{+}$	22.44	8.85	-7.09	0.8166	0.2392
N^{3}-$MedG^{+}$	56.85	43.26	-13.54	0.8517	0.2270
O^{6}-$MedG^{+}$	60.69	47.10	-13.21	0.8388	0.2414
O^{2}-$MedT^{+}$	37.68	24.09	-16.77	0.8089	0.2469
O^{4}-$MedT^{+}$	36.71	23.12	-17.39	0.8020	0.2613
dG	150.24	136.25	-13.58	0.8888	0.2489
dT	146.13	132.54	-15.12	0.8809	0.2935

Table V.2 Deglycosylation of watson-crick deprotonated alkyl nueclosides

Species	ΔH_{dg}^0	ΔH_{ig}^0	ΔH_{pg}^0	W_{cn}^0	$Q_{c1'}^0$
N^7-MedG^0	135.40	121.81	-11.41	0.9135	0.2609
N^3-MedG^0	143.73	130.14	-13.95	0.8929	0.2473
O^6-MedG^0	151.69	138.10	-16.65	0.8934	0.2616
O^2-MedT^0	144.24	130.65	-16.91	0.8891	0.2781
O^4-MedT^0	153.42	139.83	-15.03	0.8856	0.2902
dG$^-$	233.19	219.60	-25.32	0.9625	0.2770
N^7-dGH^0	132.17	118.58	-13.68	0.8749	0.2428
dT$^-$	264.92	251.33	-9.33	0.9432	0.2714

V.4.2 Results of AM1 SCF-MO study on deglycosylation

Tables V.1 and **V.2** present the AM1 generated data on the deglycosylation of the above nine systems considered in their protonated and deprotonated forms respectively. The focus here is on (1) effect of pH upon facility of deglycosylation, (2) effect of alkylation and alkylation site upon deglycosylation, and (3) effect of N^7-protonation upon glycosyl bond lability.

(1) It is immediately evident that the protonated (cationic) deoxynucleosides are predicted to be much more susceptible to deglycosylation than their deprotonated counterparts. This is seen firstly in the differentiation in values of ΔH_{dg} and ΔH_{ig} for the free unalkylated nucleosides. The values in kcal/mol of ΔH_{dg} for the protonated (neutral) and deprotonated (anionic) forms are given in brackets successively as follows: dG (150.2; 233.2) and dT (146.1; 264.9). Similarly the ΔH_{ig} index shows differentiation as follows : dG (136.3; 219.6) and dT (132.5, 251.3). The ΔH_{dg} and ΔH_{ig} indices both also show a similar demarcation between the protonated and deprotonated forms of the alkylated deoxy-nucleosides. Here, the range of $\Delta H_{dg}{}^{+}$ is 22.4 to 60.7 kcal/mol, while for $\Delta H_{dg}{}^{0}$ it is appreciably higher, being 143.7 to 153.4 kcal/mol; the range of $\Delta H_{ig}{}^{+}$ is likewise 8.9 to 47.1 kcal/mol, while for $\Delta H_{ig}{}^{0}$ it is much higher, being 130.1 to 139.8 kcal/mol. This trend of demarcation between protonated and deprotonated species is independent of the type of nucleoside, whether free or alkylated. Hence, protonation (of the N^1-G or N^3-T atoms) corresponding to lowering of pH is predicted to be one factor that enhances the suscepti-bility to deglycosylation *in vitro*, regardless of whether the system is free or alkylated.

(2) The effect of alkylation upon deglycosylation facility may be gauged from comparing the values of the ΔH_{dg} and ΔH_{ig} indices for the free and alkylated nucleosides. For both the protonated and deprotonated data sets, the trend is the same, viz. that the effect of

alkylation, regardless of alkylation site, is to assist or facilitiate the process of deglyco-sylation. The values of the ΔH_{dg} index for the free neutral nucleosides are 146.1 (dT) and 150.2 (dG) kcal/mol, while the range for the cationic alkylated nucleosides is 22.4 to 60.7 kcal/mol. A similar trend holds for the deprotonated systems. The ΔH_{ig} index also puts forward the same trends. Thus, it may be predicted that the effect of nucleoside alkylation at any site would be to destabilise the glycosyl bond.

The extent of this destabilising effect of alkylation, however, depends upon the site of alkylation, as is borne out by a comparison of the ΔH_{dg} and ΔH_{ig} index values for the different alkylated nucleosides. In this, for the cationic systems, the N^7-G site upon alkylation is predicted to have a marked destabilising effect upon the glycosyl bond, which had been earlier observed experimentally as discussed above. Alkylation of the pyrimidine oxygens seems to follow next after N^7-G alkylation in this respect, which had been noted experimentally for the O^2-T site. The O^6-G site is marked for its minimal tendency to destabilise the glycosyl bond upon alkylation, which is also in line with the findings of experiment (see above).

(3) The effect of N^7-G protonation upon depurination is predicted to be destabilisation of the glycosyl linkage, as is seen from the low values of the ΔH_{dg} and ΔH_{ig} indices for the cationic N^7-protonated deoxyguanosine (44.1 and 30.5 kcal/mol respectively) and similarly for the cationic N^7-protonated deoxyadenosine, as compared with the neutral deoxy-nucleosides. Since the N^7-G site is the prime protonation site for deoxyguanosine upon acidification, it would seem most likely that deglycosylation of deoxyguanosine under moderately strong acid conditions (below a pH of 2.5, the acidic pK_a value for dG) would involve prior protonation at this site, thereby destabilising the glycoside bond. This points

to the effect of markedly acidic pH upon the glycosyl bond strength of dG, which is well-known from experimental findings (see above). These predictions also suggest the possible involvement of the N^7-protonated species as a transient intermediate of low concentration for deglycosylation (depurination) of purine nucleosides at higher pH.

For deoxyadenosine, while N^1-A is the most basic site, this does not rule out the possibility of protonation at the N^7-A site too, particularly so since the AM1 calculations presented in Chapter Three point to a definite basicity of this site as well. Thus, N^7-protonation of deoxyadenosine could also play an important role for depurination of this species under highly acidic conditions or even at biological pH through transient participation of the N^7-protonated species.

Note that in all these discussions, the thermodynamic facility of the reaction as given by the ΔH_{pg} indices is not taken into consideration to predict ease of deglycosylation, since this property is here more related to kinetics than to the thermodynamic equilibrium. The uniform negative values of the $\Delta H_{pg}{}^{+}$ and $\Delta H_{pg}{}^{0}$ indices simply indicate that the reaction would be thermodynamically favoured, with the equilibrium shifting to the right. Note also, that unlike the ΔH_{dg} and ΔH_{ig} indices, the ΔH_{pg} indices does not display such a wide range of value. This speaks of the more kinetically significant ΔH_{dg} and ΔH_{ig} indices as providing the greater differentiating factor between the various cases considered, so that the kinetics of the reaction, rather than the simple thermodynamics of the reaction, furnishes the key differentiating element. The kinetics have been modelled here rather simplistically through use of enthalpy indices related to unstable intermediates. A more rigorous use of transition states and activation energies will be described in Sec. V.6.

V.5 Dealkylation of Alkylated DNA Nucleosides

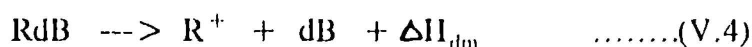
V.5.1 Approach to study of dealkylation

The reaction of *in vitro* dealkylation of alkylated DNA bases and nucleosides was studied in the following manner :

1. Determination of the relative ease of dealkylation and deglycosylation of alkylated nucleoside systems
2. Effect of pH and protonation upon dealkylation facility
3. Effect of alkylation site upon dealkylation facility

The indices used for estimating facility of dealkylation are as described below:

Firstly, use is made of the enthalpy index ΔH_{dm}, a measure of the intrinsic strength of the alkyl bond, which represents the gas-phase dissociation of the alkylated nucleoside into the products which immediately follow the bond cleavage, as given by the following formal reaction :



$$RdB \dashrightarrow R^{+} + dB + \Delta H_{dm} \qquad \text{.......(V.4)}$$

where RdB is the alkylated nucleoside, R^{+} the alkyl cation and dB the unalkylated nucleoside system as obtained immediately after bond cleavage, which is neutral for the cationic nucleosides and anionic for the deprotonated nucleosides.

Secondly, use is made of the enthalpy index ΔH_{pm}, which corresponds to the bond cleavage as assisted by a water molecule, as given by Eqn. (V.5) below :

$$RdB \dashrightarrow (ROH_2)^{+} + dB + \Delta H_{pm} \qquad \text{...... (V.5)}$$

where $(ROH_2)^{+}$ is the protonated alcohol product of hydrolysis.

Lastly, resort is made to (1) the CM1 charge Q_r of the carbon of the methyl group, and (2) the Wiberg bond order W_{cx} for the bond between the alkyl group and its alkylation

site X on the nucleoside.

V.5.2 Results of AM1 SCF-MO study on dealkylation

Tables V.3 and V.4 present the values of the above indices for the protonated and deprotonated alkyl nucleosides respectively. The results are discussed with reference to the three points of focus mentioned in Sec. V.5.1 above.

(1) By comparing these indices with the corresponding indices for deglycosylation facility, it becomes apparent that, for any one nucleoside species (whether protonated or deprotonated), the data points to a greater proneness towards deglycosylation than dealkylation. For instance, for the cationic nucleosides, the range of values for the deglycosylation index ΔH_{dg} is 22.4 to 60.7 kcal/mol, which is appreciably lower than the range for the corresponding dealkylation index ΔH_{dm}, which is about 81.3 to 114.0 kcal/mol. Similarly, the deglycosylation index ΔH_{ig} has a range of about 8.9 to 47.1 kcal/mol, which is lower than the range for the corresponding dealkylation index ΔH_{pm}, which is about 26.5 to 59.3 kcal/mol. Likewise, the Q_c index for deglycosylation is invariably higher in positive charge than the Q_r index for dealkylation, which differentiation is also reflected in the values of the Wiberg bond indices, where the W_{cn} values for the glycosyl bond are lower than the W_{cx} values for the alkyl group bond in each case. Much the same holds true for the neutral deprotonated alkyl nucleosides too. This would predict that the reaction of deglycosylation would be more facile than the dealkylation reaction for each case, which is true regardless of whether the neutral or cationic species are considered in each case. This greater facility of deglycosylation over dealkylation is reflected in the experimental findings of Singer *et al* (1978) and of Allore *et al* (1983), where the glycosyl bond cleavage invariably precedes dealkylation.

Table V.3 Demethylation of watson-crick protonated alkyl nucleosides

Species	H_{dm}^{+}	H_{im}^{+}	H_{pm}^{+}	W_{cx}^{+}	Q_{cm}^{+}
N^{7}-MedG^{+}	229.64	114.08	59.26	0.8951	-0.1004
N^{3}-MedG^{+}	209.55	93.99	39.17	0.8991	-0.0585
O^{6}-MedG^{+}	214.95	99.39	44.57	0.9025	-0.0399
O^{2}-MedT^{+}	196.90	81.34	26.52	0.9133	-0.0721
O^{4}-MedT^{+}	200.00	84.44	29.62	0.9048	-0.0465

Table V.4 Demethylation of watson-crick deprotonated alkyl nuclosides

Species	ΔH_{dm}^0	ΔH_{1m}^0	ΔH_{pm}^0	W_{cx}^0	Q_{cm}^0
N^7-MedG^0	208.92	154.10	-23.10	0.9135	-0.1051
N^3-MedG^0	214.02	159.20	-18.00	0.9270	-0.0194
O^6-MedG^0	221.88	167.06	-10.14	0.9551	-0.0454
O^2-MedT^0	197.25	142.43	-25.50	0.9391	-0.0325
O^4-MedT^0	206.67	151.85	-16.08	0.9475	-0.0459

(2) The effect of protonation upon dealkylation facility may be assessed by comparing the values of the indices of Table V.3 with those of Table V.4. Here, it becomes apparent that protonation is predicted to facilitiate ease of dealkylation for each case. The values of ΔH_{dm}^{+} (for the protonated species) have the range of about 81.3 to 114.0 kcal/mol, while the ΔH_{dm}^{0} index (for the neutral species) has the much higher range of about 197.3 to 221.9 kcal/mol. The ΔH_{pm} *index similarly displays such a clear differentiation* between protonated and deprotonated species, where the ΔH_{pm}^{+} index has the range of about 26.5 to 59.3 kcal/mol, while the ΔH_{dm}^{0} index has the much higher range of about 142.4 to 167.1 kcal/mol. These demarcations are also reflected in the more positive values of the Q_r^{+} index over the Q_r^{0} index, and the lower Wiberg bond strength W_{cx}^{+} for the protonated species as compared to the corresponding Wiberg bond strength W_{cx}^{0} for the deprotonated species in each case. All this goes to establish the consensus of prediction that lowering of pH serves to increase alkyl group lability, regardless of alkylation site. This is much in line with the findings of Singer *et al* (1978) that raising of pH strengthens the binding between alkyl group and alkylation site, so that even the more labile O-alkyl groups separate out intact in the base residues following deglycosylation at elevated pH (neutral or mildly acidic); this is not so at stronger acidic pH, when the O-alkyl groups are effectively lost during hydrolysis.

(3) The effect of alkylation site upon alkyl group lability may be gauged by inspecting the values of the dealkylation indices as evaluated for different alkylation sites. For the cationic species (Table V.3), it is seen that the O-alkylated nucleosides along with the N^3-MedG $^{+}$ species are predicted to be more susceptible to dealkylation than the N^7-MedG $^{+}$ species. This is borne out by the values of the ΔH_{dm}^{+} and ΔH_{pm}^{+} indices, as well as the charge index Q_r^{+}, but not by the Wiberg bond index W_{cx}^{+}. The ΔH_{dm}^{+} and ΔH_{pm}^{+} indices

both predict the following order for dealkylation facility with respect to alkylation site : O^2-T > O^4-T > N^3-G > O^6-G > N^7-G. This order is largely substantiated by the findings of experiment, except that the lability of the N^3-G methyl group is predicted to be too low.

The neutral (deprotonated) alkyl nucleosides present a different ordering for facility of dealkylation with respect to alkylation site, which is predicted by the $\Delta H_{dm}{}^0$ and $\Delta H_{pm}{}^0$ indices to be : O^2-T > O^4-T > N^7-G > N^3-G > O^6-G. This order is not, however, reproduced by the Wiberg bond index $W_{cx}{}^0$. So, in general, the following order of alkyl group lability with respect to alkylation site is furnished for both protonated and neutral cases : O^2-T > O^4-T > N^7-G, which experimental data corroborates.

V.6 Stabilities and Repair Modes at Biological pH

The situation at biological pH (taken to have an approximately neutral value of 7.0) is of bearing in the context of repair mechanisms and any *in vivo* process. This situation with respect to cleavage of the glycosyl bond (deglycosylation) and the methyl group bond (demethylation) may be studied by first selecting each alkylated nucleoside in the form in which it predominantly exists at this pH. The species existing at biological pH are those for which the protic equilibria at this pH shift to the right. The earlier study of Chapter Three dwelt at length on this, and put forward the following species as the ones existing at biological (or approximately neutral) pH : N^7-methyldeoxyguanosine (N^1-protonated : N^7-MedG $^+$); N^3-methyldeoxyguanosine (N^1-deprotonated : N^3-MedG ^{0}); O^6-methyldeoxy-guanosine (N^1-deprotonated : O^6- MedG ^{0}), O^2-methyldeoxythymidine (N^3-deprotonated - O^2-MedT ^{0}), and O^4-methyldeoxythymidine (N^3-deprotonated - O^4-MedT ^{0}).

One feature of the deglycosylation and demethylation reactions of these nucleoside systems at neutral pH and room temperature is that they occur very slowly or not all in the

absence of the activating enzymes. Hence, the situation at neutral pH is of much import insofar as the well-documentated and highly specific choice of repair mechanism is concerned, which will be dealt with at length in the closing part of this section.

V.6.1 Deglycosylation at biological pH

The indices pertaining to facility of deglycosylation of these five species along with the free nucleosides dG and dT are reassembled together in Table V.5, being the ΔH_{dg}, ΔH_{ig} and W_{cn} indices.

It is at once evident that deglycosylation at biological pH is predicted to be much more favoured for N^7-methyldeoxyguanosine over all the other species, including the free unalkylated nucleosides dG and dT. The values of the ΔH_{dg}, ΔH_{ig} and W_{cn} indices for N^7-MedG $^+$ are respectively 22.4 kcal/mol, 8.9 kcal/mol and 0.8140 a.u. In contrast, the range of values of these indices for the other systems are :ΔH_{dg} (143.7 to 153.4 kcal/ mol); ΔH_{ig} (130.1 to 139.8 kcal/mol); W_{cn} (0.881 to 0.893 a.u.). The deprotonated alkyl nucleosides all demonstrate a minimal tendency towards deglycosylation which is quite comparable with the free unalkylated nucleosides dG and dT. This prediction concerning the susceptibility of N^7-methyldeoxygunaosine to depurination at neutral pH is in accordance with experimental results. These calculations do not, however, reproduce the reported glycosyl bond destabilisations associated with alkylation at the N^3-G and O^2-T sites.

V.6.2 Transition states for deglycosylation

An attempt was made to study deglycosylation at neutral pH from a more rigorously kinetic point of view. The three systems taken into consideration here are the N^7-methyldeoxy-guanosine, O^6-methyldeoxyguanosine and N^3-methyldeoxyguanosine species, each in the

Table 5: Deglycosylation data at neutral *p*H

Species	ΔH_{dg}	ΔH_{ig}	ΔH_{pg}	W_{cn}	$Q_{c1'}$
N^7-MedG$^+$	22.44	8.85	-7.09	0.8166	0.2392
N^3-MedG^0	143.73	130.14	-13.95	0.8929	0.2473
O^6-MedG^0	151.69	138.10	-16.65	0.8934	0.2616
O^2-MedT^0	144.24	130.65	-16.91	0.8891	0.2781
O^4-MedT^0	153.42	139.83	-15.03	0.8856	0.2902
dG	150.24	136.25	-13.58	0.8888	0.2489
dT	146.13	132.54	-15.12	0.8809	0.2935

protic form as established for biological pH. It may be noted here that N^7- and N^3-methyldeoxyguanosine are prone to spontaneous depurination and glycosylase repair, while O^6- methyldeoxyguanosine has a stable glycosyl bond at neutral pH and is repaired by an alkyltransferase enzyme.

Trial transition-sate geometries were constructed by a "reaction coordinate" approach, fixing the bond being made at appropriate lengths (varied in steps of 0.2 angstroms) and the two closest points of high energy are taken and further refined by using the **SADDLE** option of the **MOPAC** package with a **BAR** value of 0.15. The refined trial transition-state geometry was taken as the approximate transtion state reflecting the transition state for the water mediated deglysoylation pathway for N7,N3 and O6 methylated deoxyguanosines. Owing to the computational cost associated with locating the accurate transition states (i.e. by characterisation of one and only one negative value of force matrix), the approximate transition states were used for the comparison of deglycosylation with demethylation. Similar procedures were used for locating the transition states for demethylation process also.

Transition states (TS) for nucleophilic substitutions may be described in various ways with regard to the state of making or breaking of the bonds involved. A TS with both the incoming and departing bonds being very weak or unmade may be termed as "loose". A TS with both incoming and departing bonds in a state in-between breaking and making may be defined as "tight". An "early" TS is one where making of the incoming bond well precedes breaking of the departing bond, while a "late" TS is one where breaking of the departing bond precedes making of the incoming bond. The state of making or breaking of a bond may be gauged from the bond lengths as well as from the Wiberg bond indices.

Table V.8 presents data pertaining to transition states for the deglycosylation

Table 6: Demethylation data at neutral *pH*

Species	ΔH_{dm}	ΔH_{im}	ΔH_{pm}	W_{cx}	Q_{cm}
N^7-$MedG^+$	229.64	114.08	59.26	0.8951	-0.1004
N^3-$MedG^0$	214.02	159.20	-18.00	0.9270	-0.0194
O^6-$MedG^0$	221.88	167.06	-10.14	0.9551	-0.0454
O^2-$MedT^0$	197.25	142.43	-25.50	0.9391	-0.0325
O^4-$MedT^0$	206.67	151.85	-16.08	0.9475	-0.0459

reactions of N^7-MedG $^+$, N^3-MedG ^{0} and O^6-MedG ^{0}. The calculated AM1 values of the activation energy E_a, of R_b (the length of the departing bond being broken), of R_m (the length of the incoming bond being made), the corresponding Wiberg bond index values W_b and W_m, as well as the frequency of the imaginary vibration associated with the reaction coordinate are presented.

It is apparent that all the three TS described are "loose" in character, with both incoming and departing bonds being very weak and long. The bond lengths on both sides are well in excess of conventional covalent bond length ranges, and the Wiberg bond strength indices for both bonds are appreciably small in value.

The low value of the activation energy for deglycosylation of N^7-MedG $^+$, as compared with N^3-MedG ^{0} and O^6-MedG ^{0}, speaks for a much greater proneness of N^7-MedG $^+$ towards depurination. This predicted difference in glycosyl bond strength corroborates the data of Tables V.5 and V.6 (which is based upon formal reaction enthalpies and is much less rigorously derived). That N^7-MedG $^+$ is much more susceptible to depurination than O^6-MedG ^{0} is well known (as had been discussed earlier on the basis of experimental work on *in vitro* stabilities), so that these transition state calculations well reproduce the experimental result. It does not seem to work as well for the comparison between N^7-MedG $^+$ and N^3-MedG ^{0} though, where experiment indicates low glycosyl bond strength for the latter too, which is not reflected in these calculations. Fig.V.5 to Fig.V.7 illustrate the energy profile diagrams for water mediated deglycosylation pathway for N^3,N^7 and O^6 methyl deoxyguanosines.

V.5.3 Transition states for demethylation

In like manner as the above, an attempt was made to study the transition states involved in demethylation reactions for the 3 methylated deoxynucleosides in the

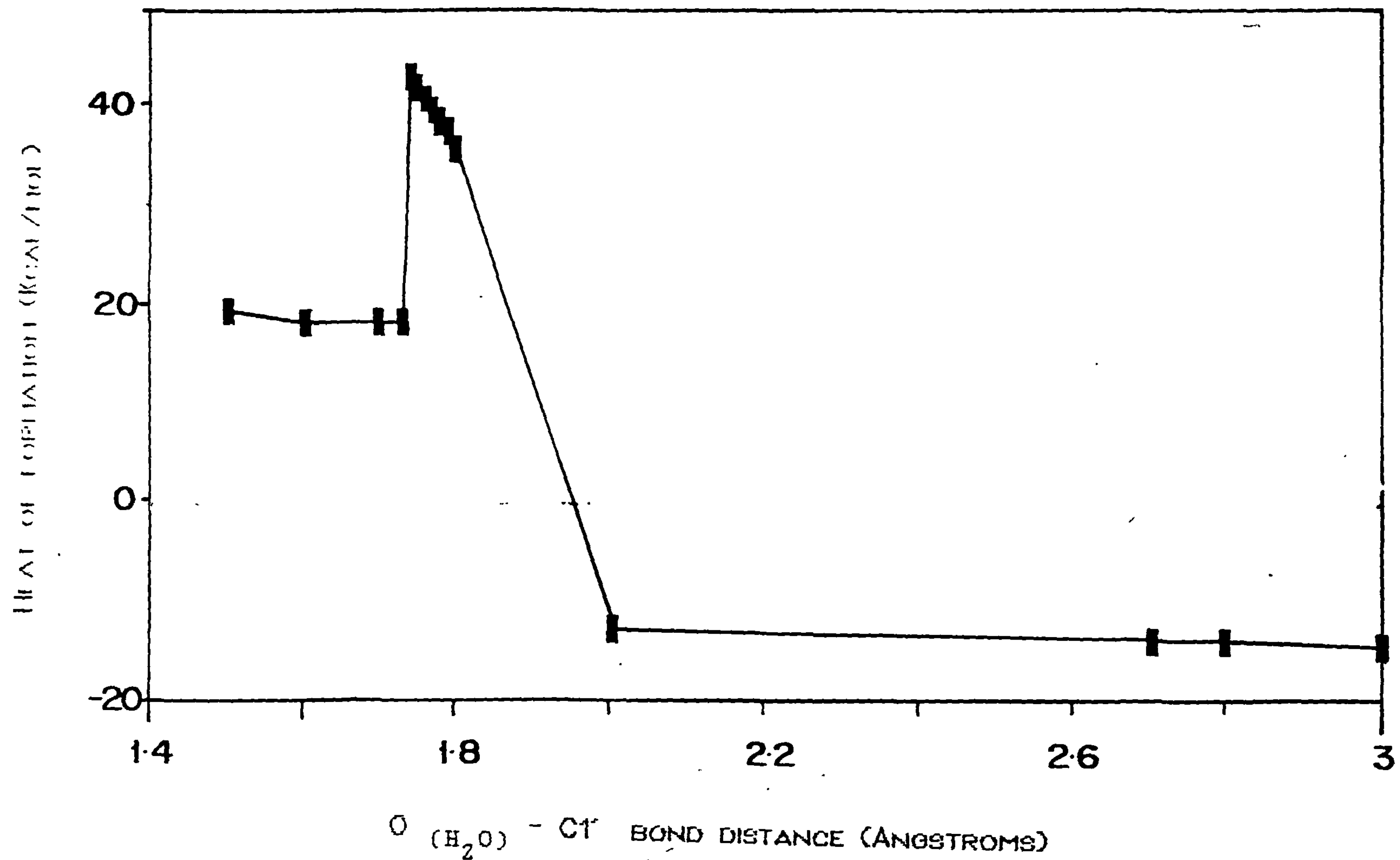


Fig. V.5: Reaction profiles for water assisted deglycosylation pathway for N^7-methyl deoxyguanosine

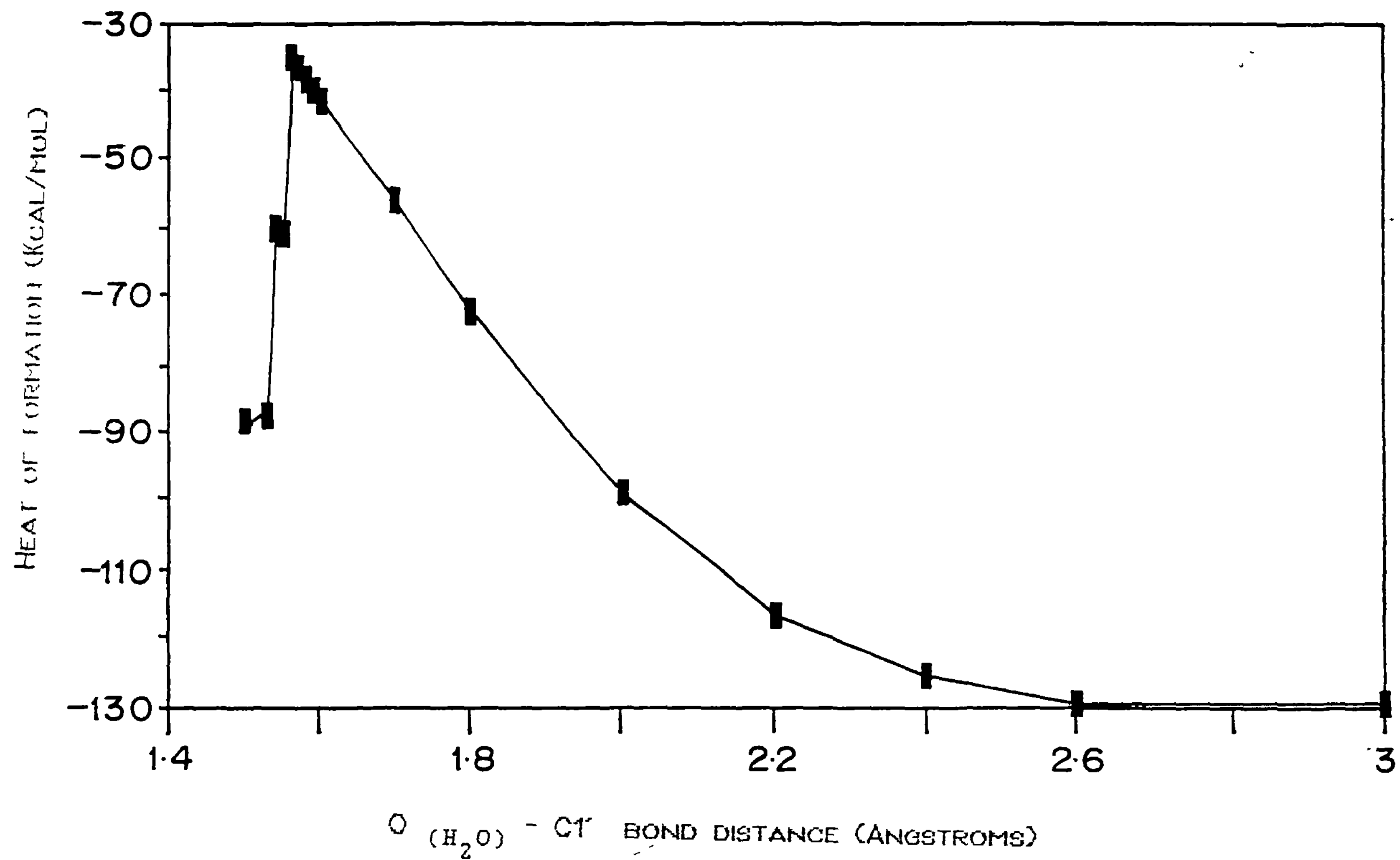


Fig. V.6: Reaction profiles for water assisted deglycosylation pathway for N^3-methyl deoxyguanosine

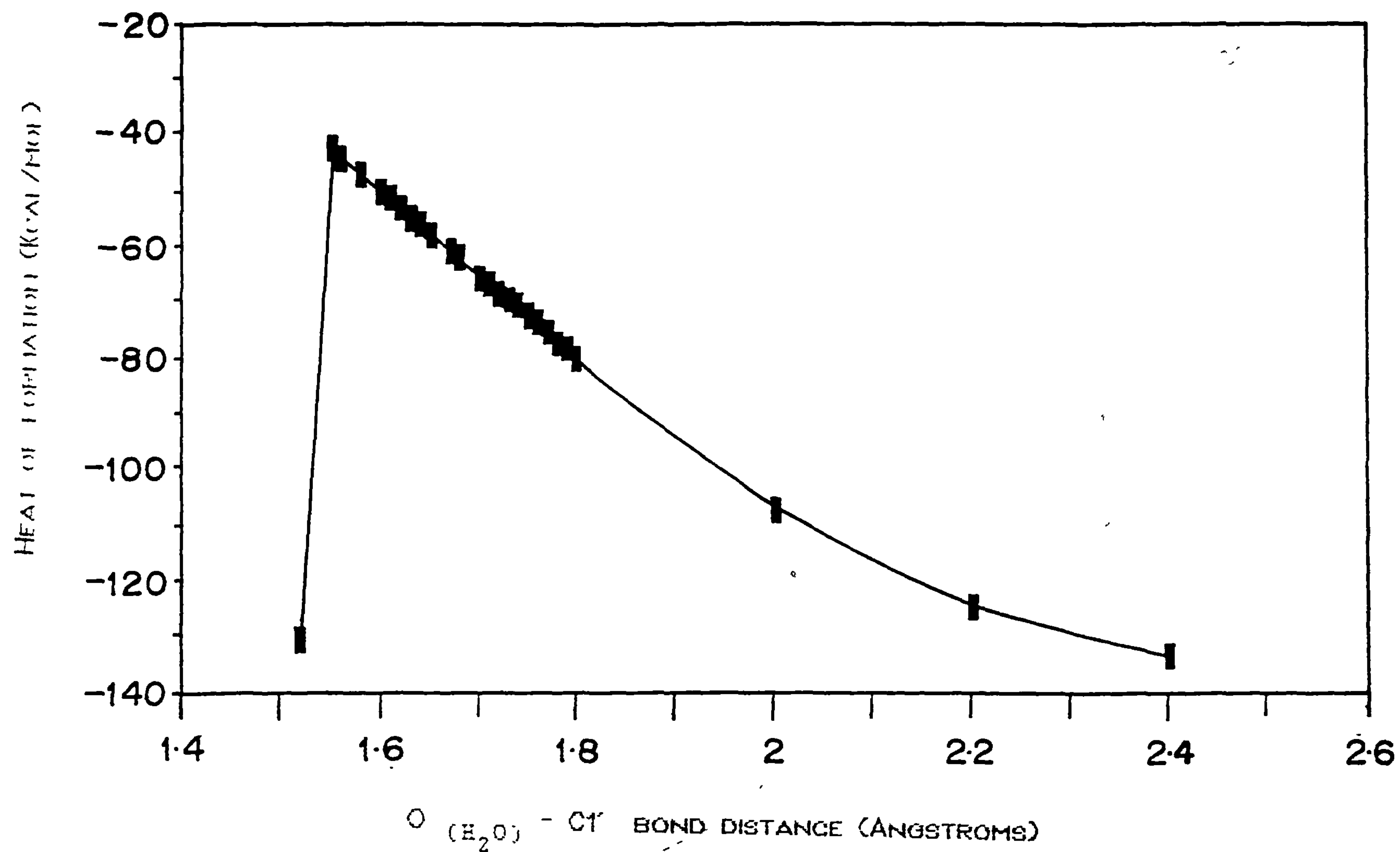


Fig. V.7: Reaction profiles for water assisted deglycosylation pathway for O[6]-methyl deoxyguanosine

Table 7: AM1 and AM1-SM2 computed demethylation data for alkylated nucleic acid bases at neutral pH

Species	ΔH_{dm}		W_{cx}		Q_{cm}	
	AM1	AM1-SM2	AM1	AM1-SM2	AM1	AM1-SM2
N^7-MeG^+	107.59	78.30	0.8929	0.9011	-0.1155	-0.0993
N^3-MeG^0	202.67	75.53	0.9323	0.8918	-0.0783	-0.1048
O^6-MeG^0	213.22	75.35	0.9558	0.9397	-0.0689	-0.0623
O^2-MeT^0	204.83	75.78	0.9414	0.9342	-0.0534	-0.0560
O^4-MeT^0	212.38	81.19	0.9485	0.9397	-0.0687	-0.0697

form available at neutral pH, viz. N^7-MedG $^+$, N^3-MedG ^{0} and O6-MedG ^{0}. It might be worth restating here that of these three, only O^6-MedG ^{0} is markedly prone to demethylation when present in its base form, being also repaired by an alkyltransferase enzyme, not by an N-glycosylase.

The AM1 generated data on the transition states for demethylation of the three nucleoside systems are presented in Table V.8. These data include the activation energy for demethylation E_a, the lengths R_b and R_m of the bonds being broken and made during demethylation, the corresponding Wiberg bond strength indices W_b and W_m, and the frequency of the imaginary frequency corresponding to the reaction coordinate.

In line with the previously stated description of a TS, the TS for demethylation of N^7-MedG $^+$ and N^3-MedG ^{0} both fit the description of "tight". The incoming and departing bonds are not too long or too short, and have bond strengths appropriate to a TS for a more or less classical SN^2 reaction. The TS for O^6-MedG ^{0}, if it may be regarded as a good approximation to a TS for demethylation, seems to earn the appelation of "early",since the incoming bond is quite well-formed, while the departing bond is long and weak. That the SN2 demethylation of O6-MedG ^{0} is "early" comes as no surprise, considering that the analogous SN2 methylation of oxygen nucleophiles by the methanediazonium ion is also similarly described by Ford and Scribner (1990) The most obvious inference from the activation energy data here in Table V.8 is that O^6-MedG ^{0} would be much more susceptible to methyl group loss than either of the two N-methylpurines studied here. This data serves to corroborate the data of Table V.6 on demethylation at neutral pH (which utilises indices derived with less rigour). This is well in line with the experimental results drawn from the hydrolysis of methylated base residues, where the O-methyl groups are noticeably more labile than the N-methyl ones.

V.5.4 In vitro stability and in vivo repair

Since at biological pH and temperature, neither deglycosylation nor demethylation proceed at any appreciable rate for the three systems noted above, the chief bearing the above data could have would be upon the highly specific choices of mechanism adopted for the enzyme-mediated repair of these lesions. Lyngdoh (1992) had dwelt on the remarkable correlation evident between in vitro stability and *in vivo* repair for alkylated DNA base systems. By comparing the innate proneness or preference towards deglycosylation or demethylation for any particular lesion, one may arrive at a good assessment of what would be the repair mechanism likely to operate enzymatically *in vivo*.

V.5.5 Comparison of deglycosylation and demethylation

Table V.9 presents data on the three systems studied above, giving the activation energies for deglycoslyation (E_{dg}) and for demethylation (E_{dm}) as well as quoting the reaction kinetically favoured in each case, adding also the choice of repair mechanism adopted *in vivo*.

It is predicted that N^7-MedG$^+$ would be more prone to deglycosylation than to demethylation at neutral pH, where the respective activation energies are 21.7 and 47.9 kcal/mol. This marked preference for deglycosylation at neutral pH has bearing upon the highly specific choice of N-glycosylase action as the repair mode for this lesion.

O^6-MedG^0 is, however, predicted to be much more susceptible to demethylation than to deglycosylation at neutral pH, the respective activation energies being 1.1 and 52.3 kcal/mol. Since demethylation does not occur spontaneously at room temperature and neutral pH, this data would have bearing upon the well-documentated choice of alkyltransferase repair for this lesion *in vivo*.

Table 8 Transition state characteristics of deglycosylation and demethylation of various methylatd deoxyribonucleosides

System	E_a	R_b	R_m	W_b	W_m	Character
Deglycosylation						
N^7-MedG$^+$	21.67	2.330	2.579	0.098	0.008	Loose
N^3-MedG^0	66.92	2.329	2.140	0.090	0.105	Loose
O^6-MedG^0	52.23	2.460	2.373	0.066	0.031	loose
Demethylation						
N^7-MedG$^+$	47.93	2.116	1.999	0.234	0.253	Tight
N^3-MedG^0	68.29	2.255	1.911	0.255	0.318	Tight
O^6-MedG^0	1.11	1.427	2.446	0.955	0.033	late

Table 9 Theoretically predicted mode of *in vivo* repair choice

Species	E_a(dg)	E_a(dm)	Reaction favoured	Repair mode
N^7-MedG$^+$	21.67	47.93	deglycosylation	Glycosylase
N^3-MedG^0	66.92	68.29	deglycosylation	Glycosylase
O^6-MedG^0	52.33	1.11	demethylation	Alkyltransferase

E_a values are in Kcal/mol
R_b and R_m values are in angstroms

The situation for N^3-MedG^0 seems to indicate only a slight preference for deglycosylation over demethylation, the respective activation energies being 66.9 and 68.9 kcal/mol. The difference is not marked, and does not go to furnish a reliable prediction of proneness towards deglycosylation, which in fact is observed *in vitro* as well as *in vivo*, being mediated through action of an N-glycosylase enzyme as being the repair mode for this lesion.

V.7 Concluding Remarks

This Chapter has dwelt in fair detail upon the reported *in vitro* stabilities of alkylated DNA nucleosides and sought to provide rationalisations and predictions for the behaviour of these systems under conditions of acidic and neutral hydrolysis. The semiempirical AM1 SCF-MO calculations utilised here served very well in this capacity to distinguish between deglycosylation and demethylation as the choice for various systems, varying conditions of pH being taken into consideration. The study at neutral or biological pH was then interpreted in terms of *in vivo* repair mechanisms, and the results indicated a correct assignment of choice of repair mode for various alkylated lesions. The connection between *in vitro* stability and *in vivo* repair was thus quite satisfatorily made by these theoretical models.

References

Abbott P J & Saffhill R (1979) *Biochim. Biophys. Acta.* **562** 51

Ahmed S, Kintanar A & Henderson E (1994) *Nature Struct. Biol* **1** 83-88

Aida & Nishimura (1987) *Mutat. Res.* **192** 83

Alberty R A, Smith R M & Bock R M (1951) *J.Biol.Chem.* **193** 425

Allore B D, Queen A, Blonski W J & Hruska F E (1983) *Can.J.Chem.* **61** 2397

Bargmann C I, Hung M C & Weinberg R A (1986) *Cell* **45** 649

Basu A K & Essigmann J M (1988) *Chem. Res. Toxicol.* **1** 1

Beers R F & Steiner R F (1957) *Nature (London)* ,**179** 1076

Bender M L, Bergeron R J & Komiyama M 1984 *The Bioorganic Chemistry of Enzymatic Catalysis* (New York, Wiley)

Benoit R L & Frechette M (1984) *Can. J. Chem* **62** 995

Benoit R L, Boulet D, Seguin L & Frechette M (1985) *Can. J. Chem.* **63** 1228

Benoit R L & Frechete M (1985) *Can. J. Chem* **63** 3053

Beranek D T, Weis C C & Swenson D H (1980) *Carcinogenesis* **1** 595

Berger N A & Eichhorn G L (1971) *Biochemistry* **10** 1857

Bhanot O S & Ray A (1986) *Proc.Natl.Acad.Sci.USA.* **83** 7348

Bishop J M (1985) *Cell* **42** 23

Bonifas J M, Morley B J, Oakey R E, Kan Y W & Epstein E H (1987) *Proc. Natl. Acad. Sci. USA* **84** 9248

Bonnaccorsi R, Scrocco E, Tomasi J & Pullman A (1972) *Theor. Chim. Acta (Berl.)* **24** 51

Bonnaccorsi R, Scrocco E, Tomasi J & Pullman A (1975) *Theor. Chim. Acta (Berl.)* **36** 339

Boveri T (1914) *Zen Frage der Entschung maligner Tumoren* (Gustav Fischer, Jena)

Breimer L H (1988) *Br. J. Cancer* **57** 6

Brent T P (1979) *Biochemistry* **18** 911

Brookes P, Dipple A & Lawley P D (1968) *J.Chem.Soc.C.* 2026

Burnout D, Koehl P & Fuchs R P P (1989) *Proc. Natl. Acad.Sci. U.S.A.* **86**

Carroll S S & Benkovic S J (1990) *Chem. Rev.* **90** 1291

Cathcart R & Goldthwait D A (1981) *Biochemistry* **20** 273

Cerutti, P.A. (1985) *Science* 227, 335.

Chetsanga C J & Lindahl T (1979) *Nucleic Acid Res.* **6** 3673

Cho B P, Kadlubar F F, Culp S J & Evans F E (1990) *Chem. Res. Toxicol.* **3** 445

Christensen J J & Izatt R M (1962) *J. Phys. Chem.* **66** 1030

Christensen J J, Rytting J H & Izatt R M (1967) *J. Phys. Chem.* **71** 2700

Christensen J J, Rytting J H & Izatt R M (1970) *Biochemistry* **9** 4907

Christensen J J, Rytting J H & Izatt R M (1970) *J. Chem. Soc. B.* 1643

Cleaver J E & Kaufman W K (1980) *Nature* **284** 378

Cochran W (1951) *Acta Crystallogr.* **4** 81

Cohn W E (1955) In: *The Nucleic Acids*. Vol.I.(eds) Chaigraff E & Davidson J N (New York, Academic Press) p. 217

Cox R A (1966) *Biochem. J.* **100** 146

Cox R & Irving C C (1977) *Cancer Lett.* **3** 265

Cramer C J & Truhlar D G (1991) *J.Am.Chem.Soc.* **113** 8305

Cramer C J & Truhlar D G (1991) *J.Am.Chem.Soc.* **113** 8553

Cramer C J & Truhlar D G (1992) *Science* **256** 213

Cramer C J & Truhlar D G (1992) *J. Comput.-Aided Mol. Des.* **6**, 629

Cramer C J, Lynch G C, Hawkins G D, Giesen D J, Truhlar D G & Liotard D A (1994) *QCPE Bulletin* **14** 55

Cramer C J, Truhlar D G (1995) In *Reviews in Computational Chemistry* Boyd, D B, Lipkowitz K B, Eds. VCH: New York, Vol 6.

Cruse W B T, Egert E, Kennard O, Sala G B, Salibury S A & Viswamitra M A (1983) *Biochemistry* **22** 1833

Culp S J, Cho B P, Kadlubar F F & Evans F E (1989) *Chem. Res. Toxicol* **2** 416

Davidon W C (1968) *Comput. J.* **11** 406

Devasagayam T P A, Steenken S, Obendorf M S W, Schulz W A & Sies H (1991) *Biochemistry* **25** 6283

Dewar M J S & Thiel W (1977) *J.Am.Chem.Soc.* **99**,4899,4907

Dewar M J S & Storch D M (1985) *J. Am. Chem. Soc.* **107** 3898

Dewar M J S, Zoebisch E G, Healy E F & Stewart J J P (1985) *J.Am.Chem.Soc.* **107** 3902

DiGuiseppi J & Fridovich I (1984) *CRC Crit. Rev.* **12** 315

Dizdaroglu M (1985) *Biochemistry* **24** 4476

Dodson M L Michaels M L & Lloyd R S (1994) *J. Biol. Chem.* **52** 32709

Dosanjh M K, Essigmann J M, Goodman M F & Singer B (1990) *Biochemistry* **29** 4698

Duncan R H & Davies G S (1989a) *J. Theor. Biol* **145** 345

Duncan R H (1989b) *J.Chem.Soc. Perkin Trans.* 2 2127

Duncan R H & Davies G S (1989c) *Int.J.Quantum.Chem.* **35**, 665

Dunn D B & Hall R H. *Handbook of biochemistry & molecular biology. Nucleic acids.* Vol.1, CRC Press, Cleaveland. 1975.

Fazakerley G V, Gdaniec Z & Sowers L C (1993) *J. Mol. Biol* **230** 6

Fletcher R & Powell M J D (1963) *Comput.J.* **6** 163

Ford G P & Scribner J D (1983) *J.Am.Chem.Soc.* **105** 349

Ford G P & Scribner J D (1990) *Chem.Res.Toxicol.* **3** 219

Ford G P & Wang B (1992a) *J. Comp. Chem.* **13** 229

Ford G P & Wang B (1992b) *J. Chem. Phys.* **97** 4162

Ford G P & Wang B (1992c) *Int.J.Quantum Chem.* **44** 587

Ford G P & Wang B (1992d) *Int.J.Quantum Chem.* **44** 604

Ford G P & Wang B (1993) *J. Mol.Struct.(THEOCHEM)* **283** 49

Fox J J & Shugar D (1952) *Biochim. Biophys.acta.* **9** 369

Fox J J, Cavalieri L F & Chang N (1953) *J. Am. Chem. Soc.* **75** 4315

Fraga C G, Shigenaga M K, Park J W, Dagan P & Ames B N (1990) *Proc. Natl. Acad. Sci, USA* **87** 4533

Franks L M & Teich N (Eds.) (1986) *Introduction to the Cellular and Molecular Biology of Cancer* (Oxford University press, Oxford)

Frei J V and Lawley P D (1975) *Chem. Biol. Interact.* **10** 413

Frei J V, Swenson D H, Warren W and Lawely P D (1978) *Biochem. J.* **174** 1031

Frescht & Knill-jones (1994)

Fried V, Hameka H F & Blukis U *Physical Chemistry* . Macmillan Publication Co. Inc., New York. 1977. p834

Friedberg E C (1985) *DNA rapair*, Freeman Publishers: New York

Gaffney B L & Jones R A (1989) *Biochemistry* **28** 5881

Gehring K, Leroy J -L & Gueron M (1993) *Nature (London)* **363** 561

Georgiadis P, Xu Y -Z & Swann P F (1991) *Biochemistry* **30** 11725
Gerson D (1989) *QCPE Bull.* **9** 99

Giesen D J, Storer J W, Cramer C J & Truhlar D G (1995) *J.Am.Chem. Soc* . **117** 1057

Goth R & Rajewsky M F (1974) *Proc.Natl.Acad.Sci.U.S.A.* **71** 639

Goth-Goldstein R (1980) *Cancer Res.* **40** 2623

Ishiwata K & Oikawa A (1979) *Biochim Biophys.Acta* **563** 375

Hariharan P C, Popkie H E & Kauffmann J J (1979) *Int.J.Quantum Chem.* **13** 255

Hehre, W J, Radon L, Schleyer, P. v. R & Pope J A (1986) *Ab initio molecular Orbital Theory*. Wiley, New York.

Hendler S S, Fuerer E & Srinivasan P R (1970) *Biochemistry* **9** 4141

Hoogsteen K (1963) *Acta Crystallogr*. **16** 907

Hutterman J, Kohnlein W, Teoule R & Bertinchesen A J (1978) Eds. *Effect of Ionizing Radiation on DNA* Springer-Verlag New York

Iball J & Wilson H R (1965) *Proc. Roy. Soc. Ser.A.* **288** 418

Inman R B (1964) *J. Mol.Biol* **9** 624

Izatt R M, Hansen L D, Rytting J H & Christensen J J (1965) *J. Am. Chem. Soc.* **87** 2760

Izatt R M, Rytting J H, Hansen L D & Christensen J J (1966) *J. Am. Chem. Soc.* **88** 2641

Jardetzky C D & Jardetzky O (1960) *J. Am. Chem. Soc.* **82** 222

Jordan D O, Mathieson A R & Matty S (1956) *J. Chem. Soc.* 158

Kang C H et al (1992) *Nature (London)* **356** 126

Karran P, Lindahl T & GriffinB (1979) *Nature* **280** 76

Kasai H & Nishimura S (1984a) *Nucleic Acids Res.* 2137

Kasai H & Nishimura S (1984b) *Gann* **75** 564

Kasai H & Nishimura S (1984c) *Gann* **75** 841

Kasai H, Yamizumi Z, Berger M & Cadet J (1992) *J. Am. Chem.Soc.* **114** 9692

Katritzky A R & Lagowski J M (1963) *Adv. Heterocycl. Chem.* **1** 2

Katritzky A R & Waring A J (1963) *J. Chem. Soc.* 3046

Kleihues P, Doerjer G, Keefer L.K, Rice J M, Roller P P & Hodgson R M (1979) *Cancer Res.* **39** 5136

Kochetkov N K & Budovskii E I (1972) *Organic Chemistry of Nucleic Acids Part A* (New York: Plenum Press)

Kohn K W (1972) *Cancer Res.* **37** 1450

Kouchakdjian M, Bodepudi V, Shibutani S, Eisenberg M, Johnson F, Grollman A P & Patel D J (1991) *Biochemistry* **30** 1403

Kuchino Y, Mori F, Kassi H, Inove H, Iwai S, Miura K, Ohtsuka E & Nishimura S (1987) *Nature* **327** 77

Kwiatkowski J S, Zielinski T J & Rein R (1986a) *Adv. Quantum Chem.* **18** 85

Kwiatkowski J S, Zielinski T J & Rein R (1986b) *Adv. Quantum Chem.* **18** 85

Lane M A, Sainten A & Cooper G M (1981) *Proc. Natl. Acad. Sci. USA* **78** 5185

Langridge R & Rich A (1963) *Nature (London)* **198** 725

Laval J (1977) *Nature* **269** 829

Laval J, Pierre J & Laval F (1981) *Proc. Natl. Acad. Sci. USA* **78** 852

Lawley P D (1973) *Chem.-Biol. Interact.* **7** 127

Lawley P D & Thatcher C J (1970) *Biochem. J* **116** 696

Leroy J -L, Kettani A, Gehring K & Gucron M (1993) *Biochemistry* **32** 6019

Leroy J -L, Gueron M, Mergny J L & Helene C (1994) *Nucl. Acids Res.* **22** 1600

Lewin S (1964) *J. Chem. Soc.* 792

Lewin S & Humphreys D A (1966) *J. Chem. Soc.B.* 210

Lindahl T (1979) *Prog. Nucleic Acid Res. Mol. Biol.* **22** 135

Lindahl T (1982) *Annu Rev Biochem* **51** 61

Lindahl T, Sedgwick B, Sekiguchi M & Nakabeppu Y (1988) *Annu. Rev. Biochem.* **57** 133

Lindahl T (1993) *Nature* **362** 709

Liotard D A, Healy E F, Ruiz J M & Dewar M J S. AMPAC QCPE No. 506-version 2.1. (1989) *QCPE Bulletin*. 9 123

Loeb L A & Kunkel T A (1982) *Annu. Rev. Biochem.* **52** 429

Loeb L A (1985) *Cell* **40** 483

Loeb L A & Preston B D (1986) *Ann. Rev. Genet.* **20** 201

Loechler E L, Green C L & Essigmann J M (1984) *Proc. Natl. Acad Sci. USA* **81** 6271

Loechler E L, Benasutti M, Basu A K, Green C L & Essigmann J M (1990) *Mutation & Environment* part A p51

Loveless A (1969) *Nature* **223** 206

Lowdin P -O (1965) *Adv. Quantum Chem.* **2** 213
Lowdin P -O (1977) *Int.J.Quantum Chem. Quantum Biol Symp*. **4** 185

Ludlum D B (1970) *J.Biol.Chem*. **245** 477

Luque F J, Negre M & Orozco M (1993) *J.Phys.Chem.* **97** 4386

Luque F J, Bachs M & Orozco M (1994) *J. Comput. Chem.* **15** 847

Lyngdoh R H D (1992) *J. Chem. Soc. Perkin Trans. 2* 1173

Lyngdoh R H D (1993) *Proc. Ind. Acad. Sci. (Chem.Sci.)* **105** 253

Lyngdoh R H D (1994) *J. Biosci*. **19** 31

Magee & Farber 1962

Maher L J, Wold B & Dervan P (1989) *Science* **245** 725

Makar G K R & Wiliams D R J (1974) *Inorg. Nucl. Chem*. **36** 1675

Margison G P & Kleihues P (1975) *Biochem.J.* **148** 521

McCann J, Spingard N E, Kobori J & Ames B N (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72** 979

Miertus S, Scrocco E & Tomasi J (1981) *J. Chem. Phys.* **55** 117

Miertus S & Tomasi J (1982) *Chem.Phys.* **65** 239

Miles H T, Howard F B & Frazier J (1963a) *Science* **142** 1458

Miles H T,Bradley R B & Becker E D (1963b) *Science* **142** 1569

Miller E C & Miller J A (1981) *Cancer* **47** 2327

Miller E C (1978) *Cancer Res.* **38** 1479

Modrich P (1994) *Science* **266** 1959

Mohammed S N & Hopfinger A J (1980) *J.Theor.Biol.* **87** ,401

Mulliken R S (1955) *J. Chem. Phys.* **23** 1833

Nakabeppu Y, Kondo H, Sekiguchi M (1984) *J. Biol.Chem.* **259** 13723

Negre M, Orozco M & Luque F J (1992) *J. Chem. Phys.* 1992 **196** 27

O'Connor P C, Capps M J & Craig A W (1973) *Br. J. Cancer* **27** 153

O'Connor P C, Capps M J, Craig A W , Lawley P D & Shah S A (1973) *Biochem. J.* **129** 519

Oda Y, Uesugi S, Ikehara M, Nishimura S, Kawase Y, Ishikawo H, Inove H & Ohtsuka E (1991) *Nucleic Acid Res.* **19** 1407

Olsson M & Lindahl T (1980) *J. Biol. Chem.* **255** 1069

Orkin S (1983) *In Recombinant DNA Applications to Human Disease* Caskey C T & White R L eds., pp 19-28, Cold Spring Harbor Laboratory

Orozco M & Luque F J (1994) J. Comp. Chem.

Ozment J L & Shmiedekamp A M (1992) *Int. J. Quantum Chem.* **43** 783

Pegg A E (1977) *Adv.Cancer Res.* **25** 195

Pegg A E, Wiest L, Foote R S, Mitra S & Perry W (1983) *J. Biol. Chem.* **258** 2327

Pegg A E, Scicchitano D & Dolan M E (1984) *Cancer Res*. **44** 3806

PiccirilliJ A, Krauch T, Moroney S E & Benner S A (1990) *Nature*, **343** 33

Pieree J & Laval J (1980) *Biochemistry* **19** 5018

Pohorille A & Loew G H (1985) *Biophys.Chem*. **22** 45

Preston B D, Singer B & Loeb L A (1986) *Proc.Natl.Acad.Sci.USA*. **83** 8501

Psoda A, Kierdaszuk D, Pohorille A, Geller M, Kusmirek J T & Shugor D (1981) *Int.J.Quantum Chem*. **20**, 543

Pullman A, Pullman B & Berthier G (1956) *C. R. Acad.Sci*. **243** 380

Pullman B (1959) *J. Chem. Soc.* 1621

Pullman B & Pullman A (1971) *Adv. Heterocycl. Chem*. **13** 77

Reinert H &Weiss R (1969) *Z.Physiol.Chem*. **350** 1321

Ripley L S, Dubins J S, deBoer J G, DeMarini D M, Bogerd A M & Kreuzer K N *J. Mol. Biol.* **200** 665

Robins P, Harris A L, Goldsmith T & Lindahl T (1983) *Nucleic Acids Res.* **22** 7743

Saenger W (1984) *Principles of nucelic acid strucutre.* New York: Springer-Verlag.

SaffhillR, Margison G P & Connor P J O' (1985) *Biochim. Biophys. Acta* **823** 111

Sancar A (1994) *Science* **266** 1954

Schendel P & Robins P E (1978) *Proc. Natl. Acad. Sci USA* **75** 6017

Schoental R (1969) *Biochem.J.* **114** 55

Sedgewick B & Lindahl T (1982) *J. Mol.Biol* **154** 169

Shapiro R (1968) *Progr. Nucleic Acid Res .* **8** 73

Shapiro R & Danzig M (1972) *Biochemistry* **11** 23

Shibutani S, Takeshita M & Grollman A P (1991) *Nature* **349** 431

Shugar D & Fox J J (1952) *Biochim. Biophys. Acta.* **9** 199

Singer B (1972) *Biochemistry* **11** 3939

Singer B (1975) *Prog.Nucleic Acids Res.Mol.Biol.* **15** 219

Singer B (1976) *Nature* **264** 333

Singer B, Kroger M & Carrano M (1978a) *Biochemistry* **17** 2346

Singer B, Fraenkel-Conrat H & Kusmierek J T (1978b) *Proc.Natl. Acad. Sci.USA* **75** 1722

Singer B, Spengler S & Bodell W J (1981) *Carcinogenesis* **2** 1069

Singer B & Grunberger D (1983) *Molecular Biology ofMutagens & Carcinogens* (New York, Plenum) p.23

Singer B, Spengler S J, Fraenkel-Conrat H & Kusmierek J T (1986) *Proc.Natl.Acad.Sci.USA* . **83** 28

Singer B (1986) *Cancer Res.* **46** 4879

Skipper H E (1983) *Adv. Cancer Res.* **40** 249

Stewart R (1985) *The Proton: Applications to Organic Chemistry* (Orlando FL, Academic Press).

Stewart J J P (1983) QCPE Program No.455 (version3.1) *QCPE Bull* **3** 43

Stewart J J P (1989) *J. Comput. Chem.* **10** 209,221

Storer J W, Giesen D J, Cramer C J & Truhler D G (1995) *J.Comput.-Aided Mol Des.* **9**

Straus D.S. (1981) *J. Natl. Cancer Inst.* **67** 233

Strauss B, Rabkin S, Sagher D & Moore P (1982) *Biochemie* **64** 829

Strazewski P & Tamm C (1990) *Angew. Chem. Int. Ed.* **29** 36

Suchorukow B T,Poltew V I & Blumenfeld L (1964) *Abh. Deut. Akad. Wiss. Berlin Kl.Med.* 381

Swann P F & Magee P N (1971) *Biochem. J* **125** 841

Sygula A & Buda A (1983) *J.Mol.Struct.(THEOCHEM)* **9** 267

Tapia O, Daudel R, Pullman A, Salem L, Veillard A; Eds. *Quantum Theory of Chemical Reactions* ; Reidel: Dordrecht, The Netherlands, Vol 2, p 25

Taylor H F W (1948) *J. Chem. Soc* . 765

Tomasi J & Persico M (1994) *Chem. Rev.* **94** 2027

Topal M D & Fresco J R (1976a) *Nature(London)* **263** 285

Topal M D & Fresco J R (1976b) *Nature(London)* **263** 289

Treager J C & McLoughlin R G (1981) *J.Am.Chem.Soc.* **103** 3647

Ts'O P.O.P, *In Carcinogenesis: Fundamental Mechanisms & Environmental effects* , Edited by Pullman B, Ts'O P O P & Gelboin H, Academic Press, New York, 1980.

Ueda T & Fox J J (1963) *J. Am. Chem. Soc* . **85** 4024

Varmus H (1984) *Ann. Rev. Genet* . **18** 553

Venkateswarlu D & Lyngdoh R H D (1995) *J. Chem. Soc. Perk.Trans.II* **4** 839

Cruzeiro-Hannson L, Swann P F, Pearl L and Goodfellow J M (1992) *Carcinogenesis*, 13, 2067.

Engel J D and von Hippel P H (1987) *J. Biol. Chem.*, 253, 927.

Fersht A R & Knill-Jones J W (1983) *J. Mol. Biol*, 165, 633

Goodfellow J M and M.A. Williams M A (1992) *Curr. Opin. Struct. Biol.*, 2, 211.

Parker K, Cruzeiro-Hannson L and Goodfellow J M (1993) *J. Chem. Soc. Farad. Trans.* 2637.

Pederson L G , Darden T A, Deerfield II D W, Anderson H M W and Hoel D G (1988) *Carcinogenesis*, 9, 1553.

Lively T N, Jurema M W and Shields G C (1994) *Int. J. Quant. Chem.* 21, 95.

NEHU LIBRARY
Acc. No. 103235
Acc. by

Voigt J M & Topal M D (1990) *Biochemistry* **29** 5012

Wani A A, Wani G & Ambrosio S M D (1990) *Carcinogenesis* **11** 1419

Ward J F (1988) *Prog. Nucleic Acid Res. Mol. Biol.* 95

Watson J D & Crick F H C (1953a) *Nature (London)* **171** 737

Watson J D & Crick F H C (1953b) *Nature (London)* **171** 964

Wood M L, Esteve A, Morningstar M L, Kuzsiemko G M & Essigmann J M (1992) *Nucleic Acid Res.* **20** 6023

Yarosh D B 1985 *Mutat. Res.* **145** 1

Zubay G (1958) *Biochim.Biophys.Acta.* **28** 644

Acc. by ....
Date ....
Class by..........
Sub Heading by ....
Enter by..........
Transcribed by

ERRATA

The following corrections were made in the thesis according to the suggestions made by one of the examiner.

1. In cases, where the tables headings are not very clear, appropriate explanations were given in the tables. Tables 12 and 13 have been modified and the missing table 14 was added.

2. While AM1 method in both gas-phase and aqueous-phase is sufficiently able to preidict the tautomeric stabilities and the proper order of the stabilties, the inaccuracy in reproducing the quantitative values is due the artefact of AM1 method's parametrization schemes, where the electronic correlation is not taken into consideration and only the valence shell electrons are considered in estimating the heats of formation and other thermodynamic parameters. One may feel that the choice of *ab initio* basis sets such as 6-31G* might improve the situation, however such methods were not available to the candidate while carrying out the doctoral studies.

3. Figure III.7 refers to the linear correlation plot between experimental and AM1-SM2 estimated pKa values. While the abscissa in this plot differs from the rest two plots such as III.6 and III.8, the main inference and the anology is essentially same. The typograhical errors in expressing the abscissa were corrected.

4. Figure IV.1 was changed to represent the mispairs among natural bases.

5. Unfortunately, the candidate had taken wrong direction cosines in estimating the angle between the glycosydic bond vectors and C1'-C1' distances. The actual values, as the examiner suggested should be close to 70°. However, since this mistake does not lead to any adverse effect on the discussion on the base-pairing properties of alkylated bases, the θ_{NH} values may not be considered while analyzing the configurational markers among various base-(mis)pairs.

6. Some mistakes in the data of IV.1 to IV.3 were corrected.

7. The missing references in the thesis were appended to the list of references.

8. Figures IV.8 and IV.9 were changed to represent the more meaniningful rotational barriers for 8-oxodeoxyguanosine and 8-oxodeoxyadenosine.

9. Some duplicated lines on page 136 have been deleted.

10. While discussing the stability and repair of various alkylated nucleosides in Chapter V, we have employed AM1 method instead of PM3 method, since AM1 method has been shown to be more promosing in studying transition state properties and various theoreitical indices such as bond orders and charges of particularly nitrogen contaning species such as DNA bases (as in Chapter III). However, in chapter IV, we have used the PM3 method owing to its recent promise in reproducing the accurate hydrogen bonding energies (as mentioned in chapter IV, page 88).

11. We could not properly isolate the transtion state for O^6MedG. This may be explained partly because of more SN^1 character of demethylation of O^6MedG. Whereas, in other cases of N-alkylated species, the water mediated dealkylation would be more prone to SN^2 type reaction. However, we could succeed in isolating the approximate transition states for other alkylated species.