ลิควิดโครมาโทกราฟี และแมสสเปกโทรเมตรีสำหรับตรวจสอบครอสลิงก์ระหว่างเมคลอเรทามีนกับดีเอ็นเอ

นาย ณัฐพล จงอรุณงามแสง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาเภสัชศาสตรมหาบัณฑิต
สาขาวิชาเภสัชเคมี ภาควิชาอาหารและเภสัชเคมี
คณะเภสัชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2552

LIQUID CHROMATOGRAPHIC AND MASS SPECTROMETRIC DETERMINATIONS OF MECHLORETHAMINE-DNA CROSSLINKS

Mr. Nutthapon Jongaroonngamsang

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Pharmacy Program in Pharmaceutical Chemistry
Department of Food and Pharmaceutical Chemistry
Faculty of Pharmaceutical Sciences
Chulalongkorn University
Academic Year 2009

Thesis Title	LIQUID CHROMATOGRAPHIC AND MASS SPECTROMETRIC DETERMINATIONS OF MECHLORETHAMINE-DNA CROSSLINKS
By	Mr. Nutthapon Jongaroonngamsang
Field of Study	Pharmaceutical Chemistry
Thesis Advisor	Assistant Professor Pornchai Rojsitthisak, Ph.D.
Thesis Co-Advisor	Associate Professor Ian S. Haworth, Ph.D.

Accepted by the Faculty of Pharmaceutical Sciences, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

Pornpen Pramyothin ..... Dean of the Faculty of Pharmaceutical Sciences

(Associate Professor Pornpen Pramyothin, Ph.D.)

THESIS COMMITTEE

..... Chairman

(Assistant Professor Chamnan Patarapanich, Ph.D.)

Pornchai Rojsitthisak ..... Thesis Advisor

(Assistant Professor Pornchai Rojsitthisak, Ph.D.)

..... Thesis Co-Advisor

(Associate Professor Ian S. Haworth, Ph.D.)

Nuansri Niwattisaiwong ..... Examiner

(Associate Professor Nuansri Niwattisaiwong)

Opa Vajragupta ..... External Examiner

(Associate Professor Opa Vajragupta, Ph.D.)

ณัฐพล จงอรุณงามแสง :ลิควิดโครมาโทกราฟี และแมสสเปกโทรเมตรีสำหรับตรวจสอบครอสลิงก์ระหว่างเมคลอเรทามีนกับดีเอ็นเอ. (LIQUID CHROMATOGRAPHIC AND MASS SPECTROMETRIC DETERMINATIONS OF MECHLORETHAMINE-DNA CROSSLINKS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.พรชัย โรจน์สิทธิศักดิ์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : ASSOC. PROF. IAN S. HAWORTH, Ph.D. 125 หน้า.

ปฏิกิริยาครอสลิงก์ระหว่างเมคลอเรทามีนกับสายดีเอ็นเอที่ตำแหน่งคู่มิสแมตช์ไซโตซีน ถูกนำมาใช้สำหรับตรวจสอบโครงสร้างดีเอ็นเอที่มีลำดับเบส d[GCC]•d[GCC] ซ้ำกัน ซึ่งเป็นลำดับเบสที่เกี่ยวข้องกับการเกิดโรคทางพันธุกรรมที่เรียกว่า Fragile X syndrome วัตถุประสงค์ของงานวิจัยนี้ เพื่อตรวจสอบโครงสร้าง และปริมาณครอสลิงก์ระหว่างดีเอ็นเอสายคู่กับเมคลอเรทามีนที่ตำแหน่งคู่มิสแมตช์ไซโตซีนด้วยวิธีไฮเพอร์ฟอร์แมนซ์ลิควิดโครมาโทกราฟี และ มัลดิ-ทอฟ-แมสสเปกโทรเมตรี และ ตรวจสอบเป้าหมายของเมคลอเรทามีนบนดีเอ็นเอสายคู่ที่มีคู่มิสแมตช์ไซโตซีนด้วยวิธีการย่อยด้วยเอนไซม์ ไฮเพอร์ฟอร์แมนซ์ลิควิดโครมาโทกราฟี และ อีเอสไอ-แมสสเปกโทรเมตรี การตรวจสอบผลิตภัณฑ์ซึ่งเกิดจากปฏิกิริยาระหว่างเมคลอเรทามีนกับดีเอ็นเอสายคู่ที่มีลำดับเบส d[CTCACACCGTGGTTC]•d[GAACCACCGTGTGAG] (เบสที่ขีดเส้นใต้แสดงคู่มิสแมตช์ไซโตซีน) นั้นด้วยวิธีไฮเพอร์ฟอร์แมนซ์ลิควิดโครมาโทกราฟี และการตรวจสอบมวลโมเลกุลด้วยวิธีมัลดิ-ทอฟ-แมสสเปกโทรเมตรี แสดงให้เห็นว่าผลิตภัณฑ์ดังกล่าวเป็นสารประกอบครอสลิงก์ของเมคลอเรทามีนกับดีเอ็นเอ เมื่อย่อยสารประกอบครอสลิงก์ที่เกิดขึ้นด้วยเอนไซม์ snake venom phosphodiesterase (SVPD) และ calf intestinal phosphatase (CIP) สามารถพบผลิตภัณฑ์ที่เกิดจากการย่อยด้วยเอนไซม์จำนวน 3 ชนิด คือ dC-mechlorethamine-Cl (dC-mech-Cl), dC-mechlorethamine-OH (dC-mech-OH) และ dC-mechlorethamine-dC (dC-mech-dC) การตรวจสอบด้วยเทคนิคอีเอสไอ-แมสสเปกโทรเมตรี พบว่าเมคลอเรทามีนเกิดครอสลิงก์ระหว่างสายดีเอ็นเอที่ไนโตรเจนตำแหน่งที่ 3 ของคู่มิสแมตช์ไซโตซีน ซึ่งสอดคล้องกับรายงานก่อนหน้านี้ที่ตรวจสอบด้วยเทคนิค gel electrophoresis และที่สำคัญที่สุด คือ งานวิจัยนี้ประสบความสำเร็จในการพิสูจน์โครงสร้างระดับโมเลกุลของสารประกอบครอสลิงก์ของเมคลอเรทามีนกับดีเอ็นเอ

ภาควิชา อาหารและเภสัชเคมี	ลายมือชื่อนิสิต ณัฐพล จงอรุณงามแสง
สาขาวิชา เภสัชเคมี	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก
ปีการศึกษา 2552	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม

#5076564633 : MAJOR PHARMACEUTICAL CHEMISTRY
KEYWORDS : FRAGILE X SYNDROME, C-C MISMATCH PAIR, MECHLORETHAMINE, HPLC, MALDI-TOF-MS, ESI-MS

NUTTHAPON JONGAROONNGAMSANG: LIQUID CHROMATOGRAPHIC AND MASS SPECTROMETRIC DETERMINATIONS OF MECHLORETHAMINE-DNA CROSSLINKS. THESIS ADVISOR : ASSIST. PROF. PORNCHAI ROJSITTHISAK, Ph.D., THESIS CO-ADVOSOR : ASSOC. PROF. IAN S. HAWORTH, Ph.D., 125 pp.

The mechlorethamine crosslinking reaction with DNA at a cytosine-cytosine (C-C) mismatch pair has been used to probe the structure of DNA containing the d[GCC]•d[GCC] fragment associated with the triplet repeat expansion disease, Fragile X syndrome. The objectives of this work are to determine the structure and amount of the crosslink formed by mechlorethamine with a DNA duplex containing a C-C mismatch pair using HPLC and MALDI-TOF-MS, and to determine the target of mechlorethamine crosslinking in this duplex by enzymatic digestion, HPLC and ESI-MS/MS. HPLC analysis and molecular weight determination by MALDI-TOF-MS of the reaction products formed by mechlorethamine and the 15-mer DNA duplex d[CTCACACCGTGGTTC]•d[GAACCACCGTGTGAG] (underlined bases form a C-C mismatch pair) indicated formation of a mechlorethamine interstrand crosslink. HPLC and ESI-MS/MS analysis showed three products following enzymatic digestion of the crosslink by snake venom phosphodiesterase (SVPD) and calf intestinal phosphatase (CIP): dC-mechlorethamine-Cl (dC-mech-Cl), dC-mechlorethamine-OH (dC-mech-OH) and dC-mechlorethamine-dC (dC-mech-dC). Mass spectrometry also indicated that mechlorethamine forms the interstrand C-C crosslink through N^3 of each cytosine base. These results are consistent with results from published gel electrophoretic data. Most importantly, they provide the first evidence of the structural connectivity of the mechlorethamine C-C crosslink, and establish this crosslink as a new chemical species.

Department : Food and Pharmaceutical Chemistry Student's Signature Nutthapon J.
Field of Study : Pharmaceutical Chemistry Advisor's Signature Pornchai Rojsitthisak
Academic Year : 2009 Co-Advisor's Signature

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Assistant Professor Pornchai Rojsitthisak, Ph.D., for his valuable advice, continual guidance, kindness and understanding throughout my graduate study. His scientific open-mindedness encouraged me to complete this research successfully.

I am sincerely grateful to my co-adivisor, Associate Professor Ian S. Haworth, Ph.D. for his valuable suggestions and discussions that made my research more accurate. I also wish to express my appreciation to Assistant Professor Chamnan Patarapanich, Ph.D., Associate Professor Nuansri Niwattisaiwong and Associate Professor Opa Vajragupta, Ph.D. for serving on my thesis committee and providing valuable suggestions that improved this thesis.

I would like to thank all staffs at Faculty of Pharmaceutical Sciences, Chulalongkorn University, for their helps throughout my study.

Furthermore, I am grateful to Miss Prapapimol Pipitpiyapakorn for her help and encouragement. I also thank my colleagues in Dr. Pornchai's laboratory and my friends for their encouragement. Finally, I would like to thank my beloved parents and sister for their support throughout my studies.

CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

LIST OF ABBREVIATIONS

A	Adenine
ACN	Acetonitrile
C	Cytosine
CIP	Calf Intestinal Phosphatase
^{o}C	Degree Celsius
dA	2´-deoxyadenosine
dAMP	2´-deoxyadenosine monophosphate
dC	2´-deoxycytidine
dCMP	2´-deoxycytidine monophosphate
dC-mech-Cl	dC-mechlorethamine-Cl
dC-mech-OH	dC-mechlorethamine-OH
dC-mech-dC	dC-mechlorethamine-dC
dG	2´-deoxyguanosine monophosphate
dGMP	2´-deoxyguanosine
DMS	Dimethyl Sulfate
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DPAGE	Denaturing Polyacrylamide Gel Electrophoresis
dT	2´-deoxythymidine
dTMP	2´-deoxythymidine monophosphate
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray Ionization
FMR1	Fragile X Mental Retardation 1 gene

FMRP	Fragile X Mental Retardation Protein
FXRP	Fragile X–Related Proteins
G	Guanine
g	Gram
hnRNPs	Heterogenous Nuclear Ribinucleoproteins
HPLC	High Performance Liquid Chromatography
M	Molar
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization Time-of-Flight
mg	Milligram
min	Minutes
ml	Milliliter
mm	Millimeter
mM	Millimolar
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
MW	Molecular Weight
m/z	mass to charge ratio
nm	Nanometer
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
OD	Optical Density
Poly-dA	Polydeoxyadenylate
Poly-dC	Polydeoxycytidylate
Poly-dG	Polydeoxyguanylate

Poly-dT	Polydeoxythymidylate
RGG box	A Cluster of Arginine and Glycine Residues
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
SPR	Surface Plasmon Resonance
SVPD	Snake Venom Phosphodiesterase
T	Thymine
TEA	Triethylamine
TEAA	Triethylamonium acetate
TREDs	Triplet Repeat Expansion Diseases
UV	Ultraviolet
5′ UTR	5′ untranslated region
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar

CHAPTER I

INTRODUCTION

1.1 Background and significance

Fragile X syndrome is a genetic disease characterized by large expansions of $d[CGG]_n \bullet d[CCG]_n$ triplet repeat sequences within the 5′ untranslated region (5′ UTR) of the fragile X mental retardation gene (*FMR1*) (Bagni and Greenough, 2005; Di Prospero, 2005; Haas, 2007; Orr and Zoghbi, 2007). In the fragile-X sequence, each strand of the $d[CGG]_n \bullet d[CCG]_n$ duplex can form unusual intrastrand hairpin DNA conformations containing multiple G-G and C-C mismatch pairs. The structures formed by $d[CCG]_n$ repeat sequences have dynamic conformational properties due to the conformational fluctuation of C-C mismatch base pairs within the hairpin stem (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Gao et al., 1995; Mariappan et al., 1996, 1998; Mitas et al., 1995; Mitas, 1997; Nadel et al., 1995; Romero et al., 1999; Yu et al., 1997; Zheng et al, 1996). A schematic representation of the possible conformers with different alignments of $d[CCG]_n$ is shown in Figure 1.1. The dynamic conformational properties of $d[CCG]_n$ and the conformational fluctuation of C-C mismatch base pairs make it difficult to study the true structure of these conformers. Hence, a chemical agent that could stabilize these structures through covalent bond formation would make it easier to determine the preferred conformations of $d[CCG]_n$ sequences.

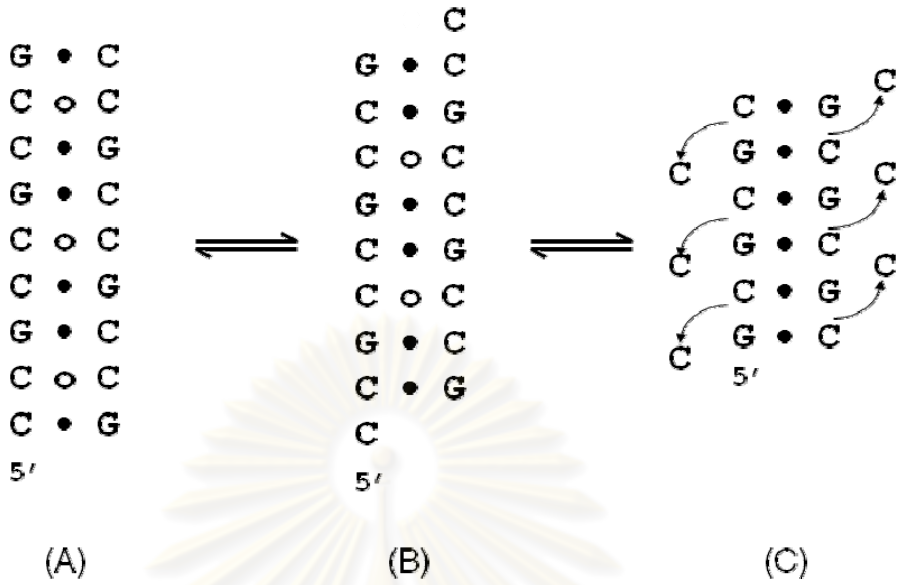


Figure 1.1 Schematic representation of the possible conformers of d[CCG]$_n$, showing molecules containing Watson-Crick pairs (●) and C-C mismatch pairs (○). Intrahelical (A and B) and extrahelical (C) C-C mismatch pairs within the DNA duplex.

Mechlorethamine (Figure 1.2) is a nitrogen mustard alkylating agent used for cancer treatment that reacts with nucleophilic centers, such as the guanine N7 atom, on the DNA duplex via an aziridinium intermediate. Because mechlorethamine is bifunctional, it can form an interstrand crosslink with suitable DNA sequences, leading to inhibition of DNA replication and cell division. Formation of a mechlorethamine-DNA interstrand crosslink has been suggested to occur through the N^7 position of the guanine (G) bases in the d[GXC]•d[GYC] duplex sequence, a so-called 1,3 G-G crosslink, when X-Y = C-G or T-A, as shown in Figure 1.3 (Rink et al., 1993; Rink and Hopkins, 1995[a]; Rink and Hopkins, 1995[b]). Recent studies by gel electrophoresis have shown that mechlorethamine can also form an interstrand DNA crosslink at a cytosine-cytosine (C-C) intrahelical mismatch pair and at an extrahelical C-C pair (Rojsitthisak et al., 2001). The detailed molecular structures of the C-C

interstrand crosslinks formed by mechlorethamine have not been proved, but it is likely that the reaction occurs in the DNA minor groove through the N^3 atom of cytosine (Figure 1.4) (Romero et al., 1999, 2001). Mechlorethamine crosslinking of a C-C pair coupled with gel electrophoresis has been used to probe the structure of DNA duplexes containing C-C mismatch pairs, with Rojsitthisak et al. (2001) finding that mechlorethamine forms crosslinks with both intrahelical and extrahelical C-C mismatch pairs in DNA containing d[GCC]•d$[GCC]_n$ (n=2 or 3) repeats sequences. These studies established several interesting features of the crosslinking reaction between mechlorethamine and a DNA duplex with intrahelical and extrahelical C-C mismatch pairs, but the detailed structures of these crosslinks have not been identified. In particular, the bond connectivity between the cytosine and the mechlorethamine is unknown.

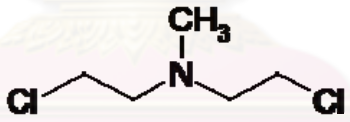


Figure 1.2 Structure of mechlorethamine

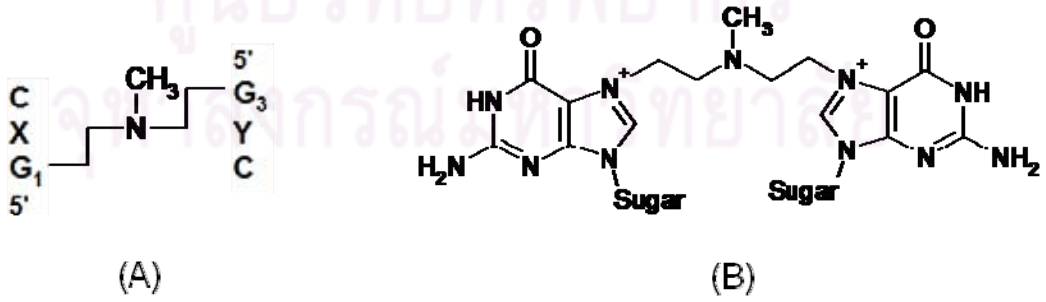
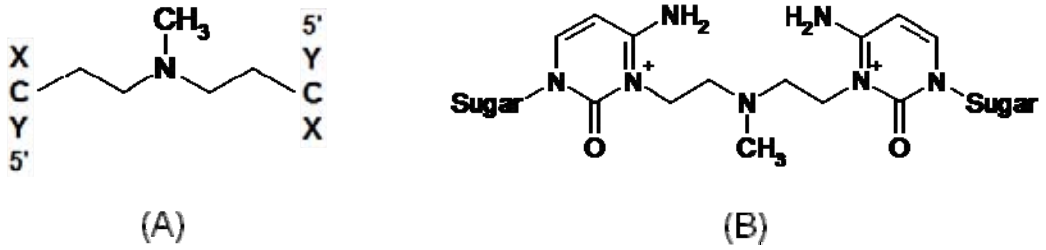


Figure 1.3 Representations of possible DNA interstrand crosslinks formed by mechlorethamine at a 1,3 G-G site (A) and the probable mechlorethamine-DNA interstrand crosslink through the guanine N^7 atoms at a 1,3 G-G site (B)



(A) (B)

Figure 1.4 Representation of DNA interstrand crosslink formed by mechlorethamine at a C-C mismatch pair (A) and the probable connectivity of the crosslink through the cytosine N^3 atoms of a C-C mismatch pair (B)

A suitable method for purification of the mechlorethamine C-C crosslink is necessary for structural characterization. High performance liquid chromatography (HPLC) and mass spectrometry (MS) are less time- and labor-intensive than gel electrophoresis and have been widely used for detection, purification and characterization of DNA crosslinks (Cummings et al., 1993; Gupta et al., 2005; Gut, 2004; Hartley et al., 1993; Hecker et al., 2001; Singh and Farmer; 2006; Tost and Gut, 2006; Winds et al., 2006). In 2001, Noll et al. succeeded in the use of enzymatic digestion, HPLC and MALDI-TOF-MS to purify and identify an N^4C-Ethyl-N^4C crosslink in a synthetic duplex (Noll et al., 2001). These techniques were subsequently used to purify and identify N^3T-Alkyl-N^3T and O^6G-Alkyl-O^6G crosslinks (Winds et al., 2004; Winds et al., 2006). In 2005, enzymatic digestion, HPLC and ESI-MS were used to determine the detailed structures of intrastrand crosslinks between cisplatin and single-strand DNA at GA and AG sites (Gupta et al., 2005), and the detailed structures of intrastrand and interstrand crosslinks formed by mechlorethamine between adenine and guanine bases have been characterized using HPLC and ESI-MS/MS (Balcome et al., 2004). These studies show that HPLC and mass spectrometry provide interesting and convincing results in DNA crosslink analysis. Therefore, in this study, methods of enzymatic digestion, HPLC and mass

spectrometry were developed to determine the structure of the DNA C-C crosslink formed by mechlorethamine.

1.2 Objectives

The goals of this research are to determine the connectivity, structure and amount of a mechlorethamine crosslink in a DNA duplex containing a C-C mismatch pair using HPLC, MALDI-TOF-MS, ESI-MS and enzymatic digestion.

1.3 Thesis overview

In this research, two designed 15-mer single-strand DNAs with sequences d[CTC ACA CCG TGG TTC] and d[GAA CCA CCG TGT GAG] were annealed to obtain double-strand DNA containing a central C-C mismatch pair (underlined C in each sequence). The DNA duplex was subsequently reacted with mechlorethamine. Crosslink formation between mechlorethamine and DNA duplex were proved by HPLC and MALDI-TOF-MS analysis. The mechlorethamine-DNA crosslink was purified by HPLC and the reaction site was identified by enzymatic digestion. The digested products were further characterized by HPLC and ESI-MS to examine the connectivity and structure of the crosslink.

1.4 Benefits

1. The molecular structure of the crosslink between mechlorethamine and DNA duplex containing a C-C mismatch pair will be identified and characterized.

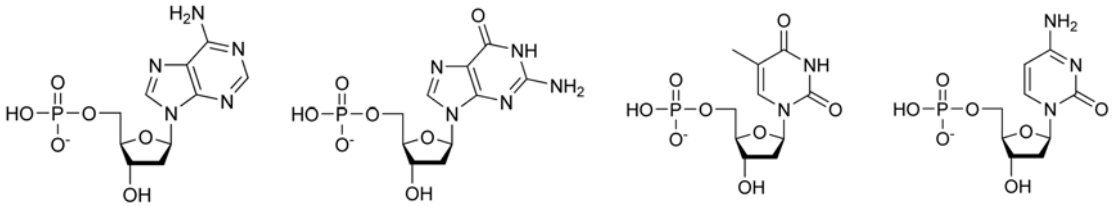
2. The developed method could be applied to determine the structure of DNA crosslinks with other alkylating agents such as nitrogen mustards and cisplatin.
3. The developed method could be used to compare the reactivity of different alkylating agents with DNA.
4. In a broader context, the work provides a basis for examination of complex DNA conformations that may be important in the pathogenesis of Fragile X syndrome.

CHAPTER II

LITERATURE REVIEW

2.1 DNA structures and properties

Deoxyribonucleic acid (DNA) is the molecular repository for genetic information. The structures of proteins and other cell constituents is a product of information programmed into the nucleotide sequence of the cell's DNA. DNA has three characteristic components: (1) nitrogenous bases, (2) a 2´-deoxyribose and (3) a phosphate group. The nitrogenous bases are derivatives of two parent compounds, purine and pyrimidine. The bases and deoxyribose found in common nucleotides are heterocyclic compounds. The carbon and nitrogen atoms in the parent structures are numbered to facilitate naming and identification of many derivative compounds. In the deoxyribonucleotide, the carbon atoms are given a prime (´) designation to distinguish them from the numbered atoms of nitrogenous bases. The base is joined covalently (at N^1 of pyrimidines and N^9 of purines) through an N-glycosidic linkage to the 1´ carbon of deoxyribose and the phosphate group is linked to the 5´ carbon through an ester. The glycosidic bond is formed by removal of a water molecule (a hydroxyl group from deoxyribose and hydrogen from base), as in O-glycosidic bond formation. Without the phosphate group, the molecule is called a deoxyribonucleoside. DNA contains two major purine bases, adenine (A) and guanine (G), and two major pyrimidines, cytosine (C) and thymine (T) (Lehninger et al., 1993). The structures of four nucleotides are shown in Figure 2.1 and the nomenclature of the nucleotides and nucleosides are summarized in Table 2.1.



Deoxyadenylate A, dA, dAMP | Deoxyguanylate G, dG, dGMP | Deoxythymidylate T, dT, dTMP | Deoxycytidylate C, dC, dCMP

Figure 2.1 Structures and nomenclature of the four deoxynucleotides

Table 2.1 Nucleotide and nucleoside nomenclature

Bases	Nucleosides	Nucleotides
Purine		
Adenine	Deoxyadenosine	Deoxyadenylate
Guanine	Deoxyguanosine	Deoxyguanylate
Pyrimidine		
Thymine	Deoxythymidine	Deoxythymidylate
Cytosine	Deoxycytidine	Deoxycytidylate

The nucleotide units of DNA are covalently linked through phosphate group. Specifically, the 5´-hydroxyl group of one nucleoside is joined to the 3´-hydroxyl group of the next unit by a phosphodiester linkage, as shown in Figure 2.2. Thus the covalent backbone of DNA consists of phosphate and deoxyribose residues and the characteristic bases may be regarded as side groups joined to the backbone at regular intervals. The backbone of DNA is hydrophilic because the hydroxyl groups of deoxyribose residues form hydrogen bonds with water. The phosphate groups in the polar backbone have a low pK_a and are completely ionized and negatively charged at pH 7; thus, DNA is an acid. These negative charges are generally neutralized by ionic interactions with positive charges on proteins, metal ions and polyamines. All phosphodiester linkages in DNA strands have the same orientation along the chain, giving the linear nucleic acid strand a specific polarity and distinct 5´ and 3´ ends. By

definition, the 5′ end lacks a nucleotide at the 5′ position and the 3′ end lacks a nucleotide at the 3′ position. Other groups may be present on one or both ends. By convention, the structure of a single strand of DNA is always written from the 5′ end to the 3′ end. As an example, the following are some simple representations of a pentadeoxyribonucleotide: p-A-C-G-T-AOH, pApCpGpTpA and pACGTA. However, the most common nomenclature in current use would represent this molecules as d[ACGTA], where d indicates deoxy (DNA, as opposed to RNA). A short DNA molecule of about 50 bases or less is referred to as an oligodeoxynucleotide. A longer DNA is called a polydeoxynucleotide (Lehninger et al., 1993). In each case, the deoxy is often omitted, but this does not correctly distinguish between DNA and RNA

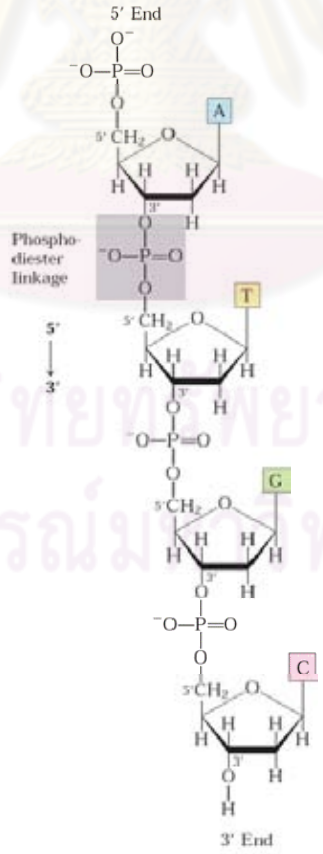


Figure 2.2 Phosphodiester linkages in the backbone of DNA (Nelson and Cox, 2004)

The bases have a variety of chemical properties that affect the structure and ultimately the function of nucleic acids. Free purines and pyrimidines are weakly basic compounds. The purines and pyrimidines common in DNA are highly conjugated molecules. This property has important effects on the structure, electron distribution and light absorption of nucleic acids. Resonance involving many atoms in the ring gives most of the bonds a partially double-bonded character. One result is that pyrimidines are plannar molecules; purines are very nearly plannar with a slight pucker. Free pyrimidine and purine bases may exist in two or more tautomeric forms depending upon the pH. The structures of purines and pyrimidines shown in Figure 2.1 are the tautomers that predominate at pH 7.0. As a result of resonance, all of the bases absorb UV light and nucleic acids are characterized by a strong absorption at wavelengths near 260 nm (Figure 2.3). The purines and pyrimidines are hydrophobic and relatively insoluble in water at the near neutral pH of the cell. At acidic and alkaline pH, the purine and pyrimidine become charged and their water solubility increases. Hydrophobic stacking interactions in which two or more bases are positioned with the planes of their rings parallel represent one of two important modes of interaction between two bases. The stacking involves a combination of van der Waals and dipole-dipole interactions between the bases. These base-stacking interactions help to minimize contact with water and are very important in stabilizing the three-dimensional structure of nucleic acids. The close interaction between stacked bases in DNA has the effect of decreasing the absorption of UV light relative to a solution with the same concentration of free nucleotides. The most important functional groups of purines and pyrimidines are ring nitrogens, carbonyl groups and exocyclic amino groups. Hydrogen bonds involving the amino and carbonyl groups are the second important mode of interaction between bases. Hydrogen bonds

between bases permit a complementary association of two and occasionally three strands of nucleic acid. The most important hydrogen-bonding patterns are those defined by James Watson and Francis Crick in 1953, in which A bonds specifically to T and G bonds to C (Figure 2.4). These two types of base pairs predominate in double-strand DNA due to these patterns. This specific pairing of bases permits the duplication of genetic information by the synthesis of nucleic acid strands that are complementary to existing strands (Lehninger et al., 1993).

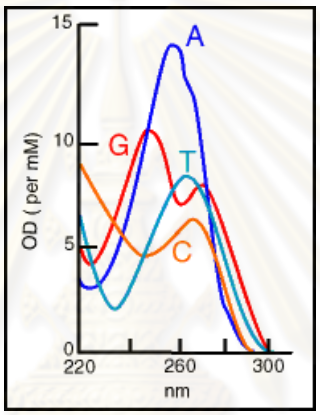


Figure 2.3 UV absorption spectra of the common nucleotides (From: http://www.scienceisart.com/A_DNA/UVspectrum_2.html)

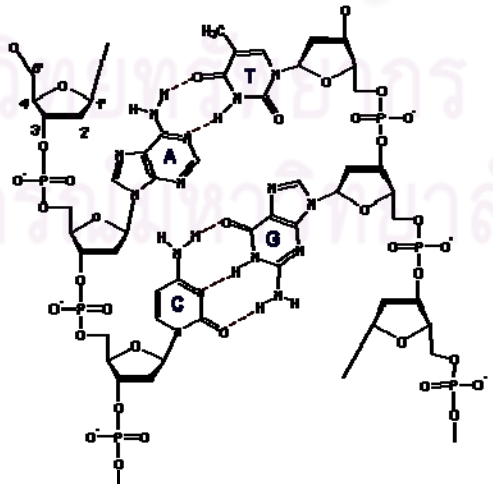


Figure 2.4 Watson-Crick hydrogen-bonding patterns in the base pairs (From: http://www.chemistry.nmsu.edu)

The most important clue to the structure of DNA came from the work of Erwin Chargaff and his colleagues in the late 1940s. They found that the four nucleotide bases in DNA occur in different ratios in DNA of different organisms and that the amounts of certain bases are closely related. These data, collected from DNAs of a great deal of different species, led Chargaff to the following conclusions:

1. The base composition of DNA generally varies from one species to another.
2. DNA specimens isolated from different tissues of the same species have the same base composition.
3. The base composition of DNA in a given species does not change with the organism's age, nutritional state or changing environment.
4. In all DNAs, regardless of the species, the number of adenine residues is equal to the number of thymine residues and the number of guanine residues is equal to the number of cytosine residues. From these relationships, it follows that the sum of the purine residues equals the sum of the pyrimidine residues.

These quantitative relationships, sometimes called "Chargaff's rules", were confirmed by many subsequent researchers. They were a key to establishing the three-dimensional structure of DNA and yielded clues to how genetic information is encoded in DNA and passed from one generation to the next (Lehninger et al., 1993).

Rosalind Franklin and Maurice Wilkins used the powerful method of X-ray diffraction to analyze DNA crystals. From this pattern, it was deduced that DNA polymers are helical with two periodicities along their axis, a primary one of 0.34 nm and a secondary one of 3.4 nm. The pattern also indicated that the molecule contains

two strands, a clue that was crucial to determine the structure. The problem then was to formulate a three-dimensional model of the DNA molecule that could account not only for the x-ray diffraction data but also for the specific A-T and C-G base equivalences discovered by Chargaff and for the other chemical properties of DNA (Lehninger et al., 1993).

In 1953, Watson and Crick postulated a three-dimensional model of DNA structure (Figure 2.5) that accounted for all of the available data. It consists of two helical DNA strands coiled around the same axis to form a right-handed double helix. The hydrophilic backbones of alternating deoxyribose and negatively charged phosphate groups are on the outside of the double helix, facing the surrounding water. The purine and pyrimidine bases of both strands are stacked inside the double helix with their hydrophobic and nearly planar ring structures very close together and perpendicular to the long axis of the helix. The spatial relationship between these strands creates a major groove and minor groove between two strands. Each base of one strand is paired in the same plane with a base of other strands. Watson and Crick found that the hydrogen-bonded base pairs illustrated in Figure 2.4 are fit best within the structure providing a rationale for Chargaff's rules. Three hydrogen bonds can form between G and C but only two can form between A and T. Other pairings of bases tend to destabilize the double helix structure. In Watson and Crick structure, the two chains or strands of the helix are antiparallel; their 5′, 3′-phosphodiester bonds run in opposite directions. The vertically stacked bases inside the double helix are 0.34 nm apart and the secondary repeat distance of about 3.4 nm could be accounted for by the presence of 10.5 nucleotide residues in each complete turn of double helix. The two antiparallel polynucleotide chains of double-helical of DNA are not identical

in either base sequence or composition. They are complementary to each other. Wherever adenine appears in one chain, thymine is found in the other; similarly, wherever guanine is found in one chain, cytosine is found in the other. The double helix or duplex is held together by two sets of forces, hydrogen bonding between complementary base pairs and base-stacking interactions. The specificity that maintains a given base sequence in each DNA strand is contributed entirely by hydrogen bonding between base pairs. The base-stacking interactions, which are largely nonspecific with respect to the identity of the stacked bases, make the major contribution to the stability of double helix (Lehninger et al., 1993).

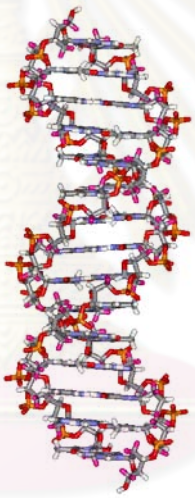
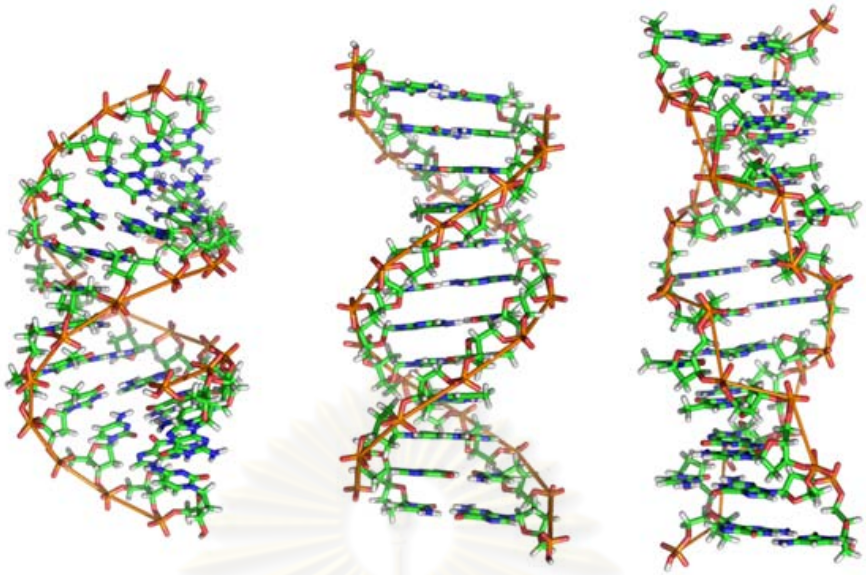


Figure 2.5 Watson and Crick model for the structure of DNA

DNA is a remarkably flexible molecule. Considerable rotation is possible around a number of bonds in the sugar-phosphate backbone and thermal fluctuation can produce bending, stretching and unpairing (melting) in the structure. Many significant deviations from the Watson-Crick DNA structure are found in cellular DNA and some or all of those may play important roles in DNA metabolism. These structural deviations generally do not affect the key properties of DNA defined by Watson and Crick (Lehninger et al., 1993).

The Watson-Crick structure is also called B-form DNA (Figures 2.5 and 2.6). The B form is the most stable structure for a random-sequence DNA molecule under physiological conditions and is therefore the standard point of reference in any study of properties of DNA. Two DNA structural deviants that have been well characterized in crystal structures are the A and Z forms (Figure 2.6). The A form is favored under conditions that are relatively devoid of water. The DNA is still arranged in a right-handed double helix but the rise per base pair is 0.23 nm and the number of base pairs per helical turn is 11, relative to the 0.34 nm rise and 10.5 base pairs per turn found in B-DNA. For a given DNA molecule, the A form is shorter and has a higher diameter than the B form. The reagents used to promote crystallization of DNA tend to dehydrate it and this leads to a tendency for many DNAs to crystallize in the A form. Z-DNA is a more radical departure from the B structure. The most obvious distinction is the left-handed helical rotation. There are 12 base pairs per helical turn with a rise of 0.38 nm per base pair. The DNA backbone takes on a zig-zag appearance. Certain nucleotide sequences fold up into left-handed Z helices more readily than others. Prominent examples are sequences in which pyrimidines alternate with purines, especially alternating C and G or 5-methyl-C and G. Whether A-form DNA actually occurs in cells is uncertain but there is evidence for some short stretches of Z-DNA in both prokaryotes and eukaryotes. These Z-DNA tracts may play an as yet undefined role in the regulation of expression of some genes or in genetic recombination. The three dimensional structure of A, B and Z form DNA are demonstrated in Figure 2.6.



(A-form DNA) (B-form DNA) (Z-form DNA)

Figure 2.6 Comparison of A, B and Z forms of DNA (From: http://www.mun.ca)

A number of other sequence-dependent structural variations have been detected that may serve locally important functions in DNA metabolism. For example, some sequences cause bends in the DNA helix. Bends are produced whenever four or more adenine residues appear sequentially in one of the two strands. Six adenines in a row produce a bend of about 18°. The bending observed with this and other sequences may be important in the binding of some proteins to DNA. A rather common type of DNA sequence is a palindrome, which is a region of DNA with inverted repeats of base sequences having two-fold symmetry over two strands of DNA (Figure 2.7). Such sequences are self-complementary within each strand and therefore have the potential to form the hairpin or cruciform (cross-shaped) structures shown in Figure 2.8. When the inverted repeat occurs within each individual strand of the DNA, the sequence is called a mirror repeat (Figure 2.7). Mirror repeats do not have complementary sequences within the same strand and cannot form hairpin or cruciform structures. Sequences of these types are found in virtually every large DNA

molecule and can encompass a few base pairs or thousands. The extent to which palindromes occur as cruciforms in cells is not known, although some cruciform structures have been demonstrated in vivo in *E.coli*. Self-complementary sequences cause isolated single strands of DNA in solution to fold into complex structures containing multiple hairpins (Lehninger et al., 1993; Nelson and Cox, 2004).

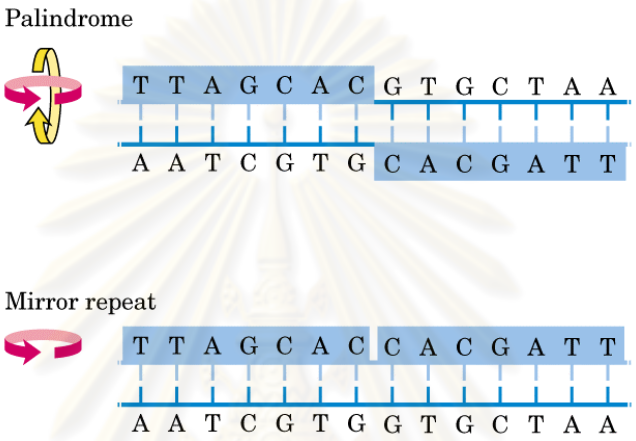


Figure 2.7 Palindromes and mirror repeats (Nelson and Cox, 2004)

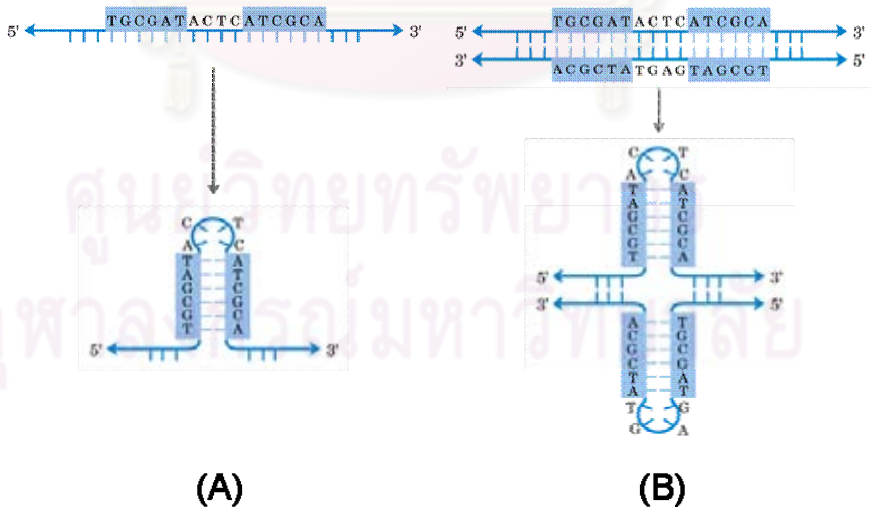


Figure 2.8 Structures of hairpin (A) and cruciform (B) (Nelson and Cox, 2004)
Several unusual DNA structures involve three or even four DNA strands.

These structural variations merit investigation because there is a tendency for many of them to appear at sites where important events in DNA metabolism (replication, recombination, transcription) are initiated or regulated. Nucleotides participating in a Watson-Crick base pair illustrated in Figure 2.4 can form a number of additional hydrogen bonds, particularly with functional groups arrayed in the major groove. For example, a protonated cytidine residue can pair with the guanosine residue of a G≡C nucleotide pair, and a thymidine can pair with the adenosine of an A=T pair (Figure 2.9). The N^7, O^6, and N^6 of purines, the atoms that participate in the hydrogen bonding of triplex DNA, are often referred to as Hoogsteen positions, and the non-Watson-Crick pairing is called Hoogsteen pairing, after Karst Hoogsteen, who in 1963 first recognized the potential for these unusual pairings. Hoogsteen pairing allows the formation of triplex DNAs. The triplexes shown in Figure 2.9 are most stable at low pH because the C≡G•C^+ triplet requires a protonated cytosine. In the triplex, the pKa of this cytosine is 7.5, altered from its normal value of 4.2. The triplexes also form most readily within long sequences containing only pyrimidines or only purines in a given strand. Some triplex DNAs contain two pyrimidine strands and one purine strand; others contain two purine strands and one pyrimidine strand. Four DNA strands can also pair to form a tetraplex or quadruplex, but this occurs readily only for DNA sequences with a very high proportion of guanosine residues (Figure 2.10A). The guanosine tetraplex, or G tetraplex, is quite stable over a wide range of conditions. The orientation of strands in the tetraplex can vary as shown in Figure 2.10B. A particularly exotic DNA structure, known as H-DNA, is found in polypyrimidine or polypurine tracts that also incorporate a mirror repeat. A simple example is a long stretch of alternating T and C residues (Figure 2.11). The H-DNA structure features the triple-stranded form illustrated in Figure 2.10. Two of the three

strands in the H-DNA triple helix contain pyrimidines and the third contains purines (Nelson and Cox, 2004).

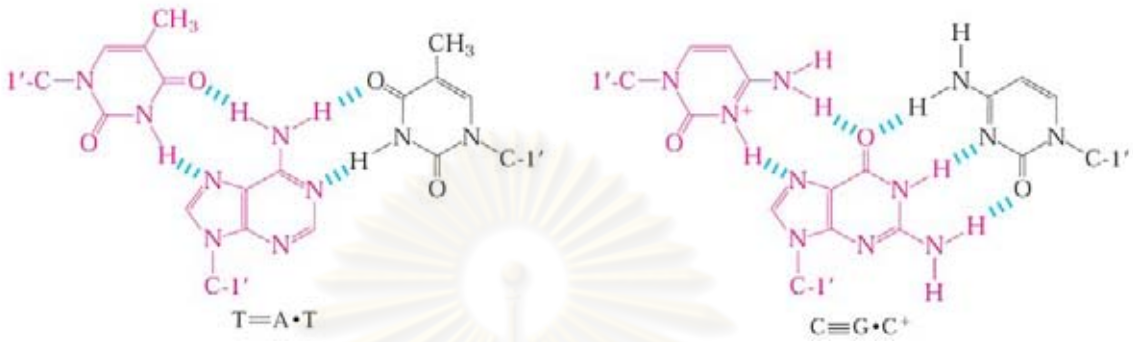


Figure 2.9 Base-pairing patterns in a well-characterized form of triplex DNA (Nelson and Cox, 2004)

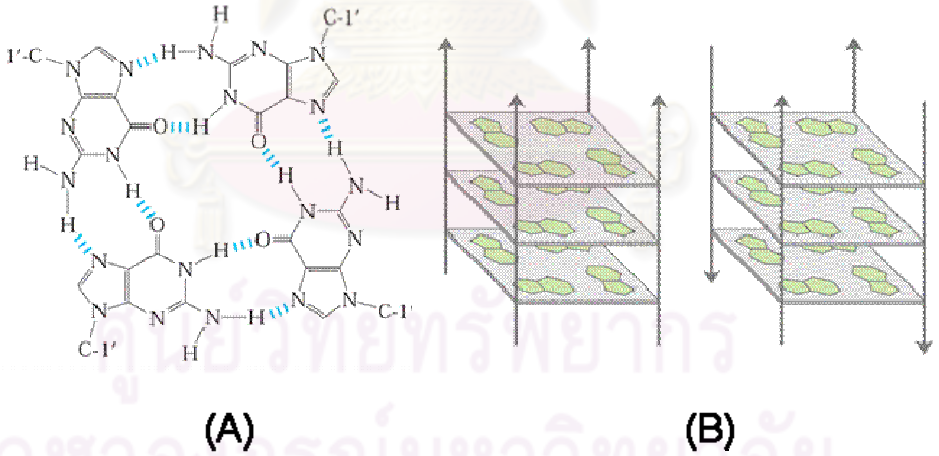


Figure 2.10 Base-pairing pattern in the guanosine tetraplex structure (A) and possible orientations of strands in a G tetraplex (B) (Nelson and Cox, 2004)

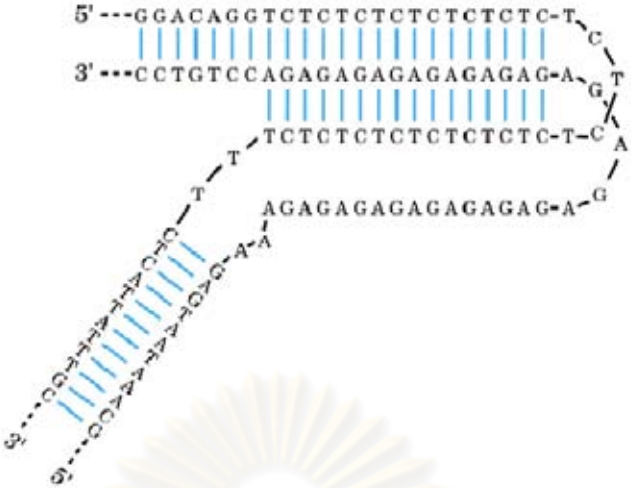


Figure 2.11 Structure of H-DNA (Nelson and Cox, 2004)

DNA replication and recombination may cause unusual base pairs that are referred to as “mismatch” pairs. Generally, mismatch pairs of DNA can be repaired by DNA repair systems in cells; however, the appearance of unrepaired mismatch pairs of DNA could cause genetic diseases and/or mutation. DNA mismatch pairs are also related to the secondary structure of DNA, such as hairpins, etc. Genes at the position of the secondary structure will function abnormally or not be expressed. The abnormal functioning or silencing of genes could cause genetic diseases. On such example of this phenomenon is caused by trinucleotide repeat sequences of DNA. The diseases caused by trinucleotide repeat sequences are called “trinucleotide repeat sequence disorders” or “triplet repeat expansion diseases” (TREDs).

2.2 Triplet repeat expansion diseases (TREDs)

The discovery that expansion of unstable triplet repeats can cause neurological disorders (Fu et al., 1991, La Spada et al., 1991) provided the first evidence that not all diseases causing mutations are stably transmitted from parent to offspring.

Moreover, the discovery of these dynamic mutations provided a molecular explanation for the variability in expressivity or severity of the disease phenotype: the larger the expansion, the earlier the onset and the more severe the course.

Triplet repeat expansions range from developmental childhood disorders such as X-linked mental retardation syndromes to the late onset neurodegenerative disorders such as Huntington disease and the inherited ataxias. The variability in repeat size underlies the broad spectrum of phenotypes seen in each of these disorders. The repeats show somatic and germ line instability. Successive generations of families affected by such dynamic mutations experience anticipation or earlier age of onset and more rapid disease progression owing to intergenerational repeat instability. For example, the onset of the neuromuscular disorder myotonic dystrophy ranges from birth in children and grandchildren to adulthood in parents and grandparents, depending on the size of the repeat (Orr and Zoghbi, 2007). At least 20 disorders have been identified as trinucleotide repeat diseases (Cummings et al., 2000; Orr and Zoghbi, 2007) or TREDs (Di Prospero and Fischbeck, 2005), as summarized in Table 2.2. Disorders that have triplet repeat expansions in non-coding regions typically cause a loss of gene function or toxic effects at the mRNA level, whereas those that occur in coding regions result in an expanded polyglutamine or polyalanine tract in the protein product, which causes the protein to become toxic, with or without the loss of its normal function (Di Prospero and Fischbeck, 2005).

Table 2.2 Triplet repeat expansion diseases (TREDs)

Disease	Gene	Locus	Protein	Repeat Sequence
Non-coding repeats				
Friedreich ataxia	*FXN*	9q13-q21.1	Frataxin	[GAA]•[TTC]
Fragile X syndrome A	*FMR1*	Xq27.3	FMR1 protein	[CGG]•[CCG]
Fragile X syndrome E	*FMR2*	Xq28	FMR2 protein	[CGG]•[CCG]
Dystrophia myotonica 1	*DMPK*	19q13	DMPK	[CTG]•[CAG]
Spinocerebella ataxia 8	Antisense to *KLHL1*	13q21	Undetermined	[CAG]•[CTG]
Spinocerebella ataxia 12	*PPP2R2B*	5q31-q33	Regulatory subunit of the protein phosphatase PP2A	[CAG]•[CTG]
Huntington disease-like 2	*JPH3*	16q24.3	Junctophillin 3	[CAG]•[CTG]
Polyglutamine disorders				
Spinal and bulbar muscular atrophy	*AR*	Xq13-q21	Androgen receptor	[CAG]•[CTG]
Huntington disease	*IT15*	4p16.3	Huntingtin	[CAG]•[CTG]
Dentatorubral-pallidoluysian atrophy	*DRPLA*	12p13.31	Atrophin 1	[CAG]•[CTG]
Spinocerebella ataxia 1	*SCA1*	6p23	Ataxin 1	[CAG]•[CTG]
Spinocerebella ataxia 2	*SCA2*	12q24.1	Ataxin 2	[CAG]•[CTG]
Spinocerebella ataxia 3 (Machado-Joseph disease)	*SCA3/MJD*	14q32.1	Ataxin 3	[CAG]•[CTG]
Spinocerebella ataxia 6	*CACNA1A*	19p13	α_{1A}-voltage-dependent calcium channel subunit	[CAG]•[CTG]
Spinocerebella ataxia 7	*SCA7*	3p12-p13	Ataxin 7	[CAG]•[CTG]
Spinocerebella ataxia 17	*TBP*	6q27	TATA box binding protein	[CAG]•[CTG]
Polyalanine disorders				
Oculopharyngeal dystrophy	*PABPN1*	14q11.2-q13	Poly(A)-binding protein 2	[GCG]•[CGC]
Congenital central hypoventilation syndrome	*PHOX2B*	4p12	Paired-like homeobox 2B	[GCN]•[MGC][a]
Infantile spasms	*ARX*	Xp22.13	Aristaless-related homeobox, X-linked	[GCN]•[MGC][a]
Synpolydactyly	*HOXD13*	2q31-q32	Homeobox D13	[GCN]•[MGC][a]

[a] N = A, T, C or G; M = base complementary to N

This table was modified from Di Prospero and Fischbeck, 2005 and Orr and Zoghbi, 2007 with additional information from Mitas, 1997; Urtizberea, 2004; Weese-Mayer et al., 2004 and Kato et al., 2007.

2.3 Intramolecular hairpin structure of DNA containing triplet repeat

Replication of duplex B-DNA requires separation of the two parental strands at the replication fork. It is during this time that stable single-strand DNA structures, such as hairpins comprised of inverted or triplet repeat sequences, have the opportunity to form. The presence of mismatches in these hairpins provides the ostensibly flexible B-DNA helix an opportunity to sample different H-bonding and/or stacking arrangements at the site of the mismatch pair. One goal in understanding mechanisms of triplet repeat expansion is to characterize the most stable base pairing arrangements of hairpins containing triplet repeat sequences and to determine the relevance of these structures to DNA replication and/or DNA repair. (Mitas, 1997)

The triplet repeat sequences, such as $d[CTG]_n$, $d[CAG]_n$, $d[CCG]_n$ and $d[CGG]_n$, in single strand DNA molecules can form intrastrand hairpin conformations (Darlow and Leach, 1998[a], 1998[b]; Mitas, 1997; Nadel et al., 1995). These hairpin conformations contain a mismatch pair (T-T, A-A, G-G and C-C) flanked by two normal base pairs (C-G and G-C) in every three base pairs. Triplet repeat sequences of $d[CTG]_n$ and $d[CAG]_n$ in single-strand DNA can separately form hairpin structures containing T-T and A-A mismatch pairs, respectively (Figure 2.12A and B) (Hagihara and Nakatani, 2006; Mitas, 1997; Nakayabu et al., 1998; Petruska et al., 1996). In contrast, triplet repeat sequences of $d[CGG]_n$ and $d[CCG]_n$ in single-strand DNA can each form hairpin structures containing G-G and C-C mismatch pairs, respectively, with two different alignments (Figure 2.12C and D) (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Mitas, 1997; Yu et al., 1997).

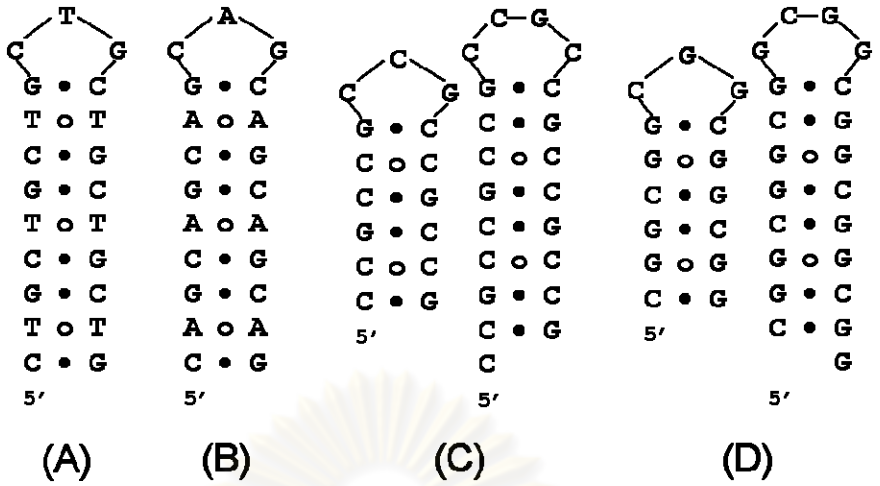


Figure 2.12 Hairpin conformations formed by triplet repeat sequences, d[CTG]$_n$ (A), d[CAG]$_n$ (B), d[CCG]$_n$ (C) and d[CGG]$_n$ (D) in single-strand DNA. For (C) and (D), the left and right hairpins include d[GC]•d[GC] and d[CG]•d[CG], respectively. A filled bullet (•) and an unfilled bullet (**o**) indicate normal and mismatched base pairs, respectively

2.4 Fragile X Syndrome

Fragile X syndrome, a triplet repeat expansion disease, is the most common forms of inherited mental retardation, affecting approximately 1 in 4000 males and 1 in 8000 in females (Haas, 2007; Orr and Zoghbi, 2007). The syndrome is caused by an expansion of the triplet repeats, d[CCG]$_n$•d[CGG]$_n$, at the 5′ untranslated region (5′ UTR) of the fragile X mental retardation gene (*FMR1*), which encodes fragile X mental retardation protein (FMRP). Hypermethylation of the d[CCG]$_n$•d[CGG]$_n$ region and the upstream CpG islands usually lead to decreased transcription and silencing of the gene (Bagni and Greenough, 2005; Di Prospero and Fischbeck, 2005; Haas, 2007; Mariappan et al., 1998). The syndrome is transmitted as an X-linked dominant trait and is more severe in males (Kaufmann and Reiss, 1999). The clinical

features and medical problems of fragile X syndrome are summarized in Table 2.3, Figure 2.13 and Figure 2.14.

Table 2.3 Clinical features and medical problems of fragile X males*

Clinical features	**%**
Long face	70
Prominent ears	70
High arched palate	52
Hyperextensible finger joints	67
Double jointed thumbs	53
Single palmar creases	25
Hand calluses	29
Flat feet	71
Heart murmur or click	18
Macro-orchidism	70
Medical problems	**%**
Vomiting	31
Failure to thrive in infancy	15
Strabismus	36
Myopia or hyperopia	22
Hernia	15
Joint dislocation	3
Orthopaedic problems	21
Otitis media	85
Sinusitis	23
Seizures	22
Mitral valve prolapsed	35
Apnoea	10
Autism	20
ADHD	80
Motor tics	19

*Majority of items scored in 100-280 patients. This table was modified from De Vries et al., 1998.

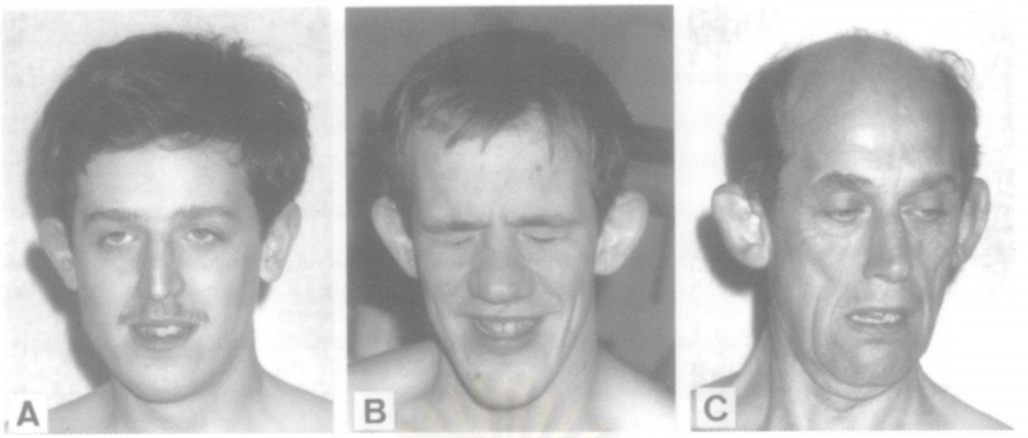


Figure 2.13 Three males with characteristic facial features of fragile X syndrome: long, narrow face with large, everted ears and poor eye contact (De Vries et al., 1998)

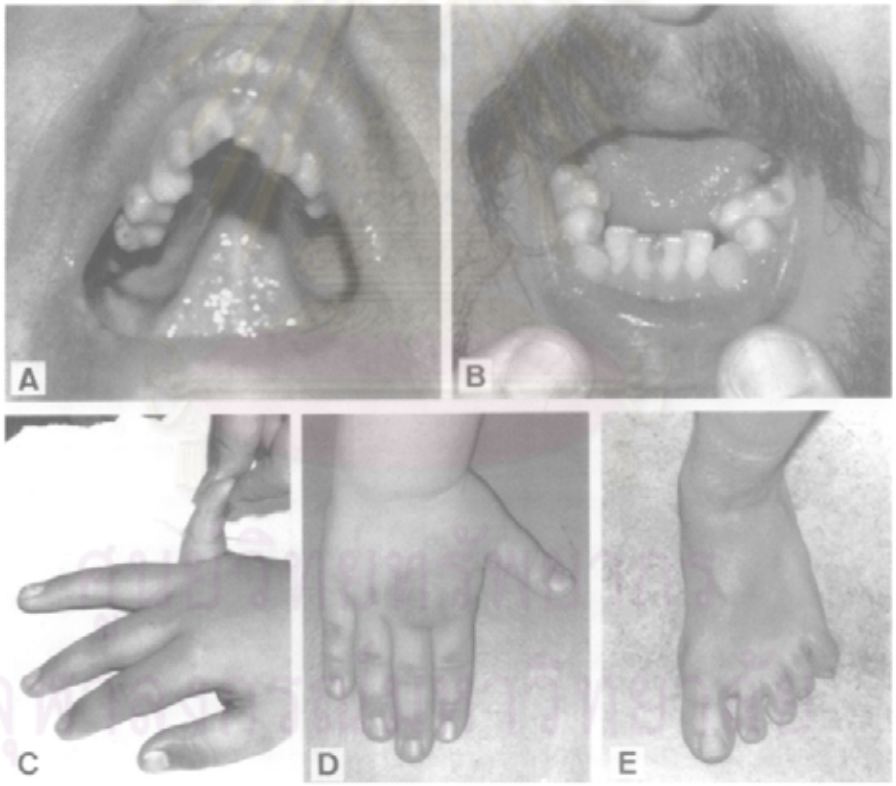


Figure 2.14 Some clinical features in fragile X patients. High arched palate (A), dental crowding (B), hyperextensible MP joints (C), hand calluses (D), and pes planus (E) (De Vries et al., 1998)

Fragile X syndrome is diagnosed by cytogenic detection of the fragile site at Xq27.3, termed FRAXA (FRAgile site, X chromosome, A site) (De Vries et al., 1998; Kaufmann and Reiss, 1999; Orr and Zoghbi, 2007). The *FMR1* within the fragile site at Xq27.3, which is responsible for fragile X syndrome, has a size of 33 kb with 17 exons and a 5′UTR containing $d[CCG]_n \bullet d[CGG]_n$ triplet repeats in the first exon. (Bagni and Greenough, 2005; De Vries et al., 1998). The CpG island, the regulatory region of *FMR1*, is located in the 5′ upstream region of $d[CCG]_n \bullet d[CGG]_n$ triplet repeats (Bagni and Greenough, 2005).

At the molecular level, massive expansion of $d[CCG]_n \bullet d[CGG]_n$ triplet repeats in the 5′UTR of *FMR1* are associated with fragile X syndrome (Bagni and Greenough, 2005; Haas, 2007; Mariappan et al., 1998; Orr and Zoghbi, 2007). Patients with fragile X syndrome have more than 200 repeats (full mutation). In contrast, the number of $d[CCG]_n \bullet d[CGG]_n$ repeats varies from 7 to 50 (with an average of 30 units) in the normal population and from 50 to 200 in fragile X carriers (premutation). The CGG-related mutations are summarized in Figure 2.15. The molecular consequence of repeat expansions is RNA-transcriptional silencing of *FMR1* because of hypermethylation of the regulatory CpG island of the gene (Bagni and Greenough, 2005; De Vries et al., 1998; Haas, 2007; Orr and Zoghbi, 2007). Examination of fragile X patients has consistently demonstrated that hypermethylation of the CpG island is the primary factor in the fragile X phenotype, with methylation being responsible for approximately 99% of known fragile X phenotypes (Haas, 2007). This increased methylation coupled with decreased histone acetylation in the 5′ of the gene leads to an absence or reduction of *FMR1* expression and FMRP function as the

molecular basis of fragile X syndrome (Bagni and Greenough, 2005; Haas, 2007; Orr and Zoghbi, 2007).

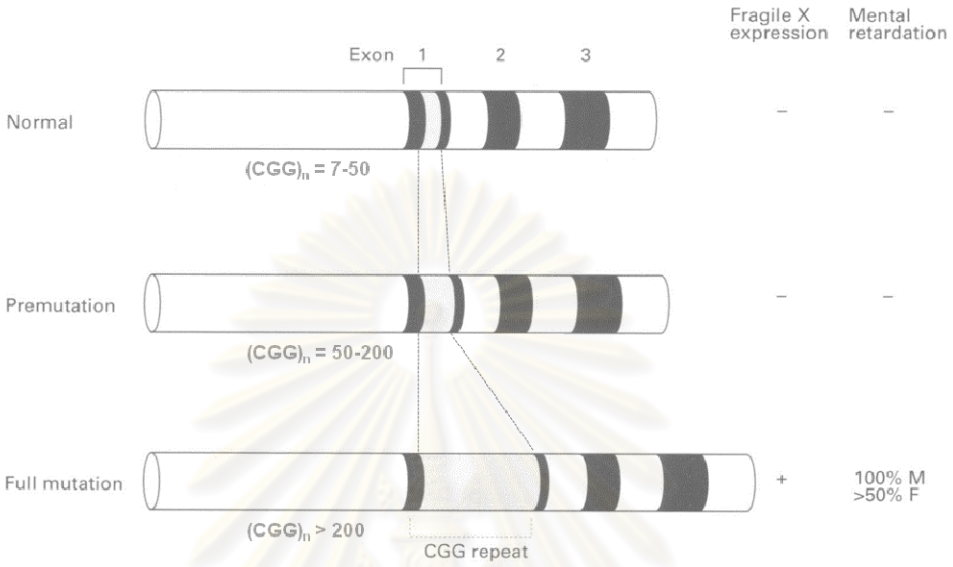


Figure 2.15 CGG-related mutations in the first exon of *FMR1* (modified from De Vries et al., 1998)

FMRP is one of a family of RNA-binding proteins known as heterogenous nuclear ribonucleoproteins (hnRNPs) that are involved in many aspects of mRNA metabolism and biology (Bagni and Greenough, 2005). The domain structure of FMRP includes three RNA-binding domains, two ribonucleoprotein K homology domains (KH1 and KH2 domains) and a cluster of arginine and glycine residues (RGG box) that supports RNA binding (Bagni and Greenough, 2005; Darnell et al., 2001; Siomi et al., 1993). In 1995, two proteins produced from *FMR1* (FXR1 and FXR2 gene) were identified and their structures are similar to that of FMRP. Because of their similarities, it has been predicted that FXR1P and FXR2P, members of this family of fragile X–related proteins (FXRP), might compensate for the functions of FMRP in fragile X patients (Orr and Zoghbi, 2007; Zhang et al., 1995). In order to

bind mRNAs, FMRP forms an RNP complex containing FMRP, FRX1P, FRX2P, nucleolin and three other proteins (Haas, 2007). This RNP complex is involved in the transport and translation of mRNA in neurons, and selectively binds mRNA by through association of G quartet structure of mRNA with the RGG box of FMRP (Darnell et al. 2001; Haas, 2007). In addition, the KH2 domain recognizes a more intricate tertiary structure in RNA targets, termed the FMRP-kissing complex (Darnell et al. 2005).

FMRP is required for export of mRNAs from the nucleus and their subcellular localization to cytoplasm. These two processes are believed to be connected (Van de Bor and Davis, 2004): when RNA processing is complete, the RNA is exported through nuclear pores. At this stage, some RNA-binding proteins are released whereas others remain attached (Farina and Singer, 2002). In the nucleus, one role of FMRP could be to associate with mRNAs and escort them out of the nucleus. FMRP contains both a functional nuclear localization signal (NLS) and a nuclear export signal (NES) which indicates that it can shuttle between the nucleus and the cytoplasm. Chromatin remodeling may be a second role for FMRP in the nucleus (Bagni and Greenough, 2005). FMRP has also been found to be associated with ribosomes in the dendritic structure of neurons, indicating a possible role of the protein in dendritic structure and neuronal plasticity (Feng et al., 1997), and FMRP may play a major role in development of neuronal structure throughout the body (Haas, 2007).

Because the expansion of $d[CCG]_n \bullet d[CGG]_n$ triplet repeats cause fragile X syndrome, the secondary structure of DNA containing $d[CCG]_n \bullet d[CGG]_n$ triplet repeats has been studied by several research groups.

2.5 Intramolecular hairpin structures of $d[CCG]_n \bullet [CGG]_n$

Expansion and hypermethylation of the fragile X DNA triplet repeat, $d[CCG]_n \bullet [CGG]_n$, are correlated with the ability of the individual G- and C-rich strands to form hairpin structures independently. Because of this, the conformational properties of single-strand $d[CCG]_n$ and $d[CGG]_n$ have been studied (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Gao et al., 1995; Mariappan et al., 1996, 1998; Mitas et al., 1995; Mitas, 1997; Nadel et al., 1995; Yu et al., 1997; Zheng et al, 1996). Two-dimensional NMR and gel electrophoresis have shown that both the G- and C-rich single strands, $d[CGG]_n$ and $d[CCG]_n$ respectively, form hairpins under physiological conditions. The G-rich strand forms hairpins and duplexes, depending on the repeat number and salt concentration. For short $d[CGG]_n$ (n = 4-11), the hairpin is the predominant conformation at low salt (5 mM NaCl) and the duplex is the major conformer at high salt (200 mM NaCl). When the repeats are longer than n = 11, the hairpin is the predominant conformation at all salt concentrations. In contrast, the C-rich strand forms hairpins only at all salt concentrations. Therefore, the C-rich strand more readily forms hairpin or slippage structures compared to the G-rich strand. NMR data also demonstrated that hairpins formed by the C-rich strand fold in such a way that the cytosine of the CpG step of the stem is a C-C mismatch pair (Chen et al., 1995 and Mariappan et al., 1996). The presence of C-C mismatch pairs generates local flexibility, thereby providing methyltransferase transition state analogs (Chen et al., 1995). These results suggest that the hairpins of the C-rich strand act as better methylation substrates for human methyltransferase, compared to the Watson-Crick duplex and G-rich strand, whereas those of the G-rich strand are specific methylation substrates for bacterial methyltransferase (Table 2.4) (Chen et al, 1995).

The duplex and hairpin structures of the C-rich strand, $d[CCG]_n$, and the G-rich strand, $d[CGG]_n$ are shown in Figure 2.16.

Table 2.4 Rate of methylation of the cytosine at the CpG step of triplet repeats (Chen et al. 1995)

Methyltransferase	Repeats (n)	Methylation rate (fmol/min)		
		CCG	CGG	Watson-Crick
Sss I (Bacterial)	5	-	-	-
	6	-	2.6	2.0
	7	-	24	23
	11	7.1	30	21
Human	5	3.9	1.1	1.6
	6	12	1.7	4.9
	7	23	9.5	14
	11	290	27	54

All values are the averages of four measurements. A dash indicates too low to measure. Hairpins contain half the potential methylation sites present in the Watson-Crick duplex.

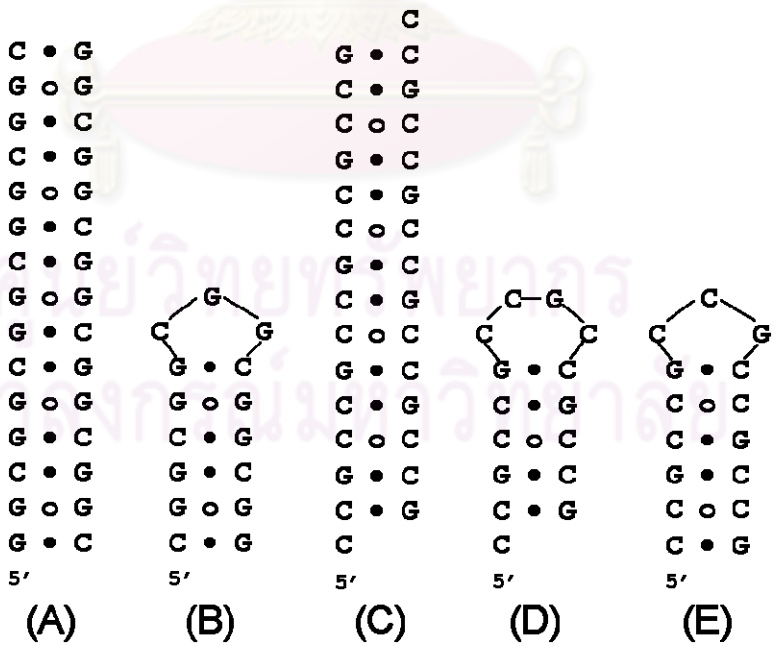


Figure 2.16 Self-assembled structures of $d[CGG]_5$ duplex (A), $d[CGG]_5$ hairpin (B), $d[CCG]_5$ slipped duplex (C), $d[CCG]_5$ hairpins (D), and blunt $d[CCG]_5$ hairpin (E)

Mitas et al. used gel electrophoresis, chemical modification with DMS and computer modeling to study the conformations of single-strand $d[CGG]_{15}$. The results suggested that the single-strand $d[CGG]_{15}$ forms an intramolecular hairpin that contains $G^{syn}•G^{anti}$ base pairs in the stem and the G residues in G•G base pairs that alternate between *anti* and *syn* conformations (Mitas et al, 1995). The conclusion that $G^{syn}•G^{anti}$ base pairs are contained in the stem of hairpin was supported by NMR data on the duplex structures of $d[CGG]_n$ (Chen et al, 1995). Gel electrophoresis and computer modeling studies performed by Nadel et al. indicated that the $d[GGC]_n$ hairpins are stabilized by guanine-guanine Hoogsteen hydrogen bonds or by Hoogsteen Watson-Crick bonds (Nadel et al, 1995).

The above-mentioned NMR, gel electrophoresis and computer modeling data were used to distinguish between two different alignments of hairpins formed by single-strand $d[CGG]_n$ triplet repeat sequences. These two alignments contain either a d[GC]•d[GC] (Figure 2.17A) or a d[CG]•d[CG] (Figure 2.17B) dinucleotide steps with normal base pairs, together with a G-G mismatch pair (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Mitas et al., 1995; Nadel et al., 1995; Mariappan et al., 1996). The hairpin with the d[GC]•d[GC] step is the preferred alignment (Figure 2.17A) (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Mitas, 1997).

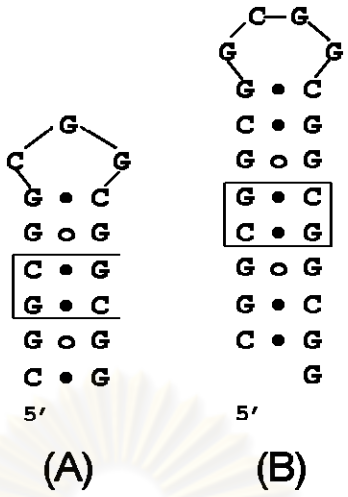


Figure 2.17 Hairpin conformations formed by d[CGG]$_n$ with d[GC]•d[GC] (A) and d[CG]•d[CG] (B) dinucleotide steps in single-strand DNA. Filled (•) and unfilled (**o**) bullets exhibit normal and mismatched base pairs, respectively.

The conformational properties of d[CCG]$_n$ repeat sequences have been described by various groups as dynamic hairpins that are excellent substrates for human methyltransferases (Chen et al, 1995; Mariappan et al., 1996) and are not effectively repaired (Mitas, 1997; Peyret et al., 1999). In 1995, Gao et al. characterized the short DNA duplex of d[CCG]$_2$ containing two C-C mismatch pairs (Figure 2.18A) by NMR spectroscopy. The NMR data demonstrated that the two C-C mismatch pairs generated flexibility of the duplex. Under physiological conditions, the position of the C-C mismatch pairs changed dynamically and subsequently formed a slipped DNA duplex with 5´-cytosine overhangs (Figure 2.18B) and an extrahelical cytosine in the minor groove of DNA, termed an e-motif (Figure 2.18C). The slipped DNA duplex and formation of e-motif demonstrates the extreme flexibility of a DNA duplex containing CCG repeats (Gao et al., 1995).

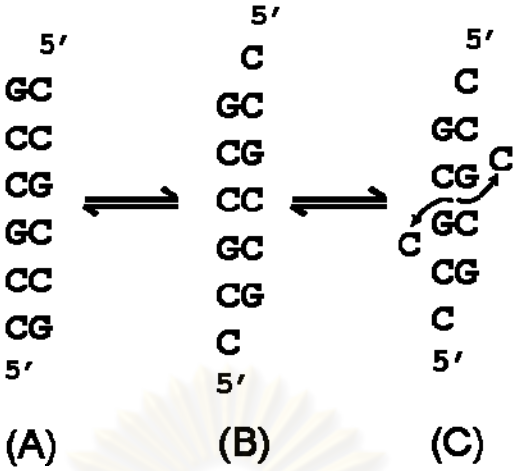


Figure 2.18 DNA duplex conformations of d[CCG]$_2$. Normal duplex containing two C-C mismatch pairs (A), slipped DNA duplex (B), and extrahelical cytosine or e-motif (C).

In 1997, Yu et al. investigated the potential structures of a single-strand oligonucleotide containing d[CCG]$_{15}$ using the pH and temperature dependence of electrophoretic mobility, UV absorbance, circular dichroism, chemical modification and P1 nuclease digestion. Single-strand d[CCG]$_{15}$ has a pK_a of 7.7±0.2. At pH 8.5, d[CCG]$_{15}$ forms a relatively unstable hairpin containing a CpG base-pair step. At physiological pH (pH 7.5), the hairpin contains protonated cytosines but no detectable C•C$^+$ base pairs, increased thermal stability, increased stacking of the CpG base-pair steps and a single cytosine that is flipped away from the central portion of the helix. Examination of single-strand d[CCG]$_{18}$ and d[CCG]$_{20}$, which were designed to adopt hairpins containing alternative GpC base-pair steps, revealed hairpins containing a CpG base-pair step, had pK_as of 8.2 and 8.4, respectively, and distorted helices. The results suggested that DNA sequences containing d[CCG]$_n$ ($n \geq 15$) adopt hairpin conformations containing CpG rather than GpC base-pair steps. The formation of a

slipped hairpin and an extrahelical cytosine (Figure 2.19) is consistent with the NMR data for the duplex structure of $d[CCG]_2$ (Gao et al., 1995). Additionally, the C-C mismatch pairs of the stem are protonated at physiological pH, but not hydrogen-bonded (Yu et al, 1997).

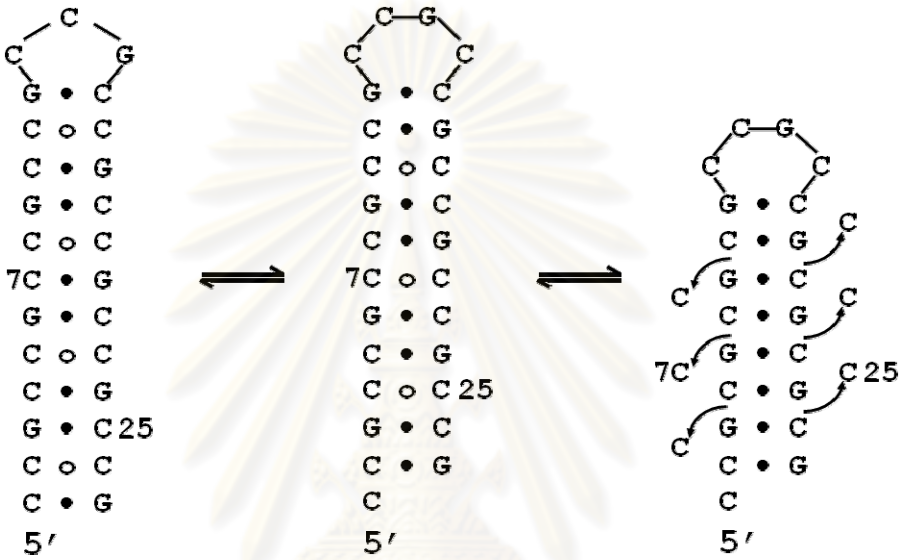


Figure 2.19 Hairpin conformations of single-strand $d[CCG]_n$

In 1998, Mariappan et al. performed heteronuclear (^{15}N-^{1}H) magnetic resonance spectroscopy to probe the structure of the CpG sites in the $d[CCG]_n$ hairpins that were ^{15}N-labeled at the amino (N^4) groups of specific cytosine bases. The analysis of chemical shift, pH-induced chemical exchange and NOE pattern of the (^{15}N-labeled) amino protons of cytosines revealed that the cytosine bases at CpG sites are intrahelical and well-stacked with the neighboring G•C base-pairs in the stem of these hairpins and possibly form single hydrogen-bonded C-C mismatch pairs. One simple explanation of single hydrogen bond formation at C-C mismatch pairs was offered: a single broad resonance was found corresponding to the amino proton of

cytosine that is hydrogen-bonded in a C-C mismatch pair, whereas the non-hydrogen-bonded proton is constantly undergoing solvent exchange and would not be observed in the spectrum. When the areas under the amino signal corresponding to the C-C mismatch pair were computed at different pH values and compared with the internal amino signal of the G•C pair, the relative area was 60 to 75% within the pH range of 6 to 7. Because of the pH-sensitive amino proton of the C-C mismatch pair, the relative area of 60 to 75% may be contributed by a single proton. The amino protons from the C-C mismatch pair exchange-broaden beyond detection above pH 7.66, whereas those originating from the G•C pair are intact up to pH 8.11 and completely disappeared only above pH 8.5. The measurements of the pH-dependent ^{1}H line-width also demonstrated that the C-C mismatch pairs are more susceptible to opening/closure than G•C base-pairs. Thus, the cytosines at the CpG sites of the $d[CCG]_n$ hairpin are flipped out more easily to the activated state than those in the corresponding Watson-Crick duplex, $d[CCG]_n$•$d[CGG]_n$ and this makes the hairpin a better target for methylation by the human methyltransferase (Mariappan et al., 1998).

According to the experimental data, two different alignments of hairpin formed by $d[CCG]_n$ triplet repeat sequences in single-strand DNA were determined. These two alignments contained either a d[GC]•d[GC] or a d[CG]•d[CG] normal base pair with a C-C mismatch pair (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Mitas, 1997; Mariappan et al., 1996; Yu et al., 1997). For 5-7 repeats of $d[CCG]_n$, the hairpin with d[GC]•d[GC] dinucleotide step is the preferred conformation (Figure 2.20A); however, for 15 or longer repeats, $d[CCG]_n$ prefers a distort hairpins with d[CG]•d[CG] dinucleotide step (Figure 2.20B) (Chen et al., 1995; Darlow and Leach, 1998[a], 1998[b]; Mitas, 1997; Yu et al., 1997).

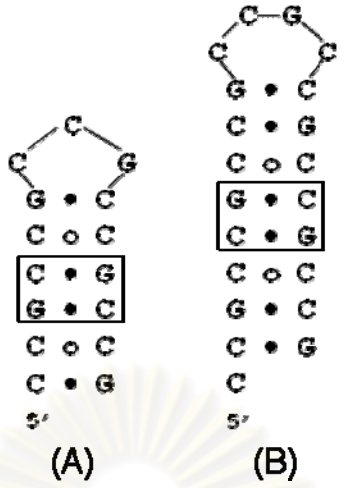


Figure 2.20 Hairpin conformations formed by d[CCG]$_n$ with d[GC]•d[GC] (A) and d[CG]•d[CG] (B) dinucleotide steps in single-strand DNA. Filled (•) and unfilled (**o**) bullets indicate normal and mismatched base pairs, respectively.

2.6 Mechlorethamine-DNA crosslinking reaction

Mechlorethamine (bis(2-chloroethyl)methylamine, Figure 1.2) was discovered in the 1940s as the first clinically useful anticancer drug. It is a nitrogen mustard that alkylates nucleic acids predominantly at the N7 of deoxyguanosine (dG) residues in a d[GXC]•d[GYC] duplex sequence (a so-called 1,3 G-G crosslink), when X-Y = C-G or T-A, as shown in Figure 1.3A and B (Rink et al., 1993; Rink and Hopkins, 1995[a]; Rink and Hopkins, 1995[b]). The ability of mechlorethamine to inhibit protein, RNA and DNA synthesis provided early evidence for the hypothesis that DNA is the biologically relevant target of mechlorethamine (Rink and Hopkins, 1995[a]). Mechlorethamine is a bifunctional electrophilic molecule that can form both intrastrand and interstrand covalent crosslinks in DNA by nucleophilic substitution reactions via an aziridinium ion intermediate (Balcome et al., 2004; Rink et al., 1993;

Rink and Hopkins, 1995[a]; Rink and Hopkins, 1995[b], Williams and Lemke, 2002). The formation of the crosslinks is shown in Figure 2.21.

The antitumor activity of mechlorethamine and other nitrogen mustards may be due to formation of guanine-guanine interstrand crosslinks in DNA duplexes (Rink and Hopkins, 1995[a], Rink et al., 1993). This reaction results in a 1,3 G-G crosslink in the major groove of DNA duplex. The 1,3 G-G crosslink is preferentially formed over the 1,2 G-G crosslink (Rink et al., 1993; Williams and Lemke, 2002). Molecular mechanics energy minimization indicated that the mechlorethamine-DNA interstrand crosslink induces a bend in the helical axis of the DNA duplex (Rink and Hopkins, 1995[a]). Once DNA alkylation occurs, the alkylated sites become prone to cleavage, resulting in formation of single strand DNA breaks (Williams and Lemke, 2002).

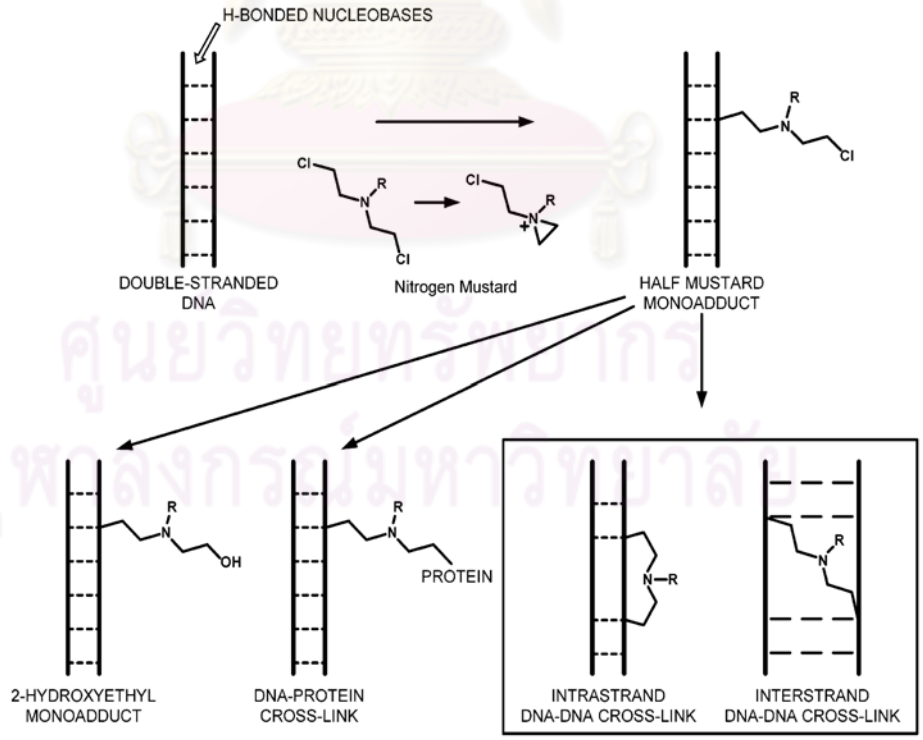


Figure 2.21 Formation of mechlorethamine-crosslinked DNA duplex (Balcome et al., 2004)

Other sites of DNA bases and phosphate oxygens of DNA backbone may also be alkylated by mechlorethamine (Williams and Lemke, 2002). For example, the formation of guanine-adenine (G-A) and adenine-adenine (A-A) adducts of mechlorethamine in double-strand DNA has been demonstrated. 1,3 G-A interstrand crosslinks of N^3A-N^7G-EMA, N^6A-N^7G-EMA and N^1A-N^7G-EMA at position (Figure 2.22A, B and C), while N^3A-N^7G-EMA is predominant in the intrastrand 1,2 G-A crosslink. The prevalent adenine-adenine crosslink is *bis*-N^3A-EMA, as shown in Figure 2.22D (Balcome et al, 2004).

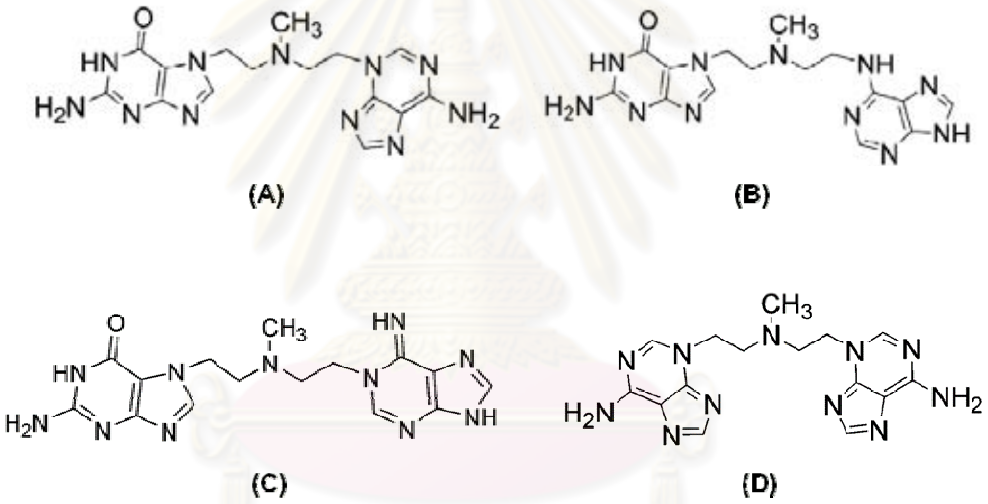


Figure 2.22 Structures of N^3A-N^7G-EMA (A), N^6A-N^7G-EMA (B), N^1A-N^7G-EMA (C) and *bis*-N^3A-EMA (D) (Balcome et al, 2004)

2.7 The mechlorethamine-DNA interstrand crosslink at a C-C mismatch pair

In 1999, Romero et al. discovered the formation of a mechlorethamine-DNA interstrand crosslink at a C-C mismatch pair (Figure 1.4A) using gel electrophoresis. The mechlorethamine C-C crosslink formed preferentially between two mismatched cytosine bases rather than between guanine bases and occurred regardless of the

flanking base pairs (Romero et al, 1999). Piperidine cleavage of crosslinked species containing the $d[GCC]_n \bullet d[GCC]_n$ sequence gives DNA fragments consistent with an alkylation product at the mismatched cytosine bases. The detailed structure of the crosslink was not determined yet, but the crosslink was suggested to form between N^3 of the two cytosine bases since these positions are the most nucleophilic atoms of the mismatch pair. This was supported by pH-dependent C-C crosslink formation (Figure 1.4B). The crosslink is much more efficiently formed at a pH above the pK_a of cytosine N^3 (pK_a = 6.95). DMS probing of the crosslinked $d[GCC]_n \bullet d[GCC]_n$ fragment showed that the major groove of guanine adjacent to the C-C mismatch pair is still accessible to DMS. In contrast, the known minor groove binder, Hoechst 33258, inhibited crosslink formation at the C-C mismatch. The results suggested that the C-C mismatch is crosslinked by mechlorethamine in the minor groove (Romero et al., 1999). The discovery of the C-C crosslink reaction suggested that mechlorethamine may serve as a useful probe for detection of structures formed by $d[CCG]_n$ repeat sequence containing C-C mismatch pairs (Romero et al., 1999).

In 2001, Romero et al. reported the kinetics and sequence dependence of crosslinking species using a series of model duplexes. The results of kinetic studies using gel electrophoresis (Table 2.5 and Figure 2.23) indicated that the mechlorethamine C-C crosslink forms more rapidly than the 1,3 G-G crosslink and reaches a higher final yield. The rate constant (k_t) and the extent of crosslink formation of the C-C and 1,3 G-G crosslinks are shown in Table 2.5. Stability studies demonstrated that the C-C crosslink is more stable than the 1,3 G-G crosslink (Figure 2.24). For the C-C crosslink in duplex A, the final yield of 27.5% of the total DNA attained after 2 hours is still maintained at 24 hours (Figure 2.24B). In contrast, the

maximum level of the 1,3 G-G crosslink in duplex B (10.0% of total DNA after 2 hours) decreases to only 6% after 24 hours (Figure 2.24B) (Romero et al., 2001). The sequence dependence of crosslink formation was also examined by gel electrophoresis using 19-mers with a central $d[M_4M_3M_2M_1Cn_1n_2n_3n_4]\bullet d[N_4N_3N_2N_1Cm_1m_2m_3m_4]$ sequence, where M-m and N-n were complementary base pairs. The results (Table 2.6) showed that the amount of crosslink increases with increasing G-C content of eight base pairs neighboring the C-C mismatch pair and with the proximity of the G-C pairs to the C-C mismatch (Romero et al., 2001).

Table 2.5 Kinetic parameters for mechlorethamine crosslinking of duplexes containing a C-C mismatch crosslinking site and a 1,3 G-G crosslinking sites (Romero et al., 2001)

	Duplex sequence[a]	**Crosslink**	**Rate constant (%crosslinks min^{-1})**[b]	**%Crosslink**[b,c]
A	5′-CTCTCACAG**C**CTGGTTCAG GAGAGTGTC**C**GACCAAGTC-5′	C-C	0.05 ± 0.01	27.5
B	5′-CTCTCACA**G**CCTGGTTCAG GAGAGTGTCG**G**ACCAAGTC-5′	1,3 G-G	0.02 ± 0.008	10.0

[a]The position of the crosslink is indicated in each duplex with crosslinked beses shown in bold.
[b] Average values obtained from three separate experiments.
[c] The amount of crosslinked DNA expressed as a percentage of the total DNA.

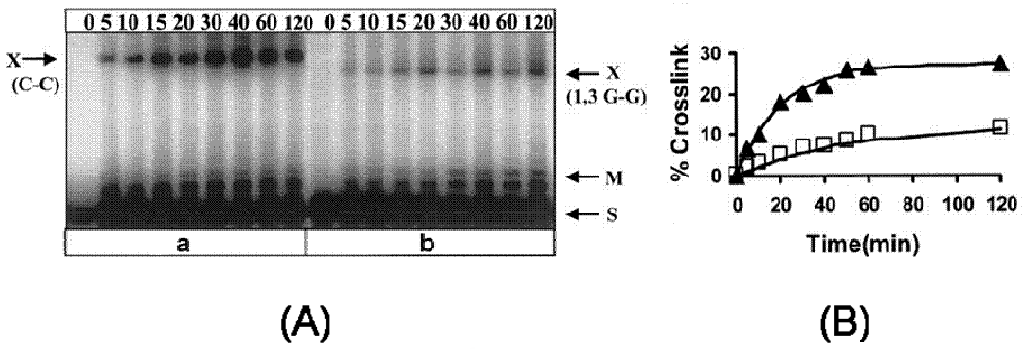


Figure 2.23 Kinetics of mechlorethamine-DNA interstrand crosslink formation for reaction times up to 2 hours. (A) Autoradiogram of 20% DPAGE gel. For each duplex, lane 0 is a control (no mechlorethamine) and all lanes show the products of incubation for a specific time (in minutes). Bands due to the crosslinks, monoadducts and unreacted single strands are identified as X, M and S, respectively. (B) The time course of total crosslink formation when ▲ is a C-C crosslink and □ is a 1,3 G-G crosslink. (Romero et al., 2001)

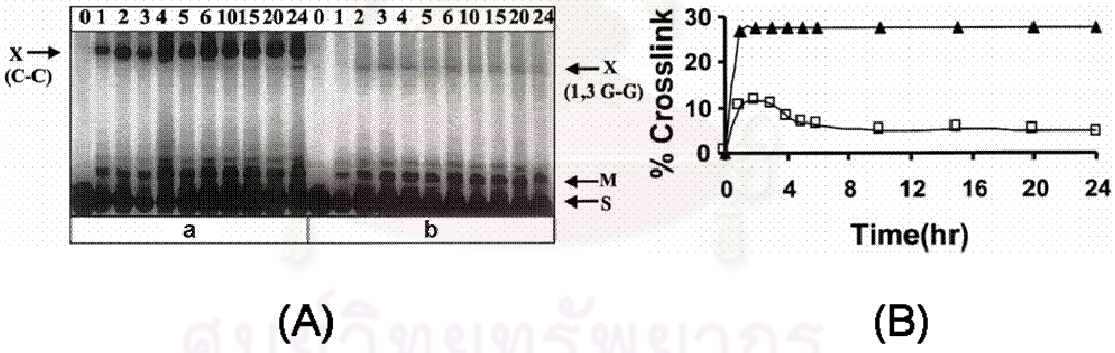


Figure 2.24 Kinetics of mechlorethamine-DNA interstrand crosslink formation for reaction times up to 24 hours. (A) Autoradiogram of 20% DPAGE gel. For each duplex, lane 0 is a control (no mechlorethamine) and all lanes show the products of incubation for a specific time (in hours). Bands due to the crosslinks, monoadducts and unreacted single strands are identified as X, M and S, respectively. (B) The time course of total crosslink formation when ▲ is a C-C crosslink and □ is a 1,3 G-G crosslink. (Romero et al., 2001)

Table 2.6 Mechlorethamine C-C mismatch crosslink formation in 19-mer duplexes

5′-CTCCCM$_4$M$_3$M$_2$M$_1$**C** n$_1$n$_2$n$_3$n$_4$CCCAG GAGGGm$_4$m$_3$m$_2$m$_1$**C** N$_1$N$_2$N$_3$N$_4$GGGTC-5′	**%Crosslink**
A A T T **C** A A T T T T A A **C** T T A A	8.7
G A T T **C** A A T C C T A A C T T A G	9.2
A G T T **C** A A C T T C A A **C** T T G A	12.5
A A C T **C** A G T T T T G A **C** T C A A	12.9
A A T C **C** G A T T T T A G **C** C T A A	16.9

This table was modified from Romero et al., 2001.

As mentioned above, C-C mismatch pairs are only weakly hydrogen-bonded (Mariappan et al., 1998), which generates flexibility of the duplex (Chen et al., 1995; Gao et al., 1995; Yu et al., 1997) and allows the cytosine bases to adopt an extrahelical location (Gao et al., 1995; Yu et al., 1997). Rojsitthisak et al. used 19-mer DNA duplexes with d[GCC]$_n$•d[GCC]$_n$ fragments containing C-C mismatches in a 1,4 base-paired relationship and demonstrated that cytosine bases of different formal mismatch pairs can be crosslinked by mechlorethamine. The gel electrophoretic analysis indicated that in the duplex containing two formal C-C mismatch pairs C_7-C_{32} and C_{10}-C_{29} (Figure 2.25), a mechlorethamine crosslink formed between C_{10} and C_{32} is detected (Figure 2.25E), in addition to C_{10}-C_{29} (Figure 2.25B) and C_7-C_{32} (Figure 2.25C) intrahelical crosslinks and double intrahelical crosslink at C_{10}-C_{29} and C_7-C_{32} (Figure 2.25D) (Rojsitthisak et al., 2001). In a duplex containing three formal C-C mismatch pairs C_7-C_{32}, C_{10}-C_{29} and C_{13}-C_{26} (Figure 2.26A), the results revealed

that two mechlorethamine crosslinks formed between C_{10} and C_{32} (Figure 2.26E) and between C_{13} and C_{29} (Figure 2.26F) in addition to intrahelical crosslinks at C_{10}-C_{29} (Figure 2.26B), C_{13}-C_{26} (Figure 2.26C) and C_7-C_{32} (Figure 2.26D) (Rojsitthisak et al., 2001). The formation of the C_{10}-C_{32} crosslink and the C_{13}-C_{29} and C_{10}-C_{32} crosslinks in the duplex containing two and three cytosine mismatch pairs, respectively, provides evidence of an extrahelical location of crosslinkable cytosines. The absence of C_7-C_{29} and C_{10}-C_{29} also suggested that the extrahelical cytosines are folded back towards the 5´-end of the duplex (Rojsitthisak et al., 2001).

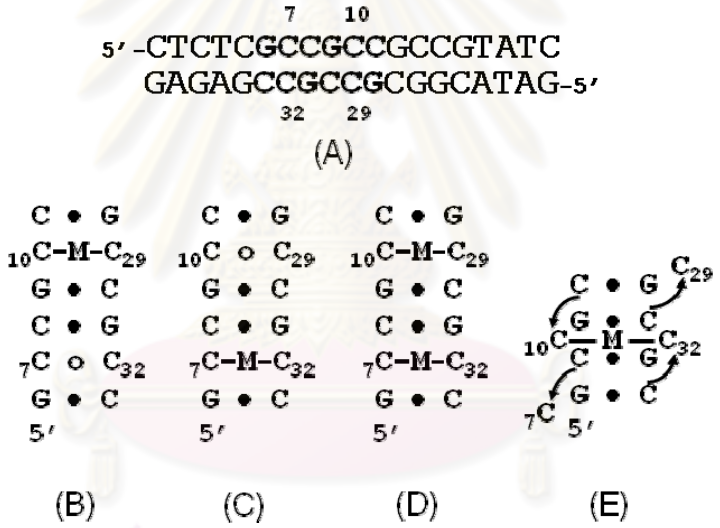


Figure 2.25 Sequence of a DNA duplex containing two C-C mismatch pairs (A). Six base pairs (bold) of the duplex are shown and -M- indicates the location of the mechlorethamine crosslink. C_{10}-C_{29} intrahelical crosslink (B), C_7-C_{32} intrahelical crosslink (C), C_{10}-C_{29} and C_7-C_{32} double intrahelical crosslink (D) and C_{10}-C_{32} extrahelical crosslink with 5´ fold-back of the extrahelical cytosine bases (E) (Modified from Rojsitthisak et al., 2001)

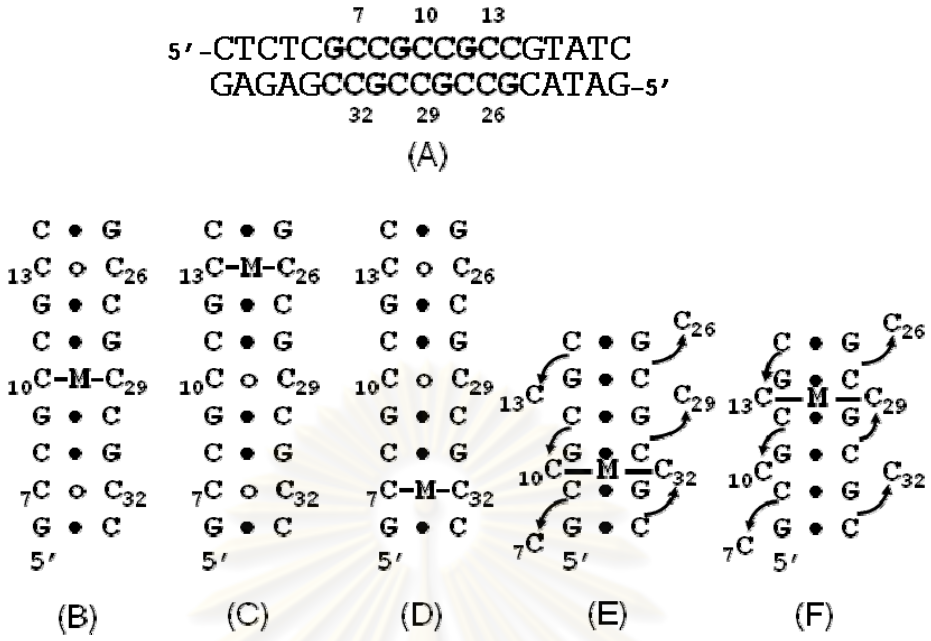


Figure 2.26 Sequence of a DNA duplex containing three C-C mismatch pairs (A). The 9 central base pairs (bold) of the duplex are shown and -M- indicates the location of mechlorethamine crosslink. C_{10}-C_{29} intrahelical crosslink (B), C_{13}-C_{26} intrahelical crosslink (C), C_7-C_{32} intrahelical crosslink (D), C_{10}-C_{32} extrahelical crosslink (E) and C_{13}-C_{29} extrahelical crosslink with 5′ fold-back of the extrahelical cytosine bases (F) (Modified from Rojsitthisak et al., 2001)

2.8 NMR determination of DNA containing mismatch base pairs

In addition to the studies of the intramolecular hairpin structure of the fragile X DNA triplet repeat sequence containing C-C and G-G mismatches and those of mechlorethamine-DNA crosslinking reaction described in sections 2.5 and 2.6, NMR techniques have been useful for characterizing the structures of DNA containing other mismatch base pairs. For DNA containing A-C mismatches, NMR studies show two possible structural arrangements of the d[CGCCGCAGC]$_2$ duplex (Figure 2.27) and two models of A-C pairing (Figure 2.28) (Sarma et al., 1987). The models of A-C

pairing are consistent with the NMR studies of a 12-mer DNA duplex containing an A-C mismatch pair (Gao et al., 1987). The NOE distance connectivity demonstrated that both A and C at the mismatch site are stacked into the right-handed helix between flanking G-C base pairs and have glycosidic torsion angles in an anti conformation (Gao et al., 1987). Additionally, the A-C pairing is independent of flanking sequence (Gao et al., 1987).

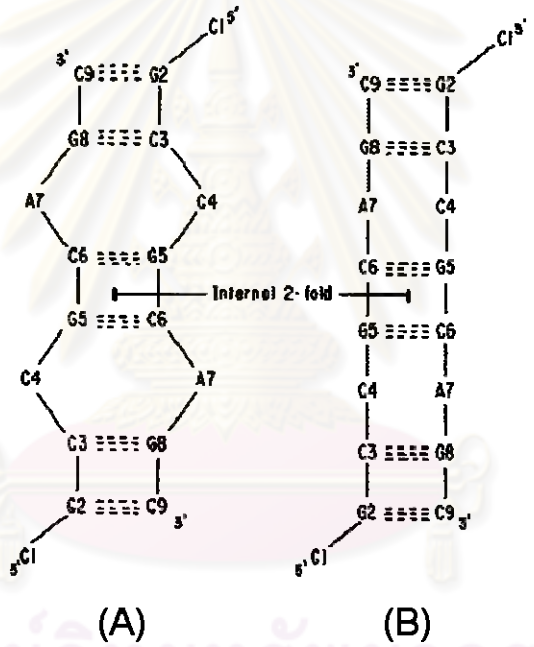


Figure 2.27 Two possible arrangements of d[CGCCGCAGC]$_2$ duplex; in one arrangement (A) C4 and A7 are an integral part of the double helix while in another (B) C4 and A7 loop out. (Sarma et al., 1987)

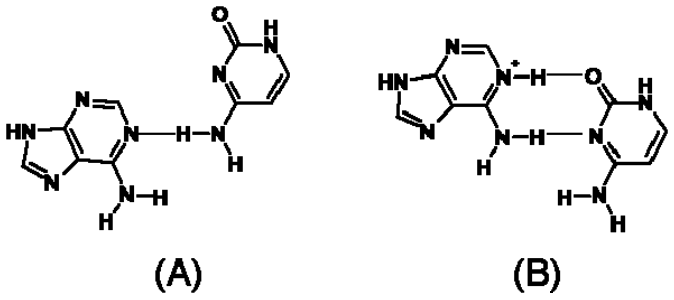


Figure 2.28 Models of A-C pairing involving a single hydrogen bond (A) and two hydrogen bonds (B) (Sarma et al., 1987)

For DNA containing G-A mismatches, several structural studies by NMR have shown that G-A mismatches are conformationally variable (Li et al., 1991). ^{1}H-NMR and ^{31}P-NMR studies of the dodecamer d[CGCAAATTGGCG]$_2$ indicated that at basic pH, mispaired bases are each in *anti* conformation and are stacked in the B-like duplex shown in Figure 2.29A (Lane et al., 1991; Li et al., 1991). When the pH is decreased, two-dimensional nuclear-overhauser-enhancement spectroscopy demonstrated that the dominant conformation is one in which the mismatched G residues are in a *syn* conformation and are hydrogen-bonded to A residues that remain in the *anti* conformation (Figure 2.29B). The residues not adjacent to G-A mismatch pair are almost unaffected by the transition or the mispairing, suggesting considerable local flexibility of the unconstrained duplexes. The conformation of the A (*anti*)-G (*anti*) duplex is very similar to the native dodecamer, whereas the AH^+ (*anti*)-G (*syn*) duplex shows a greater variation in the backbone conformation at the mismatched site. The results confirm that the A (*anti*)-G (*anti*) and AH^+ (*anti*)-G (*syn*) conformations are favored at high pH and low pH, respectively (Lane et al., 1991).

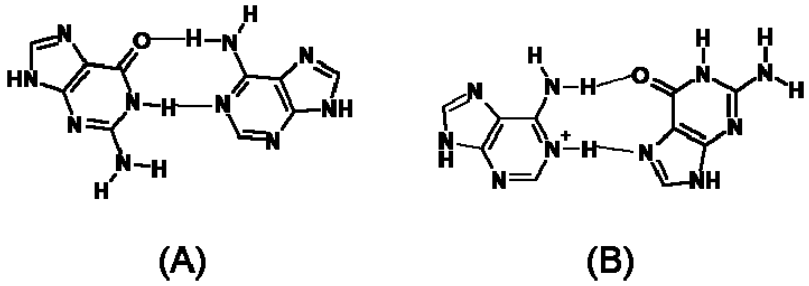


Figure 2.29 Conformations of the G-A mismatch pair: A (*anti*)-G (*anti*) (A) and AH^+ (*anti*)-G (*syn*) (B) (Lane et al., 1991)

For DNA containing G-T mismatch pairs, the three-dimensional structure of the self-complementary DNA dodecamer $d[CGTGACGTTACG]_2$ was determined using two-dimensional NMR (Allawi and SantaLucia et al., 1998). The NMR data illustrated that the G-T pair has two hydrogen bonds (Figure 2.30), with guanine projecting into the minor groove and thymine projecting into the major groove (a wobble pair). The G-T mismatch has little effect on the backbone torsion angles and helical parameters compared to standard B-DNA (Allawi and SantaLucia et al., 1998).

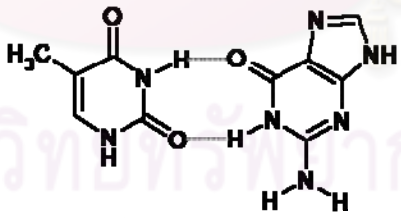


Figure 2.30 Structure of the G-T mismatch pair (Allawi and SantaLucia et al., 1998)

2.9 Use of gel electrophoresis in DNA crosslink studies

Gel electrophoresis is a sensitive method for the detection and separation of crosslinked products (Hartley et al., 1993). In addition to the studies of the

intramolecular hairpin structure of the fragile X DNA triplet repeat sequence and those of mechlorethamine-DNA crosslinking reactions described in sections 2.5-2.7, gel electrophoresis has been used to study many other examples of DNA crosslink formation and repair. A few of these applications are described below.

Electrophoresis has been used to characterize DNA crosslinking by antitumor cisplatin analogs containing enantiomeric ligands, *cis*-[PtCl2(RR-DAB)] and *cis*-[PtCl2(SS-DAB)], where DAB is 2,3-diaminobutane (Marina et al., 2000). The major differences resulting from modification of DNA by the two enantiomers were thermodynamic destabilization and conformational distortions induced by the 1,2-d(GpG) intrastrand cross-link. These differences are associated with a different biological activity of the two enantiomers. The results also indicated that the 1,2-d(GpG) intrastrand crosslink plays an important role in determining the character of the distortion induced in DNA by this lesion (Marina et al., 2000). DNA damage induced by cisplatin crosslinking of human leukocyte DNA (Hovhannisyan et al., 2004) and butadiene diepoxide-induced N^2-N^2 guanine intrastrand crosslinks have also been identified by gel electrophoresis (Carmical et al., 2000).

Gel electrophoresis has also been used to study the repair of DNA crosslinks. For example, in 2002, the repair of DNA interstrand crosslinks was shown to be a mechanism of clinical resistance to mephalan in multiple myeloma (Spanswick et al., 2002), and gel electrophoretic analysis showed that enhanced repair of DNA interstrand crosslinks by platinum compounds in ovarian tumor cells may contribute to clinically acquired resistance (Wynne et al., 2007).

2.10 Detection of DNA containing C-C and other mismatch pairs by surface plasmon resonance

Surface plasmon resonance (SPR) is an analytical technique that detects changes in the refractive index caused by variation of mass on the sensor chip surface (Figure 2.31A) when an analyte binds to an immobilized ligand on the surface. The change in the SPR signal (the SPR response, given in resonance units (RU)) is directly related to the change in surface concentration of biomolecules. The shift in resonance signal is plotted against time and displayed in a sensorgram (Figure 2.31B) (Nakatani et al., 2001). In most SPR studies, macromolecules such as proteins and DNAs are immobilized on the sensor surface to detect protein-protein and protein-DNA interactions. SPR can detect mismatched base pairs when a low molecular weight ligand that specifically binds to the mismatch is immobilized (Kobori et al., 2004).

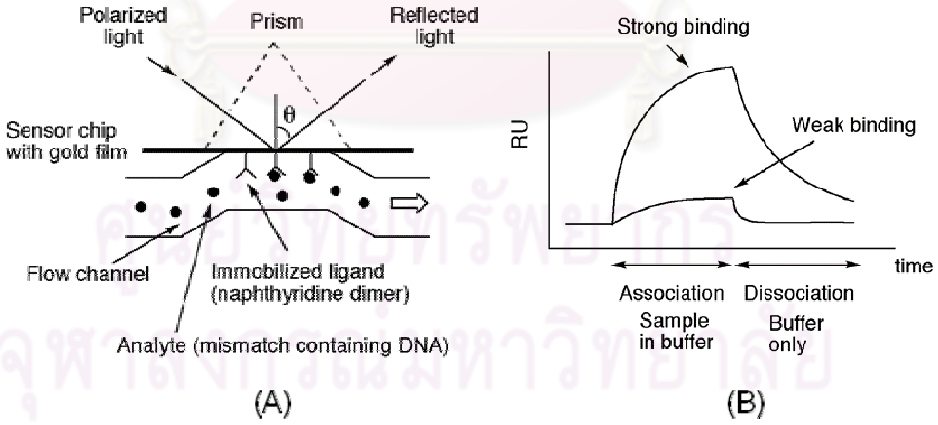


Figure 2.31 Illustration of the SPR assay using a sensor chip with a covalently immobilized ligand on the surface. (A) The angle for reflection of polarized light is changed by binding of DNA to the sensor surface. (B) The change of angle is computed to the change of response unit (RU) and plotted against time (Nakatani et al., 2001).

In 2004, Kobori et al. synthesized the first SPR sensor for C-C mismatch pair detection in DNA duplex by immobilizing an aminonaphthyridine dimer on the gold surface (Figure 2.32). The ligand consists of two 2-aminonaphthyridine chromophore and a strong connecting alkyl linker that stabilizes the C-C mismatch regardless of flanking sequences. Fully matched duplexes are not stabilized at all. Other mismatches such as C-T, C-A and T-T pairs are stabilized with reduced efficiency (Figure 2.33). The response for the C-C mismatch in a 27-mer DNA duplex containing $d[GCC]_n \bullet d[GCC]_n$ is approximately 83 times stronger than that for the fully matched duplex (Figure 2.33) and the sensor detects the C-C mismatch pair at 10 nM. The facile protonation of 2-aminonaphthyridine at pH 7 produces a hydrogen-bonding surface complementary to that of cytosine (Figure 2.34) and probably causes the high selectivity of the aminonaphthyridine dimer for the C-C mismatch (Kobori et al., 2004).

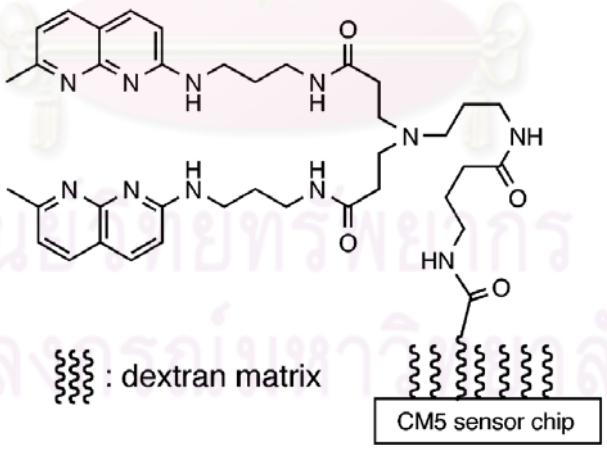


Figure 2.32 The aminonaphthyridine dimer immobilized sensor (Kobori et al., 2004)

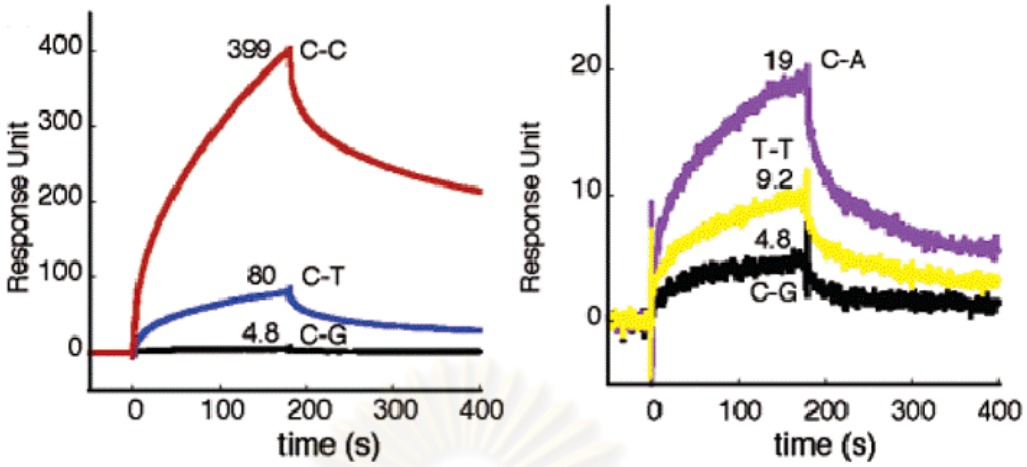


Figure 2.33 SPR assay of a 27-mer duplex containing C-C, C-T, C-A, and T-T mismatch pairs and a C-G base pair with the aminonaphthyridine dimer-immobilized sensor surface. The measurement included binding for 180 seconds and dissociation for 220 seconds (Kobori et al, 2004)

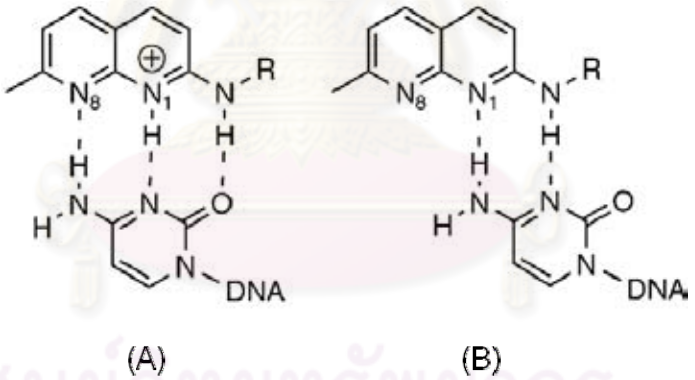


Figure 2.34 Proposed models of base paring of cytosine with protonated (A) and unprotonated (B) aminonaphthyridine (Kobori et al., 2004)

In addition to the detection of a DNA duplex containing C-C mismatch pairs (Kobori et al., 2004), an SPR sensor chip was also developed to detect other types of DNA mismatches. Nakatani et al. synthesized SPR sensor chip that specifically binds to the G-G mismatch by immobilizing a naphthyridine dimer on the surface (Figure 2.35). The SPR assay showed a significantly higher SPR response for the 27-mer

DNA containing a G-G mismatch. In contrast, DNAs containing G-A and G-T mismatch pairs as well as a fully matched duplex produced only a weak response (Nakatani et al., 2001). A kinetic analysis revealed that the binding of the G-G mismatch is dependent on the flanking base pairs, with G-G mismatches with at least one flanking G-C pair binding to the sensor surface via a two-step process with a 1:1 DNA-ligand stoichiometry (Nakatani et al., 2004).

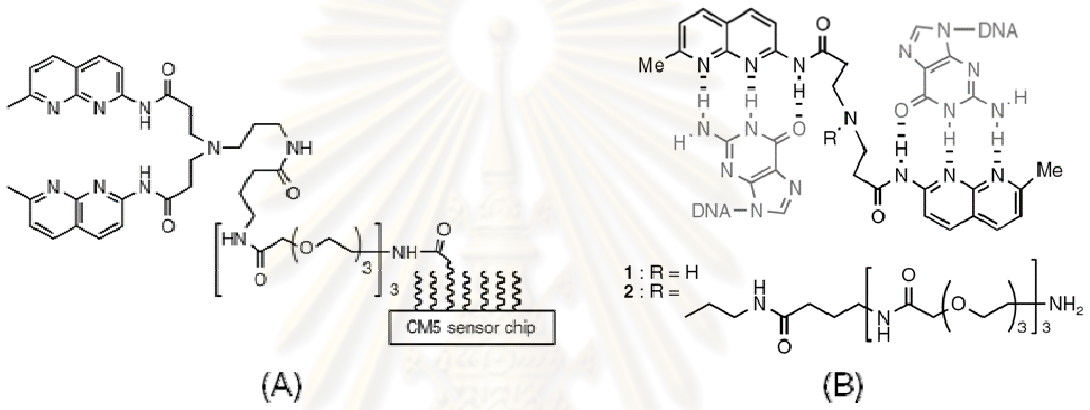


Figure 2.35 Structures of the naphthyridine dimer immobilized sensor (A) and naphthyridine dimer hydrogen bonding to guanine (B) (Modified from Nakatani et al., 2001, 2004)

An SPR assay also showed the ability of f-ImImIm, an imidazole-containing polyamide trimer (Figure 2.36), to discriminate T-G mismatch pairs from Watson-Crick and other mismatch base pairs (Lacy et al., 2002). Kinetic and thermodynamic studies showed that f-ImImIm binds significantly more strongly to T-G mismatch-containing oligonucleotides than to sequences with other mismatch pairs or with Watson-Crick base pairs. Compared with the Watson-Crick CCGG sequence, f-ImImIm associates more slowly with DNA containing T-G mismatches in place of one or two C-G base pairs and, more importantly, the dissociation rate (k_d) from the T-G oligonucleotides is very slow (Lacy et al., 2004). The results clearly showed the

binding selectivity and enhanced affinity of side-by-side imidazole/imidazole pairing for T-G mismatches and showed that the increased affinity and specificity arises from much lower k_d values with the T-G duplexes (Lacy et al., 2002).

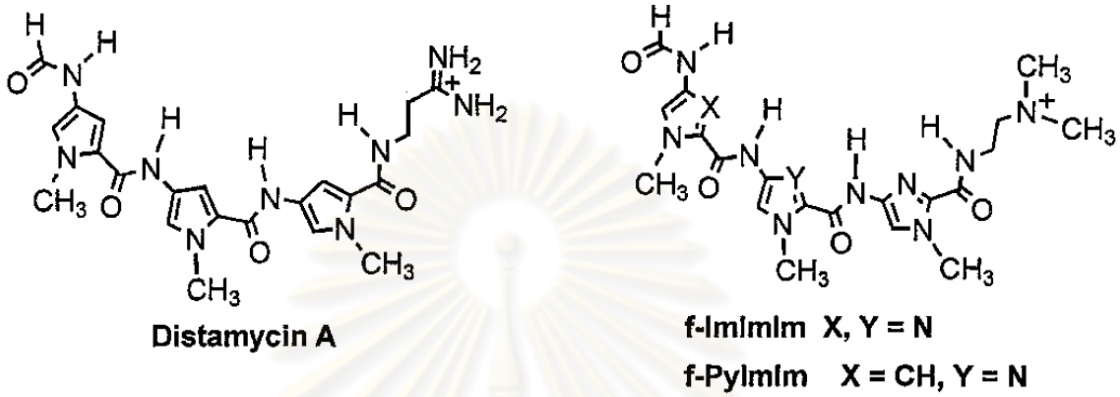


Figure 2.36 Chemical structures of an imidazole-containing polyamide trimer. Py = Pyrole; Im = Imidazole; f = formamido (NHCOH) group (Lacy et al., 2002)

In 2004, Hagihara et al. discovered a new naphthyridine-azaquinolone (Npt-Azq) hybrid that strongly stabilized the G-A mismatch pair in DNA duplex (Hagihara et al., 2004). Npt-Azq was synthesized and immobilized on the surface of a sensor chip (Figure 2.37) for SPR assay to examine SPR detection of duplexes containing a G-A mismatch. The distinct SPR response was observed when 27-mer DNA containing a G-A mismatch pair was analyzed using an Npt-Azq immobilized sensor surface, whereas the SPR response of the fully matched duplex was approximately 6-fold weaker (Figure 2.38). The SPR signals for the G-A mismatch are proportional to the concentration of DNA in a range up to 1 μM, confirming that the SPR signal is in fact due to the binding of the G-A mismatch to Npt-Azq immobilized on the sensor surface (Hagihara et al., 2004). Npt-Azq was proposed to form a complementary hydrogen-bonding surface for guanine and adenine (Figure 2.39) (Hagihara et al., 2004).

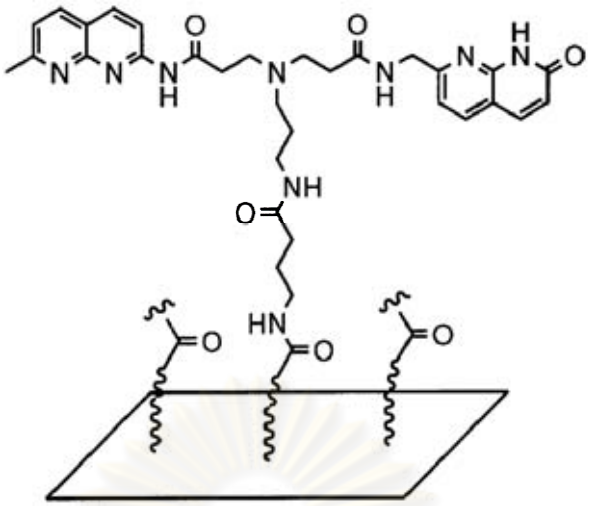


Figure 2.37 The Npt-Azq immobilized sensor (Hagihara et al., 2004)

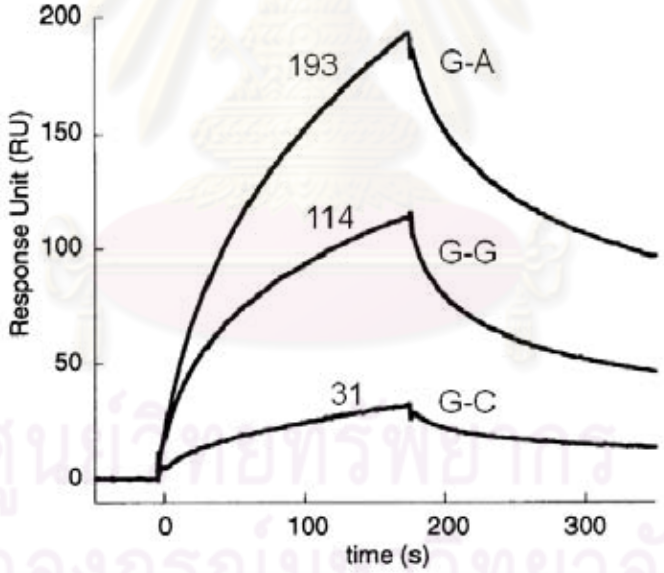


Figure 2.38 SPR assay of a 27-mer duplex containing G-A and G-G mismatch pairs and a G-C base pair with the Npt-Azq immobilized sensor surface. The measurement included binding for 180 seconds and dissociation for 220 seconds (Modified from Hagihara et al, 2004)

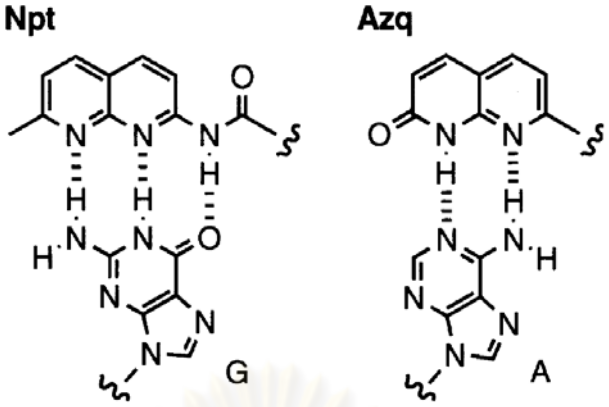


Figure 2.39 Hydrogen-bonding patterns of Npt-G and Azq-A (Hagihara et al., 2004)

2.11 Determination of crosslinked DNA at mismatch base pairs and DNA bearing modifications by liquid chromatography and mass spectrometry

High performance liquid chromatography (HPLC) and mass spectrometry (MS) have been applied widely for the determination of mismatch base pairs and crosslinking sites in DNA, as well as for several DNA adducts (Balcome et al., 2004; Dorr et al., 2007; Gaskell et al., 2007; Gupta et al., 2005; Park et al., 2005; Singh and Farmer, 2006; Tretyakova et al., 2007; Wilds et al., 2004; Winds et al., 2006). Advances in technologies now allow the direct coupling of HPLC to MS (Singh and Farmer, 2006). In addition, the introduction of matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) have revolutionalized the application of MS for characterization of biomolecules including modified oligonucleotides. Enzymatic digestion followed by MS analysis has been used for characterization of oligonucleotides bearing a number of modifications (Wang and Wang, 2003).

MALDI-TOF-MS is one of most versatile tools in the post-genomic era for analysis of biomolecules (Tost and Gut, 2006). Initially, MALDI-TOF-MS was predominantly used for analysis of proteins and peptides. However, advances in the application of MALDI-TOF-MS for more demanding DNA analysis, such as molecular haplotyping, DNA methylation analysis, expression profiling and mutation detection, have demonstrated its potential for versatile analysis of nucleic acids, besides simple SNP genotyping (Gut, 2004; Tost and Gut, 2006). MALDI-TOF-MS is suitable for analysis of complex mixtures, and the current generation of instruments is capable of recording a single spectrum in less than 1 second (Gut, 2004). Although DNA is a monotonous molecule with different sequences displaying little or no differences in desorption efficiency with current matrix systems, in contrast to proteins and peptides, quantitative analysis of DNA is possible with MALDI-TOF-MS (Gut, 2004).

ESI is especially useful for analyzing large biomolecules such as proteins, peptides and oligonucleotides and can also be used for smaller molecules. ESI is performed at atmospheric pressure and results in negligible dissociation of the molecular ion into fragment ions; it is a soft ionization technique, in which ions are protonated or deprotonated by application of a potential. DNA adduct analysis is commonly performed using a triple quadrupole instrument, and less often with an ion-trap mass spectrometer (Singh and Farmer; 2006).

For the C-C mismatch pair, Noll et al. synthesized and characterized DNA duplexes containing an N^4dC-Ethyl-N^4dC interstrand crosslink (Figure 2.40). The crosslinked duplexes were purified by strong anion-exchange HPLC and then characterized by MALDI-TOF-MS (Table 2.7). The purified crosslinked duplexes were digested by snake venom phosphodiesterase (SVPD) and calf intestinal phosphatase (CIP). The digested products were subsequently analyzed by reversed-phase HPLC and the N^4dC-Ethyl-N^4dC peak was obtained from HPLC analysis, as shown in Figure 2.41 (Noll et al., 2001).

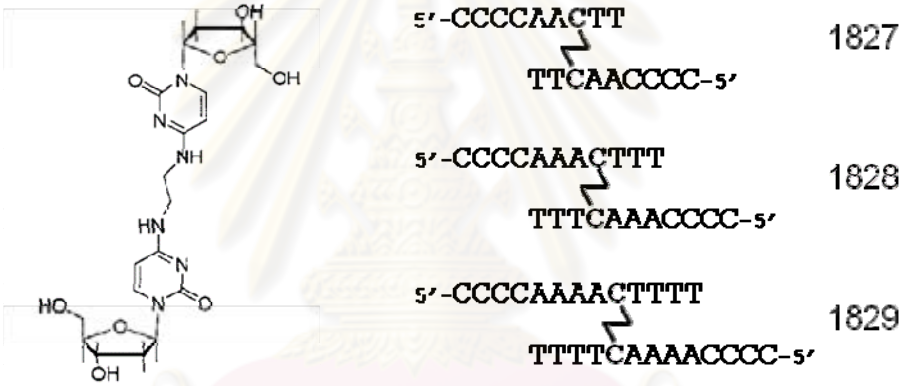


Figure 2.40 Structure of the N^4dC-Ethyl-N^4dC interstrand crosslink and sequences of the C-C crosslinked duplexes (Modified from Noll et al., 2001)

Table 2.7 Mass spectral data for crosslinked duplexes (Noll et al., 2001)

Crosslinked duplexes	Mass	
	Expected	Found
1827	5264	5269
1828	6498	6487
1829	7733	7735

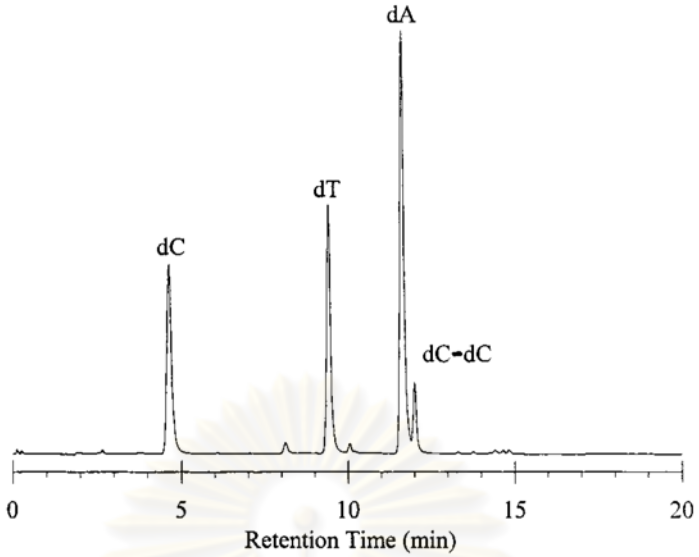
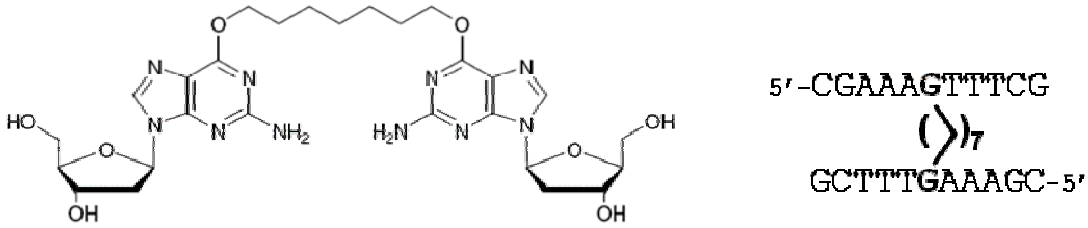


Figure 2.41 Reversed-phase HPLC chromatogram of crosslinked duplex 1829 after digestion with SVPD and CIP (Noll et al., 2001)

In addition to the detection of the crosslinked DNA duplex at a C-C mismatch pair (Noll et al., 2001), liquid chromatographic and mass spectrometric techniques have been applied to characterize the O^6dG-Heptyl-O^6dG interstrand crosslink shown in Figure 2.42. A DNA duplex containing the crosslink between two O^6 atoms of deoxyguanosine were synthesized and subsequently purified by strong anion-exchange HPLC. The purified crosslinked duplex was characterized by MALDI-TOF-MS, which indicated a molecular weight of 6806.9 (expected 6805.4). The purified crosslinked duplex was then digested by SVPD and CIP. The digest products were analyzed by HPLC and O^6dG-Heptyl-O^6dG was obtained from the HPLC analysis, as shown in Figure 2.43 (Winds et al., 2006).



(A) (B)

Figure 2.42 Structure of the O^6dG-Heptyl-O^6dG interstrand crosslink (A) and sequence of the C-C crosslinked duplex (B) (Modified from Winds et al., 2006)

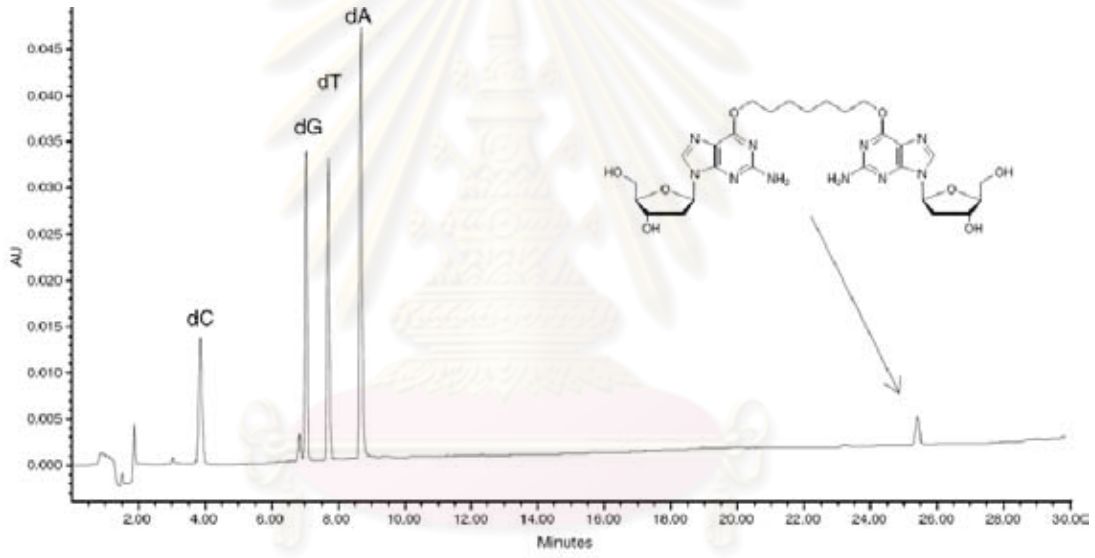


Figure 2.43 Reversed-phase HPLC chromatogram of the crosslinked duplex after digestion with SVPD and CIP (Winds et al., 2006)

In addition to the characterization of crosslinked DNA duplexes at C-C and G-G mismatch pairs, liquid chromatography and mass spectrometry have been used to analyze other types of crossklinks and DNA adducts. For example, Balcome et al. characterized the G-A and A-A interstrand crosslinks and G-A intrastrand crosslink formed by mechlorethamine and DNA duplexes using HPLC-ESI-MS/MS (Balcome et al., 2004), and Gupta et al. identified bifunctional GA and AG intrastrand

crosslinks formed between cisplatin and DNA using HPLC and ESI-MS (Gupta et al., 2005). Most recently, Dorr et al. synthesized DNA containing N6-(2-hydroxy-3-buten-1-yl)-adenine adducts of 3,4-epoxy-1-butene and characterized these adducts by HPLC-ESI-MS/MS (Dorr et al., 2007).

CHAPTER III

EXPERIMENTAL

Chemicals

1. Synthetic 15-mer oligodeoxyribonucleotides (Sigma-Proligo, USA). Sequences of the oligodeoxyribonucleotides are as follow:
   1.1 d[CTC ACA CCG TGG TTC] (referred to as top-strand DNA)
   1.2 d[GAA CCA CCG TGT GAG] (referred to as bottom-strand DNA)
   1.3 d[AAA AAA AAA AAA AAA] (polydeoxyadenylate (poly-dA))
   1.4 d[TTT TTT TTT TTT TTT] (polydeoxythymidylate (poly-dT))
   1.5 d[CCC CCC CCC CCC CCC] (for polydeoxycytidylate (poly-dC))
   1.6 d[GGG GGG GGG GGG GGG] (polydeoxyguanylate (poly-dG))
2. Monodeoxyribonucleosides (Sigma-Aldrich, USA): 2´-deoxyadenosine monohydrate (dA; MW 269.26) 2´-deoxythymidine (dT; MW 242.23), 2´-deoxycytidine (dC; MW 227.22) and 2´-deoxyguanosine (dG; MW 285.26)
3. Mechlorethamine (2-chloro-N-[2-chloroethyl]-N-methylethanamine) hydrochloride (MW 192.5) (Sigma Aldrich, USA)
4. Snake venom phosphodiesterase (SVPD) from *Crotalus adamanteus* (Sigma Aldrich, USA)
5. Calf intestinal phosphatase (CIP) (Finnzyme, Finland)
6. Sodium chloride (MW 58.44) (Merck, Germany)
7. Tris(hydroxymethyl)-aminomethane (Tris base, MW 121.14) (Merck, Germany)
8. Acetonitrile, HPLC grade (ACN, MW 41.05) (Honeywell Burdick & Jackson, USA)

9. Dimethylsulfoxide (DMSO, MW 78.13) (Sigma Aldrich, USA)
10. Triethylamine (TEA, MW 101.19) (Sigma Aldrich, USA)
11. Glacial acetic acid (MW 60.05) (Lab-Scan, Thailand)
12. 3-hydroxypicolinic acid (MW 139.11) (Fluka, Switzerland)
13. Ethylenediaminetetraacetic acid (EDTA, MW 292.24) (Merck, Germany)
14. Monobasic sodium phosphate (NaH_2PO_4, MW 141.96) (Merck, Germany)
15. Dibasic sodium phosphate dihydrate (MW 156.03) (Merck, Germany)
16. 85% phosphoric acid (MW 98.00) (Merck, Germany)
17. Magnesium chloride hexahydrate (MW 203.30) (Merck, Germany)

DNA solutions contain approximately 5 OD of oligonucleotides. The OD units are obtained by spectrophotometric measurement at 260 nm. One OD corresponds to the amount of oligonucleotide in a 1 ml volume that results in an optical density of 1 in a 1 cm path-length cuvette. This corresponds approximately to 33 µg or 5 nmols of oligonucleotide for a 15mer. The properties of the synthetic oligodeoxyribonucleotide solutions from Sigma-Proligo are shown in Table 3.1

Table 3.1 Properties of synthetic DNA

Name	**Sequence(5´-3´)**	**MW (g/mole)**	**Qty/tube**		**Conc. (µM)**	**T_m* (^oC)**
			OD	**nmole**		
Top strand	5´-CTC ACA CCG TGG TTC-3´	4505	5.0	37.7	205	48
Bottom strand	5´-GAA CCA CCG TGT GAG-3´	4603	5.0	33.7	215	48
Poly-dA	5´-AAA AAA AAA AAA AAA-3´	4639	5.0	27.3	190	30
Poly-dT	5´-TTT TTT TTT TTT TTT-3´	4501	5.0	41.0	273	30
Poly-dC	5´-CCC CCC CCC CCC CCC-3´	4276	5.2	47.7	2076	60
Poly-dG	5´-GGG GGG GGG GGG GGG-3´	4876	5.1	33.4	927	60

*T_m (Melting Temperature) is defined as the temperature at which 50% of the DNA molecules form a stable double helix and the other 50% have been separated into single strands

Instruments

1. HPLC was performed using a Shimadzu-VP system (Shimadzu, Japan) consisting of the following modules:
   1.1 SCL-10A VP system controller
   1.2 LC-10AD VP pumps
   1.3 SPD-10AD VP UV-VIS detector
   1.4 SIL-10AD VP auto-injector
   1.5 CTO-10A VP column oven
   1.6 FRC-10A fraction collector
   1.7 DGU-14A degasser
   1.8 Shimadzu CLASS-VP software
2. HPLC columns
   2.1 Biobasic-C4 column (4.6 x 250 mm, 5 µm) (Thermo Electron, UK)
   2.2 Rainin Microsorb-C18 column (4.6 x 150 mm, 5 µm) (Varian, USA)
3. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometer (Autoflex II, Bruker, USA) with FlexAnalysis software
4. Triple quadrupole mass spectrometer (API 4000, Applied Biosystem, USA) with Analyst software version 4.4.2
5. Microcon® YM-3 (centrifugal filter device for concentration and purification of biological samples) with 3000 molecular weight cut-off (Millipore, USA)
6. Centrifugal instrument (Marathon MicroA, Fischer Scientific, UK)
7. Vortex mixer (Baxter, USA)
8. Hot plate (Heidolph MR2002, Germany)
9. Heating block (AccuBlock, Labnet, USA)

Methods

3.1 Determination of the structure and amount of crosslink formed by mechlorethamine with a DNA duplex containing a C-C mismatch pair

3.1.1 Preparation of 2 M triethylammonium acetate (TEAA) solution:

200 ml of 2 M TEAA solution was prepared by adding 55 ml of triethylamine (TEA) and 24 ml of glacial acetic acid into a 200-ml volumentric flask. The solution was diluted to volume with distilled water.

3.1.2 Preparation of Buffer A pH 7.0:

Buffer A was prepared by dissolving 75 mg of EDTA and 100 ml of 2 M TEAA with distilled water to obtain 2000 ml of Buffer A. The solution was adjusted to pH 7.0.

3.1.3 Preparation of 1.0 M sodium chloride solution:

1.0 M sodium chloride solution was prepared by dissolving 2.93 g of sodium chloride with distilled water to obtain 50 ml of 1.0 M sodium chloride solution.

3.1.4 Preparation of 0.5 M Tris buffer solution pH 7.5:

50 ml of 0.5 M Tris buffer solution was prepared by dissolving approximately 3.03 g of Tris base with distilled water. The solution was adjusted to pH 7.5.

3.1.5 Preparation of DNA stock solutions:

DNA stock solutions were prepared by diluting each synthetic DNA solution (top strand and bottom strand) with distilled water. Distilled water was added into the tube of top-strand or bottom-strand DNA solution to obtain single-strand DNA stock solutions at final concentrations of 100 µM.

3.1.6 Preparation of single-strand DNA sample solutions:

Each single-strand DNA sample solution (10 µM, 100 µl) containing 0.1 M sodium chloride and 0.05 M Tris buffer was prepared by diluting 10 µl of DNA stock solution with 70 µl of distilled water. The solution was combined with 10 µl of 1.0 M sodium chloride stock solution and 10 µl of 0.5 M Tris buffer stock solution, and mixed, centrifuged and kept at 4°C before use. The solution (20 µl) was analyzed using reversed-phase HPLC.

3.1.7 Preparation of a mixture containing top-strand and bottom-strand DNAs:

The solution was prepared by mixing an equal amount of top-strand and bottom-strand DNA in 0.1 M sodium chloride and 0.05 M Tris buffer (pH 7.5). In this experiment, a 100-µl solution containing 10 µM top-strand DNA and 10 µM bottom-strand DNA was prepared by mixing 10 µl of top-strand DNA stock solution and 10 µl of bottom-strand DNA stock solution. The mixture was combined with 60 µl of distilled water, 10 µl of 1.0 M sodium chloride stock solution, and 10 µl of 0.5 M Tris buffer stock solution to obtain the final concentrations (10 µM) of top-strand and bottom-strand DNA. The solution was

mixed, centrifuged, and kept at 4°C before use. The solution (20 µl) was analyzed using reversed-phase HPLC.

3.1.8 Preparation of DNA duplex (DNA annealing process):

The scheme of the DNA annealing process is shown in Figure 3.1. DNA duplex was prepared by mixing an equal amount of complementary oligonucleotides in 0.1 M sodium chloride and 0.05 M Tris buffer (pH 7.5). In this experiment, 100 µl of 10 µM DNA duplex solution was prepared by mixing 10 µl of top-strand DNA stock solution and 10 µl of bottom-strand DNA stock solution. The mixture was combined with 60 µl of distilled water, 10 µl of 1.0 M sodium chloride stock solution, and 10 µl of 0.5 M Tris buffer stock solution to obtain a final concentration of DNA duplex solution of 10 µM. The solution was mixed, centrifuged, heated at 90°C for 2 minutes, and then slowly cooled to room temperature to allow annealing of DNA. The solution was kept in refrigerator (4°C) overnight to stabilize the duplex before use.

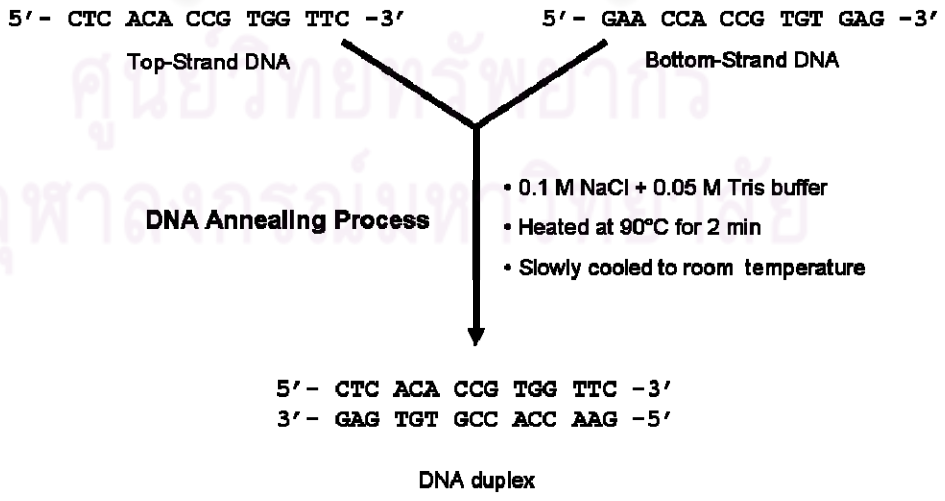


Figure 3.1 Scheme showing the DNA annealing process.

3.1.9 Preparation of the mechlorethamine-crosslinked DNA duplex:

The scheme of the mechlorethamine-DNA crosslinking reaction is shown in Figure 3.2. Following annealing of the strands, 100 µl of DNA duplex solution (10 µM) was incubated with 1 µl of 0.1 M mechlorethamine solution in DMSO for an hour at room temperature. The final concentration of mechlorethamine in the solution was about 100 times higher than that of the DNA duplex. The mechlorethamine solution in DMSO was freshly prepared by dissolving approximately 2 mg of mechlorethamine in 100 µl of DMSO, and was immediately used. After incubation, 20 µl of the mixture was analyzed by reversed-phase HPLC.

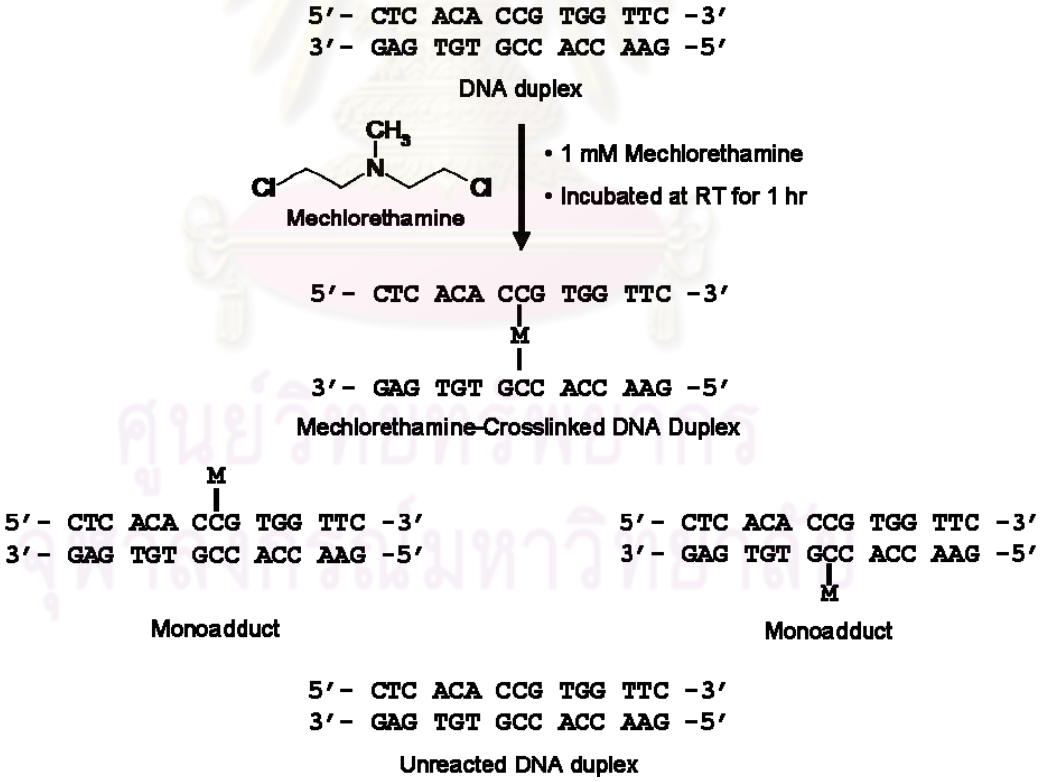


Figure 3.2 Scheme showing the mechlorethamine-DNA crosslinking reaction.

3.1.10 Determination and purification of the mechlorethamine-crosslinked DNA duplex by HPLC

Top-strand DNA, bottom-strand DNA, a mixture of the two, DNA duplex and mechlorethamine-crosslinked DNA duplex were analyzed by HPLC. Chromatography was performed using a Biobasic-C4 column (4.6x250mm, 5µm) at 33°C with gradient elution of 5-15% ACN in Buffer A pH 7.0 containing 100 mM TEAA and 0.1 mM EDTA in 60 minutes at a flow rate of 1.0 ml/min. UV-detection was set at 260 nm and the injection volume was 20 µl. The HPLC condition and the plot of the gradient elution of the mobile phase are shown in Figure 3.4.

Column	: Biobasic-C4 column (4.6x250 mm, 5 µm)
Mobile phase	: 5-15% ACN in Buffer A pH 7.0 containing 100 mM TEAA and 0.1 mM EDTA in 60 minutes
Flow rate	: 1 ml/min
Oven temperature	: 33°C
Detector	: UV 260 nm
Injection volume	: 20 µl

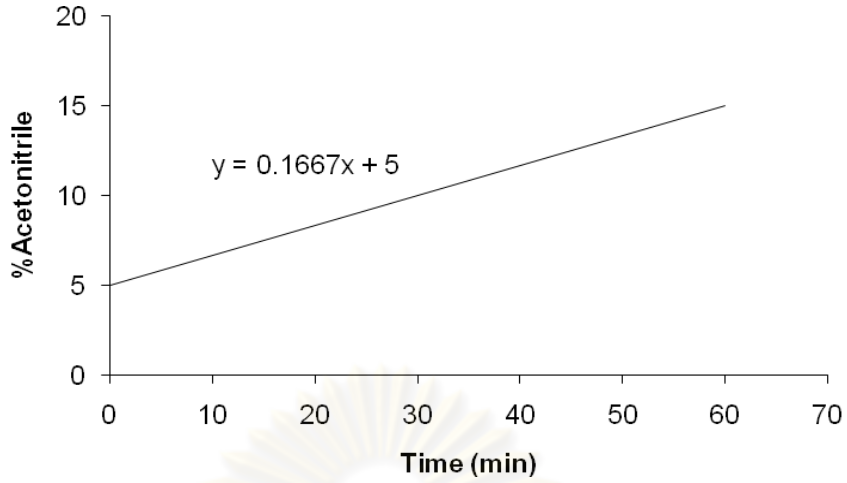


Figure 3.3 Gradient elution of the mobile phase; 5-15% ACN in 60 minutes

The mechlorethamine-crosslinked DNA duplex was purified from the reaction solution by collecting the eluent at the retention time of the crosslink according to the HPLC chromatogram (Figure 3.4).

The amount of crosslinked DNA duplex formed by mechlorethamine was calculated from the peak areas of the crosslinked DNA duplex and the unreacted top-strand and bottom-strand DNA in HPLC analysis. The amount of crosslinked DNA duplex was expressed as % crosslink using the following equation.

$$\% \text{ crosslink} = \frac{(\text{crosslink}) \times 100}{[(\text{unreacted top}) + (\text{unreacted bottom}) + (\text{crosslink})]}$$

where (crosslink), (unreacted top), and (unreacted bottom) represent the peak areas of the crosslinked DNA, unreacted top-strand DNA and unreacted bottom-strand DNA, respectively.

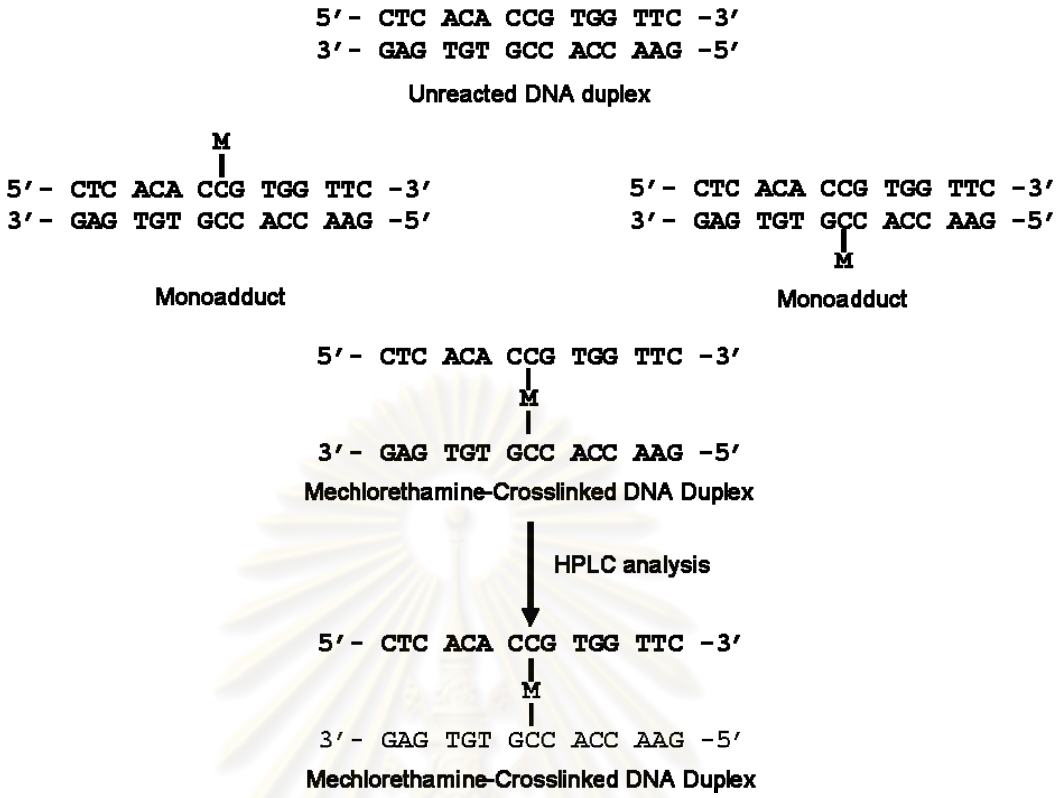


Figure 3.4 Scheme showing the purification and quantification of the mechlorethamine-crosslinked DNA duplex by HPLC.

3.1.11 Characterization of the mechlorethamine-crosslinked DNA duplex by MALDI-TOF-MS

After the purified mechlorethamine-crosslinked DNA duplex was collected from HPLC analysis (section 3.1.10), the purified crosslinked DNA duplex solution was desalted by centrifugal filtration through Microcon® YM-3 at 10,000 rpm for 5 min. Subsequently, the crosslink was prepared for MALDI-TOF-MS analysis by removing solvent using lyophilization and mixing the crosslinked DNA duplex with matrix substances such as 3-hydroxypicolinic acid. The mixture of crosslinked DNA and matrix substance was applied to the sample plate and analyzed by MALDI-TOF-MS. The scheme of MALDI-TOF-MS

determination of the mechlorethamine-crosslinked DNA duplex is shown in Figure 3.5.

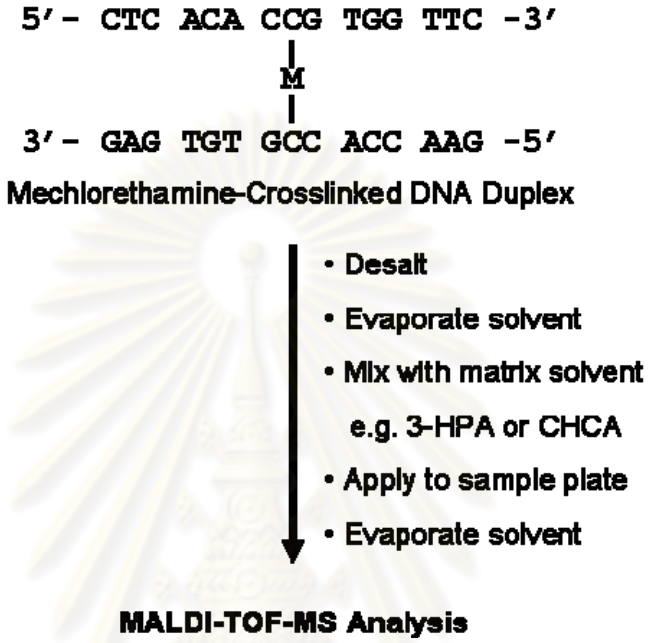


Figure 3.5 Scheme showing MALDI-TOF-MS determination of the mechlorethamine-crosslinked DNA duplex.

3.2 Determination of the target of mechlorethamine in a DNA duplex containing a C-C mismatch pair

3.2.1 Preparation of 0.2 M monobasic sodium phosphate solution:

250 ml of 0.2 M monobasic sodium phosphate solution was prepared by dissolving 7.80 g of monobasic sodium phosphate dihydrate in distilled water.

3.2.1 Preparation of 0.2 M dibasic sodium phosphate solution:

50 ml of 0.2 M dibasic sodium phosphate solution was prepared by dissolving 1.42 g of dibasic sodium phosphate in distilled water.

3.2.3 Preparation of phosphate buffer (pH 5.8):

A liter of phosphate buffer (pH 5.8) was prepared by mixing 20 ml of 0.2 M monobasic sodium phosphate solution and 230 ml of 0.2 M dibasic sodium phosphate solution in distilled water. The solution was adjusted to pH 5.8 using 85% phosphoric acid.

3.2.4 Preparation of 50 mM Tris buffer solution pH 8.1:

100 ml of 50 mM Tris buffer solution was prepared by dissolving approximately 606 mg of Tris base in distilled water. The solution was adjusted to pH 8.1.

3.2.5 Preparation of 0.01 M magnesium chloride solution:

100 ml of 0.01 M magnesium chloride solution was prepared by dissolving 203 mg of magnesium chloride hexahydrate in distilled water.

3.2.6 Preparation of monodeoxyribonucleoside stock standard solutions:

Stock standard solutions of dT, dA, dC and dG were prepared by diluting each monodeoxyribonucleoside standard in distilled water to obtain a final concentration of each monodeoxyribonucleoside of 10 mM.

3.2.7 Preparation of monodeoxyribonucleoside working standard solutions:

Working standard solutions (100 µM) of dT, dA, dC and dG were prepared by diluting each monodeoxyribonucleoside stock standard solution with distilled water.

3.2.8 Preparation of polydeoxyribonucleoside solutions:

Sample solutions (100 µM) of poly-dA, poly-dT, poly-dC and poly-dG were prepared by dissolving each polydeoxyribonucleoside in distilled water.

3.2.9 Preparation of DNA duplex (DNA annealing):

Top-strand DNA and bottom-strand DNA were annealed using the process described in section 3.1.8 to obtain DNA duplex.

3.2.10 Preparation of the mechlorethamine-crosslinked DNA duplex:

DNA duplex crosslinked by mechlorethamine was prepared using the crosslinking reaction described in section 3.1.9. The mechlorethamine-crosslinked DNA duplex was separated and purified by HPLC using the conditions described in section 3.1.10. The crosslinked DNA collected from HPLC was desalted by centrifugal filtering through Microcon® YM-3 and solvent was completely removed by lyophilization. The mechlorethamine-crosslinked DNA duplex was subsequently dissolved in water to obtain a final concentration of mechlorethamine-crosslinked DNA duplex solution of 10 µM. This solution was then used for enzymatic digestion.

3.2.11 Enzymatic digestion:

Single-strand DNAs (top strand and bottom strand), DNA duplex, polydeoxyribonucleotides and the mechlorethamine-crosslinked DNA duplex were subjected to enzymatic digestion with a combination of SVPD and CIP. The scheme of enzymatic digestion of the mechlorethamine-crosslinked DNA duplex is shown in Figure 3.6 and 3.7. In this experiment, 6.6 µl of 10 µM top-strand DNA, 10 µM bottom-strand DNA, 10 µM DNA duplex, 100 µM polydeoxyribonucleotides or 10 µM mechorethamine-crosslinked DNA duplex was added with 0.67 µl of 10 mM magnesium chloride solution and 0.67 µl of 50 mM Tris buffer pH 8.1, and then treated with 10 µl of SVPD (0.4 units) and 2 µl of CIP (20 units). Each reaction mixture was adjusted to 100 µl with distilled water. The reaction mixtures were then mixed, centrifuged and incubated at 37°C for 48 hours. After incubation, the enzymes and salts were removed by centrifugal filtering of the reaction mixtures through Microcon® YM-3 and each mixture was analyzed by reversed-phase HPLC and ESI-MS/MS.

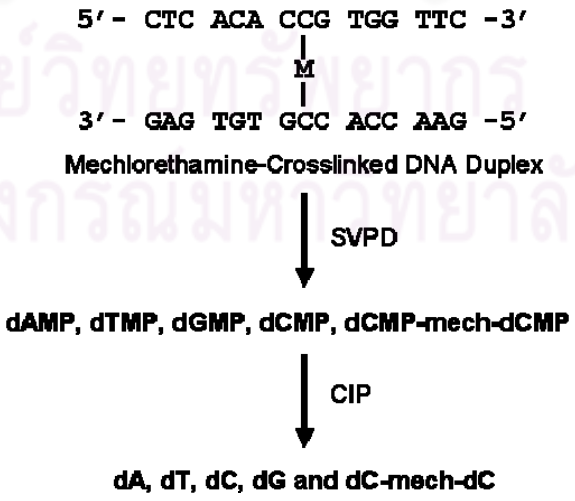


Figure 3.6 Scheme showing enzymatic digestion of the mechlorethamine-crosslinked DNA duplex.

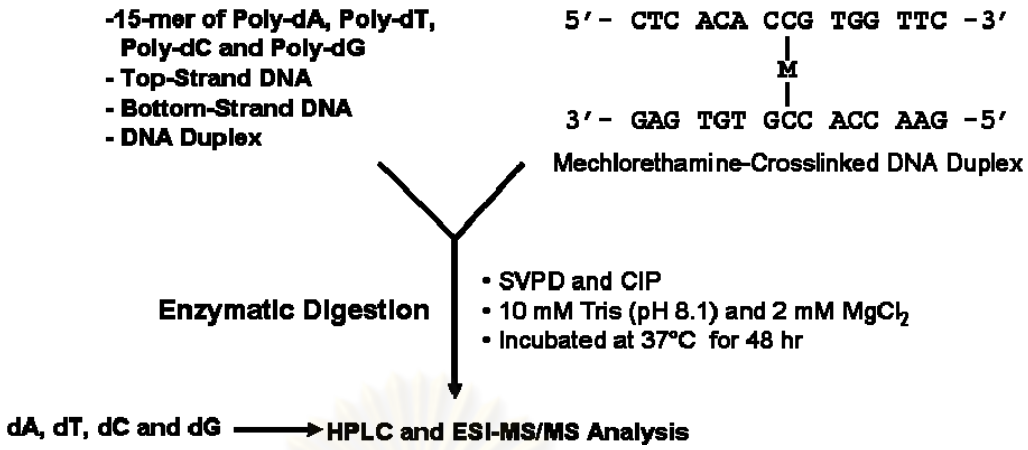


Figure 3.7 HPLC and ESI-MS/MS determination of digested products and target of mechlorethamine on the DNA duplex

3.2.12 Determination of digested products by HPLC:

Monodeoxynucleoside working standard solutions and the digested products of oligonucleotides and mechlorethamine-crosslinked DNA duplex were analyzed by HPLC. Chromatography was performed using a Rainin Microsorb-C18 column (4.6x250mm, 5µm) at 25°C with gradient elution of 2-20% ACN in phosphate buffer (pH 5.8) containing 50 mM sodium phosphate in 20 minutes at a flow rate of 1.0 ml/min. UV-detection was set at 260 nm and the injection volume was 20 µl. The HPLC condition and the plot of the gradient elution of the mobile phase are shown in Figure 3.8.

Column : Rainin Microsorb-C18 column (4.6x150mm, 5 µm)

Mobile phase : 2-20% ACN in phosphate buffer (pH 5.8) containing 50 mM sodium phosphate in 20 minutes

Flow rate : 1 ml/min

Oven temperature : 25°C

Detector : UV 260 nm

Injection volume : 20 µl

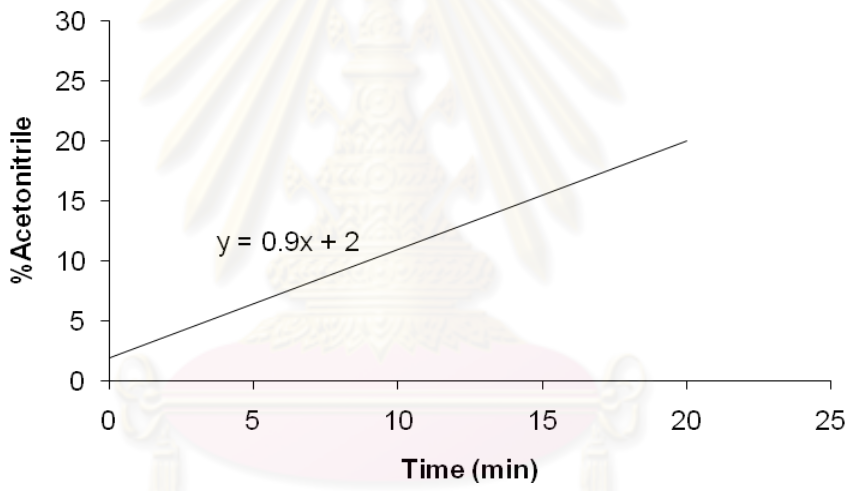


Figure 3.8 Gradient elution of mobile phase; 2-20% ACN in 20 minutes

3.2.13 Determination of the target of mechlorethamine in a DNA duplex containing a C-C mismatch pair:

The mechlorethamine-crosslinked DNA duplex digested by SVPD and CIP and purified by Microcon® YM-3 (section 3.2.11) was analyzed by ESI-MS/MS using an API 4000 triple quadrupole instrument in the product ion scan mode, with a turbo-ion spray source in positive mode. The parameters for the

mass spectrometer are summarized in Table 3.2. The samples were directly infused into the mass spectrometer at an infusion flow rate of 10 µl/minutes. Data processing was performed using Analyst software (version 4.4.2). In this experiment, three digested products were focused: dC-mechlorethamine-Cl (dC-mech-Cl, MW = 347.82), dC-mechlorethamine-OH (dC-mech-OH, MW = 329.38) and dC-mechlorethamine-dC (dC-mech-dC, MW = 539.59).

Table 3.2 Mass spectrometer parameters

Parameters	**Value**		
	dC-M-Cl	**dC-M-dC**	**dC-M-OH**
GS1 (psi)	20	20	20
GS2 (psi)	0	0	0
CAD gas (psi)	4	4	4
Curtain gas (psi)	10	10	10
Temperature (°C)	0	0	0
Ion spray voltage (V)	4500	4500	4500
Declustering Potential (V)	20	20	20
Entrance Potential (V)	10	10	10
Collision cell enhance potential (V)	10	20	10
Collision cell exit potential (V)	10	35	15
Start (amu)	50	50	50
Stop (amu)	400	300	350
Time (s)	0.5	0.5	0.5
Mode of analysis	Positive	Positive	Positive

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Determination of the structure and amount of crosslink formed by mechlorethamine with a DNA duplex containing a C-C mismatch pair

4.1.1 HPLC analysis of DNA duplex formation:

To ensure that the DNA was annealed before incubation with mechlorethamine, three control experiments were performed with top-strand DNA, bottom-strand DNA, and a mixture of the two, each prepared in an identical buffer to that used for the DNA duplex. The samples were subjected to HPLC analysis using the condition described in section 3.1.10. Top-strand and bottom-strand DNA showed peaks with retention times of 27.87 and 25.07 minutes, respectively, as shown in Figures 4.1 and 4.2. The mixture of top-strand and bottom-strand DNA in Figure 4.3 shows two significant peaks with retention times of 25.56 and 27.93 minutes, which correspond to bottom-strand and top-strand DNA, respectively.

The double-strand DNA solution obtained from the annealing process was analyzed by reversed-phase HPLC using the conditions described in section 3.1.10. The chromatogram of the DNA duplex (Figure 4.4) shows two significant peaks with retention times of 25.60 and 28.15 minutes. Interestingly, a shoulder was observed on the second peak on the chromatogram in Figure 4.4.

Comparison of the chromatogram in Figure 4.4 with the chromatograms for the top strand and bottom strand in Figures 4.1 and 4.2, respectively, suggests that the first peak in Figure 4.4 is bottom-strand DNA and the second peak (with a shoulder) is top-strand DNA. This peak may overlap with incompletely denatured secondary structures of the top or bottom strand. However, the absence of the shoulder in the second peak in the chromatogram of the mixture of top and bottom strands (Figure 4.3) suggests that the second peak with a shoulder in the chromatogram of the DNA duplex (Figure 4.4) is top-strand DNA overlapped with incompletely denatured secondary structures of the top strand or bottom strand.

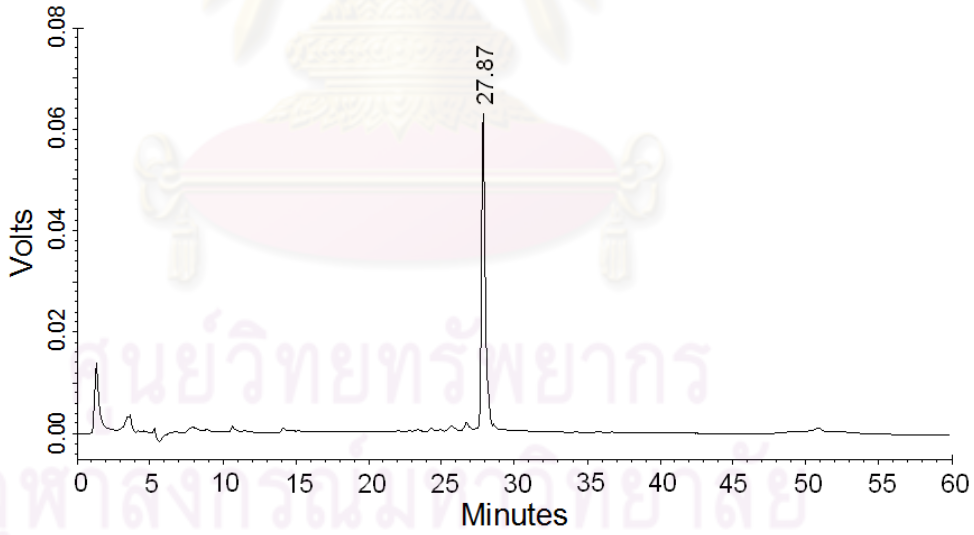


Figure 4.1 Chromatogram of top-strand DNA at 33°C showing a peak with a retention time of 27.87 minutes (peak area = 1187428)

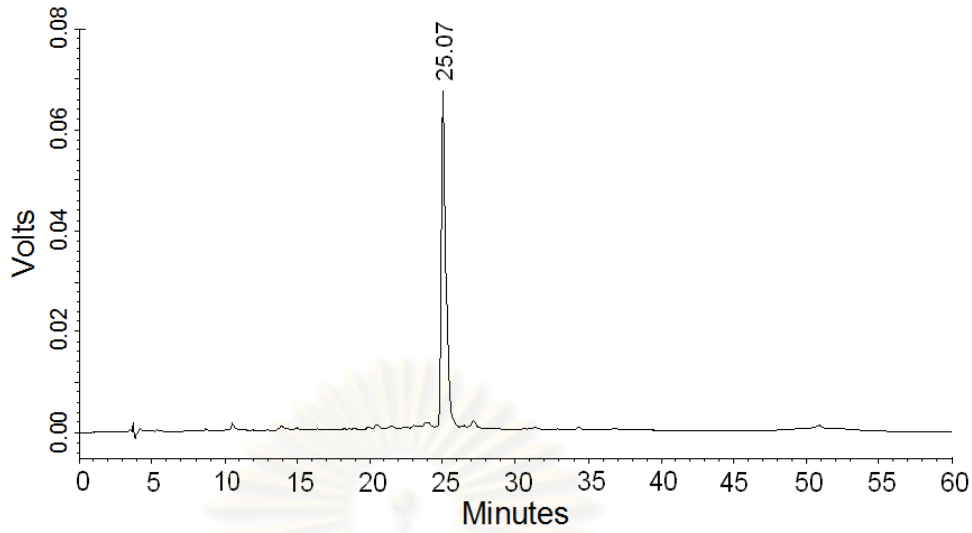


Figure 4.2 Chromatogram of bottom-strand DNA at 33°C showing a peak with a retention time of 25.07 minutes (peak area = 1480675)

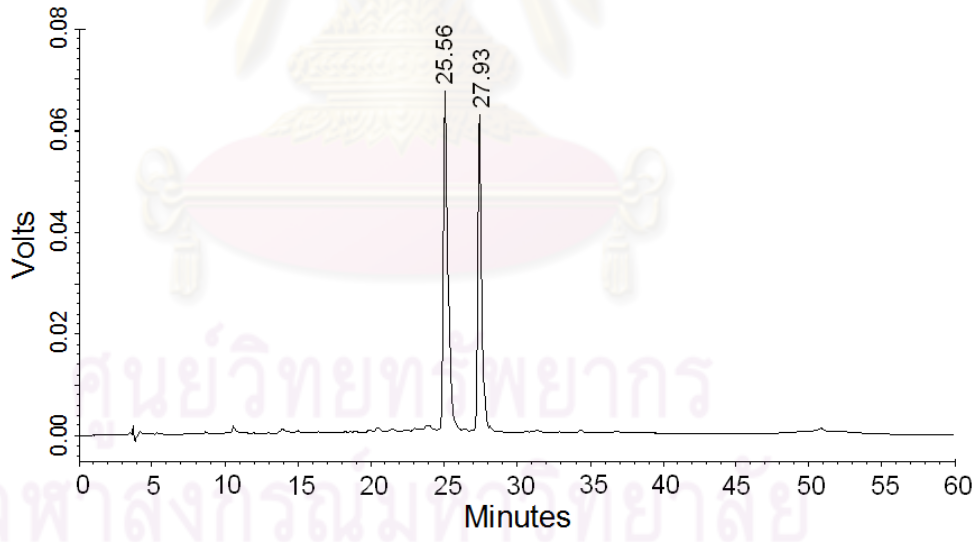


Figure 4.3 Chromatogram of the mixture of top- and bottom-strand DNA at 33°C showing two peaks with retention times of 25.56 (peak area = 1265532) and 27.93 minutes (peak area = 1457943) corresponding to the bottom strand and top strand, respectively

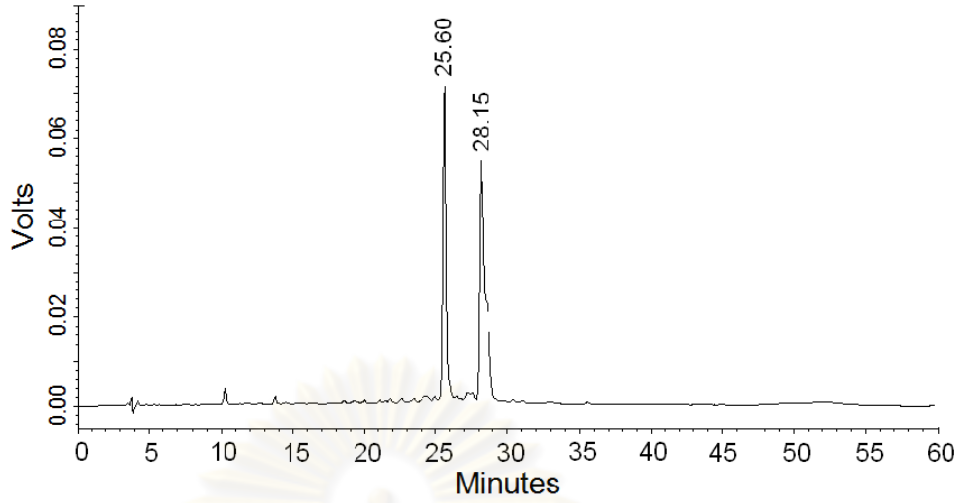


Figure 4.4 Chromatogram of DNA duplex at 33°C showing a peak corresponding to the bottom strand with a retention time of 25.60 minutes (peak area = 1273939) and a peak corresponding to the top strand with a retention time of 28.15 minutes (peak area = 1486238). Top-strand DNA peak may overlap with a peak for incompletely denatured DNA duplex, which appears as a shoulder on the top-strand peak.

The retention times of top- and bottom-strand DNA on ion-pair reversed-phase HPLC are slightly different. Comparing the chromatogram of top- and bottom-strand DNA at 33°C (Figure 4.1 and Figure 4.2), the retention time of top-strand DNA is a little longer than that of bottom-strand DNA (27.87 minutes for the top strand compared with 25.07 minutes for the bottom strand), although the molecular weight of bottom-strand DNA (MW = 4603) is higher than that of top-strand DNA (MW = 4505). The different retention time is possibly due to length of DNA sequences, base composition and order of bases within the sequences (Gilar et al., 2002).

Top- and bottom-strand DNA are 15-mers with different type and order of bases within the sequences. Therefore, the number of sugar and phosphate residues and the number of hydrogen bonding groups are similar, and the type and order of bases within the sequences are possibly responsible for making the retention times slightly different (Gilar et al., 2002).

Oligonucleotides of the same length but with different base composition have different retention times. The hydrophobicity contribution to the oligonucleotide retention time increases in order C<G<A<T (Gilar et al., 2002). Thymine is more retained than cytosine, and guanine is less retained than adenine. These suggest that for pyrimidine bases, thymine is more hydrophobic than cytosine, which is consistent with the presence of a methyl group at the 5´ position of thymine. For purine bases, guanine is less hydrophobic than adenine because guanine contains more polar groups (carbonyl and amino groups) compared to adenine (only an amino group). Top-strand DNA consists of 6 cytosines, 3 guanines, 2 adenines and 4 thymines (C:T = 1.5:1) whereas bottom-strand DNA consists of 4 cytosines, 5 guanines, 4 adenines and 2 thymines (C:T = 2:1). Therefore, the higher C:T ratios of the bottom strand make it less retained in HPLC column than the top strand.

For order of bases within the sequences, steric effects such as secondary structure formations presumably contribute to the retention of oligonucleotides (Gilar et al., 2002). The difference in order of bases in the sequences possibly makes different DNA conformers. These may be responsible for different

retention times between top- and bottom-strand DNA in addition to the impact of base composition.

The HPLC condition used in this experiment appears to affect the stability of the DNA duplex during the analysis. This may have resulted in the denaturation of DNA duplex. As shown in Figure 4.4, the chromatogram of DNA duplex analyzed by HPLC is almost identical to the mixture of top-strand and bottom-strand DNA in Figure 4.3. Comparison of the chromatograms of DNA duplex (Figure 4.4) and the mixture of top- and bottom-strand DNA (Figure 4.3) with the chromatograms of top-strand DNA (Figure 4.1) and bottom-strand DNA (Figure 4.2) indicates that the first peak with retention time 25.07-25.56 minutes in the chromatograms of DNA duplex and the mixture of top and bottom-strand DNA is due to bottom-strand DNA. Similarly, the second peak with retention times of 27.93-28.25 minutes is due to top-strand DNA. These results indicate that the HPLC conditions can denature the duplex DNA. This is very useful since the analysis of mechlorethamine-DNA crosslink is not interfered by the duplex DNA.

4.1.2 Determination and separation of mechlorethamine-crosslinked DNA duplex formation by HPLC:

The crosslinking reaction between DNA duplex and mechlorethamine was performed at room temperature for 1 hour, and 20 µl of the reaction mixture was then analyzed by HPLC using the conditions described in section 3.1.10. The chromatogram (Figure 4.5) shows three significant peaks with retention times of 25.73 (peak area = 425204), 28.37 (peak area = 721866) and 40.18 (peak area =

382883) minutes. A similar chromatogram (Figure 4.6) was obtained when the DNA duplex and mechlorethamine were incubated at room temperature for 2 hours, with retention times of 25.81 (peak area = 448024), 28.72 (peak area = 531988) and 40.54 (peak area = 332898) minutes. Based on comparison to the chromatograms of top-strand DNA, bottom-strand DNA and the mixture of the two strands in Figures 4.1, 4.2 and 4.3, respectively, the peak with a retention time of about 40 minutes is due to the mechlorethamine-crosslined DNA duplex. The retention times of all compounds are summarized in Table 4.1.

The amount of the mechlorethamine-crosslined DNA duplex was also determined based on peak areas of the crosslink product and unreacted (or monoadduct) DNA. The percentages of crosslink were 25.03% and 25.36% for 1- and 2-hour incubations, respectively. The results are summarized in Table 4.2.

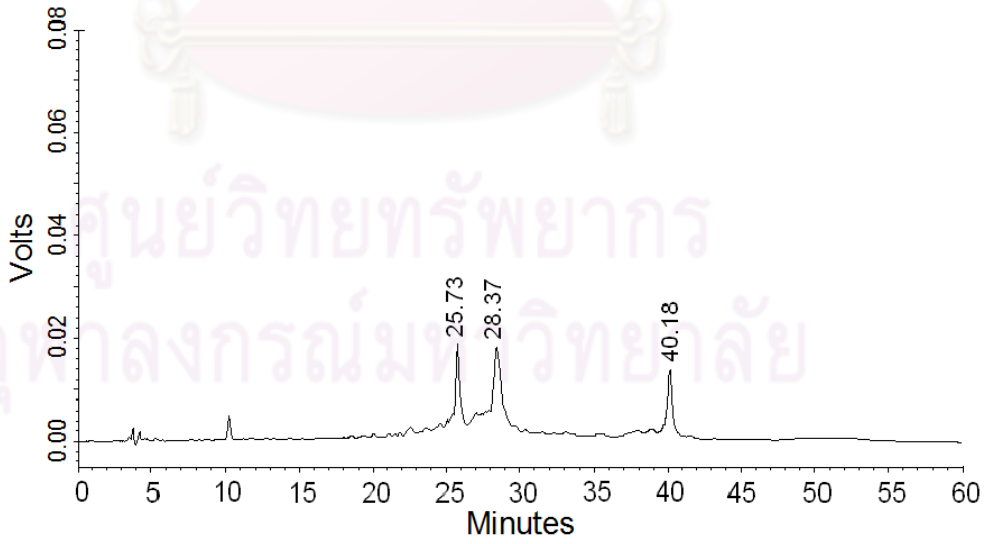


Figure 4.5 Chromatogram of the DNA duplex after reaction with mechlorethamine for 1 hour, showing three significant peaks with retention times of 25.73 (peak area = 425204), 28.37 (peak area = 721866) and 40.18 (peak area = 382883) minutes.

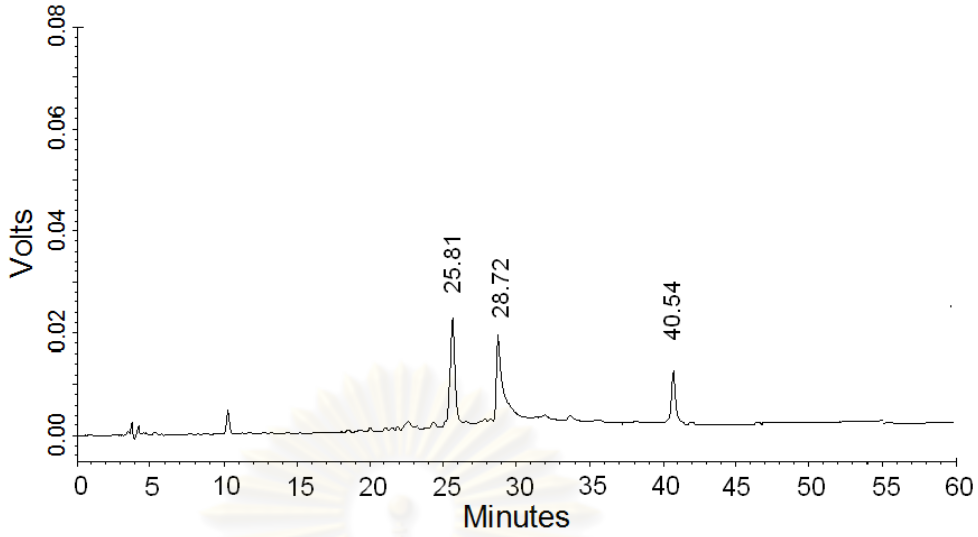


Figure 4.6 Chromatogram of the DNA duplex after reaction with mechlorethamine for 2 hours, showing three significant peaks with retention times of 25.81 (peak area = 448024), 28.72 (peak area = 531988) and 40.54 (peak area = 332898) minutes.

Table 4.1 Retention times of unreacted top-strand DNA, unreacted bottom-strand DNA and a mechlorethamine-crosslinked DNA duplex analyzed by reversed-phase HPLC

Incubation time (hr)	Retention time (minutes)		
	Top-strand	Bottom strand	Crosslink
Control	27.87	25.07	-
1	28.37	25.73	40.18
2	28.19	25.66	40.30

Table 4.2 Peak areas of unreacted top-strand DNA, unreacted bottom-strand DNA, crosslinked DNA and percentage of crosslink after 1 and 2-hour mechlorethamine-DNA reactions

Incubation time (hr)	Peak area			% Crosslink*
	Unreacted top strand	Unreacted bottom strand	Crosslink	
1	425204	721866	382883	25.03
2	448024	531988	332898	25.36

$$^{*}\%\ \text{crosslink} = \frac{(\text{crosslink}) \times 100}{[(\text{unreacted top}) + (\text{unreacted bottom}) + (\text{crosslink})]}$$

where (crosslink), (unreacted top), (unreacted bottom) are the peak areas of the crosslinked DNA, unreacted top-strand DNA and unreacted bottom-strand DNA, respectively.

The DNA duplex was reacted with mechlorethamine and the mixture was analyzed using reversed-phase HPLC after incubation periods of 1 and 2 hours. After a DNA duplex is reacted with mechlorethamine, the mixture of DNA and mechlorethamine consists of many compounds: unreacted top-strand DNA, unreacted bottom-strand DNA, top-strand DNA monoadducts, bottom-strand DNA monoadducts, unreacted DNA duplex, the DNA duplex monoadducts, and the mechlorethamine-crosslinked DNA duplex; as well as some single and DNA duplex that reacted with more than one molecule of mechlorethamine (Balcome et al., 2004; Rink and Hopkins, 1995[b]; Rink et al., 1993). However, when the reaction mixture was analyzed by HPLC, unreacted DNA duplex and DNA duplex-monoadducts may have been denatured and separated into top and bottom single-strand DNA and top or bottom-strand monoadducts, as shown in Figure 4.7. Therefore, the DNA duplex denaturation under the chosen HPLC conditions

is helpful for clear separation of the crosslink. The separation of unreacted DNA duplex and DNA duplex-mechlorethamine monoadducted complex into single-strand DNA is important because the peaks of unreacted DNA duplex and DNA duplex-monoadducts may not be separated from the crosslink under other HPLC conditions.

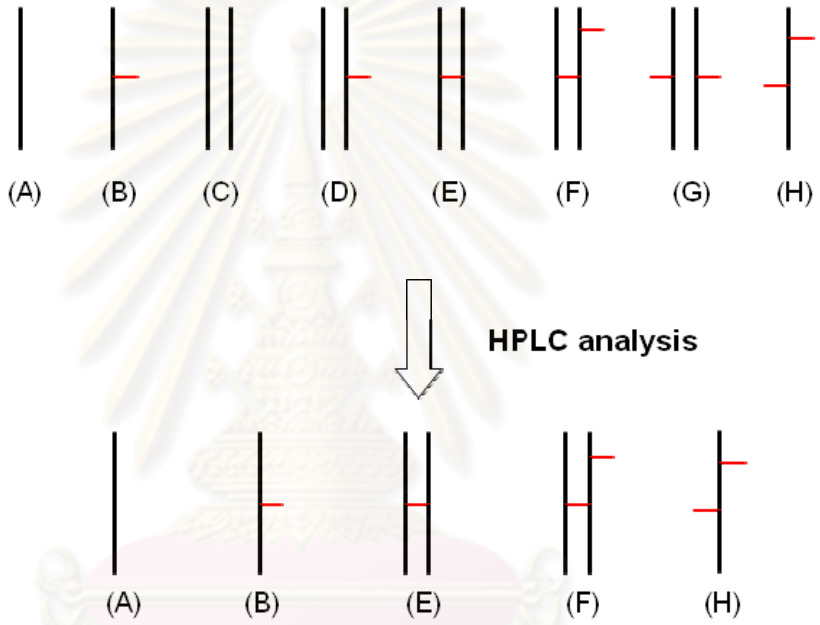


Figure 4.7 Compounds in the mixture after the DNA duplex was reacted with mechlorethamine and after HPLC analysis: the unreacted top or bottom single-strand DNA (A), top or bottom strand monoadducts (B), unreacted DNA duplex (C), DNA duplex monoadduct (D), mechlorethamine-crosslinked DNA duplex (E), examples of double and single-strand DNA that reacted with more than one mechlorethamine (F), (G) and (H).

After the incubation period of 1 hour, DNA duplex reacted with mechlorethamine was analyzed by reversed-phase HPLC. The chromatogram (Figure 4.5) shows three peaks with retention times of 25.73, 28.37 and 40.18 minutes, respectively. In comparison with the control chromatograms of top, bottom and DNA duplex recorded under the same conditions, the chromatograms of top-strand DNA (Figure 4.1) and bottom-strand DNA (Figure 4.2) have retention times of 27.87 and 25.07 minutes, respectively, and the chromatogram of DNA duplex (Figure 4.4) has two peaks for top-strand and bottom-strand DNA with retention time of 28.15 and 25.60 minutes respectively. Hence, the first two peaks with retention times of 28.37 and 25.73 minutes in the chromatogram of the DNA duplex after reaction with mechlorethamine correspond to top-strand and bottom-strand DNA, respectively, and the third peak with a retention time of 40.18 minutes may be the mechlorethamine-crosslinked DNA duplex at the C-C mismatch pair (the DNA duplex does not contain a potential 1,3 G-G crosslink site or a 1,2 G-G site). The small peak close to the peak for top-strand DNA may be a single-strand monoadduct.

The reaction mixture of DNA duplex and mechlorethamine after 2-hour incubation was analyzed using the same HPLC conditions. The chromatogram in Figure 4.6 has three peaks that resemble those in the chromatogram after 1-hour incubation. The ratio of the mechlorethamine-DNA crosslink peak area to the summed areas of unreacted DNA and monoadducts was also similar for the two incubation times.

The amount of the mechlorethamine-crosslinked DNA duplex was calculated based on the peak area of the crosslink and the unreacted DNA/monoadducts. The percentages of the crosslink were 25.03 and 25.36% in the 1- and 2-hour incubations, respectively, with no significant difference between these results. Hence, the DNA duplex should be incubated with mechlorethamine for 1 hour, rather than 2 hours, to enhance purity, since the longer incubation times can lead to formation of multiple adducts (Romero et al., 2001). Therefore, the incubation time is a compromise between obtaining a high yield of the expected crosslink with the lowest amount of side-products.

This amount of crosslink found in this study (25.03% and 25.36% after 1- and 2-hour incubations, respectively) is close to the percentages of crosslink that have been determined previously (27%) (Romero et al., 2001). The slight differences may be due to the lengths of duplexes and the incubation temperature. In this study, a 15-mer DNA duplex was incubated with mechlorethamine at room temperature, while Romero et al. performed reactions with a 19-mer duplex and an incubation temperature of 37°C (Romero et al., 1999; Romero et al., 2001). The amount of crosslink formation depends on local stability and conformational fluctuation around the C-C mismatch (Romero et al., 2001), and a longer sequence may increase the duplex stability and allow more crosslinking. The mechlorethamine-DNA reaction is also increased at higher temperature.

4.1.3 Characterization of the mechlorethamine-crosslinked DNA duplex by MALDI-TOF-MS:

The mechlorethamine-crosslinked DNA duplex was purified and collected from HPLC analysis, desalted and prepared for MALDI-TOF-MS analysis to determine the molecular weight of the crosslinked DNA. The mass spectrum of bottom-strand DNA in Figure 4.8A shows a significant peak with m/z of 4604.219 (expected MW = 4603) whereas top-strand DNA in Figure 4.8B show a significant peak with m/z 4505.410 (expected MW = 4505). The mass spectrum in Figure 4.8C shows a significant peak with m/z 9222.088 (expected MW = 9193.2). Comparison of the m/z and the calculated mass suggests that the peak with m/z 9222.088 corresponds to the mechlorethamine-crosslinked DNA duplex.

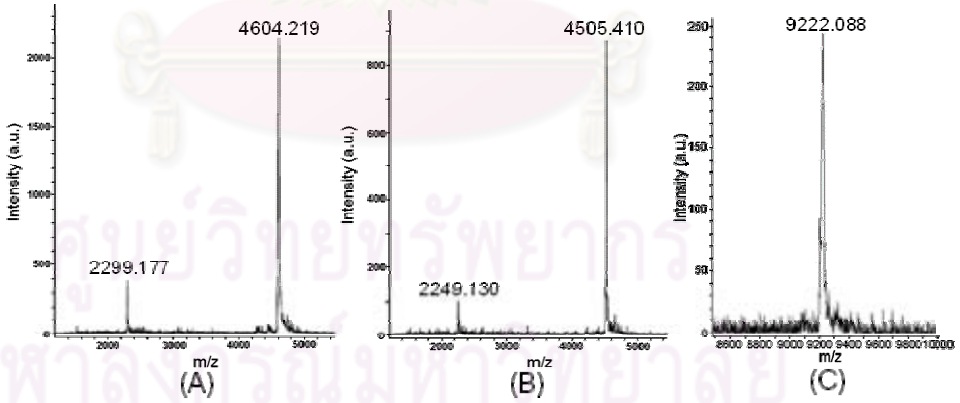


Figure 4.8 MALDI-TOF-MS spectra of bottom-strand DNA (A), top-strand DNA (B) and the mechlorethamine-DNA crosslink (C) with a peak at m/z 4604.219, 4505.410 and 9222.088, respectively

MALDI-TOF-MS is a versatile tool in the post-genome era for analysis of biomolecules (Tost and Gut, 2006) and advances in the application of MALDI-TOF-MS for more demanding DNA analysis demonstrates the potential of this method for analysis of nucleic acids (Gut, 2004; Tost and Gut, 2006). To examine mechlorethamine-crosslinked DNA duplex formation, the crosslink peak at 40.18 minutes (Figure 4.5) was collected and analyzed by MALDI-TOF-MS. In the mass spectrum (Figure 4.8), the m/z ratio (m/z 9222.088) is close to the expected molecular weight of 9193.2. Although the m/z ratios obtained from the experiment are close to the expected molecular weight, these results of MALDI-TOF-MS analysis are not as good as expected. This problem has been explained by Gut based on the monotonous nature of DNA, with different sequences displaying little or no difference in desorption efficiency with current matrix substances (Gut, 2004). This may cause peak broadening and reduction of resolution, sensitivity and accuracy due to the large kinetic energy differences of ions generated at the surface of the sample target (Bakhtiar and Nelson, 2000; Gut, 2004; Tost and Gut, 2006). Hence, deviations of m/z ratios in MALDI-TOF-MS analysis of nucleic acids are common.

4.2 Determination of the target of mechlorethamine in a DNA duplex containing a C-C mismatch pair

4.2.1 Enzymatic digestion of the mechlorethamine-crosslinked DNA duplex:

To determine the target of mechlorethamine in a DNA duplex containing a C-C mismatch pair, top-strand DNA, bottom-strand DNA, DNA duplex, control sequences of poly-dA, poly-dT, poly-dC and poly-dG and the mechlorethamine

crosslinked DNA duplex were subjected to enzymatic digestion with a combination of SVPD (3´-exonuclease) and CIP at 37°C for 48 hours. The methods of enzymatic digestion by SVPD and CIP were optimized from those used in three earlier studies (Noll et al., 2001; Winds et al., 2004; Winds et al., 2006). SVPD can selectively cleave phosphodiester bonds in the DNA backbone from the 3´-end of DNA, with production of monodeoxyribonucleotides (dAMP, dTMP, dGMP and dCMP) after digestion. The phosphate groups of the monodeoxyribonucleotides are then removed by CIP to give the equivalent monodeoxyribonucleosides. The scheme of enzymatic digestion by SVPD and CIP is shown in Figure 4.9. Based on these reactions, the mechlorethamine-crosslinked DNA duplex should be hydrolyzed to its component nucleosides containing dA, dT, dG and dC as well as a dC-mechlorethamine-dC (dC-mech-dC) fragment, as shown in Figure 4.10.

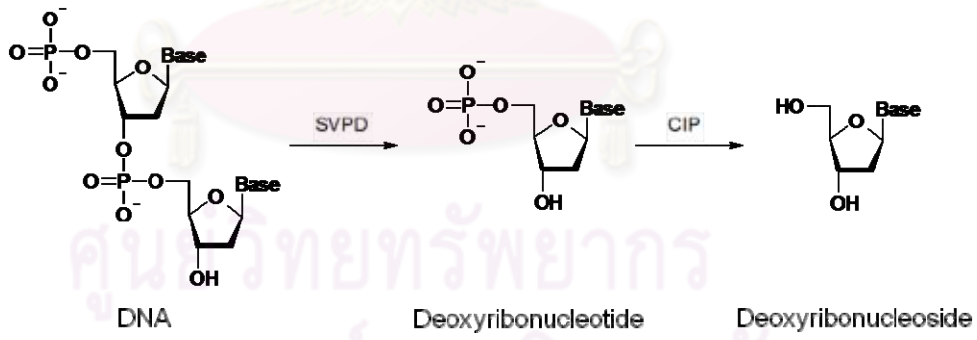


Figure 4.9 Scheme showing enzymatic digestion by SVPD and CIP

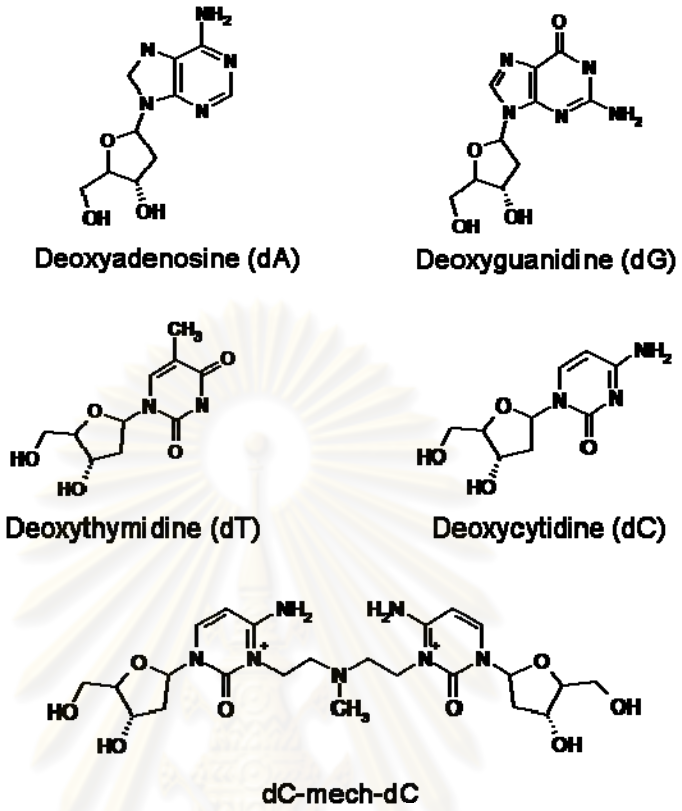


Figure 4.10 Expected digestion products of a mechlorethamine-crosslinked DNA duplex containing a C-C mismatch pair after enzymatic digestion by SVPD and CIP

4.2.2 HPLC analysis of products obtained from enzymatic digestion:

The control sequences of 15-mer polydeoxyribonucleotides (poly-dA, poly-dT, poly-dC and poly-dG) were prepared in the identical buffer and digested by a combination of SVPD and CIP at 37°C for 48 hours. The digested polydeoxyribonucleotides were each analyzed by HPLC using the conditions described in section 3.1.12. The HPLC conditions were modified from those used to determine the N^4C-Ethyl-N^4C interstrand crosslink (Noll et al., 2001). According to the chromatograms in Figures 4.11, 4.12, 4.13 and 4.14, the digested products of poly-dC, poly-dG, poly-dT and poly-dA show significant

peaks with retention times of 3.83 (peak area = 1532776), 8.15 (peak area = 29914), 9.32 (peak area = 1268076), and 11.13 (peak area = 86462) minutes, respectively. Interestingly, the chromatogram of poly-dA shows a small peak at 7.88 minutes, which may be a produce due to incomplete digestion. The other chromatograms suggest that enzymatic digestion of poly-dC, poly-dG and poly-dT yields only one product. The identities of the digested products of poly-dC, poly-dG, poly-dT and poly-dA were confirmed by comparison with standard monodeoxyribonucleotides. The overlay chromatogram of digested poly-dC, poly-dG, poly-dT and poly-dA in Figure 4.15 showed that the digested products of these four control sequences are separated from each other using the HPLC conditions in the study.

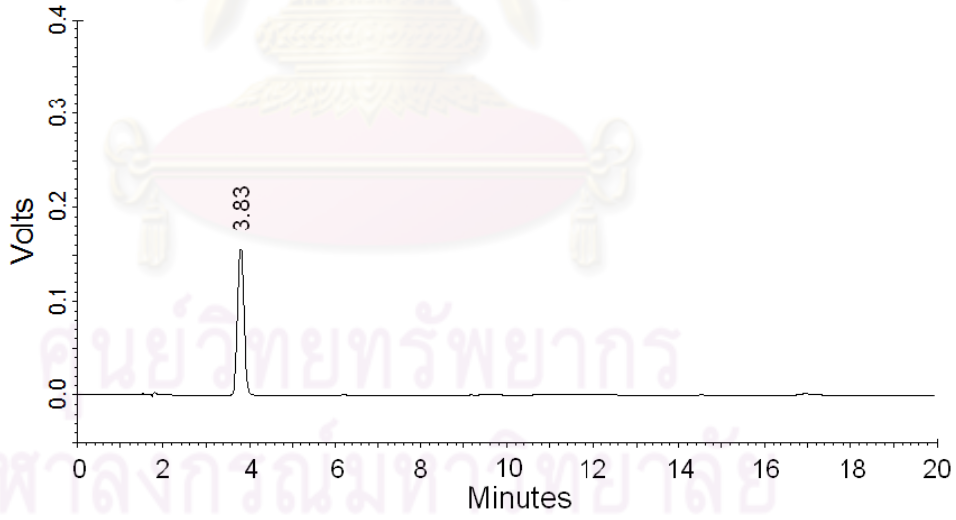


Figure 4.11 Chromatogram of poly-dC after digestion with SVPD and CIP at 37°C for 48 hours

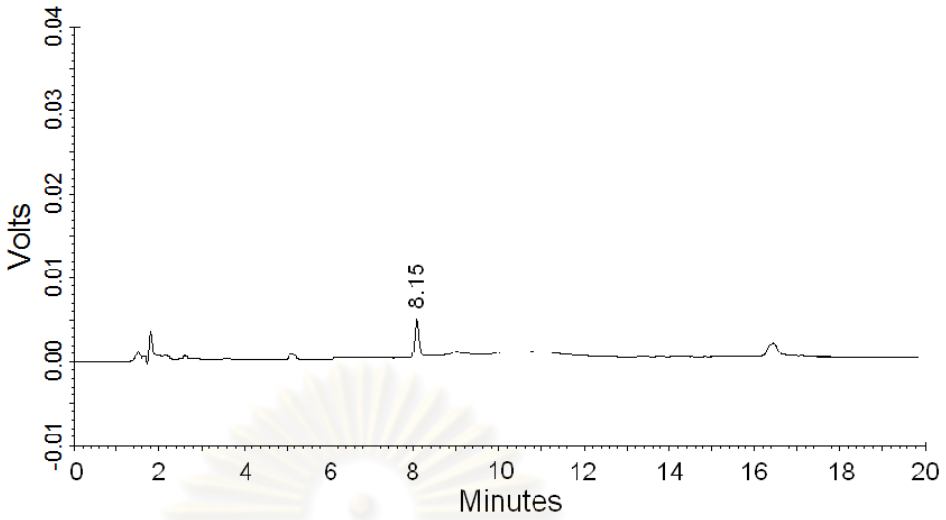


Figure 4.12 Chromatogram of poly-dG after digestion with SVPD and CIP at 37°C for 48 hours

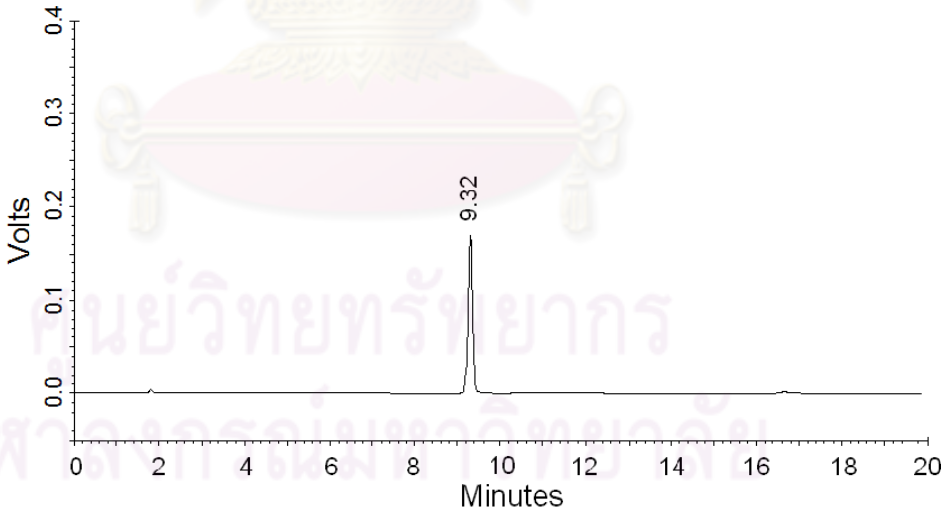


Figure 4.13 Chromatogram of poly-dT after digestion with SVPD and CIP at 37°C for 48 hours

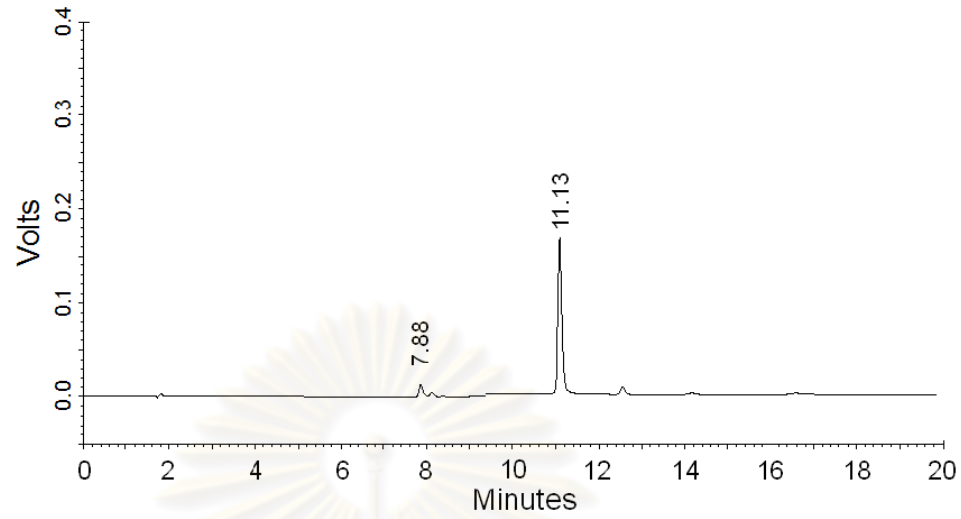


Figure 4.14 Chromatogram of poly-dA after digestion with SVPD and CIP at 37°C for 48 hours

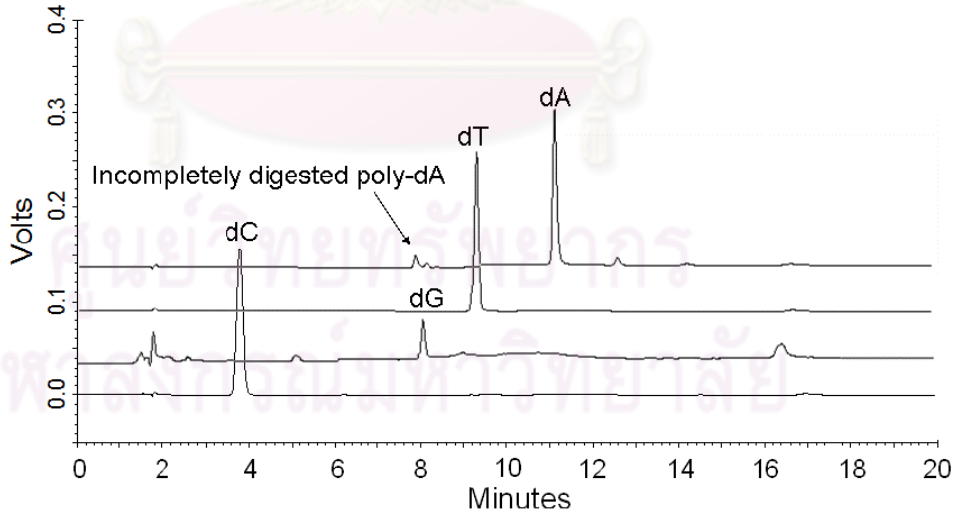


Figure 4.15 Overlay chromatogram of digested poly-dC, poly-dG, poly-dT and poly-dA

To identify the digested products of the control sequences, working standard solutions of dA, dT, dG and dC, were prepared and analyzed by HPLC using the condition described in section 3.1.12. The chromatograms in Figures 4.16, 4.17, 4.18 and 4.19 show that standard dC, dT, dG and dA each show a single peak with retention times of 3.81 (peak area = 1235993), 8.18 (peak area = 1829518), 9.25 (peak area = 1075144) and 11.08 (peak area = 1192023) minutes, respectively. The retention times of the standard dC, dT, dG and dA were used as references for HPLC analysis of the digested products of the control sequences and mechlorethamine-crosslinked DNA duplex. The overlay chromatogram of dC, dG, dT and dA in Figure 4.20 showed that these four monodeoxynucleosides are separated from each other using the HPLC conditions in the study. Comparison of the chromatograms of digested control sequences with those of standard monodeoxyribonucleosides shows that the digested products of poly-dC, poly-dG and poly-dT are dC, dG and dT at retention times of about 3.8, 8.1 and 9.3 minutes, respectively, whereas poly-dA shows two digested products: the peak with retention time about 11.1 minutes is dA and the peak at 7.8 minutes is incompletely digested poly-dA. These results suggested that polydeoxyribonucleosides are mostly digested to their component monodeoxyribonucleosides by a combination of SVPD and CIP at 37°C for 48 hours.

The retention time on reversed-phase HPLC depends on the hydrophobicity of the solute. In this experiment, dT was eluted after dC and dG was eluted before dA. These results suggest that thymine is the more hydrophobic pyrimidine, which is consistent with the presence of a methyl group at the 5´

position of thymine. For purine bases, guanine is less hydrophobic than adenine because guanine contains more polar groups (carbonyl and amino groups) compared to adenine (an amino group).

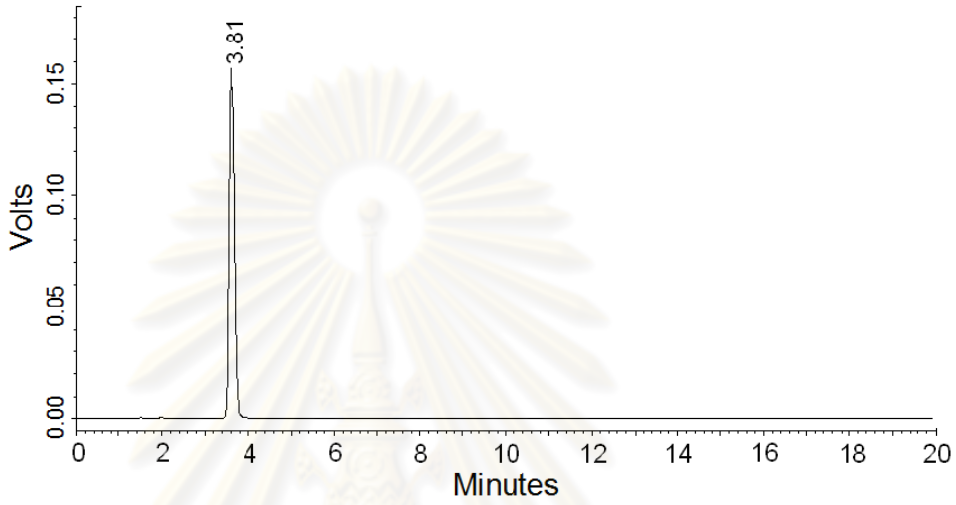


Figure 4.16 Chromatogram of standard dC showing a significant peak with a retention time of 3.81 minutes (peak area = 1235993)

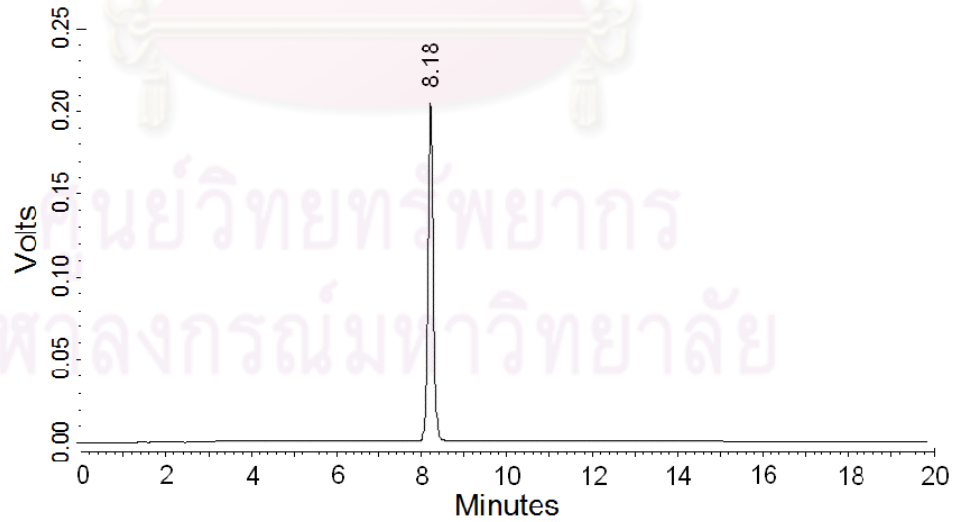


Figure 4.17 Chromatogram of standard dG showing a significant peak with a retention time of 8.18 minutes (peak area = 1829518)

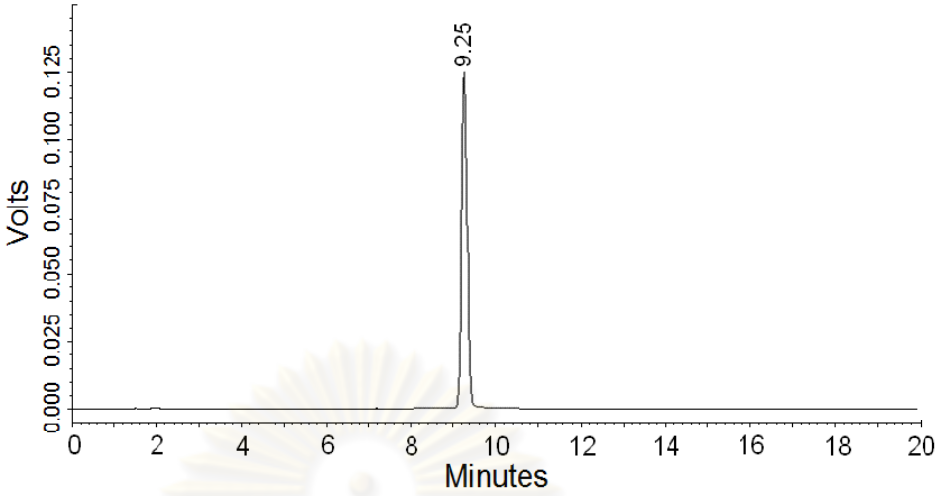


Figure 4.18 Chromatogram of standard dT showing a significant peak with a retention time of 9.25 minutes (peak area = 1075144)

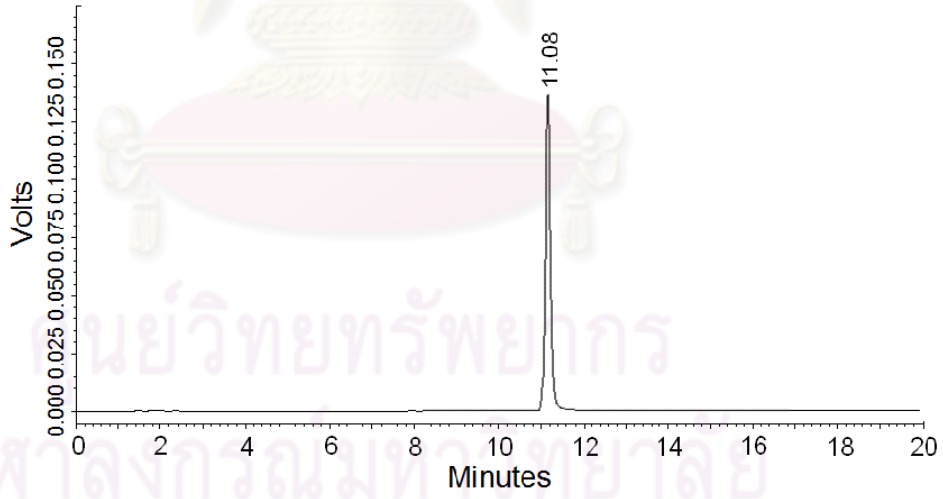


Figure 4.19 Chromatogram of standard dA showing a significant peak with a retention time of 11.08 minutes (peak area = 1192023)

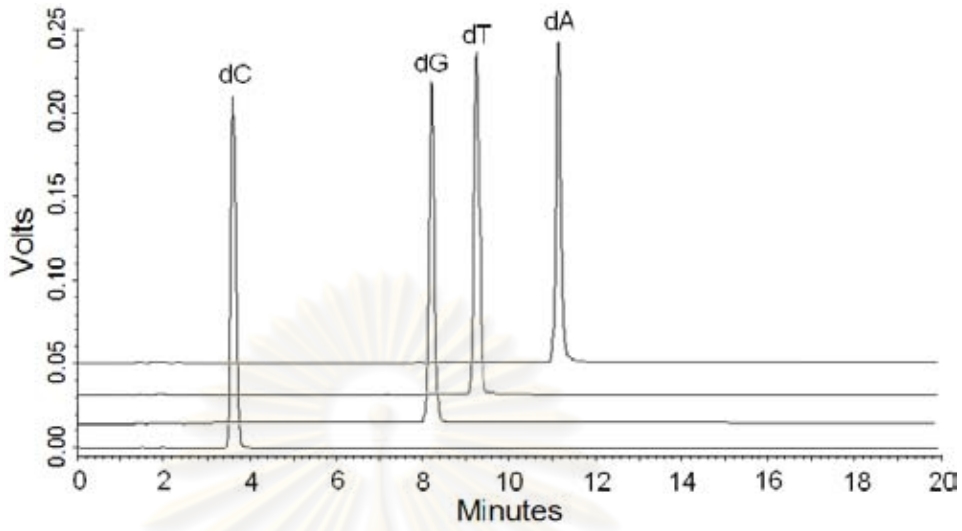


Figure 4.20 Overlay chromatogram of standard dC, dG, dT and dA

To ensure that the two designed DNA sequences are digested by SVPD and CIP and to confirm the nucleoside ratio of these sequences, top-strand DNA, bottom-strand DNA and DNA duplex were prepared and digested by SVPD and CIP at 37°C for 48 hours. The digested products were subjected to HPLC analysis using the conditions described in section 3.1.12. The chromatograms in Figures 4.21, 4.22 and 4.23 show five significant peaks corresponding to dC, dG, dT, dA and incompletely digested poly-dA. For top-strand DNA, five peaks with retention times of 3.84, 7.89, 8.27, 9.28 and 11.14 minutes were identified as dC, incompletely digested poly-dA, dG, dT and dA, respectively. For bottom-strand DNA, five peaks with retention times of 3.82, 7.91, 8.31, 9.26 and 11.16 minutes are identified as dC, incompletely digested poly-dA, dG, dT and dA, respectively. For DNA duplex, five peaks with retention times of 3.88, 7.91, 8.32, 9.33 and 11.19 minutes were identified as dC, incompletely digested poly-dA, dG, dT and dA, respectively. The HPLC analysis of top-strand DNA, bottom-strand DNA

and DNA duplex indicated that the DNA single strands and duplex are digested by a combination of SVPD and CIP at 37°C for 48 hours. The nucleoside ratios are summarized in Table 4.6. The percentage of dA cannot be determined due to the presence of incompletely digested poly-dA.

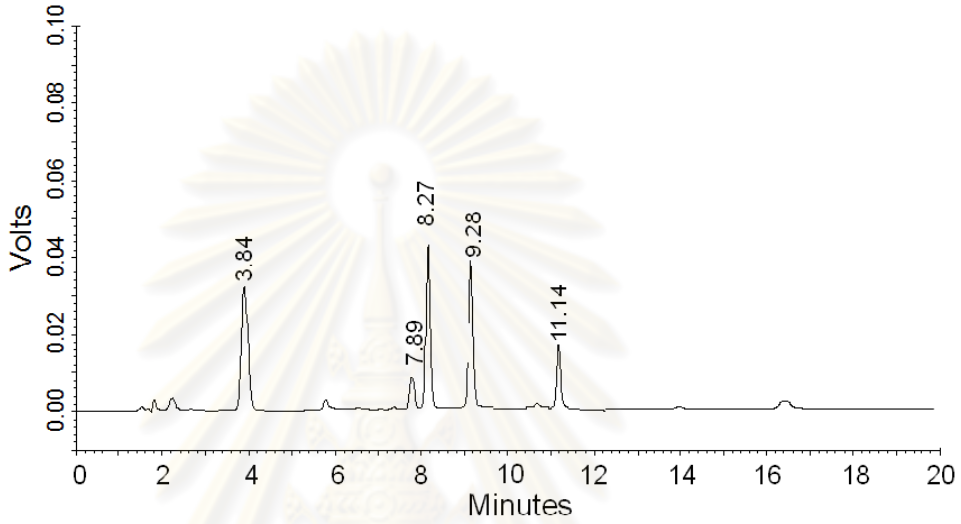


Figure 4.21 Chromatogram of top-strand DNA after digestion with SVPD and CIP at 37°C for 48 hours

Table 4.3 Retention time, peak area, resolution and tailing factor of each significant peak for top-strand DNA after digestion with SVPD and CIP at 37°C for 48 hours

Peak No.	Retention time (min)	Peak area	Resolution	Tailing factor
1	3.84	334297	11.04	1.11
2	7.89	56305	18.98	1.14
3	8.27	288470	1.84	1.14
4	9.28	219496	4.92	1.11
5	11.14	128439	9.54	1.13

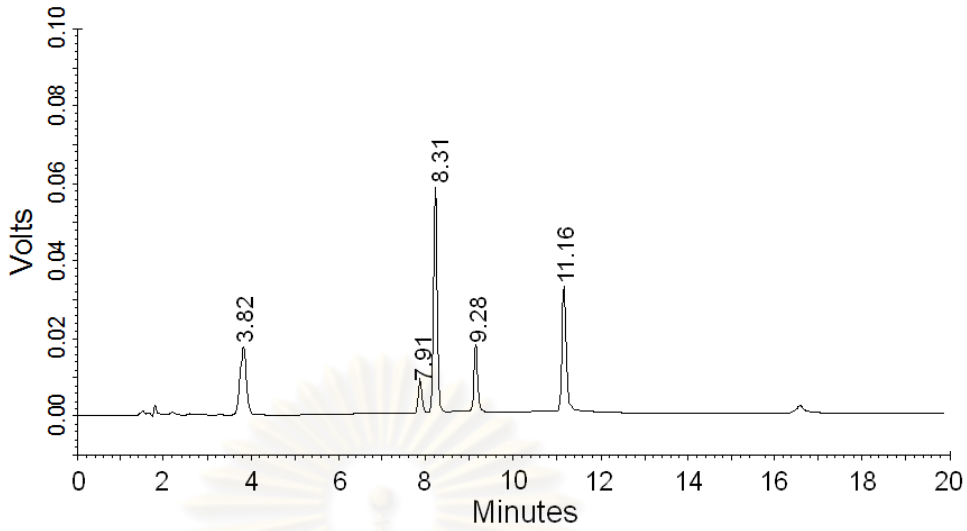


Figure 4.22 Chromatogram of bottom-strand DNA after digestion with SVPD and CIP at 37°C for 48 hours

Table 4.4 Retention time, peak area, resolution and tailing factor of each significant peak for bottom-strand DNA after digestion with SVPD and CIP at 37°C for 48 hours

Peak No.	Retention time (min)	Peak area	Resolution	Tailing factor
1	3.82	176746	10.98	1.10
2	7.91	61409	19.16	1.15
3	8.31	373152	2.00	1.15
4	9.28	113262	4.96	1.10
5	11.16	238991	9.62	1.12

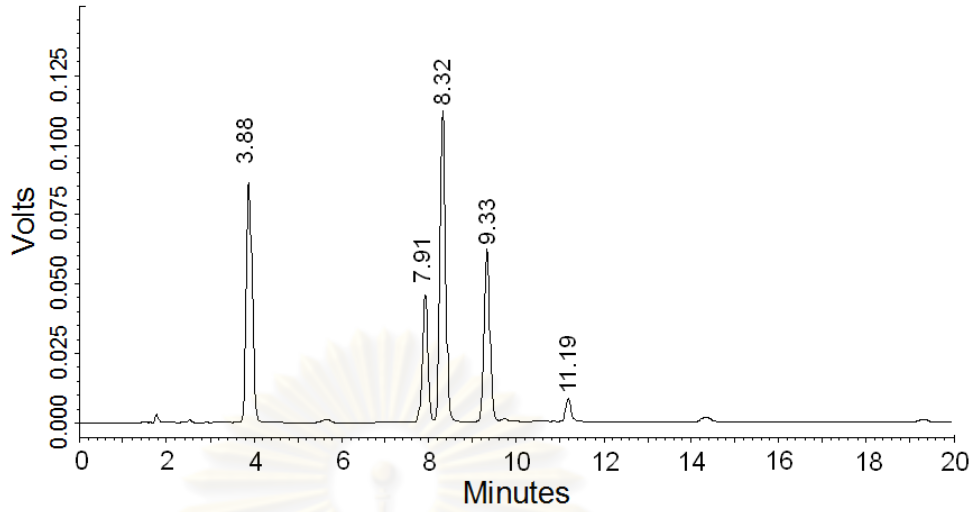


Figure 4.23 Chromatogram of DNA duplex after digestion with SVPD and CIP at 37°C for 48 hours

Table 4.5 Retention time, peak area, resolution and tailing factor of each significant peak for the DNA duplex after digestion with SVPD and CIP at 37°C for 48 hours

Peak No.	Retention time (min)	Peak area	Resolution	Tailing factor
1	3.88	804893	11.16	1.14
2	7.91	436202	18.80	1.10
3	8.32	1034704	1.99	1.10
4	9.33	593605	4.92	1.10
5	11.19	73541	9.54	1.08

Table 4.6 Nucleoside ratios of top-strand DNA, bottom-strand DNA and DNA duplex

	Nucleoside ratio	
	Expected	**Found**
Top-strand DNA		
- dC	2.00	1.99
- dT	1.33	1.08
- dG	1.00	1.00
Bottom-strand DNA		
- dC	2.00	1.88
- dT	1.00	1.00
- dG	2.50	2.32
DNA duplex		
- dC	1.67	1.63
- dT	1.00	1.00
- dG	1.33	1.23

The mechlorethamine-crosslinked DNA duplex was digested by a combination of SVPD and CIP at 37°C for 48 hours. The digested products were subjected to HPLC analysis using the condition described in section 3.1.12. After enzymatic digestion, the expected products are the four monodeoxyribonucleosides (dC, dG, dT and dA) and dC-mech-dC. The chromatogram of the digested mechlorethamine-crosslinked DNA duplex (Figure 4.24) shows eight significant peaks with retention times of 3.87, 7.88, 8.29, 9.31, 10.72, 11.19, 11.45 and 11.86 minutes, as shown in Table 4.7. Comparing the chromatogram of digested mechlorethamine-crosslinked DNA duplex with the chromatograms of four digested polydeoxyribonucleosides (Figure 4.11-4.14), four standard monodeoxyribonucleosides (Figure 4.16-4.19), digested top-strand

DNA (Figure 4.21), digested bottom-strand DNA (Figure 4.22) and digested DNA duplex (Figure 4.23), the five peaks with retention times of 3.87, 7.88, 8.29, 9.31 and 11.19 minutes are identified as dC, incompletely digested poly-dA, dG, dT and dA, respectively. This leaves three interesting peaks with retention times of 10.72, 11.45 and 11.86 minutes that are still unidentified. These peaks should be crosslinking products formed by mechlorethamine.

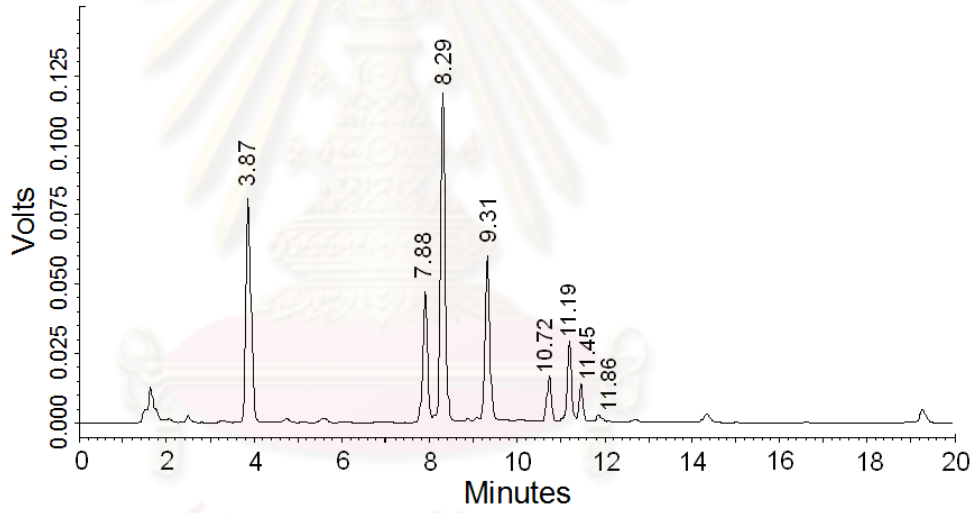


Figure 4.24 Chromatogram of the mechlorethamine-crosslinked DNA duplex after digestion with SVPD and CIP at 37°C for 48 hours

Table 4.7 Retention time, peak area, resolution and tailing factor of each significant peak for the mechlorethamine-crosslinked DNA duplex after digestion with SVPD and CIP at 37°C for 48 hours

Peak No.	Retention time (min)	Peak area	Resolution	Tailing factor
1	3.87	656077	11.13	1.14
2	7.88	377269	18.79	1.10
3	8.29	926624	1.99	1.10
4	9.31	322399	4.97	1.10
5	10.72	118052	7.28	1.08
6	11.19	208300	2.65	1.10
7	11.45	95973	1.49	1.24
8	11.86	25769	1.84	1.35

4.2.3 ESI-MS/MS analysis of the target of mechlorethamine in a DNA duplex containing a C-C mismatch pair:

Enzymatic digestion of mechlorethamine-crosslinked DNA duplex was complete after 48 hours of incubation at 37°C with a combination of SVPD and CIP. Mass spectra of a solution containing digested mechlorethamine-crosslinked DNA duplex are shown in Figures 4.25, 4.26 and 4.27, and indicate crosslink formation with a C-C mismatch pair. These mass spectra show molecular ions at m/z 347.3, 329.6 and 269.2 corresponding to dC-mech-Cl, dC-mech-OH and dC-mech-dC, respectively. The proposed structure of dC-mech-dC contains two positive charges at N^3 of each cytosine base (Romero et al., 1999). The detected m/z of dC-mech-dC is a half of its molecular weight because m/z represents mass to charge ratio of the ion. Thus the mass of dC-mech-dC detected by mass spectrometric technique is 538.4 g/mole. The mass spectral data are summarized

in Table 4.8. The results indicate that mechlorethamine forms an interstrand crosslink with the DNA duplex at a C-C mismatch pair. These results are also consistent with data from gel electrophoresis (Romero et al., 1999; Romero et al., 2001).

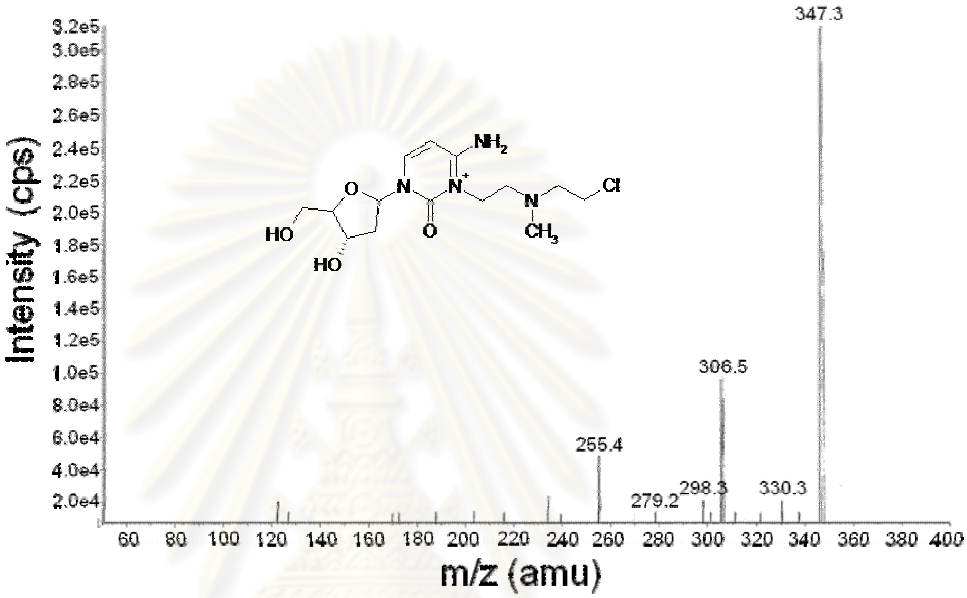


Figure 4.25 Mass spectrum of dC-mech-Cl

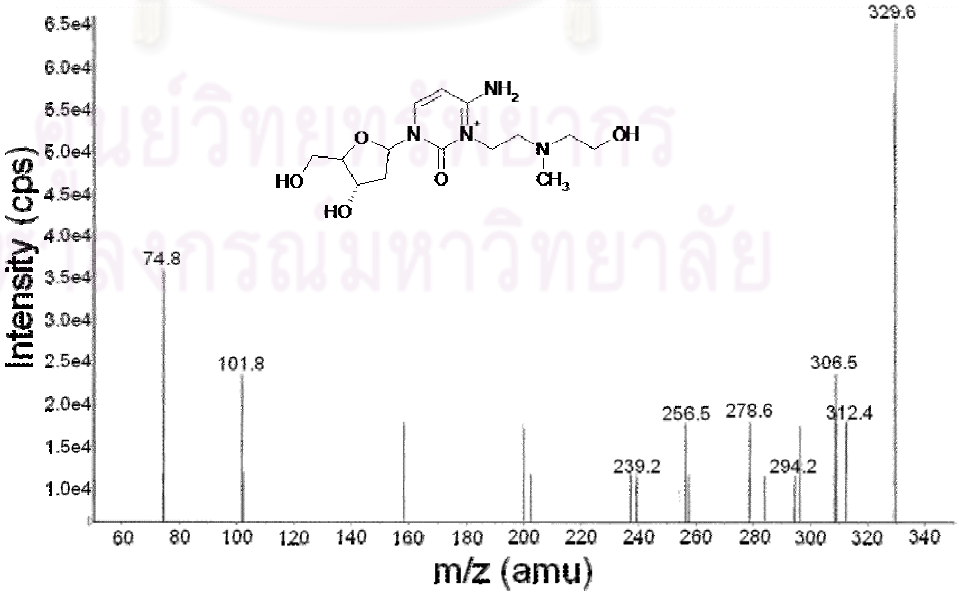


Figure 4.26 Mass spectrum of dC-mech-OH

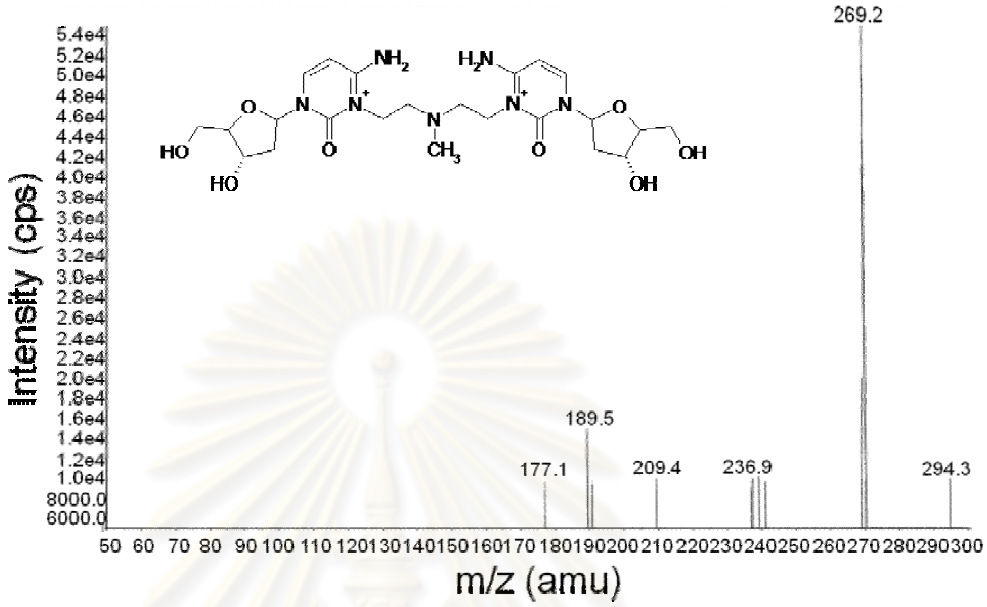


Figure 4.27 Mass spectrum of dC-mech-dC

Table 4.8 Mass spectral data for dC-mech-Cl, dC-mech-OH and dC-mech-dC

Products	m/z	
	Expected	Found
dC-mech-Cl	347.8	347.3
dC-mech-OH	329.4	329.6
dC-mech-dC	269.8	269.2

Identification of the daughter ions is important for determining the target atom of mechlorethamine on the cytosine bases of the mismatch pair and to characterize the molecular structures of dC-mech-Cl, dC-mech-OH and dC-mech-dC. For dC-mech-Cl, the mass spectrum (Figure 4.25) shows significant daughter ions at m/z 330.3, 306.5, 298.3 and 255.4 resulting from the fragmentation of dC-mech-Cl at m/z 347.3 $[M]^+$, corresponding to $[M\text{-}NH_3]^+$, $[M–(Cl+CH_2=CH_2)+Na)]^+$, $[M\text{-}(CH_2\text{-}Cl)+H]^+$ and $[M–(Cl+HN(CH_3)CH=CH_2)]^+$, respectively, as shown in Figure 4.28. For dC-mech-OH, the mass spectrum (Figure 4.26) shows significant daughter ions at m/z 312.4, 306.5, 294.2 and 237.2 resulting from fragmentation of dC-mech-OH at m/z 329.6 $[M]^+$, corresponding to $[M\text{-}NH_3]^+$, $[M–(H_2O+CH_2=CH_2)+Na)]^+$, $[M–(NH_3+H_2O)]^+$ and $[M–(NH_3+H_2O+HN(CH_3)CH=CH_2)]^+$, respectively, as shown in Figure 4.29. For dC-mech-dC, the mass spectrum (Figure 4.27) shows significant daughter ions at m/z 294.3, 236.9 and 177.1 resulting from fragmentation of dC-mech-dC at m/z 269.2 $[M]^{2+}$, corresponding to $[M–(dC+NH_3)]^+$, $[M–(dC+NH_3+ HN(CH_3)CH=CH_2)]^+$ and $[M–(dC+NH_3+\text{deoxyribose})]^+$, respectively, as shown in Figure 4.30.

Benzo[a]pyrene diol epoxide adduct formation at the exocyclic-NH_2 group of nucleotide base has been proved by the absence of ions resulting from the neutral loss of 17 u corresponding to NH_3 (Gaskell et al., 2007). In contrast, ions resulting from the neutral loss of NH_3 were observed from mass spectra of dC-mech-Cl, dC-mech-OH and dC-mech-dC in this experiment. This indicated that the exocyclic-NH_2 group of cytosine (N^4) does not react with mechlorethamine. In addition, the molecular ions and daughter ions of dC-mech-

Cl, dC-mech-OH and dC-mech-dC, and especially the presence of the m/z corresponding to $[M]^{2+}$ from mass spectrum of dC-mech-dC, provide strong evidence of crosslink formation between the N^3 atoms of the C-C mismatch pair.

The crosslink is formed between the cytosine N^3 atoms because these are the most nucleophilic atoms of the C-C mismatch pair. This is supported by the observation that C-C crosslink formation is pH-dependent, with crosslinking being much more efficient at a pH above pK_a of cytosine N^3 (pK_a = 6.95). However, it is also possible that N^3 protonation could induce a DNA conformational change that could prevent crosslink formation at another atom of cytosine (Romero et al., 1999), and therefore the pH dependence does not confirm that mechlorethamine forms an interstrand crosslink between the N^3 atoms of C-C mismatch pair. In contrast, the molecular structures of dC-mech-Cl, dC-mech-OH and dC-mech-dC (Figure 4.31) characterized in this study provide very strong evidence for the connectivity of the crosslink through the N^3 atoms of cytosine.

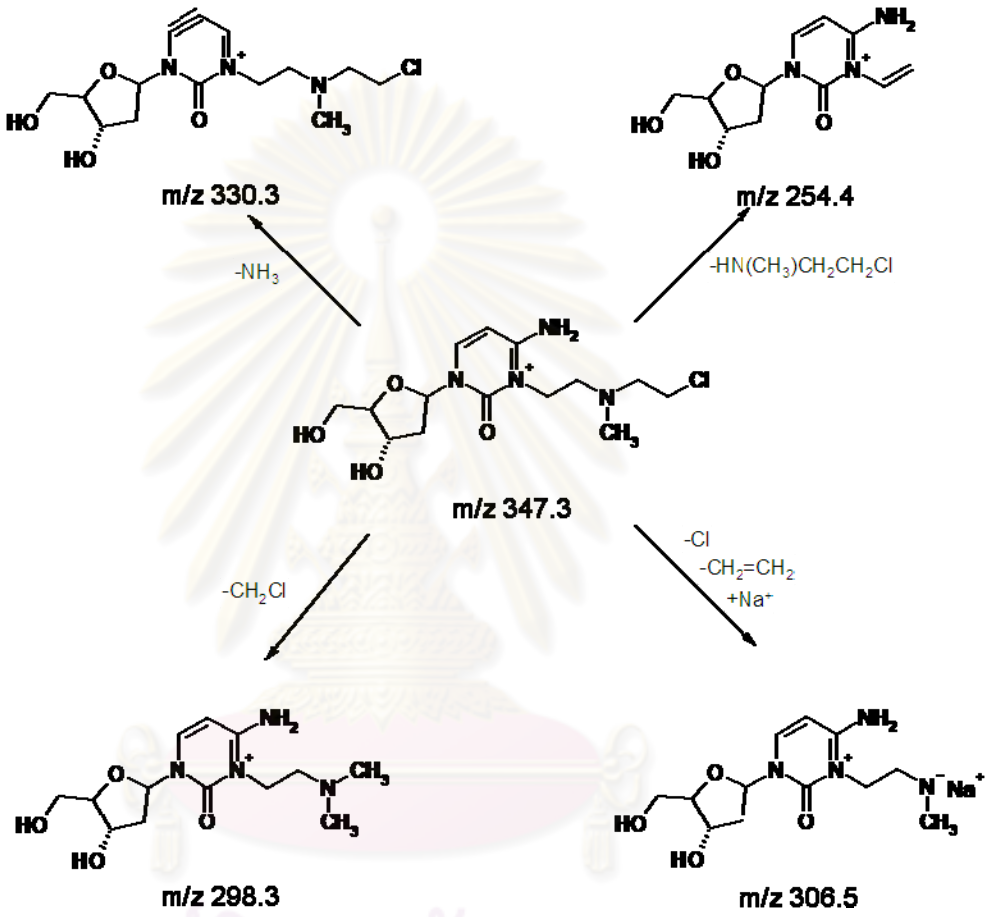


Figure 4.28 Scheme showing fragmentations of dC-mech-Cl

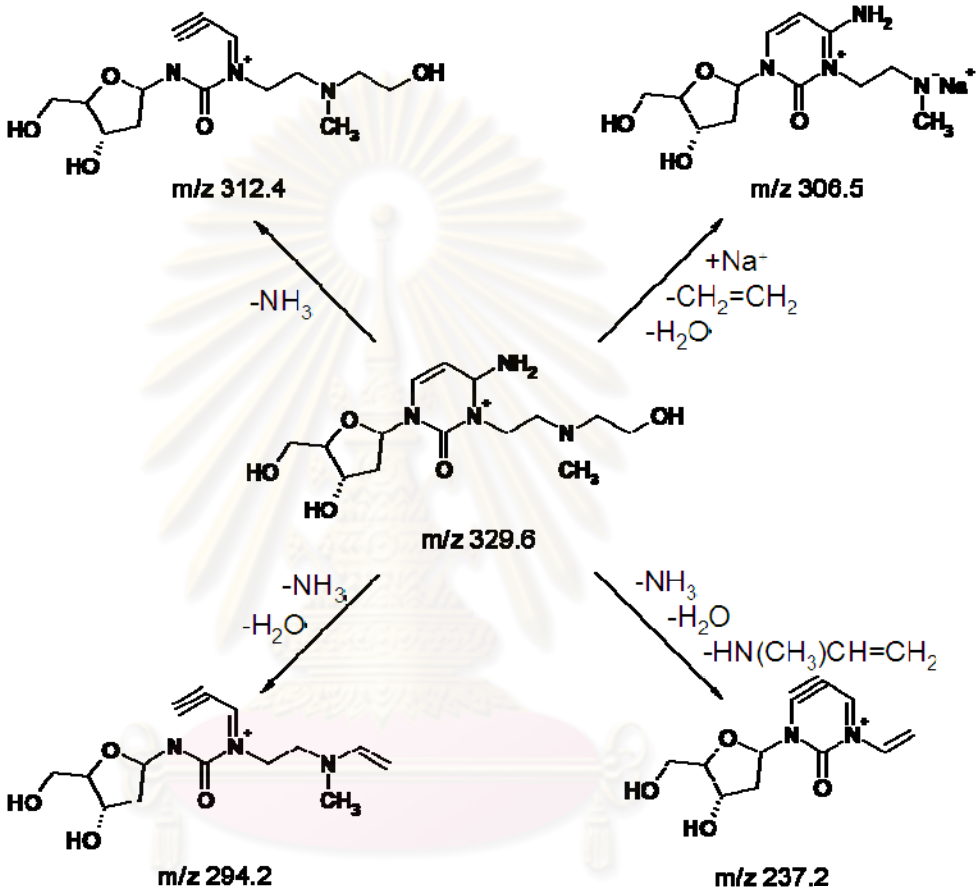


Figure 4.29 Scheme showing fragmentations of dC-mech-OH

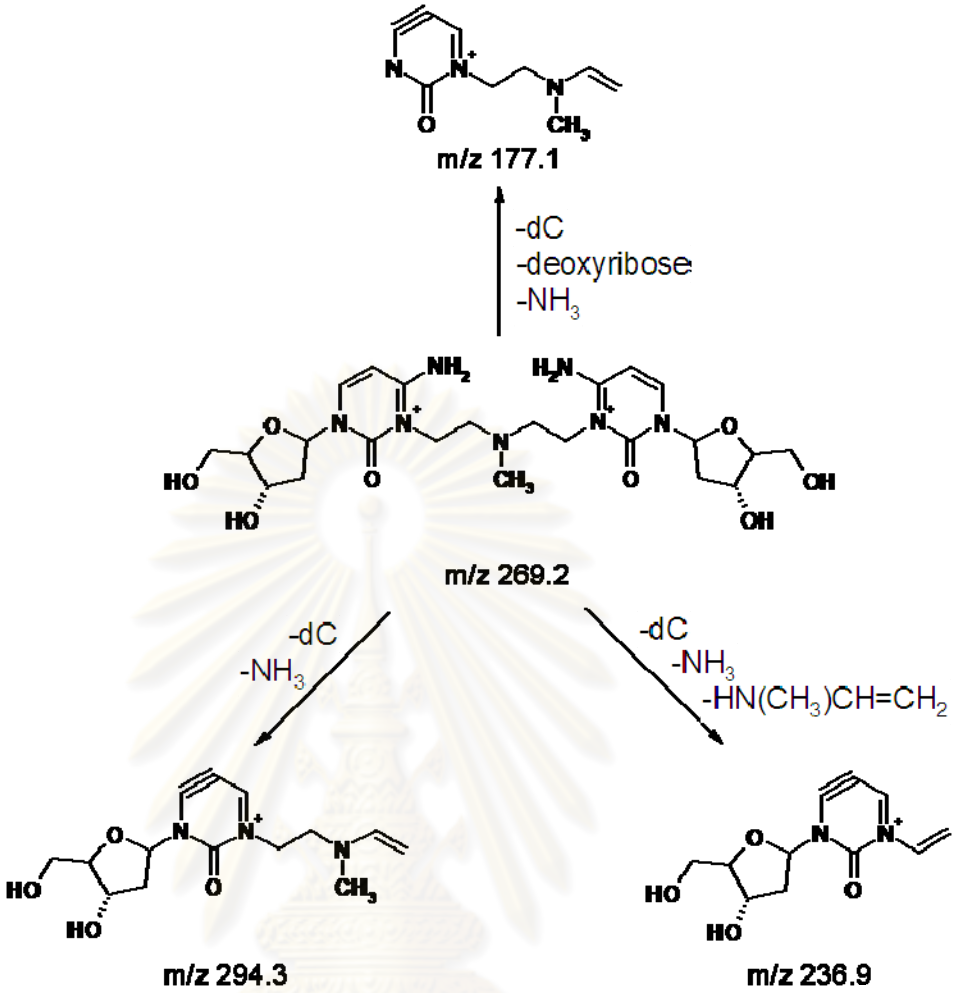


Figure 4.30 Scheme showing fragmentation of dC-mech-dC

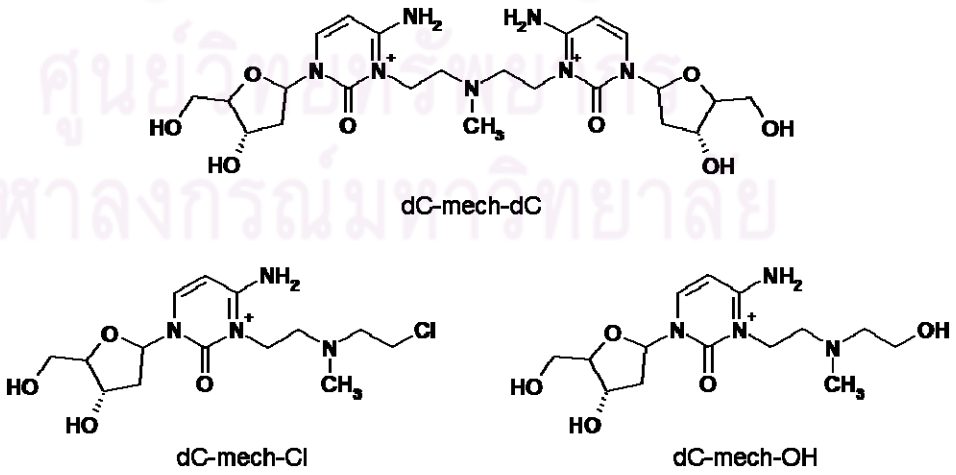


Figure 4.31 Molecular structures of dC-mech-dC, dC-mech-Cl and dC-mech-OH

CHAPTER V

CONCLUSION

Separation and identification of the mechlorethamine-DNA crosslink by HPLC are based on the relationship between the type and order of bases within the sequences and retention times of the DNA and the crosslink. By using a Biobasic-C4 column (4.6x250 mm, 5 μm) with a column temperature of 33^{o}C, the crosslinked DNA can be separated from unreacted DNA using two mobile phase gradient elutions of 5-15% ACN in 100 mM TEAA and 0.1 mM EDTA in 60 minutes at a flow rate of 1 ml/min with UV detection at 260 nm. The peak for crosslinked DNA had a retention time of about 40 minutes under these conditions. Separation of the crosslinked DNA using this approach allowed for collection of the purified crosslinked duplex for further characterization by MALDI-TOF-MS and ESI-MS/MS to establish the formation, the target bases, and molecular connectivity of the mechlorethamine-DNA crosslink at a C-C mismatch pair.

Enzymatic digestion of the mechlorethamine-DNA crosslink with a combination of SVPD and CIP at 37°C for 48 hours results in cleavage of the phosphodiester bond by SVPD and removal of the phosphate group by CIP. This enzymatic digestion yields monodeoxyribonucleosides and mechlorethamine-crosslinked products.

Identification and separation of the digested products by HPLC was achieved based on the relationship between hydrophobicity and retention time of each product. Using a Rainin Microsorb-C18 column (4.6x150 mm, 5 μm) maintained at a column temperature of 25°C, the digested products of mechlorethamine-DNA crosslink were separated with two mobile phase gradient elutions of 2-20% ACN in phosphate buffer (pH 5.8) containing 50 mM sodium phosphate in 20 minutes at a flow rate of 1 ml/min with UV detection at 260 nm. Eight significant peaks for digested products were obtained, five of which were identified as dC, incompletely digested poly-dA, dG, dT and dA with retention times of about 3.8, 7.8, 8.2, 9.3 and 11.2 minutes, respectively, based on comparison of standard monodeoxyribonucleosides and digested polydeoxyribonucleosides. The three unidentified peaks were tentatively identified as dC-mechlorethamine-Cl (dC-mech-Cl, MW = 347.82), dC-mechlorethamine-OH (dC-mech-OH, MW = 329.38) and dC-mechlorethamine-dC (dC-mech-dC, MW = 539.59).

The three digestion products were identified by analysis of the reaction mixture using mass spectrometry. In the mass spectra, the peaks corresponding to dC-mech-Cl, dC-mech-OH and dC-mech-dC gave molecular ions of m/z 347.3, 329.6 and 269.2, respectively. Daughter ions resulting from the neutral loss of NH_3 were observed, which indicates that the exocyclic-NH_2 group of cytosine (N^4) does not react with mechlorethamine. In addition, the most nucleophilic atoms of C-C mismatch pair are N^3. These results confirm that mechlorethamine forms an interstrand crosslink with a C-C mismatch in a DNA duplex via reaction with the N^3 atoms of the mismatch pair.

In conclusion, the molecular structure of the crosslink between mechlorethamine a DNA duplex containing a C-C mismatch pair was characterized by HPLC, enzymatic digestion and mass spectrometry. The method developed in this work could be applied to determine DNA crosslinks with other alkylating agents such as N-mustards and cisplatin and to compare the reactivity of different alkylating agents with DNA.

REFERENCES

Allawi, H. T., and SantaLucia, Jr, J. 1998. NMR Solution Structure of a DNA Dodecamer containing Single G•T Mismatches. Nucleic Acids Res. 26: 4925-4934.

Bagni, C., and Greenough, W. T. 2005. From mRNP Trafficking to Spine Dysmorphogenesis: The Roots of Fragile X Syndrome. Nat. Rev. Neurosci. 6: 376-387.

Bakhtiar, R., and Nelson, R. W. 2000. Electrospray Ionization and Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Emerging Technologies in Biomedical Sciences. Biochem. Pharmacol. 59: 891–905.

Balcome, S., Park, S., Quirk Dorr, D. R., Hafner, L., Phillips, L., and Tretyakova, N. 2004. Adedine-Containing DNA-DNA Cross-links of Antitumor Nitrogen Mustards. Chem. Res. Toxicol. 17: 950-962.

Carmical, J. R., Kowalczyki, A., Zou, Y., Van Houten, B., Nechevi, L. V., Harrisi, C. M., Harrisi, T. M., and Lloyd, R. S. 2000. Butadiene-induced Intrastrand DNA Cross-links: A Possible Role in Deletion Mutagenesis. J. Biochem. 275: 19482-19489.

Chen, X., Santhana Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. 1995. Hairpins are Formed by the Single DNA Strands of the Fragile X Triplet Repeats: Structure and Biological Implications. Proc. Natl. Acad. Sci. 92: 5199-5203.

Cummings, J., French, R. C. and Smyth, J. F. 1993. Application of High-Performance Liquid Chromatography for Recognition of Covalent Nucleic Acid Modification with Anticancer Drugs. J. Chromatogr. 618: 251-276.

Cummings, C. J., and Zoghbi, H. Y. 2000. Fourteen and Counting: Unraveling Trinucleotide Repeat Diseases. Hum. Mol. Genet. 9: 909–916.

Darlow, J. M., and Leach, D. R. F. 1998. Secondary Structures in d(CGG) and d(CCG) Repeat Tracts. J. Mol. Biol. 275: 3-16.[a]

Darlow, J. M., and Leach, D. R. F. 1998. Evidence for Two Preferred Hairpin Folding Patterns in d(CGG) •d(CCG) Repeat Tracts in vivo. J. Mol. Biol. 275: 17-23.[b]

Darnell, J. C., Johnson, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. 2001. Fragile X Mental Retardation Protein Targets G Quartet MRNAs Important for Neuronal Function. Cell 107: 489-499.

Darnell, J. C., Fraser, C. E., Mostovetsky, O., Stefani, G., Jones, T.A., Eddy, S. R., and Darnell, R. B. 2005. Kissing Complex RNAs Mediate Interaction between the Fragile-X Mental Retardation Protein KH2 Domain and Brain Polyribosomes. Genes Dev. 19: 903–918.

De Vries, B. B., Halley, D. J., Oostra, B. A., and Niermeijer, M. F. 1998. The Fragile X Syndrome. J. Med. Genet. 35: 579-589.

Di Prospero, N. A. and Fischbeck, K. H. 2005. Therapeutics Development for Triplet Repeat Expansion Diseases. Nat. Rev. Genet. 6: 756-765.

Dorr, D. Q., Murphy, K., and Tretyakova, N. 2005. Synthesis of DNA Oligonucleotides Containing Structurally Defined N^6-(2-hydroxy-3-buten-1-yl)-adenine Adducts of 3,4-epoxy-1-butene. Chem-Biol. Interact. 166: 104-111.

Farina, K. L., and Singer, R. H. 2002. The Nuclear Connection in RNA Transport and Localization. Trends Cell Biol. 12: 466–472.

Feng, Y., Gutekunst, C. A., Eberhart, D. E., Hong, Y., Warren, S. T., and Hersch, S. M. 1997. Fragile S Mental Retardation Protein: Nucleocytoplasmic Shuttling and Associatoin with Somatodendritic Ribosomes. J. Neurosci. 17: 1539-1547.

Fu, Y. H., Kuhl, D. P. A., Pizutti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G. Jr., Warren, S. T., Oostra, B. A., Nelson, D. L., and Caskey, C. T. 1991.Variation of the CGG Repeat at the Fragile X Site Results in Genetic Instability: Resolution of the Sherman Paradox. Cell 67: 1047–1058.

Gaskell, M., Kaur, B., Farmer P. B., and Singh, R. 2007. Detection of Phosphodiester Adducts Formed by the Reaction of Benzo[a]pyrene diol epoxide with 2´-Deoxynucleotide using Collision-Induced Dissociation Electrospray Ionization Tandem Mass Spectrometry. Nucleic Acids Res. 35: 5014-5027.

Gao, X., Huang, X., Smith, G.K., Zheng, M., and Liu, H. 1995. New Antiparallel Duplex Motif of DNA CCG Repeats That Is Stabilized by Extrahelical Bases Symmetrically Located in the Minor Groove. J. Am. Chem. Soc. 117: 8883-8884.

Gao, X., and Patel, D. J. 1987. NMR Studies of A•C Mismatches in DNA Dodecanucleotides at Acidic pH. J. Biochem. 262: 16973-16984.

Gilar, M., Fountain, K. J., Budman, Y., Neue, U. D., Yardley, K. R., Rainville, P. D., Russel II, R. J., and Gebler, J. C. 2002. Ion-Pair Reversed-Phase High-Performance Liquid Chromatography Analysis of Oligonucleotides: Retention Prediction. J.Chromatogr. A 958: 167-182.

Gupta R., Beck, J. L., Sheil, M. M., and Ralph, S. F. 2005. Identification of bifunctional GA and AG intrastrand crosslinks formed between cisplatin and DNA. J. Inorg Biochem. 99: 552-559.

Gut, I. G. 2004. DNA Analysis by MALDI-TOF Mass Spectrometry. Hum. Mutat. 23: 437-441.

Haas, J. G. 2007. History Tends to Repeat: FMR-1 Silencing in Fragile X Syndrome. Eukaryon 3: 41-46.

Hagihara, S., Kumasawa, H., Goto, Y., Hayashi, G., Kobori, A., Saito, I., and Nakatani, K. 2004. Detection of Guanine-Adenine Mismatches by Surface Plasmon Resonance Sensor Carrying Naphthyridine-Azaquinolone Hybrid on the Surface. Nucleic Acids Res. 32: 278-286.

Hagihara, M., and Nakatani, K. 2006. Inhibition of DNA Replication by a d(CAG) Binding Ligand. Nucleic Acid Symp. 50: 147-148.

Hartley, J. A., Souhami, R. L., and Berardini, M. D. 1993. Electrophoretic and Chromatographic Separation Methods Used to Reveal Interstrand Crosslinking of Nucleic Acids. J. Chromat. Biomed. Appl. 618: 277-288.

Hecker, K. L., Green, S. M., and Kobayashi, K. 2001. Analysis and Purification of Nucleic Acids by Ion-Pair Reversed-Phase High-Performance Liquid Chromatography. J. Biochem. Biophys. Methods. 46: 83-93.

Hovhannisyan, G. G., Haroutunyan, T. S., Arutyunyan, R. M. 2004. Evaluation of DNA Crosslinks Formation with UV-C Application by the Alkaline COMET-Assay. Exp. Oncol. 26: 240-242.

Kato, M., Saitoh, S., Kamei, A., Shiraishi, H., Ueda, Y., Akasaka, M., Tohyama, J., Akasaka, N., and Hayasaka, K. 2007. A Longer Polyalanine Expansion Mutation in the *ARX* Gene Causes Early Infantile Epileptic Encephalopathy with Suppression-Burst Pattern (Ohtahara Syndrome). Am. J. Hum. Genet. 81: 361-366.

Kaufmann, W. E., and Reiss, A. L. 1999. Molecular and Cellular Genetics of Fragile X Syndrome. Am. J. Med. Genet. 88: 11-24.

Kobori, A., Horie, S., Suda, H., Saito, I., and Nakatani, K. 2004. The SPR Sensor Detecting Cytosine-Cytosine Mismatches. J. Am. Chem. Soc. 126: 557-562.

Lacy, E. R., Cox, K. K., Wilson, W. D., and Lee, M. 2004. Recognition of T•G Mismatched Base Pairs in DNA by Stacked Imidazole-Containing Polyamides: Surface Plasmon Resonance and Circular Dichroism Studies. Nucleic Acids Res. 30: 1834-1841.

Lane, A. N., Jenkins, T. C., Brown, D. J. S., and Brown, T. 1991. NMR Determination of the Solution Conformation and Dynamics of the A•G Mismatch in the d[CGCAAATTGGCG]$_2$ dodecamer. Biochem. J. 279: 269-281.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, H. 1991. Androgen Receptor Gene Mutations in X-linked Spinal and Bulbar Muscular Atrophy. Nature 352: 77–79.

Lehninger, A. L., Nelson, D. I., and Cox, M. M. 1993. Principle of Biochemistry. 2nded. New York: Worth Publishers.

Li, Y., Zon, G., and Wilson, W. D. 1991. NMR and Molecular Modeling Evidence for a G•A Mismatch Base Pair in a Purine-Rich DNA Duplex. Proc. Natl. Acad. Sci. 88: 26-30.

Malina, J., Hofr, C., Maresca, L., Natile, G., and Brabec, V. 2000. DNA Interactions of Antitumor Cisplatin Analogs Containing Enantiomeric Amine Ligands. Biophys. J. 78: 2008–2021.

Mariappan, S. V., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Bradbury, E. M., and Gupta, G. 2000. Solution structures of the individual single strands of the fragile X DNA triplets (GCC)$_n$• (GGC)$_n$. Nucleic Acids Res. 24: 784-792.

Mariappan, S. V., Silks III, L.A., Bradbury, E. M., and Gupta, G. 1998. Fragile X DNA Triplet Repeats, (GCC)$_n$, Form Hairpins with Single Hydrogen-bonded Cytosine•Cytosine Mispairs at the CpG Sites: Isotope-edited Nuclear Magnetic Resonance Spectroscopy on (GCC)$_n$ with Selective ^{15}N4-labeled Cytosine Bases. J. Mol. Biol. 283: 111-120.

Mitas, M. 1997. Trinucleotide Repeats Associated with Human Disease. Nucleic Acids Res. 25: 2245-2253.

Mitas, M., Yu, A., Dill, J., and Haworth, I. S. 1995. The Trinucleotide Repeat Sequence d(CGG)15 Forms a Heat-Stable Hairpin Containing G^{syn}•G^{anti} Base Pairs**.** Biochemistry 34: 12803-12811.

Nadel, Y., Weisman-Shomer, P., and Fry, M. 1995. The Fragile X Syndrome Single Strand d(CGG)$_n$ Nucleotide Repeats Readily Fold Back to Form Unimolecular Hairpin Structures. J. Biol.Chem. 270: 28970–28977.

Nakatani, K., Kobori, A., Kumasawa, H., Gotoa, Y., and Saito, I. 2004. The Binding of Guanine–Guanine Mismatched DNA to Naphthyridine Dimer Immobilized Sensor Surfaces: Kinetic Aspects. Bioorgan. Med. Chem. 12: 3117–3123.

Nakatani, K., Sando, S., and Saito, I. 2001. Scanning of Guanine-Guanine Mismatches in DNA by Synthetic Ligands Using Surface Plasmon Resonance. Nat. Biotechnol. 19: 51-55.

Nakayabu, M., Miwa, S., Suzuki, M., Izuta, S., Sobue, G., and Yoshida, S. 1998. Mismatched Nucleotides may Facilitate Expansion Trinucleotide Repeats in Genetic Diseases. Nucleic Acids Res. 26: 1980–1984.

Nelson, D. I., and Cox, M. M. 2004. Lehninger Principle of Biochemistry. 4th ed. Hampshier: W. H. Freeman.

Noll, D. M., Noronha, A. M., and Miller, P. S. 2001. Synthesis and Characterization of DNA Duplexes Containing an N^4C-Ethyl-N^4C Interstrand Cross-Link. J. Am. Chem. Soc. 123: 3405-3411.

Orr, H. T. and Zoghbi, H. Y. 2007. Trinucleotide Repeat Disorders. Annu. Rev. Neurosci. 30: 575-621.

Park, S., Anderson, C., Loeber, R., Seetharaman, M., Jones, R., and Tretyakova, N. 2005. Interstrand and Intrastrand DNA-DNA Cross-Linking by 1,2,3,4-Diepoxybutane: Role of Stereochemistry. J. Am. Chem. Soc. 127: 14355-14365.

Petruska, J., Arnheim, N., and Goodman, M. F. 1996. Stability of Intrastrand Hairpin Structures Formed by the CAG/CTG Class of DNA Triplet Repeats Associated with Neurological Diseases. Nucleic Acids Res. 24: 1992-1998.

Peyret, N., Seneviratne, A., Allawi, H. T., and SantaLucia J. Jr. 1999. Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A•A, C•C, G•G and T•T Mismatches. Biochemistry 38: 3468-3477.

Rink, S. M., and Hopkins, P. B. 1995. A Mechlorethamine-Induced DNA Interstrand Cross-link Bends Duplex DNA. Biochemistry 34: 1439-1445.[a]

Rink, S. M., and Hopkins, P. B. 1995. Direct Evidence for DNA Intrastrand Cross-linking by the Nitrogen Mustard Mechlorethamine in Synthetic Oligonucleotides. Bioorganic Med. Chem. Lett. 5: 2845-2850.[b]

Rink, S. M., Solomon, M. S., Taylor, M. J., Raja, S. B., Mclaughlin, L. W., and Hopkins, P. B. 1993. Covalent Structure of a Nitrogen Mustard-Induced DNA Interstrand Cross-link, an N_7-to-N_7 Linkage of Deoxyguanosine Residues at the Duplex Sequence 5´-d(CNG). J. Am. Chem. Soc. 115: 2551-2557.

Rojsitthisak, P., Romero, R. M., and Haworth, I. S. 2001. Extrahelical Cytosine Bases in DNA Duplexes Containing $d[GCC]_n$•$d[GCC]_n$ Repeats: Detection by a Mechlorethamine Crosslinking Reaction. Nucleic Acids Res. 29: 4716-4723.

Romero, R. M., Mitas, M., and Haworth, I. S. 1999. Anomalous Cross-Linking by Mechlorethamine of DNA Duplexes Containing C-C Mismatch Pairs. Biochemistry 38: 3641-3648.

Romero, R. M., Rojsitthisak, P., and Haworth, I. S. 2001. DNA Interstrand Crosslink Formation by Mechlorethamine at a Cytosine-Cytosine Mismatch Pair: Kinetics and Sequence Dependence. Arch. Biochem. Biophys. 386: 143–153.

Sarma, M. H., Gupta, G., and Sarma, R. H. 1987. DNA Structure in Which an Adenine-Cytosine Mismatch Pair Forms an Integral Part of the Double Helix. Biochemistry 26: 7707-7715.

Singh, R., and Farmer, P. B. 2006. Liquid Chromatography-Electrospray Ionization-Mass Spectrometry: the Future of DNA Adduct Detection. Carcinogenesis 27: 178-196.

Siomi, H., Siomi, M. C., Nussbaum, R. L., and Dreyfuss, G. 1993. The Protein Product of Fragile X Gene, FMR1, Has Characteristics of an RNA-Binding Protein. Cell 74: 291-298.

Spanswick, V. J.,Craddock, C., Sekhar, M., Mahendra, P., Shankaranarayana, P., Hughes, R. G., Hochhauser, D., and Hartley, J. A. 2002. Repair of DNA Interstrand Crosslinks as a Mechanism of Clinical Resistance to Melphalan in Multiple Myeloma. Blood 100: 224-229.

Tost, J., and Gut, I. G. 2006. DNA Analysis by Mass Spectrometry-Past, Present and Future. J. Mass Spectrom. 41: 981-995.

Tretyakova, N., Livshits, A., Park, S., Bisht, B., and Goggin, M. 2007. Structural Elucidation of a Novel DNA-DNA Cross-Link of 1,2,3,4-Diepoxybutane. Chem. Res. Toxicol. 20: 284-289.

Urtizberea, J. A. 2004. Oculophayngeal Muscular Dystrophy. Orphanet Encyclopedia.

Van de Bor, V., and Davis, I. 2004. mRNA Localisation gets More Complex. Curr. Opin. Cell Biol. 16: 300–307.

Wang, Y., and Wang Y. 2003. Structure Elucidation of DNA Interstrand Cross-Link by a Combination with Mass Spectrometry. Anal. Chem. 75: 6306-6313.

Weese-Mayer, D. E., Marazita, M. L., and Berry-Kravis, E. M. 2004. Congenital Central Hypoventilation Syndrome. GeneReviews [Online]. Available from: http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=ondine [2009, May 15].

Williams, D. A., Lemke, T. A. 2002. Foye's Principles of Medicinal Chemistry. 5th ed. Philadelphia: Lippincott Williams.

Wilds, C. J., Booth, J. D., and Noronha, A. M. 2006. Synthesis of Oligonucleotides Containing an O^6-G-alkyl-O^6-G Interstrand Cross-link. Tetrahedron Lett. 47: 9125-9128.

Wilds, C. J., Noronha, A. M., Robidoux, S., and Miller, P. S. 2004. Mispair-Aligned N^3T-alkyl-N^3T Interstrand Cross-Linked DNA: Synthesis and Characterization of Duplexes with Interstrand Cross-Links of Variable Lengths. J. Am. Chem. Soc. 126: 2957-2965.

Wynne, P., Newton, C., Ledermann, J. A., Olaitan, A., Mould T. A., and Hartley, J. A. 2007. Enhanced Repair of DNA Interstrand Crosslinking in Ovarian Cancer Cells from Patients Following Treatment with Platinum-based Chemotherapy. Brit. J. Cancer. 97: 927-933.

Yu, A., Barron, M. D., Romero R. M., Christy, M., Gold, B., Dai, J., Gray, D. M., Haworth I. S., and Mitas, M. 1997. At Physiological pH, $d[CCG]_n$ Forms a Hairpin Containing Protonated Cytosines and a Distort Helix. Biochemistry 36: 3687-3699.

Zhang, Y., O'Connor, J. P., Siomi M. C., Srinivasan, S., Dutra, A. Nussbaum, R. L., and Dreyfuss, G. 1995. The Fragile X Mental Retardation Syndrome Protein Interacts with Novel Homologs FXR1 and FXR2. EMBO J. 14: 5358–66.

BIOGRAPHY

Mr. Nutthapon Jongaroonngamsang was born in Bangkok, Thailand on October 7th, 1982. He received his Bachelor of Science in Pharmacy with 1st class honors from Chulalongkorn University, Bangkok, Thailand in 2006. He has worked as a research scientist at Drug Quality Research and Development Laboratory Unit, Faculty of Pharmaceutical Sciences, Chulalongkorn University throughout his graduate study. In 2008, he was a research follow of ICS-UNIDO to conduct his research in the field of combinatorial chemistry and molecular design at CISI, University of Milan, Milan, Italy. From 2006 to 2009, he has four published papers, as follows.

1. Wichitnithad, W., Jongaroonngamsang, N., Pummangura, S., and Rojsitthisak, P. A Simple Isocratic HPLC Method for the Simultaneous Determination of Curcuminoids in Commercial Turmeric Extracts. Phytochem. Anal. 20 (2009): 314–319.
2. Lertsutthiwong, P., Noomun, K., Jongaroonngamsang, N., Rojsitthisak, P., and Nimmannit, U. Preparation of alginate nanocapsules containing turmeric oil. Carbohydrate Polymers 74 (2008): 209-214.
3. Rojsitthisak, P., Jongaroonngamsang, N., Thampanitchawong, B., Thanawattanawanich, P., and Wongtavatchai, J. Quality Control of Formulated Clove Oil Solution Aquanes®. Thai. J. Vet. Med. 37 (2007): 114.
4. Jongaroonngamsang, N., Wongjesda, T., Theeraroungchaisri, A., Tatong, W., and Rojsitthisak, P. Molecular Modeling of Renieramycin-DNA Complexes. Thai J. Pharm. Sci. 30 (suppl.) (2006): 36.