

The Selection of DNA Damage Targets In Vivo and In Vitro  
by Eneidyne Antitumor Antibiotics

by

Jinghai Xu

B. A. Biology and Chemistry  
Denison University, 1990

SUBMITTED TO THE DIVISION OF TOXICOLOGY IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Ph.D. IN TOXICOLOGY  
AT THE  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 1998

[June 1998]

© 1998 Massachusetts Institute of Technology. All rights reserved.

Signature of Author: \_\_\_\_\_  
Division of Toxicology  
February, 1998

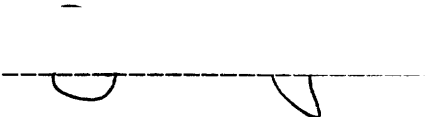
Certified by: \_\_\_\_\_  
Peter C. Dedon  
Associate Professor of Toxicology  
Thesis Supervisor

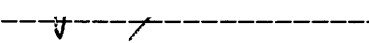
Accepted by: \_\_\_\_\_  
Peter C. Dedon  
Associate Professor of Toxicology  
Chairman, Committee on Graduate Students

FEB 27 1998

Science

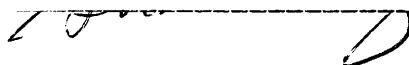
This doctoral thesis has been examined by a Committee of the Division of Toxicology as follows:

Professor John Essigmann  \_\_\_\_\_ Chairman

Professor Peter Dedon  \_\_\_\_\_ Supervisor

Professor William Thilly  \_\_\_\_\_

Professor Ram Sasisekharan  \_\_\_\_\_

Professor Gerald Wogan  \_\_\_\_\_

# The Selection of DNA Damage Targets In Vivo and In Vitro by Eneidiyne Antitumor Antibiotics

by  
Jinghai Xu

Submitted to the Division of Toxicology  
on February 12, 1998 in Partial Fulfillment of the  
Requirements for the Degree of  
Ph.D. in Toxicology

## ABSTRACT

Many environmental mutagens and chemotherapeutic agents exert their biological effects by damaging DNA. An important determinant of cellular response to DNA damage is the location and quantity of damage. Therefore, the goal of my thesis research is to elucidate those mechanisms that govern how genotoxins select their DNA targets in cells.

To this end, I have performed studies to identify how the cellular environment of DNA modulates damage produced by three enediayne antitumor antibiotics: esperamicin A1, esperamicin C and calicheamicin  $\gamma_1$ I. Using a modified ligation-mediated PCR (LMPCR) technique, DNA damage along the human phosphoglycerate kinase (PGK1) promoter was mapped. The results show that location and quantity of the DNA damage are affected by nucleosomes, transcription factors and DNA methylation. Nucleosome cores in the inactive PGK1 promoter suppressed DNA damage produced by esperamicin A1 (an intercalator) but not by esperamicin C and calicheamicin (both minor groove binders). Transcription factors in the active PGK1 promoter suppressed damage by esperamicin A1, but only minor groove binding transcription factors suppressed damage by esperamicin C and calicheamicin. Some major groove binding factors even enhanced damage caused by esperamicin C and calicheamicin. Enhancement of damage was also observed with all three enediaynes between transcription factor binding sites. In addition, cytosine methylation enhanced DNA damage made by esperamicins. These results are consistent with a model in which target accessibility, DNA dynamics, DNA bending and minor groove width are important factors affecting enediayne target selection *in vivo*. Finally, a technique was developed to isolate DNA damage fragments produced by nanomolar calicheamicin concentrations. The technique can potentially be used to map low levels of DNA damage in the entire genome.

This work extends our *in vitro* observations of enediayne target selection to the complex milieu of the whole cell. These results lay the foundation for understanding target selection mechanisms of other genotoxins *in vivo*.

Thesis Supervisor: Peter C. Dedon  
Title: Associate Professor of Toxicology

**Dedicated to  
my family and my teachers**



## ACKNOWLEDGMENTS

I am especially grateful to Professor Peter Dedon, whose intellect, support and understanding have guided me through the past five and half years. Your expertise in the field of DNA biochemistry has benefited me tremendously. Most importantly, you encouraged and challenged me to be an independent scientist and a critical thinker. I thank you for always being an encouraging critic and a great mentor.

I am particularly fortunate to have the opportunity to interact on a daily basis with some of the brightest people I have known. These are the ones who have helped me in difficult tasks, who provided advice whenever I needed it, and most importantly, made the long hours tolerable and even enjoyable. I hope our friendship will continue even after I graduate. To Bill, I am especially grateful for your extra effort in keeping everything running so smoothly and for providing invaluable advice for my manuscript and thesis. To Aaron, thank you for having always been there whenever I needed you. To Jiongru, you are the best colleague/roommate I could ever ask for. To Punam, thank you for being so understanding and supportive throughout these years; and a special thanks for performing the methylation studies on the [<sup>32</sup>P]-labeled PGK1 restriction fragment. And to Laura Trudel, John Zhuang, Ken, Qi, Elaine, Daniel, William, and many others, thank you for your help and countless lessons.

I would like to express my heart-felt gratitude to my thesis committee members for your advice and service: Drs. John Essigmann, Ram Sasisekharan, William Thilly and Gerald Wogan. To Dr. John Essigmann, thank you so much for chairing my committee in your sabbatical year. Your guidance and support in my research project and scientific career is most appreciated.

I would like to thank Debra Luchanin for her constant interest in my career development ever since my arrival at MIT. To Marci, thank you so much for your secretarial assistance.

I am grateful to Dr. Gerd Pfeifer for his support in acquiring the LMPCR technology and for providing an X-chromosome-enriched clone of the X86T2 cells. Further thanks are extended to Dr. Stanley Gartler for providing the original hamster-human hybrid cells, and to Dr. Steven Baits for subcloning the X86T2 and Y162-11C cells. I am also grateful to Dr. Judith Singer-Sam for providing the cloned PGK1 gene.

I am indebted to MIT and the National Institute of Health (NIH) for financial support during the course of this work. The funding of my thesis research was provided by the Whitaker Health Sciences Fund Fellowship from MIT, and research grants awarded to my advisor from NIH.

Finally, I am grateful to my wife, Li, for being a constant provider of love, patience, support, advice, sanctuary, and much more.

## TABLE OF CONTENTS

|  |    |
|--|----|
| ABSTRACT   | 3  |
| ACKNOWLEDGMENTS  | 5  |
| TABLE OF CONTENTS  | 7  |
| LIST OF ABBREVIATIONS  | 10 |
| LIST OF FIGURES  | 11 |
| LIST OF TABLE  | 13 |
| BIOGRAPHICAL NOTE  | 14 |
| PUBLICATIONS   | 15 |
| <br>   |    |
| CHAPTER 1. INTRODUCTION  | 16 |
| I. Cellular Environment of DNA   | 19 |
| A. Sequence Context in Naked DNA   | 19 |
| B. DNA Methylation   | 21 |
| C. The Nucleosome  | 22 |
| D. Higher Order Chromatin Structure in Transcriptionally Inactive<br>Genes                                 | 24 |
| E. Transcriptionally Active Chromatin  | 25 |
| F. The Human PGK1 As a Model Gene for Target Selection Studies   | 29 |
| II. Eneidyne As Model Genotoxins   | 31 |
| III. Earlier Studies of Eneidyne Target Selection <i>In Vitro</i>  | 34 |
| IV. Significance   | 36 |
| V. Figures   | 37 |
| <br>   |    |
| CHAPTER 2. DEVELOPMENT OF A MODIFIED LIGATION-MEDIATED PCR<br>STRATEGY TO MAP ENEDIYNE-SPECIFIC DNA DAMAGE | 52 |
| I. Introduction  | 53 |

|                             |    |
|-----------------------------|----|
| II. Materials and Methods   | 55 |
| III. Results and Discussion | 60 |
| IV. Figures                 | 63 |

### CHAPTER 3. THE ROLE OF NUCLEOSOMES IN ENEDIYNE TARGET

|                           |    |
|---------------------------|----|
| SELECTION                 | 76 |
| I. Introduction           | 77 |
| II. Materials and Methods | 79 |
| III. Results              | 82 |
| IV. Discussion            | 85 |
| V. Figures                | 89 |

### CHAPTER 4. THE ROLE OF TRANSCRIPTION FACTORS IN ENEDIYNE

|                           |     |
|---------------------------|-----|
| TARGET SELECTION          | 98  |
| I. Introduction           | 99  |
| II. Materials and Methods | 101 |
| III. Results              | 103 |
| IV. Discussion            | 105 |
| V. Figures                | 113 |

### CHAPTER 5. THE ROLE OF DNA METHYLATION IN ENEDIYNE TARGET

|                           |     |
|---------------------------|-----|
| SELECTION                 | 122 |
| I. Introduction           | 123 |
| II. Materials and Methods | 126 |
| III. Results              | 128 |
| IV. Discussion            | 130 |
| V. Figures                | 134 |

|   |     |
|---|-----|
| CHAPTER 6. DEVELOPMENT OF TECHNIQUES TO MAP LOW LEVELS OF<br>DNA BREAKS IN WHOLE CELLS----- | 141 |
| I. Introduction-----  | 142 |
| II. Materials and Methods-----  | 144 |
| III. Results-----   | 148 |
| IV. Discussion-----   | 150 |
| V. Figures-----   | 157 |
| <br>  |     |
| APPENDIX: Mathematical Derivations-----   | 168 |
| BIBLIOGRAPHY-----   | 170 |

## LIST OF ABBREVIATIONS

(in alphabetical order)

|               |   |
|---------------|---|
| ATP           | Adenine triphosphate                        |
| bp            | Base pairs                                  |
| C/EBP         | CCAAT enhancer binding protein              |
| Calicheamicin | calicheamicin $\gamma_1$ I                  |
| DMS           | Dimethyl sulfate                            |
| DNA           | Deoxyribonucleic acid                       |
| DS            | Double-strand                               |
| GSH           | Glutathione                                 |
| LMPCR         | Ligation-mediated polymerase chain reaction |
| MNase         | Micrococcal nuclease                        |
| NF1           | Nuclear factor 1                            |
| NT            | Nick translation                            |
| PBS           | Phosphate-buffered saline                   |
| PCR           | Polymerase chain reaction                   |
| PGK-1         | Phosphoglycerate kinase gene                |
| SS            | Single-strand                               |
| TDT           | Terminal deoxynucleotidyl transferase       |
| X cell        | X86T2 cell, contains Xi                     |
| Xa            | Active X chromosome                         |
| Xi            | Inactive X chromosome                       |
| Y cell        | Y162-11C cell, contains Xa                  |

## LIST OF FIGURES

|            |     |
|------------|-----|
| Figure 1.1 | 38  |
| Figure 1.2 | 40  |
| Figure 1.3 | 42  |
| Figure 1.4 | 44  |
| Figure 1.5 | 46  |
| Figure 1.6 | 48  |
| Figure 1.7 | 50  |
| Figure 2.1 | 64  |
| Figure 2.2 | 66  |
| Figure 2.3 | 68  |
| Figure 2.4 | 70  |
| Figure 2.5 | 72  |
| Figure 3.1 | 90  |
| Figure 3.2 | 92  |
| Figure 3.3 | 94  |
| Figure 3.4 | 96  |
| Figure 4.1 | 114 |
| Figure 4.2 | 116 |
| Figure 4.3 | 118 |
| Figure 4.4 | 120 |
| Figure 5.1 | 135 |

|            |     |
|------------|-----|
| Figure 5.2 | 137 |
| Figure 5.3 | 139 |
| Figure 6.1 | 158 |
| Figure 6.2 | 160 |
| Figure 6.3 | 162 |
| Figure 6.4 | 164 |
| Figure 6.5 | 166 |



## LIST OF TABLE

Table 6.1. Buffer effects on DNA binding to Dynabeads -----156

## BIOGRAPHICAL NOTE

Jinghai Xu was born on November 6, 1969, in Guangzhou (Canton), P. R. China. He was admitted without the College Entrance Examination to Fudan University in 1987, where he majored in Biophysics. He transferred to Denison University in 1990, where he completed an honor thesis under Prof. Thomas Evans (Denison) and Dr. Michelle Buchanan (Oak Ridge National Laboratory), and was awarded BA degrees in Biology and Chemistry in 1992. He pursued a Ph.D. in Toxicology under the guidance of Prof. Peter Dedon at the Massachusetts Institute of Technology, where he investigated the selection of DNA damage targets *in vivo* and *in vitro* by enediyne antitumor antibiotics. He defended his Ph. D. Thesis on February 12, 1998 and will soon join a genomics firm, CuraGen, as a research scientist.

Jinghai now lives with his beloved wife, Li, in Stonington, Connecticut.

## PUBLICATIONS

- Xu, J. and Dedon, P. C. (1997) "DNA Damage Produced by Eneidyne in the Human Phosphoglycerate Kinase Gene *in vivo*: Esperamicin A1 as a Nucleosome Footprinting Agent," *Biochemistry*, in press.
- Mathur, P., Xu, J. and Dedon, P.C. (1997) "Cytosine Methylation Enhances DNA Damage Produced by Groove Binding and Intercalating Eneidyne: Studies with Esperamicins A1 and C," *Biochemistry*, Vol. 36, 14868-14873.
- Xu, J. and Dedon, P. C. (1997) "Technologies for Detection of DNA Damage and Mutations," *Photochemistry and Photobiology*, Vol. 65, 603-604.
- Dedon, P. C., Salzberg, A. S. and Xu, J. (1993) "Exclusive Production of Bistranded DNA Damage by Calicheamicin," *Biochemistry*, Vol. 32, 3617-3622.

## CONFERENCE PUBLICATIONS & PRESENTATIONS

- Xu, J. and Dedon, P. C. (1997) "Eneidyne-induced DNA Damage Is Affected by Chromatin Structures and Transcriptional Activity in Whole Cells," *Proceedings of the American Association for Cancer Research*, Vol. 38, 5.
- Mathur, P., Xu, J. and Dedon, P.C. (1997) "The Effect of DNA Methylation on DNA Damage Caused by Eneidyne Anti-cancer Drug," *Proceedings of the American Association for Cancer Research*, Vol. 38, 228.
- Xu, J. and Dedon, P. C. (1995) "A General Method for Isolating Oxidative DNA Damage From Cells: Application to Calicheamicin," *Proceedings of the American Association for Cancer Research*, Vol. 36, 549.
- Yu, L.; Xu, J. and Dedon, P. C. (1993) "Chromatin Structure Modulates the DNA Damage Produced by Eneidyne Antitumor Agents," *Proceedings of the American Association for Cancer Research*, Vol. 34, 350.
- Xu, J. and Dedon, P. C. (1994) "Development of methods to isolate and clone cellular DNA targets of anticancer drugs," Gordon Research Conference: *Chemotherapy of Experimental and Clinical Cancer*, Colby Sawyer College.

## **CHAPTER 1**

### **INTRODUCTION**

The integrity of DNA is critical not only for the viability but also for the functional activity of all organisms. However, cellular DNA is constantly subjected to a variety of potentially harmful insults, from both endogenous and exogenous agents (reviewed in [1]). The endogenous sources include spontaneous depurination and cytosine deamination, as well as reactions with oxidants and alkylating agents generated during cell metabolism [2]. The exogenous sources include chemicals (such as aflatoxin B1) and physical agents (such as UV light and  $\gamma$ -radiation). Damage to the cellular genome has also been used as a double-edged sword for the treatment of cancer, since destruction of cancer cells is an intended outcome in cancer chemotherapy. The DNA damage caused by drugs and toxins, if unrepaired, can lead to cell death or DNA mutation. The premise for my thesis research is that the location and quantity, *i.e.*, the target selection of the damage, plays an important role in determining the cellular response to DNA damage. Therefore, the goal of my thesis research is to elucidate those mechanisms that govern how genotoxins select their DNA targets in cells.

To this end, I have performed studies to identify how the cellular environment of DNA modulates the distribution of DNA damage produced by genotoxins. Using enediyne antitumor antibiotics as model genotoxins, I focused on the following aspects of the cellular environment of DNA: nucleosomes, transcription factors and DNA methylation. In the following chapters, I will first provide an overview of the subject matter (Chapter 1) and describe a technique to map enediyne-induced DNA damage at single-nucleotide resolution in a single-copy gene in mammalian cells (Chapter 2). I will then examine these three aspects of the cellular environment of DNA in greater detail (Chapters 3-5). Finally, I will move beyond a single-copy gene

and describe a genome-based technique that can potentially be used to map low levels of DNA damage in the genome (Chapter 6).

## I. Cellular Environment of DNA

DNA in mammalian nuclei is packaged along with various proteins into the compact mass of chromatin [3, 4]. For reasons of simplicity, most genotoxin-DNA interactions have been studied with purified (protein-free) DNA *in vitro*. However, chromatin is the true target for genotoxins and the packaging of DNA as chromatin may affect the target selection by genotoxins. One of the most important goals of my thesis research has been to understand how genotoxins select their targets in cells compared to naked DNA.

Many factors of DNA structure, dynamics, and accessibility in mammalian nuclei may interact collectively with genotoxin structure to determine which targets are selected *in vivo* (Figure 1.1). These factors include sequence selectivity, methylation modification, binding of transcription factors and other nucleoproteins, nucleosome structure, and high order chromatin folding (Figure 1.1). In the following introduction, I will discuss these factors in order of increasing organizational complexity, from naked DNA to nuclear architecture.

### A. Sequence Context in Naked DNA

Duplex DNA exemplifies the union of similarity and diversity in life. While the sugar-phosphate backbone of DNA remains chemically conserved throughout evolution, differences in base sequence are fundamental to the vast diversity in all life forms. Sequence context and its related structure and dynamics in DNA form the bases for sequence-specific interactions with proteins as well as sequence-dependent interactions with genotoxins.

A genotoxin can interact with duplex DNA in two primary ways that are significantly different (Figure 1.1):

- 1) groove-binding interactions which involve direct contact of the bound molecule with the edges of base-pairs in either of the (major or minor) grooves of DNA; and
- 2) intercalation of planar or approximately planar aromatic ring systems between base-pairs.

Regarding the former, the major and minor grooves differ significantly in electrostatic potential, hydrogen-bonding characteristics, steric effects, and hydration [5]. Many proteins exhibit binding specificity primarily through major groove interactions while small genotoxins in general prefer the minor groove. Any specificity which arises in the binding comes from contacts between the bound molecule and the edges of the base pairs on the “floor” of the groove. Intercalation results from rotation about torsional bonds in the DNA backbone, separation of base-pairs, and insertion of a planar or near planar molecule in between base-pairs. DNA is unwound at the site of an intercalation complex and the normal approximately  $36^\circ$  rotation of one base-pair with respect to the next is decreased as a result of intercalation [5].

In comparing groove-binding and intercalating genotoxins, it is clear that the groove-binders, as a class, display significantly greater binding selectivity than intercalators. Intercalation cavities created at A:T or at G:C base-pairs are quite similar in their potential for interaction with planar aromatic ring systems. Electrostatic, van der Waals, hydrophobic, etc. contributions to binding are similar for the two sites. On the other hand, groove-binding molecules can contact more base-pairs as they lie along the groove in a DNA helix and thus have more selectivity in terms of sequence context [5].



Using enediyne antitumor antibiotics as model genotoxins, I have the opportunity to compare and contrast the effects of these two types of DNA interactions on genotoxin target selection (see section II of this chapter).

## B. DNA Methylation

In mammalian cells, DNA is covalently modified by attachment of methyl groups to the C5-position of the cytosines in CpG dinucleotides. Cytosine methylation has significant biological functions in mammalian development, including genomic imprinting and X-chromosome inactivation (reviewed in [6]). Abnormalities in the distribution of methylated sites on DNA may contribute to the pathogenesis of certain human diseases, such as fragile X syndrome and cancer [6].

Fundamental to these biological functions is the ability of some methylated CpG repeats to suppress transcription [7]. Two mechanisms may underlie such gene silencing phenomena: site-specific methylation may decrease the binding of essential transcription activators, or enhance the binding of transcription inhibitors. Regarding the first hypothesis, methylation at a CpG site centrally located within the recognition sequence of a HeLa cell transcription factor strongly inhibited both the factor binding and gene transcription (e.g., [8]). This hypothesis also forms the basis for the methylation interference assay (e.g., [9]), although not all transcription factors are affected by DNA methylation [10]. Meanwhile, the discoveries of methyl-CpG-binding proteins, MeCP1 [11] and MeCP2 [12], their inhibitory effects on transcription [13], and their location in heterochromatin [14] all strongly support the second mechanism *in vivo*.

The observation that methylated DNA alters the binding of proteins essential in transcription regulation underscores the changes in DNA

structure that occur upon CpG-methylation. First and most obvious, the C5-methyl groups protrude into the major groove, where most regulatory proteins contact DNA. Secondly, cytosine methylation causes many more subtle changes in DNA structure and dynamics that include helical unwinding [15], increased base stacking and helical stability [16], reduction in major groove charge density near the methyl group [17] and, in certain sequence contexts, modulation of DNA bending [18-20]. Such structural changes may be the bases of altered protein interactions with methylated DNA.

Similar to the relationship between proteins and methylated-CpG sequences, the interactions of genotoxins with methylated DNA may also be altered. Both benzo[a]pyrene diol epoxide [21] and mitomycin C [22] show enhanced reactivity with methylated sequences, while damage produced by bleomycin [23] and N-methyl-N-nitrosourea [24] is inhibited by cytosine methylation.

I have examined the effect of DNA methylation on enediyne-mediated DNA damage in Chapter 5, with an emphasis on the relationship between genotoxin structure and methylation effects.

### **C. The Nucleosome**

Human cells contain ~1 m of DNA packaged into a nucleus with a diameter of 5-10  $\mu\text{m}$ . To accomplish this  $10^5$ -fold reduction in length while maintaining functionality, nuclear DNA is organized into a hierarchy of chromatin structures with over 100 different proteins. The main constituents of nucleoproteins in eukaryotes are a class of highly basic proteins known as histones, which consist of five distinct subtypes: H4, H3, H2A, H2B and H1. In addition to histones, a heterogeneous group of proteins with high cell-type

specificity is present in the nucleus. These proteins, which are grouped simply as nonhistone proteins, consist of several hundred different proteins, most of which are present in trace amounts [3].

The highest level of chromatin structure divides the genome into two distinct forms: heterochromatin, containing transcriptionally silent DNA, and euchromatin, which represents transcriptionally competent DNA [3]. Not all genes are packaged to the same level of compaction inside eukaryotic nuclei. Transcriptionally inactive genes in heterochromatin are more compacted than active genes. The latter are characterized by a more open chromatin structure and the presence of a variety of nonhistone proteins essential for the regulation and maintenance of transcription.

At the base of this organization is the nucleosome, a structure common to both hetero- and euchromatin [3, 4]. The canonical nucleosome consists of two regions: core and linker. The core is composed of 146 base pairs (bp) of DNA wrapped twice into a left-handed superhelix around an octamer of histone proteins consisting of pairs of histones H2A, H2B, H3 and H4. The structure of a nucleosome core particle, crystallized under physiological relevant ionic conditions, was recently solved by X-ray crystallography at 2.8 Å resolution [25], a significant improvement over previous studies [26]. One of the most important findings to emerge from this three-dimensional analysis was that the DNA does not take a perfectly regular path in winding about the histone octamer. Instead, the double helix is bent fairly sharply at several specific locations. The histones interact with the phosphodiester backbone on the inner face of the DNA superhelix, with the 146 bp of the core DNA contacting the histone octamer in the following order: H2A, H2B, H4, H3, H3, H4, H2B, and H2A. Thus, a nucleosome has a dyad axis of pseudo-symmetry with two sharp bends flanking an S-shaped jog of the DNA at the dyad axis.

The bends occur adjacent to points of substantial contact between the DNA and histones H3 and H4. At these sites, the regular B helix conformation is distorted over several neighboring base pairs rather than being kinked abruptly [25]. Another finding was that DNA in nucleosome core is overwound compared to protein-free DNA. The helical repeat of DNA in nucleosome core is 10.2 bp/turn compared to 10.5 bp/turn in the canonical B-form DNA in solution [27].

Depend on the mode of interaction with duplex DNA, various genotoxins may recognize nucleosomal DNA differently. DNA minor groove binders can bind to DNA in nucleosome core at sites where the minor groove faces away from the core histones, while intercalator binding is restricted to the linker regions [28]. The reduced binding of intercalators to nucleosome core DNA is most likely caused by constrained DNA dynamics due to histone-DNA contacts [29-32].

I have examined the effect of nucleosome structure on enediyne-mediated DNA damage in a model gene *in vivo* (Chapter 3), with an emphasis on the relationship between genotoxin structure and DNA conformation in nucleosomes.

#### **D. Higher Order Chromatin Structure in Transcriptionally Inactive Genes**

In transcriptionally inactive DNA in interphase nuclei, the linear array of nucleosomes is further folded into high order chromatin structures. According to the currently accepted models, a string of nucleosomes is folded into a solenoid fiber containing six nucleosomes per turn and a diameter of 300 Å. This superhelical structure is stabilized by histone H1, which contacts linker DNA [3].

According to the loop model of genomic organization,  $10^4 - 10^5$  bp loops of solenoidal chromatin in interphase nuclei have their bases attached to a nucleoprotein/RNA scaffold called the nuclear matrix [33]. During mitosis, the matrix takes the form of a metaphase scaffold with chromatin loops arranged spirally around the scaffold to form the arms of chromosomes [34]. This provides the highest level of compaction of DNA.

The structure of a canonical nucleosome is well understood because of the high resolution crystallographic studies [25]. However, the exact way that nucleosomes are folded into higher order chromatin structure is not as clear. The solenoidal structure in the 300 Å fiber is the most accepted model, but other models do exist [3]. The exact mechanism by which the 300 Å solenoid fiber forms chromatin loops is even less certain [35].

As in the nucleosome structure, higher order chromatin compaction is expected to affect interactions with genotoxins. However, currently there is no good method to quantitatively map DNA damage in a region beyond the size of  $10^4$  bp. Depending on the level of selectivity of a genotoxin, mapping DNA damage in  $10^9$  bp of the human genome is a daunting task. However, in Chapter 6, I will present a technique that allows such issues to be addressed at physiologically-relevant levels of genotoxin-induced DNA damage.

## **E. Transcriptionally Active Chromatin**

The hierarchical organization of transcriptionally inactive chromatin presents a problem for proteins involved in DNA metabolism. During DNA transcription, interphase chromatin must undergo local decondensation to allow the access of transcription factors and the passage of transcription machinery. Therefore, transcriptionally active chromatin has a more open structure than its inactive counterpart. So far, four major differences between

active and inactive chromatin have been reported: (1) acetylation of histones, (2) depletion of H1, (3) chromatin remodeling factors, and (4) transcription factor binding. I will address each of them briefly.

### 1. Acetylation of histones

Histones in transcriptionally active chromatin are hyperacetylated compared to inactive chromatin [3, 4, 36]. The acetylation of the positively charged  $\epsilon$ -amino groups of lysines has been proposed to neutralize electrostatic interactions between the histones and the phosphodiester backbone, and thus reduce contacts between histones and DNA [35, 37].

Recently, histone acetylation was found to be a prerequisite for transcription rather than a downstream effect of transcription. Hebbes *et al.* demonstrated that hyperacetylation was associated with genes poised for transcription rather than those that are actively transcribing [38]. In addition, the discovery that acetyltransferases form complexes with certain transcription factors provides a mechanism by which the appropriate region of chromatin (*i.e.*, transcription factor binding region) can be targeted for "remodeling" [39]

### 2. Depletion of histone H1

Many studies have shown a deficiency of linker histones in transcriptionally active regions of chromatin (reviewed in [40]). The conclusion from these studies is that linker histones are present on transcribed or transcribable genes to a much reduced extent, in some cases by as much as 50%. Several other studies indicate an inverse relationship between transcriptional activity and linker histone content [40].

The linker histone H1 is important for the formation of the condensed 300 Å fiber [3]. Thus, the partial depletion of H1 in active chromatin may facilitate the formation or maintenance of active chromatin.

### 3. Action of chromatin remodeling factors

Nucleosome remodeling triggered by a primary transcription factor is a prerequisite in many promoters for the binding of additional factors and the assembly of transcription machinery (reviewed in [41]). However, in many systems additional protein factors are required that utilize energy from adenine triphosphate (ATP) to remodel nucleosomes.

One example is the SWI/SNF complex, a general gene activator complex that is required for the transcriptional induction of many genes in yeast. It has a DNA-stimulated ATPase activity and is capable of relieving nucleosome repression in an ATP-dependent manner [42]. Recently, the SWI/SNF complex has been shown to be associated with the yeast RNA polymerase II holoenzyme [43], providing strong support for its role as a chromatin remodeling factor in either transcription initiation or elongation.

Another protein involved in nucleosome remodeling, the nucleosome remodeling factor or NURF, which is distinct from the SWI/SNF complex, is found in *Drosophila*. NURF is also capable of perturbing nucleosome structure in the presence of ATP [44].

### 4. Transcription factor binding

A prerequisite for transcriptional activation is the binding of transcription factors to their recognition sequences (reviewed in [45]). The accessibility of the recognition sequences for transcription factors is restricted in nucleosomes [3]. Depending on the mode of interaction between transcription factors and their cognate DNA, some transcription factors can recognize their binding sites within a nucleosome while others require a nucleosome-free region to bind [46]. In the mouse mammary tumor virus (MMTV) promoter, the glucocorticoid receptor (GR) is capable of binding to the glucocorticoid response elements (GREs) located at the nucleosome dyad

facing away from the histone octamer. In contrast to GR, nuclear factor 1 (NF1) cannot bind to its recognition sequence within the core DNA of nucleosomes (reviewed in [41]). The exclusion of NF1 from nucleosomes is probably due to the binding mode of NF1 to its cognate DNA. NF1 requires the whole circumference of the DNA for binding [47]. When DNA is packaged into nucleosomes, the histone-DNA interaction occludes one side of the DNA path on the histone surface, thereby causing steric hindrance for NF1 binding [47]. Therefore, the model for glucocorticoid-induced gene activation is that GR serves as the primary transcription factor capable of binding to its target sequences within the nucleosome cores. The binding of GR triggers nucleosome remodeling and allows the subsequent binding of NF1 [41].

Although nucleosomes inhibit the binding of certain transcription factors, not all promoter regions are entirely nucleosome-free. In fact, there are several examples of nucleosome-mediated folding of regulatory DNA, which results in a conformation that permits efficient interaction of the transcription machinery [48].

The exact mechanism by which the above four characters of active chromatin cooperate to allow or facilitate transcription has not yet been completely defined. However, one of the more likely scenarios is as follows. The binding of a primary transcription factor to its target sequence triggers nucleosome remodeling. Associated with this process is the acetylation of histones and loss of histone H1 from the nucleosome linker. The remodeled nucleosome then allows secondary factors to associate with their recognition sequences and to stimulate the assembly of the transcription initiation complex, which subsequently leads to transcription (reviewed in [41, 49]).



In summary, the highly compact chromatin structure found in inactive gene is perturbed in active genes. Active chromatin is characterized by stretches of open, nucleosome-free regions bound by transcription factors. Since the binding of transcription factors is so fundamental to chromatin structure in active genes and it represents the next level of complexity beyond the DNA helix, its effects on genotoxin target selection will be addressed in my thesis study (Chapter 4).

#### **F. The Human PGK1 As a Model Gene for Target Selection Studies**

The X-linked phosphoglycerate kinase gene (PGK1) provides an excellent opportunity to study the role of nucleosome structure, transcription factor binding and DNA methylation on the selection of DNA targets by genotoxins.

The PGK1 gene is a highly conserved housekeeping gene that codes for phosphoglycerate kinase, an essential enzyme in glycolysis that catalyzes the reversible conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate, generating one molecule of ATP [50, 51]. In its inactive state, the human PGK1 gene has two nucleosomes positioned in the promoter region, and these nucleosomes are replaced by transcription factors in the transcriptionally active copy of the gene (Figure 1.2). With regard to cytosine methylation, the cytosines in the CpG dinucleotides are completely methylated in the promoter region of the inactive PGK1 but unmethylated in the active PGK1. Finally, the X-linked state of the PGK1 gene permits study of both the inactive and active forms of the gene in human female cells. In the following studies, I have made use of the human-hamster hybrid cell lines containing either a single copy of an active human X chromosome (Y162-11C cells) or an inactive X chromosome (X86T2 cells) [52]. The human X

chromosomes in these hybrid cells have been shown to maintain activity states, methylation patterns, and protein-DNA footprints (*i.e.*, the experimental indicators of *in vivo* chromatin structure) very similar to normal human lymphocytes [52, 53]. Thus, enediyne-mediated DNA damage can be studied as a function of chromatin structure and DNA methylation in the PGK1 model.

## II. Eneidyne As Model Genotoxins

My thesis research has focused on the enediyne family of antitumor antibiotics, a structurally diverse family of DNA-cleaving molecules that share a common mechanism for damaging DNA. The structural diversity has allowed me to define the relationship between drug structure and genomic target selection.

Eneidyne are naturally occurring antitumor antibiotics isolated from several eubacterial species (reviewed in [54]). The name of the family derives from the common ene-diyne moiety in the core of these molecules that can form carbon-centered radicals and abstract hydrogens from the DNA backbone (Figure 1.3). Among the most studied members of the enediyne family are calicheamicin [55], esperamicin [56, 57], neocarzinostatin [58, 59], C-1027 [60], dynemicin [61], kedarcidin [62, 63], and maduropeptin [64] (for a review, see [54, 65]). Eneidyne are extremely potent cytotoxins that are lethal to tumor cells at  $10^{-9}$  to  $10^{-15}$  M [66]. While this potency is credited to the ability of eneidyne to damage DNA, both from the *in vitro* DNA cleavage studies (reviewed in [54, 65]) and the existence of DNA repair-deficient cell lines that are hypersensitive to the compounds (e.g., [67]), relatively little is known about the nature of enediyne DNA target(s) *in vivo*, or the roles of chromatin structure and DNA methylation in the target selection process. These questions form the basis of my thesis research.

Eneidyne bind to DNA with high affinity and produce high levels of double-stranded (DS) DNA damage by a common mechanism involving the enediyne core structure. The affinity between eneidyne and DNA has been estimated to range over  $10^{-6}$ - $10^{-8}$  M, depending on the sequences of DNA used in the studies [58, 68]. The DNA damage mechanism of the enediyne core in calicheamicin and esperamicin is shown in Figure 1.4. Initiated by a

nucleophilic attack at the methyltrisulfide trigger by thiols (presumably glutathione *in vivo*), the enediyne core undergoes a Bergman-Masamune rearrangement to form a diradical intermediate [69]. When the drug binds in the minor groove of DNA, the diradical abstracts hydrogen atoms from deoxyribose of both strands. Deoxyribose lesions then undergo oxygen-dependent reactions, yielding a spectrum of degradation products unique to each abstracted hydrogen. The final degradation products include strand breaks with 5'- and 3'-phosphates, 3'-phosphoglycerate, and modified abasic sites (Figure 1.5) [58, 65]. Calicheamicin produces DS lesions exclusively, with the lesions consisting of either two direct breaks or an abasic site opposite a single-strand (SS) break (Figure 1.6, left panel) [70, 71]. DNA damage by the parent esperamicin, esperamicin A1, consists of both SS breaks (75%) and DS lesions (25%) with an abasic site opposite a strand break (Figure 1.6, right panel) [71]. Esperamicin C, the analog of esperamicin A1 missing the anthranilate-deoxyfucose moiety (Figure 1.7), produces mostly DS lesions (>95%) like calicheamicin (Figure 1.6, left panel). In all cases, polyamines (e.g., putrescine) and lysines (either free or on nucleoproteins) cleave the abasic sites and nucleoside aldehydes, leaving ~85% of the ends of the DS breaks with both 3'- and 5'-phosphates and 3'-three nucleotide overhang (3-nt 3'-overhang) (Figure 1.6, middle panel) [72-74]. These drug-specific DNA ends will be exploited throughout my studies for the mapping of enediyne-specific DNA damage in the genome.

In spite of this common mechanism for damaging DNA, the enediynes differ significantly in the structure and arrangement of functional groups attached to the enediyne core (Figure 1.3). In my thesis research, I have focused on calicheamicin and esperamicins, which include esperamicin A1 and esperamicin C (Figure 1.7). Calicheamicin and esperamicin have an

identical aglycone core, the bicyclic ring that contains the enediyne moiety. However, the terminal carbohydrate-aromatic group in calicheamicin is positioned on the opposite side of the enediyne core in esperamicin A1, in the form of an anthranilate-deoxyfucose moiety (Figure 1.7). This anthranilate-deoxyfucose group is missing in esperamicin C. In my studies, I have used these structural differences to define the role of various enediyne structural motifs in the selection of cellular DNA targets *in vivo*.

### III. Earlier Studies of Eneidyne Target Selection *In Vitro*

The aim of my thesis research is to explore enediyne target selection in the complex milieu of the whole cell, and to test the relevance of *in vitro* models of enediyne target selection.

Previous studies in this laboratory have established models by which enediynes select their DNA targets in naked DNA and chromatin *in vitro*. This is illustrated by our studies of calicheamicin and esperamicins A1 and C. Using isolated nuclei and nucleosome core particles, Yu *et al.* demonstrated that these drugs target different regions of isolated nucleosomes [28]. Esperamicin A1 was limited to damaging the linker DNA that joins adjacent nucleosome cores, a behavior similar to that of an intercalating agent. It was subsequently shown that the anthranilate of esperamicin A1 was a novel DNA intercalator [71, 75].

However, calicheamicin and esperamicin C, an esperamicin A1 analog missing the sugar-anthranilate moiety, cleave both the linker and core DNA [28]. The aromatic ring in the side chain of calicheamicin did not intercalate in DNA and calicheamicin was capable of binding to DNA when the minor groove faced away from the nucleosome core [76]. Further studies in reconstituted nucleosomes revealed that calicheamicin damage at one site in the nucleosome occurred 4-fold more frequently than in naked DNA [76]. This finding suggested that there may be sites in the nucleus where protein-DNA complexes create high affinity targets for calicheamicin and other DNA-damaging agents.

Two observations led to the proposal that calicheamicin targets curved or flexible DNA sequences. First was the ability of calicheamicin to bind to the bent region of nucleosome core DNA. Second, we and others have also observed that calicheamicin produces damage mainly at the 3'-ends of purine

runs [76, 77]. One explanation for this behavior is that calicheamicin targets an unusual structure at the 3'-end of a purine tract, such as a kink or hinge at the purine/pyrimidine junction. This hypothesis is supported by the studies of Salzberg *et al.*, who found that calicheamicin bends DNA upon binding and that the binding sites may be flexible DNA sequences [78]. It is thus possible that calicheamicin will target curved or flexible DNA in certain regions of the chromatin *in vivo*.

In my thesis research, I have tested the relevance of these *in vitro* observations in the setting of the whole cell. Using the human PGK1 as a model gene, I have studied the effects of positioned nucleosomes (Chapter 3) and transcription factors (Chapter 4) on enediyne target selection *in vivo*.

#### IV. Significance

The research presented in this study is significant for several reasons. First, I have carried our earlier studies in naked DNA and reconstituted nucleosomes into the nucleus of a living cell. This allowed us to test the relevance of DNA target recognition established in these *in vitro* models. Secondly, the structural diversity of the enediynes provides an opportunity to define the roles of the various enediyne structural elements in target selection *in vivo*. Third, the techniques developed and the results obtained with enediynes will be broadly applicable to understanding the target selection mechanisms of other genotoxins. Finally, the knowledge gained from these studies will provide important information for the design of targeted anticancer drugs and for understanding the mechanism of genotoxin-induced mutagenesis.



## **V. Figures for chapter one**

**Figure 1.1.** Multiple factors of DNA structure in the mammalian nuclei may interact collectively with genotoxin structure to determine genotoxin target selection *in vivo*.

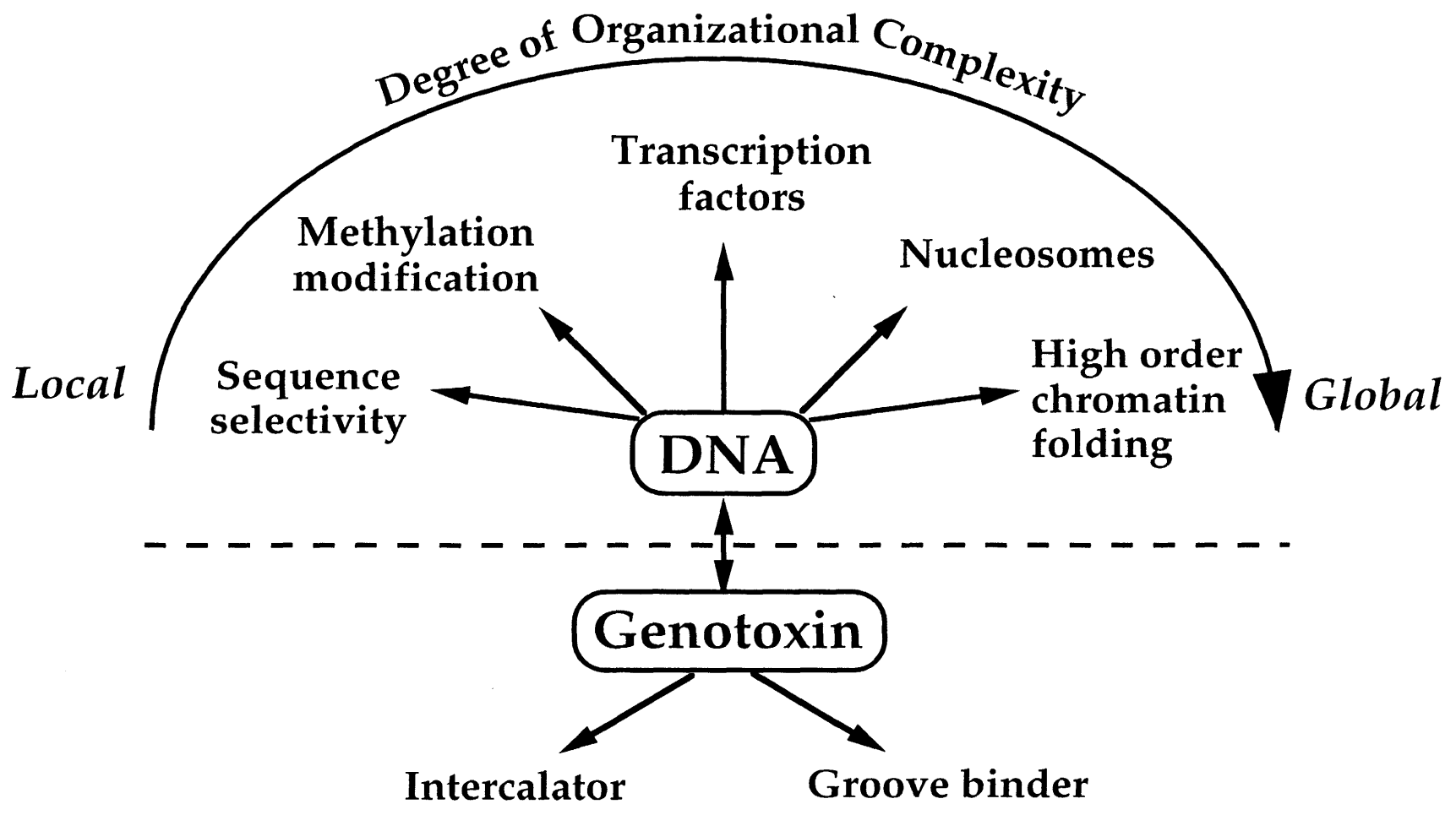


Figure 1.1

**Figure 1.2.** Chromatin structures of the human X chromosome-linked phosphoglycerate kinase gene (PGK1). The positions of nucleosomes in the inactive PGK1 and transcription factors in the active PGK1 were derived from the *in vivo* DNase I footprinting studies of Pfeifer and coworkers [79, 80]. The X chromosome location of PGK1 allows a comparison of DNA damage in the same gene in different states of chromatin structure.

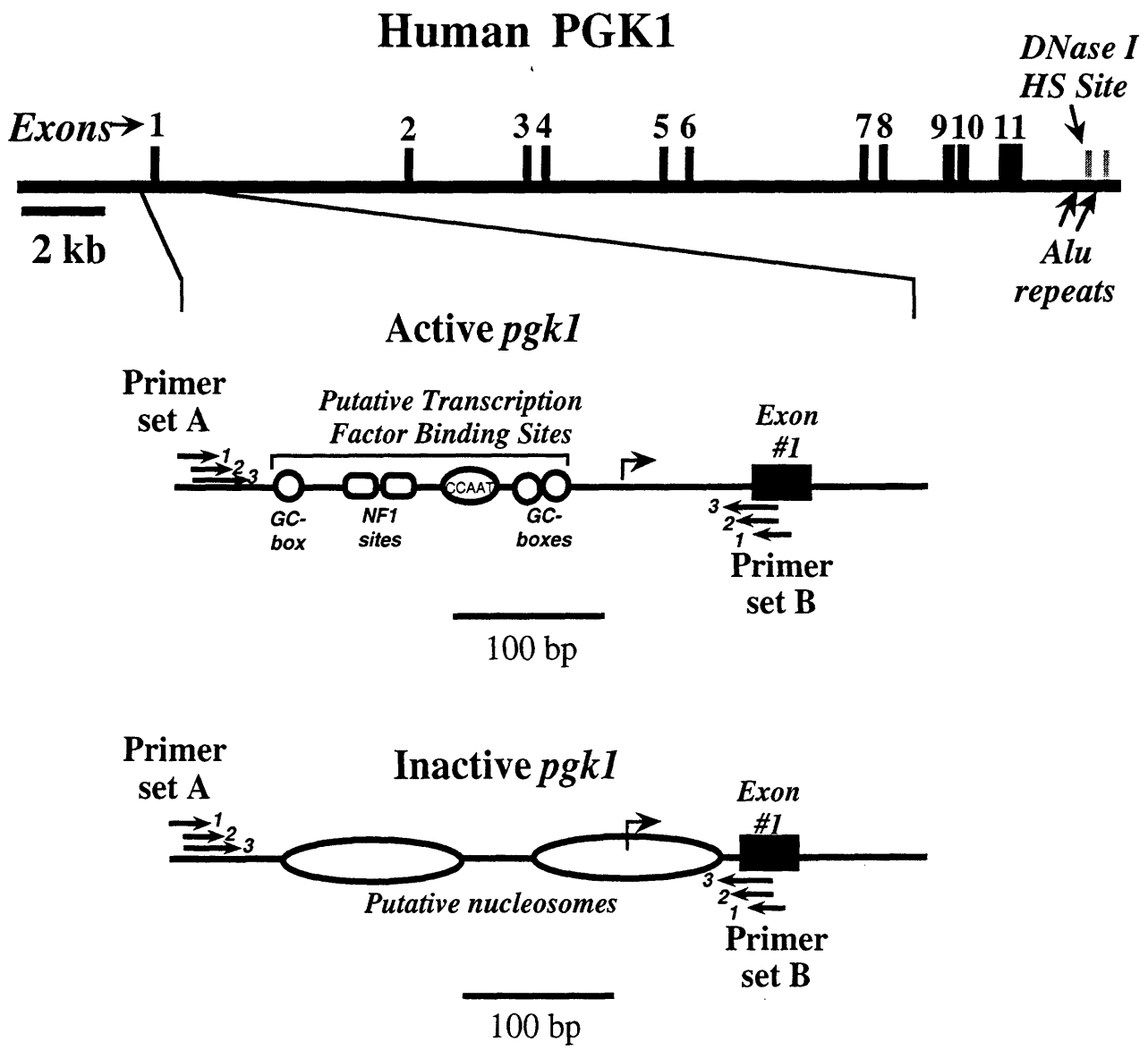


Figure 1.2

**Figure 1.3.** Members of the enediyne family. Enediynes are naturally occurring antitumor antibiotics isolated from several eubacterial species (reviewed in [54]). The name of the family derives from the common enediyne moiety in the core of these molecules that can form carbon-centered radicals and abstract hydrogens from DNA backbones.

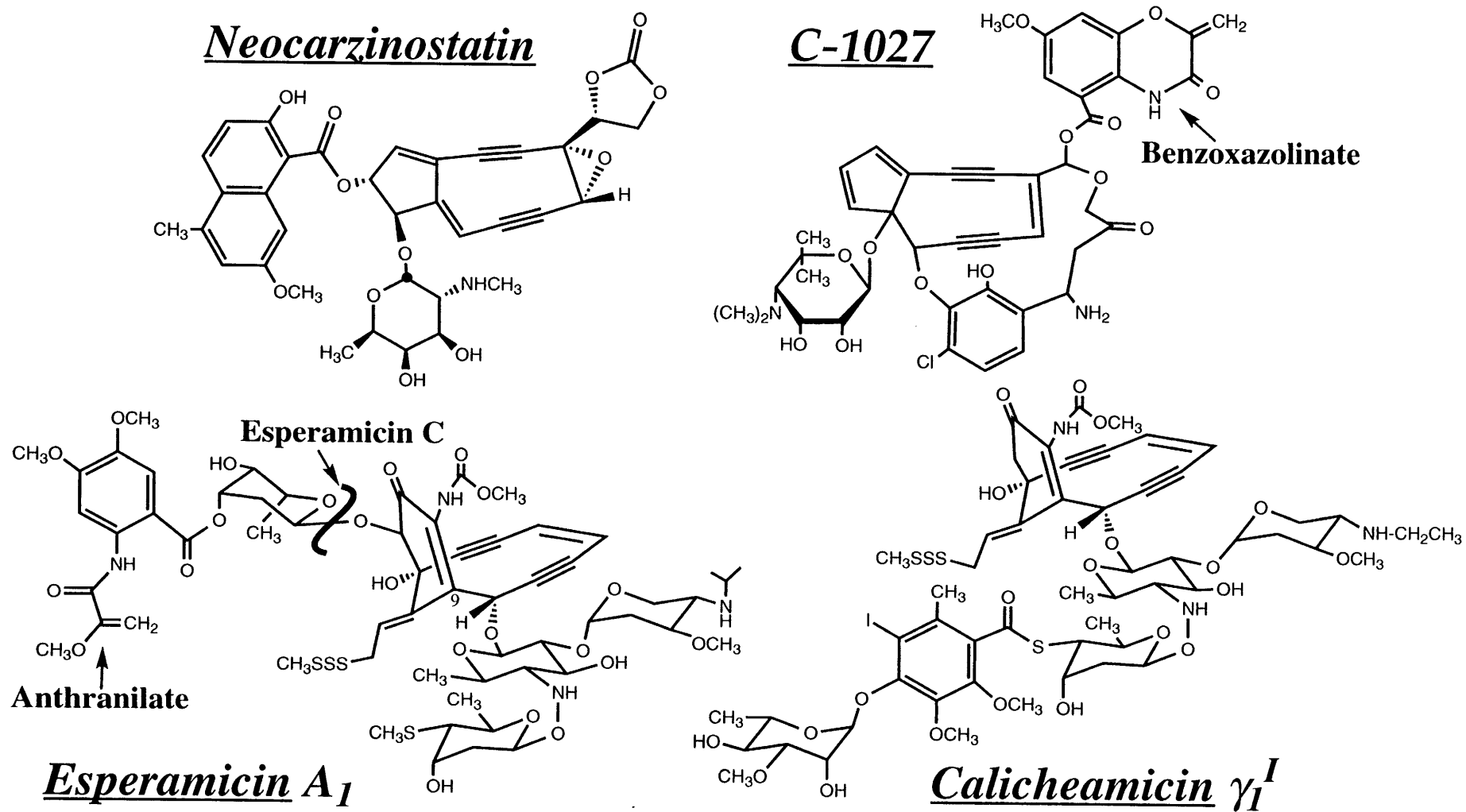


Figure 1.3

**Figure 1.4.** DNA damage mechanism of the enediyne core in calicheamicin and esperamicin. Initiated by a nucleophilic attack at the methyltrisulfide trigger by thiols (presumably glutathione *in vivo*), the enediyne core undergoes a Bergman-Masamune rearrangement to form a diradical species. When the drug binds in the minor groove of DNA, the diradical abstracts hydrogen atoms from deoxyribose of both strands to cause DNA lesions. The drug is deactivated after hydrogen abstraction [69].



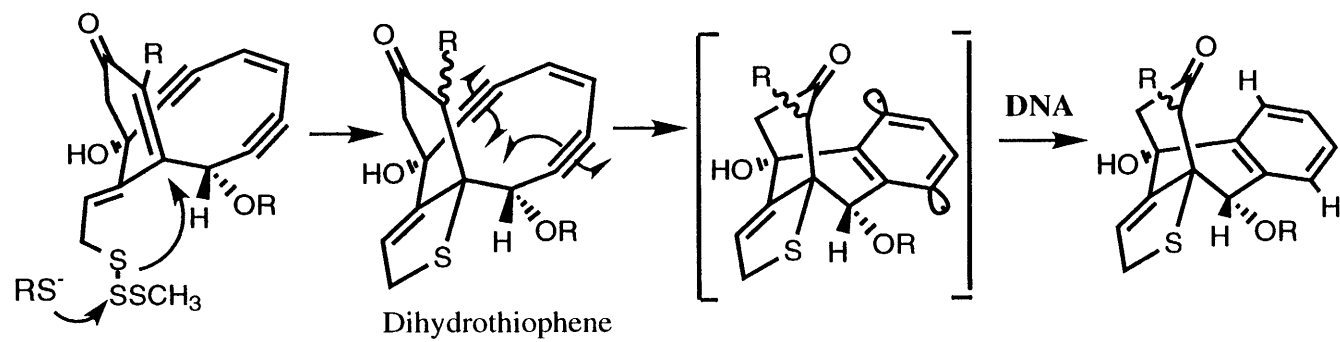
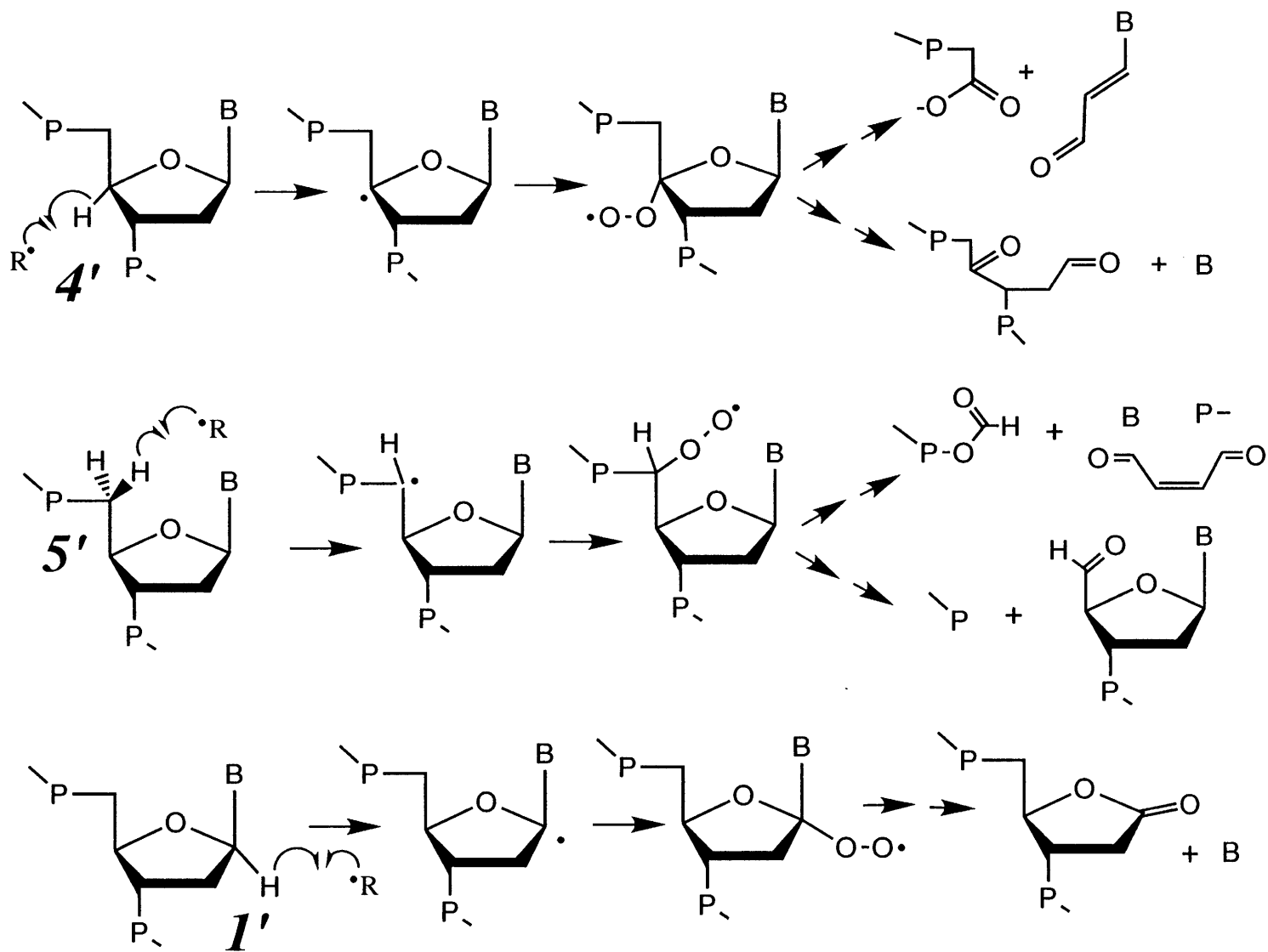


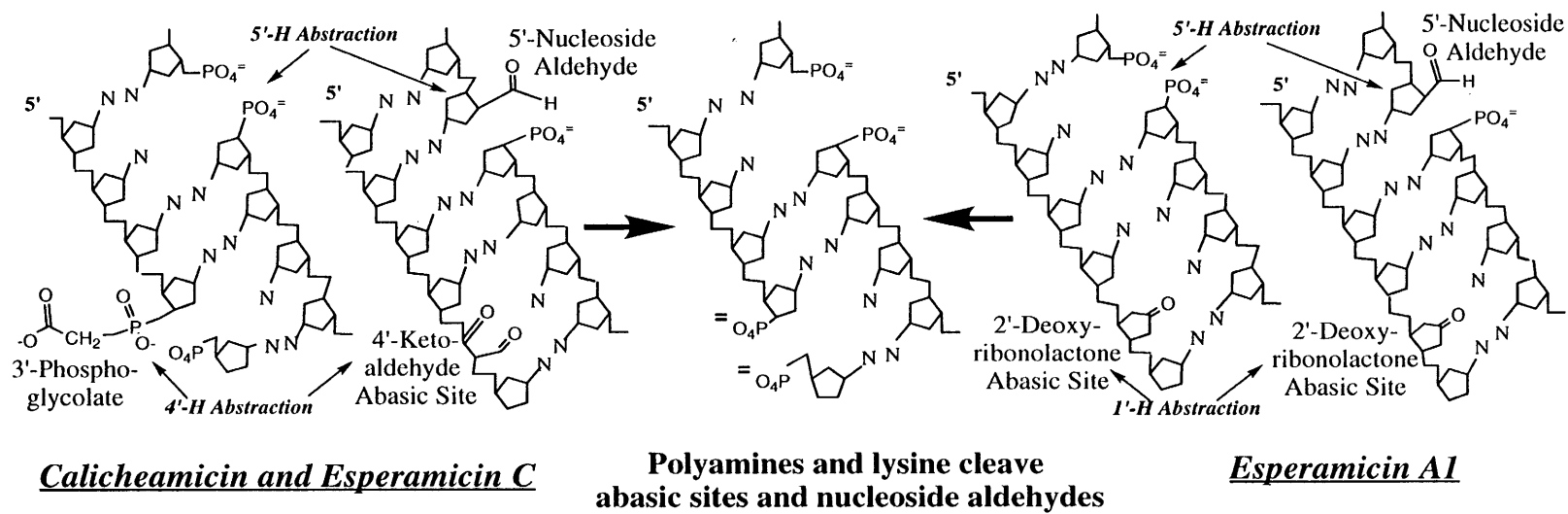
Figure 1.4

**Figure 1.5.** The expression of DNA lesions after hydrogen abstraction by enediynes. Enediynes can abstract hydrogens protrude toward the minor groove of a duplex DNA, which are the 1', 4', and 5' hydrogens. Deoxyribose lesions then undergo oxygen-dependent reactions, yielding a spectrum of degradation products unique to each abstracted hydrogen [58, 65].



**Figure 1.5**

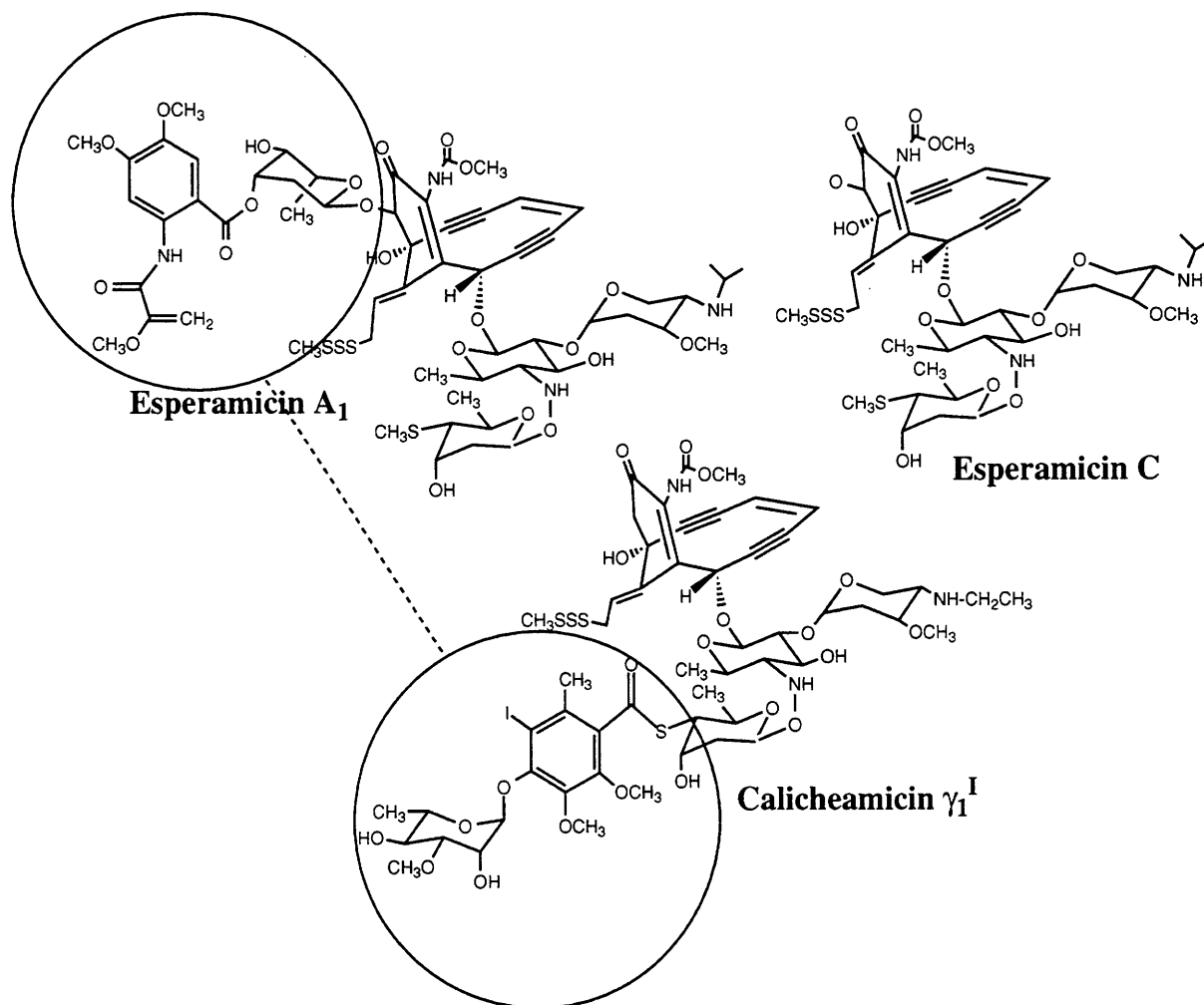
**Figure 1.6.** DNA lesions produced by calicheamicin, esperamicin A1 and C. Calicheamicin produces DS lesions exclusively, with the lesions consisting of either two direct strand breaks or an abasic site opposite a single-strand (SS) break (left panel) [70, 71]. DNA damage by the parent esperamicin, esperamicin A1, consists of both SS breaks (75%) and DS lesions (25%) with an abasic site opposite a strand break (right panel) [71]. Esperamicin C, the analog of esperamicin A1 missing the anthranilate-deoxyfucose moiety, produces mostly DS lesions (>95%) like calicheamicin (left panel). In all cases, polyamines (e.g., putrescine) and lysines (either free or on nucleoproteins) cleave the abasic sites and nucleoside aldehydes, leaving ~85% of the ends of the DS breaks with 3 nucleotide (3-nt) 3'-overhangs opposite 5'-phosphates (middle panel) [72-74].



**Figure 1.6**

**Figure 1.7.** Members of the enediyne family used in my thesis study.

Calicheamicin and esperamicin have an identical aglycone core, the bicyclic ring that contains the enediyne moiety. However, the terminal carbohydrate-aromatic group in calicheamicin is positioned on the opposite side of the enediyne core in esperamicin A1, in the form of an anthranilate-deoxyfucose moiety. This anthranilate-deoxyfucose group, which is an DNA intercalator [71], is missing in esperamicin C. This structural diversity allows one to define the relationship between drug structure and target selection.



**Figure 1.7**

## **CHAPTER 2**

### **DEVELOPMENT OF A MODIFIED LIGATION- MEDIATED PCR STRATEGY TO MAP ENEDIYNE- SPECIFIC DNA DAMAGE**



## I. Introduction

In Chapter two, a modified ligation-mediated polymerase chain reaction (LMPCR) is described that can accommodate the structure of enediyne-induced double-strand breaks. This technique, along with the conventional LMPCR, is used to identify the DNA damage distributions along the human PGK1 gene and the factors that modulate damage distribution (Chapters 3-5).

To define the location and relative quantity of drug-induced damage in the single-copy PGK1 gene, I needed a very sensitive method for mapping DNA damage. Traditionally, Southern-blot has been used to measure the extent of DNA damage in specific DNA sequences in cells (reviewed in [81]). However, due to its limited sensitivity, only damage in multi-copy sequences could be studied and resolution was limited to hundreds of base pairs [81].

LMPCR, on the other hand, allows the mapping of DNA damage at single nucleotide resolution in single-copy genes [82-84]. This technique was initially developed to address a problem that limits the use of PCR in genomic footprinting and sequencing studies: only one end of a DNA sequence is defined when strand breaks are introduced along the sequence. LMPCR solves the problem by ligation of a second defined sequence at sites of strand breaks [85, 86].

The technique of LMPCR is illustrated in Figure 2.1. Cells or purified DNA are first treated with a DNA damaging agent and DNA lesions or adducts are converted to strand breaks with 5'-phosphate ends compatible with DNA ligase. In the "first-strand synthesis", a gene-specific primer is annealed to the purified DNA and extended with a DNA polymerase until it reaches a strand break to create a blunt-ended DNA fragment. A duplex oligonucleotide linker is then ligated to the blunt-ended fragments. This

creates a substrate suitable for PCR and allows for the simultaneous amplification of all blunt-ended DNA fragments, using a primer complementary to the linker sequence and a second gene-specific primer. After amplification, the resulting damage ladder will be uniformly longer due to the length of the added linker. However, the Maxam-Gilbert sequencing standards [87] will also be amplified by the same procedure. Thus, the locations of genotoxin-induced damage sites are determined relative to the sequencing standards on a sequencing gel. All damage and sequence ladders can be visualized by extension of a [<sup>32</sup>P]-labeled gene-specific primer (Figure 2.1). Alternatively, the damage sites can be identified by blotting and hybridization with a radio-labeled gene-specific probe. The frequency of DNA damage at each site is determined by phosphorimager analysis.

I have modified the conventional LMPCR method to make it more specific for enediyne-induced DNA damage. As shown in Figure 2.2 (middle panel), after putrescine (a polyamine) treatment, calicheamicin and esperamicin produce double strand breaks with 5'-phosphates and three nucleotide 3'-overhangs (3-nt 3'-overhangs). Drug-induced double strand breaks can be directly ligated to oligonucleotide linkers with a randomized 3-nt 3'-overhang that complements the drug-induced overhangs (Figure 2.3). The advantage of this approach is that it circumvents the need for first-strand synthesis. In addition, direct ligation of the double-strand linker prevents any background single-strand nicks from being ligated and amplified, thus making the technique more specific for drug-induced DNA damage. The technique is also applicable to the 2-nt overhangs produced by other enediynes (e.g., C-1027 [60, 88]), as long as the end of the randomized linkers are changed accordingly.

## II. Materials and Methods

**Sources of the enediynes.** Calicheamicin was kindly provided by Dr. G. Ellestad, Wyeth-Ayerst Research; esperamicin A1 and C were kindly provided by Dr. J. Golik, Bristol-Myers Squibb.

**Cell culture.** HeLa S3 cells were grown as a suspension culture in Joklik modified Eagle's minimum essential medium with 10% newborn calicheamicin serum (Sigma) and 1% penicillin/streptomycin (Sigma). Human-hamster hybrid cell lines containing either an inactive human X chromosome (X8/6T2 cells) or an active human X chromosome (Y162/11C cells) were kind gifts of Dr. S. Gartler, Univ. Washington, Seattle. They were grown as a monolayer in RPMI 1640 medium with 10% fetal calicheamicin serum (Sigma) and 40 µg/ml Gentamicin (Sigma) [52].

**Treating cells with enediynes.** Cells are harvested at ~80% confluency and resuspended in phosphate-buffered saline (PBS) at  $10^7$  cells/ml. An aliquot of calicheamicin or esperamicin in methanol was added (final methanol concentration <1%), and the reaction was allowed to proceed for 30 min at 37°C. This incubation time was long enough to ensure that DNA damage was virtually complete, yet short enough to avoid significant repair or apoptosis [28]. We have previously demonstrated that enediyne-mediated DNA damage in nuclei was a direct result of the drug, and was not mediated by topoisomerases or nucleases [28]. The drug-damaged DNA was purified using a QiaAmp blood kit (Qiagen) and then treated with 100 mM putrescine for 1 hr at 37°C to cleave abasic sites to strand breaks with phosphate-ends [69, 89, 90]. DNA samples were finally ethanol precipitated and redissolved in 1 mM Tris, 0.1 mM EDTA (pH 7.8) at 1 mg/ml.

**Treating purified DNA with enediynes.** Genomic DNA was purified using a QiaAmp blood kit and dissolved in 50 mM HEPES, 5 mM EDTA, 10

mM glutathione (pH 7.0) at 0.1 mg/ml. An aliquot of calicheamicin or esperamicin in methanol was added (final methanol concentration <1%), and the reaction was allowed for 30 min at 37°C. The drug-damaged DNA was treated with putrescine as described above. DNA samples were then ethanol precipitated and redissolved in 1 mM Tris, 0.1 mM EDTA (pH 7.8) at 1 mg/ml.

**Maxam-Gilbert reactions.** DNA sequence ladders generated by base-specific DNA cleavage and conventional LMPCR are essential for the precise mapping of damage distribution at single-nucleotide resolutions. For Maxam-Gilbert sequencing reactions [87], purified genomic DNA was concentrated by ethanol precipitation to 5-10 mg/ml. The reactions were performed according to a protocol optimized for LMPCR [91].

**Oligonucleotide sequences.** Gene-specific primers #1 and #2 for the PGK1 promoter region were identical to that of Pfeifer *et al.* [53]. Labeling primers #3 was chosen to meet the criteria established by Mueller and Wold [92], and designed with the aid of Primer 0.5 software [93] using 50 mM salt concentration and 0.25 nM DNA concentration [94]. An upstream primer set A and a downstream primer set B were used. Each primer set consisted of three primers: primer A1 (first strand synthesis): AAGTCGGGAAGGTTTCCTT; primer A2 (PCR amplification): AAGGTTTCCTTGCGGTTTCGCGGCG; primer A3 (labeling): AAGGTTTCCTTGCGGTTTCGCGGCGTG. Primer B1 (first strand synthesis): CGTCCAGCTTGTCCAGC; primer B2 (PCR): TCCAGCGTCAGCTTGTTAGAAAGCG; primer B3 (labeling): TTCAGCGTCAGCTTGTTAGAAAGCGAC. The positions of these primers along the promoter region of PGK1 gene were shown in Figure 2.4. The sequence of the linker primer was as described elsewhere [85].

For conventional LMPCR, the linker consisted of a 25-mer annealed to an 11-mer [85, 86]. For modified LMPCR, the long strand of the linker

contained a randomized 3-nt 3'-overhang (Figure 2.3). Randomization of the three terminal nucleotides of the 28-mer was achieved using an equimolar mixture of all four nucleotides during synthesis (Oligos Etc). The resulting linker thus contained equal amounts of all 64 permutations of 3-nt 3'-overhangs, allowing it to accommodate all possible sequences damaged by calicheamicin and esperamicins.

**Conventional LMPCR.** To ensure the quantitative nature and reproducibility of the LMPCR process, all samples in each experiment were prepared in parallel. Master reaction cocktails were used to minimize pipetting error.

Conventional LMPCR was performed on drug-treated samples and Maxam-Gilbert sequencing markers. The protocol was a hybrid of Pfeifer's and Mueller's protocols [92, 95]. First strand synthesis: 2  $\mu\text{g}$  of DNA, 0.6 pmoles of primer 1, 3  $\mu\text{l}$  of 5x Sequenase buffer (250 mM NaCl, 200 mM Tris-HCl pH 7.7), and sufficient water were mixed in a final volume of 15  $\mu\text{l}$ . The DNA was denatured at 95°C for 3 min and primer annealing was allowed to occur at 48°C for 30 min. The sample was then chilled on ice and spun briefly in a microcentrifuge. To the annealed sample, 7.5  $\mu\text{l}$  of freshly prepared Mg-dNTP solution (made by mixing 47  $\mu\text{l}$  of water, 1  $\mu\text{l}$  of 1 M  $\text{MgCl}_2$ , 1  $\mu\text{l}$  of 1 M DTT, and 1  $\mu\text{l}$  of 25 mM dNTP) and 1.5  $\mu\text{l}$  of Sequenase at 3.25 units/ $\mu\text{l}$  [made by diluting stock Sequenase 2.0 (13 units/ml, (United States Biochemical-Pharmacia) four-fold in ice cold TE pH 8.0) were added. The following temperature ramp was applied to the sample in a thermocycler (MJ Research): 48°C for 5 min, 50°C for 1 min, 51°C for 1 min, 52°C for 1 min, 54°C for 1 min, 56°C for 1 min, 58°C for 1 min, and 60°C for 1 min. This temperature ramp during the first strand synthesis step was found to minimize PCR background [96]. The sample was cooled on ice and 6  $\mu\text{l}$  of 300 mM Tris-HCl pH 7.7 was

added. The sample was finally heat inactivated at 67°C for 15 min, cooled on ice and spun briefly.

Linker ligation and PCR amplification were performed according to Pfeifer *et al.* [95, 97]. Briefly, 45 µl of freshly prepared ligation mix (13.33 mM MgCl<sub>2</sub>, 30 mM DTT, 1.66 mM ATP, 83 mg/ml bovine serum albumin, 3 units per reaction T4 DNA ligase (Promega), and 100 pmoles linker) was added to the above sample. Ligation was allowed to occur at 18°C for 16 hours, and terminated by heating to 70°C for 10 min. DNA was precipitated by ethanol, dried under vacuum, and dissolved in 50 µl of water. A 50 µl aliquot of freshly prepared master Taq polymerase mix was added to each sample. This polymerase mix consisted of 20 mM Tris-HCl (pH 8.9), 80 mM NaCl, 0.02% gelatin, 4 mM MgCl<sub>2</sub>, 0.4 mM of each dNTP, 10 pmoles of primer #2, 10 pmoles of linker-primer, and 3 units of Taq polymerase (Boehringer-Mannheim). PCR (20 cycles) was performed on a thermocycler using the following temperature profile: 95°C for 1 min, 67°C for 2 min, 76°C for 3 min; with an additional 7 min of 76°C in the last cycle.

After PCR, samples were transferred to ice and 5 µl of the following labeling mix was added per reaction: mix 1 µl of 5x Taq buffer (200 mM NaCl, 50 mM Tris-HCl pH 8.9, 0.05 % (w/v) gelatin), 2.5 pmoles of the 5'-[<sup>32</sup>P] end-labeled primer #3, 6.25 nmoles of each dNTP, 1.5 units of Taq polymerase, and sufficient water to bring the volume to 5 µl. PCR products were then labeled by 2-4 repetitions of the following thermocycler program: 95°C for 1 min, 69°C for 2 min, and 76°C for 10 min [92].

Note: In certain cases, background smearing posed a problem in this direct labeling strategy. A gel transfer and hybridization method was used to eliminate the problem (see Chapter 3 and [91, 97, 98]).

**Modified LMPCR.** Eneidyne-treated DNA samples were also subjected to an alternative LMPCR strategy, which exploited the unique three nucleotide 3'-overhangs present on all double-strand breaks produced by calicheamicin and esperamicins A1 and C [70, 71].

Drug-damaged DNA was ligated directly to a modified linker possessing a three nucleotide 3'-overhang complementary to the drug-induced overhang. The modified linker consisted of a 28-mer oligonucleotide (GCGGTGACCCGGGAGATCTGAATTCNNN, where N represents a randomized nucleotide) annealed to an 11-mer oligonucleotide (5'-GAATTCAGATC) at 20 pmoles/ $\mu$ l. Randomization of the three terminal nucleotides of the 28-mer was achieved using an equimolar mixture of all four nucleotides during synthesis (Oligos Etc.). The resulting population of oligonucleotides provided complementary 3'-overhangs for all possible damage sites (Figure 2.3). The use of the modified linker obviated the need for first-strand synthesis. Ligation reactions consisted of 2  $\mu$ g of DNA, 7.5  $\mu$ l of 10x ligase buffer (Promega), 1  $\mu$ l of ligase (Promega, 3 units/ $\mu$ l), 100 pmoles of modified-linker, and sufficient water to yield a 75  $\mu$ l final volume. Ligation was allowed to occur at 18°C for 16 hours. The PCR amplification and labeling steps were the same as the conventional LMPCR described above.

**Sequencing gel analysis.** All LMPCR samples were phenol/chloroform extracted, ethanol precipitated, and resuspended in 4  $\mu$ l formamide loading buffer (95% v/v formamide, 20 mM EDTA pH 8.0, 0.05% xylene cyanol, 0.05% bromophenol blue). Half of each sample (2  $\mu$ l) was loaded onto an 8% polyacrylamide gel. Electrophoresis was carried out at 80 watts constant power (~50°C gel temperature) for 2 to 6 hours. Gel was dried under vacuum and exposed to a phosphor image screen over night. The image was scanned by a PhosphorImager (Molecular Dynamics).

### III. Results and Discussion

**Comparison of conventional LMPCR and modified LMPCR.** It was hypothesized that the modified LMPCR would make the technique more specific for enediyne-induced DNA damage by reducing the amplification of background single-strand nicks, due to the specificity of the modified linker for enediyne-induced DNA breaks. This hypothesis was tested with calicheamicin, an enediyne that produces double strand breaks exclusively [70]. As shown in Figure 2.5, modified LMPCR produced less background and cleaner signals than did conventional LMPCR, resulting in a more distinguishable signal at 100 nM calicheamicin concentration (lane 8 *vs.* lane 3).

Also shown in Figure 2.5 are the results of amplifying *Bsl* I-digested DNA by modified (lane 13) and conventional LMPCR (lane 14). *Bsl* I produces a 3-nt 3'-overhang that is compatible with the modified linker. Notice that in both *Bsl* I-digested and calicheamicin-damaged DNA, the products from the modified LMPCR are 3 bases longer than those from the conventional LMPCR. This is because the long strand of the mixed-linker is 3 nucleotides longer than the conventional linker, due to its 3-nt overhang (Figure 2.3). This should be kept in mind when determining the location of enediyne-mediated DNA damage by modified LMPCR.

Taking into account the 3-nt size difference and the gain of signal-to-background ratio in the modified LMPCR, most of the major DNA damage sites detected by the conventional LMPCR was also detected by the modified LMPCR. This indicates the validity of modified LMPCR in mapping enediyne-induced DNA damage.

**Single-hit conditions are important when mapping DNA damage sites by LMPCR.** In LMPCR, it is important to work under conditions in which the



gene region of interest is damaged only once. Multiple lesions in the amplified region will cause a bias toward shorter DNA fragments since only break sites closest to the gene-specific primer will be amplified. Figure 2.5 also shows that single-hit conditions are important for the quantitative comparison of damage frequency along DNA sequences. Good dose-response was obtained with modified LMPCR from 10 nM to 1  $\mu$ M calicheamicin concentrations (lanes 7 to 9). However, when calicheamicin was used at 10  $\mu$ M (lane 10), multiple damage sites in the region of interest bias the LMPCR toward shorter fragments.

To ensure that drug damage is within single-hit conditions in the DNA region of interest, the size of DNA fragments in the drug-treated sample is assessed by agarose gel electrophoresis. The molecular weight corresponding to the mobility of the peak mass distribution should be 6000 bp or greater. This ensures that greater than 99% of the DNA substrate receives less than or equal to one hit within a 400 bp region starting from the gene-specific primer (see appendix for mathematical derivations). In practice, the peak mass distribution of 6000 bp or greater is readily achieved by treating purified DNA with 1  $\mu$ M concentration or lower and by treating whole cells with 20  $\mu$ M or lower drug concentrations.

**The quantitative nature of LMPCR.** Although PCR is an exponential process, the amount of PCR product is linearly proportional to the number of template molecules, as long as the following two conditions are satisfied. First, all reaction components except template must be in vast excess, as is usually the case for at least the first 20 cycles [99]. Second, statistical sampling errors must be minimized by starting with at least 100 molecules of each template fragment [84]. This is readily achieved by starting with 2  $\mu$ g of DNA ( $\sim 6 \times 10^5$  copies of the genome; [100]) damaged to an average fragment size of

6000 bp. This damage frequency ensures single-hit conditions within the 400 bp region analyzed by any gene-specific primer (see above), and ensures that there are at least 100 molecules of each damage site fragment (see appendix for mathematical derivations).

The quantitative nature of the LMPCR is shown in Figure 2.6, in which HeLa genomic DNA was cleaved with *Ras* I. In lanes 1 and 3, 1 or 2  $\mu\text{g}$  of digested DNA was amplified by LMPCR, respectively, using primers specific to the PGK1 promoter. LMPCR signals in these lanes indicated fragments of the expected size, and the signals were quantitated using the ImageQuant software. The signal intensities in lanes 1 and 3 were 1729 and 3587 counts, respectively; non-digested controls showed very little background. Thus, two-fold differences in the amount of the starting material can be detected. Pfeifer *et al.* also reported that the two-fold differences between one X chromosome-containing male DNA and two X chromosome-containing female DNA could be reliably detected [53, 80].

LMPCR is also highly reproducible. The last two lanes in Figure 2.5 (lanes 15-16) are both T+C reaction markers, amplified with primer annealing temperatures of 69°C and 67°C, respectively. The results are indistinguishable. Thus, LMPCR is a robust technique and slight differences in amplification conditions still yield the same results.

In conclusion, a reliable and quantitative LMPCR procedure was established. This allowed mapping of enediynes-induced DNA damage in any DNA region where primer sequences were known. In subsequent Chapters, I will apply the conventional and modified LMPCR to the human PGK1 gene model to study the roles of nucleosomes, transcription factors, and DNA methylation in the selection of DNA targets by enediynes.

#### **IV. Figures for chapter two**

**Figure 2.1.** Schematic comparison of the conventional and modified LMPCR.  
See main text for a detailed description.

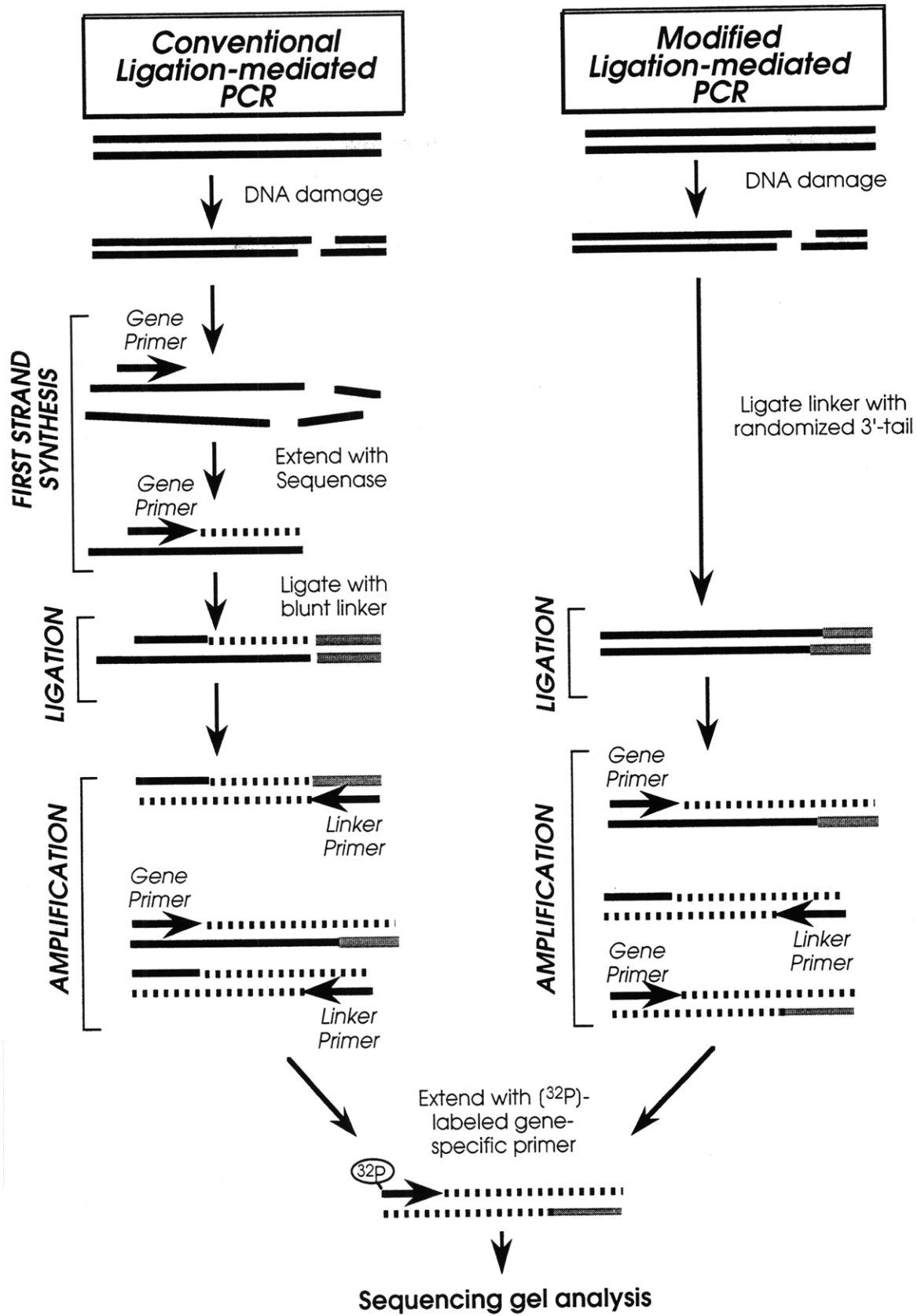
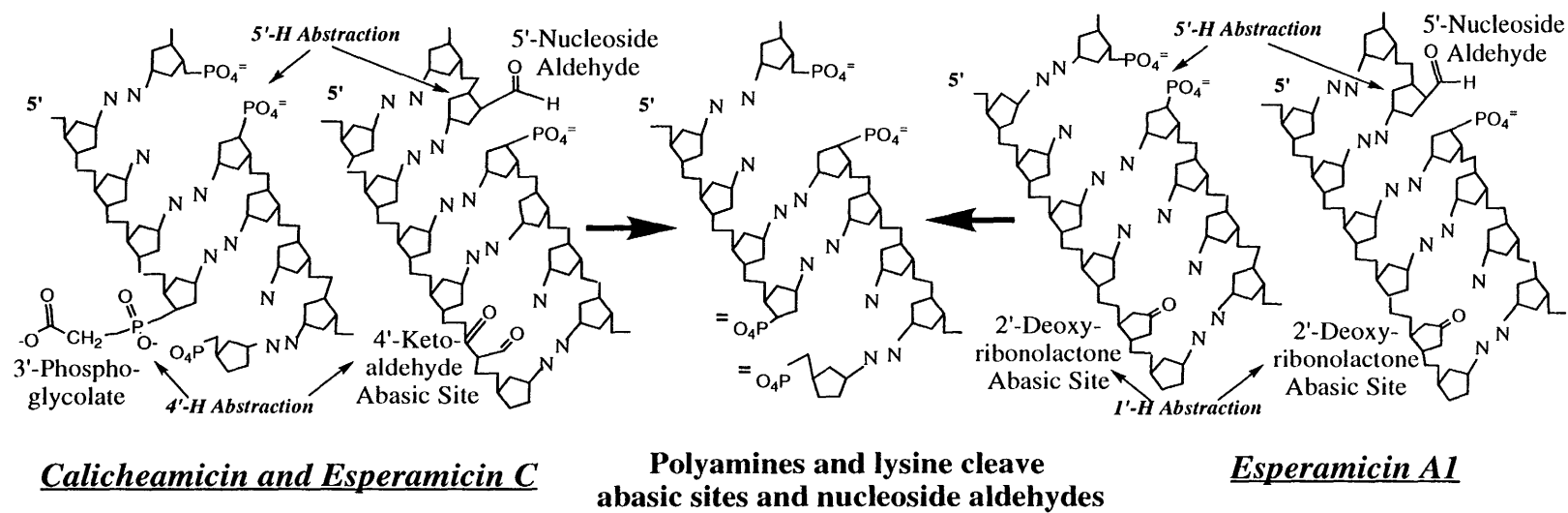


Figure 2.1

**Figure 2.2.** DNA lesions produced by calicheamicin, esperamicin A1 and C. Calicheamicin produces DS lesions exclusively, with the lesions consisting of either two direct strand breaks or an abasic site opposite a single-strand (SS) break (left panel) [70, 71]. DNA damage by the parent esperamicin, esperamicin A1, consists of both SS breaks (75%) and DS lesions (25%) with an abasic site opposite a strand break (right panel) [71]. Esperamicin C, the analog of esperamicin A1 missing the anthranilate-deoxyfucoose moiety, produces mostly DS lesions (>95%) like calicheamicin (left panel). In all cases, polyamines (e.g., putrescine) and lysines (either free or on nucleoproteins) cleave the abasic sites and nucleoside aldehydes, leaving ~85% of the ends of the DS breaks with 3 nucleotide (3-nt) 3'-overhangs opposite 5'-phosphates (middle panel) [72-74].



**Figure 2.2**

**Figure 2.3.** Structure of the modified and conventional linkers. The long strand of the modified linker is three nucleotides longer than the conventional linker, due to the randomized 3-nt 3'-overhang. The nucleotide randomization was created by using an equimolar mixture of all four nucleotides during synthesis. The resulting linker thus contains equal amounts of all 64 permutations of 3-nt 3'-overhangs, allowing it to accommodate all possible sequences damaged by calicheamicin and esperamicins.



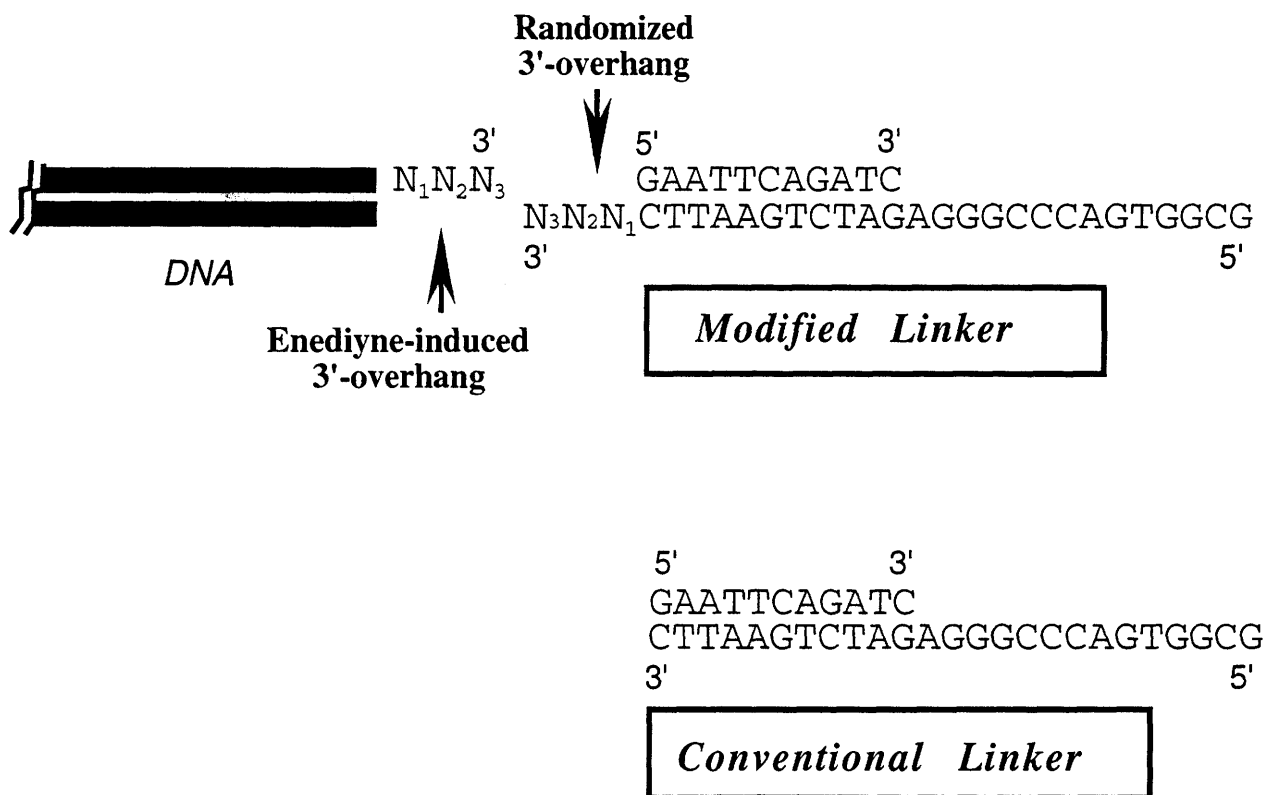


Figure 2.3

**Figure 2.4.** The position of primer set A and B on the human PGK1 gene. Each primer set consists of three primers for first strand synthesis (primer 1), PCR amplification (primer 2), and labeling (primer 3).

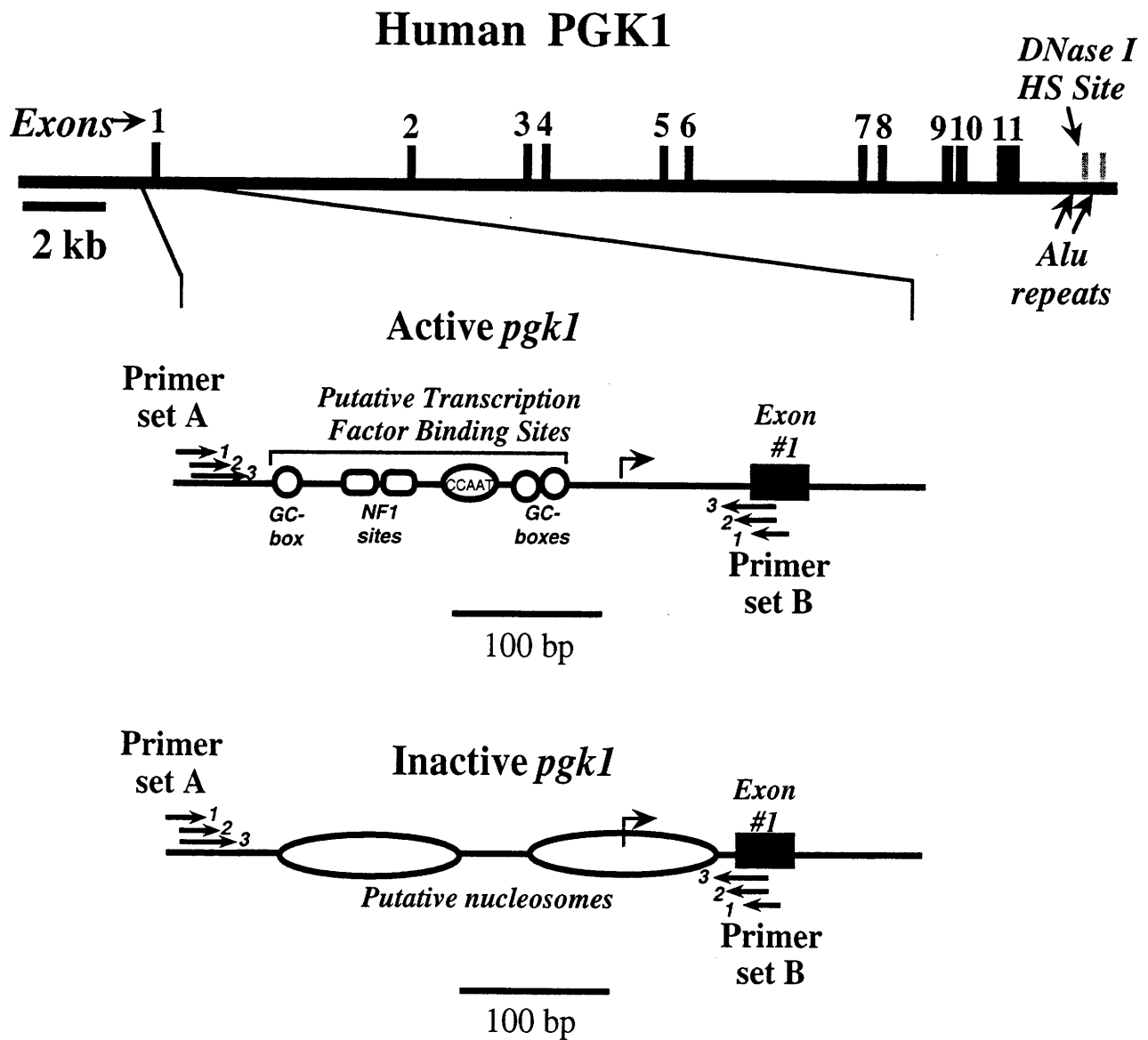


Figure 2.4

**Figure 2.5.** Results of conventional and modified LMPCR. Lanes 1-5 were conventional LMPCR amplification of HeLa DNA damaged by 0, 10 nM, 100 nM, 1  $\mu$ M, and 10  $\mu$ M of calicheamicin, respectively. Lanes 6-10 were modified LMPCR amplification of HeLa DNA damaged by 0, 10 nM, 100 nM, 1  $\mu$ M, and 10  $\mu$ M of calicheamicin, respectively. Lanes 11 and 12 are GA- and TC-specific reactions amplified by conventional LMPCR. Lanes 13 and 14 are *Bsl* I-digested DNA amplified by modified and conventional LMPCR, respectively. *Bsl* I produces 3-nt 3'-overhangs that are also ligatable by the modified linker. Notice that the product from the modified LMPCR is 3 bases longer than that from the conventional LMPCR, as expected. Lanes 15 and 16 are both TC-specific reactions amplified at 69 and 67°C, respectively.



Figure 2.5

**Figure 2.6.** The quantitative nature of LMPCR. Lanes 1 and 3 are modified LMPCR amplification of different amounts of *Rsa* I digested HeLa DNA: 1 and 2  $\mu$ g, respectively. Lanes 2 and 4 are modified LMPCR amplification of undigested HeLa DNA: 1 and 2  $\mu$ g, respectively. The LMPCR bands in lanes 1 and 3 are at the expected size (pointed by the arrow), and are subjected to volume quantitation by the ImageQuant software: bands A and B has 1729 and 3587 counts, respectively. Blunt linkers were used to ligate directly to *Rsa* I-induced blunt ends in these experiments.

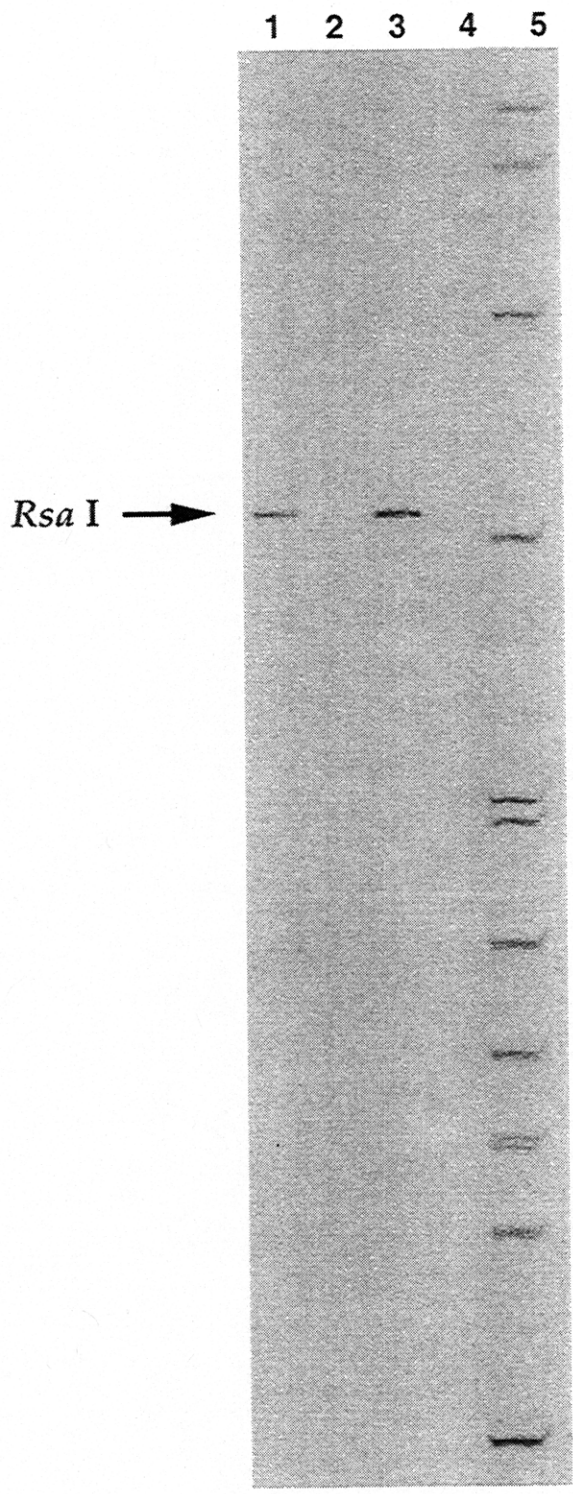


Figure 2.6

## **CHAPTER 3**

### **THE ROLE OF NUCLEOSOMES IN ENEDIYNE TARGET SELECTION**



## I. Introduction

The packaging of DNA as chromatin in cells alters both its conformation and dynamics (Chapter 1). These perturbations introduce another level of complexity to the mechanisms by which DNA-damaging chemicals select their targets. To better understand how genomic organization affects target selection by genotoxins, I have compared the DNA damage produced by two enediynes, calicheamicin and esperamicin, in a single-copy human gene. In this Chapter, I will focus on the effects of nucleosome structure in the inactive human PGK1 gene on enediyne target selection *in vivo*.

The nucleosome is a basic element of chromatin structure [3] as well as a model for the effects of protein binding on DNA structure and dynamics (Chapter 1). It consists of two regions, core and linker. The core is composed of ~146 base pairs of DNA wrapped ~1.8 times in a left-handed superhelix around four pairs of histone proteins, and the linker represents the 20-60 base pairs of DNA joining adjacent cores [3, 4]. In addition to the bending-induced changes in DNA structure, the nucleosomal DNA is constrained in its dynamics by protein-DNA contacts. This constraint contributes to the reduced binding of intercalating agents to nucleosome core DNA [29-32].

The enediyne family of antitumor antibiotics presents a unique opportunity to study the relationship between drug structure and selection of DNA targets in chromatin. Calicheamicin and esperamicins share a common mechanism for producing DNA damage, yet differ in the organization of functional groups attached to the enediyne core (Chapter 1). Specifically, the terminal carbohydrate-aromatic group of the calicheamicin tetrasaccharide side chain is positioned on the opposite side of the enediyne core in

esperamicin A1, in the form of a deoxyfucose-anthranilate moiety. This group is missing in esperamicin C (Figure 1.7).

The diversity of enediyne structure suggested that the drugs would target different regions of chromatin in cells, an hypothesis supported by our studies in isolated chromatin [28] and reconstituted nucleosomes [76, 101]. These studies revealed that esperamicin A1 was limited to damaging the linker region between nucleosome cores due to intercalation of an anthranilate moiety [71]. However, calicheamicin and esperamicin C, which are nonintercalating groove-binders, damaged both the core and linker DNA [28, 76, 101]; core DNA damage in this case was limited to sites where the minor groove faced away from the histone proteins. These observations serve as a model for the relationship between enediyne structure and *in vivo* target selection, and they suggest that the structural diversity of the enediynes can be exploited to probe chromatin structure in cells.

To test the validity of these *in vitro* models, I have now examined enediyne-induced damage in a single copy gene in living cells. Using LMPCR, I demonstrate that esperamicin A1-induced DNA damage is suppressed in two putative nucleosome cores in the transcriptionally silent human PGK1 gene, while calicheamicin and esperamicin C produced damage throughout the nucleosome. These results confirm hypotheses generated from studies about enediyne target selection in isolated chromatin and reconstituted nucleosomes, and suggest that enediynes may prove useful as chromatin footprinting agents.

## II. Materials and Methods

**Materials and cell lines.** Calicheamicin  $\gamma_1^I$  and esperamicin A1 and C were provided by Dr. George Ellestad (Wyeth-Ayerst Research) and Dr. Jerzy Golik (Bristol-Myers Squibb), respectively. Chinese hamster-human hybrids containing either an inactive (cell line X86T2) were provided by Dr. Stanley Gartler (University of Washington, Seattle; ref. [52]). A clone of X86T2 cells enriched in the human X chromosome was provided by Dr. Gerd Pfeifer (City of Hope Medical Center, Duarte, CA). Cells were grown as a monolayer in RPMI 1640 with 10% fetal calicheamicin serum and 40  $\mu\text{g}/\text{ml}$  Gentamicin [52]. A clone of the upstream region of the human PGK1 gene, pBSHPGK1, was provided by Dr. Judith Singer-Sam (City of Hope Medical Center, Duarte, CA; [102]).

**Treatment of cells with enediynes.** Cells were harvested at ~80% confluency and resuspended in phosphate-buffered saline (PBS) at  $10^7$  cells/ml. Drug treatment, DNA purification, and putrescine treatment were as described before (Chapter 2). DNA samples were finally ethanol precipitated and redissolved in 1 mM Tris, 0.1 mM EDTA (pH 7.8) at 1 mg/ml.

**Treatment of purified DNA with enediynes.** Genomic DNA was purified using a QiaAmp blood kit and dissolved in 50 mM HEPES, 5 mM EDTA, 10 mM glutathione (pH 7.0) at 0.1 mg/ml. Drug treatment, putrescine treatment, and DNA purification were as described before (Chapter 2). After ethanol precipitation, DNA samples were dissolved in 1 mM Tris, 0.1 mM EDTA (pH 7.8) at 1 mg/ml.

For DNA sequencing reactions, purified genomic DNA was concentrated by ethanol precipitation to 5-10 mg/ml. Maxam-Gilbert sequencing reactions were performed according to an LMPCR-optimized protocol [91].

**LMPCR.** We employed two LMPCR techniques that used the following primers in the 5' region of the human PGK1 gene: CGTCCAGCTTGTCAGC (+134 to +118, primer 1); TCCAGCGTCAGCTTGTTAGAAAGCG (+123 to +99, primer 2); TGGGGAGAGAGGTCCGGTGATTCGGTCA (+80 to +54, primer 3); TCCAGCGTCAGCTTGTTAGAAAGCGACAT (+123 to +95, primer 4). The sequences of the blunt linker and linker primer were as described elsewhere [85]. Conventional LMPCR was performed as described in Chapter 2. Half of each sample was resolved on a 6% sequencing gel. Electroblothing and hybridization were performed as described elsewhere [95]. The hybridization probe was made by repeated primer extension from a cloned PGK1 template (pBSHPGK1) using primer 3 and *Taq* polymerase [103]. Hybridized membranes were subjected to phosphorimager analysis (Molecular Dynamics).

Modified LMPCR was also performed as described in Chapter 2. After 20 cycles of amplification, labeling mix was added and each sample was subjected to 4 cycles of labeling (Chapter 2). Half of each sample was resolved on a 4-6% sequencing gel, which was then dried and subjected to phosphorimager analysis (Molecular Dynamics).

**Data analysis.** In order to compare damage frequency in isolated and cellular DNA, I performed the experiments with drug concentrations that produced roughly similar levels of DNA damage in the two situations. This amounts to roughly a 10-fold higher concentration of drug for treating cells than isolated DNA (*e.g.*, Figure 3.1). The difference is likely due to factors such as accessibility of DNA in higher order chromatin structures, sequestration of the drug in lipid membranes in the cells, or deactivation of the drugs in the cytoplasm.

I then accounted for unavoidable differences in the levels of DNA damage in different DNA samples and for lane-to-lane variation in gel loading. To do this, I normalized the phosphorimager signal intensities in each lane so that damage in putative nucleosome linker regions was the same in both isolated and cellular DNA. For example, the signal intensities for damage produced by esperamicin A1 in cellular DNA (Figure 3.1, lane 5) were multiplied by a factor of three so that damage frequencies in the region ~+20 to ~+70 were the same in both naked and cellular DNA. In this case, the factor of three represents the average difference in signal intensity for four major peaks in this region in isolated and cellular DNA (see Figure 3.2).

The validity of the normalization process is illustrated by three points. First, the expectation of equivalent amounts of damage in the nucleosome linkers has firm foundations in our previous *in vitro* studies [28, 76]. Second, normalization of the data to the linker peaks is used consistently in all of our studies to avoid any biases. Finally, normalization of the data to either of the two linker regions in the inactive PGK1 gene, (*i.e.*, +20 to +70 and around position -200 in Figure 3.2A) produces the same result.

### III. Results

**Mapping esperamicin-induced DNA damage in the PGK1 gene by modified LMPCR.** In this experiment, a modified linker was ligated directly to the double-strand breaks produced by esperamicins A1 and C. It should be noted that single-strand breaks produced by esperamicin A1 would not be detected using the modified linker.

The damage produced by esperamicins A1 and C in naked DNA and in cells containing an inactive PGK1 is shown in the gel in Figure 3.1 and in the line graphs in Figure 3.2. Examination of Figure 3.1 reveals that 10- to 20-fold higher concentrations of esperamicins A1 and C were required to produce comparable amounts of DNA damage in cellular DNA compared to isolated DNA. This was the case for all of the enediynes and, as discussed earlier, it was likely due to factors such as sequestration of the drug in lipid membranes in the cells or deactivation of the drugs in the cytoplasm.

To address the role of the cellular environment of DNA on the damage produced by the enediynes, we have made quantitative comparisons of the various lanes in Figure 3.1 following normalization of the damage frequency data. In the inactive PGK1 gene, damage produced by esperamicin A1 in cells is reduced compared to that in naked DNA in two regions indicated by the bars in Figure 3.1. This phenomenon is shown more clearly in Figure 3.2A, in which regions of reduced damage lie between positions -330 to -200 and positions -150 to +1, and they are flanked by regions in which cellular and isolated DNA experience similar levels of damage. The sizes of the protected regions (~130 and ~150 base pairs) are consistent with the 146 base pair length of nucleosome core DNA [3]. Furthermore, the identification of two nucleosomes in this region of the inactive PGK1 gene is consistent with the DNase I digestion studies of Pfeifer and Riggs [80].

The profile of esperamicin C-induced DNA damage in the inactive PGK1 gene in cells is similar to that in isolated DNA (Figure 3.2B). There is no general reduction in the level of damage between positions -330 to -200 and positions -150 to +1 as observed with esperamicin A1.

**Mapping enediyne-induced DNA damage in the PGK1 gene by conventional LMPCR.** Studies were also performed with conventional LMPCR to compare the *in vivo* and *in vitro* damage produced by enediynes, since this method recognizes both single- and double-strand breaks [86]. Representative gels are shown in Figure 3.3 for esperamicins A1 and C and calicheamicin. As shown in the line graphs in Figures 3.4B and 3.4C, calicheamicin and esperamicin C produce damage in both the putative linker and core regions of the downstream nucleosome, while damage produced by esperamicin A1 occurs mainly outside the core DNA (Figure 3.4A).

There are several sites at which the damage frequency differs for isolated and cellular DNA treated with calicheamicin and esperamicin C. These sites may be located where the minor groove faces the histone proteins, thus making the site inaccessible to the minor groove-specific enediynes. We have observed this phenomenon in isolated core particles [28] and reconstituted nucleosomes [76]. However, many of the sites are subjected to similar levels of damage in both isolated and cellular DNA, which is consistent with damage at sites where the minor groove faces away from the histone proteins. Furthermore, there is no generalized suppression of the damage between positions -150 and +1 as observed with esperamicin A1.

The reproducibility of the results is demonstrated in Figures 3.3 and 3.4A. In Figure 3.3, duplicate samples of esperamicin A1-damaged isolated (lanes 1) and cellular DNA (lanes 2) are shown. While there are some minor bands that differ between the duplicate lanes, probably due to statistical

sampling errors with low levels of damage [84, 86], the majority of bands are of similar intensity in each lane. Data beyond position -80 was not used for analysis due to unreliable amplification of long fragments by conventional LMPCR, as discussed later. Of greater importance, I have indicated the ratios of the level of esperamicin A1-induced DNA damage in cellular DNA to that in isolated DNA in Figure 3.4A. The damage ratio at each site is an average value ( $\pm$  SD) for three different experiments. The relatively small errors attest to the accuracy of the data.



#### IV. Discussion

The goal of the present work was to test the hypothesis that different enediynes would recognize different chromatin structures in living cells. This hypothesis arose from our previous studies in isolated chromatin [28] and reconstituted nucleosomes [76, 101], in which we observed that damage produced by esperamicin A1 was limited to the linker regions between nucleosome cores. The basis for this linker selectivity was determined to be intercalation by an anthranilate moiety [71, 75], which caused the drug to bind poorly to the dynamically constrained DNA of the nucleosome core. Consistent with this hypothesis was our observation that removal of the anthranilate to form esperamicin C caused the drug to damage both the core and linker DNA of the nucleosome. Furthermore, calicheamicin, another nonintercalating groove binder, was also capable of damaging both the core and linker DNA [28, 76].

The present results in whole cells confirm these *in vitro* observations. In the two nucleosome core-sized regions upstream of the inactive human PGK1 gene, damage produced by esperamicin A1, but not by esperamicin C and calicheamicin, is reduced in cells compared to naked DNA, while flanking regions (putative linkers) show similar levels of damage. One such region of suppressed damage lies between positions -330 and -200, the other between -150 and +1. The two nucleosome cores would thus be joined by about 50 base pairs of linker DNA, which is typical for nucleosome linkers in mammalian cells [3]. The second nucleosome core appears to end near position +1, as suggested by the results with both conventional and modified LMPCR (Figures 3.2 and 3.4).

The presence of two nucleosomes detected by esperamicin A1 in the inactive PGK1 gene and the absence of detectable nucleosomes in the active

PGK1 gene are consistent with the nuclease digestion studies of Pfeifer and Riggs [80]. They observed DNase I hypersensitive regions spaced at roughly 10 base pair intervals between positions -330 to -200 and -90 to +50 in the inactive gene. However, this spacing would require a ~110 base pair linker between the two nucleosomes, which is inconsistent with the observed linker sizes of ~40-60 base pairs in vertebrate organisms and mammalian cells in culture (reviewed in [3]). There are at least two explanations for this discrepancy in the position of the downstream nucleosome. One is that detergent-induced disruption of the cell and nuclear membranes, which is required to allow entry of DNase I into the nucleus, causes sliding of the downstream nucleosome(s). Such sliding has been observed by several groups during nuclease digestions in isolated nuclei (reviewed in [3]). It is also possible that portions of the DNase I digestion pattern were influenced by factors other than accessibility of nucleosome core DNA, such as sequence selectivity of the enzyme or the presence of other chromosomal proteins. Confirmation of the nucleosome positions by micrococcal nuclease digestion was not possible in the studies of Pfeifer and Riggs, probably due to the fact that the region under study is very G-C rich and micrococcal nuclease shows a marked preference for A-T rich sequences [80].

In all of my studies, I observed low levels of esperamicin A1-induced damage in the nucleosome core DNA *in vivo*. This is likely due to transient disruption of nucleosomal structure during DNA replication and repair. Cells used in the present studies were monolayers at ~80% confluency, so it is likely that some of the cells were in S-phase. While DNA transcription has also been shown to disrupt nucleosome structure [48], the X86T2 cells used in our studies contained only the inactive X chromosome. Thus the low levels

of nucleosome core DNA cleavage by esperamicin A1 are not likely due to transcription.

While the quantity of damage at each site is modulated by chromatin structure *in vivo*, the positions of damage sites are remarkably similar to that in naked DNA (Figures 3.1-3.4). This indicates that DNA *in vivo* retains some of the structural features present *in vitro*. Specifically, the *in vitro* sequence selectivity still plays a significant role in target selection by genotoxins in cells.

It is noteworthy that, under my conditions, the LMPCR technique modified for enediyne-induced DNA damage consistently resulted in longer amplification products than conventional LMPCR. This allowed me to examine longer regions of the PGK1 gene for the location of drug-induced DNA breaks (compare Figures 3.2 and 3.4). The basis for this difference appears to lie in the use of Sequenase for the primer extension step of conventional LMPCR, since, in modified LMPCR, this step is eliminated by direct ligation of damage sites to the linker DNA; subsequent steps are identical in both techniques. Sequenase is a highly processive enzyme at 37°C [104]. However, first-strand synthesis was performed at 48°C or above to reduce non-specific amplification in conventional LMPCR [96] and it is possible that the elevated temperature reduced the processivity of the polymerase. Whatever the basis, LMPCR modified for enediyne-induced DNA damage allows longer regions of DNA to be examined on a single sequencing gel. The modified-linker technique is also applicable to the two base pair overhangs produced by other enediynes (e.g., C-1027; [60, 88]), as long as the ends of the linkers are changed to account for the different overhangs.

Enediynes offer several advantages over other methods for defining nucleosome positions. The main advantage is their utility with intact cells, since treatment of cells with enediynes does not disrupt the architecture of the cell or nucleus. The enediynes are lipophilic molecules that readily diffuse into cells. On the contrary, use of DNase I and micrococcal nuclease requires cell permeabilization to allow access of the enzymes to nuclear chromatin. In addition, DNase I-treated DNA often gives poor LMPCR signals due to polymerase extension from the enzyme-induced 3' hydroxyl-ended strand breaks [105], and micrococcal nuclease is biased toward AT-rich regions [80]. Methidiumpropyl EDTA produces linker-selective DNA damage in isolated nuclei [106], though evidence is lacking for its utility in whole cells, and dimethyl sulfate, which readily penetrates the cell membrane, does not footprint nucleosomes [107]. UV photofootprinting has the advantage of minimal cell perturbation and good sensitivity, but it is limited to the presence of dipyrimidines at the protein-DNA contact sites, and the need for additional DNA treatment by T4 endonuclease V and *E. coli* photolyase [83]. In comparison to these other agents, treatment of intact cells by enediynes is very straightforward and results in strand breaks with 5'-phosphate ends compatible with LMPCR.

In conclusion, I have found that esperamicin A1 can recognize locally positioned nucleosomes on the inactive human PGK1 gene *in vivo*, while esperamicin C and calicheamicin cleave both the core and linker DNA of the nucleosome. The results of these structure/function studies are consistent with *in vitro* models and serve as a benchmark for future explorations of the role of genomic organization in the selection of targets by enediynes and other genotoxins. The results also suggest that enediynes may prove useful as chromatin footprinting reagents.



## **V. Figures for chapter three**

**Figure 3.1.** Sequencing gel analysis of the DNA damage produced by esperamicins in the inactive human PGK1 genes *in vitro* and *in vivo*. Modified LMPCR was employed to amplify sites of enediyne-induced strand breaks in the inactive PGK1 gene (X86T2 cells). Drug damage was studied in both isolated DNA (lanes 1-3) and in intact cells (lanes 4-6). Lanes 1 and 4: untreated controls; lanes 2 and 3: 0.1  $\mu\text{M}$  of esperamicin A1 and C, respectively; lanes 5 and 6: 1  $\mu\text{M}$  of esperamicin A1 and C, respectively. The amplified DNA was resolved on a 4% sequencing gel. The position in the PGK1 gene is shown in the right margin and the proposed positions of two nucleosomes are denoted by black bars. The weak cleavage apparent with esperamicin C in the inactive PGK1 gene in cells (lane 6) was due to a low drug concentration in this particular experiment; in other experiments with higher drug concentrations (10  $\mu\text{M}$ ), the damage was similar to that in isolated DNA (*e.g.*, Figure 3.3).

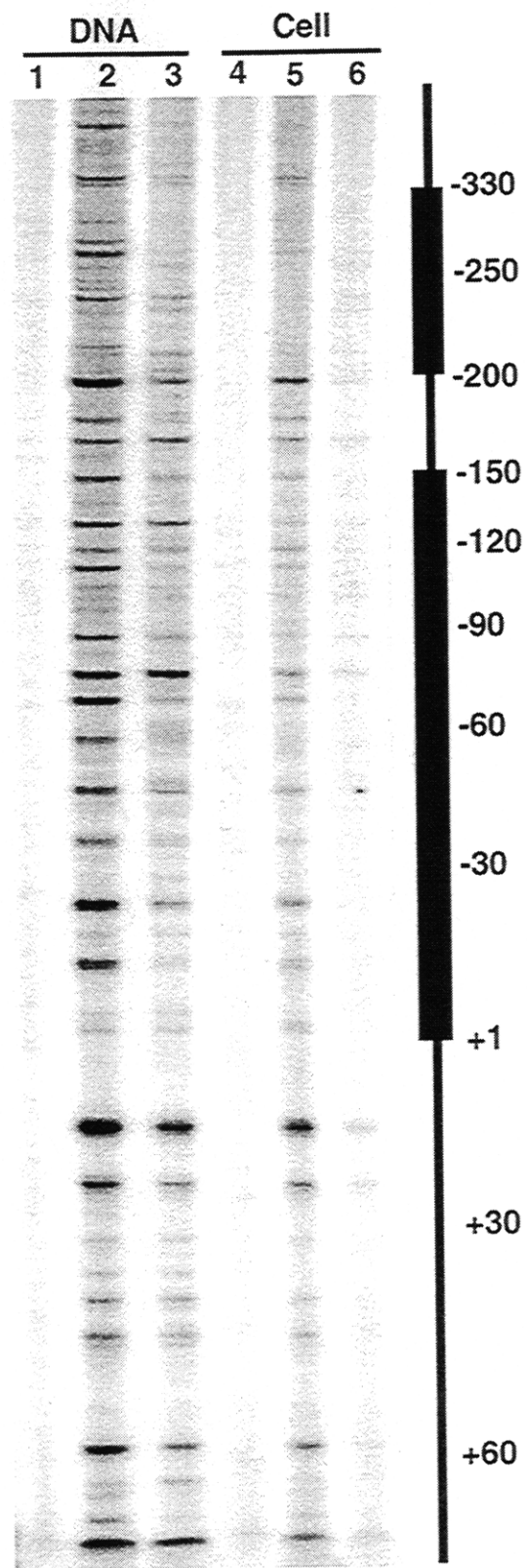


Figure 3.1



**Figure 3.2.** Comparison of DNA damage produced by esperamicin A1 (panel A) and esperamicin C (panel B) in the inactive human PGK1 gene. The gel shown in Figure 3.1 was subjected to phosphorimager analysis and the normalized data are presented as overlaying line graphs of damage frequency along the human PGK1 gene. Black lines represent damage in isolated DNA and gray lines represent damage in cells. The position in the PGK1 gene is indicated below each graph and the proposed positions for two nucleosomes are indicated below the graph.

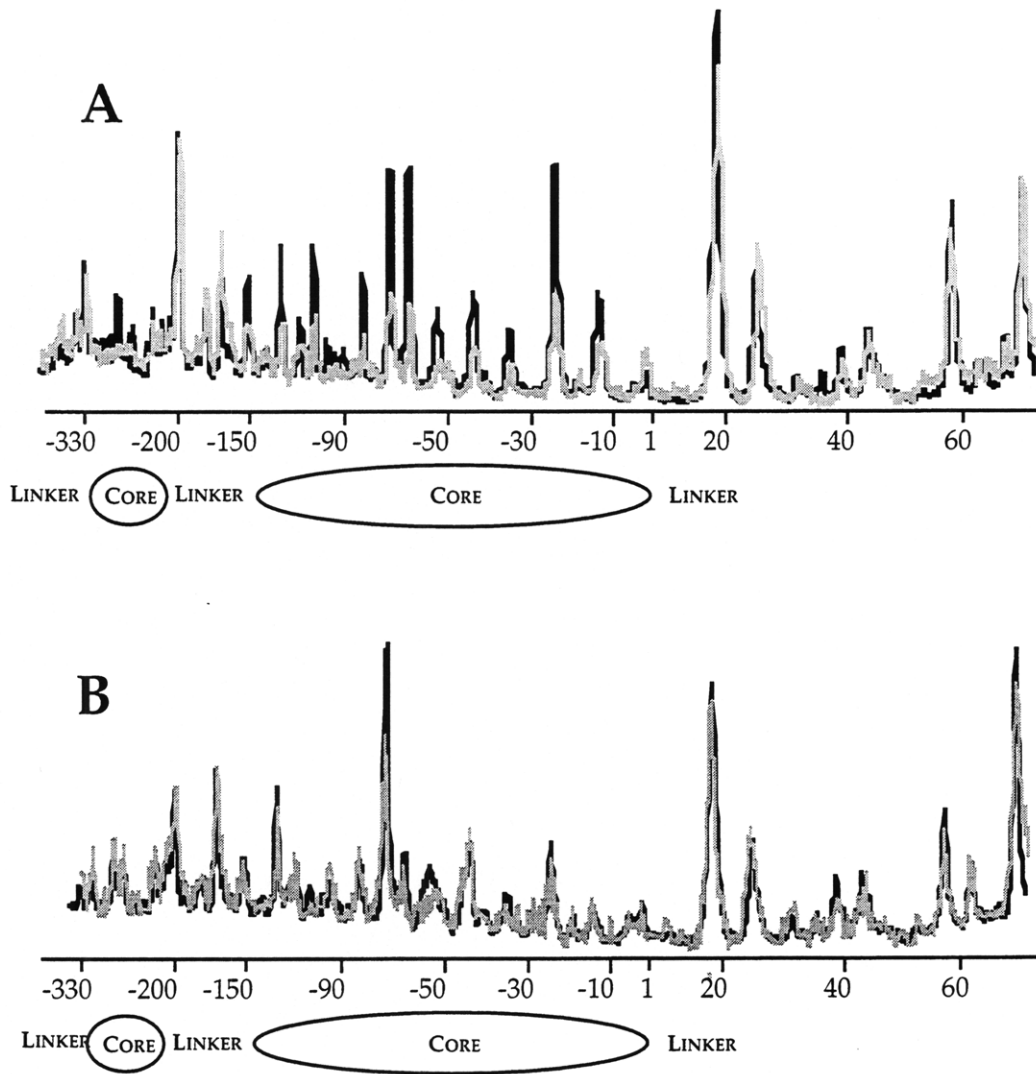
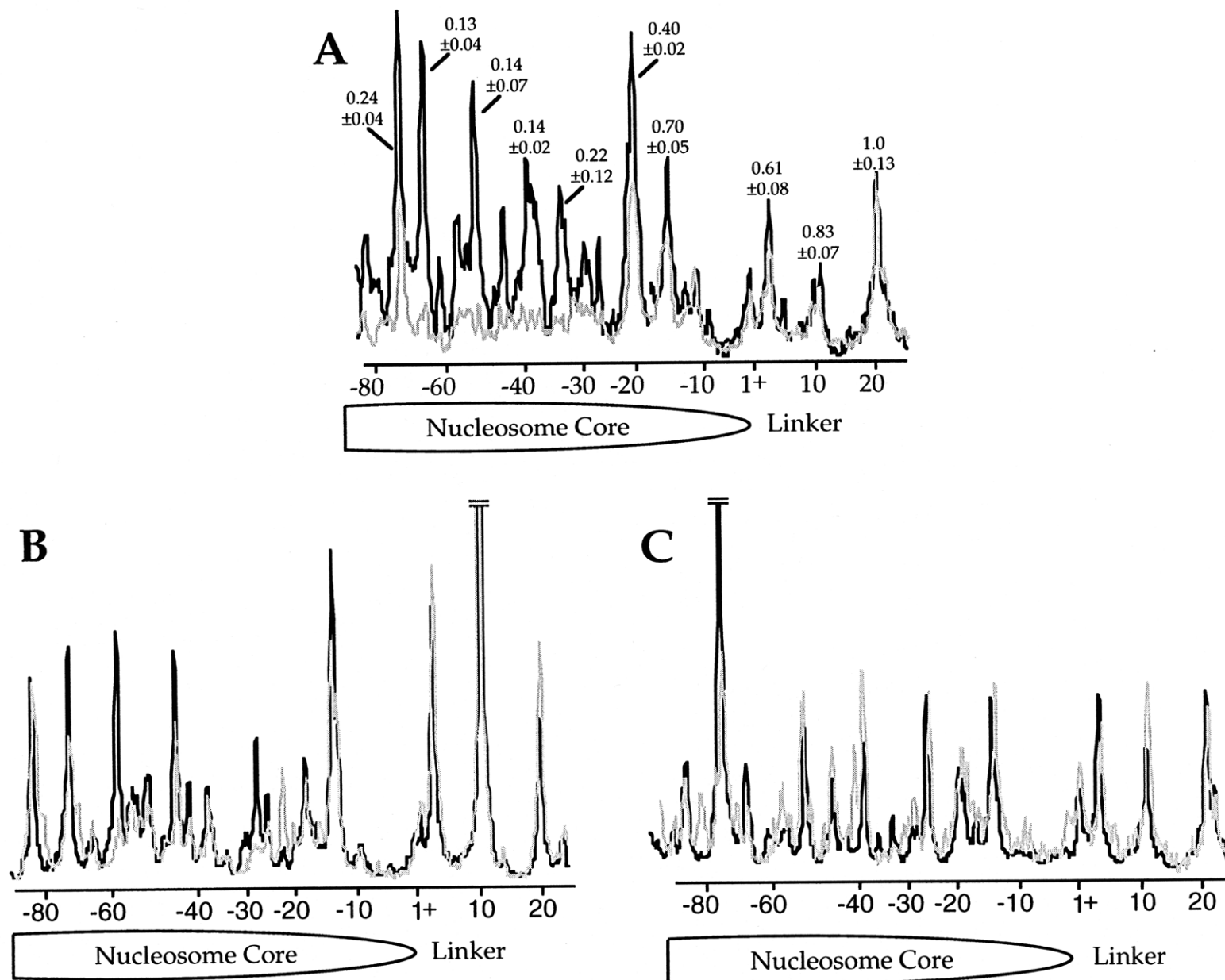


Figure 3.2

**Figure 3.3.** Sequencing gel analysis of the DNA damage produced by esperamicins A1 and C and calicheamicin in the inactive human PGK1 gene *in vitro* and *in vivo*. Conventional LMPCR was used to amplify sites of DNA damage produced by the enediynes in X86T2 cells (lane 2 in all panels) or in DNA isolated from these cells (lane 1 in all panels). Esperamicin A1 concentrations were 0.1  $\mu\text{M}$  (*in vitro*, lane 1) and 1  $\mu\text{M}$  (*in vivo*, lane 2); esperamicin C concentrations were 0.4  $\mu\text{M}$  (*in vitro*, lane 1) and 10  $\mu\text{M}$  (*in vivo*, lane 2); and calicheamicin concentrations were 0.1  $\mu\text{M}$  (*in vitro*, lane 1) and 4  $\mu\text{M}$  (*in vivo*, lane 2). The amplified DNA was resolved on a 6% sequencing gel. AG and CT are Maxam-Gilbert sequencing standards. The numbers on the right indicate the position in the human PGK1 gene. Please note that a lane to the right of the CT sequencing standard in the esperamicin A1 panel and a lane between the calicheamicin-treated samples were removed for clarity.



**Figure 3.4.** Comparison of DNA damage produced by esperamicin A1 (A), calicheamicin (B), and esperamicin C (C) in the inactive human PGK1 gene. The gels shown in Figure 3.2 were subjected to phosphorimager analysis and the normalized data are presented as an overlay of line graphs of damage frequency along the human PGK1 gene. Black lines represent damage in isolated DNA and gray lines represent damage in cells. The numbers above the peaks in Panel A represent the average ratios of the level of esperamicin A1-induced DNA damage in cellular DNA to that in isolated DNA at each site; the indicated errors are standard deviations for n=3. The position in the PGK1 gene is noted below the graphs along with the proposed position of the nucleosome.



**Figure 3.4**

## **CHAPTER 4**

# **THE ROLE OF TRANSCRIPTION FACTORS IN ENEDIYNE TARGET SELECTION**

## I. Introduction

Transcriptionally active genes in eukaryotic nuclei exist in a decondensed state of chromatin structure that allows access of transcription factors and polymerases to their recognition sequences [3, 4]. Active chromatin is characterized by stretches of open, nucleosome-free regions bound by transcription factors (Chapter 1). Since the binding of transcription factors is so fundamental to chromatin structure in active genes and it represents the next level of complexity beyond the DNA helix, its effects on enediyne target selection is addressed in this chapter.

The binding of transcription factors to their recognition sequences alters DNA structure and dynamics [108]. The structures of several transcription factor-DNA complexes have been solved, and there are models for other proposed structures [45]. Our understanding of the interactions between transcription factors and their cognate DNA sequences has reached the sophistication of the nucleosome structure. With this in mind, I have undertaken structure-function studies with enediynes to define the effects of transcription factors on genotoxin target selection.

The promoter of the human PGK1 gene is an ideal model to study active chromatin. Unlike inducible genes that require transient modification of the nucleosomal structure, the active PGK1 maintains nucleosome-free regions bound by putative transcription factors [109]. Such stable rather than fluctuating chromatin structure facilitates the interpretation of *in vivo* DNA damage patterns produced by genotoxins.

On the basis of footprinting studies, Pfeifer and coworkers have identified several putative transcription factor binding sites in the active PGK1 gene, including a CCAAT-binding protein, nuclear factor 1 (NF1)-like protein, and GC-box binding proteins [80]. Previous biochemical studies have



demonstrated that the binding of these factors to their cognate DNA significantly alters the local environment of the DNA. For example, NF1 was shown to completely embrace the circumference of the DNA helix [47], while CCAAT and GC-box binding proteins bind in the major groove and bend the DNA helix [110-113]. Therefore, these transcription factors in the human PGK1 promoter provide an excellent model to study the effects of different types of transcription factors on enediyne target selection.

I have mapped the DNA damage produced by calicheamicin and esperamicins in the active human PGK1 gene *in vivo*. A comparison of the damage in the same gene in its purified form and in cells indicated that DNA damage *in vivo* is modulated by transcription factor binding. Esperamicin A1-induced DNA damage is suppressed at sites within all transcription factor footprints, but it is enhanced at several sites flanking the footprints. This damage enhancement is also observed with esperamicin C and calicheamicin. Furthermore, damage caused by esperamicin C and calicheamicin is suppressed in the NF1- like site but not within the CCAAT- and GC boxes. These observations are consistent with the types of drug-DNA and protein-DNA interactions among the participating molecules. The results demonstrate that both sequence selectivity and chromatin structure are important in defining target selection by enediynes *in vivo*.

## II. Materials and Methods

**Materials and cell lines.** Calicheamicin  $\gamma_1^I$  and esperamicin A1 and C were provided by Dr. George Ellestad (Wyeth-Ayerst Research) and Dr. Jerzy Golik (Bristol-Myers Squibb), respectively. Chinese hamster-human hybrids containing an active human X chromosome (cell line Y162-11C) were provided by Dr. Stanley Gartler (University of Washington, Seattle) [52]. Cells were grown as a monolayer in RPMI 1640 with 10% fetal calicheamicin serum and 40  $\mu\text{g}/\text{ml}$  Gentamicin [52]. A clone of the upstream region of the human PGK1 gene, pBSHPGK1, was provided by Dr. Judith Singer-Sam (City of Hope Medical Center, Duarte, CA) [102].

**Treatment of cells with enediynes.** Cell treatment was as described before (Chapters 2). After conversion of abasic sites to strand breaks by putrescine, DNA samples were ethanol precipitated and redissolved in 10 mM HEPES, 0.1 mM EDTA (pH 7.7) at 1 mg/ml. The usage of HEPES instead of Tris for DNA storage was recommended to limit background oxidative damage [114].

**Treatment of purified DNA with enediynes.** DNA reactions were as described before (Chapters 2). After ethanol precipitation, DNA samples were finally stored in 10 mM HEPES, 0.1 mM EDTA (pH 7.7) at 1 mg/ml.

DNA sequencing reactions were performed as described (Chapters 2).

**LMPCR.** We used the following primers in the 5' region of the human PGK1 gene: CGTCCAGCTTGTCCAGC (+134 to +118, primer B1); TCCAGCGTCAGCTTGTAGAAAGCG (+123 to +99, primer B2); TGGGGAGAGAGGTCGGTGATTCGGTCA (+80 to +54, primer B3). The sequences of the blunt linker and linker primer were as described elsewhere [85]. Conventional LMPCR was used to amplify DNA sequencing reactions,

while modified LMPCR was used to amplify drug-induced DNA damage. Both methods were performed as described in Chapters 2 and 3.

After PCR amplification, half of each sample was resolved on a 6% sequencing gel. Electroblothing and hybridization were performed as described before [95]. The hybridization probe was made by repeated primer extension from a cloned PGK1 template (pBSHPGK1) using primer B3 and *Taq* polymerase [103]. Hybridized membranes were subjected to phosphorimager analysis (Molecular Dynamics).

**Data analysis.** In order to compare damage frequency in isolated and cellular DNA, I performed the experiments with drug concentrations that produced roughly similar levels of DNA damage in the two situations. This necessitates a higher drug concentration for treating cells than isolated DNA. I then performed phosphorimager analysis essentially as described elsewhere [76]. Each pixel value was first divided by the sum of the total pixel values in the lane to account for lane-to-lane variations in sample loading. Normalized pixel values in each lane were then plotted as a line graph. Line graphs of the same sample from different sequencing gels (due to varied running time to resolve different-sized fragments) were combined so that the overlapping signal intensities were the same. The combined line graphs from *in vitro* and *in vivo* drug treatment were presented as an overlay to better visualize differences between the two substrates.

### III. Results

**DNA damage by esperamicin A1 in the active PGK1 gene.** The damage produced by the DNA intercalator esperamicin A1 in naked DNA and in cells containing an active PGK1 is shown in the right half of Figures 4.1 and 4.2, and summarized in the overlaid line graphs in Figure 4.3A. The positions of the putative transcription factors are derived from the *in vivo* DNase I and dimethyl sulfate (DMS) footprinting studies of Pfeifer and coworkers [53, 80, 115]. In all cases, damage produced by esperamicin A1 in cells is reduced compared to naked DNA within the putative transcription factor binding sites. In other words, *in vivo* damage suppression regions colocalize with transcription factor footprints [53, 80].

While damage within the transcription factor footprints is suppressed, there is an enhancement of cellular damage by esperamicin A1 in regions between transcription factor binding sites (asterisks in Figures 4.1, 4.2, and 4.3A).

**DNA damage by esperamicin C in the active PGK1 gene.** I also compared *in vivo* and *in vitro* damage produced by esperamicin C, an A1 analog missing the deoxyfucose-anthranilate intercalator. Esperamicin C also exhibits increases in cellular DNA damage in the region between NF1 and CCAAT-binding sites (Figures 4.1 and 4.3B). These sites are identical to those noted earlier for esperamicin A1. Therefore, it appears that the enhancement of damage in the regions between transcription factors is not dependent on DNA intercalation.

Within the transcription factor binding sites, damage by esperamicin C in cells is suppressed in the middle of the putative NF1 site, but not in the CCAAT- and GC-binding sites. In fact, damage in the two GC-boxes is enhanced in cellular DNA (asterisks in Figures 4.1 and 4.3B). The

enhancement of damage within protein footprints by esperamicin C is in direct contrast to the suppressed damage in the same regions by esperamicin A1 (compare Figures 4.3B and 4.3A).

**DNA damage by calicheamicin in the active PGK1 gene.** A comparison of damage produced by calicheamicin in cells and isolated DNA is shown in the left half of Figure 4.2 and the line graph in Figure 4.3C. Once again, two sites of enhanced damage are observed in the regions between bound transcription factors (indicated by asterisks). In addition, one site of enhanced damage is observed within the downstream GC-box. On the other hand, damage in cells is suppressed within the putative NF1 site.

Compared to the esperamicins, calicheamicin damage sites are fewer in number, probably due to its extended sugar side chain which increases the sequence selectivity of the molecule [116]. This may explain the lack of damage in the CCAAT site and the upstream GC-box. In the ~400 bp PGK1 promoter region studied, one hotspot of calicheamicin damage was detected at position -224 in naked DNA. The sequence context of this damage hotspot is shown on the top of Figure 4.4. Position -224 lies at the 3'-end of a run of eight purines, which is typical of calicheamicin damage hotspot reported in the literature (Figure 4.4). In cellular DNA, two calicheamicin damage hotspots were detected, one at position -224 and the other at -165, both of which are located in the region between the putative GC- and NF1-sites (Figure 4.3C).

#### IV. Discussion

The results with the active human PGK1 promoter indicate that genotoxin damage *in vivo* is modulated by transcription factors. The cellular and DNA damage data for three structurally different enediynes are summarized in the line graphs in Figure 4.3, where *in vivo* damage (gray lines) is overlaid on *in vitro* damage (black lines). Several conclusions can be drawn from these comparisons. First, cellular DNA damage by esperamicin A1 is inhibited within the putative transcription factor binding sites, but enhanced between the binding sites. Secondly, these sites of damage enhancement are also observed with esperamicin C and calicheamicin. Finally, DNA damage produced by both esperamicin C and calicheamicin in cells is suppressed within the putative NF1 site but enhanced in the GC-boxes. I will now discuss these results in light of the altered DNA accessibility, dynamics, and conformation caused by binding of these transcription factors.

Previous studies have demonstrated that the binding of transcription factors to their cognate sequences significantly alters the local environment of the DNA [108]. These alterations include steric hindrance [41], constraints on DNA dynamics [117], and changes in DNA conformation [118]. The ubiquitous transcription factor Sp1 regulates a variety of genes by binding to the GC-boxes *in vivo* [112]. Sp1 family proteins belong to the zinc-finger family of DNA binding proteins [119] that bind to the major groove of DNA and induce an asymmetric bend directed towards the major groove, with a bend center located towards the 3' end of the GC box [113]. Another family of transcription factors, CCAAT/enhancer-binding proteins (C/EBP), bind to the consensus CCAAT sequences [111]. They belong to the leucine zipper family and are proposed to bind to the major groove of DNA by a "scissors grip" model: the paired set of basic peptides track DNA in opposite directions along

the major groove, forming a molecular clamp around DNA [110]. The C/EBP family of proteins were shown to induce a small directed DNA bend of 1-4 degrees toward the minor groove [120]. Finally, the DNA binding behavior of the NF1 family of proteins was studied *in vitro* by footprinting with dimethyl sulfate (for G contacts), ethylnitrosourea (for phosphate contacts), and potassium permanganate (for T contacts) [47]. The NF1 protein almost completely surrounds the double helix, establishing a large number of contacts with the bases and backbone [47].

The DNA damage results obtained with calicheamicin and esperamicins are consistent with the binding modes of these transcription factors, which is discussed in more detail below.

**Damage modulation within transcription factor binding sites.** Within the boundaries of the putative transcription factor binding sites [53, 80], damage produced by the intercalator esperamicin A1 is reduced in cells compared to naked DNA. This may be explained by steric hindrance and/or constraints on DNA dynamics upon protein binding. This is analogous to the results with nucleosomes in inactive chromatin, in which damage produced by esperamicin A1 is suppressed in the core DNA (Chapter 3). Reduced interaction with DNA intercalators has been observed both in transcription factor-DNA complex [117] and in nucleosome core DNA [29-32].

On the other hand, for minor groove agents calicheamicin and esperamicin C, binding of proteins to the major groove might not inhibit drug-DNA interactions. Cellular damage in the core of the putative NF1 binding sites is reduced for both calicheamicin and esperamicin C, consistent with NF1 contacts in both the major and minor grooves [47]. This damage inhibition is not observed in the CCAAT site and GC-boxes, consistent with the major groove binding of these two factors.

Furthermore, I observed an enhancement of drug-induced DNA damage by the two minor groove agents in the putative Sp1 binding sites (GC-boxes) *in vivo* (Figures 4.3B and 4.3C). One possible explanation is that Sp1 binding bends the DNA towards the major groove [113], making it a better target for calicheamicin. Our previous studies with reconstituted nucleosomes revealed a 4-fold increase in calicheamicin-induced damage at a site of sharp DNA bending near the dyad axis of the nucleosome [76]. This site is located with its minor groove facing away from the histones [76], therefore the direction of the bend is towards the major groove, similar to the type of bend in Sp1-DNA complexes. Alternatively, protein-induced DNA bending may simply create an optimal minor groove width for enediyne binding, since minor groove width was shown to be important in target selection by both calicheamicin [77, 121] and esperamicin [122].

#### **Damage enhancement between transcription factor binding sites.**

While effects on DNA damage within protein binding sites varied according to structures of enediynes and protein-DNA complexes, the hypersensitivity of damage in sequences flanking transcription factor binding sites was observed with both groove binding and intercalating enediynes. This indicates that such damage enhancement is not dependent on the presence or absence of an intercalator but rather is more likely related to some other effect of the protein-DNA interactions. Pfeifer et al. observed that these same regions in PGK1 were hypersensitive to DNase I [80]. Two possible mechanisms may explain these observations: (1) the regions are free of proteins and thus more accessible to genotoxins than the rest of the DNA; or (2) DNA in these regions has a unique conformation targeted by enediynes and DNase I. Although these two mechanisms need not be mutually exclusive, the fact that not all sequences flanking transcription factor binding



sites are hypersensitive to enediyne- and DNase I-induced damage favors the second mechanism.

What could be the unique conformation targeted by these genotoxic agents? From structural studies of the enediyne- and DNase I-DNA complexes, it appears that the minor-groove width and depth are important parameters affecting the interactions of calicheamicin [121], esperamicin [122] and DNase I [123, 124] with their recognition sequences. In addition, DNA bending is important for the interactions of calicheamicin [78] and DNase I [124] with their DNA targets. It is known that the transactivating domains of many bound transcription factors interact with each other, causing the intervening DNA to form a loop [45]. These interactions put mammalian genes under multiple controls for the level of transcription activity [45]. Such looping may induce DNA bending, change minor groove width and possibly unwind the helix in regions between transcription factors. Calicheamicin [78] and DNase I [125] target bent DNA, DNA intercalators prefer underwound DNA [126-129], while all three enediynes and DNase I require a widened minor groove for favorable binding [121-124]. Therefore, similar to the enhanced damage observed within Sp1 binding sites, enhanced DNA damage in flanking sequences is consistent with a model in which target accessibility, DNA dynamics, DNA bending and minor groove width are important factors affecting enediyne target selection *in vivo*.

**Both sequence selectivity and chromatin structure determine enediyne target selection *in vivo*.** The results with the human PGK1 indicate that esperamicin A1 and C share similar sequence selectivities (Figures 4.1 and 4.3). Calicheamicin, with its extended sugar side chain, displays a higher degree of sequence selectivity [116] that results in fewer damage sites and at different locations (Figure 4.3). The sequence selectivity of calicheamicin is

determined by its sugar side chain [130], which targets flexible DNA sequences and bend the DNA upon binding [131].

Several examples in the studies illustrate that enediyne target selection *in vivo* is determined by both sequence selectivity and chromatin structure. The first example is in regard to the calicheamicin damage hotspot in the ~400 bp PGK1 promoter region studied. In naked DNA, one hotspot of calicheamicin damage was detected at position -224. In cellular DNA, two damage hotspots are detected: one at position -224 and the other at position -165, both of which lie in the region between GC-box and NF1-binding sites (Figure 4.3C). Position -224 is consistent with a common feature of calicheamicin damage sites: the drug targets the 3'-end of a purine tract (Figure 4.4). According to Salzberg *et al.*, this target selection is due to the preference of calicheamicin for bent or flexible sequences [78, 131]. Position -165 does not lie at the end of a purine run and is a weak damage site for calicheamicin *in vitro*. However, damage in this site is increased *in vivo*. The site is located in the flanking region near the end of the NF1 site (Figure 4.3C). Therefore, both the *in vitro* sequence selectivity and *in vivo* chromatin structure may determine the location of calicheamicin damage hotspots in cells.

The second example of the importance of both sequence context and chromatin structure comes from a comparison of DNA damage within the two GC-boxes. An increase in damage by calicheamicin was observed only in the downstream GC-box (Figure 4.3C). Esperamicin C, which has different sequence selectivity than calicheamicin, exhibit an increase in damage in both GC-boxes *in vivo* (Figure 4.3B). Sp1, which binds to GC-boxes, can tolerate considerable sequence variations while retaining function as well as three-dimensional structure [119]. The two GC-boxes, although containing the

“GGGCGG” consensus sequence [112], have different sequence contexts. The fact that the structure and dynamics of DNA are significantly affected by flanking sequences [132] may explain the lack of enhancement of calicheamicin-induced damage in the upstream GC-box. Analogous to this situation, Yu *et al.* found that histone-induced DNA bending alone cannot be responsible for defining the damage site in nucleosomal DNA, since there was no similar increase in calicheamicin damage one helical turn on the other side of the nucleosome dyad [76]. Calicheamicin thus appears to select a combination of both sequence-dependent conformation and protein-induced structural alteration.

**In vivo applications of enediynes: footprinting and transcription modulation.** The results of the present studies suggest that enediynes may prove useful as footprinting reagents for the study of chromatin structure and as modulators of transcription factor binding and functioning.

Along with the previous results with nucleosomes, the suppression of cellular damage by esperamicin A1 within transcription factor binding sites indicates that this drug is an effective footprinting agent for defining chromatin structure in intact cells. As discussed earlier (Chapter 3), enediynes offer several advantages over other agents. The main advantage is their utility with intact cells, since treatment of cells with enediynes does not disrupt the architecture of the cell or nucleus. Also, the resulting DNA double strand breaks are directly ligatable to modified linkers for the LMPCR procedure developed in Chapter 2.

With regard to the modulation of transcription factor binding and functioning, my studies suggest that unlike DNA intercalators, the minor groove agents can still interact with DNA within the transcription factor-DNA complex. In fact, the interactions between esperamicin C and

calicheamicin with DNA in the GC-boxes are enhanced *in vivo* (Figures 4.3B and C), suggesting that these agents could be used to inhibit transcription factor binding and function. The oligosaccharide portion of calicheamicin was shown to inhibit both the binding of transcription factors and transcription [133]. The dimerized form of the calicheamicin oligosaccharide was also synthesized and demonstrated to interfere with transcription factor function [130]. The enediyne portion of the antibiotic contributes to the drug-DNA interaction by increasing binding energy and widening the spectrum of the target sequence [116]. The results from the present studies suggest that at ~10  $\mu$ M concentrations, calicheamicin binding is enhanced within the Sp1 binding site *in vivo*. This concentration is lower than most of the *in vitro* concentrations used to demonstrate the inhibitory effect towards transcription factor binding in the earlier studies with calicheamicin oligosaccharides [130, 133]. Therefore, the deactivated form of the total calicheamicin, which bind to its target DNA with the same sequence selectivity and energy as the parent calicheamicin [116], may be a better transcription inhibitor.

In conclusion, I have found that enediyne-induced DNA damage in the active human PGK1 gene *in vivo* is modulated by bound transcription factors. Transcription factors suppressed damage by esperamicin A1, but only minor groove binding transcription factors suppressed damage by esperamicin C and calicheamicin. Some major groove binding factors even enhanced damage caused by esperamicin C and calicheamicin. Enhancement of damage was also observed with all three enediynes between transcription factor binding sites. These results are consistent with a model in which target accessibility, DNA dynamics, DNA bending and minor groove width are important factors affecting enediyne target selection *in vivo*.

## **V. Figures for chapter four**

**Figure 4.1.** Sequencing gel analysis of the DNA damage produced by the esperamicins in the active human PGK1 gene *in vitro* and *in vivo*. The left four lanes contain duplicate samples of cellular DNA or isolated DNA damaged by esperamicin C, 20  $\mu$ M and 2  $\mu$ M, respectively. The right four lanes contain duplicate samples of cellular DNA and isolated DNA damaged by 2  $\mu$ M and 0.4  $\mu$ M esperamicin A1, respectively. These two different drug-treated samples are separated by a no-drug control lane (lane 0) and by Maxam-Gilbert TC and GA sequencing markers. The LMPCR-amplified DNA was resolved on an 8% sequencing gel, electroblotted onto a nylon membrane, and hybridized with a human PGK1 probe. The positions in the PGK1 gene are shown in the right margin and the positions of putative transcription factors are denoted by open rectangles; transcription factor positions were defined by Pfeifer and coworkers by DMS and DNase I footprinting [53, 80, 115]. Sites of enhanced damage *in vivo* are denoted with asterisks.



**Figure 4.2.** Sequencing gel analysis of the DNA damage produced by esperamicin A1 and calicheamicin in the active human PGK1 gene *in vitro* and *in vivo*. Cells were treated with 14 or 8  $\mu\text{M}$  calicheamicin and 2  $\mu\text{M}$  esperamicin A1 (duplicate samples), while isolated DNA was treated with 0.4 and 0.1  $\mu\text{M}$  calicheamicin and 0.4  $\mu\text{M}$  esperamicin A1 (duplicate samples). Lane 0 is the no-treatment control, and TC and GA lanes are Maxam-Gilbert sequencing reactions. The LMPCR-amplified DNA was resolved on an 8% sequencing gel, electroblotted onto a nylon membrane, and hybridized with a human PGK1 probe. The positions in the PGK1 gene are shown in the right margin and the positions of putative transcription factors are denoted by open rectangles. Sites of enhanced damage *in vivo* are denoted with asterisks. Please note that a lane between the calicheamicin- and esperamicin A1-damaged samples was removed for clarity.



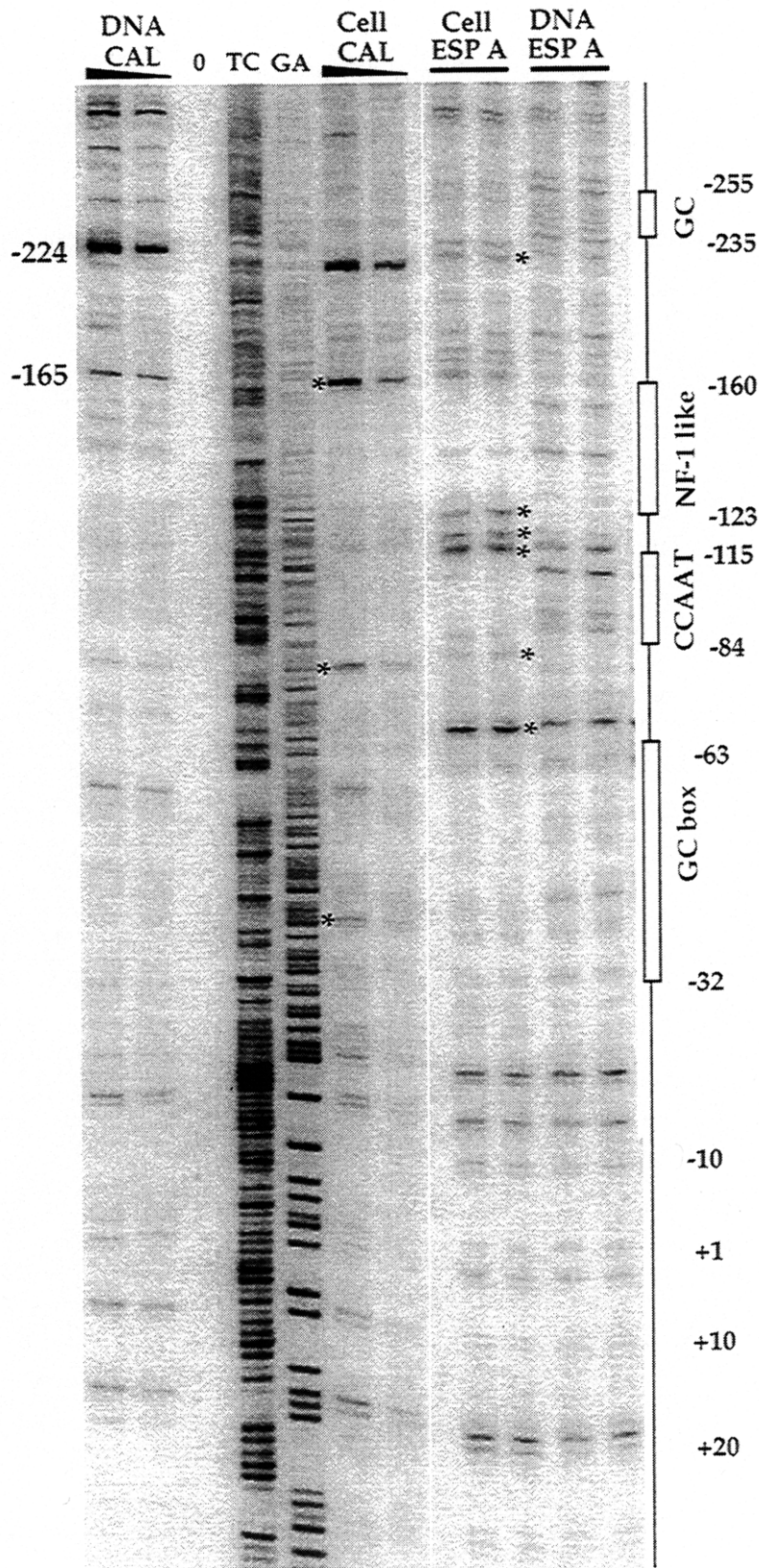


Figure 4.2

**Figure 4.3.** Comparison of enediyne-induced DNA damage in the active human PGK1 gene. The gels shown in Figures 4.2 to 4.3 were subjected to phosphorimager analysis and the data are presented as overlaid line graphs of relative damage frequency along the human PGK1 gene, as described in Materials and Methods. Black lines represent damage in isolated DNA and gray lines represent damage in cells. The position in the PGK1 gene and the proposed positions for transcription factors are indicated below each graph in panels A (esperamicin A1), B (esperamicin C) and C (calicheamicin). Damage enhancement sites in cellular DNA compared to naked DNA are indicated by asterisks. Two *in vivo* calicheamicin damage hotspots, one at position -224 and the other at -165, are also indicated in panel C. The data for esperamicin C from position -63 to +1 in panel B are derived from additional sequencing gels (gels not shown).

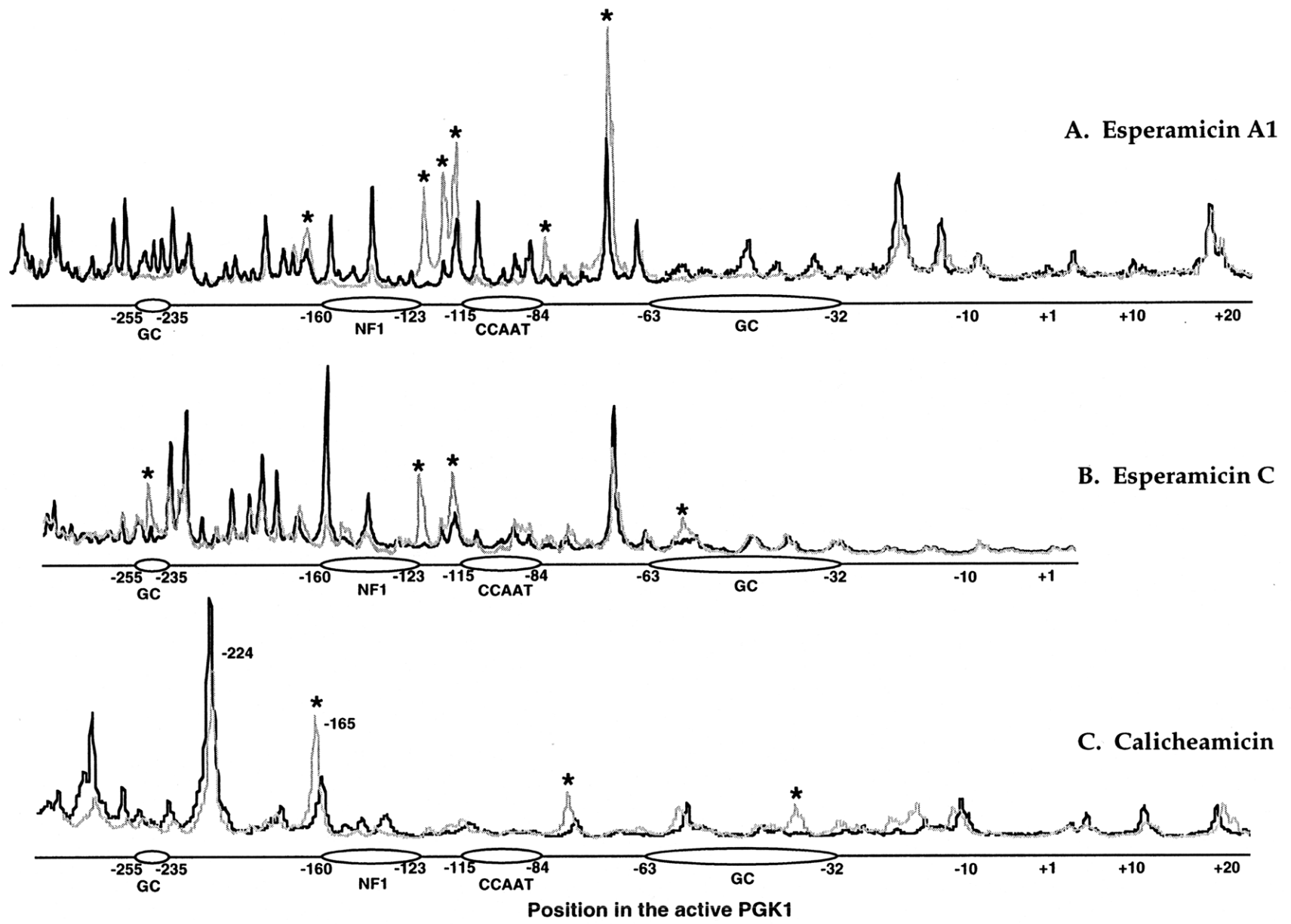


Figure 4.3

**Figure 4.4.** Sequence context of calicheamicin damage hotspots. The hotspots for calicheamicin damage in human PGK1 are listed together with a collection of calicheamicin recognition sequences from the literature. Clearly, calicheamicin has a tendency to recognize the 3'-ends of purine runs (underlined). Position -165 is the damage hotspot in cellular DNA but not in isolated DNA. Both positions -224 and -165 lie between the GC-box and the NF1-site in the active PGK1 *in vivo*.



## **CHAPTER 5**

### **THE ROLE OF DNA METHYLATION IN ENEDIYNE TARGET SELECTION**

## I. Introduction

Modification of mammalian DNA by methylation at the C5 position of cytosine plays a critical role in both the regulation of gene expression [134-136] and in cancer [137]. Given the known effects of cytosine methylation on DNA structure and dynamics, I have examined its effect on DNA damage produced by two structurally-related enediyne antitumor antibiotics, esperamicins A1 and C [56, 65, 138].

Cytosine methylation occurs in the underrepresented CG dinucleotides in eukaryotic cells. While 1-2% of the genome consists of clusters of nonmethylated CG sequences, mainly in the 5' regions of certain genes, approximately 70% of CG dinucleotides are methylated [135]. The general observation is that transcriptionally active regions are undermethylated in comparison to their inactive counterparts in other cells. For the human PGK1 gene, all the CpG dinucleotides are methylated in the inactive copy while none is methylated in the active copy [53]. Thus, the human PGK1 is a good model to examine the effect of DNA methylation on genotoxin-induced DNA damage. There are exceptions to the association of methylation with transcriptional silencing since methyl cytosine also occurs in active genes such as p53 [139].

Although the mechanism by which cytosine methylation causes changes in gene expression is unclear, it is believed to involve differential recognition of methylated DNA sequences by proteins due to changes in DNA conformation or by a direct methyl group-protein interaction ([12], Chapter 1). Regarding the former hypothesis, cytosine methylation causes changes in DNA structure and dynamics that include helical unwinding [15], increased base stacking and helical stability [16], reduction in major groove charge density near the methyl group [17] and, in certain sequence contexts,

modulation of DNA bending [18-20]. These effects may account for the propensity of methylated CG repeats to induce the formation of Z-DNA [140].

The changes in DNA structure and dynamics associated with cytosine methylation also affect the interaction of small molecules with DNA. Both benzo[a]pyrene diol epoxide [21] and mitomycin C [22] show enhanced reactivity with methylated sequences while damage produced by bleomycin [23] and N-methyl-N-nitrosourea [24] is inhibited by cytosine methylation. However, the molecular basis for this altered reactivity remains unclear.

To better understand the role of cytosine methylation in the selection of DNA targets by small molecules, I have undertaken studies with two enediynes, esperamicins A1 and C. Enediynes are extremely potent cytotoxins that produce high levels of double-strand DNA damage by forming a benzenoid diradical intermediate that binds in the minor groove and abstracts hydrogen atoms from the deoxyribose backbone (Chapter 1). While they share a common mechanism for damaging DNA, enediynes differ in the structure and arrangement of functional groups attached to the enediyne core. For example, esperamicin A1 binds in the minor groove of DNA with intercalation of its anthranilate moiety [71, 75], while removal of the sugar-anthranilate to produce esperamicin C results in a shift in the chemistry of the DNA damage [71, 141]. In spite of this structural difference, esperamicins A1 and C have similar preferences for damaging CG dinucleotides among other recognition sequences [57]. These two analogs thus provide an opportunity to investigate the relationship between cytosine methylation and DNA intercalation.

I found that cytosine methylation in CG sequences increases the reactivity of esperamicin A1. However, it was also observed that damage produced by esperamicin C, the nonintercalating analog of esperamicin A1,



was enhanced by DNA methylation at multiple methylated but not single methylated CpG site noted with esperamicin A1. The results with the esperamicins illustrate the complexity of methylation-induced alterations of DNA reactivity and suggest that the altered reactivity extends beyond the CG sequence.

## II. Materials and Methods

**Materials and cell lines.** Esperamicin A1 was generously provided by Dr. Jerzy Golik, Bristol-Myers Squibb. Esperamicin C was prepared by acid catalyzed methanolysis as described elsewhere [71]. Chinese hamster-human hybrids containing an active (cell line Y162-11C) or inactive (cell line X86T2) human X chromosome were provided by Dr. Stanley Gartler (University of Washington, Seattle) [52]. Cells were grown as a monolayer in RPMI 1640 with 10% fetal calicheamicin serum and 40 µg/ml Gentamicin [52]. A plasmid (pBSHPGK1) containing a fragment of the upstream region of the human phosphoglycerate kinase gene (PGK1) was provided by Dr. Judith Singer-Sam, City of Hope [102].

**Treatment of purified DNA with enediynes.** Genomic DNA was purified using a QiaAmp blood kit and dissolved in 50 mM HEPES, 5 mM EDTA, 10 mM glutathione (pH 7.0) at 0.1 mg/ml. An aliquot of calicheamicin or esperamicin in methanol was added (final methanol concentration <1%), and the reaction was allowed for 30 min at 37°C. The drug-damaged DNA was treated with putrescine as described in the previous Chapters. DNA samples were then ethanol precipitated and redissolved in 10 mM HEPES, 0.1 mM EDTA (pH 7.7) at 1 mg/ml.

For DNA sequencing reactions, purified genomic DNA was concentrated by ethanol precipitation to 5-10 mg/ml. Maxam-Gilbert sequencing reactions were performed according to an LMPCR-optimized protocol [91]. 5-methyl-cytosine is resistant to Maxam-Gilbert sequencing reactions [87]. Thus, a gap in the sequence ladder of the methylated DNA as compared to non-methylated DNA indicates sites of DNA methylation [86].

**LMPCR and data analysis.** We used the following primers in the 5' region of the human PGK1 gene: AAGTCGGGAAGGTTTCCTT (-238 to -221,

primer A1); AAGGTTCCCTTGCGGTTTCGCGGCG (-230 to -208, primer A2); CGGCGTGCCGGACGTGACAAAC (-212 to -191, primer A3). The sequences of the blunt linker and linker primer were as described elsewhere [85]. Conventional LMPCR (described in Chapter 2) was used to amplify DNA sequencing reactions as well as drug-treated DNA, since it amplifies both SS and DS damage.

After PCR amplification, half of each sample was resolved on an 8% sequencing gel. Electroblothing and hybridization were performed as described before [95]. The hybridization probe was made by repeated primer extension from a cloned PGK1 template (pBSHPGK1) using primer A3 and *Taq* polymerase [103]. Hybridized membranes were subjected to phosphorimager analysis (Molecular Dynamics).

Phosphorimager analysis was performed essentially as described elsewhere [142]. I accounted for minor differences in the levels of DNA damage in different DNA samples and for lane-to-lane variation in gel loading. To do this, I normalized the phosphorimager signal intensities in each lane so that damage in both DNA substrates was the same at cytosine -143. This cytosine resides in a run of pyrimidines (bottom of the gel in Figure 5.1), and is not methylated in either DNA substrate. The validity of the normalization process is illustrated by two points. First, the expectation of equivalent amounts of damage in the non-methylated regions has firm foundations in our previous studies [142]. Second, normalization of the data to this non-methylated damage site resulted in the same damage in either substrate in virtually every other non-methylated sites (Figure 5.2).

### III. Results

The effect of cytosine methylation on esperamicin-induced DNA damage was studied in the promoter region of the human PGK1 gene. In Figure 5.1, both the Maxam-Gilbert sequencing markers and drug-treated samples are amplified by conventional LMPCR techniques to examine both single-strand and double-strand breaks produced by enediynes. An additional band in the C-reaction lane of the unmethylated DNA confirms the presence of methylated Cs in the methylated DNA, as methylated Cs are resistant to Maxam-Gilbert reactions [87]. The positions of the methylated Cs are indicated by the asterisks next to the C-reaction lane (Figure 5.1). At cytosine position -125, which is in the middle of several methylatable CG-repeats, DNA damage by esperamicin A1 and C are enhanced in the methylated DNA compared to their unmethylated counterpart. Calicheamicin-induced lesions are also enhanced in methylated DNA, although this is a relatively weak damage site for calicheamicin.

Quantitative comparisons of these damages are expressed in Figure 5.2. To account for minor differences in lane-to-lane variation in gel loading, we normalized the phosphorimager signal intensities in each lane so that damage in the non-methylated cytosine -143 in both DNA substrates was the same. Methylation-dependent increase of esperamicin A1-induced damage is apparent in cytosine positions -125 and -99 (panel A). There is also a methylation-dependent enhancement at -119, although the LMPCR signals in this site is rather weak. Esperamicin C-induced damage is enhanced in multiple methylated cytosines -125 and -119, but not in singly-methylated cytosine -99 (panel B). The enhancement of damage is observed only at methylatable sites, since damages are the same at sites where there is no methylation differences. One anomaly is at guanine -100, in which LMPCR

signal is apparently enhanced in the unmethylated substrate (Figure 5.1.). This difference was not observed with direct DNA damage (discussed below).

To control that the observed differences are indeed caused by the drugs and not by the LMPCR procedure, we have also studied the direct DNA fragmentation caused by esperamicins in a cloned fragment of the human PGK1 gene [142]. Cytosines in CG sequences were uniformly methylated using *Sss* I methylase and verified by Maxam-Gilbert sequencing reactions [143]. Both non-methylated and methylated PGK1 fragments were 5'-end labeled by  $^{32}\text{P}$ , and DNA damage were directly visualized on sequencing gels [142]. In Figure 5.3, damage patterns examined by LMPCR is shown above the PGK1 sequences (as in Figure 5.2), while damage patterns detected by direct sequencing gel analysis of the PGK1 fragment is indicated underneath the nucleotide sequences. It is evident that: (1) the patterns of damage identified by either method are very similar, especially the sites of damage are the same; (2) methylation-dependent enhancement of damage at positions -125, -119 and -99 for esperamicin A1 and at -125 and -119 for esperamicin C are consistent. There is a 1.5 to 2-fold increase in damage produced by esperamicin A1 at both CG repeats and a single CG sequence, while damage enhancement with esperamicin C are observed at CG repeats only. One difference between the two methods was observed at position -119, however. The LMPCR produced a very weak signal at this site compared to direct damage. The LMPCR signal of the Maxam-Gilbert sequencing maker is also weak at this site (Figure 5.1).

#### IV. Discussion

The role of DNA methylation in the selection of DNA targets by small molecules has been the subject of several studies, e.g., with benzo[a]pyrene diol epoxide [21], mitomycin C [22], bleomycin [23], and N-methyl-N-nitrosourea [24]. Here I demonstrate that methylation enhances the damage produced by two equilibrium-binding DNA-cleaving molecules, esperamicins A1 and C. Both enediynes bind in the minor groove and share similar sequence selectivities. However, they differ structurally in that esperamicin C does not possess the intercalating anthranilate of esperamicin A1. The results with these structural analogs illustrate the complexity of methylation-induced alterations of DNA reactivity.

There are several effects of cytosine methylation on DNA structure that could influence the interactions of small molecules with DNA. These include helical unwinding [15], increased base stacking and helical stability [16], reduction in major groove charge density near the methyl group [17] and, in certain sequence contexts, modulation of DNA bending [18-20]. Methylation-induced torsional flexibility or helical unwinding could explain the increased binding of esperamicin A1 since intercalators unwind the helix upon binding and prefer to bind to the underwound helix of negatively supercoiled DNA [126-129]. This effect would be consistent with the observation of Denissenko *et al.* that cytosine methylation increases adduct formation by benzo[a]pyrene diol epoxide [21]. However, plasmid unwinding studies with esperamicin C suggest that the drug does not alter DNA twist [71]. Esperamicin C would thus not be expected to be sensitive to methylation-induced changes in DNA twist unless such changes affected other features of DNA structure, such as minor groove width, thereby creating a more attractive binding site for the drug. A similar argument could

also be made for the methylation-induced increase in cross-linking by the nonintercalating mitomycin C [22].

In studies with mitomycin C, Johnson *et al.* proposed that the enhancement of mitomycin C cross-linking by cytosine methylation is due to an electronic effect transmitted by hydrogen-bonding between the methyl group and the N2 of guanine [22]. It could be argued that such local alterations in minor groove charge density increase the binding of esperamicins A1 and C at position -125 in the PGK1 sequence (Figure 5.3). However, this could not be the case at position -99 (Figure 5.3) where esperamicin A1-induced damage is enhanced by cytosine methylation while damage produced by esperamicin C is not.

Alternatively, axial flexibility or curvature could be involved in the enhancement of esperamicin C-induced DNA damage by methylation. We have observed that adenine methylation at the 3'-end of a target sequence for calicheamicin, a related enediyne, increases drug-induced DNA damage by 1.5- to 2-fold (unpublished observations), an effect that may be related to the propensity of calicheamicin to bend its target sequences [78]. While esperamicin C is a structural analog of calicheamicin missing the terminal sugar and aromatic ring, it does not share calicheamicin's selectivity for the 3'-end of purine tracts [77, 78].

Regardless of the local changes in DNA conformation caused by DNA methylation, the effects appear to be transmitted along the helix over at least several base pairs. Kim *et al.* have observed that CG repeats exert long-range influences on DNA secondary structure [144], and it is possible that cytosine methylation affects these long-range structural changes. Furthermore, Hertzberg *et al.* demonstrated that the effect of methylation on bleomycin cleavage can extend as far as 14 base pairs away from the methylated cytosine

[23], while Hodges-Garcia and Hagerman observed that cytosine methylation affects curvature of A-tracts up to three base pairs away [19]. Eneidyne are predicted to bind to DNA up to four base pairs away from their DNA damage sites [75], in this case methylated CG sites. This suggests that the methylation-induced changes in DNA conformation extend beyond the CG site to effect drug binding.

The effects of methylation on the conformation of DNA targets of esperamicins A1 and C also appear to be dependent on sequence context, most notably at the single CG site (-99 in Figure 5.3). At this site, damage produced by esperamicin A1 is enhanced, but there was no enhancement of esperamicin C-induced damage. Such sequence-dependent effects have also been noted in studies of methylated cytosine and DNA curvature [18, 19, 145].

My comparison of the data by the LMPCR and the direct damage procedure indicated that the patterns of damage identified by either method are very similar; especially the sites of damage and the degrees of methylation-dependent damage enhancement are the same (Figure 5.3). However, LMPCR produced a very weak signal at -119 compared to direct damage. The LMPCR signal of the Maxam-Gilbert sequencing maker is also weak at this site (Figure 5.1), suggesting this is an LMPCR-related artifact. Another anomaly was observed at guanine -100, in which the LMPCR signal intensity by esperamicin C is enhanced in unmethylated substrate. Such enhancement is not found in the direct damage studies (Figure 5.3). This may be caused by the LMPCR efficiency differences between the two substrates at this particular site. Therefore, caution should be taken when comparing two different damage sites or two differentially-methylated DNA substrates (notice that methylation at cytosine -99 occurs adjacent to guanine -100). The verification of methylation effects by direct damage studies served its



necessary purpose in this regard. However, all of my previous LMPCR studies were based on comparisons of damage in naked DNA and in cells. LMPCR was performed on DNA substrates of the same methylation status, and conclusions were made from comparing the same damage sites. Therefore, the rare LMPCR anomaly observed here with different damage sites or different methylation states should not be a problem in those studies.

In conclusion, the present studies broaden the repertoire of DNA-damaging molecules for which target selection is affected by DNA methylation. The results with esperamicins A1 and C suggest that the rules governing the effects of cytosine methylation on DNA structure and dynamics are complex and remain to be elucidated. Future studies in our laboratory will address the mechanism of cytosine methylation in enhancing enediyne-induced DNA damage.

## V. Figures for chapter five

**Figure 5.1** The effect of cytosine methylation on enediynes-induced DNA damage in the promoter region of the human PGK1 gene. Both the Maxam-Gilbert sequencing markers and drug-treated samples were amplified by conventional LMPCR techniques. Damages in isolated X86T2 DNA (methylated) and Y162-11C DNA (unmethylated) were produced by three enediynes as indicated above each lanes of drug-treated samples. The "methyl +" and "methyl -" signs indicate methylated and unmethylated DNA, respectively. Drug concentrations: esperamicin A1 (1  $\mu$ M), esperamicin C (2  $\mu$ M), and calicheamicin  $\gamma$  (1  $\mu$ M). TC, GA and C lanes are Maxam-Gilbert sequencing markers. An additional band in the pyrimidine-reaction lane of the unmethylated DNA confirms the presence of methylated cytosines in X86T2 DNA, since methylated Cs are resistant to Maxam-Gilbert reactions [87]. Therefore, an additional bands in lanes 4 vs. 3 indicates methylated cytosines, which are marked by the asterisks to the left of the C-reaction lane. The amplified DNA was resolved on an 8% sequencing gel, electroblotted onto a nylon membrane, and hybridized to a human PGK1 probe. The positions in the PGK1 gene is shown in the left margin, and the nucleotide sequences are shown between the first two sequencing marker lanes. At cytosine position -125 in the middle of several methylatable CG-repeats, DNA damage by all three enediynes are enhanced in the methylated DNA compared to their unmethylated counterpart, although this is a relatively weak damage site for calicheamicin.

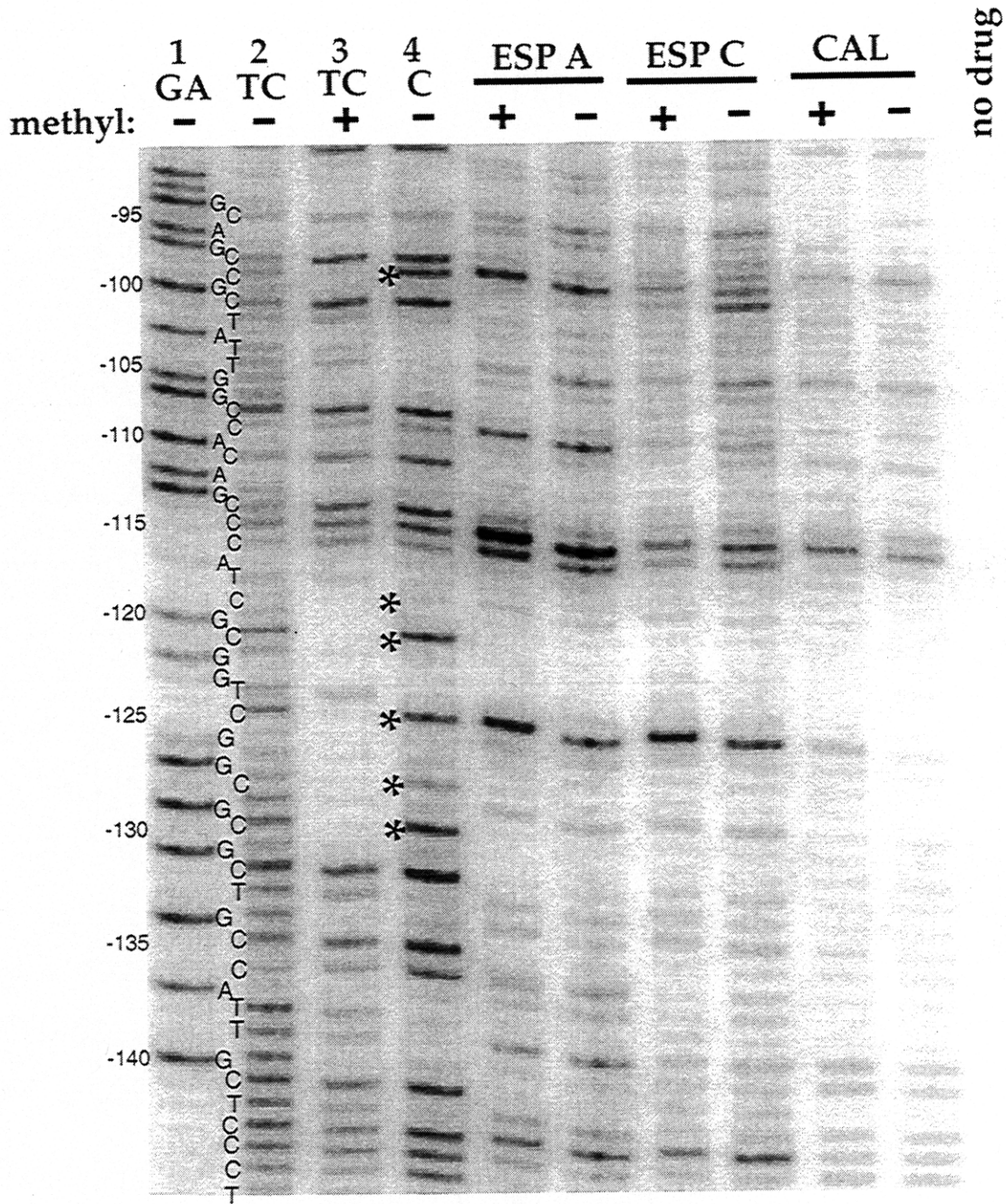
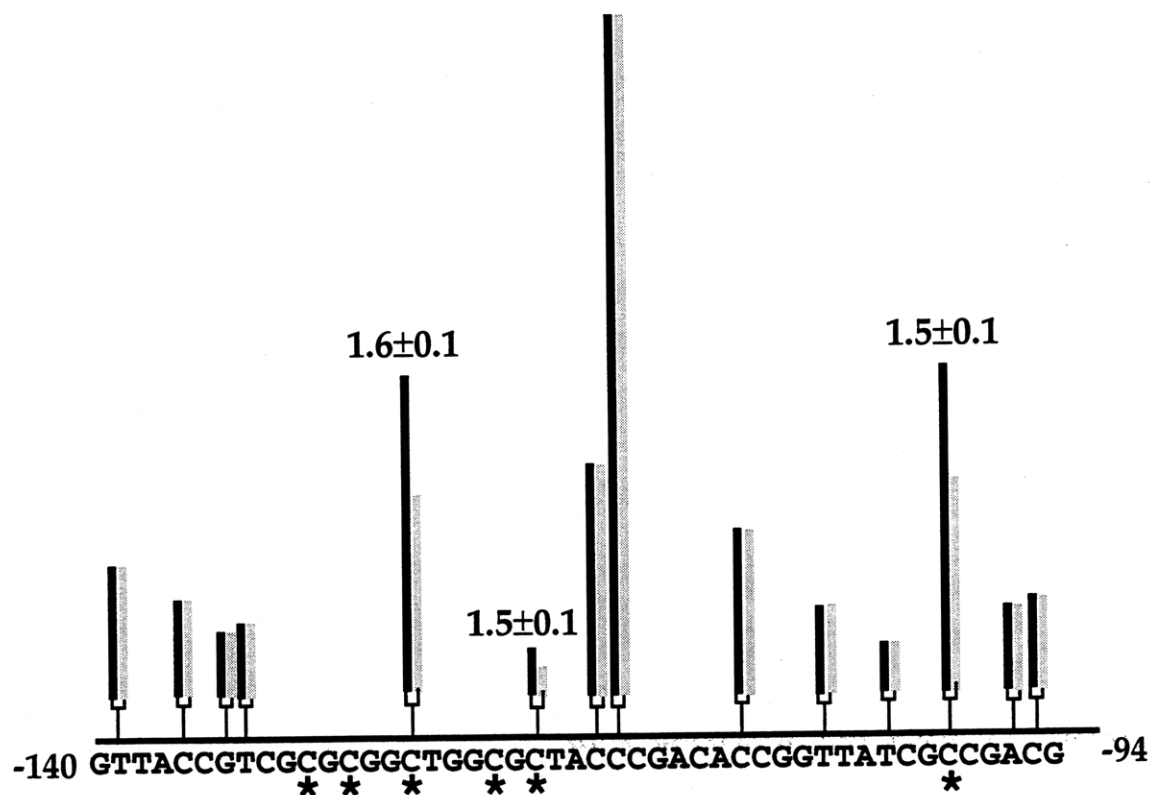
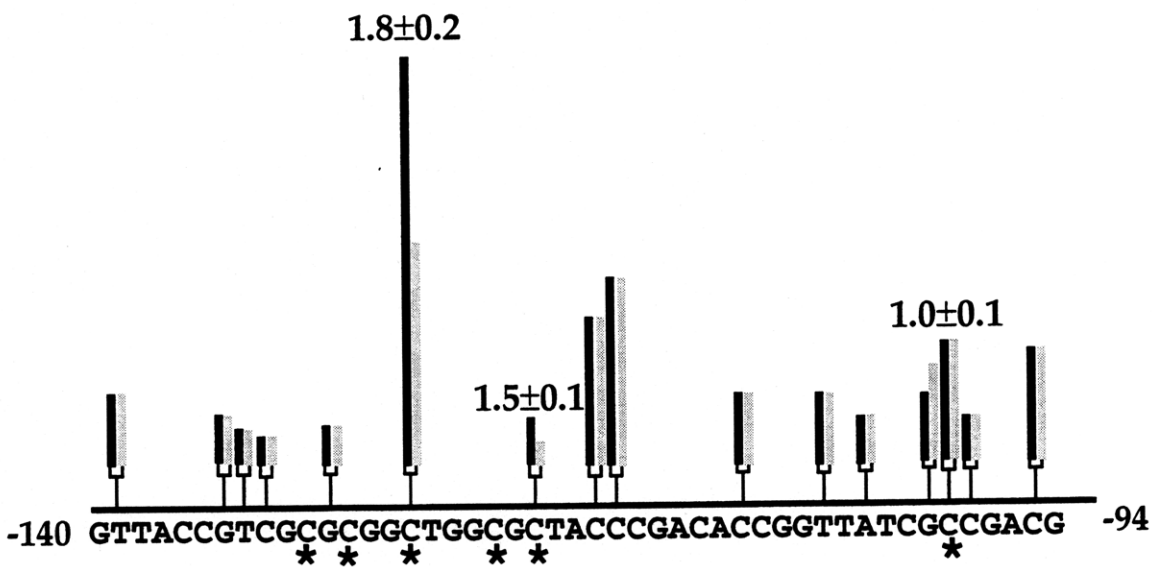


Figure 5.1

**Figure 5.2.** Quantitative comparisons of the effect of methylation on DNA damage caused by esperamicin A1 (panel A) and esperamicin C (panel B). The gel in Figure 5.1 was subjected to phosphorimager analysis. To account for minor differences in lane-to-lane variation in gel loading, signal intensities were normalized so that damage in the unmethylated cytosine -143 (near the bottom of Figure 5.1) in both DNA substrates was the same. Damage intensities were plotted over the human PGK1 sequence between positions -140 to -94. Only the bottom strand of the PGK1 sequence was plotted due to the unidirection of the gene-specific primer. The height of the bar at each position in the PGK1 sequence is proportional to the damage intensity. The black bars and gray bars indicate damage intensities in methylated and unmethylated DNA, respectively. The numbers above the bars represent average ratios of the damage frequency in methylated DNA to that in unmethylated DNA from two independent experiments; the indicated errors represent deviations about the mean for duplicate samples. Methylated cytosines are denoted with asterisks.



A. Esperamicin A1 damage along the PGK1



B. Esperamicin C damage along the PGK1

Figure 5.2

**Figure 5.3.** Frequency of DNA damage produced by esperamicin A1 (panel A) and esperamicin C (panel B). Data from both the LMPCR procedure (presented above the nucleotide sequences) and from direct damage in a cloned human PGK1 fragment (presented underneath the sequences) were included. The LMPCR data above the nucleotide sequences were plotted the same way as in Figure 5.2. The direct damage data utilized a cloned fragment of the human PGK1 gene as DNA substrate and Sss I methylase to methylate cytosines in all CG sequences *in vitro* [142]. Both non-methylated and methylated PGK1 fragments were 5'-end labeled by  $^{32}\text{P}$ , and DNA damage were directly visualized on DNA sequencing gels [142]. Damage intensities were plotted over the human PGK1 sequence between positions -140 to -94. The height of the bar at each position in the PGK1 sequence is proportional to the damage frequency. The black bars and gray bars indicate damage intensities in methylated and unmethylated DNA, respectively. The numbers under the bars in the direct damage data represent average ratios of the damage frequency in methylated DNA to that in unmethylated DNA; the indicated errors are standard deviations for  $n=3$ . Methylated cytosines are denoted with asterisks.

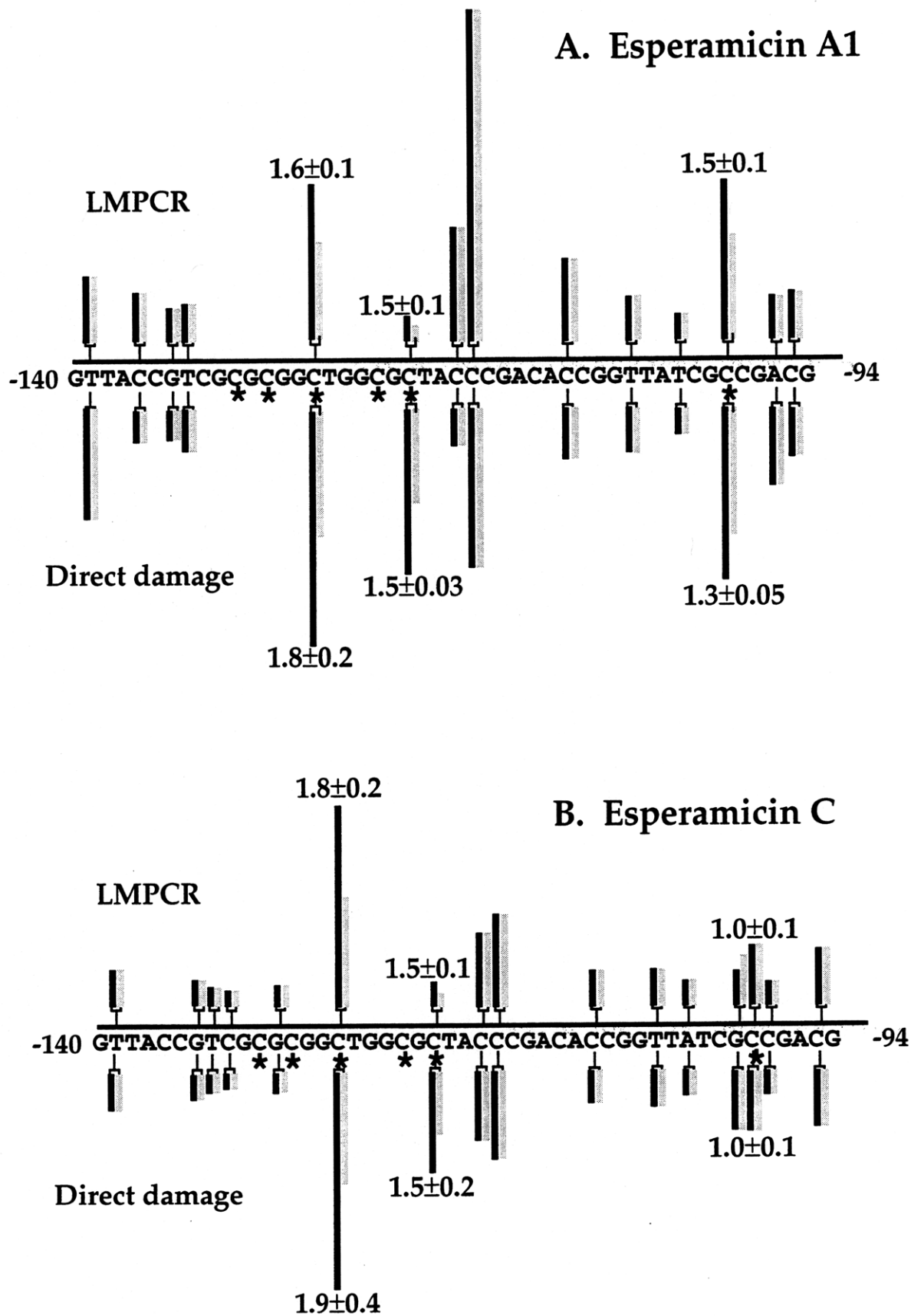


Figure 5.3



## **CHAPTER 6**

### **DEVELOPMENT OF TECHNIQUES TO MAP LOW LEVELS OF DNA BREAKS IN WHOLE CELLS**

## I. Introduction

So far, my approach to studying the genomic targets of DNA-damaging agents makes use of the human PGK1 as a model gene. These studies provide significant insights into the effects of local chromatin structure and DNA methylation on the target selection by genotoxins. However, the studies are limited by two factors. First, even with the power of LMPCR, the observation of gene-specific damage on DNA sequencing gels requires much higher levels of chemical treatment than experienced by cells during chemotherapy or environmental exposure. High drug concentrations could saturate potential hotspots for DNA damage and cause damage in low affinity or low frequency sites. Secondly, the conclusions are limited to the possible unique chromatin structure of a single gene, and do not provide significant clues as to the location of the damage sites in the entire nucleus.

It has become increasingly clear that genotoxic chemicals and other damaging agents such as radiation do not damage DNA randomly in cells [146, 147]. For example, many chemicals attack transcriptionally active DNA more frequently than inactive DNA. Included among these are the anticancer drugs bleomycin and neocarzinostatin [146]; the carcinogenic aflatoxin B1, which targets the nucleolar rRNA genes [148]; and the classical probes of chromatin structure, micrococcal nuclease and DNase I [3]. Therefore, hotspots of DNA damage may exist in the mammalian genome. The ability to map low levels of DNA damage in the nucleus will provide an opportunity to detect DNA damage hotspots in the genome.

To this end, I have developed a technique that is broadly applicable to the mapping of low levels of DNA breaks in the entire genome. Rather than screening a battery of model chromatin structures and genes, we will isolate all damage sites from whole cells subjected to low levels of DNA damage.

The isolated damage sites, collectively and individually, contain a wealth of information about the characteristics of the DNA damage at all levels of genomic organization, from nuclear location to local DNA structure. Strategies to further manipulate these damage sites (amplification and cloning) to form a "genomic damage library" will be discussed.

I have chosen calicheamicin as the model agent for these studies for several reasons. Its unique mechanism of action, involving abstraction of deoxyribose hydrogen by a diradical intermediate, results exclusively in lethal double-strand (DS) lesions with high efficiency [70]. Second, it is a potent genotoxin,  $\sim 10^3$ -fold more toxic than adriamycin [66] and has an  $LD_{50}$  of  $\sim 30$  pM in several human cell lines [66, 67, 149]. Even with the power of LMPCR, I need to treat the cells with  $\sim 1$   $\mu$ M drug concentration in order to resolve DNA damage along the PGK1 gene. The goal of the studies presented in this Chapter is to develop techniques to isolate all DNA damage sites at much lower concentrations than those used for the LMPCR studies.

As shown in Figure 6.1, the technique involves preserving the cellular DNA by agarose embedment, *in situ* labeling of DNA damage with biotinylated nucleosides, and affinity purification of DNA damage fragments with streptavidin-coated magnetic beads (Dynabeads). After drug treatment, cells are embedded in agarose to provide structural support and to prevent mechanical damage to DNA during subsequent steps [150]. I then exploit the 3'-phosphatase activity of T4 polynucleotide kinase (T4 kinase; ref. [151]) to convert calicheamicin-induced 3'-phosphates to 3'-hydroxyl ended fragments. The 3'-ends of DS breaks can finally be labeled with biotinylated nucleosides and terminal deoxynucleotidyl transferase (TDT).

The purification strategy we have devised, involves restriction enzyme digestion to produce manageable-sized fragments, followed by affinity

purification of biotin-labeled fragments by Dynabeads (Figure 6.1). The purified damage fragments can be used to make DNA probes for mapping DNA damage sites, or can be amplified and cloned for the generation of a "DNA damage library".

## II. Materials and Methods

**Materials and cell lines.** Calicheamicin  $\gamma_1^I$  was provided by Dr. George Ellestad (Wyeth-Ayerst Research). HeLa S3 cells were grown as a suspension culture in Joklik modified Eagle's minimum essential medium with 10% newborn calicheamicin serum (Sigma) and 1% penicillin/streptomycin (Sigma). T4 kinase and TDT were obtained from New England Biolabs and Amersham Life Science (formerly United States Biochemical), respectively. Dynabeads (streptavidin-coated magnetic beads) and magnetic separator were purchased from Dynal, Inc.

**Drug treatment and agarose-embedding of cells.** HeLa S3 cells were grown overnight in cell culture media with 1  $\mu\text{Ci/ml}$  of  $^3\text{H}$ -thymidine, then harvested the next day at ~80% confluency and resuspended in phosphate-buffered saline (PBS) at  $10^7$  cells/ml. An aliquot of calicheamicin in methanol was added (final methanol concentration <1%), and the reaction was allowed to proceed for 20 min at 37°C. The drug-treated cells were next embedded in agarose essentially as described before [150]. A suspension of the drug-treated cells (100  $\mu\text{l}$ ) was gently mixed with an equal volume of molten 1.6% low-melting agarose in PBS at 37°C (Type VII agarose, Sigma), and dispersed in a plastic agarose plug mold (BioRad) with pipet tips that have been cut off 5 mm at the ends to minimize mechanical damage to cells. The mixture was allowed to solidify for 10 min at 4°C, after which agarose blocks were gently pushed out of the plug mold onto a glass slide. The agarose block was then sliced into ~1  $\text{mm}^3$  "chops" using a slide cover glass. Such small chops of agarose minimize processing time by permitting rapid diffusion of chemicals and enzymes into the cells.

**DNA purification and labeling in agarose.** Cells embedded in agarose chops were treated with 1% SDS in 10 mM Tris, 30 mM EDTA, pH 8, for 3 hr

at room temperature by gently rotating samples. This treatment lyses membranes, denatures nucleases, and removes a significant portion of the cytosolic and nuclear proteins [152]. However, unlike proteinase K, SDS does not cause strand breaks [152]. SDS was removed by 4 x 10 min washes in 10 mM Tris, 1 mM EDTA, 50 mM NaCl, pH 8 (TES buffer), and one wash in water. Buffer exchanges with agarose chops were facilitated by using screen caps (BioRad) fitted to 50 ml Falcon tubes perforated at the other ends. The chops were then equilibrated with 3'-phosphatase buffer (0.2 M NaCl, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 0.1 M MES, pH 6) twice for 10 min. T4 kinase (20 U) was then added and dephosphorylation proceeded for 1 hr at 33°C. The chops were washed 4 x 10 min in TES and 2 x 10 min in TDT buffer (2 mM CoCl<sub>2</sub>, 0.2 mM β-mercaptoethanol, 100 mM sodium cacodylate, pH 7.2). Biotinylated-dCTP (4 μM, Gibco BRL) and TDT (17 U) were added and the labeling reaction allowed to proceed for 1 hr at 33°C. Labeling was followed by washes in TES (4 x 10 min), RNase A digestion (0.1 mg/ml, 33°C, 1 hr), proteinase K digestion (0.1 mg/ml, 33°C, 1 hr), and washes in TES (4 x 10 min). After labeling, DNA was digested with a restriction enzyme *Dpn* II, a 4-cutter enzyme that cut genomic DNA to manageable-sized fragments (average size ~300 bp). Following digestion, the DNA fragments were removed from the agarose chops by electroelution and stored in 10 mM Tris, 1 mM EDTA, pH 8 until use.

**Assays for the levels of biotinylation.** An aliquot of each DNA sample representing equal amount of DNA (by <sup>3</sup>H count) were spotted on a neutral Nylon membrane (New England Nuclear). The levels of biotinylation were assayed by a PhosphorImager detection kit (New England Biolabs) and following the manufacture's protocol. The chemiluminescent signals were detected by X-ray films and quantitated by a densitometer (BioRad).

**Optimization of DNA binding buffer.** In order to affinity capture minute amount of biotinylated DNA in the vast excess of non-biotinylated DNA, optimal binding buffer is required to achieve maximal specific binding and minimal non-specific binding. A biotinylated 100 bp double-stranded DNA (generated by PCR reactions with a biotinylated primer and a non-biotinylated primer) was 5'-end labeled with  $\gamma^{32}\text{P}$ -ATP and T4 kinase [153]. Non-biotinylated  $^3\text{H}$ -labeled DNA was purified from HeLa cells grown in the cell culture media with 1  $\mu\text{Ci}/\text{ml}$  of  $^3\text{H}$ -thymidine overnight. The biotinylated or non-biotinylated DNA were incubated with sufficient Dynabeads under various test buffers at room temperature for 30 min by gently rotating samples. Washes in the same test buffer were performed by flipping the samples 10 times in room temperature. Dynabeads were concentrated on one side of the tube by the magnetic separator and each supernatant was transferred to a new tube. Half of each supernatant was mixed with 10-time excess volume of Scinti Verse I (Fisher Scientific) and radioactivity was counted by a liquid scintillation counter (Beckman). The washes were repeated until radioactivity in the last three supernatants had reached background levels (i.e., any subsequent washes did not further elute more radioactivity). The final Dynabeads were subjected to scintillation counting, and the result of each sample was calculated as percentage of total radioactivity remaining on the beads.

The presence of detergent, monovalent salt and salt concentrations were among the factors considered in various test buffers. The following buffers were tested: (1) TE (10 mM Tris, 1 mM EDTA, pH 8) + 1 M NaCl + 0.1% Tween 20, (2) 6xSSC (0.9 M NaCl, 0.09 M  $\text{Na}_3\text{citrate}$ , pH 7) + 0.1% Tween 20, (3) TE + 1 M NaCl + 0.1% SDS, (4) 6xSSC, (5) TE + 3 M LiCl + 0.2% Tween 20, (6) 20xSSC (3 M NaCl, 0.3 M  $\text{Na}_3\text{citrate}$ , pH 7) + 0.2% Tween 20.

**Affinity purification of biotin-labeled DNA fragments.** After labeled DNA samples were electroeluted out of agarose chops, they were allowed to bind to Dynabeads for 30 min at room temperature in the optimal binding buffer (10 mM Tris, 1 mM EDTA, 1 M NaCl, 0.1% Tween 20, pH 8) with gentle rotation. Non-biotinylated DNA was removed by repeated washing in the binding buffer, until radioactivity in the last three supernatants had reached background levels. Finally, an aliquot of the beads were used to quantitate the amount of  $^3\text{H}$ -DNA remaining on the Dynabeads by scintillation counting. The counts from calicheamicin-treated samples were then compared with that from control samples.



### III. Results

#### **3'-phosphatase-dependent TDT labeling of calicheamicin damage sites.**

The relative levels of biotin labeling with and without the 3'-phosphatase activity of T4 kinase is presented in Figure 6.2. The control cells embedded in agarose chops displayed equal levels of labeling with and without T4 kinase. This indicates that most of the background DNA breaks have 3'-hydroxyl groups rather than 3'-phosphates at their ends. In the absence of T4 kinase, labeling of calicheamicin-induced damage sites occurred at the level of the control cells. With the 3'-phosphatase activity of T4 kinase, however, the level of biotin labeling increased significantly in cells treated with 1 nM calicheamicin. The kinase-dependent labeling of the drug-treated sample indicates that most of the calicheamicin-damage sites *in vivo* possess 3'-phosphate ends, which is in agreement with our *in vitro* studies [70]. The results of these biotin labeling studies demonstrate that the agarose embedment and kinase/TDT labeling strategy permits labeling of calicheamicin-induced DNA breaks at low drug concentrations.

**Optimization of DNA binding buffer.** In order to achieve maximal specific binding and minimal non-specific binding to drug-induced damage sites, I tested the effects of detergents, monovalent salts and ionic strength on the binding of a  $^{32}\text{P}$ -labeled biotinylated DNA as well as  $^3\text{H}$ -labeled non-biotinylated DNA. Table 6.1 shows the percentages of radioactivity remaining on the Dynabeads after extensive washes with each test buffer. We hypothesized that the presence of detergent would minimize non-specific binding to the beads. Indeed, with a non-ionic detergent Tween 20 at 0.1%, the non-specific binding decreased from 0.0011% in buffer 4 to 0.00007% in buffer 2. Meanwhile, Tween 20 did not sacrifice specific binding. Ironically, the ionic detergent SDS decreased the binding of biotinylated DNA from

~70% to ~20% (compare buffer 1 to 3). The presence of monovalent salts at high concentrations (500 mM or above) was important to minimize non-specific binding (data not shown). However, lithium chloride at 3 M (buffer 5) or sodium chloride at 3 M in 20xSSC (buffer 6) did not enhance specific binding when compared to sodium chloride at 1 M (buffer 1) or 0.9 M (buffer 2). In conclusion, buffers 1 and 2 both afforded similar levels of specific and non-specific binding. Since buffer 1 tended to have longer shelf life due to the presence of EDTA, I decided to use it as the DNA binding buffer.

**Affinity purification of calicheamicin-damage fragments.** The calicheamicin dose-response for both the biotin labeling and Dynabeads purification is presented in Figure 6.3. Clearly, with increasing concentrations of calicheamicin, more biotinylated DNA were captured on Dynabeads. At 1 nM calicheamicin, about 3 times more DNA was captured compared to non-drug treated control. Also at this concentration only about 0.01% of the starting DNA was captured on Dynabeads, thus representing a  $10^4$  fold enrichment of drug-induced damage fragments.

#### IV. Discussion

In this Chapter, I have presented a technique applicable to the mapping of low levels of DNA breaks in the entire nucleus. Rather than screening a battery of model chromatin structures and genes, we are now able to isolate all damage sites from whole cells subjected to low levels of DNA damage. The isolated damage sites contain a wealth of information about the characteristics of DNA damage and can be used collectively or individually to study the distribution of DNA damage at all levels of genomic organization, from nuclear location to DNA fine structure.

The technique involves enzymatic labeling of DNA damage with biotinylated nucleosides and affinity purification of DNA damage fragments with streptavidin-coated magnetic beads (Dynabeads), since the binding affinity between biotin and streptavidin ( $K_d=10^{-15}$  M) is near the strength of covalent bonds [154]. For *in situ* labeling of DNA damage, two techniques have traditionally been employed: nick translation (NT) and 3'-tailing with terminal deoxynucleotidyl transferase (TDT). The former exploits both the 5'-to-3' exonuclease and the polymerase activity of *E. Coli* DNA polymerase I (DNA pol I) to incorporate labeled nucleotides at DNase I-induced nicks [153]. It has been used to study DNase I hypersensitive sites in nuclei and chromosomes [155, 156]. However, NT labels only SS nicks. TDT does not require a template and adds nucleosides to the 3'-ends of DS breaks, and to a lesser extent SS breaks [157]. It has been employed to label apoptotic DNA damage in whole cells (e.g., [158]). Other methods to label damaged DNA exploit cellular repair enzymes to incorporate labeled nucleotides (e.g. [159]). Drawbacks to these approaches include inefficient repair of DS lesions and selective repair of transcribed genes [160, 161]. Because the focus of my study

is DS breaks, I have chosen to use TDT to label calicheamicin damage sites with biotinylated nucleosides.

However, there is an impediment to the enzymatic labeling of oxidative DNA damage: the presence of phosphate and sugar residues on the 3'-ends of the DNA fragments. Indeed, 3'-phosphate-ended fragments represent the bulk (~85%) of DNA damage occurring by calicheamicin and esperamicin in the presence of naturally-abundant polyamines and lysine amino groups *in vitro* [70, 71]. To overcome this limitation, I have exploited the 3'-phosphatase activity of the T4 polynucleotide kinase [151]. The combined strategy of dephosphorylation with T4 kinase and labeling with biotinylated nucleotides and TDT is shown in Figure 6.1A.

The fact that labeling of calicheamicin damage sites was kinase-dependent (Figure 6.2) proves the feasibility of our approach. It also demonstrates that enediyne-induced DNA damage in cells is a direct result of the drug and is not mediated by apoptosis, since apoptotic DNA damage processes strand breaks with 3'-hydroxyl ends (e.g., [162]). It further rules out topoisomerase II as a target for the drugs *in vivo* [163], since drug-induced inhibition of topoisomerase II also results in strand breaks with 3'-hydroxyl groups that would not require dephosphorylation for end-labeling [163].

The success of our strategy relies not only on the specific biotinylation of calicheamicin damage sites, but also on an affinity purification system that can capture biotinylated DNA with minimum background binding. The presence of Tween 20 (0.1%) and monovalent salt (sodium chloride at 1 M) helps to minimize non-specific binding to ~0.00001%, while not sacrificing specific binding (Table 6.1). On the other hand, the ionic detergent SDS (0.1%) decreases specific binding from ~70% to ~20%, probably because SDS partially denatures the streptavidin tetramer thus releasing some biotinylated DNA. It

is known that the binding affinity to biotin is  $10^{-15}$  M for native streptavidin tetramer and only  $10^{-8}$  M for streptavidin monomer [154].

The fact that only up to ~70% of biotinylated DNA remained on the beads in Table 6.1 is a reflection of the biotinylation efficiency of the biotinylated primer used to generate the 100-bp DNA, since it is not uncommon that the biotin modification during oligonucleotide synthesis is only 50 to 70% efficient [164]. When an HPLC-purified biotin-dCTP was added to the 3'-ends of DNA by TDT, 95% of the DNA was captured by Dynabeads in our optimum DNA binding buffer (data not shown).

With our current approach, it is possible to detect the levels of DNA damage at 1 nM for  $10^7$  cells/ml (Figure 6.3). The drug/DNA ratio in this case is similar to the  $LD_{50}$  concentration of calicheamicin (30 pM for  $10^5$  cells/ml, Gentest, personal communication). The purified damage fragments ( $10^4$ -fold enrichment) contain a broad spectrum of calicheamicin-induced DNA damage sites, and can be used to make DNA probes by random-primer extension [165]. These probes can be used for mapping DNA damage at all levels of genomic organization, from nuclear location (by fluorescent *in situ* hybridization) to DNA fine structure (by dot-blot hybridization) (Figure 6.1C).

However, the presence of significant amounts of biotinylated DNA in non-drug treated samples (Figures 6.2 & 6.3) may be problematic. Initial dot-blot and fluorescence hybridization experiments with DNA probes generated from isolated damage fragments showed no significant differences between the drug-treated (1 nM) and the control samples for several sequences including telomeres, matrix attachment regions and human ribosomal DNA. The reasons for such similarity may be: (1) background DNA break fragments represent about one third of damage fragments in the 1 nM calicheamicin sample (Figure 6.3), masking any subtle differences between the two; and/or

(2) there is indeed no significant differences between the two DNA populations, at least not detectable by the kind of gross hybridization experiments performed. In order to differentiate these possibilities, we need to lower the amount of DNA fragments isolated from control cells in the future.

The DNA fragments isolated from control cells stem from background TDT labeling in these cells (Figure 6.2), and are most likely due to the presence of background strand breaks and labeling of telomeres. At 1 nM calicheamicin concentration, the average size of DNA fragments is  $10^6$  bp (data not shown). Therefore, there are on average  $10^3$  double-strand (DS) breaks per cell. Extreme care was taken to minimize DNA breaks introduced during DNA purification, including embedding cells in low melting agarose and using pipet tips with ends cut off. However, a living cell contains on the order of  $10^3$  single-strand (SS) nicks due to endogenous DNA damage, DNA replication and DNA repair [166]. These SS breaks can also be labeled by TDT, although to a lesser extent than DS breaks [157]. Human telomeres, on the other hand, exist as "DS breaks" with extended 3'-overhangs [167]. I have found that these DS breaks are good substrates for TDT: dot-blot hybridization experiments with isolated damage fragments from both control and drug-treated cells showed more than 20-fold enrichment of telomere sequences compared to total genomic DNA fragments (data not shown). Therefore, telomeres are highly enriched in the isolated damage fragments by the current protocol.

One way to minimize the labeling of background breaks and telomeres in the future is by taking advantage of the modified linker used successfully in the LMPCR studies (Chapters 2 to 4). However, in this case, the linker would possess a biotin on one end (Figure 6.4). This double-strand linker will

ligate only to calicheamicin-induced DS breaks or 3-nt 3'-overhangs, but not background SS nicks or telomeres (without the compatible 3'-overhang). The presence of double-strand (DS) breaks in living cells is much rarer, presumably because DS breaks are more lethal to cells than SS damage [161]. The frequency of DS breaks with 3-nt 3'-overhangs, as produced by calicheamicin, must be even rarer. Therefore, the majority of the background breaks will not be labeled. Assuming the occurrence of breaks containing 3-nt 3'-overhangs is a thousand times less than the frequency of background breaks in control cells (not an unreasonable assumption), this improved strategy (Figure 6.5) may substantially decrease background labeling. This has significant implications for the detection of calicheamicin-induced DNA damage at sub-lethal drug concentrations.

Furthermore, the presence of a universal linker places the first primer sequence adjacent to the DNA damage sites. If a second linker is ligated on the other end created by a restriction enzyme, PCR amplifications by both linker primers would be possible. The amplified fragments could be cloned into TA Cloning vectors from Invitrogen [168] to form a "genomic damage library" (Figure 6.5). Such a library would contain all the information regarding the quantity and location of DNA damage sites at very low drug concentrations. This will provide significant insights into the mechanisms of genotoxin target selection. For example, calicheamicin induces severe chromosome rearrangement in CHO cells (G. Ellestad, personal communication). It would be of great interest to examine whether the locations of DS break sites correlate with sites of chromosome rearrangement in these cells. In addition, the DNA damage probes generated by the techniques presented here could be used collectively for hybridization to

GeneChips from Affymetrix to obtain "DNA damage patterns" across the entire genome [169, 170].

In conclusion, a method has been developed to isolate calicheamicin-induced DNA breaks at nanomolar drug concentrations. This represented a  $10^4$  fold enrichment of drug damage fragments. Further improvement of the technique may make it possible to generate a collection of DNA damage sites at sub-lethal drug concentrations. Such a "genomic damage library" would prove useful to decipher the mechanisms of genomic target selection at the level of the whole genome.



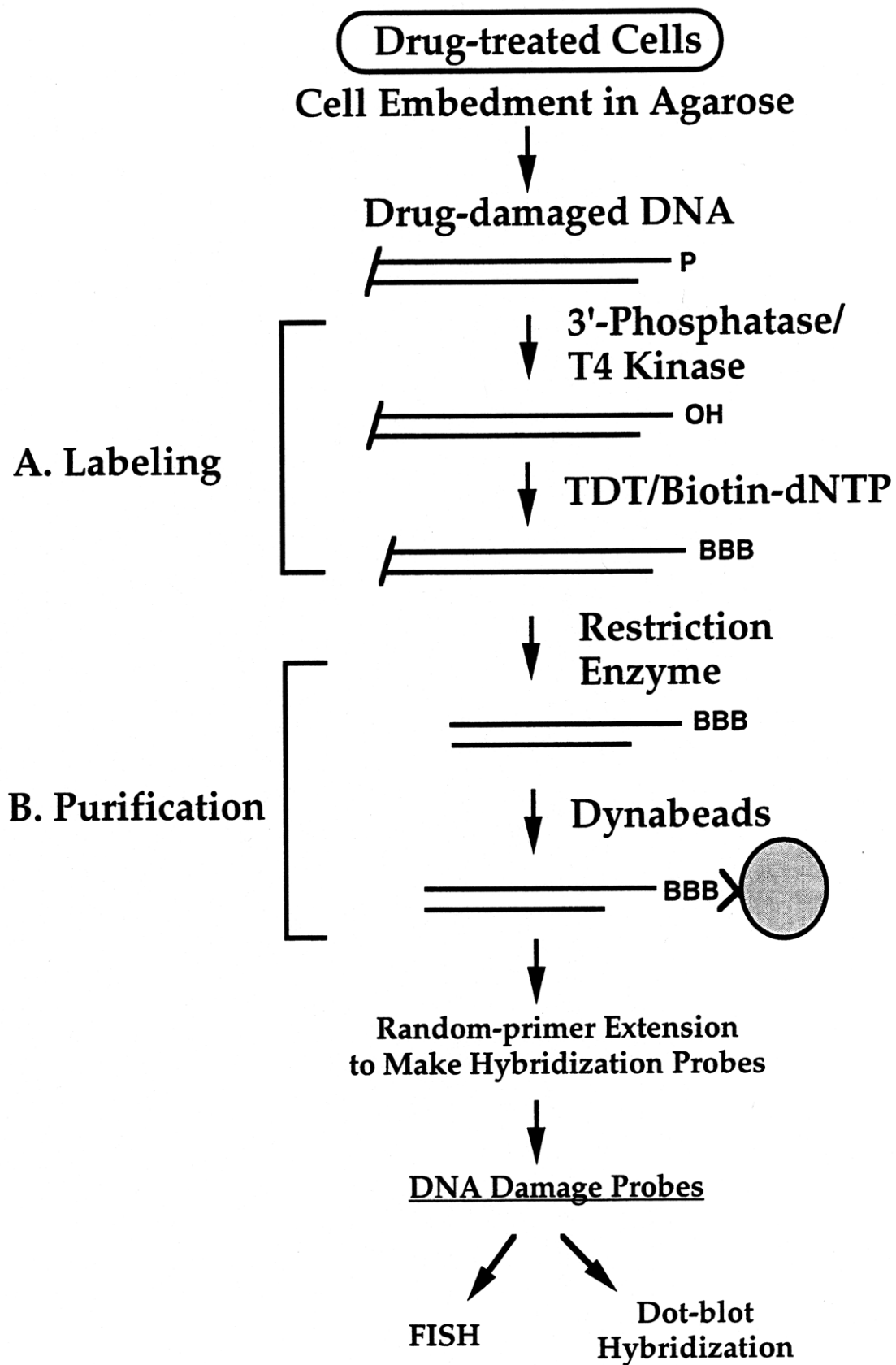
**Table 6.1.**

|   | <u>Type of buffers</u> |                |                   |       |                     |                 |
|---|------------------------|----------------|-------------------|-------|---------------------|-----------------|
|   | 1                      | 2              | 3                 | 4     | 5                   | 6               |
|   | TE<br>NaCl<br>Tween    | 6xSSC<br>Tween | TE<br>NaCl<br>SDS | 6xSSC | TE<br>LiCl<br>Tween | 20xSSC<br>Tween |
| <b>Specific<br/>binding</b>                         | 72.5±0.5               | 70±1           | 18±2              | 70    | 67                  | 58              |
| <b>Non-specific<br/>binding<br/>x10<sup>4</sup></b> | 1.6±1.0                | 0.7±0.3        | nt                | 11±1  | nt                  | nt              |

**Table 6.1.** Buffer effects on DNA binding to Dynabeads. Specific binding was tested with a <sup>32</sup>P-labeled biotinylated DNA fragment, while non-specific binding was tested with <sup>3</sup>H-labeled HeLa genomic DNA. The numbers indicate the percentages of radioactivity remaining on the Dynabeads after extensive washes with each buffer. Errors represent deviations about the mean for duplicate samples. nt = not tested. The following buffers were tested: (1) TE (10 mM Tris, 1 mM EDTA, pH 8) + 1M NaCl + 0.1% Tween 20, (2) 6xSSC + 0.1% Tween 20, (3) TE + 1M NaCl + 0.1% SDS, (4) 6xSSC, (5) TE + 3M LiCl + 0.2% Tween 20, (6) 20xSSC + 0.2% Tween 20.

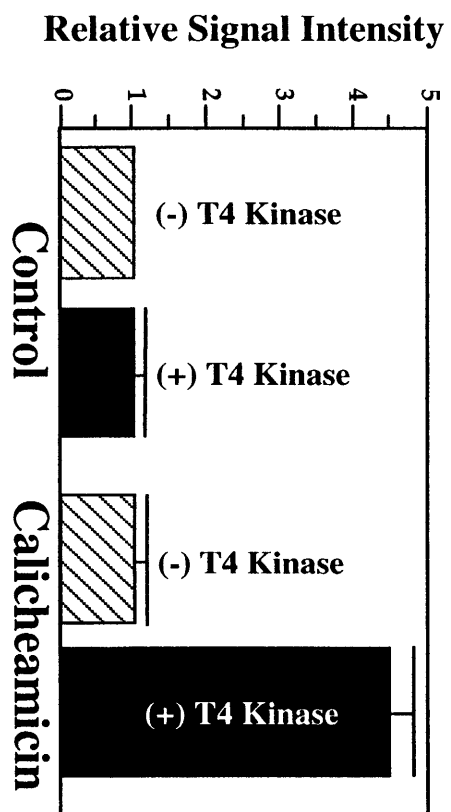
## V. Figures for chapter six

**Figure 6.1.** TDT labeling technique to capture low levels of DNA damage fragments in whole cells. After drug treatment, cells are embedded in agarose to provide structural support and to prevent mechanical damage to DNA during subsequent steps. The 3'-phosphatase activity of T4 kinase is exploited to convert calicheamicin-induced 3'-phosphates to 3'-hydroxyl ended fragments. The 3'-ends of DS breaks can then be labeled with biotinylated nucleosides and TDT. The purification strategy we have devised, involves restriction enzyme digestion to produce manageable-sized fragments, followed by affinity purification of biotin-labeled fragments by Dynabeads. Non-biotinylated DNA is removed from the Dynabeads by extensive washes. The purified damage fragments can be used to make DNA probes by random primer extension. These probes can be used for mapping DNA damage at all levels of genomic organization, from nuclear location (by fluorescent *in situ* hybridization) to DNA fine structure (by dot-blot hybridization).



**Figure 6.1**

**Figure 6.2.** Kinase-dependent TdT labeling of calicheamicin-damage ends. Both  $^3\text{H}$ -thymidine-labeled control cells and cells treated with 1 nM calicheamicin were embedded in agarose chops, and processed as described in the material and method section. Half of the samples were treated with T4 kinase (20 U) for 1 hr at 33°C. Afterwards, all samples were treated with biotinylated-dCTP (4  $\mu\text{M}$ ) and TdT (17 U) for 1 hr at 33°C. Digestions by RNase A (0.1 mg/ml) and proteinase K (0.1 mg/ml) and *Dpn* II (100 U) were followed to cut genomic DNA to manageable-sized fragments (average size ~300 bp). The DNA fragments were removed from the agarose chops by electroelution and stored in 10 mM Tris, 1 mM EDTA, pH 8. An aliquot of each DNA sample representing equal amount of DNA (equal cpm of  $^3\text{H}$ ) were spotted on a Nylon membrane. And the levels of biotinylation were assayed by a Phototope detection kit. The chemiluminescent signals were detected by X-ray films and quantitated by a densitometer. The relative levels of biotin labeling with and without the 3'-phosphatase activity of the T4 kinase is presented. Error bars represent deviations about the mean for duplicate samples.



**Figure 6.3.** Calicheamicin dose-response of biotin labeling and affinity purification. Both  $^3\text{H}$ -thymidine-labeled control cells and cells treated with increasing concentrations of calicheamicin were embedded in agarose chops, and processed as described in the material and method section. All samples were treated with T4 kinase (20 U) for 1 hr at 33°C and labeled with biotinylated-dCTP (4  $\mu\text{M}$ ) and TDT (17 U) for 1 hr at 33°C. Digestions by RNase A (0.1 mg/ml) and proteinase K (0.1 mg/ml) and *Dpn* II (100 U) were followed to cut genomic DNA to manageable-sized fragments (average size ~300 bp). The DNA fragments were removed from the agarose chops by electroelution and stored in 10 mM Tris, 1 mM EDTA, pH 8. About half of each DNA sample representing equal amount of DNA (equal cpm of  $^3\text{H}$ ) were allowed to bind to Dynabeads for 30 min at room temperature in the optimal binding buffer with gentle rotation. Non-biotinylated DNA was removed by extensive washes in the binding buffer. Finally, an aliquot of the beads were used to quantitate the amount of  $^3\text{H}$ -DNA remaining on the Dynabeads by scintillation counting. The percentages of total  $^3\text{H}$ -DNA remained on the Dynabeads from two independent experiments were pooled and presented with increasing calicheamicin concentrations.

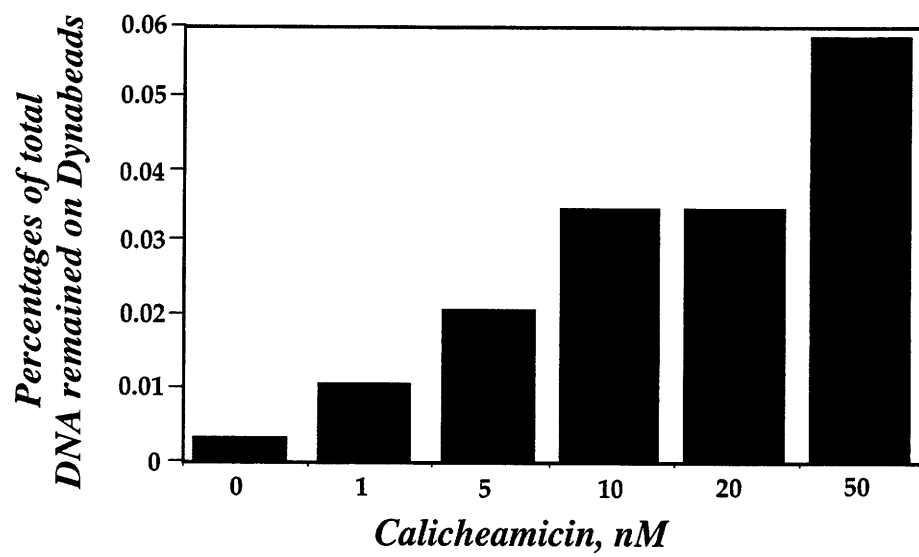


Figure 6.3



**Figure 6.4.** The structure of the modified biotinylated linker. This double-strand linker has a randomized 3-nt 3'-overhang on one end and a biotin tag on the other. It will ligate only to calicheamicin-induced DS breaks or a 3-nt 3'-overhang breaks, but not background SS nicks or DS breaks without the compatible 3'-overhang. Therefore, the majority of the background breaks will not be labeled. This improvement is expected to significantly decrease the background labeling and aid the detection of calicheamicin-induced DNA damage at extremely low drug concentrations.

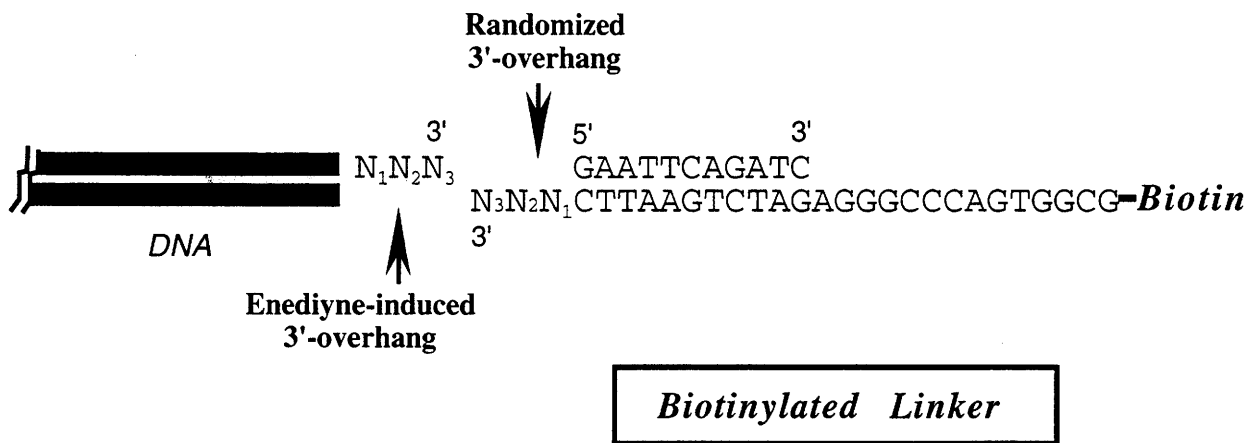


Figure 6.4

**Figure 6.5.** The improved ligation strategy to capture extremely low levels of DNA damage fragments in whole cells. After drug treatment, cells are embedded in agarose to provide structural support and to prevent mechanical damage to DNA during subsequent steps. Alternatively, since the modified linker is not compatible with background SS nicks, DNA is simply purified by conventional method. The 3'-phosphatase activity of T4 kinase is exploited to convert calicheamicin-induced 3'-phosphates to 3'-hydroxyl ended fragments. The 3'-ends of DS breaks can then be ligated directly to the biotinylated linker with randomized 3'-ends (A: ligation). The purification strategy is the same as in Figure 6.1 (B: purification), which involves restriction enzyme digestion to produce manageable-sized fragments, followed by affinity purification of biotin-labeled fragments by Dynabeads. Non-biotinylated DNA are removed from the Dynabeads by extensive washes. In C: amplification and cloning, a second linker is ligated on the other end created by the restriction enzyme. PCR amplifications by both linker primers can follow. The amplified fragments can be cloned into TA Cloning® vectors (Invitrogen) to form a "genomic damage library". Such a library would contain all the information regarding the quantity and location of DNA damage sites at very low drug concentrations. This collection of DNA damage fragments can be further studied by standard DNA mapping techniques, from probe hybridization to DNA sequencing. This will provide significant insights into the mechanisms of genotoxin target selection by these DNA damaging chemicals.

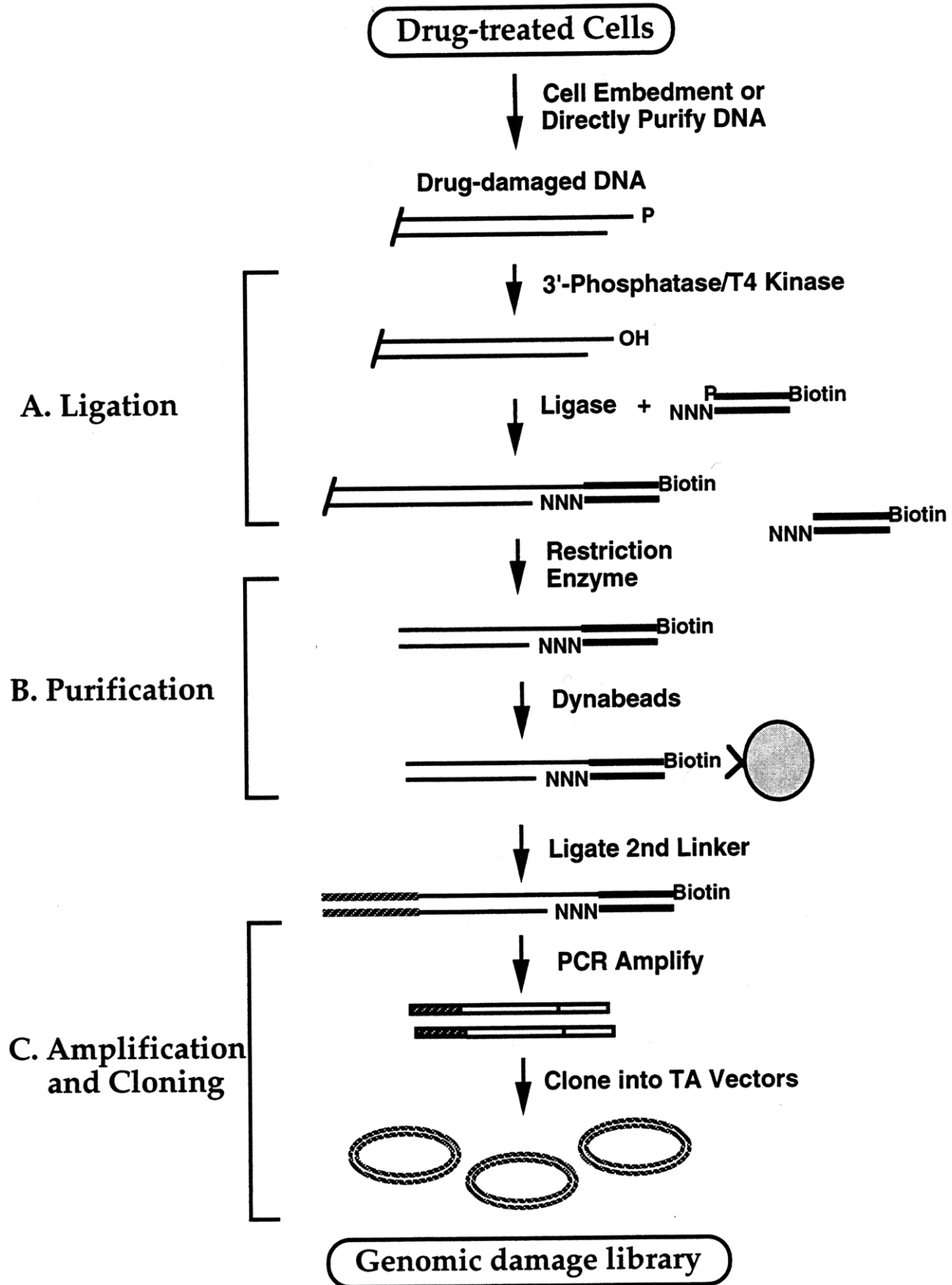


Figure 6.5

## Appendix: Mathematical Derivations

**1. Single-hit condition in the 400 bp DNA region analyzed by each primer set is achieved by cleaving genomic DNA to an average molecular size of 6000 bp.**

For simplicity, let us assume that in DNA damage experiments, each phosphodiester bond in the DNA backbone is damaged with the same probability  $P$ . Then the fragment sizes of the resulting DNA molecules approach a certain distribution, as resolved on an agarose gel. In such a distribution, the weight-average molecular weight  $M_w$  is approximately equal to the molecular weight corresponding to mobility of the peak of the mass distribution, and the probability of a bond being broken is the reciprocal of half the  $M_w$  [171]. Thus, if the molecular weight corresponding to the mobility of the peak mass distribution is 6000 bp as determined by agarose gel electrophoresis, the probability  $P$  is  $1/(6000/2)$ , or  $1/3000$ .

A DNA region of 400 bp will be examined with each primer set on a sequencing gel. In such 400 phosphodiester bonds, the probability of having "n" bonds broken is determined by the binomial distribution [172]:

$$P(n) = \frac{P^n(1-P)^{N-n} N!}{(N-n)!n!}, \text{ where } N = 400, \text{ and } P = 1/3000.$$

Thus,  $P(0) = 0.875$ ,  $P(1) = 0.117$ , and  $P(0) + P(1) = 0.992$ . This means that greater than 99% of the DNA substrate receives less than or equal to one hit in any 400 bp region (i.e., single-hit condition).

The single-hit condition is maintained as long as the molecular weight of the peak mass distribution is 6000 bp or greater. In practice, this is readily satisfied by treating purified DNA with 1  $\mu\text{M}$  or lower and whole cells with 10  $\mu\text{M}$  or lower drug concentrations.

**2. Starting with 2  $\mu\text{g}$  of mammalian genomic DNA ensures the presence of at least 100 molecules of each damage fragment.**

A diploid mammalian genome contains approximately  $6 \times 10^9$  bp, which is equivalent to  $6.8 \times 10^{-12}$  grams [173]. Thus, 2  $\mu\text{g}$  of mammalian genomic DNA represents about  $2.94 \times 10^5$  cells or  $5.88 \times 10^5$  autosomal gene copies. For the purpose of my studies, I use  $3 \times 10^5$  cells or  $6 \times 10^5$  autosomal gene copies per 2  $\mu\text{g}$  of genomic DNA, as did before in other studies [100].

As discussed above, genomic DNA damaged to peak mass distribution of 6000 bp ensures single-hit conditions within the 400 bp region analyzed by any gene-specific primer. Under these conditions, the probability P (any bond being broken) is  $1/3000$ . Consider an extreme situation, where damage occurs randomly at every single nucleotide. The number of starting molecules at every possible damage site is equivalent to the number of gene copies multiplied by the probability of a bond being broken:  $(6 \times 10^5) \times (1/3000) = 200$ . Thus, by starting with 2  $\mu\text{g}$  of DNA, there are at least 100 molecules of each possible damage site in the beginning of ligation-mediated PCR.

## BIBLIOGRAPHY

1. Bohr, V.A. and R.M. Anson, *DNA damage, mutation and fine structure DNA repair in aging*. Mutation Research, 1995. 338: p. 25-34.
2. Lindahl, T., *Recognition and processing of damaged DNA*. J. Cell Sci. Suppl., 1995. 19: p. 73-77.
3. van Holde, K.E., *Chromatin*. 1989, New York: Springer-Verlag.
4. Wolffe, A., *Chromatin Structure and Function*. 1992, San Diego: Academic Press.
5. Wilson, W.D., *Reversible interactions of nucleic acids with small molecules*, in *Nucleic Acids in Chemistry and Biology*, G.M. Blackburn and M.J. Gait, Editors. 1990, IRL Press: Oxford.
6. Bestor, T.H., *Methylation patterns in the vertebrate genome*. J. NIH Res., 1993. 5: p. 57-60.
7. Boyes, J. and A. Bird, *Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein*. EMBO J, 1992. 11: p. 327-333.
8. Watt, F. and P.L. Molloy, *Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter*. Genes Dev., 1988. 2: p. 1136-1143.
9. Bi, W., et al., *DNA binding specificity of the CCAAT-binding factor CBF/IN-F-Y*. J. Biol. Chem., 1997. 272: p. 26562-26572.
10. Harrington, M.A., et al., *Cytosine methylation does not affect binding of transcription factor Sp1*. Proc. Natl. Acad. Sci., 1988. 85: p. 2066-2070.
11. Boyes, J. and A. Bird, *DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein*. Cell, 1991. 64: p. 1123-1134.
12. Lewis, J.D., et al., *Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA*. Cell, 1992. 69: p. 905-914.
13. Nan, X., F.J. Campoy, and A. Bird, *MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin*. Cell, 1997. 88: p. 471-481.
14. Nan, X., et al., *DNA methylation specifies chromosomal localization of MeCP2*. Mol. Cell Biol., 1996. 16: p. 414-421.
15. Zacharias, W., T.R. O'Connor, and J.E. Larson, *Methylation of cytosine in the 5-position alters the structural and energetic properties of the supercoil-induced Z-helix and of B-Z junctions*. Biochemistry, 1988. 27(8): p. 2970-2978.
16. Sowers, L.C., B.R. Shaw, and W.D. Sedwick, *Base stacking and molecular polarizability: Effect of a methyl group in the 5-position of pyrimidines*. Biochem. Biophys. Res. Comm., 1987. 148(2): p. 790-794.
17. Hausheer, F.H., et al., *Computational analysis of structural and energetic consequences of DNA methylation*. Carcinogenesis, 1989. 10: p. 1131-1137.
18. Diekmann, S., *DNA methylation can enhance or induce DNA curvature*. EMBO J., 1987. 6(13): p. 4213-4217.
19. Hodges-Garcia, Y. and P.J. Hagerman, *Cytosine methylation can induce local distortions in the structure of duplex DNA*. Biochemistry, 1992. 31: p. 7595-7599.
20. Muiznieks, I. and W. Doerfler, *The topology of the promoter of RNA polymerase II- and III-transcribed genes is modified by the methylation of 5'-CG-3' dinucleotides*. Nucleic Acids Res., 1994. 22(13): p. 2568-2575.
21. Denissenko, M.F., et al., *Cytosine methylation determines not spots of DNA damage in the human P53 gene*. Proc. Natl. Acad. Sci. USA, 1997. 94: p. 3893-3898.
22. Johnson, W.S., Q.-Y. He, and M. Tomasz, *Selective recognition of the m<sup>5</sup>CpG dinucleotide sequence in DNA by mitomycin C for alkylation and cross-linking*. Bioorg. Med. Chem., 1995. 3(6): p. 851-860.
23. Hertzberg, R.P., M.J. Caranfa, and S.M. Hecht, *DNA methylation diminishes bleomycin-mediated strand scission*. Biochemistry, 1985. 24(20): p. 5285-5289.

24. Mathison, B.H., B. Said, and R.C. Shank, *Effect of 5-methylcytosine as a neighboring base on methylation of DNA guanine by N-methyl-N-nitrosourea*. *Carcinogenesis*, 1993. **14**(2): p. 323-327.
25. Lugar, K., *et al.*, *Crytal structure of the nucleosome core particle at 2.8 Å resolution*. *Nature*, 1997. **389**: p. 251-260.
26. Richmond, T.J., *et al.*, *Structure of the nucleosome core particle at 7Å resolution*. *Nature*, 1984. **311**: p. 532-537.
27. Hayes, J.J., T.D. Tullius, and A.P. Wolffe, *The structure of DNA in a nucleosome*. *Proc. Natl. Acad. Sci. USA*, 1990. **87**: p. 7405-7409.
28. Yu, L., I.H. Goldberg, and P.C. Dedon, *Enediyne-mediated DNA damage in nuclei is modulated at the level of the nucleosome*. *J. Biol. Chem.*, 1994. **269**(February 11): p. 4144-4151.
29. Hurley, I., *et al.*, *Torsional motion and elasticity of deoxyribonucleic acid double helix and its nucleosomal complexes*. *Biochemistry*, 1982. **21**: p. 4999-5009.
30. Chaires, J.B., N. Dattagupta, and D.M. Crothers, *Binding of daunomycin to calf thymus nucleosomes*. *Biochemistry*, 1983. **22**: p. 284-292.
31. McMurray, C.T. and K.E. van Holde, *Binding of ethidium to the nucleosome core particle. 1. Binding and dissociation reactions*. *Biochemistry*, 1991. **30**: p. 5631-5643.
32. McMurray, C.T., E.W. Small, and K.E. van Holde, *Binding of ethidium to the nucleosome core particle. 2. Internal and external binding modes*. *Biochemistry*, 1991. **30**: p. 5644-5652.
33. Capco, D.G., K.M. Wan, and S. Penman, *The nuclear matrix: three-dimensional architecture and protein composition*. *Cell*, 1982. **29**(3): p. 847-858.
34. Nasedkina, T.V. and S.I. Slesinger, *The structure of partly decondensed metaphase chromosomes*. *Chromosoma*, 1982. **86**(2): p. 239-249.
35. van Holde, K. and J. Zlatanova, *What determines the folding of the chromatin fiber?* *Proc Natl Acad Sci USA*, 1996. **93**: p. 10548-10555.
36. Turner, B.M., *Cell*, 1993. **75**: p. 5-8.
37. Lee, D.Y., *et al.*, *A positive role for histone acetylation in transcription factor access to nucleosomal DNA*. *Cell*, 1993. **72**: p. 73-84.
38. Hebbes, T.R., *et al.*, *Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain*. *EMBO J.*, 1994. **13**: p. 1823-1830.
39. Tsukiyama, T. and C. Wu, *Chromatin remodeling and transcription*. *Curr. Opin. Genet. Dev.*, 1997. **7**: p. 182-191.
40. Zlatanova, J. and K. van Holde, *Histone H1 and transcription: still an enigma?* *J. Cell Sci.*, 1992. **103**: p. 889-895.
41. Qiao, L., O. Wrange, and P. Eriksson, *The role of chromatin in transcriptional regulation*. *Int. J. Biochem. Cell Biol.*, 1997. **29**(5): p. 731-742.
42. Imbalzano, A.N., *et al.*, *Facilitated binding of TATA-binding protein to nucleosomal DNA*. *Nature*, 1994. **370**: p. 481-485.
43. Wilson, C.J., *et al.*, *RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling*. *Cell*, 1996. **84**: p. 235-244.
44. Tsukiyama, T. and C. Wu, *Purification and properties of an ATP-dependent nucleosome remodeling factor*. *Cell*, 1995. **83**: p. 1011-1020.
45. Lodish, H., *et al.*, *Molecular Cell Biology*. 1995, New York: Scientific American Books, Inc.
46. Beato, M. and K. Eisefeld, *Transcription factor access to chromatin*. *Nucleic Acids Research*, 1997. **25**: p. 3559-3563.
47. Eisefeld, K., *et al.*, *Binding of NF1 to the MMTV promoter in nucleosomes: influence of rotational phasing, translational positioning and hitone H1*. *Nucleic Acids Res.*, 1997. **25**(18): p. 3733-3742.
48. Wolffe, A.P., *Regulation of chromatin structure and function*. 1994, Austin: R.G. Landes Company.



49. Steger, D.J. and J.L. Workman, *Remodeling chromatin structures for transcription: what happens to the histones?* Bioessays, 1996. 18: p. 875-884.
50. Michelson, A.M., A.F. Markham, and S.H. Orkin, *Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase.* Proc. Natl. Acad. Sci. USA, 1983. 80: p. 472-476.
51. Michelson, A.M., et al., *Structure of the human phosphoglycerate kinase gene and the intron-mediated evolution and dispersal of the nucleotide-binding domain.* Proc. Natl. Acad. Sci. USA, 1985. 82: p. 6965-6969.
52. Hansen, R.S., N.A. Ellis, and S.M. Gartler, *Demethylation of specific sites in the 5' region of the inactive X-linked human phosphoglycerate kinase gene correlates with the appearance of nuclease sensitivity and gene expression.* Mol. Cell. Biol., 1988. 8: p. 4692-4699.
53. Pfeifer, G.P., et al., *In vivo footprinting and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1.* Genes Dev., 1990. 4: p. 1277-1287.
54. Smith, A.L. and K.C. Nicolaou, *The enediyne antibiotics.* J. Med. Chem., 1996. 39: p. 2103-2117.
55. Lee, M.D., G.A. Ellestad, and D.B. Borders, *Calicheamicins: Discovery, structure, chemistry, and interaction with DNA.* Acc. Chem. Res., 1991. 24: p. 235-243.
56. Long, B.H., et al., *Esperamicins, a class of potent antitumor antibiotics: mechanism of action.* Proc. Natl. Acad. Sci. USA, 1989. 86: p. 2-6.
57. Sugiura, Y., et al., *Nucleotide-specific cleavage and minor-groove interaction of DNA with esperamicin antitumor antibiotics.* Proc. Natl. Acad. Sci. USA, 1989. 86: p. 7672-7676.
58. Dedon, P.C. and I.H. Goldberg, *Free-radical mechanisms involved in the formation of sequence-dependent bistranded DNA lesions by the antitumor antibiotics bleomycin, neocarzinostatin, and calicheamicin.* Chem. Res. Tox., 1992. 5(3): p. 311-332.
59. Goldberg, I.H., *Mechanism of neocarzinostatin action: role of DNA microstructure in determination of chemistry of bistranded oxidative damage.* Accts. Chem. Res., 1991. 24: p. 191-198.
60. Xu, Y.-j., Y.-s. Zhen, and I.H. Goldberg, *C1027 chromophore, a potent new enediyne antitumor antibiotic, induces sequence-specific double-strand DNA cleavage.* Biochemistry, 1994. 33: p. 5947-5954.
61. Konishi, M., et al., *Dynemicin A, a novel antibiotic with the anthraquinone and 1,5-diyne-3-ene subunit.* J. Antibiot., 1989. 42: p. 1449-1452.
62. Leet, J.E., et al., *Kedarcidin, a new chromoprotein antitumor antibiotic: Structure elucidation of kedarcidin chromophore.* J. Am. Chem. Soc., 1992. 114: p. 7946-7948.
63. Zein, N., et al., *Kedarcidin chromophore: An enediyne that cleaves DNA in a sequence-specific manner.* Proc. Natl. Acad. Sci. USA, 1993. 90: p. 2822-2826.
64. Zein, N., et al., *Maduropeptin: an antitumor chromoprotein with selective protease activity and DNA cleaving properties.* Biochemistry, 1995. 34: p. 11591-11597.
65. Dedon, P.C. and I.H. Goldberg, *Mechanism of DNA damage by neocarzinostatin and other bicyclic enediyne antibiotics, in Nucleic Acid Targeted Drug Design, C. Probst and T. Perun, Editors. 1992, Marcel Dekker, Inc.: New York. p. 475-523.*
66. Zein, N., et al., *Calicheamicin  $\gamma$ <sup>I</sup>: an antitumor antibiotic that cleaves double-stranded DNA site specifically.* Science, 1988. 240: p. 1198-1201.
67. Sullivan, N. and L. Lyne, *Sensitivity of fibroblasts derived from ataxia-telangiectasia patients to calicheamicin  $\gamma$ <sup>I</sup>.* Mutation Res., 1990. 245: p. 171-175.
68. Mah, S.C., C.A. Townsend, and T.D. Tullius, *Hydroxyl radical footprinting of calicheamicin. Relationship of DNA binding to cleavage.* Biochemistry, 1994. 33: p. 614-621.
69. Dedon, P.C. and I.H. Goldberg, *Influence of thiol structure on neocarzinostatin activation and expression of DNA damage.* Biochemistry, 1992. 31: p. 1909-1917.

70. Dedon, P.C., A.A. Salzberg, and J. Xu, *Exclusive production of bistranded DNA damage by calicheamicin*. *Biochemistry*, 1993. **32**: p. 3617-3622.
71. Yu, L., et al., *The deoxyfucose-anthranilate of esperamicin A1 confers intercalative DNA binding and causes a switch in the chemistry of bistranded DNA lesions*. *J. Am. Chem. Soc.*, 1994. **116**(21): p. 9733-9738.
72. Povirk, L.F. and C.W. Houlgrave, *Effect of apurinic/apyrimidinic endonucleases and polyamines on DNA treated with bleomycin and neocarzinostatin: specific formation and cleavage of closely opposed lesions in complementary strands*. *Biochemistry*, 1988. **27**: p. 3850-3857.
73. Bennett, R.O., P.S. Swerdlow, and L.F. Povirk, *Spontaneous cleavage of bleomycin-induced abasic sites in chromatin and their mutagenicity in mammalian shuttle vectors*. *Biochemistry*, 1993. **32**: p. 3188-3195.
74. Dedon, P.C., Z.-W. Jiang, and I.H. Goldberg, *Neocarzinostatin-mediated DNA damage in a model AGT·ACT site: mechanistic studies of thiol-sensitive partitioning of C4' DNA damage products*. *Biochemistry*, 1992. **31**: p. 1917-1927.
75. Ikemoto, N., et al., *Esperamicin A1 intercalates into duplex DNA from the minor groove*. *J. Am. Chem. Soc.*, 1994. **116**(20): p. 9387-9388.
76. Yu, L., A.A. Salzberg, and P.C. Dedon, *New insights into calicheamicin-DNA interactions derived from a model nucleosome system*. *Bioorg. Med. Chem.*, 1995. **3**(6): p. 729-741.
77. Mah, S.C., et al., *Features of DNA recognition for oriented binding and cleavage by calicheamicin*. *Tetrahedron*, 1994. **50**(5): p. 1361-1378.
78. Salzberg, A., P. Mathur, and P. Dedon, *The intrinsic flexibility and drug-induced bending of calicheamicin DNA targets*, in *DNA and RNA Cleavers and Chemotherapy of Cancer and Viral Diseases*, B. Meunier, Editor. 1996, Kluwer Academic Publishers: Dordrecht. p. 23-36.
79. Riley, D.E., M.A. Goldman, and S.M. Gartler, *Chromatin structure of active and inactive human x-linked phosphoglycerate kinase gene*. *Somatic Cell Mol. Genet.*, 1986. **12**(1): p. 73-80.
80. Pfeifer, G.D. and A.D. Riggs, *Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR*. *Genes Dev.*, 1991. **5**: p. 1102-1113.
81. Wassermann, K., *Intragenomic heterogeneity of DNA damage formation and repair: a review of cellular responses to covalent drug DNA interaction*. *Crit. Rev. in Tox.*, 1994. **24**: p. 281-322.
82. Pfeifer, G.D., et al., *In vivo mapping of a DNA adduct at nucleotide resolution: Detection of pyrimidine (6,4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction*. *Proc. Natl. Acad. Sci. USA*, 1991. **88**(4): p. 1374-1378.
83. Pfeifer, G.P., et al., *Binding of transcription factors creates hot spots for UV photoproducts in vivo*. *Mol Cell Biol*, 1992. **12**(4): p. 1798-804.
84. Pfeifer, G.D., R. Drouin, and G.P. Holmquist, *Detection of DNA adducts at the DNA sequence level by ligation-mediated PCR*. *Mutation Res*, 1993. **288**: p. 39-46.
85. Mueller, P.R. and B. Wold, *In vivo footprinting of a muscle specific enhancer by ligation mediated PCR*. *Science*, 1989. **246**: p. 780-786.
86. Pfeifer, G.P., et al., *Genomic sequencing and methylation analysis by ligation mediated PCR*. *Science*, 1989. **246**: p. 810-813.
87. Maxam, A. and W. Gilbert, *Sequencing end-labeled DNA with base-specific chemical cleavages*. *Methods Enzymol.*, 1980. **65**: p. 499-560.
88. Sugiura, Y. and T. Matsumoto, *Some characteristics of DNA strand scission by macromolecular antitumor antibiotic C-1027 containing a novel enediyne chromophore*. *Biochemistry*, 1993. **32**: p. 5548-5553.
89. Lindahl, T. and A. Andersson, *Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid*. *Biochemistry*, 1972. **11**: p. 3618-3623.

90. Povirk, L.F. and I.H. Goldberg, *Endonuclease-resistant apyrimidinic sites formed by neocarzinostatin at cytosine residues in DNA: evidence for a possible role in mutagenesis*. Proc. Natl. Acad. Sci. USA, 1985. **82**: p. 3182-3186.
91. Pfeifer, G.P. and A.D. Riggs, *Genomic sequencing*. Methods Mol Biol, 1993. **23**: p. 169-81.
92. Mueller, P.R. and B. Wold, *Ligation-mediated PCR: Applications to genomic footprinting*. Methods: a companion to methods in enzymology, 1991. **2**(1): p. 20-31.
93. Lincoln, S.E., M.J. Daly, and E.S. Lander, *Primer: a computer program for automatically selecting PCR primers*, . 1991, MIT.
94. Rychlik, W., W.J. Spencer, and R.E. Rhoads, *Optimization of the annealing temperature for DNA amplification in vitro*. Nucleic Acids Res., 1990. **18**: p. 6409-6412.
95. Tornaletti, S. and G.P. Pfeifer, *Ligation-mediated PCR for analysis of UV damage*, in *Technologies for Detection of DNA Damage and Mutations*, G.P. Pfeifer, Editor. 1996, Plenum Press: New York and London. p. 199-209.
96. Rodriguez, H., *et al.*, *Mapping of copper/hydrogen peroxide-induced DNA damage at nucleotide resolution in human genomic DNA by ligation-mediated polymerase chain reaction*. J. Biol. Chem., 1995. **270**: p. 17633-17640.
97. Pfeifer, G.P., *Analysis of chromatin structure by ligation-mediated PCR*. PCR Methods Appl., 1992. **2**(2): p. 107-11.
98. Pfeifer, G.P., J. Singer-Sam, and A.D. Riggs, *Analysis of methylation and chromatin structure*. Methods in Enzymology, 1993. **225**: p. 567-83.
99. Singer-Sam, J., *et al.*, *Measurement by quantitative PCR of changes in HPRT, PGK-2, APRT, MTase, and Zfy gene transcripts during mouse spermatogenesis*. Nucleic Acids Res., 1990. **18**: p. 1255-1259.
100. Cha, R.S., *et al.*, *Mismatch amplification mutation assay (MAMA): Application to the c-H-ras gene*. PCR Methods and Applications, 1992. **2**: p. 14-20.
101. Liang, Q., D.-J. Choi, and P.C. Dedon, *Calicheamicin-mediated DNA damage in a reconstituted nucleosome is not affected by histone acetylation: the role of drug structure in the target recognition process*. Biochemistry, 1997. **36**(42): p. 12653-12659.
102. Singer-Sam, J., *et al.*, *Sequence of the promoter region of the gene for human 3-phosphoglycerate kinase*. Gene, 1984. **32**: p. 409-417.
103. Sturzl, M. and W.K. Roth, *"Run-off" synthesis and application of defined single-stranded DNA hybridization probes*. Anal. Biochem., 1990. **185**: p. 164-169.
104. Tabor, S., H.E. Huber, and C.C. Richardson, *Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7*. J. Biol. Chem., 1987. **262**: p. 16212-16223.
105. Tormanen, V.T., *et al.*, *Extension product capture improves genomic sequencing and DNase I footprinting by ligation-mediated PCR*. Nucleic Acids Res., 1992. **20**(20): p. 5487-8.
106. Cartwright, I.L., *et al.*, *Cleavage of chromatin with methidiumpropyl-EDTA.iron(II)*. Proc. Natl. Acad. Sci. USA, 1983. **80**: p. 3213-3217.
107. McGhee, J.D. and G. Felsenfeld, *Reaction of nucleosome DNA with dimethyl sulfate*. Proc. Natl. Acad. Sci., 1979. **76**: p. 2133-2137.
108. Travers, A.A., in *DNA-Protein: Structural Interactions*, D.M.J. Lilley, Editor. 1995, Oxford Univ. Press: New York. p. 49-75.
109. Riley, D.E., T.K. Canfield, and S.M. Gartler, *Chromatin structure of active and inactive human X chromosome*. Nucleic Acids Res., 1984. **12**: p. 1829-1845.
110. Shuman, J.D., C.R. Vinson, and S.L. McKnight, *Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP*. Science, 1990. **249**: p. 771-774.
111. Garlatti, M., *et al.*, *CCAAT/enhancer-binding protein-related proteins bind to the unusual promoter of the aspartate aminotransferase housekeeping gene*. J. Biol. Chem., 1993. **268**: p. 6567-6574.

112. Gidoni, D., W.S. Dynan, and R. Tjian, *Multiple specific contacts between a mammalian transcription factor and its cognate promoters*. *Nature*, 1984. **312**: p. 409-413.
113. Sjøttem, E., C. Andersen, and T. Johansen, *Structural and functional analyses of DNA bending induced by Sp1 family transcription factors*. *J. Mol. Biol.*, 1997. **267**: p. 490-504.
114. Drouin, R., *et al.*, *Ligation-mediated PCR for analysis of oxidative DNA damage*, in *Technologies for Detection of DNA Damage and Mutations*, G.P. Pfeifer, Editor. 1996, Plenum Press: New York and London. p. 211-225.
115. HersHKovitz, M. and A.D. Riggs, *Metaphase chromosome analysis by ligation-mediated PCR: Heritable chromatin structure and a comparison of active and inactive X chromosomes*. *Proc. Natl. Acad. Sci. USA*, 1995. **92**: p. 2379-2383.
116. Uesugi, M. and Y. Sugiura, *New insights into sequence recognition process of esperamicin A<sub>1</sub> and calicheamicin γ<sub>1</sub><sup>I</sup>: Origin of their selectivities and "induced fit" mechanism*. *Biochemistry*, 1993. **32**: p. 4622-4627.
117. Spassky, A. and H.C. Buc, *Physico-chemical properties of a DNA binding protein: Escherichia coli factor HI*. *Eur. J. Biochem.*, 1977. **81**: p. 79-90.
118. Werner, M.H., A.M. Gronenborn, and G.M. Glore, *Intercalation, DNA kinking, and the control of transcription*. *Science*, 1996. **271**: p. 778-784.
119. Shi, Y. and J.M. Berg, *A direct comparison of the properties of natural and designed zinc-finger proteins*. *Chem. & Biol.*, 1995. **2**(2): p. 83-89.
120. Avitahl, N. and K. Calame, *The C/EBP family of proteins distorts DNA upon binding but does not introduce a large directed bend*. *J. Biol. Chem.*, 1994. **269**: p. 23553-23562.
121. Kumar, R.A., N. Ikemoto, and D.J. Patel, *Solution structure of the calicheamicin gamma II-DNA complex*. *J. Mol. Biol.*, 1997. **265**(2): p. 187-201.
122. Kumar, R.A., N. Ikemoto, and D.J. Patel, *Solution structure of the esperamicin A<sub>1</sub>-DNA complex*. *J. Mol. Biol.*, 1997. **265**(2): p. 173-186.
123. Lahm, A. and D. Suck, *DNase I-induced DNA conformation. 2 A structure of a DNase I-octamer complex*. *J. Mol. Biol.*, 1991. **222**(3): p. 645-667.
124. Weston, S.A., A. Lahm, and D. Suck, *X-ray structure of the DNase I-d(GGTATACC)<sub>2</sub> complex at 2.3 Å resolution*. *J. Mol. Biol.*, 1992. **226**(4): p. 1237-1256.
125. Mendoza, R., *et al.*, *DNase I susceptibility of bent DNA and its alteration by ditercalinium and distamycin*. *Biochemistry*, 1990. **29**(21): p. 5035-5043.
126. Bauer, W.R. and R. Gallo, *Physical and topological properties of closed circular DNA*, in *Chromosomes: Eukaryotic, Prokaryotic, and Viral*, K.W. Adolph, Editor. 1990, CRC Press, Inc.: Boca Raton. p. 87-126.
127. Lerman, L.S., *Structural considerations in the interaction of DNA and acridines*. *J. Mol. Biol.*, 1961. **3**: p. 18-30.
128. Waring, M., *Variation of the supercoils in closed circular DNA by binding of antibiotics and drugs: Evidence for molecular models involving intercalation*. *J. Mol. Biol.*, 1970. **54**: p. 247-279.
129. Vologodskii, A.V. and N.R. Cozzarelli, *Conformational and thermodynamic properties of supercoiled DNA*. *Annu. Rev. Biophys. Biomol. Struct.*, 1994. **23**: p. 609-643.
130. Liu, C., *et al.*, *Sequence-selective carbohydrate-DNA interaction: dimeric and monomeric forms of the calicheamicin oligosaccharide interfere with transcription factor function*. *Proc. Natl. Acad. Sci. USA*, 1996. **93**: p. 940-944.
131. Salzberg, A. and P. Dedon, *Calicheamicin target recognition: DNA structure and dynamics in small molecule-DNA interactions*. *Proc. Natl. Acad. Sci. USA*, 1998. submitted.
132. Sinden, R.R., *DNA Structure and Function*. 1994, San Diego: Academic Press.

133. Ho, S.N., *et al.*, *Specific inhibition of formation of transcription complexes by a calicheamicin oligosaccharide: a paradigm for the development of transcriptional antagonists*. Proc. Natl. Acad. Sci. USA, 1994. **91**(20): p. 9203-9207.
134. Turker, M.S. and T.H. Bestor, *Formation of methylation patterns in the mammalian genome*. Mutat. Res., 1997. **386**: p. 119-130.
135. Lewis, J. and A. Bird, *DNA methylation and chromatin structure*. FEBS Letters, 1991. **285**(2): p. 155-159.
136. Bird, A., *The essentials of DNA methylation*. Cell, 1992. **70**: p. 5-8.
137. Laird, P.W. and R. Jaenisch, *DNA methylation and cancer*. Hum. Mol. Genet., 1994. **3**: p. 1487-1495.
138. Nicolaou, K.C. and W.-M. Dai, *Chemistry and Biology of the Eneidyne Anticancer Antibiotics*. Angew. Chem. Int. Ed. Engl., 1991. **30**(11): p. 1387-1530.
139. Rideout, W.M., G.A. Coetzee, and P.A. Jones, *5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes*. Science, 1990. **249**: p. 1288-1290.
140. Behe, M.J. and G. Felsenfeld, *Effects of methylation on a synthetic polynucleotide: the B--Z transition in poly(dG-m5dC).poly(dG-m5dC)*. Proc. Natl. Acad. Sci. USA, 1981. **78**: p. 1619-1623.
141. Christner, D.F., *et al.*, *Unmasking the chemistry of DNA cleavage by the esperamicins: Modulation of 4'-hydrogen abstraction and bistranded damage by the fucose-anthranilate moiety*. J. Am. Chem. Soc., 1992. **114**: p. 8763-8767.
142. Mathur, P., J. Xu, and P.C. Dedon, *Cytosine methylation enhances DNA damage produced by groove binding and intercalating enediynes: studies with esperamicins A1 and C*. Biochemistry, 1997. **36**: p. 14868-14873.
143. Nur, I., *et al.*, *Prokaryotic and eukaryotic traits of DNA methylation in spiroplasmas (mycoplasmas)*. J. Bacteriol., 1985. **164**(1): p. 19-24.
144. Kim, U.S., *et al.*, *Dynamics and structures of DNA: long-range effects of a 16 base-pair (CG)<sub>8</sub> sequence on secondary structure*. Biopolymers, 1993. **33**(11): p. 1725-1745.
145. Hagerman, P.J., *Pyrimidine 5-methyl groups influence the magnitude of DNA curvature*. Biochemistry, 1990. **29**: p. 1980-1983.
146. Kuo, M.T., *Preferential damage of active chromatin by bleomycin*. Cancer Res., 1981. **41**: p. 2439-2443.
147. Beckmann, R.P., *et al.*, *Assessment of preferential cleavage of an actively transcribed retroviral hybrid gene in murine cells by deoxyribonuclease I, bleomycin, neocarzinostatin, or ionizing radiation*. Biochemistry, 1987. **26**: p. 5409-5415.
148. Irvin, T.R. and G.N. Wogan, *Quantitation of aflatoxin B<sub>1</sub> adduction within the ribosomal RNA gene sequences of rat liver DNA*. Proc. Natl. Acad. Sci. USA, 1984. **81**: p. 664-668.
149. Zhao, B., *et al.*, *Modulation of nicotinamide adenine dinucleotide and poly(adenosine diphosphoribose) metabolism by calicheamicin  $\gamma$ 1 in human HL-60 cells*. Cancer Lett., 1990. **50**: p. 141-147.
150. Jackson, D.A., P. Dickson, and P.R. Cook, *The size of chromatin loops in HeLa cells*. EBMO J., 1990. **9**(2): p. 567-571.
151. Cameron, V. and O.C. Uhlenbeck, *3'-Phosphatase activity in T4 polynucleotide kinase*. Biochemistry, 1977. **16**(23): p. 5120-5126.
152. Cook, P.R., *A general method for preparing intact nuclear DNA*. EMBO J., 1984. **3**(8): p. 1837-1842.
153. Ausubel, F.M., *et al.*, *Current Protocols in Molecular Biology*. 1989, New York: John Wiley and Sons.
154. Pierce, *Avidin-Biotin Chemistry: A Handbook*. 1994, Rockford, IL: Pierce Chemical Company.
155. Hutchison, N. and H. Weintraub, *Localization of DNase I-sensitive sequences to specific regions of interphase nuclei*. Cell, 1985. **43**: p. 471-482.

156. Kerem, B.-S., *et al.*, *In situ nick-translation distinguishes between active and inactive X chromosomes*. *Nature*, 1983. **304**: p. 88-90.
157. Deng, G.R. and R. Wu, *Terminal transferase: use in the tailing of DNA and for in vitro mutagenesis*. *Methods in Enzymology*, 1983. **100**: p. 96-116.
158. Beerman, T., G. Mueller, and H. Grimmond, *Effects of neocarzinostatin on chromatin in HeLa S<sub>3</sub> nuclei*. *Mol. Pharmacol.*, 1983. **23**: p. 493-499.
159. Smerdon, M.J., *DNA repair and the role of chromatin structure*. *Curr. Opin. Cell Biol.*, 1991. **3**: p. 422-428.
160. Friedberg, E.C., *DNA Repair*. 1985, New York: W. H. Freeman and Co.
161. Friedberg, E.C., G.C. Walker, and W. Siede, *DNA Repair and Mutagenesis*. 1995, Washington, D.C.: ASM Press.
162. Bromidge, T.J., *et al.*, *Adaptation of the TdT assay for semi-quantitative flow cytometric detection of DNA strand breaks*. *Cytometry*, 1995. **20**(3): p. 257-260.
163. Liu, L.F., *Anticancer drugs that convert DNA topoisomerases into DNA damaging agents*, in *DNA Topology and its Biological Effects*, N.R. Cozzarelli and J.C. Wang, Editors. 1990, Cold Spring Harbor Press: Cold Spring Harbor. p. 371-390.
164. Mitchell, L.G., R. Vaghmar, and J.J. Murtagh, Jr., *Biotin-enhanced sequencing of PCR products and generation of single-strand DNA probes*, in *Advances in Biomagnetic Separation*, M. Uhlen, E. Hornes, and O. Olsvik, Editors. 1994, Eaton Publishing Co.: Natick, MA. p. 31-48.
165. Feinberg, A.P. and B. Vogelstein, *A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity*. *Anal. Biochem.*, 1983. **132**: p. 6-13.
166. Lindahl, T., *Repair of intrinsic DNA lesions*. *Mut. Res.*, 1990. **238**: p. 305-311.
167. Blackburn, E.H., *Structure and function of telomeres*. *Nature*, 1991. **350**: p. 569-773.
168. Mead, D.A., *et al.*, *A universal method for the direct cloning of PCR amplified nuclei acid*. *Biotechnology (NY)*, 1991. **9**: p. 657-663.
169. Lockhart, D.J., *et al.*, *Expression monitoring by hybridization to high-density oligonucleotide arrays*. *Nature Biotech.*, 1996. **14**: p. 1675-1680.
170. Chee, M., *et al.*, *Accessing genetic information with high-density DNA arrays*. *Science*, 1996. **274**: p. 610-614.
171. Drouin, R., S. Gao, and G.P. Holmquist, *Agarose gel electrophoresis for DNA damage analysis*, in *Technologies for detection of DNA damage and mutations*, G.P. Pfeifer, Editor. 1996, Plenum Press: New York, NY. p. 37-43.
172. Holman, J.P. and W.J. Gajda, *Experimental methods for engineers*, . 1978, McGraw-Hill: New York. p. 56.
173. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning, A Laboratory Manual*. Second ed. 1989, Cold Spring Harbor: Cold Spring Harbor Press.