DNA-PROTEIN CROSS-LINKING BY
BIFUNCTIONAL DNA ALKYLATED AGENTS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Erin Denise Michaelson-Riche

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Natalia Tretyakova, Advisor

March 2010
Acknowledgements

First and foremost, I would like express my deep gratitude to my advisor, Dr. Natalia Tretyakova, for providing guidance, support, and encouragement through the last three years. I am also grateful to my co-advisor, Dr. Colin Campbell, for providing support, guidance, and hands-on training throughout this project. I also wish to thank Dr. Rory Remmel for serving on my committee and providing support and valuable feedback through the years. I would also like to acknowledge Dr. Edgar Arriaga, for serving as my coadvisor while I was a trainee on the NIH Chemistry-Biology Interface Training Grant (T32-GM08700) from 2007–2008. In addition, the work in this thesis would not have been possible without the contributions of Dr. Daniel Liebler and Dr. Simona Codreanu (Vanderbilt University), who performed all of the proteomics analyses.

I also wish to thank those who directly contributed to the development of this work, especially Dr. Rachel Loeber, whose hard work and dedication laid the groundwork for this thesis. I would also like to acknowledge Jamie Kurtz for invaluable contributions to the AGT monooxide work, and Xun Ming, B.J. Valente, and Khavong Pha for helping process countless samples. I am also grateful to former and current members of the Tretyakova laboratory for providing friendship, moral support, technical assistance, and insightful discussions, especially Dr. Melissa Goggin, Brock Matter, Dr. Rebecca Guza, Dr. Bhaskar Malayappan, Dr. Delshanee Kotandeniya, Uthpala Seneviratne, Cate Tsufis, and Srikanth Kotapati. I would especially like to acknowledge Melissa and Brock for providing training and assistance with MS and HPLC. I am also grateful to Bob Carlson for help preparing figures for publication.

Finally, I wish to thank my family and friends, especially my husband, Michael Richie, and my parents, Doug and Gwen Michaelson, for their unwavering love, support, and encouragement, without which I could not have completed this degree.
Dedication

This thesis was written in loving memory of my grandfather, Edward W. Shultz.

You will forever remain in my heart.

(1928 – 1997)
Abstract

Many common DNA alkylating agents, such as environmental toxins and chemotherapeutic drugs, are bis-electrophiles capable of covalently cross-linking cellular biomolecules. While DNA-DNA cross-linking by such compounds is well-characterized, the identities and the biological effects of the corresponding DNA-protein cross-links (DPCs) are poorly understood. Furthermore, because bis-electrophiles produce DNA-DNA cross-links and DNA monoadducts in addition to DPCs, it is difficult to establish the biological outcomes specifically resulting from DPC lesions. The purpose of the present work was to characterize DNA-protein cross-linking by two bis-electrophiles, 1,2,3,4-diepoxybutane (DEB) and bis(2-chloroethyl)methylamine (mechlorethamine), and to evaluate the ability of DPCs to induce cytotoxic and mutagenic effects. Mass spectrometry-based proteomics and immunological detection methods identified 41 proteins participating in DPC formation in the presence of DEB in nuclear protein extracts prepared from human cervical carcinoma (HeLa) cells, and 38 proteins which formed DPCs to the chromosomal DNA of human fibrosarcoma (HT1080) cells treated with mechlorethamine. Relative to their cellular abundance, a disproportionately high number of the proteins involved in DPC formation were nuclear proteins with known nucleic acid-binding capabilities which participate in cellular processes such as transcriptional regulation and DNA repair. HPLC-ESI\(^+\)-MS/MS analysis of total proteolytic digests of DPCs revealed the chemical structures of the cross-links produced by DEB and mechlorethamine to be 1-(S-cysteinyl)-4-(guan-7-yl)-2,3-butanediol (Cys-N7G-BD) and N-[2-(S-cysteinyl)ethyl]-N-[2-(guan-7-yl)ethyl]methylamine (Cys-N7G-EMA), respectively. In order to analyze the biological consequences of DPC lesions, we selectively induced DPCs in mammalian cell cultures by electroporating them in the presence of epoxide-containing protein reagents. Significant levels of cell death and mutations were observed, suggesting that DPC lesions contribute to the biological effects of bis-electrophiles.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>I. LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview of DNA-Protein Cross-linking</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Types of DPCs</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Physical and Chemical Agents Involved in DPC Formation</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3 Stability and Repair of DPCs</td>
<td>8</td>
</tr>
<tr>
<td>1.1.4 Biological Consequences of Unrepaired DPCs</td>
<td>12</td>
</tr>
<tr>
<td>1.2 Methods for the Detection and Characterization of DNA-Protein Cross-links</td>
<td>15</td>
</tr>
<tr>
<td>1.2.1 Classical Techniques Employed in the Study of DPCs</td>
<td>15</td>
</tr>
<tr>
<td>1.2.2 Mass Spectrometry: Characterization of DNA-Protein Cross-linking</td>
<td>18</td>
</tr>
<tr>
<td>1.2.3 Proteomics Technologies: Identification of Proteins which Form DPCs</td>
<td>20</td>
</tr>
<tr>
<td>1.3 DNA-Protein Cross-linking by O(^6)-Alkylguanine DNA Alkyltransferase</td>
<td>23</td>
</tr>
<tr>
<td>1.3.1 Structure and Function of AGT Protein</td>
<td>24</td>
</tr>
<tr>
<td>1.3.2 DNA-Protein Cross-linking by AGT</td>
<td>26</td>
</tr>
<tr>
<td>1.4 Cross-linking Agents Examined in this Work</td>
<td>28</td>
</tr>
<tr>
<td>1.4.1 1,2,3,4-Diepoxybutane</td>
<td>29</td>
</tr>
<tr>
<td>1.4.2 Antitumor Nitrogen Mustards</td>
<td>31</td>
</tr>
<tr>
<td>1.5 Goals of this Work</td>
<td>34</td>
</tr>
</tbody>
</table>
## II. DNA-PROTEIN CROSS-LINKING BY 1,2,3,4-DIEPOXYBUTANE

2.1 Introduction .................................................................................................................. 35

2.2 Materials and Methods .................................................................................................. 38

2.3 Results .......................................................................................................................... 46

2.3.1 Strategy for Affinity Purification of Proteins Cross-linked to DNA by DEB .......................................................... 46

2.3.2 Concentration-Dependent Formation of DPCs in Nuclear Protein Extracts Following DEB Treatment ................................................. 47

2.3.3 Mass Spectrometric Identification of Cross-linked Proteins ....................................... 47

2.3.4 Identification and Quantitation of Individual Proteins that form Cross-links to DNA in the Presence of DNA by Western Blot Analysis ......................................................... 56

2.3.5 Identification of Amino Acid Side Chains that Participate in DPC Formation .......................................................... 59

2.4 Discussion .................................................................................................................... 61

## III. DNA-PROTEIN CROSS-LINKING IN HUMAN FIBROSARCOMA HT1080 CELLS BY THE ANTITUMOR NITROGEN MUSTARD MECLRETHAMINE

3.1 Introduction .................................................................................................................. 65

3.2 Materials and Methods .................................................................................................. 69

3.3 Results .......................................................................................................................... 78

3.3.1 Cytotoxicity Experiments to Establish Optimal Mechlorethamine Concentrations ........................................................................... 78

3.3.2 Strategy for the Isolation of DPCs from Mammalian Cells Treated with Mechlorethamine ........................................................................... 78
V. SUMMARY AND CONCLUSIONS

VI. FUTURE DIRECTIONS

6.1 Quantitation of DPCs and Other Bifunctional DNA Adducts In Vivo

6.2 Identification of Cross-linking Sites Within Proteins

6.3 Quantitative Proteomic Analysis of DPCs in Nuclear Protein Extracts and Cell Cultures

VII. BIBLIOGRAPHY
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Nuclear extract proteins from human cervical carcinoma (HeLa) cells that form DPCs in the presence of DEB</td>
<td>52</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Proteins that form cross-links to chromosomal DNA in human fibrosarcoma (HT1080) cells in the presence of mechlorethamine</td>
<td>87</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Cytotoxicity and mutagenicity of Cys\textsuperscript{150} AGT monoepoxide in human fibrosarcoma (HT1080) cells</td>
<td>128</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Chart 1.1  Chemical structures of representative bis-electrophiles known to induce DPCs……………………………………………………………………………………………………3

Chart 1.2  Chemical structures of selected DPC lesions………………………………………………………………………………………………………5

Scheme 1.1  DNA damage by bis-electrophiles……………………………………………………………………………………………………………………14

Scheme 1.2  Experimental scheme for a mass spectrometry-based proteomics experiment…21

Scheme 1.3  Mechanism for the removal of O6-alkylguanine lesions by AGT………………25

Scheme 1.4  AGT-DNA cross-linking by DEB and antitumor nitrogen mustards…………………………………………………………………………………………………27

Scheme 1.5  Metabolic activation of BD and formation of DNA adducts by DEB………30

Scheme 1.6  Activation and reactivity of nitrogen mustards………………………………………32

Scheme 2.1  Formation of DPCs by DEB……………………………………………………………………………………………………………………………36

Scheme 2.2  Experimental scheme for biotin capture enrichment of DPCs from nuclear protein extracts incubated with DEB in the presence of double-stranded DNA………39

Figure 2.1  Concentration-dependent formation of DPCs in HeLa nuclear protein extracts following exposure to DEB……………………………………………………………………48

Figure 2.2  SDS-PAGE analysis of DPCs produced in HeLa nuclear protein extracts following exposure to DEB …………………………………………………….50

Figure 2.3  Examples of HPLC-ESI+ -MS/MS spectra of tryptic peptides used for protein identification of Fen-1, GAPDH, and PARP……………………………………51

Figure 2.4  Cellular functions of human proteins that form DPCs in the presence of DEB…………………………………………………………………………………………55

Figure 2.5  Western blot analysis of DEB-induced DPCs in nuclear protein extracts from HeLa cells……………………………………………………………………………57

Figure 2.6  HPLC-ESI+ -MS/MS analysis of a tryptic peptide from actin containing a DEB-induced cross-link to guanine…………………………………………………………60

Figure 2.7  HPLC-ESI+ -MS/MS analysis of Cys-N7G-BD conjugates in total proteolytic digests of DEB-induced DPCs……………………………………………………………62
Scheme 3.1  Formation of DPCs by antitumor nitrogen mustards………………………….66
Scheme 3.2  Strategy for the isolation and analysis of DPCs from mechlorethamine-treated mammalian cell cultures……………………………………………………………………………………………………………………………………………………………………68
Figure 3.1  Mechlorethamine-induced cell death in HT1080 cells……………………………79
Figure 3.2  Representative HPLC-UV chromatogram for the separation of nucleoside mixtures resulting from the enzymatic digestion of nucleic acids that were isolated from HT1080 cells treated with mechlorethamine to induce DNA-protein cross-linking……………………………………………………………..81
Figure 3.3  SDS-PAGE analysis of concentration-dependent mechlorethamine-induced DNA-protein cross-linking in HT1080 cells………………………………………………………………………………………………………………………………………………82
Figure 3.4  SDS-PAGE analysis of DPCs from HT1080 cells treated with mechlorethamine……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
| Figure 4.4 | HPLC-ESI-MS/MS analysis of tryptic peptides from Cys\textsuperscript{150} AGT monoepoxide | 122 |
| Figure 4.5 | SDS-PAGE analysis of the time course of AGT monoepoxide formation | 124 |
| Figure 4.6 | SDS-PAGE analysis of the concentration-dependence of AGT monoepoxide formation | 126 |
| Figure 4.7 | Cytotoxicity of Cys\textsuperscript{150} AGT monoepoxide in Chinese hamster lung fibroblast cells | 130 |
| Figure 4.8 | Cytotoxicity of BSA monoepoxide in human fibrosarcoma HT1080 cells | 132 |
| Scheme 6.1 | HPLC-ESI-MS/MS analysis of bifunctional DEB-DNA adducts | 142 |
| Scheme 6.2 | Experimental scheme for biotin capture enrichment of modified peptides from nuclear extract proteins cross-linked to DNA in the presence of bis-electrophiles | 144 |
| Scheme 6.3 | Experimental scheme for a 4-plex iTRAQ quantitative proteomics experiment | 146 |
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT</td>
<td>$O^6$-alkylguanine DNA alkyltransferase</td>
</tr>
<tr>
<td>aza-dC</td>
<td>5-aza-2’-deoxycytidine</td>
</tr>
<tr>
<td>BD</td>
<td>1,3-butadiene</td>
</tr>
<tr>
<td>Bis-N7G-BD</td>
<td>1,4-\textit{bis}(guan-7-yl)-2,3-butanediol</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CHL</td>
<td>Chinese hamster lung cell line</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary cell line</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>Cisplatin</td>
<td>\textit{cis}-diaminedichloroplatinum</td>
</tr>
<tr>
<td>Cys (or C)</td>
<td>cysteine</td>
</tr>
<tr>
<td>Cys-N7G-BD</td>
<td>1-(S-cysteinyl)-4-(guan-7-yl)-2,3-butanediol</td>
</tr>
<tr>
<td>Cys-N7G-EMA</td>
<td>N-[2-(S-cysteinyl)ethyl]-N-[2-(guan-7-yl)ethyl]methylamine</td>
</tr>
<tr>
<td>Cys-N7G-PBA</td>
<td>N-[2-(S-cysteinyl)ethyl]-N-[2-(guan-7-yl)ethyl]-p-aminophenylbutyric acid</td>
</tr>
<tr>
<td>Cys-[^{15}\text{N}]-N7G-EMA</td>
<td>N-[2-(S-cysteinyl)ethyl]-N-[2-(^{15}\text{N}-guan-7-yl)ethyl]methylamine</td>
</tr>
<tr>
<td>dA</td>
<td>deoxyadenosine</td>
</tr>
<tr>
<td>DEB</td>
<td>1,2,3,4-diepoxybutane</td>
</tr>
<tr>
<td>dG</td>
<td>deoxyguanosine</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle’s Medium</td>
</tr>
<tr>
<td>DPC</td>
<td>DNA-protein cross-link</td>
</tr>
<tr>
<td>dT</td>
<td>deoxothymidine</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EB</td>
<td>3,4-epoxy-1-butene</td>
</tr>
<tr>
<td>EBD</td>
<td>3,4-epoxy-1,2-butanediol</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>EF-1α1</td>
<td>elongation factor 1-α1</td>
</tr>
<tr>
<td>EMSA</td>
<td>electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>ESI⁺</td>
<td>elestrospray ionization, positive mode</td>
</tr>
<tr>
<td>FA</td>
<td>formaldehyde</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>Fen-1</td>
<td>flap endonuclease-1</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>Glu (or E)</td>
<td>glutamine</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>GSH</td>
<td>glutathione</td>
</tr>
<tr>
<td>Gua (or G)</td>
<td>guanine</td>
</tr>
<tr>
<td>hAGT</td>
<td>human recombinant AGT</td>
</tr>
<tr>
<td>HEB</td>
<td>2-hydroxy-3,4-epoxybut-1-yl</td>
</tr>
<tr>
<td>HeLa</td>
<td>human cervical carcinoma cell line</td>
</tr>
<tr>
<td>His (or H)</td>
<td>histidine</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogeneous nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HPRT (or hprt)</td>
<td>hypoxanthine guanine phosphoribosyl transferase</td>
</tr>
<tr>
<td>HR</td>
<td>homologous recombination</td>
</tr>
<tr>
<td>HT1080</td>
<td>human fibrosarcoma cell line</td>
</tr>
<tr>
<td>Ku</td>
<td>ATP-dependent DNA helicase, subunit 2</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>Lys (or K)</td>
<td>lysine</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix assisted laser desorption/ionization</td>
</tr>
<tr>
<td>Mechlorethamine</td>
<td>bis(2-chloroethyl)methylamine</td>
</tr>
<tr>
<td>6-MP</td>
<td>6-mercaptopurine</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>NER</td>
<td>nucleotide excision repair</td>
</tr>
<tr>
<td>NSI</td>
<td>nanospray ionization</td>
</tr>
<tr>
<td>1,N\textsuperscript{6}-HMHP-dA</td>
<td>1,N\textsuperscript{6}-(1-hydroxymethyl-2-hydroxypropan-1,3-diyl)-2\textsuperscript{-deoxyadenosine}</td>
</tr>
<tr>
<td>N7G-N1A-BD</td>
<td>1-(guan-7-yl)-4-(aden-1-yl)-2,3-butanediol</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PARP</td>
<td>poly(ADP-ribose) polymerase</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethanesulfonyl fluoride</td>
</tr>
<tr>
<td>pol</td>
<td>polymerase</td>
</tr>
<tr>
<td>Ref-1</td>
<td>DNA-(apurinic- or apyrimidinic-site) lyase</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Ser (or S)</td>
<td>serine</td>
</tr>
<tr>
<td>SRM</td>
<td>selected reaction monitoring</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-buffered saline</td>
</tr>
<tr>
<td>THB</td>
<td>trihydroxybutyl</td>
</tr>
<tr>
<td>TIC</td>
<td>total ion current</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>XRCC</td>
<td>x-ray cross-complementing protein</td>
</tr>
</tbody>
</table>
I. LITERATURE REVIEW

1.1 Overview of DNA-Protein Cross-linking

DNA-protein cross-links (DPCs) are structurally diverse DNA lesions that form when a protein becomes irreversibly trapped on DNA as a result of exposure to endogeneous or exogeneous physical and chemical agents. These bulky lesions impose a steric hindrance that can interfere with DNA-protein interactions and block critical cellular processes. Despite their potential to harm cells, the role of DPCs in mutagenesis and carcinogenesis is not well understood, and their identities have not been well characterized. A comprehensive understanding of DPC formation, repair, and biological effects is necessary in order to fully realize the biological consequences of DNA damaging agents.

1.1.1 Types of DPCs

In living cells, DPCs are generated by two major pathways: exposure to endogenous and environmental alkylating agents and normal DNA metabolism. Class I DPCs occur when topoisomerase I becomes trapped at the 3’ end of a DNA single strand nick, a process which is believed to occur when the covalent reaction intermediate consisting of a tyrosine-phosphodiester linkage between topoisomerase I and DNA is frozen by nearby DNA damage or exposure to topoisomerase poisons. Class I DPCs are removed by the tyrosine-DNA phosphodiesterase Tdp1.

Class II DPCs occur when topoisomerase II is trapped at the 5’ ends of a double-strand break, likely through a mechanism similar to that of class I DPC formation. Alternatively, class II DPCs can involve Spo11, a meiotic protein that initiates recombination by making a staggered double-strand break in the DNA. Each side of the dimeric protein may become transiently linked
to the resulting 5’ ends. Class II DPCs are removed by an endonuclease that makes an incision in
the DNA several nucleotides from the cross-link site.\textsuperscript{5}

The formation of Class III DPCs occurs during attempted base excision repair.\textsuperscript{2} Cleavage of
an abasic site by DNA-(apurinic- or apyrimidinic-site) lyase (Ref-1) generates a 2-
deoxyribonolactone which can become cross-linked to repair proteins through an amide linkage at
the 5’ end of the DNA strand.\textsuperscript{6} A likely candidate from cross-linking via this mechanism is DNA
polymerase (pol) β, which is the primary human protein involved in the removal of 2-
deoxyribonolactone.\textsuperscript{7}

The fourth major class of DPCs is perhaps the most common form of DPCs, and comprises
those DPCs which are induced by reactions with endogenous and exogenous chemical and
physical agents.\textsuperscript{2} Different agents induce Class IV DPCs by many different mechanisms, resulting
in the formation of cross-links with a broad range of distinct chemical and biological
characteristics.\textsuperscript{1} The resulting diversity in the properties of the cross-links and the identities of the
cross-linked proteins makes it challenging to fully understand the biological consequences of
these lesions. The present work focuses on Class IV DPCs.

\subsection*{1.1.2 Physical and Chemical Agents Involved in DPC Formation}

Class IV DPCs can form as a result of exposure to a wide variety of cytotoxic, mutagenic,
and carcinogenic agents, including environmental pollutants, industrial chemicals,
chemotherapeutic drugs, and endogenous physiological metabolites (Chart 1.1).\textsuperscript{1} Of these,
formaldehyde (FA) is the most thoroughly studied DPC-inducing agent. FA is a genotoxic
chemical employed in the manufacture of thousands of household, medicinal, and industrial
products, leading to widespread human exposure to FA in industrial and household settings.\textsuperscript{8,9}
Additionally, FA can be formed in the liver through the metabolism of ingested or inhaled
**Chart 1.1** Chemical structures of representative *bis*-electrophiles known to induce DPCs

- **Nitrogen mustards**
- **1,2,3,4-Diepoxybutane (DEB)**
- **Methylglyoxal**
- **Cisplatin**
- **Formaldehyde (FA)**
- **Acrolein**
- **Alkylnitrosoureas**
- **2'-Deoxyxanosine**
- **Mitomycin C**
compounds.\textsuperscript{9} It has long been known that FA exerts genotoxic effects in bacterial and eukaryotic cells. Early research demonstrated its ability to induce point mutations and gene deletions.\textsuperscript{8} Although FA can induce several types of DNA lesions, the vast majority of FA-induced DNA lesions are DPCs.\textsuperscript{10,11} FA adducts are formed \textit{via} Schiff base reactions at amide, amine, hydroxyl, and thiol functionalities, such as those found on arginine and lysine residues within proteins and adenine, cytosine, and guanine bases within DNA (\textit{Chart 1.2}).\textsuperscript{1,12,13} Early evidence of FA-induced DPC formation was discovered in \textit{Escherichia coli} which, when exposed to low doses of FA, exhibited interstrand DNA cross-links mediated by a covalent protein bridge.\textsuperscript{14} Rhesus monkeys exposed to inhaled FA exhibited a rate of formation of DPCs which was proportional to the tissue concentration of FA,\textsuperscript{15} an observation which has enabled the utilization of DPCs as a biomarker of human FA exposure.\textsuperscript{16} Proteins in close proximity to chromosomal DNA are more likely to form cross-links to DNA in the presence of FA, as has been observed with all five major histone proteins (H1, H2a, H2b, H3, and H4).\textsuperscript{17} Histone-DNA cross-links are considered a major contributor to FA tumorigenesis and carcinogenicity.\textsuperscript{18}

Metal compounds comprise another class of DPC-inducing agents commonly present in environmental and industrial pollutants.\textsuperscript{1} Metals such as nickel, arsenic, and chromium have been implicated in a variety of cancers, including those of the lung, liver, bladder, intestine, mouth, skin, and blood.\textsuperscript{19} Nickel (II) is a carcinogenic compound employed in industry for the production of alloys used to make consumer products such as coins, jewelry, and stainless steel.\textsuperscript{20} The most significant route of human exposure to nickel is through inhalation of nickel-containing dust and fumes. Studies in rat lymphocytes suggest that nickel exposure results in the generation of hydroxyl radicals capable of inducing DPCs \textit{in vivo}.\textsuperscript{21} Arsenic is an environmental contaminant present in soil, water, and air particles,\textsuperscript{22,23} enabling human exposure \textit{via} inhalation and ingestion in the form of arsenite (As\textsuperscript{III}) or arsenate (As\textsuperscript{V}).\textsuperscript{19} Arsenite treatment has been shown to induce DPCs in multiple human cell lines,\textsuperscript{24,25} suggesting that DPCs may contribute to the toxic effects
**Chart 1.2** Chemical structures of selected DPC lesions produced by formaldehyde (A), oxidation (B), acrolein (C), ionizing radiation (D), reactive oxygen species (E), cisplatin (F), and reactive nitrogen species (G).
associated with human arsenic exposure. Chromium compounds are common industrial pollutants resulting from the large-scale manufacture of products such as stainless steel and pressure-treated wood. Hexavalent chromium is a contaminant often found in drinking water. Recent work in the erythrocytes of freshwater fish with acute exposure to hexavalent chromium and divalent nickel has shown that DPCs are a potential biomarker for exposure to metal cations.

DPCs can also form as a result of exposure to ultraviolet (UV) and ionizing radiation. In the 1970s, Hawkins et al. employed a phenol extraction-based strategy to study DPC formation in bacteriophage exposed to ionizing radiation. DNA and proteins were isotopically labeled with $^{14}$C and $^{32}$P, respectively. DNA containing cross-linked proteins was isolated by phenol extraction, and the number of DPC adducts formed was estimated by monitoring the mobility of the radiolabel into the different extraction phases. Mass spectrometry studies have shown that radiation-induced DPCs frequently involve thymine bases within DNA and tyrosine residues within proteins (Chart 1.2). Interestingly, ionizing radiation can induce DPCs under both aerated and hypoxic conditions, although fewer DPCs are formed in the presence of oxygen. Cress and colleagues employed a filter-binding assay to study DNA-protein cross-linking in Chinese hamster ovary (CHO) cells, and found that radiation-induced DPC formation is dose-dependent and may involve specific proteins. Recently, Barker et al. employed mass spectrometry-based proteomics to identify 29 proteins in CHO AA8 and human fibroblast GM00637 cells which became cross-linked to chromosomal DNA following exposure to ionizing radiation. A disproportionate number of the identified proteins were involved in cellular functions such as transcriptional regulation, RNA splicing, and chromatin regulation, supporting the findings of Cress et al. that radiation-mediated DPC formation involves specific proteins.

Several endogenous chemicals can also contribute to DPC formation, including reactive oxygen species and lipid peroxidation. Oxidation of DNA nucleobases produces electron-deficient species that can react with nearby proteins, giving rise to covalent DPCs. For example,
the bacterial DNA repair enzyme MutY can become cross-linked to DNA via the 8-oxoguanosine bases which it targets for repair. Xu and colleagues have reported that 8-oxoguanosine-mediated DPC formation involves cross-linking to tyrosine residues, and Ban et al. have suggested that reactive oxygen species may form DPCs involving tyrosine and cytosine (Chart 1.2). Induction of DPCs can also result from peroxidation of proteins by radiation-generated free radicals, as has been observed in vitro for insulin and bovine serum albumin (BSA). In addition, DPCs can also form between oxanine, the nitric oxide-induced deamination product of guanine, and DNA binding proteins (Chart 1.1). Oxanine-mediated DPCs can involve several amino acid residues, including lysine, arginine, and serine (Chart 1.2), and may be repaired by an excision nuclease.

Endogenous aldehydes, such as methylglyoxal, comprise another class of endogenous DPC-inducing agents (Chart 1.1). Methylglyoxal is a highly reactive α-ketoaldehyde which can react with guanine bases in DNA and lysine and arginine residues within proteins. This compound has been shown to covalently cross-link bacterial DNA pol I to DNA in vitro, suggesting that methylglyoxal-induced DPCs may severely inhibit DNA replication. Another endogenous aldehyde which is known to induce DPCs is acrolein (Chart 1.1). This α,β-unsaturated aldehyde is an environmental toxin and a product of endogenous lipid peroxidation. It is likely that acrolein-mediated cross-links form via Michael addition or Schiff base reactions, and may involve histidine side chains in proteins and the N2 position of guanine bases within DNA (Chart 1.2). Acrolein-mediated DNA-peptide cross-links may be subject to error-prone replication bypass, suggesting that DPCs may contribute to acrolein-mediated genotoxicity.

Many bis-electrophilic chemotherapeutic agents are also capable of forming covalent cross-links between cellular biomolecules (Chart 1.1). One such class of drugs are the antitumor platinum compounds, including cis-diamminedichloroplatinum(II) (cisplatin) and its trans-isomer. Cisplatin undergoes sequential chloride displacement steps to produce a positively-
charged bis-hydrated species that can interact with negatively-charged DNA through electrostatic interactions. Although the cytotoxic properties of these agents have been attributed to their ability to form guanine-guanine cross-links in DNA, it is likely that DPCs (Chart 1.2) also contribute to the cytotoxicity of these common platinum drugs. Haloalkynitrosoureas are another class of chemotherapeutic drugs which can form several types of DNA damage, including DNA strand breaks and DPCs. Antitumor antibiotics, such as mytomycin C, also form DPCs which may contribute to the cytotoxic effects of these agents. Two classes of bis-electrophiles examined in this thesis are antitumor nitrogen mustards, such as bis(2-chloroethyl)methylamine (mechloretamine), and epoxide-containing compounds, such as 1,2,3,4-diepoxybutane (DEB) (see Section 1.4).

### 1.1.3 Stability and Repair of DPCs

DPC lesions produced by different cross-linking agents have unique chemical structures (Chart 1.2) and significantly vary in their hydrolytic stability. For example, DPCs produced by acetaldehyde are hydrolytically unstable, with less than 25% of cross-links remaining after only 8 h at 37°C in vitro. In contrast, DNA-histone cross-links involving malondialdehyde have a half-life of 13.4 days under the same conditions. Many aldehyde-mediated DPCs, such as those produced by FA, are easily reversed by spontaneous hydrolysis or by heating at high temperatures. This property was recently exploited by Qiu and Wang, who employed FA-mediated cross-linking to identify DNA-binding proteins in human acute promyelocytic leukemia HL-60 cells. In contrast, DNA-histone cross-links formed in nucleosome core particles under conditions of oxidative stress are stable for at least 48 h under physiological conditions. There is evidence that DPCs formed by metal cations persist even longer. For example, Kuykendall and colleagues studied DNA-protein cross-linking in freshwater fish erythrocytes, and found that
divalent nickel- and hexavalent chromium-induced DPCs were present 10 and 20 days post-exposure, respectively, suggesting that residual metal cations may contribute to DPC formation long after acute exposure.\textsuperscript{26} Interestingly, Chen \textit{et. al.} found that the stability of oxanine-mediated DPCs depends in part on the identity of the amino acid which participates in the cross-link, with cysteine and glycine cross-links persisting longer than those involving tyrosine.\textsuperscript{39}

Many factors may contribute to the stability and persistence of DPCs both \textit{in vitro} and \textit{in vivo}. While some DPC lesions, such as those produced by aldehydes, can be eliminated through rapid hydrolysis,\textsuperscript{9} hydrolytically stable DPCs require active repair in order to be removed. The large size of DPCs compared to other DNA adducts poses unique challenges for cellular repair machinery. Because of the broad substrate specificity of nucleotide excision repair (NER) pathways toward other bulky DNA adducts,\textsuperscript{59,60} many groups have proposed roles for NER in the repair of DPCs.\textsuperscript{1} In NER, a damaged DNA base is removed by making incisions at phosphodiester bonds in the DNA backbone both 5’ and 3’ to the damage site.\textsuperscript{59} In prokaryotes, the UvrABC nuclease complex excises an oligomer that is 12-13 nucleotides in length,\textsuperscript{59} while eukaryotic excision nucleases produce 24-32 nucleotide-long oligomers. Following removal of the adduct-containing oligonucleotide, the resulting gap is filled by replication proteins involved in repair synthesis. Defects in NER have been linked to several human diseases, most notably xeroderma pigmentosum.\textsuperscript{61}

Several reports have implicated NER pathways in the removal of DPC lesions. For example, Minko and colleagues generated structurally-defined DPC lesions by covalently trapping T4 pyrimidine DNA glycosylase on apurinic and apyrimidinic sites within duplex DNA.\textsuperscript{62} Incubation of the resulting DPC substrate with bacterial UvrABC nuclease \textit{in vitro} resulted in the formation of a 12-mer DNA incision product containing the covalently cross-linked protein, suggesting a role for NER in the repair of bulky DPC lesions.\textsuperscript{62} Interestingly, the same group later found that DNA-peptide cross-links are removed by this pathway with even
greater efficiency.\textsuperscript{63} Several other groups have reported on the effective removal of DNA-peptide cross-links by NER pathways, and suggested that proteosomal degradation of large DPCs is important to NER-mediated repair.\textsuperscript{9,64-66}

The experimental evidence described above is consistent with a DPC repair model in which NER is coupled with proteolytic degradation of the cross-linked proteins to remove DPCs from duplex DNA. In the replication-coupled repair model proposed by Reardon \textit{et. al.}, a replication complex encounters a DPC lesion, triggering the recruitment of proteolytic enzymes necessary for the degradation of the protein to a polypeptide.\textsuperscript{2} Simultaneously, the necessary repair factors are recruited to the lesion site so that once the DPC is partially digested, a 24-32 nucleotide long segment of the DNA which contains the cross-linked peptide can be removed by NER.\textsuperscript{2} This model is consistent with evidence that normal and NER-deficient cells treated with a proteasome inhibitor suffer a loss in DPC repair activity.\textsuperscript{9} A similar transcription-dependent repair mechanism has been proposed for the removal of Class I and Class II DPCs involving topoisomerases.\textsuperscript{67,68}

Nakano and colleagues studied DPC formation in repair deficient bacteria by two cross-linking agents known to form DPCs, FA and 5-aza-2’-deoxycytidine (aza-dC).\textsuperscript{69} They found that NER-deficient uvrA mutants were hypersensitive to FA but not aza-dC, suggesting that bacterial DPC tolerance requires multiple repair pathways.\textsuperscript{69} Interestingly, the excision efficiency of uvrABC-catalyzed NER was dependent on the size of the DPC lesion, with only DPCs involving small proteins (molecular weight less than 12-14 kDa) repaired by NER \textit{in vivo}.\textsuperscript{69} Similar results were obtained in mammalian cells, where the size limit for NER removal of DPCs was 8-10 kDa.\textsuperscript{70} At first glance, these findings were consistent with the proteolysis-coupled NER repair model described by Reardon \textit{et. al.}.\textsuperscript{2} However, further analysis of the polyubiquitination status of the cross-linked proteins, a requirement for 26 S proteasome-mediated recognition and degradation of proteins, revealed that no detectable level of ubiquitination was present in DPC-
containing DNA from mammalian cells. These findings suggest that DPCs are not polyubiquitinated, and therefore are unlikely to be subject to proteasomal degradation in mammalian cells. Similar conclusions were drawn in studies conducted in bacteria, which contain four cytosolic ATP-dependent proteases which are functionally homologous to eukaryotic proteasomes. Although mutants with deficiencies in cytosolic ATP-dependent proteases were sensitive to both FA and aza-dC, the introduction of an additional sulA mutation almost completely suppressed the sensitivity, suggesting that the observed sensitivity was due to a mutation in the gene encoding for the Lon protease that results in lethal effects which are unrelated to DNA repair capabilities.

An alternative mechanism by which cells may remove bulky DPC lesions is through the homologous recombination (HR) repair pathway. HR repair processes take advantage of the sequence identity between sister chromatids or the maternal and paternal copies of the same chromosomal region to repair severe DNA damage, such as double-strand breaks and cross-link lesions. HR repair of double-strand breaks begins with the formation of a heteroduplex involving the 3’ end of the DNA strand from the damaged region and the undamaged double-strand homologue region. This is followed by repair synthesis using the undamaged homologue as a template to restore intact DNA. In yeast, this process involves the Rad family of proteins, which are structurally similar to bacterial Rec proteins. In mammalian cells, HR repair involves the x-ray repair cross complementing (XRCC) family of proteins, named for a series of mammalian mutant cell lines which are hypersensitive to x-ray-induced DNA damage.

Nakano et. al. have suggested that HR has a more prominent role than does NER in cellular tolerance of DPCs, especially those involving large proteins. HR-deficient bacterial and mammalian cells were hypersensitive to DPC-inducing agents, but the sensitivity was eliminated when HR repair capabilities were restored. As determined by immunological detection methods, the treatment of mammalian cells with FA and aza-dC resulted in nuclear accumulation
of RAD51, a protein with a central role in the HR repair pathway, and phosphorylated histone H2AX, which is believed to form in conjunction with chromatin remodeling. These findings suggest that HR plays an important role in the cellular processing of FA- and aza-dC-induced DPCs. Furthermore, the comet assay revealed that double-strand breaks, which likely form when the replication fork stalls upon encountering a DPC lesion, accumulated in HR-deficient but not HR-proficient cells, suggesting that HR is required to repair double-strand breaks formed as a result of DNA-protein cross-linking.\textsuperscript{70} In bacterial cells exposed to DPC-inducing agents, repair of double-strand breaks proceeded through a RecBCD-dependent HR repair pathway.\textsuperscript{69}

It is likely that the cellular repair of DPCs is a complex process which involves multiple repair pathways. For example, de Graaf and colleagues recently studied cellular repair of FA-induced DPCs in a yeast non-essential gene deletion library.\textsuperscript{76} They found that following chronic, low-dose exposure to FA, yeast strains containing deficiencies in HR repair capabilities displayed the greatest sensitivity.\textsuperscript{76} In contrast, NER-deficient yeast strains were more sensitive to acute FA exposure, suggesting that the intermediates through which cells process FA-induced DPCs depends on the dose and duration of exposure, and that these intermediates are differentially recognized by NER and HR repair enzymes.\textsuperscript{76} Taken together with the results of Nakano \textit{et al.}\textsuperscript{69,70} these findings substantiate the hypothesis that NER and HR contribute differentially to the cellular repair and tolerance of DPCs. Further studies are needed to more fully elucidate the mechanisms by which cells repair DPC lesions.

### 1.1.4 Biological Consequences of Unrepaired DPCs

The covalent cross-linking of proteins to chromosomal DNA is likely to disrupt normal cellular functions. Because of their bulky nature and helix-distorting effects, DPC lesions can physically block the formation and progression of protein complexes involved in critical cellular
processes such as transcription, replication, chromatin remodeling, and DNA repair. Many DPC-inducing agents are known to interfere with DNA replication, and DPCs are believed to contribute to the antitumor properties of many common chemotherapeutic agents.

Recently, Yamanaka and colleagues reported on the role of lesion bypass DNA polymerases in translesion synthesis past DNA-peptide cross-links. They employed an N6-deoxyadenosine (dA) acrolein-mediated DNA-peptide cross-link as a model for bulky major groove DNA adducts. The DNA-peptide cross-link was fixed on a 30-mer synthetic DNA duplex oligomer, which was subsequently annealed to a radio-labeled primer containing a free 3’-OH three nucleotides upstream from the lesion site. Interestingly, pol ν was able to fully extend the primer past the lesion and correctly incorporated a deoxythymidine (dT) opposite the adducted site, suggesting that pol ν catalyzed error-free bypass past this major groove adduct. In contrast, the presence of a DNA-peptide cross-link situated in the minor groove through a linkage at N2-deoxyguanosine (dG) resulted in complete blockage of primer extension, suggesting that pol ν is not capable of translesion synthesis past bulky minor groove adducts. These findings may have significant implications for DPCs involving proteins known to bind in the minor groove of DNA.

Although there is ample evidence demonstrating that agents capable of causing DPCs are genotoxic to cells, the majority of these agents also produce other types of DNA damage, including monoadducts and DNA-DNA cross-links, which also contribute to the observed cytotoxic and mutagenic effects of these agents (Scheme 1.1). Therefore, it is difficult to fully comprehend the biological implications of unrepaired DPCs. In order to effectively study the mechanisms of DPC formation and repair, as well as the biological consequences of unrepaired DPCs, it is necessary to develop novel reagents capable of selectively inducing DPCs in cell while avoiding other types of damaging DNA lesions (see Chapter IV).
Scheme 1.1 DNA damage by *bis*-electrophiles\textsuperscript{49}

![Diagram of DNA damage by *bis*-electrophiles](image)
1.2 Methods for Detection and Characterization of DNA-Protein Cross-links

Numerous methodologies have been applied to the study of DNA-protein cross-linking. Early studies provided valuable insight into the abilities of several cross-linking agents to form DPCs, both in vitro and in vivo, and enabled the relative quantitation of cellular proteins which associate with DNA following exposure to cross-linking agents. More recently, advanced technologies such as mass spectrometry have been employed for the identification of specific proteins which participate in DNA-protein cross-linking and to characterize the chemical structures of the resulting amino acid-nucleobase conjugates.

1.2.1 Classical Techniques Employed in the Study of DPCs

One of the earliest techniques to be employed in the study of DNA-protein cross-linking was alkaline filter elution. This biophysical technique, which was developed in the 1970s by Kohn and colleagues, can be used to detect several types of DNA damage, including single-strand breaks, interstrand DNA-DNA cross-links, and DPCs. In this methodology, cells are metabolically radiolabeled with $^{14}$C-thymidine prior to exposure to DNA-damaging agents. DNA damage is then measured by monitoring the rate at which DNA from the lysed cells migrates through membrane filters under alkaline conditions via scintillation counting. Unbound DNA migrates more quickly through the filter than does DNA containing DPCs. Because interstrand DNA-DNA cross-links display an elution profile similar to that of DPCs, sample pretreatment with proteolytic enzymes is required to distinguish the two types of DNA damage. This technique has been successfully employed to study DPCs formed by several cross-linking agents, including platinum compounds, antitumor nitrogen mustards, DEB, and alkylnitrosoureas.
Another biophysical technique which has been employed in the study of DNA-protein cross-linking in cell cultures is the comet assay. In this technique, cells are first exposed to a DNA-damaging agent. To investigate bulky DNA-DNA cross-links and DPCs, which would reduce the electrophoretic mobility of the DNA, the cells are further subjected to γ-irradiation to induce DNA strand breaks. The cells are then suspended in agarose, fixed on a microscope slide, and placed in a lysis buffer to break open the cells and disperse all cellular components into the agarose matrix, leaving the DNA behind in the void formerly filled by the cell. The DNA is then subjected to electrophoretic separation and visualized using a DNA-specific fluorescent stain such as ethidium bromide. During electrophoresis, intact genomic DNA is too large to migrate into the gel and forms the comet’s head, while the damaged, fragmented DNA migrates into the gel to form the comet’s tail. Radiation-induced strand breaks enable DPCs and DNA-DNA cross-links to migrate into the gel. The length of the comet’s tail, called the tail moment, is a measure of the extent of DNA damage. As with alkaline elution, DPCs and interstrand DNA-DNA cross-links can only be distinguished by treatment with proteolytic enzymes. Merk and colleagues have employed this simple, inexpensive technique to analyze DPC formation by several cross-linking agents, including formaldehyde and chromate. However, this method cannot be used to isolate DPCs, establish the cross-link structure, or identify the cross-linked proteins.

The electrophoretic mobility shift assay (EMSA), commonly referred to as the gel-shift assay, is another technique which has been utilized for the detection of DPCs formed in vitro. In a typical EMSA experiment, proteins are incubated with radiolabeled oligonucleotides in the absence or the presence of a cross-linking agent to induce DPC formation. The reaction mixture is then subjected to electrophoretic separation on a denaturing gel under conditions that dissociate non-covalent DNA-protein complexes, and the radiolabeled DNA is visualized by autoradiography. DNA containing covalently cross-linked proteins appears as a slowly moving band relative to free DNA due to the molecular weight increase caused by the
cross-link lesion. This simple technique has been applied to demonstrate DPC formation involving specific proteins by a number of cross-linking agents, including cisplatin, nitrogen mustards, and DEB.\textsuperscript{51,93-95} Although the EMSA is a valuable tool for detecting DPCs involving specific proteins, a major limitation of the technique is that it is limited to known biological targets and cannot be employed in the study of DPC formation in more complex biological systems.

In the 1990s, Zhitkovich and Costa reported a novel technique for detecting DPCs in cell cultures and \textit{in vivo} tissue samples.\textsuperscript{96} This method is based on the observation that sodium dodecyl sulfate (SDS) binds tightly to proteins but not to DNA. Cells and tissues which have been exposed to a cross-linking agent are lysed with an SDS-containing buffer, and proteins and detergent-resistant DNA-protein complexes are co-precipitated by addition of potassium chloride.\textsuperscript{96} Upon centrifugation, free DNA remains in the supernatant and the protein pellet is subjected to proteolytic degradation to release protein-bound DNA, which is subsequently recovered and quantified. DPC content is determined by quantifying the amount of DNA which precipitated following addition of SDS. Potassium-SDS precipitation has been employed in numerous experiments to detect DPCs in cells exposed to a variety of chemical cross-linking agents.\textsuperscript{9,55} The major advantage of this technique is that it can be employed to study DPCs formed \textit{in vivo}. For example, potassium-SDS precipitation has been employed to measure DPCs formed in the blood of lead workers from a battery-manufacturing plant in Taiwan,\textsuperscript{97} and in Bulgarian workers exposed to hexavalent chromium.\textsuperscript{98} These studies revealed that DPC levels correlated with blood lead and chromium levels, respectively, suggesting that DPCs can serve as a biomarker for occupational exposure to hazardous metal cations.\textsuperscript{98,99}
1.2.2 Mass Spectrometry: Characterization of DNA-Protein Cross-linking

Based on the successful mass spectrometric characterization of DNA-DNA cross-link lesions formed by a variety of cross-linking agents,\textsuperscript{100-104} it is not surprising that mass spectrometry has recently emerged as a valuable tool in the study of DNA-protein cross-linking.\textsuperscript{105} A major advantage of this technology is that it can be used to obtain detailed structural information that is unattainable using the biophysical methods described in Section 1.2.1. Many mass spectrometry-based studies have focused on establishing the chemical structures of the amino acid-nucleobase conjugates and mapping cross-linking sites within specific proteins.

Dizdaroglu and colleagues previously employed gas chromatography-mass spectrometry (GC-MS) to identify amino acid residues which participate in DNA-protein cross-linking following exposure to hydroxyl radicals produced by γ-irradiation.\textsuperscript{106,107} Following γ-irradiation to induce cross-linking \textit{in vitro}, nucleohistones were subjected to acid hydrolysis. The resulting hydrolysates were derivatized with trimethylsilane prior to GC-MS analysis in the selective ion monitoring mode. Using this methodology, they detected cross-links involving thymine bases within DNA and tyrosine and lysine residues of the proteins.\textsuperscript{106} Quantitative analysis of these conjugates yielded a linear correlation between irradiation dose and conjugate formation. More recently, Olinski \textit{et. al.} employed this methodology to detect thymine-tyrosine cross-links in the chromatin of human cells subjected to γ-irradiation and H$_2$O$_2$ treatment (\textit{Chart 1.2}).\textsuperscript{108} These findings confirm that mass spectrometry is a valuable tool for the determination of cross-link structures.

Chen and colleagues recently employed mass spectrometry to study the ability of lysozyme amino acid side-chains to react with oxanine bases to form DPCs.\textsuperscript{39} 2’-Deoxyoxanosine is a DNA lesion formed when dG reacts with nitric oxide or nitrous acid, typically as a result of cancer or inflammatory diseases (\textit{Chart 1.1}).\textsuperscript{109} To identify amino acids that participate in DPC formation, lysozyme was incubated with 2’-deoxyoxanosine, digested with trypsin, and the resulting tryptic
peptides were analyzed by nanoscale liquid chromatography coupled with nanospray ionization tandem mass spectrometry (nanoLC-NSI-MS/MS). Several peptides were observed to have a mass increase of +152 Da, and MS/MS analysis revealed the presence of oxanine adducts at Lys, Lys, Lys, Ser, and Ser. In contrast, no mass shift of +268 Da was observed, indicating that 2′-deoxyoxanosine becomes depurinated to form oxanine adducts. Interestingly, only adducts at Lys, Ser, and Ser were observed when lysozyme was reacted with oxanine-containing calf thymus DNA, suggesting that DNA-protein interactions contribute to the formation of DPCs at specific amino acid residues. Additionally, the use of a selected reaction monitoring (SRM) mass spectrometry method revealed that the extent of modification of specific sites on lysozyme correlated with 2′-deoxyoxanosine concentration.

Mass spectrometric analysis of DPCs has also been applied to study non-covalent nucleic acid-protein interactions, which are important to numerous cellular processes, such as DNA replication, transcription, and DNA repair. Doneanu et. al. subjected DNA-protein complexes to photochemical cross-linking to determine which amino acid residues in the human replication protein A are necessary to its DNA-binding capabilities. Structural studies suggest that human replication protein A, which has roles in many DNA metabolic processes, interacts with DNA via stacking interactions involving aromatic amino acid residues and hydrogen bonding between amino acid side chains and the DNA backbone. A combination of matrix-assisted laser desorption/ionization (MALDI) MS and nanoLC-NSI-MS/MS analyses were used to identify amino acid residues that reside in the protein’s DNA-binding domain. Following UV irradiation of human replication protein A in the presence of a 30-mer oligodeoxynucleotide duplex to fix the DNA-protein complex, they were able to identify three tryptic peptides which contained covalent cross-links to DNA: V, I, and A. C-terminal proteolytic degradation of the peptides revealed the cross-linking sites to be at or near the Phe, Phe, and Phe residues, consistent with x-ray
crystal structures which placed these residues in the protein’s DNA-binding domain. These findings suggest that DNA-protein cross-linking in combination with mass spectrometry may provide information which can complement data obtained from x-ray crystallography and site-directed mutagenesis studies of non-covalent DNA-protein interactions.

1.2.3 Proteomics Technologies: Identification of Proteins which Form DPCs

Recent studies which have employed recombinant proteins to study DNA-protein cross-linking have suggested that the biological consequences of DPCs depend, in part, on the identity of the cross-linked protein. Therefore, many recent studies have focused on identifying proteins which participate in DNA-protein cross-linking. Advances in the field of mass spectrometry-based proteomics have greatly facilitated the identification and quantitation of proteins for a number of important applications, including the study of post-translational modifications, characterization of protein-protein interactions, and identification of disease biomarkers. In a typical experiment, complex protein mixtures are separated via gel electrophoresis or liquid chromatography and subjected to enzymatic digestion to peptides (Scheme 1.2). The peptides are then analyzed by mass spectrometry to obtain information regarding the peptides’ masses and amino acid sequences. Finally, this information is subjected to database searching in order to identify the proteins present in the original sample. Several approaches can be utilized for the quantitative comparison of protein levels in two or more samples, including two-dimensional gel electrophoresis and stable isotope labeling. To date, mass spectrometry-based proteomics has been employed in several experiments to identify proteins which form DPCs in the presence of a variety of cross-linking agents.

Our laboratory recently developed novel methodology which combined affinity capture with mass spectrometry-based proteomics for the analysis of DPCs formed in vitro in nuclear protein extracts exposed to bis-electrophilic cross-linking agents. Nuclear extract proteins from
Scheme 1.2 Experimental scheme for a mass spectrometry-based proteomics experiment
human cervical carcinoma (HeLa) and CHO cells were incubated with biotinylated oligodeoxynucleotides in the presence of the antitumor nitrogen mustard mechlorethamine to induce DPC formation. The DNA containing covalently cross-linked proteins was affinity captured on streptavidin beads and washed to remove non-covalently bound proteins prior to elution in an SDS-containing gel loading buffer. Following SDS-polyacrylamide gel electrophoresis (PAGE) separation and in-gel tryptic digestion of the proteins, the tryptic peptides were analyzed by high performance liquid chromatography electrospray ionization tandem mass spectrometry (HPLC-ESI⁺-MS/MS). Database searching revealed that 53 HeLa and 15 CHO nuclear proteins became cross-linked to DNA by mechlorethamine. Many of these proteins were known nucleic acid-binding proteins involved in cellular functions such as transcription, DNA replication, chromatin remodeling, and DNA repair.

Barker and colleagues have used mass spectrometry-based proteomics to identify cellular proteins which became cross-linked to chromosomal DNA in mammalian cells in the presence of ionizing radiation. Based on a method previously developed in their laboratory, chromosomal DNA was isolated from irradiated cells using the commercially available DNAzol reagent (Invitrogen, Carlsbad, CA), a proprietary chaotrope-detergent which lyses cells, hydrolyzes RNA, and dissociates non-covalent DNA-protein interactions for the isolation of genomic DNA. The recovered genomic DNA was washed with SDS, urea, and sodium chloride to further strip away any non-covalently bound proteins leaving only DNA containing DPCs. After enzymatic digestion of the DNA, the proteins were subjected to SDS-PAGE separation followed by in-gel tryptic digestion. MS/MS analysis of the tryptic peptides was coupled with database searching to identify 29 proteins which formed DPCs following exposure to ionizing radiation. These proteins encompassed several cellular functions, including cell structure and architecture, cell cycle, chromatin regulation, cellular homeostasis, transcriptional regulation, and RNA splicing.
Qiu and Wang recently employed DNA-protein cross-linking coupled with mass spectrometry-based proteomics to study non-covalent DNA-protein interactions. Human acute promyelocytic leukemia HL-60 cells were treated with FA to fix non-covalently bound proteins to the chromosomal DNA, and the DNA containing DPCs was isolated using DNAzol following a protocol modified from Barker et al. FA was used as a cross-linking agent because FA-induced cross-links are easily reversible by thermal hydrolysis, enabling rapid release of the cross-linked proteins from the DNA. The released proteins were separated by SDS-PAGE and digested with trypsin. The resulting peptides were subjected to mass spectrometric analysis and database searching to identify the proteins. Using this methodology, they were able to identify 780 DNA binding proteins. Interestingly, classification of the proteins based on their Gene Ontology (GO) annotations for molecular function revealed that only 14.2% of the identified proteins were classified as DNA binding. An additional 23.8% of proteins were annotated as nucleotide-binding, and 20.0% were considered RNA-binding. That many of the identified proteins were not specifically annotated as DNA-binding suggests that the available GO annotational information is incomplete, or that FA treatment resulted in the formation of ternary covalent DNA-protein-protein complexes.

1.3 DNA-Protein Cross-linking by O6-Alkylguanine DNA Alkyltransferase

O6-alkylguanine DNA alkyltransferase (AGT) is a DNA repair protein that protects the genome from alkylation damage by repairing promutagenic O6-alkylguanine adducts induced by environmental toxins and chemotherapeutic drugs, such as cyclophosphamide, tobacco-specific nitrosamines, and haloalkynitrosoureas. AGT is ubiquitously expressed in nearly every tissue type, with greatest activity in the liver, kidneys, and lungs. In contrast to its repair
capabilities, AGT is known to form biologically relevant DPCs which produce damaging genotoxic effects.

1.3.1 Structure and Function of AGT Protein

Human AGT is a small protein whose tertiary structure consists of two distinct domains.126 The N-terminal domain, which contains residues 1–85, comprises an α/β-roll consisting of a four-stranded anti-parallel β-sheet, an α-helix, and a 3_10-helix. Within this domain resides a zinc ion which is coordinated to Cys^5, Cys^24, His^29, and His^85 and functions to stabilize the interface between the N- and C-terminal domains. The C-terminal domain, which comprises residues 86–207, is made up of a short, two-stranded parallel β-sheet and five α-helices. Within the C-terminal domain is a helix-turn-helix motif which functions as the protein’s DNA-binding domain, as well as the protein’s active site, which contains the solvent-accessible catalytic Cys^145 residue.126 Unlike most DNA-binding proteins, AGT binds to the minor groove of DNA, rather than the major groove.127 The binding of AGT protein to DNA results in a conformational change of the DNA, which likely facilitates the protein’s repair activity.

Structural and biophysical studies have shown that AGT’s repair capabilities result from the protein’s ability to bind to the DNA helix and scan along the backbone, flipping nucleotides out of the base stack and into the protein’s active site in search of O^6-alkylguanine lesions.127 The conformational change induced by the nucleotide flipping places these lesions in close proximity to the side chain of Cys^145, which is activated to a thiolate anion through a hydrogen bonding network consisting of Glu^172, His^146, H_2O, and Cys^145. During repair reactions, the O^6-alkylguanine substituent is subjected to nucleophilic attack by Cys^145, resulting in the transfer of the O^6-alkyl lesion from the guanine to AGT, thereby restoring intact guanine within the DNA duplex (Scheme 1.3). The irreversible alkylation of Cys^145 destabilizes the protein’s tertiary
Scheme 1.3  Mechanism for removal of O\textsuperscript{6}-alkylguanine lesions by AGT\textsuperscript{126}
structure, serving as a signal for ubiquitination and subsequent proteosomal degradation of AGT.128

1.3.2 DNA-Protein Cross-linking by AGT

Although AGT functions primarily to protect cells from DNA damage induced by simple alkylating agents, numerous studies have suggested that the expression of human AGT protein in bacteria enhances the toxicity and mutagenicity of several alkylating agents, including N-methyl-N’-nitro-N-nitrosoguanidine, 1,2-dibromoethane, and dibromomethane.129-134 Experimental evidence suggests that the observed enhancement of genotoxic effects is due to the formation of covalent AGT-DNA cross-links involving Cys135,136. Valadez et. al. characterized the AGT-mediated activation of several dihaloalkanes to mutagenic conjugates, and found that reactivity followed halide order (I > Br > Cl), with at least one bromine or iodine needed for AGT-DNA cross-linking and AGT-mediated enhancement of genotoxicity.113 Furthermore, genotoxicity and cross-linking were maximal when the methylene chain was between one and three carbons in length. The group also investigated other bis-electrophiles, including haloforms and DEB. Importantly, they observed that DEB exhibited strong AGT-dependent genotoxicity, suggesting that DEB can also induce AGT-DNA cross-links which may contribute to its cytotoxic and mutagenic effects.113

Our laboratory employed mass spectrometry to characterize AGT-DNA cross-linking by DEB (Scheme 1.4, panel A).94 HPLC-ESI+ -MS/MS analysis of tryptic peptides resulting from digestion of wild type protein and the C145A hAGT mutant following treatment with N7-(2’-hydroxy-3’,4’-epoxybut-1’-yl)-deoxyguanosine (dG monoepoxide) revealed the ability of the protein to form butanediol-dG cross-links at the catalytic alkyl acceptor site Cys145 and the nearby active site residue Cys150.94 Similar results were observed when hAGT was cross-linked to synthetic DNA oligodeoxynucleotides in the presence of DEB.94 Mass spectrometric analysis of
Scheme 1.4 AGT-DNA cross-linking by DEB (A) and antitumor nitrogen mustards (B)
amino acid mixtures resulting from total proteolytic digestion of alkylated proteins established the
chemical structure of the cross-link as 1-(S-cysteiny1)-4-(guan-7-yl)-2,3-butanediol (Cys-N7G-
BD) (Scheme 1.4, panel A). Further studies established the ability of AGT to become cross-linked
to DNA in the presence of another class of bis-electrophiles, the antitumor nitrogen mustards
(Scheme 1.4, panel B). These cross-links also involved the two cysteine residues, Cys145 and
Cys150, which mediated AGT-DNA cross-linking in the presence of DEB. The cross-link
structures formed by two representative nitrogen mustards, mechlorethamine and chlorambucil,
were established as N-[2-[S-cysteinyl]ethyl]-N-[2-(guan-7-yl)ethyl]methylamine (Cys-N7G-
EMA) and N-(2-[S-cysteinyl]ethyl)-N-(2-[guan-7-yl]ethyl)-p-aminophenylbuycric acid (Cys-N7G-
PBA), respectively (Scheme 1.4, panel B). More recently, our laboratory has demonstrated the
ability of AGT to form DPCs in vitro in the presence of other cellular proteins as a result of
exposure to mechlorethamine and DEB (See Chapter II).

1.4 Cross-linking Agents Examined in this Work

Previously, our laboratory has employed mass spectrometry to study DNA-protein cross-
linking in vitro by two types of bifunctional alkylating agents which are known to react with
cellular biomolecules: the bis-epoxide DEB and the antitumor nitrogen mustards
mechlorethamine and chlorambucil. In order to gain further insight into the biological effects of
DNA-protein cross-linking, we now extend these studies to investigate DEB- and
mechlorethamine-mediated DNA-protein cross-linking in nuclear protein extracts and cultured
mammalian cells.
1.4.1 1,2,3,4-Diepoxybutane

DEB is a prominent bis-electrophile hypothesized to be the ultimate carcinogenic metabolite of 1,3-butadiene (BD). BD is an important industrial chemical employed in the production of plastic and synthetic rubbers, including styrene, butadiene, and nitrile rubbers and styrene-butadiene latex. BD is also a byproduct of commercial ethylene production. United States industries produce billions of pounds of BD annually, leading to occupational exposure affecting thousands of American workers. BD is also present in cigarette smoke and automobile exhaust, resulting in widespread human exposure. The carcinogenic effects of BD are well-established in rodents, where inhalation exposure induces tumorigenesis in multiple target sites, including the lung, liver, and heart. Interestingly, there are large interspecies differences in the biological response to BD which may be attributed to differences in metabolic activity. Several studies have demonstrated that BD can exert damaging genotoxic effects. Furthermore, epidemiological studies have suggested a correlation between chronic occupational BD exposure and hematopoietic cancers. Potential biomarkers for BD exposure include urinary metabolites and hemoglobin adducts.

Metabolic activation of BD proceeds via a cytochrome P450 monooxygenase-catalyzed pathway to form several metabolites, including DEB. Initially, DB is oxidized by CYP2E1 and CYP2A6 to form 3,4-epoxy-1-butene (EB), which can then be further oxidized by CYP2E1 or CYP3A4 to form DEB (Scheme 1.5). Alternatively, EB can be hydrolyzed by epoxide hydrolase (EH) to form 1-butene-3,4-diol. EH-mediated hydrolysis to form 3,4-epoxy-1,2-butanediol (EBD) and glutathione (GSH) conjugation serve as detoxification pathways for DEB. There are remarkable species differences in BD metabolism which give rise to pronounced differences in metabolic products and susceptibility to BD-induced cancer. For example, the rate of cytochrome P450-catalyzed oxidation of BD to EB and EB to DEB is greater in mice compared to rats, whereas hydrolysis of EB and DEB is faster in rats and humans.
**Scheme 1.5** Metabolic activation of BD and formation of DNA adducts by DEB
compared to mice. The net result is that mice form greater levels of DEB and are thus more susceptible to BD-induced tumorigenesis than are rats or humans.

The high level of genotoxicity exerted by DEB relative to the more abundant monoepoxide-containing BD metabolites has been attributed to its ability to form covalent cross-links involving cellular biomolecules such as DNA and proteins (Scheme 1.5). Initial alkylation of DNA by DEB produces 2-hydroxy-3,4-epoxybut-1-yl (HEB) adducts on the N7-position of guanine and the N1, N3, N6, and N7 positions of adenine. If not hydrolyzed to form trihydroxybutyl (THB) monoadducts, the second epoxide can subsequently react with another nucleophilic site on the DNA to form a variety of DNA-DNA cross-links and exocyclic adducts, many of which have been extensively studied in our laboratory. Alternatively, the HEB lesion can react with nucleophilic sites of amino acid side chains within nearby proteins, giving rise to covalent DPCs. To further understand DEB-mediated DNA-protein cross-linking, the identities of nuclear proteins which form DPCs in vitro in the presence of DEB were established by mass spectrometry-based proteomics (see Chapter II).

1.4.2 Antitumor Nitrogen Mustards

Nitrogen mustard compounds comprise a class of bifunctional alkylating agents which were originally developed during World War I as part of an initiative to design novel agents to be employed in chemical warfare. These compounds have since had a long history of use as therapeutic agents in chemotherapy for the treatment of a variety of cancers, including leukemia, lymphomas, and carcinomas. Nitrogen mustards are a diverse class of bis-(2-chloroethyl)amines which differ in the identity of the alkyl group occupying the third valence of the nitrogen atom (Scheme 1.6). Mechlorethamine, the original nitrogen mustard, has a methyl group on the central nitrogen. Mechlorethamine’s high reactivity and propensity for off-target toxicity led to the development of several analogs which are less reactive and more selective.
Scheme 1.6 Formation of DNA and protein adducts by antitumor nitrogen mustards
towards cancerous cells.\textsuperscript{164} For example, the drug chlorambucil contains a phenylbutyric acid moiety in place of the methyl group. The presence of this substituted benzene ring decreases the reactivity and toxicity of the nitrogen mustard.

The antitumor effects of nitrogen mustards have been attributed to their ability to form interstrand DNA-DNA cross-links which inhibit DNA replication. The $N$-(2-chloroethyl) groups can spontaneously cyclize, giving rise to highly reactive aziridinium ions capable of reacting with nucleophilic sites within DNA, such as the N7 position of guanine (\textit{Scheme 1.6}).\textsuperscript{165} Formation of DNA-DNA cross-links occurs via a two step process in which one aziridinium ion forms and alkylates one guanine base, followed by repetition of the cycle to form the guanine-guanine cross-link.\textsuperscript{166} Other reactive sites within DNA include the N1, N3, N$^6$, and N7 positions of adenine, rendering it possible to form a variety of interstrand and intrastrand cross-links involving guanine and adenine nucleobases.\textsuperscript{167} Interestingly, only one cross-link is formed per approximately 20 monoalkylation events.\textsuperscript{168} In an effort to improve cross-linking efficiency, Prakash \textit{et. al.} developed a series of aniline mustards tethered to intercalating chromophores intended to increase the affinity of the mustard compound for the DNA substrate.\textsuperscript{168} While the presence of the intercalating moiety increased the frequency of monoadduct formation, there was no improvement in the cross-linking efficiency.

While the role of DNA-DNA cross-linking in the cytotoxicity of nitrogen mustards is well-established,\textsuperscript{166,169-171} DNA-protein cross-linking by these agents is not as well understood. Previously, Kohn and colleagues detected nitrogen mustard-mediated DPCs by alkaline elution.\textsuperscript{172} More recently, our laboratory employed mass spectrometry to characterize AGT-DNA cross-linking by two nitrogen mustards, mechlorethamine and chlorambucil,\textsuperscript{95} and used mass spectrometry-based proteomics to identify other nuclear proteins which formed DPCs \textit{in vitro} in the presence of mechlorethamine.\textsuperscript{117} In the present work, mass spectrometry-based proteomics
was employed to characterize mechlorethamine-mediated DNA-protein cross-linking in mammalian cell cultures (see Chapter III).

### 1.5 Goals of This Work

When nuclear proteins become covalently bound to chromosomal DNA following exposure to damaging cross-linking agents, the resulting lesions are likely to interfere with normal DNA metabolic functions, possibly leading to cell death or a variety of mutagenic outcomes. However, the protein composition and the biological consequences of DPCs are not well understood. The goal of this research was to more fully characterize the formation and biological effects of damaging DPC lesions produced by two important bis-electrophilic cross-linking agents: the antitumor nitrogen mustard mechlorethamine, and the genotoxic bis-epoxide DEB. Specifically, we sought to (1) identify proteins which become cross-linked to DNA *in vitro* in the presence of DEB (see Chapter II); (2) characterize mechlorethamine-mediated DNA-protein cross-linking in mammalian cell cultures (see Chapter III); and (3) analyze the cytotoxic and mutagenic effects of DNA-protein cross-linking in cells in the absence of other types of DNA damage (see Chapter IV).
II. DNA-PROTEIN CROSS-LINKING BY 1,2,3,4-DIEPOXYBUTANE

2.1 Introduction

DPCs are bulky, helix-distorting lesions that are hypothesized to block the binding and progression of protein complexes, interfering with DNA replication, transcription, DNA repair, recombination, and chromatin remodeling. DPCs can form endogenously as a result of oxidative stress and lipid peroxidation, or can be induced by exposure to ionizing radiation, metals, or common chemotherapeutic agents such as nitrogen mustards, platinum drugs, and alkyl nitrosoureas. Although their biological relevance is poorly understood, certain types of DPCs persist through several cycles of DNA replication, potentially leading to cytotoxic and mutagenic outcomes such as sister chromatid exchanges and large deletions.

One prominent bis-electrophile capable of inducing DPCs is DEB, the proposed ultimate carcinogenic metabolite of BD. BD is a known animal and human carcinogen present in automobile exhaust and in cigarette smoke. The adverse biological effects of DEB have been attributed to its ability to cross-link cellular biomolecules. Initial alkylation of adenine and guanine bases in DNA by DEB produces HEB lesions, which contain an inherently reactive oxirane group that can alkylate neighboring nucleobases within the DNA duplex to form DNA-DNA cross-links. Alternatively, the 3,4-epoxy ring can be subject to nucleophilic attack by amino acid side chains of neighboring proteins, giving rise to DPCs (Scheme 2.1).

DNA-protein cross-linking by DEB was first observed by Jelitto et al., who employed alkaline elution methodology to detect DPC formation in liver tissue of B6C3F1 mice following exposure to BD. This observation was subsequently confirmed by other groups that utilized similar biophysical methods for detecting DEB-induced DPCs. Although these authors have
Scheme 2.1  Formation of DPCs by DEB

1,3-Butadiene (BD) → 1,2-Epoxy-3-butene (EB) → 1,2,3,4-Diepoxybutane (DEB)

- CYP 2E1 & CYP 2A6
- CYP 2E1 & CYP 3A4

HEB Monoadducts

Spontaneous or Thermal Depurination

PROTEIN
established the ability of DEB to covalently cross-link proteins to DNA, they did not identify specific proteins which participate in DPC formation, nor did they provide information regarding the chemical structure of the resulting amino acid-nucleobase conjugates.

In recent years, mass spectrometric methods have become an increasingly valuable tool in the study of DNA-protein interactions. Recent \textit{in vitro} studies with purified recombinant proteins have identified three proteins that can form covalent cross-links to DNA in the presence of DEB: AGT, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and histone H3. Our laboratory has described DEB-mediated cross-linking between the DNA repair protein AGT and DNA. Mass spectrometric analysis of tryptic peptides identified two cross-linking sites within the AGT protein: the catalytic alkyl acceptor site (Cys\textsuperscript{145}), and a neighboring active site residue (Cys\textsuperscript{150}). Loecken and colleagues have reported that DEB also forms DPCs involving GAPDH and histones H2b and H3. Like AGT, DPC formation by GAPDH involves a cysteine residue (Cys\textsuperscript{246}). Alkylation of Cys\textsuperscript{246} inhibits \textit{in vitro} GAPDH activity. Similarly, DEB is capable of cross-linking the Cys\textsuperscript{111} residue of histone H3 to DNA. In \textit{vitro} DPC formation by all three proteins required treatment with relatively high concentrations of DEB (20 mM for GAPDH and histone proteins and 15 mM for AGT). Interestingly, the over-expression of human AGT in bacteria enhanced the cytotoxic and mutagenic effects of DEB, presumably through the formation of toxic DPC lesions. In contrast, no enhanced mutagenesis was observed in cells over-expressing GAPDH or histone H2b. These results suggest that the identity of the cross-linked protein influences the biological effects of DPCs. However, the identities of other nuclear proteins that participate in cross-linking to DNA in the presence of DEB, as well as the abundance with which DNA-protein lesions are produced in mammalian cells following DPC exposure, have not been established, limiting our understanding of the role of DNA-protein cross-linking in the genotoxicity and cytotoxicity of DEB.
The purpose of the current study was to identify human proteins participating in DEB-mediated DPC formation. We have recently developed an affinity capture technique which can be coupled with mass spectrometry-based proteomics and immunological detection to identify nuclear extract proteins involved in DPC formation in the presence of bis-electrophiles. In the present study, the new methodology was used to investigate DEB-mediated DNA-protein cross-linking (Scheme 2.2). The identities of nuclear proteins which become covalently cross-linked to DNA in the presence of DEB were established by mass spectrometry-based proteomics, and immunological approaches were employed to estimate the cross-linking efficiency for specific protein targets. These results are significant because the identification of specific nuclear proteins that become cross-linked to DNA in the presence of DEB will allow a better understanding of their role in DEB-mediated cytotoxicity.

2.2 Materials and Methods

Safety Statement – DEB is a known human carcinogen and should be handled with caution in a well-ventilated fume hood with appropriate personal protective equipment.

Chemicals and Reagents – DEB, phenylmethanesulfonyl fluoride (PMSF), pepstatin, leupeptin, aprotinin, dithiothreitol (DTT), and iodoacetamide were purchased from Sigma-Aldrich (St. Louis, MO). Mass spectrometry-grade Trypsin Gold was purchased from Promega (Madison, WI). Proteinase K was obtained from Worthington Biochemical Corp. (Lakewood, NJ). Primary polyclonal antibodies to human actin, Ref-1, GAPDH, and poly(ADP-ribose) polymerase (PARP) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). The primary polyclonal antibody to the ATP-dependent DNA helicase subunit 2 (Ku) was purchased from Lab Vision/NeoMarkers (Fremont, CA). The primary monoclonal antibody to AGT was purchased
Scheme 2.2 Experimental scheme for biotin capture enrichment of DPCs from nuclear protein extracts incubated with DEB in the presence of double-stranded DNA.
from Millipore (Temecula, CA). Alkaline phosphatase-conjugated anti-mouse and anti-rabbit IgG secondary antibodies were obtained from Sigma-Aldrich (St. Louis, MO). Synthetic DNA oligodeoxynucleotides were prepared at the University of Minnesota’s Biomedical Genomics Center (Minneapolis, MN). Synthetic Cys-N7G-BD was prepared as described previously.\(^94\)

**Cell Culture** – Human cervical carcinoma (HeLa) cells were generously provided by Dr. Jonathan Marchant (University of Minnesota). The cells were maintained as exponentially growing monolayer cultures in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 9% fetal bovine serum (FBS), in a humidified incubator at 37°C with 5% CO₂.

**Preparation of Nuclear Extracts** – Nuclear protein extracts were prepared from HeLa cells as described in the literature.\(^{178}\) Briefly, ~10⁸ cells were harvested, washed three times with ice cold phosphate-buffered saline (PBS), and suspended in hypotonic buffer (10 mM Tris-HCl – pH 7.4/10 mM MgCl₂/10 mM KCl/1 mM DTT) containing 1 mM PMSF. After incubating for 5 min on ice, cells were broken by 20 strokes in a Dounce homogenizer and centrifuged at 2,000 g for 10 min. The sedimented nuclei were re-suspended in hypotonic buffer (see above) containing 350 mM NaCl and a protease inhibitor cocktail (1 μg/mL pepstatin; 0.5 μg/mL leupeptin; 0.75 μg/mL aprotinin; 1 mM PMSF), and incubated on ice for 1 h. The resulting nuclear lysate was centrifuged at 160,000g at 4 °C for 30 min, and the nuclear proteins were isolated in a clear supernatant. Extracts were dialyzed for 2 h at 4°C against 10 mM Tris-HCl – pH 7.4/10 mM KCl/10 mM MgCl₂ buffer using Slide-A-Lyzer dialysis cassettes with a 3.5 kDa molecular weight cut-off (Pierce Biotechnology, Rockford, IL). Protein concentrations were determined using the Bio-Rad Protein Assay (Bio-Rad, Hercules, CA).
DNA-Protein Cross-linking and Biotin Capture – 5'-Biotinylated double-stranded oligodeoxynucleotides (5'-GGA GCT GGT GGC GTA GGC-3' (+) strand, 3.12 nmol) were combined with nuclear protein extracts from HeLa cells (500 μg total protein) in the absence or in the presence of DEB (5-500 mM, 1 mL total volume) and incubated at 37°C for 3 h to induce cross-linking. Biotinylated DNA, along with covalently cross-linked proteins, was captured by incubation with Streptavidin Sepharose High Performance beads (750 μL slurry, GE Healthcare, Piscataway, NJ) overnight at 4°C, with rotation. To remove any non-covalently bound proteins, the beads were washed twice with 1% SDS (1 h followed by 30 min, with rotation, room temperature), twice with each 4 M urea and 1 M NaCl (both 30 min followed by 15 min, with rotation, room temperature), and twice with phosphate-buffered saline (no incubation). Following each washing step, the beads were centrifuged at 2,000g for 1 min, and the supernatant was discarded. DNA containing covalently bound proteins was released from the beads by adding 110 μL NuPage 4X LDS Sample Buffer (Invitrogen, Carlsbad, CA) and heating to 90°C for 15 min. Due to the thermal instability of DEB-induced cross-links, which form specifically at the N7 of guanine,94 these conditions enabled a quantitative release of the proteins from DNA in the form of protein-guanine conjugates. The proteins were subsequently analyzed by HPLC-ESI+ MS/MS and western blotting as described below.

Protein Identification by Mass Spectrometry. For proteomic analyses, the cross-linked proteins were separated using 12% Tris-HCl Ready Gels (Bio-Rad, Hercules, CA) and stained with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA). Gel lanes were cut into slices comprising the entire molecular weight range, washed with 100 mM ammonium bicarbonate, and subjected to reduction with dithiothreitol and iodoacetamide treatment as previously described.179 Gel pieces were dehydrated with acetonitrile, dried under vacuum, and reconstituted in 25 mM ammonium bicarbonate containing 10 μg of mass spectrometry-grade trypsin. The samples were digested
overnight at 37°C. Tryptic peptides were extracted with 1% aqueous formic acid/60% acetonitrile, evaporated to dryness, and re-suspended in 0.1% aqueous formic acid for MS analysis.

Tryptic peptides were subjected to HPLC-ESI+-MS/MS analysis with a ThermoFinnigan LTQ ion trap mass spectrometer equipped with a Thermo MicroAS autosampler and a Thermo Surveyor HPLC pump, a nanospray source, and an Xcalibur 1.4 instrument control. Peptides were resolved on a 100 μm x 11 cm fused silica capillary column (Polymicro Technologies, LLC, Phoenix, CA) packed with 5 μm, 300 Å Jupiter C18 packing (Phenomenex, Torrence, CA). The column was eluted at 0.6 μL/min with a gradient of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The solvent composition was initially held at 2% B for 15 min, followed by a linear increase to 25% B over 25 min, and further to 90% B in 15 min. Liquid chromatography was carried out at an ambient temperature. centroided MS-MS scans were acquired using an isolation width of 2 m/z, an activation time of 30 ms, an activation Q of 0.250 and 30% normal collision energy using 1 microscan with a max ion time of 100 ms for each MS/MS scan. The mass spectrometer was tuned prior to analysis using the synthetic peptide TpepK (AVAGKAGAR), so that some parameters may have varied slightly from experiment to experiment. Typically, the tune parameters were as follows: spray voltage of 2 kV, a capillary temperature of 150°C, a capillary voltage of 50 V, and tube lens of 120 V. The MS/MS spectra of the peptides were collected using data-dependent scanning, in which one full MS spectrum was followed by four MS/MS spectra. MS/MS spectra were recorded using dynamic exclusion of previously analyzed precursors for 60 s.

The "ScanSifter" algorithm v0.1, an in-house developed software, read tandem mass spectra stored as centroided peak lists from Thermo RAW files and transcoded them to DTA files. Spectra that contained fewer than 6 peaks or that had less than 20 measured total ion current (TIC) did not result in DTAs. If 90% of the intensity of a tandem mass spectrum appeared at a
lower $m/z$ than the precursor ion, a single precursor charge was assumed; otherwise, the spectrum was processed under both double and triple precursor charge assumptions. Proteins were identified using the SEQUEST v.27 algorithm\textsuperscript{180,181} on a high speed, multiprocessor Linux cluster in the Advanced Computing Center for Research & Education at Vanderbilt University using the human subset of the IPI human protein database, version 331, created 7/20/07. To estimate false discovery rates (FDRs), each sequence of the database was reversed and concatenated to the database, for a total of 135,168 entries for the human database. The database search encompassed tryptic peptides with a maximum of 5 missed cleavage sites for enzyme search and with a maximum number of 10 internal cleavage sites. Cysteines were expected to undergo carboxamidomethylation (+57 Da), and methionine oxidation (+16 Da) was permitted. DEB-induced alkylation at cysteine, arginine, lysine, histidine, and the N-terminus (hydrolyzed monoadduct: +104 Da; cross-link to guanine: +237 Da) were specified as dynamic modifications to identify spectra of adducted peptides. Precursor ions were required to fall within 1.25 $m/z$ of the position expected from their average masses, and fragment ions were required to fall within 0.5 $m/z$ of their monoisotopic positions. The database searches produced raw identifications in SQT format.\textsuperscript{182}

Peptide identification, filtering, and protein assembly were done with IDPicker software 42, which filtered raw peptide identifications to a target FDR of 5%. The peptide filtering employed reversed sequence database match information to determine thresholds that yielded an estimated 5% FDR for the identifications of each charge state by the formula $FDR = (2R)/(R+F)$, where $R$ is the number of passing reversed peptide identifications and $F$ is the number of passing forward (normal orientation) peptide identifications. A second round of filtering removed proteins supported by less than two distinct peptide identifications in the analyses. Indistinguishable proteins were recognized and grouped. Parsimony rules were applied to generate a minimal list of proteins that explained all of the peptides that passed our entry criteria.\textsuperscript{183} Finally, to be
considered a positive identification, all proteins were required to have a minimum of three unique peptide spectra and 10 total spectral counts. Any protein identified in the DEB-treated samples that displayed comparable MS spectral counts to the untreated controls was disregarded.

**Western Blot Analysis of Identified Proteins** – Nuclear protein extracts from HeLa cells were exposed to 0-25 mM DEB in the presence of 5'-biotinylated oligonucleotide duplexes as described above, and the resulting DNA-protein cross-links were captured on streptavidin beads. The cross-linked proteins were released from the DNA backbone by thermal hydrolysis, separated by SDS-PAGE, and transferred to Trans-Blot nitrocellulose membranes (Bio-Rad, Hercules, CA). Membranes were immediately blocked in Tris-buffered saline (TBS) containing 5% (w/v) BSA. Following a 1–2 h incubation with the primary antibody at room temperature, the blots were washed three times with TBS and incubated overnight at 4°C with the corresponding alkaline phosphatase-conjugated secondary antibody. Following three additional washes with TBS, the blots were developed using SIGMA Fast BCIP/NBT (Sigma, St. Louis, MO). The developed blots were scanned as image files, and the optical densities of the bands were then quantified using ImageJ software available free of charge from the NIH website (www.nebi.nlm.nih.gov). The extent of DNA-protein cross-linking was estimated by comparing the band intensities in cross-linked samples to that of a known amount of nuclear proteins extract loaded on the gel as a positive control.

**HPLC-ESI+-MS/MS Analysis of Cys-N7G-BD Conjugates** – Cross-linking reactions and biotin capture were carried out as described above, with the following exceptions: 1.3 mg total protein was incubated with 8 nmol of double-stranded 5'-bionylated oligodeoxynucleotides (5'-GGA GCT GGT CGT GTA GGC-3' (+) strand) in the presence of 50 mM DEB, and the DPCs were eluted from the streptavidin beads by three incubations with 1 mL 70% acetonitrile/5% aqueous
formic acid (overnight at 4°C, 30 min at room temperature, and 30 min at 90°C). Following each incubation, the beads were centrifuged for 1 min at 2,000g, and the supernatants containing DPCs were decanted and pooled. Cross-linked proteins were released from the DNA backbone by neutral thermal hydrolysis (1 h at 70°C) to produce protein-guanine conjugates. Samples were dried under vacuum, reconstituted in 25 mM ammonium bicarbonate, and digested to peptides with trypsin (20 μg, 0.86 nmol, 37°C overnight). To achieve complete hydrolysis of the peptides to amino acids, tryptic peptides were dried under vacuum, reconstituted in water, and digested with proteinase K (20 μg, 0.70 nmol) for 48 h at room temperature. To enrich for Cys-N7G-BD, the digest mixtures were purified by solid phase extraction. ExtractClean™ Carbo cartridges (150 mg, 4mL, Grace Davison Discovery Sciences, Deerfield, IL) were equilibrated with acetonitrile (3 mL), methanol (3 mL), and water (2 x 3 mL) prior to loading samples, which were adjusted to pH 9 by the addition of 1M ammonium hydroxide (1 mL total load volume). Samples were washed with water (3 mL), 50% methanol (3 mL), and 100% methanol (3 mL) prior to elution with 2:1 acetonitrile:water containing 1% formic acid (3 mL). The eluates were dried under vacuum and re-suspended in 15 mM ammonium acetate, pH 5.0 (25 μL) prior to HPLC-ESI^+^-MS/MS analysis (injection volume, 8 μL).

An Agilent 1100 capillary HPLC system interfaced to a Thermo-Finnigan TSQ Quantum Discovery mass spectrometer was utilized for detection of Cys-N7G-BD conjugates. Chromatographic separation was accomplished with a Phenomenex Synergi C18 column (250 mm x 0.5 mm, 4 μm) eluted with 15 mM ammonium acetate, pH 5.0 (A) and 3:1 methanol:acetonitrile (B) at a flow rate of 10 μL/min. The gradient program began at 2% B, followed by a linear increase to 8% B in 10 min, further to 11% B in 17 min, and finally to 30% B in 2 min. Using this gradient, Cys-N7G-BD eluted at ~19 min. ESI was achieved at a spray voltage of 3.2 kV and a capillary temperature of 250°C. CID was performed with Ar as a collision gas (1.0 mTorr) at a collision energy of 25V. The MS parameters were optimized for
maximum response during infusion of a standard solution of Cys-N7G-BD. HPLC-ESI+-MS/MS analyses were performed in the selected reaction monitoring mode using the transition corresponding to a major fragment ion observed upon CID fragmentation of Cys-N7G-BD in a triple quadrupole mass spectrometer ($m/z$ 359.7 [M + H]$^+$ $\rightarrow$ 87.9 [M + H – Gua – C$_4$H$_8$O$_2$S]$^+$).

2.3 Results

2.3.1 Strategy for Affinity Purification of Proteins Cross-linked to DNA by DEB

Our laboratory has recently developed an affinity capture-based methodology to investigate the *in vitro* formation of DPCs between nuclear extract proteins and biotinylated oligodeoxynucleotides.[117] In this approach (*Scheme 2.2*), nuclear extract proteins are incubated with synthetic DNA duplexes containing a 5'-biotin tag in the absence or the presence of a *bis*-electrophile, such as DEB. Following affinity capture of biotinylated DNA and any covalently cross-linked proteins on streptavidin beads, the beads are subjected to stringent washing to remove any non-covalently bound proteins. Finally, the biotinylated DNA and the cross-linked proteins are eluted by heating the beads in SDS-containing gel loading buffer. Because the N7-alkylguanine adducts induced by DEB destabilize the glycosidic bond of DNA and are thus thermally labile, these elution conditions also release the cross-linked proteins from the DNA backbone, yielding protein-butanediol-guanine conjugates (*Scheme 2.2*). Following elution, the cross-linked proteins are separated by SDS-PAGE and are subjected to in-gel tryptic digestion followed by mass-spectrometry based proteomic analysis to identify the cross-linked proteins. Alternatively, the protein mixtures are analyzed by western blotting using commercial antibodies to target specific candidate proteins.
2.3.2 Concentration-Dependent Formation of DPCs in Nuclear Protein Extracts Following DEB Treatment

Previous studies have demonstrated the ability of AGT, GAPDH, and histone proteins to become cross-linked to DNA \textit{in vitro} in the presence of DEB.\textsuperscript{94,114,115} However, these studies were conducted using purified recombinant proteins in the absence of other cellular proteins which could also form DPCs or interfere with DNA-protein cross-linking. To determine whether DEB-induced DPC formation occurs in a more complex biological system, we employed the affinity capture approach developed in our laboratory\textsuperscript{117} (Scheme 2.2) to capture nuclear extract proteins from human cervical carcinoma (HeLa) cells which became cross-linked to biotinylated DNA duplexes in the presence of DEB. Following SDS-PAGE separation of the cross-linked proteins (Figure 2.1, panel A), densitometric analysis of the stained protein bands over the 25 – 250 kDa molecular weight range showed a concentration-dependent increase in protein signal in samples treated with 5, 10, 25, 50, and 100 mM DEB, indicating the formation of DPCs (panel B). Minimal DPC background was present in control samples prepared in the absence of DEB, signifying that the majority of non-covalently bounds proteins are successfully removed by the washing procedures (lane 2 in panel A). A DEB concentration of 100 mM was required to achieve cross-linking of 2% of total protein to DNA. By comparison, 100-fold lower concentrations were required to obtain similar levels of cross-linking with the antitumor nitrogen mustard, mechlorethamine,\textsuperscript{117} indicating that DEB is a less effective DPC-inducing agent than mechlorethamine.

2.3.3 Mass Spectrometric Identification of Cross-linked Proteins

In order to identify human nuclear proteins which participate in DEB-induced DPC formation, nuclear protein extracts from HeLa cells were incubated in triplicate with biotinylated DNA in the absence or in the presence of DEB (500 mM). This high drug concentration was
Figure 2.1  Concentration-dependent formation of DPCs in HeLa nuclear protein extracts following exposure to DEB. (A) Nuclear protein extracts from HeLa cells and 5'-biotinylated double-stranded oligodeoxynucleotides were incubated in the presence of 0-100 mM DEB (lanes 2-7). The resulting DPCs were subjected to biotin capture enrichment, hydrolyzed to release protein-guanine conjugates, and resolved by 12% SDS-PAGE. Gels were stained with SimplyBlue SafeStain to visualize the cross-linked proteins. (B) Densitometric analysis of protein bands in the 25 – 250 kDa molecular weight range to estimate the extent of total protein cross-linking to DNA in the presence of DEB. Band intensity was compared to staining of a known amount of nuclear protein extract analyzed as a control to estimate the cross-linking efficiency.
selected in order to maximize the number of low-abundance proteins identified in the proteomic analysis, due to the low numbers of DPCs observed in preliminary experiments (Figure 2.1). Following biotin capture enrichment (Scheme 2.2), the cross-linked proteins were separated by SDS-PAGE (Figure 2.2), and the gel lanes were separated into five fractions encompassing the entire molecular weight range and excised. Proteins in these gel pieces were subjected to in-gel tryptic digestion, and the resulting tryptic peptides were extracted and analyzed by HPLC-ESI+-MS/MS analysis to identify the cross-linked proteins. Protein identifications were based on the MS/MS spectra of unmodified tryptic peptides, as shown for representative peptides in Figure 2.3. Spectral data were subjected to parsimony analysis, resulting in the identification of a total of 39 HeLa nuclear proteins which participated in cross-linking to DNA in the presence of DEB (Table 2.1).

The proteins identified in affinity captured DPCs that were induced by DEB exposure encompass a variety of cellular functions (Figure 2.4). Over half of the identified proteins are nucleic acid binding proteins known to participate in transcriptional regulation (e.g. PARP, heterogeneous nuclear ribonucleoprotein U (hnRNP-U), GAPDH, Ref-1, actin), chromatin remodeling (e.g. actin), DNA replication (e.g. DNA polymerase δ), and DNA repair (e.g. Ref-1, AGT, PARP, flap endonuclease-1 (Fen-1), Msh2). We also observed proteins involved in cell motility (e.g. actin), protein folding (e.g. T complex proteins), and biosynthesis (e.g. histidyl tRNA synthetase). Ten of the proteins identified in biotin capture fractions of DEB-induced DPCs, including actin, GAPDH, EF-1α1, PARP, and Fen-1, have been previously shown to form cross-links to DNA in the presence of the antitumor nitrogen mustard mechlorethamine (Table 2.1). Additionally, several of the identified proteins were closely related to proteins that form DPCs in the presence of mechlorethamine (e.g. different isoforms of tubulin). Altogether, approximately 25% the 39 proteins identified in the present work also formed DPCs in the presence of mechlorethamine. The remaining proteins were not
Figure 2.2  SDS-PAGE analysis of DPCs in HeLa nuclear protein extracts following exposure to DEB. Nuclear protein extracts and 5’-biotinylated double-stranded oligodeoxynucleotides were incubated in triplicate with 0 (A) or 500 mM DEB (B). The resulting DPCs were captured on streptavidin beads and resolved by 12% SDS-PAGE. Gels were stained with SimplyBlue SafeStain to visualize the cross-linked proteins. Each lane was cut into five sections, and the proteins present in each piece were subjected to in-gel tryptic digestion prior to HPLC-ESI-MS/MS analysis and proteomic identification of cross-linked proteins.
Figure 2.3  Examples of HPLC-ESI$^+$-MS/MS spectra of tryptic peptides used for protein identification of Fen-1 (A), GAPDH (B), and PARP (C) present in affinity-captured DPCs. $^1$Cysteine carboxamidomethylation (+57).
Table 2.1  Nuclear extract proteins from human cervical carcinoma (HeLa) cells that form DPCs in the presence of DEB\textsuperscript{*}

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
<th>Forms DPCs by Mechlorethamine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>O00148</td>
<td>ATP-dependent RNA helicase DDX39</td>
<td>13</td>
<td>63</td>
<td>33</td>
<td>49130</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>Q7L1Q6</td>
<td>Basic leucine zipper and W2 domain-containing protein 1, isoform 2</td>
<td>3</td>
<td>13</td>
<td>12</td>
<td>48043</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>P68104</td>
<td>EF-1α1</td>
<td>14</td>
<td>147</td>
<td>45</td>
<td>50141</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>P26641</td>
<td>Elongation factor 1-γ</td>
<td>7</td>
<td>21</td>
<td>15</td>
<td>50119</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>P13639</td>
<td>Elongation factor 2</td>
<td>12</td>
<td>24</td>
<td>17</td>
<td>95338</td>
<td>17</td>
<td>No</td>
</tr>
<tr>
<td>P60842</td>
<td>Eukaryotic initiation factor 4A-I</td>
<td>8</td>
<td>29</td>
<td>25</td>
<td>46154</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>P78347</td>
<td>General transcription factor II-I, isoform I</td>
<td>11</td>
<td>17</td>
<td>14</td>
<td>112416</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>P61978</td>
<td>hnRNP-K, isoform 1</td>
<td>7</td>
<td>18</td>
<td>23</td>
<td>50976</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>P52272</td>
<td>hnRNP-M, isoform 1</td>
<td>5</td>
<td>11</td>
<td>9</td>
<td>77516</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>Q00839</td>
<td>hnRNP-U, isoform short</td>
<td>5</td>
<td>12</td>
<td>8</td>
<td>90584</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>A8K7X6</td>
<td>poly(rC)-binding protein 2 isoform β</td>
<td>4</td>
<td>11</td>
<td>14</td>
<td>38222</td>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td>P26599</td>
<td>Polypyrimidine tract-binding protein 1, isoform 1</td>
<td>7</td>
<td>27</td>
<td>21</td>
<td>57221</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Q9Y265</td>
<td>RuvB-like 1, isoform 1</td>
<td>6</td>
<td>14</td>
<td>17</td>
<td>50228</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>Q7KZF4</td>
<td>Staphylococcal nuclease domain-containing protein 1</td>
<td>14</td>
<td>25</td>
<td>21</td>
<td>101997</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>Q68DF1</td>
<td>91 kDa protein (putative uncharacterized protein DKFZP779B0247)</td>
<td>7</td>
<td>12</td>
<td>11</td>
<td>90555</td>
<td>12</td>
<td>No</td>
</tr>
</tbody>
</table>

\textsuperscript{*}500 mM for 3 h at 37\textdegree C
Table 2.1 (continued from p. 52)

### DNA replication/DNA repair

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein Description</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
<th>Forms DPCs by Mechlorethamine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P43246</td>
<td>DNA mismatch repair protein Msh2</td>
<td>15</td>
<td>35</td>
<td>21</td>
<td>104743</td>
<td>13</td>
<td>Yes</td>
</tr>
<tr>
<td>P52701</td>
<td>DNA mismatch repair protein Msh6, isoform GTBP-N</td>
<td>13</td>
<td>24</td>
<td>10</td>
<td>152786</td>
<td>32</td>
<td>No</td>
</tr>
<tr>
<td>P28340</td>
<td>DNA polymerase δ, catalytic subunit</td>
<td>6</td>
<td>13</td>
<td>6</td>
<td>12363</td>
<td>26</td>
<td>No</td>
</tr>
<tr>
<td>P33993</td>
<td>DNA replication licensing factor MCM7, isoform 1</td>
<td>6</td>
<td>11</td>
<td>10</td>
<td>81308</td>
<td>11</td>
<td>No</td>
</tr>
<tr>
<td>P39748</td>
<td>Fen-1</td>
<td>4</td>
<td>13</td>
<td>16</td>
<td>42593</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>P09874</td>
<td>PARP</td>
<td>10</td>
<td>20</td>
<td>13</td>
<td>113084</td>
<td>14</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Architectural/structural

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein Description</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
<th>Forms DPCs by Mechlorethamine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P60709</td>
<td>Actin</td>
<td>16</td>
<td>225</td>
<td>47</td>
<td>41737</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>P68363</td>
<td>Tubulin α-1B chain</td>
<td>12</td>
<td>57</td>
<td>39</td>
<td>50152</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>P07437</td>
<td>Tubulin β chain</td>
<td>12</td>
<td>54</td>
<td>36</td>
<td>49671</td>
<td>8</td>
<td>No</td>
</tr>
</tbody>
</table>

### Chromatin regulation

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein Description</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
<th>Forms DPCs by Mechlorethamine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9NTJ3</td>
<td>Structural maintenance of chromosomes protein 4, isoform 2</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>147182</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>Q16269</td>
<td>Regulator of chromosome condensation 1, isoform a</td>
<td>9</td>
<td>27</td>
<td>32</td>
<td>4915</td>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 2.1 (continued from p. 53)

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
<th>Forms DPCs by Mechlorethamine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P11586</td>
<td>C-1-tetrahydrofolate synthase</td>
<td>12</td>
<td>30</td>
<td>15</td>
<td>101559</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>O43175</td>
<td>D-3-phosphoglycerate dehydrogenase</td>
<td>13</td>
<td>55</td>
<td>31</td>
<td>56651</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>P04406</td>
<td>GAPDH</td>
<td>5</td>
<td>20</td>
<td>21</td>
<td>36053</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>P08107</td>
<td>Heat shock 70 kDa protein 1</td>
<td>14</td>
<td>34</td>
<td>31</td>
<td>70052</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>P52789</td>
<td>Hexokinase-2</td>
<td>14</td>
<td>32</td>
<td>20</td>
<td>102380</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>P12081</td>
<td>Histidyl-tRNA synthetase</td>
<td>6</td>
<td>18</td>
<td>15</td>
<td>57411</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>P14618</td>
<td>Pyruvate kinase isozymes M1/M2, isoform M2</td>
<td>29</td>
<td>223</td>
<td>60</td>
<td>57937</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>P17987</td>
<td>T-complex protein 1 subunit α</td>
<td>8</td>
<td>16</td>
<td>17</td>
<td>60344</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>P50991</td>
<td>T-complex protein 1 subunit δ</td>
<td>5</td>
<td>17</td>
<td>11</td>
<td>57924</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>P40227</td>
<td>T-complex protein 1 subunit ζ</td>
<td>5</td>
<td>12</td>
<td>13</td>
<td>58024</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td>Q99832</td>
<td>T-complex protein 1 subunit η</td>
<td>6</td>
<td>13</td>
<td>14</td>
<td>59367</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>P22102</td>
<td>Trifunctional purine biosynthetic protein adenosine-3, isoform long</td>
<td>7</td>
<td>15</td>
<td>10</td>
<td>107767</td>
<td>22</td>
<td>No</td>
</tr>
<tr>
<td>Q00341</td>
<td>Vigilin</td>
<td>12</td>
<td>28</td>
<td>13</td>
<td>141456</td>
<td>11</td>
<td>No</td>
</tr>
</tbody>
</table>
Figure 2.4 Cellular functions of human proteins that form DPCs in the presence of DEB, as identified by affinity capture in combination with mass spectrometry-based proteomics.
common between the two lists, which is not surprising given the distinct mechanisms of cross-link formation by these two bis-electrophiles.

The proteins identified in our proteomics screen were additionally categorized according to the molecular weight region of the gel in which they were present prior to in-gel tryptic digestion (Figure 2.2). Peptides from all identified proteins were found in the correct molecular weight region as predicted by the protein’s molecular weight (Table 2.1). In addition, many proteins were also present in higher molecular weight fractions, suggesting that DEB can form ternary DNA-protein-protein complexes.

2.3.4 Identification and Quantitation of Individual Proteins that form Cross-links to DNA in the Presence of DEB by Western Blot Analysis

The identities of a subset of proteins which were detected by mass spectrometry-based proteomics were further confirmed by western blot analysis using commercial antibodies. Nuclear protein extracts derived from HeLa cells were incubated with biotinylated DNA in the presence of 0 – 25 mM DEB to induce DPC formation. Following biotin capture (Scheme 2.2), thermal hydrolysis, and SDS-PAGE separation, protein-guanine conjugates were transferred to nitrocellulose membranes and subjected to western blot analysis using commercial antibodies against actin, AGT, GAPDH, PARP, Ref-1, and Ku (Figure 2.5). These proteins were selected based on either their identification from the proteomics screen (actin, PARP, GAPDH) or their previously demonstrated ability to form DPCs in the presence of other bis-electrophiles, such as antitumor nitrogen mustards (AGT, Ref-1, Ku). Densitometric analysis of protein signals observed in the western blots of biotin capture mixtures and total nuclear extracts was used to estimate the efficiency of DPC formation for specific proteins in the presence of DEB.

Western blot analysis confirmed the identities of three proteins detected in mass spectrometric analysis of DEB-induced DPCs: actin, GAPDH, and PARP (Figure 2.5, panel A).
Figure 2.5 Western blot analysis of DEB-induced DPCs in nuclear protein extracts from HeLa cells. (A) Nuclear extract proteins were incubated with 0-25 mM DEB (lanes 1-4) in the presence of 5'-biotinylated double-stranded oligodeoxynucleotides (5'-GGA GCT GGT GGT GGC CTA-3' (+) strand). Following biotin capture enrichment and hydrolysis to release protein-guanine conjugates, the cross-linked proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes. Western blots were performed using primary antibodies against actin, AGT, GAPDH, PARP, and Ref-1. (B) Densitometric analysis of western blots to estimate the extent of protein cross-linking to DNA in the presence of DEB.
In addition, AGT and Ref-1 were found to form DPCs in the presence of DEB (panel A). These proteins were not detected in the mass spectrometry-based proteomics screen (Table 2.1), likely due to their low abundance in the nucleus. The intensity of antigen-specific staining of streptavidin-captured fractions increased as DEB concentration was increased from 5 to 25 mM (lanes 2-4 in panel A). Densitometric analysis revealed that between 2 and 12% of total protein became cross-linked to DNA in the presence of 25 mM DEB (panel B), with GAPDH forming the largest number of DPCs (~12% of total protein). Interestingly, the overall participation of all proteins in cross-linking to DNA under these conditions was less than 1% (Figure 2.1), suggesting that actin, AGT, GAPDH, PARP, and Ref-1 have a higher propensity to form DPCs in the presence of DEB as compared to other nuclear proteins. The DNA helicase protein Ku, which is present in the nuclear protein extract at high abundance, did not form detectable levels of DPCs in the presence of DEB (results not shown).

The requirement for high concentrations of DEB (5–25 mM) in order to induce DPC formation in our experiments (Figure 2.5) is consistent with previous studies investigating DEB-induced cross-linking to recombinant proteins. However, these concentrations are significantly higher than those required to obtain similar levels of cross-linking with antitumor nitrogen mustards. For example, about 3% of total Ref-1 in HeLa nuclear protein extracts became cross-linked to DNA in the presence of 10 mM DEB, but a similar cross-linking efficiency can be achieved in the presence of 0.5 mM mechlorethamine. Our recent data indicate that even lower drug concentrations are required for detection of DPCs in analogous experiments with cisplatin, with approximately 3% of total Ref-1 proteins forming DPCs in the presence of 1 μM cisplatin (Ming et al, manuscript in preparation). These results indicate a 10^4-fold difference between the efficiency of DPC formation by different bifunctional alkylating agents. These pronounced differences may be attributable to variations in chemical reactivity, distinct mechanisms for cross-link formation, and different levels of hydrolytic stability of the resulting DPC lesions.
2.3.5 Identification of Amino Acid Side Chains that Participate in DPC Formation

The identities of protein amino acid side chains which participate in DEB-induced DPC formation were investigated by considering the MS/MS spectra of peptides containing DEB-induced adducts and by analyzing amino acid-nucleobase conjugates in total enzymatic digests of affinity captured DPCs. HPLC-MS/MS analysis of tryptic digests detected a peptide representing residues 207-238 of actin which contained a DEB-mediated cross-link to guanine (Figure 2.6). In this spectrum, the \( m/z \) values of the \( b_8 \) and \( b_9 \) ions are consistent with the theoretical values for the unmodified peptide, while the \( m/z \) values of the \( b_{12} \), \( b_{14} \), \( b_{19} \), and \( b_{24} \) fragments are all increased by 237 Da, indicating that the adduct resides on the 10th (Leu), 11th (Cys), or 12th (Tyr) residue of the peptide (Figure 2.6). Previous studies of DPC formation by DEB have identified cysteine as the main target for cross-linking to DNA,\(^94,114,115\) suggesting that the cross-linking site within actin is Cys\(^{217}\). In addition, MS/MS data for two peptides originating from actin and hnRNP-U implicated the involvement of lysine residues in DPC formation by DEB (results not shown). Furthermore, two peptides from EF-1α1 contained the hydrolyzed THB monoadducts at cysteine and lysine residues. These results suggest that cysteine and lysine side chains of proteins can participate in DEB-mediated DPC formation.

To determine the exact chemical structure of the DEB-induced amino acid-nucleobase conjugates, DNA and cross-linked proteins eluted from streptavidin beads were subjected to thermal hydrolysis to release protein-nucleobase conjugates, followed by digestion with trypsin and proteinase K to achieve complete proteolytic digestion of proteins to amino acids (Scheme 2.2). The resulting amino acid-nucleobase conjugates were enriched by solid phase extraction and analyzed by capillary HPLC-ESI\(^+\)-MS/MS. Our previous studies conducted with recombinant AGT protein indicated that DEB-induced cross-linking takes place primarily between the cysteine sulfhydryl side chain within proteins and the N7-position to form Cys-N7G-BD conjugates.\(^94\)
Figure 2.6  HPLC-ESI^+-MS/MS analysis of a tryptic peptide from actin containing DEB-induced cross-link to guanine. The peptide was present in tryptic digests of biotin captured DPCs subjected to thermal hydrolysis to release N7-alkylated guanines from the DNA backbone. *Cysteine containing cross-link to guanine (+237 a.m.u.); ^Methionine oxidation (+16).
Our SRM method for Cys-N7G-BD was based on a characteristic mass transition ($m/z$ 359.1 $\rightarrow$ 87.9) corresponding to C-S bond cleavage within the cysteine moiety of Cys-N7G-BD (Figure 2.7, panel A). Cys-N7G-BD was detected only in the total digests of DEB-treated nuclear extracts (panel C), but not in the untreated control (panel B). The presence of the Cys-N7G-BD conjugate in treated samples indicates that DEB forms covalent cross-links between the N7 position of guanine in duplex DNA and the side-chain sulfhydryls of cysteine residues in proteins. These results are consistent with our MS/MS data for alkylated peptides, which reveal DEB-mediated cross-links to guanine involving cysteine residues (Figure 2.6). Consistent with this interpretation, all of the identified proteins contain multiple cysteine residues (Table 2.1).

2.4 Discussion

We have developed a comprehensive approach for identification of the proteins that participate in DNA-protein cross-linking in the presence of bis-electrophiles in cell-free extracts (Scheme 2.2). This approach enabled us to identify 39 HeLa nuclear proteins which became cross-linked to DNA following exposure to DEB. Among these, approximately 25% of proteins are also targeted for cross-linking to DNA in the presence of nitrogen mustards, while the remaining proteins appear unique to DEB-mediated cross-linking.

The ability of specific proteins to form DPCs is likely to be influenced by a number of factors, including protein abundance, structure, and cellular and nuclear localization. For example, DNA-binding proteins are in close proximity to DNA and readily available to form DPCs in the presence of cross-linking agents such as DEB. Several of the proteins found to form DEB-mediated DPCs, such as those involved in DNA replication and repair, are known to associate with DNA in the nucleus. For example, the DNA repair protein PARP is involved in the DNA damage signaling response, and has been previously shown to form cross-links to DNA as a
Figure 2.7 HPLC-ESI+-MS/MS analysis of 1-(S-cysteinyl)-4-(guan-7-yl)-2,3-butanediol (Cys-N7G-BD) conjugates in total proteolytic digests of DEB-induced DPCs. Nuclear protein extracts from HeLa cells were exposed to DEB in the presence of biotinylated DNA duplexes. Following biotin capture enrichment, DPCs were subjected to thermal hydrolysis and enzymatic digestion of proteins to amino acids to release amino acid-nucleobase conjugates: (A) Synthetic Cys-N7G-BD; (B) enzymatic digests of HeLa nuclear protein extracts following incubation with DNA in the absence of DEB (negative control); (C) enzymatic digests of HeLa nuclear protein extracts following incubation with 50 mM DEB in the presence of DNA.
result of exposure to nitrogen mustards,117 ionizing radiation,33 and formaldehyde.57 In contrast, some other proteins identified in our screen (e.g. GAPDH, actin) (Table 2.1) are involved in cellular processes which do not involve association with DNA (e.g. glycolysis, cell motility). However, additional cellular roles have been proposed for many of these proteins, which would explain their ability to associate with DNA and form DPCs. For example, actin is primarily known for its role in cell motility, but is also believed to have regulatory roles in chromatin remodeling, DNA replication, and transcription.196;197

The biological consequences of DNA-protein cross-linking are not fully understood. It is hypothesized that the presence of bulky, helix-distorting DPC lesions interrupts critical cellular metabolic processes and results in damaging cytotoxic and genotoxic effects.1 Previous studies have shown that certain types of DPCs can persist through several cycles of DNA replication, resulting in permanent DNA alterations.1;174;198 Possible mechanisms for DPC repair include NER, HR, and proteolytic degradation.9;69;70 More than one repair mechanism may be required.1;69;70 For example, de Graaf et. al. have reported that HR-deficient yeast strains showed the greatest sensitivity to chronic, low-exposure doses of formaldehyde, whereas NER-deficient yeast strains showed increased sensitivity to acute formaldehyde exposure.76 Nakano and colleagues have reported that the DPC repair mechanism is dependent on the size of the cross-linked protein, with NER repairing cross-links involving peptides and small proteins, and HR repairing cross-links involving larger proteins.69;70 The upper size limit of cross-linked proteins for NER repair of DPCs in bacteria was 12-14 kDa,69 whereas in mammalian cells it was 8-10 kDa.70 This would suggest that the majority of DPCs induced by DEB are expected to be repaired by HR, due to their significant size (Table 2.1).

In conclusion, our study demonstrates the ability DEB to covalently cross-link a number of nuclear proteins to DNA. Although the DEB concentrations employed in this work are, for practical reasons, much higher than typical levels of human exposure, these results still hold
biological relevance due to the likelihood that DPCs are also formed in cells exposed to low levels of DEB. However, it should be noted that the induction of DPCs in vitro requires much higher concentrations of DEB relative to mechlorethamine and cisplatin. This is potentially of interest, considering that these agents are nearly equi-toxic in human fibrosarcoma HT1080 cells (Kurtz and Campbell, unpublished results). Taken together, these findings suggest that DPCs may contribute less to the cytotoxicity associated with DEB exposure as compared to other bis-electrophiles investigated so far (e.g. cisplatin, mechlorethamine). Regardless, the formation of these bulky, helix-distorting DPC lesions would have considerable potential to interfere with critical cellular processes such as replication and transcription, ultimately triggering programmed cell death or genotoxic outcomes.
III. DNA-PROTEIN CROSS-LINKING IN HUMAN FIBROSARCOMA HT1080 CELLS BY THE ANTITUMOR NITROGEN MUSTARD MECHLORETHAMINE

3.1 Introduction

DNA-protein cross-linking can result from exposure to a broad range of chemical and physical agents, including transition metals,98 ionizing radiation,33 and endogenous aldehydes.199 DPCs can also be induced by chemotherapeutic agents, such as platinum drugs,51 alkyl nitrosoureas,52 and nitrogen mustards.173 As compared to other types of DNA lesions, DPCs are unusually bulky and complex adducts which are not well characterized, and their biological consequences are not well understood. Because of their considerable size and their ability to block normal DNA-protein interactions, DPCs interfere with critical cellular processes, including DNA replication, transcription, chromatin remodeling, and DNA repair.1 If left unrepaired, DPCs may persist through several cycles of replication, resulting in permanent DNA alterations that may lead to damaging cytotoxic and mutagenic effects.1;174

Antitumor nitrogen mustards are a class of bifunctional alkylating agents which contain two N-(2-chloroethyl) groups that can spontaneously cyclize under physiological conditions. The resulting highly reactive aziridinium ions are capable of reacting with nucleophilic groups on DNA and proteins to form DPCs and DNA-DNA cross-links (Scheme 3.1).49 This class of compounds, which includes mechlorethamine, chlorambucil, melphalan, and cyclophosphamide, has a long history of clinical use in the treatment of cancer.163 Nitrogen mustard-induced DNA-DNA cross-links are well characterized and are thought to play a prominent role in the antitumor
Scheme 3.1 Formation of DPCs by antitumor nitrogen mustards
effects of these drugs. In contrast, little is known about the composition, chemical structures, and cellular abundance of the corresponding DNA-protein lesions.

The ability of nitrogen mustards to cross-link proteins to DNA was first reported in the 1970s by Ewig and Kohn, who employed alkaline elution methodology to investigate DNA damage in mouse leukemia L1210 cells exposed to mechlorethamine. These studies detected the formation of bulky, proteinase-sensitive DNA adducts hypothesized to be DPCs. Later work by Thomas et al. showed that mechlorethamine treatment resulted in covalent binding of nuclear proteins to chromosomal DNA through alkylated purine residues. Further studies revealed that while these DNA-protein complexes were resistant to separation by detergents and high salt buffers, the proteins were released from the DNA under acidic conditions. However, these primarily biophysical methods could not uncover protein identities, determine DPC amount, or identify the structures and the cross-linking sites within the protein. More recently, our laboratory employed a mass spectrometry-based methodology to demonstrate that the recombinant human DNA repair protein AGT becomes cross-linked to DNA in vitro in the presence of two nitrogen mustards, mechlorethamine and chlorambucil, and that cross-linking takes place through the protein’s active site cysteine residue (Cys) and another nearby cysteine (Cys). Furthermore, we applied affinity capture methodology coupled with mass spectrometry and immunological detection to identify mechloethamine-induced DPCs involving 53 human proteins present in nuclear extracts from human cervical carcinoma HeLa cells.

In the present work, we sought to characterize DPC formation in cultured mammalian cells exposed to mechlorethamine. A simple DPC isolation strategy consisting of phenol/chloroform extraction in the presence of proteasome inhibitors was developed to enable selective purification of the proteins covalently trapped on DNA and the removal of the bulk of non-covalently attached proteins (Scheme 3.2). This approach was coupled with thermal hydrolysis of DNA, electrophoretic separation of the proteins, mass spectrometry-based proteomics, and western
Scheme 3.2 Strategy for the isolation and analysis of DPCs from mechlorethamine-treated mammalian cell cultures

---

- **Treat Cells**
  - Isolate Nuclei
  - Extraction of Chromosomal DNA and Cross-linked Proteins

- **Neutral Thermal Hydrolysis**
  - 1) Neutral Thermal Hydrolysis
  - 2) Total Digest to Amino Acids

- **LC-MS/MS Sequencing of Tryptic Peptides for Protein Identification**

- **Western Blotting Against Candidate Protein Targets**

- **LC-MS/MS Analysis of Amino Acid-Guanine Conjugates**

---

Cys-N7G-EMA
blotting, enabling us to identify 38 proteins that became cross-linked to chromosomal DNA in human fibrosarcoma HT1080 cells following mechlorethamine treatment. Many of these proteins were classified as nuclear proteins which play important roles in critical cellular processes such as transcriptional regulation and chromatin remodeling. Furthermore, HPLC-ESI^-MS/MS analyses of total proteolytic digests revealed a concentration-dependent formation of Cys-N7G-EMA conjugates in HT1080 cells treated with cytotoxic concentrations of the drug as a result of sequential alkylation of cysteine residues within the proteins and guanine nucleobases in DNA by mechlorethamine. Taken together, these results suggest that antitumor nitrogen mustards such as mechlorethamine are capable of forming covalent DPCs involving a wide range of nuclear proteins in vivo. These finding are significant because super-bulky DPC lesions are likely to contribute to the cytotoxic and mutagenic effects of nitrogen mustards and other bis-electrophiles.

3.2 Materials and Methods

Safety Statement – Phenol and chloroform are toxic chemicals that should be handled with caution in a well-ventilated fume hood with appropriate personal protective equipment.

Chemicals and Reagents – Mechlorethamine hydrochloride, PMSF, leupeptin, pepstatin, aprotinin, DTT, iodoacetamide, chloroform, ribonuclease A, nuclease P1, and alkaline phosphatase were purchased from Sigma (St. Louis, MO). Mass spectrometry-grade Trypsin Gold was purchased from Promega (Madison, WI). Proteinase K was obtained from New England Biolabs (Beverly, MA). Primary polyclonal antibodies specific for GAPDH, PARP, and Ref-1 were obtained from Santa Cruz Biotechnology (Santa Cruz, CA). The primary polyclonal antibody to Ku and the monoclonal antibody specific for XRCC-1 were purchased from Lab Vision/NeoMarkers (Fremont, CA). The primary monoclonal antibody against AGT was
purchased from Millipore (Temecula, CA). Alkaline phosphatase-conjugated anti-mouse and anti-rabbit secondary antibodies were purchased from Sigma (St. Louis, MO). Cys-N7G-EMA and Cys-[\textsuperscript{15}N]-N7G-EMA were prepared as described previously.\textsuperscript{95}

**Cell Culture** – Human fibrosarcoma (HT1080) cells\textsuperscript{201} were obtained from the American Type Culture Collection. The cells were maintained as exponentially growing monolayer cultures in DMEM supplemented with 9% FBS. Chinese hamster ovary (CHO) cells expressing recombinant human AGT (CHO-AGT) were generously provided by Dr. Anthony E. Pegg (Pennsylvania State University), and were maintained as exponentially growing monolayer cultures in α-minimal essential medium supplemented with 9% FBS and 1 mg/mL G-418. Both cell lines were maintained in a humidified incubator at 37°C with 5% CO\textsubscript{2}.

**Assay for Cytotoxicity as a Result of Mechlorethamine Exposure** – The effect of mechlorethamine on cell survival was determined via cytotoxicity assay. HT1080 cells were plated in DMEM containing 9% FBS at a density of 5x10\textsuperscript{5} cells/dish and permitted to adhere overnight. The next morning, cells were treated (in triplicate) with mechlorethamine (0, 10, 25, 50, or 100 µM) for 3 h at 37°C. Following treatment, the drug-containing media was replaced with drug-free media, and the cells were incubated for 18 h, followed by cell counting in a haemocytometer. Cytotoxicity was expressed as the number of cells surviving mechlorethamine treatment relative to untreated controls.

**Isolation of Proteins Cross-linked to Chromosomal DNA by Mechlorethamine** – To analyze DPC formation in mammalian cells exposed to mechlorethamine, HT1080 cells were treated with mechlorethamine (0, 10, 25, 50, or 100 µM) for 3 h at 37°C. Following treatment, the cells were harvested, washed three times with ice cold PBS, and suspended in PBS to a final concentration
of ~2 x 10^6 cells/mL. To isolate nuclei, cells were lysed by adding an equal volume of 2X cell lysis buffer (10 mM Tris-HCl/100 mM EDTA/2% (w/v) SDS/0.65 M sucrose, pH 8.0), incubated on ice for 5 min, and centrifuged at 2,000 g for 10 min at 4ºC. The nuclear pellet was re-suspended to a concentration of ~5x10^6 nuclei/mL in a saline-EDTA solution (75 mM NaCl/24 mM EDTA/1% (w/v) SDS, pH 8.0) containing RNase A (10 µg/mL) and a protease inhibitor cocktail (1 mM PMSF; 1µg/mL pepstatin; 0.5 µg/mL leupeptin; 1.5 µg/mL aprotinin), and incubated for 2 h at 37ºC with gentle shaking. To isolate chromosomal DNA containing covalent DPCs, nuclear lysates were extracted by addition of two volumes of Tris-buffer saturated phenol. The resulting white emulsion was centrifuged for at 1,000g for 15 min at room temperature. The aqueous layer and interface material were subjected to a second extraction with two volumes of Tris-buffer saturated phenol:chloroform (1:1). DNA was precipitated from the aqueous layer by addition of two volumes of ice cold 95% ethanol. To collect the DNA, the samples were centrifuged at 80,000 g for 20 min at 4ºC, and the resulting DNA pellet was washed with ice cold 70% ethanol, briefly air dried, and reconstituted in 1 mL MilliPore H2O. DNA concentrations were estimated by UV spectrophotometry. DNA amounts and purity were determined by quantitation of dG in enzymatic hydrolysates as described below.

**Enzymatic Digestion and HPLC Separation of Nucleic Acids for dG Quantitation** – In order to accurately quantify DNA present in samples following phenol/chloroform extraction, approximately 5 µg DNA from each sample was subjected to neutral thermal hydrolysis (1 h at 70ºC) to release protein-guanine conjugates from the DNA backbone. The partially depurinated DNA was digested to nucleosides by incubating samples for 20 h at 37ºC in 5 mM ZnCl₂/50 mM ammonium acetate (pH 5.3) buffer containing nuclease P1 (1 U), alkaline phosphatase (10 U), and supplemented with 45 ng coformycin to prevent deamination of dA. The digest mixtures
were passed through Amicon Ultra-0.5 mL Centrifugal Filters (10K MWCO, Millipore, Temecula, CA) to remove enzymes and proteins prior to HPLC analysis.

The enzymatic digests were analyzed with an Agilent Technologies HPLC System (1100 model) incorporating a diode array detector and an autosampler. Chromatographic separation of nucleosides was achieved using a Zorbax SB-C8 (4.6 x 150 mm, 5 µm) column (Agilent Technologies, Palo Alto, CA) eluted with 150 mM ammonium acetate (A) and acetonitrile (B). The solvent composition was initially held at 0% B for 2 min, followed by a linear increase to 3% B over 13 min, and further to 30% B over 3 min where it was held for the final 7 min. UV absorbance was monitored at 260 nm. By this method, dG eluted as a sharp peak at ~13.5 min. The amount of dG present in each sample was determined by comparing the UV peak area of dG to a calibration curve resulting from injection of known amounts of dG.

Mass Spectrometric Identification of Cross-linked Proteins – To identify cellular proteins that become covalently attached to chromosomal DNA in mechlorethamine-treated cells, HT1080 cells (~10^7 cells, in triplicate) were treated with 25 µM mechlorethamine or solvent control for 3 h at 37°C, and the chromosomal DNA containing covalently cross-linked proteins was isolated by phenol/chloroform extraction and quantified as described above. DNA (30 µg) was subjected to neutral thermal hydrolysis to release protein-guanine conjugates, dried completely under vacuum, and reconstituted in 100 µL of 1X NuPAGE Sample Buffer (Invitrogen, Carlsbad, CA). Proteins were separated using 12% Tris-HCl Ready Gels (Bio-Rad, Hercules, CA) and stained with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA). The gel lanes were divided into five sections which encompassed the entire molecular weight range, and the gel sections were further diced into ~1 mm pieces. The proteins contained within the gel pieces were subjected to in-gel tryptic digestion as described previously.179 Briefly, the gel pieces were rinsed with 100 mM ammonium bicarbonate, and proteins were subjected to reduction with DTT and alkylation with
iodoacetamide. The gel pieces were then dehydrated by incubation with acetonitrile, dried under vacuum, and reconstituted in 25 mM ammonium bicarbonate. Mass spectrometry-grade trypsin (10 µg) was added, and the samples were permitted to digest overnight at 37ºC. The resulting tryptic peptides were extracted from the gel pieces using 60% acetonitrile containing 1% aqueous formic acid, evaporated to dryness, and reconstituted in 0.1% formic acid for MS analysis.

HPLC-ESI+-MS/MS analysis of tryptic peptides was performed on a Thermo LTQ ion trap mass spectrometer (Thermo Electron, San Jose, CA) equipped with an Eksigent nanoLC (Dublin, CA) and a Thermo Surveyor micro-autosampler. Peptides were resolved on a 100 µm x 11 cm fused silica capillary column (Polymicro Technologies, LLC, Phoenix, AZ) with 5 µm, 300 Å Jupiter C18 packing (Phenomenex, Torrence, CA). The column was eluted at a flow rate of 0.6 µL/min with 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The solvent composition was held at 2% B for the first 15 min, followed by a linear increase to 25% B in 35 min, and further to 90% B over 15 min, where it was held constant for the remaining 10 min. The column temperature was maintained at 25ºC. Peptides eluting from the capillary tip were introduced into the LTQ source in micro-electrospray mode with a capillary voltage of ~2 kV. MS/MS spectra were acquired using data-dependent scanning, in which a single full ESI+-MS scan (400 – 2000 m/z) was followed by four MS/MS scans. Spectra were recorded using dynamic exclusion of previously analyzed precursors for 60 s. MS/MS spectra were generated by CID of the peptide ions at a normalized collision energy of 35% in order to generate a series of b- and y-ions as major fragments. Centroided MS/MS scans were acquired using an isolation width of 2 m/z, an activation time of 30 ms, an activation Q of 0.250, and 30% normalized collision energy using 1 microscan with a maximum ion time of 100 ms for each MS/MS scan. The instrument was tuned prior to analysis using the synthetic peptide TpepK (AVAGKAGAR), so that some tune parameters may have varied slightly between experiments. Typically, the tune parameters
were as follows: spray voltage of 2 kV, capillary temperature of 150°C, capillary voltage of 50 V, and tube lens of 120 V.

The “ScanSifter” algorithm v0.1, an in-house developed software, read MS/MS spectra stored as centroided peak lists from Thermo RAW files, and transcoded them into mzML files. Spectra that contained fewer than 10 peaks, or that had less than 20 measured TIC, did not result in mzML files. Only MS/MS scans were written to the mzML files; MS scans were excluded. If at least 90% of a MS/MS spectrum appeared at a lower m/z than the precursor ion, a single precursor charge was assumed. Otherwise, the spectrum was processed under both double and triple charge assumptions. MS/MS spectra were assigned to peptides from the IPI human database version 356 (03 March 2009) by the MyriMatch algorithm, version 1.6.33. To estimate FDRs, each sequence was reversed and concatenated to the database, for a total of 153,182 entries for the human database. The database search encompassed tryptic peptides with a maximum of five missed cleavage sites for enzyme search, and with a maximum number of ten internal cleavage sites. Cysteine residues were expected to undergo carboxamidomethylation (+57 Da), and methionine residues were allowed to be oxidized (+16 Da). Mechlorethamine-induced alkylation at the N-terminus or cysteine, histidine, lysine, or arginine residues (hydrolyzed monoadduct: +102 Da; cross-link to guanine: +234 Da) were specified as dynamic modifications to identify spectra of adducted peptides. Precursor ions were required to fall within 1.25 m/z of the position expected from their average masses, and fragment ions were required to fall within 0.5 m/z of their monoisotopic positions.

Peptide identification, filtering, and protein assembly were accomplished using IDPicker software version 2.5.80, which filtered raw peptide identifications to a target FDR of 5%. The peptide filtering employed reversed sequence database match information to determine thresholds that yielded an estimated 5% FDR for the identifications of each charge state by the formula FDR = (2R)/(R+F), where R is the number of passing reversed peptide identifications, and F is the
number of passing forward (normal orientation) peptide identifications. The second round of filtering removed proteins supported by less than two distinct peptide identifications in the analyses. Indistinguishable proteins were recognized and grouped. Parsimony rules were applied to generate a minimal list of proteins that explained all of the peptides that passed our entry criteria. To ensure that the levels of proteins captured from treated samples were significantly higher than those in untreated samples, only proteins that were present in treated samples only, or displayed at least twice as many spectral counts in treated samples compared to untreated samples were considered positive identifications. Thus, any protein identified in samples from mechlorethamine-treated cells that displayed comparable spectral counts to control samples from untreated cells was disregarded.

**Western Blot Analysis of Identified Proteins** – In order to estimate protein cross-linking efficiency and to identify additional proteins which participate in DPC formation in the presence of mechlorethamine, HT1080 cells (~10^7) were treated with mechlorethamine (0, 10, 25, 50, or 100 µM) for 3 h at 37°C. Chromosomal DNA along with any covalently bound proteins was extracted as described above. In parallel, a whole cell protein lysate was prepared from an equal number of untreated cells as described below for use in estimating cross-linking efficiency. Following dG quantitation, DNA (15 µg) from each sample was subjected to neutral thermal hydrolysis (1 h at 70°C) to release protein-guanine conjugates from the DNA backbone. Proteins were separated by 12% SDS-PAGE and transferred to Trans-blot nitrocellulose membranes (Bio-Rad, Hercules, CA). After blocking in TBS containing 5% (w/v) BSA, the membranes were incubated with the primary antibody for 3 h at room temperature, rinsed three times with TBS, and incubated overnight at 4°C with the corresponding alkaline phosphatase-conjugated secondary antibody. The blots were then subjected to three additional TBS washes, followed by development with SIGMA Fast BCIP/NBT (Sigma, St. Louis, MO). The developed blots were
scanned as image files, and ImageJ software (available free of charge from the NIH website, www.ncbi.nlm.nih.gov) was used to quantify the optical densities of the protein bands. The efficiency of DNA-protein cross-linking in vivo was approximated by comparing signal intensities of the protein which was co-purified with chromosomal DNA (corresponding to total cross-linked protein) and the intensity of the corresponding protein band present in the whole cell protein lysate (representing total cellular protein). In order to detect AGT-DNA cross-linking in cell culture, analogous experiments were conducted using CHO-AGT cells treated with 0, 1, 5, 10, 25, or 50 µM mechlorethamine.

**Preparation of Whole Cell Protein Extracts** – In order to estimate the cross-linking efficiency of DPCs involving specific proteins by western blotting, whole cell protein extracts were prepared in parallel with each phenol/chloroform extraction of DPC-containing chromosomal DNA. HT1080 cells (~10^7) were harvested, washed three times with ice cold PBS, and centrifuged at 1000 g for 5 min at room temperature. The cell pellet was re-suspended in 2 mL lysis buffer (50 mM NaH2PO4/300 mM NaCl/10 mM imidazole, pH 8.0) containing a protease inhibitor cocktail (1 mM PMSF; 1 µg/mL pepstatin; 0.5 µg/mL leupeptin; 0.75 µg/mL aprotinin) and incubated on ice for 5 min prior to sonication (3 x 15 s, 1 min on ice in between). The resulting cellular lysate was centrifuged at 80,000g for 20 min at 4°C, and the cellular proteins were isolated in the clear supernatant.

**Isotope Dilution HPLC-ESI+-MS/MS Analysis of Cys-N7G-EMA in Cells Exposed to Mechlorethamine** – HT1080 cells (~10^6) were treated with mechlorethamine (0, 10, 25, 50, or 100 µM for 3 h at 37°C, in triplicate). The chromosomal DNA containing DPCs was isolated using the modified phenol/chloroform extraction procedure described above, and DNA samples (5 µg) were subjected to neutral thermal hydrolysis (1 h at 37°C) to release protein-guanine.
conjugates from the DNA backbone. Proteins were subjected to tryptic digestion (1 µg trypsin in 25 mM ammonium bicarbonate, overnight at 37°C), and the resulting peptides were further digested to amino acids in the presence of proteinase K (10 µg in 250 µL H$_2$O, overnight at 37°C). The digests were spiked with stable isotope-labeled internal standard (Cys-[${}^{15}$N]-N7G-EMA, 17.3 pmol) and HPLC retention time marker (dA, 0.5 µg), followed by HPLC purification of Cys-N7G-EMA conjugates. An Agilent Technologies HPLC system (1100 model) incorporating a diode array detector, an autosampler, and a fraction collector was fitted with a Supelcosil LC-18-DB (4.6 x 250 mm, 5 µm) column (Sigma-Aldrich, St. Louis, MO). The column was eluted at a flow rate of 1 mL/min using 20 mM ammonium acetate, pH 4.9 (A) and acetonitrile (B). The gradient program was increased linearly from 0 to 24% B over 24 min, and increased further to 60% B in 6 min. Using this method, dA eluted at ~13.5 min. HPLC fractions containing Cys-N7G-EMA (9 – 10.5 min) were collected, dried under vacuum, and reconstituted in 25 µL 15 mM ammonium acetate, pH 5.0 prior to HPLC-ESI^+-MS/MS analysis (8 µL injection).

Quantitative analyses of Cys-N7G-EMA were conducted with an Agilent 1100 capillary HPLC system interfaced to a Thermo-Finnigan TSQ Quantum Discovery mass spectrometer. Chromatographic separation was accomplished using a Phenomenex Synergi Hydro-RP C18 column (250 mm x 0.5 mm, 4 µm) eluted with 15 mM ammonium acetate, pH 5.0 (A) and methanol (B) at a flow rate of 12 µL/min. The solvent composition was linearly changed from 0 to 24% B in 15 min. Under these conditions, Cys-N7G-EMA and its internal standard (Cys-[${}^{15}$N]-N7G-EMA) eluted at ~12 min. Electrospray ionization was achieved at a spray voltage of 3000 V and a capillary temperature of 250°C. CID was performed with Ar as a collision gas (1.0 mTorr) at a collision energy of 10 V. The MS parameters were optimized for maximum response during infusion of a standard solution of Cys-N7G-EMA. HPLC-ESI^+-MS/MS analysis was performed in SRM mode by following the neutral loss of guanine from protonated molecules of
Cys-N7G-EMA (m/z 356.1 [M + H]⁺ → 204.9 [M + H – Gua]⁺) and the corresponding mass transition for Cys-[¹⁵N]-N7G-EMA (m/z 361.1 [M + H]⁺ → 204.9 [M + H – Gua]⁺). The relative response ratios of the analyte peak area compared to the internal standard peak area were used to obtain a concentration dependence curve for Cys-N7G-EMA in HT1080 cells treated with mechlorethamine.

3.3 Results

3.3.1 Cytotoxicity Experiments to Establish Optimal Mechlorethamine Concentrations

To establish optimal conditions for mechlorethamine treatment, HT1080 cells were incubated with 0, 10, 25, 50, or 100 µM mechlorethamine for 3 h. Following drug removal and overnight incubation in regular media, the cells were counted, and the cytotoxicity was measured as the percentage of cells surviving mechlorethamine treatment compared to untreated controls. There was a significant increase in cell death resulting from treatment with increasing concentrations of mechlorethamine, with approximately 50% cell death as a result of treatment with 50 µM mechlorethamine (Figure 3.1). These values are consistent with previous findings of Hardej and Billack, who studied mechlorethamine cytotoxicity in various human cell types and found LC₅₀ values ranging between 6 and 1000 µM following 24 h exposure. Based on these results, 25 µM mechlorethamine was chosen for the proteomics study.

3.3.2 Strategy for the Isolation of DPCs from Mammalian Cells Treated with Mechlorethamine

The proteomic identification of DNA-protein lesions from mechlorethamine-treated HT1080 cells required a method for isolating proteins covalently attached to DNA, while removing any non-covalently bound proteins. Our initial attempts employed the commercial
Figure 3.1 Mechlorethamine-induced cell death in HT1080 cells. Cells (5x10^5/dish) were treated in triplicate with mechlorethamine (0–100 µM). After 3 h, the cell media was replaced with drug-free media and the cells were incubated overnight before being harvested and counted. Cell survival was determined by comparing the number of cells surviving mechlorethamine treatment relative to untreated controls. Error bars represent the standard error from two independent experiments.
DNAzol reagent, a proprietary chaotrope-detergent intended for the isolation of genomic DNA from a variety of sources by lysing cells, hydrolyzing RNA, and dissociating non-covalent DNA-protein complexes reagent (Invitrogen, Carlsbad, CA). DNAzol has been successfully employed by Barker et al. to isolate DPCs induced by ionizing radiation. However, our attempts to use DNAzol to isolate mechlorethamine-induced DPCs yielded poor results due to low solubility of chromosomal DNA containing mechlorethamine-induced DPCs (results not shown). Therefore, an alternative method based on phenol/chloroform extraction was developed.

In our approach (Scheme 3.2), nuclei are isolated from control and treated cells and lysed in the presence of a protease inhibitors. Chromosomal DNA containing covalently cross-linked proteins is isolated using a modified phenol/chloroform extraction method, which incorporates repeated extraction steps in the presence of SDS to remove any non-covalently bound proteins. We found that this method consistently provides good DNA yields and minimizes RNA contamination. To ensure DNA purity and to quantify its amounts, DNA aliquots were digested in the presence of nuclease P1 and alkaline phosphatase, and the resulting nucleosides were analyzed by HPLC with UV detection. As shown in a representative chromatogram in Figure 3.2, no RNA ribonucleosides were detected in enzymatic digests, indicating that our protocol was effective at removing RNA. Quantitative analysis of dG from the same HPLC-UV runs using external calibration curves was used to determine DNA concentrations.

3.3.3 Concentration-Dependent Formation of DPCs in Human Cell Cultures Following Mechlorethamine Treatment

Human fibrosarcoma HT1080 cells were treated with mechlorethamine (0, 10, 50, or 100 µM) for 3 hours. Chromosomal DNA containing any covalent DPCs was isolated by modified phenol/chloroform extraction as described above, and equal amounts of DNA from each sample (15 µg) were dissolved in SDS-containing gel loading buffer and analyzed by 12% SDS-PAGE.
Figure 3.2  Representative HPLC-UV chromatogram for the separation of nucleoside mixtures resulting from the enzymatic digestion of nucleic acids that were isolated from HT1080 cells treated with mechlorethamine to induce DNA-protein cross-linking. After harvesting cell nuclei, DNA containing cross-linked proteins was isolated by a phenol/chloroform extraction protocol optimized for the extraction of DPCs. The identities of nucleosides were confirmed by HPLC analysis of authentic standards of the four natural deoxyribonucleosides, “dA”, “dC”, “dG”, and “dT”. Coformycin was added to digest reactions to inhibit the deaminase activity of alkaline phosphatase. Accurate DNA quantification was accomplished by comparing dG peak areas in all samples to a standard calibration curve obtained by injecting known amounts of dG.
Figure 3.3 SDS-PAGE analysis of concentration-dependent mechlorethamine-induced DNA-protein cross-linking in HT1080 cells. Cells were treated with 0-100 µM mechlorethamine (lanes 3-6) for 3 h, and chromosomal DNA and cross-linked proteins were isolated using phenol/chloroform extraction. Proteins (from 15 µg DNA) were separated by SDS-PAGE and visualized by staining with SimplyBlue SafeStain (A), and densitometric analysis of protein bands in the 25–250 kDa molecular weight region was used to approximate the extent of DNA-protein cross-linking (B).
Total nuclear extract proteins were analyzed in parallel to estimate cross-linking efficiency. The proteins were visualized by SimplyBlue SafeStain (panel A). We found that the intensities of protein bands increased with increasing drug concentration (lanes 4-6 in panel A), reaching up to 1.6% cross-linking following treatment with 100 µM mechlorethamine (panel B). Very little protein was observed in a control lane containing DNA from untreated cells (lane 3 in panel A). Based on the degree of protein staining in the gel, 25 µM mechlorethamine was selected as the drug concentration to be employed in the proteomics analysis.

### 3.3.4 Identification of Cross-linked Proteins by Mass Spectrometry-Based Proteomics

To identify the proteins participating in DPC formation in the presence of nitrogen mustard, HT1080 cells were treated (in triplicate) with 25 µM mechlorethamine, while control cells were treated (in triplicate) with the solvent blank. Chromosomal DNA was extracted by the modified phenol/chloroform extraction method described above, and equal DNA amounts from each sample (30 µg) were taken for further analysis. To simplify protein identification by MS, mechlorethamine-induced DPCs were subjected to thermal hydrolysis to induce depurination of N7-guanine adducts. This treatment releases DPCs from the DNA backbone in the form of protein-guanine conjugates (Scheme 3.2). The proteins were then separated by SDS-PAGE (Figure 3.4). Protein staining revealed distinct signals which corresponded to DNA-bound proteins in mechlorethamine-treated samples (panel B). In contrast, the untreated sample exhibited only weak protein signals (panel A), suggesting that our DNA isolation methods remove the bulk of non-covalently bound proteins. All protein bands in the molecular weight range of 20–250 kDa were excised from the gel and subjected to in-gel tryptic digestion. The resulting peptides were subjected to HPLC-ESI-MS/MS analysis for protein identification. As shown for representative peptides in Figure 3.5, HPLC-ESI-MS/MS analysis of the tryptic peptides yielded characteristic b- and y-series fragment ions that were used to determine the
Figure 3.4 SDS-PAGE analysis DPCs from HT1080 cells treated with mechlorethamine. Cells (~10^7) were treated with 0 (A) or 25 μM (B) for 3 h. Following phenol/chloroform extraction, cross-linked proteins were separated by 12% SDS-PAGE and visualized by staining with SimplyBlue SafeStain. Proteins present in the 20 – 250 kDa molecular weight range were excised from the gel, subjected to in-gel tryptic digestion, and analyzed by HPLC-ESI-MS/MS for the identification of cross-linked proteins.
Figure 3.5 Representative HPLC-ESI+-MS/MS spectra of tryptic peptides used in the identification of DPCs involving matrin-3 (A), and zinc finger Ran-binding domain-containing protein 2 (B).
amino acid sequences of the peptides and to identify the proteins co-purified with chromosomal DNA in mechlorethamine treated cells.

Database searching and parsimony analysis of the MS/MS spectral data resulted in the identification of 34 proteins that formed DPCs in the presence of mechlorethamine (Table 3.1). All proteins were required to be supported by a minimum of two unique peptides. Furthermore, only proteins which were present solely in the mechlorethamine-treated samples or displayed at least twice as many spectral counts in treated samples compared to untreated controls were considered positive identifications. These criteria ensured that only proteins which exhibited significantly increased ion counts in treated versus untreated samples were included in the list. Furthermore, proteins were classified by molecular weight (Table 3.1), and were expected to have originated in the correct molecular weight region of the gel (Figure 3.4). The vast majority of peptides from nearly all of the identified proteins were found in the correct molecular weight region of the gel as predicted by their molecular weight (Table 3.1). Peptides from some of the proteins were detected in lower molecular weight regions of the gel than would be expected based on their mass, which could be due to proteolytic degradation of the cross-linked proteins in the cell or during phenol/chloroform extraction. In addition, peptides from a few proteins were present in higher molecular weight fractions than were predicted. These gel shifts could be attributed to protein-protein cross-linking.

Of the identified proteins in Table 3.1, 23 (67.6%) were classified as nuclear proteins by the GO database available via the European Bioinformatics Institute (http://www.ebi.ac.uk/QuickGO) (Figure 3.6, Panel A). These included nucleophosmin, matrin-3, and protein DEK. This is not surprising, given that only proteins present in the nucleus are available for cross-linking to chromosomal DNA. An additional 11 (32.4%) of proteins were classified as cytoplasmic, and 5 (14.7%) were classified as membrane-bound proteins. It is
Table 3.1 Proteins that form cross-links to chromosomal DNA in human fibrosarcoma (HT1080) cells in the presence of mechlorethamine*

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9NYF8</td>
<td>Bcl-2-associated transcription factor 1, Isoform 1</td>
<td>13</td>
<td>51</td>
<td>13</td>
<td>2</td>
<td>106122</td>
</tr>
<tr>
<td>P17096</td>
<td>High mobility group protein HMG-I/HMG-Y, Isoform HMG-I</td>
<td>4</td>
<td>10</td>
<td>38</td>
<td>0</td>
<td>11545</td>
</tr>
<tr>
<td>P52926</td>
<td>High mobility group protein HMG-I-C</td>
<td>3</td>
<td>5</td>
<td>33</td>
<td>0</td>
<td>11832</td>
</tr>
<tr>
<td>P43243</td>
<td>Matrin-3</td>
<td>7</td>
<td>24</td>
<td>10</td>
<td>9</td>
<td>94492</td>
</tr>
<tr>
<td>Q9NWH9</td>
<td>Modulator of estrogen induced transcription, Isoform b</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>117017</td>
</tr>
<tr>
<td>P17480</td>
<td>Nucleolar transcription factor 1, Isoform UBF1</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>89406</td>
</tr>
<tr>
<td>P06748</td>
<td>Nucleophosmin, Isoform 2</td>
<td>6</td>
<td>29</td>
<td>17</td>
<td>3</td>
<td>32575</td>
</tr>
<tr>
<td>Q9UQ88</td>
<td>PITSLRE serine/threonine-protein kinase CDC2L2, Isoform SV1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>90974</td>
</tr>
<tr>
<td>P35659</td>
<td>Protein DEK</td>
<td>5</td>
<td>25</td>
<td>10</td>
<td>5</td>
<td>42543</td>
</tr>
<tr>
<td>Q96T23</td>
<td>Remodeling and spacing factor 1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>27</td>
<td>163821</td>
</tr>
<tr>
<td>P18583</td>
<td>SON protein, Isoform F</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>263715</td>
</tr>
<tr>
<td>Q969G3</td>
<td>SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1, Isoform 1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>46649</td>
</tr>
<tr>
<td>Q5BKZ1</td>
<td>Zinc finger protein 326, Isoform 1</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>65654</td>
</tr>
<tr>
<td>P62979</td>
<td>40S ribosomal protein S27a</td>
<td>3</td>
<td>10</td>
<td>16</td>
<td>6</td>
<td>9418</td>
</tr>
</tbody>
</table>

*25 µM for 3 h
Table 3.1 (continued from p. 87)

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
</tr>
</thead>
<tbody>
<tr>
<td>P78316</td>
<td>Nucleolar protein 14, Isoform 1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>97668</td>
</tr>
<tr>
<td>Q9Y3B9</td>
<td>Ribosomal RNA-processing protein 15</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>31353</td>
</tr>
<tr>
<td>Q07955</td>
<td>Splicing factor, arginine/serine-rich 1, Isoform ASF-1</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>27613</td>
</tr>
<tr>
<td>Q8TF01</td>
<td>Splicing factor, arginine/serine-rich 18, Isoform 1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>92577</td>
</tr>
<tr>
<td>Q16629</td>
<td>Splicing factor, arginine/serine-rich 7, Isoform 1</td>
<td>5</td>
<td>18</td>
<td>21</td>
<td>4</td>
<td>27367</td>
</tr>
<tr>
<td>P62995</td>
<td>Transformer-2 protein homolog β</td>
<td>6</td>
<td>13</td>
<td>17</td>
<td>2</td>
<td>33666</td>
</tr>
<tr>
<td>Q9Y2W2</td>
<td>WW domain-binding protein 11</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>13446</td>
</tr>
<tr>
<td>O95218</td>
<td>Zinc finger Ran-binding domain-containing protein 2, Isoform 1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>37404</td>
</tr>
</tbody>
</table>

**RNA processing/mRNA splicing**

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
</tr>
</thead>
<tbody>
<tr>
<td>P05556</td>
<td>Integirn β-1, Isoform β-1A</td>
<td>8</td>
<td>15</td>
<td>11</td>
<td>56</td>
<td>86191</td>
</tr>
<tr>
<td>P46821</td>
<td>Microtubule-associated protein 1B</td>
<td>11</td>
<td>18</td>
<td>5</td>
<td>21</td>
<td>270489</td>
</tr>
<tr>
<td>P07196</td>
<td>Neurofilament light polypeptide</td>
<td>6</td>
<td>15</td>
<td>10</td>
<td>1</td>
<td>61385</td>
</tr>
<tr>
<td>P08670</td>
<td>Vimentin</td>
<td>55</td>
<td>308</td>
<td>73</td>
<td>1</td>
<td>53520</td>
</tr>
</tbody>
</table>

**Cell signaling/motility/architecture**
Table 3.1 (continued from p. 88)

DNA damage response/DNA repair

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q7Z4V5</td>
<td>Hepatoma-derived growth factor-related protein 2, Isoform 1</td>
<td>2</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>74137</td>
</tr>
<tr>
<td>Q12888</td>
<td>Tumor suppressor p53-binding protein 1, Isoform 1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>36</td>
<td>213574</td>
</tr>
</tbody>
</table>

Cellular homeostasis/cell cycle

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
</tr>
</thead>
<tbody>
<tr>
<td>P15144</td>
<td>Aminopeptidase N</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>109408</td>
</tr>
<tr>
<td>P12270</td>
<td>Nucleoprotein TPR</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>267161</td>
</tr>
<tr>
<td>Q14568</td>
<td>Putative heat shock protein HSP 90-alpha A2</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>39365</td>
</tr>
</tbody>
</table>

Unknown cellular function

<table>
<thead>
<tr>
<th>Swiss-Prot-ID</th>
<th>Protein</th>
<th>Peptide Sequences</th>
<th>Total Spectra</th>
<th>% Coverage</th>
<th>Protein MW (Da)</th>
<th>No. of Cysteines</th>
</tr>
</thead>
<tbody>
<tr>
<td>B4DJQ5</td>
<td>cDNA FLJ59211, highly similar to Glucosidase 2, Subunit β</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>16</td>
<td>60134</td>
</tr>
<tr>
<td>A8MQ38</td>
<td>Putative uncharacterized protein NOP2</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>12</td>
<td>93983</td>
</tr>
<tr>
<td>Q86VM9</td>
<td>Zinc finger CCCH domain-containing protein 18, Isoform 1</td>
<td>6</td>
<td>13</td>
<td>7</td>
<td>4</td>
<td>106378</td>
</tr>
</tbody>
</table>
Figure 3.6  GO annotations for the cellular distributions (A), molecular functions (B), and biological processes (C) of proteins that form DPCs in the presence of mechlorethamine. The number of proteins in each category is labeled on the charts. Many proteins are counted under more than one GO annotation category.
possible that many of these proteins have additional cellular roles which are not annotated in the GO database, but would explain their presence in the nucleus and ability to interact with chromosomal DNA. DPC-forming proteins were further classified according to their GO annotations relating to molecular functions and biological processes (Figure 3.6, panels B and C). We found that a large portion of the identified proteins (23 proteins, or 67.6%) are classified as nucleic acid binding proteins (Table 3.1 and Figure 3.6, panel B). In contrast, only 10.3% of total nuclear proteins of human fibroblasts are listed as DNA- or RNA-binding. In particular, many of the identified proteins (14, or 41.2%) have roles in transcriptional regulation, including transcription factors, activators, and repressors (Figure 3.6, panel C). This group includes high mobility group protein HMG-I, Bcl-2 associated transcription factor 1, and SON. An additional 10, or 29.4% of proteins are involved in the processing and/or splicing of RNA, including zinc finger Ran-binding domain-containing protein-2 and transformer-2 protein homolog β.

It is important to note that many of the identified proteins function in more than one component of the cell due to their varied roles in multiple biological processes. This observation is reflected in their GO annotations, where the majority of proteins are counted in multiple GO categories, and therefore the sum of the percentages in Figure 3.6 is greater than 100%. For example, Bcl-2 associated transcription factor 1 is a DNA-binding transcriptional repressor that is believed to play a role in tumor suppression via the induction of apoptosis. As such, the protein is counted under two GO annotation categories, transcriptional regulation and apoptosis (Figure 3.6, panel C). Conversely, the available annotation information may fail to account for a protein’s secondary roles, rendering it possible for a protein to be classified solely according to its primary cellular localization, biological process, or molecular function.
3.3.5 Western Blot Analysis of Cross-linked Proteins

Because our previous experiments with cell free protein extracts from human HeLa cells revealed DPC formation involving many additional proteins including GAPDH, PARP, Ref-1, Ku, AGT, and XRCC-1, western blot analyses were conducted to test the ability of these proteins to become cross-linked to chromosomal DNA in cells. DPCs were isolated from HT1080 cells treated with 10, 25, and 50 µM mechlorethamine using the modified phenol/chloroform extraction methodology described above. Equal amounts of DNA (15 µg) were heated with SDS-containing gel loading buffer to release protein-guanine conjugates from the DNA backbone. The resulting proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes for western blot analysis. The blots were probed using antibodies specific for several proteins which were previously found to form mechlorethamine-induced DPCs in vitro. These experiments identified four additional proteins that participated in DPC formation: GAPDH, PARP, Ref-1, and XRCC-1 (Figure 3.7, panel A). These western blots revealed concentration-dependent formation of DPC lesions containing these target proteins (lanes 1-4 of panel A), suggesting that the observed protein signals are the result of covalent DPC formation. Among the proteins analyzed by western blotting, Ref-1 displayed the greatest cross-linking efficiency, with approximately 2.5 % of total protein becoming cross-linked to DNA following treatment with 25 µM mechlorethamine (Figure 3.7, panel B). In contrast, no DPCs involving AGT and Ku were detected (results not shown). We hypothesized that our failure to detect DPCs involving AGT was due to its low cellular abundance. Indeed, the concentration-dependent formation of AGT-DNA cross-links was detected when the western blot experiment was repeated using cross-linked proteins isolated from mechlorethamine-treated CHO cells which over-express human AGT protein (Figure 3.8). These results suggest that detection of cross-linking involving other low abundance proteins may require extremely sensitive methodology.
Figure 3.7 Western blot analysis of mechlorethamine-induced DPCs in HT1080 cells. Following treatment with 0-50 µM mechlorethamine (lanes 1-4), DNA and covalently cross-linked proteins were isolated by phenol/chloroform extraction. Samples were normalized for DNA content, and proteins from 15 µg DNA were separated by SDS-PAGE and transferred to nitrocellulose membranes. Western blots were performed using primary antibodies specific for GAPDH, PARP, Ref-1, and XRCC-1 (A). The efficiency of DPC formation in the presence of mechlorethamine was estimated by densitometric analysis of protein bands in DPC samples and a whole cell protein lysate control (B).
**Figure 3.8** Western blot analysis of AGT-DNA cross-links in mechlorethamine-treated CHO-AGT cells. Following treatment with 0-50 µM mechlorethamine (lanes 1-6), DNA and covalently cross-linked proteins were isolated by phenol/chloroform extraction. Samples were normalized for DNA content, and proteins from 15 µg DNA were separated by SDS-PAGE and transferred to nitrocellulose membranes. Western blotting was performed using a primary antibody against AGT.
3.3.6 HPLC-ESI⁺-MS/MS Analysis of Cys-N7G-EMA Conjugates as Evidence for DPC Formation

To confirm that mechlorethamine treatment induces covalent DPCs in vivo, HT1080 cells were treated with 0, 10, 25, 50, or 100 µM mechlorethamine, and the chromosomal DNA containing DPCs was extracted as described above. Equal DNA amounts (5 µg) were taken from each sample for conjugate analysis. Following neutral thermal hydrolysis of DNA to release protein-guanine conjugates from the DNA backbone, the proteins were enzymatically digested to amino acids, and the digest mixtures were subjected to HPLC purification to enrich for Cys-N7G-EMA conjugates prior to HPLC-ESI⁺-MS/MS analysis. Quantitation of Cys-N7G-EMA was achieved by isotope dilution with Cys-[¹⁵N]-N7G-EMA.

Representative extracted ion chromatograms for HPLC-ESI⁺-MS/MS analysis of Cys-N7G-EMA in samples from mechlorethamine-treated and control HT1080 cells are shown in Figure 3.9. Cys-N7G-EMA was detected DNA samples from mechlorethamine-treated cells (Figure 3.9, panel B), but not from untreated cells (panel A), consistent with conjugate formation as a result of mechlorethamine treatment. Adduct amounts increased linearly with increasing concentrations of mechlorethamine, with approximately 2.1 ± 0.7 adducts formed per 10⁶ dG following treatment with 50 µM mechlorethamine (Figure 3.10). These data indicate that mechlorethamine-induced DNA-protein cross-linking can occur between the side chain sulfhydryls of cysteine residues in proteins and the N7-position of guanine bases in chromosomal DNA. These findings are consistent with our previous in vitro studies, in which we detected Cys-N7G-EMA conjugates in digest mixtures from nuclear extract proteins which were cross-linked to synthetic oligonucleotide duplexes by mechlorethamine. Although these results confirm the participation of cysteine residues in DPC formation, we cannot exclude the possibility that other amino acids, such as lysines and arginines, also participate in mechlorethamine-induced DNA-protein cross-linking.
Figure 3.9 HPLC-ESI+-MS/MS analysis of Cys-N7G-EMA conjugates in total proteolytic digests of mechlorethamine-induced DPCs isolated from HT1080 cells. HT1080 cells were treated with mechlorethamine to induce DNA-protein cross-linking. Following extraction of the chromosomal DNA containing covalent DPCs, the proteins were subjected to thermal and enzymatic hydrolysis to release amino acid-nucleobase conjugates. Digest mixtures were spiked with isotopically labeled internal standard to enable the direct quantitation of Cys-N7G-EMA. Enzymatic digests of DPC mixtures from HT1080 cells incubated in the absence of mechlorethamine (negative control) (A); enzymatic digests of DPC mixtures treated with 10 µM mechlorethamine (B).
Figure 3.10 Concentration-dependent formation of Cys-N7G-EMA in mechlorethamine-treated HT1080 cells. HT1080 cells were exposed to 0, 10, 25, 50, or 100 µM mechlorethamine for 3 h. After extracting the chromosomal DNA containing DPCs, equal amounts of DNA from each sample were subjected to thermal and enzymatic hydrolysis to release amino acid-nucleobase conjugates. The samples were subjected to offline HPLC to enrich for Cys-N7G-EMA prior to HPLC-ESI+-MS/MS analysis. Quantification of Cys-N7G-EMA was accomplished using isotope dilution with Cys-[15N]-N7G-EMA internal standard. Error bars represent the standard error of three independent experiments.
3.4 Discussion

Many common antitumor drugs and genotoxic carcinogens are bis-electrophiles capable of sequentially reacting with two nucleophilic sites on DNA and proteins to form macromolecular cross-links. We have developed a simple and effective protocol for the isolation and characterization of chromosomal DNA containing covalent DPC lesions from cultured cells exposed to cross-linking agents (Scheme 3.2). By combining this methodology with mass spectrometry-based proteomics and immunoblotting, we examined DNA-protein cross-linking in human fibrosarcoma HT1080 cells exposed to the antitumor nitrogen mustard mechlorethamine. A total of 38 proteins were found to form cross-links to chromosomal DNA in the presence of mechlorethamine (Table 3.1 and Figure 3.7). These proteins encompass a variety of cellular functions, including transcriptional regulation, RNA processing, cell signaling, apoptosis, and DNA damage. The majority of the identified proteins are known nucleic acid binding proteins which are present in the nucleus (Figure 3.6). We identified several DPC-forming proteins which are classified as RNA-binding (Figure 3.6, panel C). However, our failure to detect RNA ribonucleosides in these samples by HPLC-UV analysis (Figure 3.2) suggests that these proteins are not the result of interactions with contaminating RNA. An alternative explanation is that these proteins have DNA-binding capabilities which have yet to be observed and/or reflected in the GO database.

In an earlier study, Barker et. al. employed DNAzol extraction of DPCs with mass spectrometry-based proteomics to identify 29 proteins which became cross-linked to chromosomal DNA as a result of exposure to ionizing radiation. Among these were proteins involved in cell structure and architecture, cell cycle, chromatin regulation, transcriptional regulation, and RNA splicing. Of these proteins, vimentin was the only protein to overlap with those identified in the present analysis. Vimentin is a structural protein which, along with
microtubules and actin, comprises the cytoskeleton of eukaryotic cells.\textsuperscript{211} The high cellular abundance of this protein likely facilitated its identification in both proteomics screens. Conversely, the majority of the identified proteins were not identified in both analyses. A likely explanation for the significant variation in protein identifications is the different mechanisms of cross-linking of ionizing radiation and mechloretamine. For example, radiation-induced DPCs often involve tyrosine residues,\textsuperscript{28} whereas mechloretamine-mediated DPCs predominantly involve cysteine residues.\textsuperscript{95,117} Therefore, the identities of the amino acid side chains present in a protein’s DNA-binding domain may influence the likelihood of that protein becoming cross-linked to the chromosomal DNA.

More recently, Qiu and Wang employed a similar methodology to study DNA-protein interactions in human acute promyelocytic leukemia cells.\textsuperscript{57} Proteins associated with chromosomal DNA were reversibly trapped in covalent DNA-protein complexes by FA treatment, enabling the identification of 780 DNA-binding proteins.\textsuperscript{57} Of these, 39.1\% were classified as nuclear proteins, and 40.8\% were annotated as cytoplasmic proteins. Many proteins were classified as both nuclear and cytoplasmic. DNA-binding proteins comprised 14.2\% of the identified proteins, while RNA-binding proteins made up 20.0\% of total proteins. Twelve of the identified proteins were also identified in the present work, including tumor suppressor p53-binding protein 1, matrin-3, high mobility group protein HMG-I, and nucleophosmin (Table 3.1). The identification of hundreds of additional proteins which were not detected in either our analysis of mechloretamine-induced DPCs or the Barker \textit{et al.} analysis of ionizing radiation-mediated DPCs\textsuperscript{33} is likely a reflection of the FA concentrations employed in the study, the high reactivity of FA, and the low specificity of FA-mediated cross-linking.\textsuperscript{57}

Our results for mechloretamine-induced DPC formation in HT1080 cell cultures differ from the previous data obtained with cell free protein extracts from human cervical carcinoma (HeLa) cells.\textsuperscript{117} Of the proteins identified in the \textit{in vitro} study,\textsuperscript{117} only two (matrin-3 and
nucleophosmin) were also detected among DPCs induced by mechlorethamine in vivo (Table 3.1). Though the exact biological function of matrin-3 is unclear, this protein is known to associate with other cellular proteins involved in multiple biological processes, including chromatin remodeling, translation, DNA replication, DNA repair, RNA metabolism, and apoptosis. \(^{212}\) Likewise, nucleophosmin is a multifunctional nuclear phosphoprotein involved in a variety of cellular functions, including transcriptional regulation, centrosome duplication, and ribosomal protein assembly. \(^{213}\) Altered expression of nucleophosmin has been observed in many cancers, including hepatocellular carcinoma, \(^{213}\) acute myeloid leukemia, \(^{214}\) and arsenic-related bladder cancers. \(^{215}\) Both matrin-3 and nucleophosmin are expressed in the nucleus at a high level, \(^{212,213}\) which likely facilitates their detection in both proteomics screens.

The list of proteins identified among the targets of DPC formation in mechlorethamine-treated HT1080 cell cultures (Table 3.1) displayed very little overlap with the proteins cross-linked to biotinylated DNA duplexes in a cell free system employing nuclear protein extracts from HeLa cells. \(^{117}\) Additionally, western blot analyses revealed that 10-fold greater concentrations of mechlorethamine were required to achieve DPC formation in nuclear protein extracts as compared to intact cells. \(^{117}\) Several factors are likely to contribute to these discrepancies, including differences in protein abundance between the two cell types, and, most importantly, chromatin structure. For example, the HeLa cells used in the in vitro studies are derived from a human cervical carcinoma cell line, whereas the HT1080 cells employed in the present study originated from a human fibrosarcoma cell line. These two cell types display differential protein expression profiles, \(^{195,205}\) which may contribute to the observed differences in protein identification. However, it is likely that the most important factor in explaining the discrepancies in protein identifications is that the DNA of living cells is organized in chromatin and involves specific interactions between DNA and proteins at recognition sites, increasing the probability of a protein becoming cross-linked to chromosomal DNA. In contrast, our in vitro
studies employed a short DNA duplex that lacked chromosomal structure and was unlikely to participate in sequence-specific interactions with DNA-binding proteins. Rather, in cell-free nuclear extracts, there was a high occurrence of DPC formation involving high abundance proteins. For example, tubulin is a major constituent of microtubules, the protein filaments which comprise the cytoskeleton of eukaryotic cells. While some forms of tubulin have been observed in the nucleus, little is known about its functional role within the nucleus or its propensity to interact with chromosomal DNA. The high cellular abundance of this protein, in conjunction with simple mass action considerations, likely facilitated its ability to become cross-linked to DNA in the presence of mechlorethamine in vitro. In contrast, even a high abundance protein would likely need to function in close proximity to chromosomal DNA within the nucleus in order to form DPCs in the cell, a possible explanation for our failure to detect DPCs involving tubulin in vivo.

Western blot analysis of DPCs from mechlorethamine-treated HT1080 cells resulted in the detection of four additional proteins (GAPDH, PARP, Ref-1, and XRCC-1) that were not identified in the proteomics screen (Figure 3.7, panel A). Cross-linking was concentration-dependent following exposure to low concentrations of mechlorethamine (Figure 3.7, panel B). However, when cells were exposed to mechlorethamine concentrations higher than 50 µM, some proteins exhibited similar or significantly lower cross-linking efficiency than they did when lower concentrations of mechlorethamine were used (results not shown). One possible explanation is that mechlorethamine exposure triggers the onset of programmed cell death. An early step in this process is the activation of initiator caspases that activate downstream proteolytic degradation of specific cellular protein targets. For example, PARP is an essential nuclear protein involved in the cellular response to irreparable DNA damage. Caspase-mediated cleavage of PARP during the onset of apoptosis is well-documented. It is therefore plausible that the induction of
apoptosis results in the degradation and loss of other DPC-forming proteins, and that these effects are significantly more pronounced in cells exposed to higher concentrations of mechlorethamine.

Cellular repair of DPC lesions may also contribute to the observed plateau in relative cross-linking efficiency following exposure to higher concentrations of mechlorethamine. Several possible mechanisms for DPC repair have been proposed, including proteolytic degradation, NER, and HR, and there is substantial evidence for the involvement of more than one repair pathway. For example, Quievryn and Zhitkovich have shown that FA-induced DPCs are at least partially removed by proteolytic degradation, and suggested that removal of bulky DPC lesions may involve an initial proteolysis step followed by NER repair of the resulting DNA-peptide cross-links. Nakano et al. have studied DPC repair in bacteria and mammalian cells, and found that NER is involved in the removal of DPCs involving small proteins and peptides, whereas HR is responsible for the repair of cross-links involving larger proteins. Additionally, repair mechanism may depend on the degree of exposure to a cross-linking agent, as demonstrated by de Graaf and colleagues. Taken together, these findings suggest that the detection of DPCs formed in a living cellular environment may be affected by normal biological processes, such as the repair of damaging DNA lesions.

Although our western blot data imply that DPC formation plateaus at higher concentrations of mechlorethamine, the concentration-dependent formation of Cys-N7G-EMA conjugates in total proteolytic digests of DPC mixtures suggests otherwise (Figure 3.10), rendering it possible that there is a secondary source of thiols capable of producing Cys-N7G-EMA conjugates. One such thiol is glutathione (GSH), a naturally occurring tripeptide (L-γ-glutamyl-L-cysteinylglycine) that, in conjunction with the enzyme glutathione S-transferase, plays a crucial role in the detoxification of a variety of endogenous and exogenous chemicals. Toxic electrophiles and reactive oxygen species are frequently subject to GSH conjugation in vivo, and microsomal enzymes often function to introduce electrophilic centers into compounds lacking
such moieties to make them more susceptible to GSH conjugation.\textsuperscript{222} Interestingly, GSH has been implicated in the resistance of tumors to a variety of chemotherapeutic drugs, including mustard compounds, by forming covalent conjugates with these agents.\textsuperscript{223,224} GSH is present in the nucleus,\textsuperscript{225} and can form cross-links to DNA in the presence of FA,\textsuperscript{226,227} rendering it possible that some fraction of the Cys-N7G-EMA detected in our studies is the result of DNA-GSH cross-linking. Because both DNA-protein and DNA-GSH cross-links would be extracted with the chromosomal DNA, released from the DNA backbone by thermal hydrolysis, and enzymatically digested to Cys-N7G-EMA conjugates, it would be impossible to distinguish conjugates resulting from digestion of DNA-GSH cross-links from those resulting from DPCs using the methods described above. Further studies are needed to establish the source of Cys-N7G-EMA conjugates and confirm that covalent DPCs are indeed formed following cellular exposure to mechlorethamine.

In conclusion, this study demonstrates that DNA-protein cross-links involving a variety of cellular proteins are formed in human fibrosarcoma cells following exposure to cytotoxic concentrations of the antitumor nitrogen mustard mechlorethamine. Many of the identified proteins are low abundance nuclear proteins involved in cellular processes that place them in close proximity to chromosomal DNA, including transcriptional regulation and DNA damage response. The formation of bulky DPC lesions on chromosomal DNA is likely to have serious consequences for the cell, blocking the progression of protein complexes involved in critical cellular processes, such as DNA replication, chromatin remodeling, transcription, and DNA repair, and ultimately resulting in cytotoxic and mutagenic effects.\textsuperscript{1} Further studies are currently underway to more fully establish the biological consequences of DNA-protein cross-linking (see Chapter IV).
IV. SELECTIVE INDUCTION OF DNA-PROTEIN CROSS-LINKS IN MAMMALIAN CELL CULTURES

4.1 Introduction

DNA-protein cross-linking occurs when a protein becomes covalently attached to chromosomal DNA. DPCs are ubiquitous lesions that can form as the result of cellular exposure to a variety of physical and chemical agents,\(^1\) including formaldehyde,\(^{16}\) ionizing radiation,\(^{33}\) and anticancer drugs.\(^{51,52,173}\) If not repaired, DPC lesions are likely to block the formation and progression of normal DNA-protein complexes which are necessary to critical cellular processes, such as transcription and DNA replication, and are hypothesized to contribute to cytotoxicity.\(^1\) However, the contribution of DPCs to the observed cellular effects of DNA-alkylating agents is poorly understood due to the propensity of these agents to induce other types of DNA damage. For example, the carcinogenic diepoxide DEB is capable of forming DNA-DNA cross-links\(^{100,101}\) and monoadducts\(^{102,228}\) (Scheme 4.1) in addition to DPCs.\(^{94,114,115}\) DPCs are estimated to constitute only 1–3% of total DNA damage following exposure to ionizing radiation and \textit{bis}-electrophiles,\(^{1,33,49}\) making it difficult to completely understand the role of DPCs in the cytotoxic and mutagenic effects observed following DEB exposure.\(^{113,177}\) To overcome this challenge, it is necessary to develop methodology for selectively introducing DPCs into cells in order to observe the biological effects of DPCs in the absence of other types of DNA damage.

The cytotoxicity and mutagenicity of several \textit{bis}-electrophiles, including 1,2-dibromoethane, dibromomethane, and DEB, are enhanced in bacteria which over-express DNA alkyltransferase repair proteins, such as human AGT protein.\(^{113,129,131-133,229}\) AGT is a DNA repair protein that functions to protect the human genome from the damaging effects of promutagenic \(O^6\)-alkylguanine lesions that form as a result of exposure to simple alkylating agents, such as
Scheme 4.1 Formation of bifunctional DNA adducts by DEB

**DNA** → **HEB monoadducts** → **DNA-protein cross-links** → **DNA-DNA cross-links** → **THB monoadducts** → **Exocyclic adducts**
chemotherapeutic drugs and environmental toxins.\textsuperscript{120} During repair reactions, the $O^\delta$-alkylguanine substituent is flipped out of the base stack into the protein’s active site, where it is subjected to nucleophilic attack by the activated side chain thiolate anion of Cys$^{\text{145,127}}$. The irreversible transfer of the $O^\delta$-alkyl lesion from guanine to Cys$^{\text{145}}$ restores intact guanine within the DNA duplex. Alkylation at Cys$^{\text{145}}$ destabilizes the protein’s tertiary structure, thereby targeting AGT for ubiquitination and proteosomal degradation. The AGT-mediated enhancement of \textit{bis}-electrophile genotoxicity in bacteria has been attributed to the formation of toxic AGT-DNA cross-links.\textsuperscript{113,135}

Recent studies have employed mass spectrometry to characterize AGT-DNA cross-linking by \textit{bis}-electrophiles. For example, Liu and colleagues detected the formation of 1,2-dibromoethane-mediated cross-links between guanine bases in DNA and the catalytic cysteine residue (Cys$^{\text{145}}$) of human AGT.\textsuperscript{135} More recently, our laboratory used mass spectrometry to characterize AGT-DNA cross-linking by DEB.\textsuperscript{94} Recombinant wild type hAGT and the C145A hAGT mutant were incubated with DEB in the presence of synthetic DNA oligonucleotides, followed by enzymatic digestion with trypsin. HPLC-ESI$^+$-MS/MS analysis of the tryptic peptides revealed the formation of DEB-mediated cross-links to guanine at Cys$^{\text{145}}$ and a nearby active site cysteine residue (Cys$^{\text{150}}$). The chemical structure of the cross-link was established as Cys-N7G-BD. More recently, our laboratory has demonstrated that AGT can participate in DEB-induced AGT-DNA cross-linking in the presence of other cellular proteins (see Chapter II).

In theory, AGT-DNA cross-linking can be achieved by initial DEB alkylation of either the DNA or the protein to form HEB intermediates that can subsequently react with the second biomolecule to form the cross-link (\textit{Scheme 4.2}). Kalapila \textit{et. al.} investigated the sequential order of reactivity of the DEB, protein, and DNA using a gel shift assay.\textsuperscript{230} Two of the three components (either DEB with protein or DEB with $^{35}$S-labelled duplex oligodeoxynucleotides) were preincubated for different lengths of time prior to the addition of the third component. DPCs
Scheme 4.2  Formation of DPCs by DEB
were formed regardless of the order of component addition, indicating that both protein-DEB and oligonucleotide-DEB intermediates are formed. Greater reactivity was observed when the DNA was preincubated with DEB prior to addition of the protein, indicating that the oligonucleotide-DEB intermediate is more stable than the corresponding protein-DEB intermediate. These findings suggest that the primary pathway for DEB-mediated DPC formation involves initial alkylation of the DNA, followed by nucleophilic attack by the Cys^{145} or Cys^{150} residues of AGT protein. Interestingly, when wild-type hAGT was preincubated with DEB prior to addition of DNA, DPC formation decreased significantly with increasing preincubation time. In contrast, preincubation of the C145S AGT mutant with DEB prior to addition of DNA yielded only a slight decrease in cross-link formation, implying that the HEB intermediate at Cys^{150} is more stable and less susceptible to hydrolysis than the corresponding adduct Cys^{145}.

Consistent with its ability to form toxic DPC lesions, AGT expression in bacteria increases the cytotoxicity and mutagenicity of DEB. Studies with wild type and mutant AGT have shown that DEB-induced mutations and cytotoxicity were greatly increased by the presence of wild type hAGT, and increased somewhat by the presence of the C145A active site mutant. Both of these protein forms are capable of forming DEB-mediated DPCs in vitro. In contrast, the presence of the C145A/C150S double mutant hAGT protein, which is capable of binding DNA but lacks the cysteine residues required for DPC formation, had no impact on the genotoxic effects of DEB exposure. Taken together, these results suggest that the increases in cell death and mutations in cells expressing wild type and C145A AGT are the result of covalent DPC formation, rather than non-covalent DNA-protein interactions. These findings are further supported by the lack of effect in bacteria expressing R128A and R128A/C145A AGT mutant proteins, which contain at least one of the cysteine residues required for DPC formation but lack DNA-binding ability. Taken together, these findings provide evidence in support of the ability of DEB to form cytotoxic and mutagenic DPC lesions involving AGT in vivo.
The goal of the present work was to analyze the effects of DPCs on cell survival and mutagenicity. Based on the known ability of AGT to form toxic DPC lesions,\textsuperscript{113,230} we selected this protein as a basis for designing a protein reagent capable of selectively inducing DPCs in living cells. Analogous experiments were also conducted using BSA protein. Upon introduction of DEB-induced protein monoepoxides into cultured cells, the highly reactive HEB group is capable of alkylating chromosomal DNA to form DPCs (Scheme 4.3). Unlike their DEB precursor, these protein derivatives are incapable of producing other types of DNA adducts. This strategy enabled the specific induction of DPCs into cultured mammalian cells and permitted analysis of the effects of DPCs in the absence of other types of DNA damage. These results are significant because they provide direct evidence for the ability of DPC lesions to exert cytotoxic effects on mammalian cells.

4.2 Materials and Methods

\textit{Safety Statement} - DEB is a known human carcinogen and should be handled with caution in a well-ventilated fume hood with appropriate personal protective equipment.

\textit{Chemicals and Reagents} – Racemic DEB and BSA were purchased from Sigma (St. Louis, MO), and mass spectrometry-grade Trypsin Gold was purchased from Promega (Madison, WI). Adenosine 5’-[\gamma^{32}\text{P}]triphosphate was obtained from Perkin-Elmer (Boston, MA), and T4 polynucleotide kinase was purchased from New England Biolabs (Beverly, MA). Synthetic DNA oligodeoxynucleotides were prepared at the University of Minnesota’s Biomedical Genomics Center (Minneapolis, MN). Recombinant C-terminal histidine-tagged wild type hAGT and an N-terminal histidine-tagged C145A mutant hAGT were a generous gift from Prof. Anthony E. Pegg at Pennsylvania State University (Hershey, PA).
Scheme 4.3 Strategy for the selective induction of DPCs in mammalian cells

**NEGATIVE CONTROL EXPERIMENTS**

<table>
<thead>
<tr>
<th>Protein</th>
<th>1,2,3,4-Diepoxynbutane (DEB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT or BSA</td>
<td>OR</td>
</tr>
</tbody>
</table>

**Scheme Details**

1. Monoepoxide-containing protein
   - AGT or BSA
2. **PROTEIN**
3. Size Exclusion (To remove unreacted DEB)
4. Electroporate Cells
5. **DPC lesion formation**
6. Determine Cytotoxicity
7. **PROTEIN**
8. Size Exclusion
9. Electroporate Cells
10. **No DPCs formed**
Cell Culture – HT1080 human fibrosarcoma cells\textsuperscript{201} were obtained from the American Type Culture Collection. The cells were maintained as exponentially growing monolayer cultures in DMEM supplemented with 9% FBS. Chinese hamster lung (CHL) cell lines (V-79 – wild type parental cell line; V-B11 – UV sensitive/NER deficient; V-H4 – Fanconi anemia-like/sensitive to bifunctional alkylating agents) were purchased from the Coriell Institute for Medical Research (Camden, NJ), and were maintained as exponentially growing monolayers in Ham’s F-12 media supplemented with 10% FBS. All cell lines were maintained in a humidified incubator at 37°C with 5% CO\textsubscript{2}.

DEB-Induced Alkylation of AGT Protein for Mass Spectrometric Analysis – Recombinant wild type hAGT (62.5 µg, 2.85 nmol) or C145A hAGT protein (61.5 µg, 2.67 nmol) was incubated with 25 molar equivalents of racemic DEB in 10 mM Tris-HCl, pH 7.4 for 2 h at 37ºC (150 µL total volume). To remove unreacted DEB, proteins were acidified to by addition of formic acid to a final concentration of 1% and subjected to size exclusion chromatography using Micro Bio-Spin 6 columns (Bio-Rad, Hercules, CA) in which the buffer was exchanged to 0.05% formic acid following manufacturers instructions. Alternatively, unreacted DEB was removed by HPLC using an Agilent 1100 HPLC system. An Agilent Zorbax 300 SB-C3 column (2.1 x 150 mm, 5 µm) was eluted with 0.05% TFA in water (A) and 0.05% TFA in acetonitrile (B) at a flow rate of 0.2 mL/min. The solvent composition began at 30% B and was increased to 80% B over 30 min. Under these conditions, both native and modified AGT proteins eluted as a single peak at ~17 min, and the acetonitrile was evaporated under vacuum. To assess epoxide stability, DEB-modified protein was incubated at 37ºC, and aliquots corresponding to 0.5 nmol protein were removed immediately and after 1 h, 2 h, 4 h, and 24 h, and frozen at -20ºC pending HPLC-ESI\textsuperscript{+}MS analysis as described below.
To identify the DEB-alkylation sites within the AGT protein and to investigate the fate of the epoxide functionalities, recombinant hAGT (62.5 μg, 2.85 nmol) or C145A hAGT (61.5 μg, 2.67 nmol) was treated with 25 molar equivalents of racemic DEB at 37 °C in 150 μL 10 mM Tris-HCl, pH 7.5. Aliquots (50 μL) were following 1 h, 2 h, or 3 h incubation and diluted to 200 μL by the addition of water and 100 mM ammonium bicarbonate pH 7.9 (final concentration 25 mM). Trypsin (2.0 μg) was added to initiate proteolytic digestion, and the samples were allowed to digest overnight at 37ºC. The tryptic digests were dried to completion under vacuum, resuspended in 25 μL aqueous 0.5% formic acid/0.01% TFA, and subjected to HPLC-ESI+-MS/MS analysis as described below.

**HPLC-ESI+-MS Analysis of AGT Monoepoxide** – All mass spectrometric analyses were performed with an Agilent 1100 capillary HPLC-ion-trap MS system operated in ESI+ mode (m/z 200-2000). For whole protein analysis, chromatography was achieved using an Agilent Zorbax Extend SB 300-C8 column (150 x 0.3 mm, 3.5 μm), eluted at a flow rate of 8 μL/min with a mobile phase of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The gradient program was held at 30% B for the first 5 min, followed by a linear increase to 80% B over 25 min, and further to 95% B in 5 min. Using these conditions, the C145A hAGT protein (both native and alkylated) eluted at ~14.5 min. Deconvolution of the protein charge envelope was performed using the deconvolution software on the Agilent ion trap.

For analysis of tryptic peptides, an Agilent Zorbax SB-C18 column (150 x 0.5 mm, 5 μm) was eluted with 0.1% formic acid/0.05% TFA in water (A) and 0.1% formic acid/0.05% TFA in acetonitrile (B) at a flow rate of 15 μL/min. The gradient program was initially held at 3% B for 3 min, followed by a linear increase to 5% B in 7 min. The solvent composition was held at 5% B for 10 min, followed by a linear increase to 35% B in 95 min, and further to 75% B in 10 min. Auto MS² was employed to select and fragment the doubly-charged ions at m/z 658.4
(unmodified peptide G^{136}NPV{\text{PILCHR}}^{147}), m/z 701.4 (HEB monoadduct on G^{136}NPV{\text{PILCHR}}^{147}), m/z 710.4 (THB monoadduct on G^{136}NPV{\text{PILCHR}}^{147}), m/z 834.4 (unmodified peptide V^{148}VCSSGAVGNYS{\text{GGLAVK}}^{165}), m/z 877.4 (HEB monoadduct on V^{148}VCSSGAVGNYS{\text{GGLAVK}}^{165}), and m/z 886.4 (THB monooadduct on V^{148}VCSSGAVGNYS{\text{GGLAVK}}^{165}).

**Gel Shift Assay** – In order to analyze the time course for DPC formation by C145A AGT monooepoxide *in vitro*, aliquots of C145A hAGT (4 μg, 0.17 nmol) were incubated in the presence of 25 mM DEB in 10 mM Tris-HCl, pH 7.4 (20 μL total volume) for various periods of time (15, 30, or 60 min) at 37ºC, and unreacted DEB was removed by size exclusion as described above. DEB-modified protein samples were combined with ^{32}P-endlabelled double-stranded oligodeoxynucleotides (5’-GGA GCT GGT GGC GTA GGC-3’ + strand, 40 pmol), and the cross-linking reaction proceeded for 2 h at 37ºC. As a negative control experiment to test the effectiveness of our size exclusion protocol for removal of DEB, a protein-free 25 mM DEB sample was incubated for 15 min at 37ºC, and half of the sample was subjected to size exclusion. Both samples were then incubated for 2 h at 37ºC in the presence of C145A hAGT (4 μg) and duplex ^{32}P-labeled oligodeoxynucleotides (40 pmol). Additional controls consisted of ^{32}P-labeled oligonucleotide alone and AGT incubated with ^{32}P-labeled oligonucleotide in the absence of DEB. All cross-linking reactions were conducted in a final volume of 30 μL in 10 mM Tris-HCl, pH 7.4 buffer. Following cross-linking incubations, samples were diluted to 45 μL with water, and 15 μL 4X NuPAGE buffer was added. Samples were heated for 15 min at 90ºC prior to separation by 12% SDS-PAGE, and radiolabeled products were visualized using a Packard Cyclone Phosphoimager (Packard BioScience, Meridan, CT). Relative quantitation of cross-linking efficiency was accomplished using the ImageJ software available free of charge from the NIH website (www.ncbi.nlm.nih.gov). 113
To analyze the concentration dependence for DPC formation, Cys^{150} AGT monooepoxide was prepared by incubating C145A hAGT protein (4 μg, 0.17 nmol) in the presence of 30, 100, 300, 1000, or 3000 molar equivalents of DEB for 1 h at 37ºC, and unreacted DEB was removed by size exclusion. ^{32}P-endlabeled oligodeoxynucleotides (40 pmol) was added, and cross-linking reactions and SDS-PAGE analysis proceeded as described above.

**Cytotoxicity Experiments** – For cytotoxicity experiments, Cys^{150} AGT monooepoxide was prepared by incubating C145A hAGT (110 μg, 4.8 nmol) with 25 molar equivalents of racemic DEB (120 nmol) in 10 mM Tris-HCl, pH 7.4 for 2 h at 37ºC. As negative controls, C145A hAGT (4.8 nmol) was incubated in the absence of DEB, or DEB (120 nmol) was incubated in the absence of protein. The total reaction volume of all samples was 220 μL. Following incubation, samples were acidified and subjected to size exclusion to remove unreacted DEB as described above. A small aliquot of protein (5 μg) was removed before and after size exclusion for a protein assay (Bio-Rad Protein Assay, Bio-Rad, Hercules, CA), to determine protein recovery. Typical recovery was >75%.

To introduce the DEB-alkylated C145A AGT protein into cells, HT1080 cells or CHL cells (~10^7) were suspended in 500 μL of ice-cold serum-free media (DMEM or Ham’s F-12, respectively) and electroporated using a BTX ECM 630 instrument in the presence of 2 μM Cys^{150} AGT monoepoxide. As internal controls, cells were electroporated in the presence of unalkylated AGT or in buffer containing residual DEB remaining after size exclusion. Immediately following electroporation, cells were plated in pre-warmed serum-containing media and allowed to recover for 24 h before being harvested and counted in a haemocytometer. Cell counts for individual experiments were the average of two independent counts. Cell survival was expressed as the number of cells surviving electroporation in the presence of Cys^{150} AGT monoepoxide compared to the number of cells surviving electroporation in the presence of
unalkylated C145A AGT protein or a buffer-only control containing any residual DEB which remained after size exclusion.

Analogous experiments were conducted with DEB-alkylated BSA. BSA monoepoxide was prepared by incubating BSA (1.25 mg, 18.9 nmol) with 130 molar equivalents of racemic DEB (2.5 mmol) in 10 mM Tris-HCl, pH 7.4 for 2 h at 37°C (total volume 50 µL). As a negative control, DEB (2.5 mmol) was incubated under identical conditions in the absence of protein. Following incubation, the samples were acidified and subjected to size exclusion to remove unreacted DEB as described above. Typical protein recovery from size exclusion was ~60%. HT1080 cells were electroporated in the presence of 10 µM BSA monoepoxide, plated, and counted as described above. Percent survival was expressed as the number of cells surviving electroporation in the presence of BSA monoepoxide compared to the number of cells surviving electroporation in the presence of a buffer-only control containing residual DEB remaining after size exclusion.

**Mutagenicity Assay** – The effect of Cys^{150} AGT monoepoxide on mutation frequency at the hypoxanthine guanine phosphoribosyltransferase (hprt) gene was determined via selection of HPRT mutants. HT1080 cells were electroporated (in duplicate) in the presence of Cys^{150} AGT monoepoxide (5 µM) or buffer containing residual DEB remaining after size exclusion and assayed for cell survival as described above. The cells were then re-plated and permitted to recover for 8 days at 37°C. Following trypsinization, 1.5 x 10^6 cells were plated onto 15 cm dishes (3 dishes plated per electroporation) in DMEM containing 9% FBS, and placed overnight at 37°C. The following day, the cell cultures were placed in media containing 1 mM 6-mercaptopurine (6-MP) and placed in the 37°C incubator for 13-15 days to allow for the formation of colonies. Mutation frequency was expressed as the number of 6-MP resistant colonies per 10^6 surviving cells.
4.3 Results

4.3.1 Strategy for the Selective Induction of DPCs in Cell Cultures

To address the challenges associated with studying the biological consequences of DPC lesions in vivo, we have developed a novel strategy for the selective induction of DPCs into living cells (Scheme 4.3). In this approach, proteins are incubated with a large excess of DEB to generate HEB “warheads” at cysteine residues. After removing any unreacted DEB by size exclusion, these protein reagents are electroporated into cultured mammalian cells, where they can react with chromosomal DNA to form DPCs. Importantly, unlike DEB and other bis-electrophiles, these protein reagents are unable to produce other types of DNA damage which could contribute to cytotoxicity and mutagenicity (Scheme 4.1). This approach enables the analysis of biological outcomes associated specifically with DPCs.

4.3.2 HPLC-ESI+-MS Analysis of AGT Monoepoxide Formation and Stability

AGT monoepoxide was prepared by incubating recombinant human wild type or C145A AGT protein with excess DEB for 2 h at 37°C. Any unreacted DEB was removed by size exclusion or HPLC, and the alkylated protein was analyzed by HPLC-ESI+-MS using an ion trap mass spectrometer. As shown in Figure 4.1 (panel A), HPLC-ESI+-MS analysis of wild type AGT protein following incubation with excess DEB revealed a single protein peak at 16.6 min with an ESI+ spectrum containing m/z signals corresponding to the protein’s multiple charge states (+20 – +27). Deconvolution of this spectrum yielded protein masses consistent with the calculated molecular weight of both wild type hAGT (calculated \( M = 21,876 \) Da; observed \( M = 21,879 \) Da) and protein containing an HEB monoadduct (calculated \( M = 21,962 \) Da; observed \( M = 21,966 \) Da). When analogous experiments were done with AGT monoepoxide prepared from C145A AGT, both the C145A hAGT protein (calculated \( M = 23,016 \) Da; observed \( M = 23,019 \) Da) and
Figure 4.1 HPLC-ESI^-MS analysis of wild type AGT monoepoxide. AGT monoepoxide was prepared by incubating recombinant hAGT with 25 molar equivalents DEB for 2 h at 37º. Unreacted DEB was removed by HPLC. (A) HPLC-ESI^-MS of AGT monoepoxide immediately after formation and (B) following a 2 h incubation under physiological conditions. (Top panel) Total ion chromatogram; (Bottom panel) ESI^- mass spectrum of the 16.6 min protein peak; (Inset) Deconvoluted mass spectrum of the 16.6 min peak. The calculated masses of unmodified hAGT protein and hAGT protein containing a single HEB monoadduct are 21,876 and 21,962 Da, respectively.
protein containing the HEB monoadduct (calculated $M = 23,102$ Da; observed $M = 23,105$ Da) were observed (Figure 4.2, panel A). Additional HEB lesions resulted in subsequent mass increases of ~86 Da, corresponding protein alkylation by multiple DEB molecules.

To determine the stability of DEB-alkylated AGT protein, a time course experiment was conducted in which AGT monoepoxide was prepared from wild type and C145A hAGT protein as described above. After removing unreacted DEB, the protein samples were placed at 37°C, and aliquots were removed immediately and following incubation periods of 1 h, 2 h, 3 h, and 4 h, and frozen pending mass spectrometric analysis. Comparison of the HPLC-ESI+-MS spectra immediately after formation of the Cys150 AGT monoepoxide (Figure 4.2, panel A) and following a 4 h incubation period under physiological conditions (Figure 4.2, panel B) revealed that the epoxide-containing protein is hydrolytically stable. The wild type AGT monoepoxide displayed no evidence of hydrolysis after 2 h (Figure 4.1, panel B). However, the significant decrease in signal intensity following the 2 h incubation (Figure 4.1) suggests that this form of the protein is more susceptible to precipitation than the C145A mutant protein. These results are consistent with previous studies which suggest that alkylation at Cys145 destabilizes the AGT protein.128

4.3.3 Peptide Mapping by HPLC-ESI+-MS/MS

Previous studies by Kalapila and colleagues suggested that alkylation by DEB at the Cys145 position results in an intermediate that is readily susceptible to hydrolysis, whereas DEB alkylation at the neighboring Cys150 is considerably more stable in an aqueous environment.230 To further investigate this possibility, AGT monoepoxide was prepared by incubating recombinant wild type hAGT with excess DEB. Aliquots were removed at different time points and digested with trypsin prior to HPLC-ESI+-MS/MS analysis of the resulting peptides. Previous peptide mapping experiments in our laboratory revealed two sites within the AGT protein, the catalytic
Figure 4.2  HPLC-ESI$^+$-MS analysis of Cys$^{150}$ AGT monoepoxide. Cys$^{150}$ AGT monoepoxide was prepared by incubating recombinant C145A hAGT with 25 molar equivalents DEB for 2 h at 37º. Unreacted DEB was removed by HPLC. (A) HPLC-ESI$^+$-MS of Cys$^{150}$ AGT monoepoxide immediately after formation and (B) following a 4 h incubation under physiological conditions. (Top pane) Total ion chromatogram; (Bottom panel) ESI$^+$ mass spectrum of the 14.5 min protein peak; (Inset) Deconvoluted mass spectrum of the 14.5 min peak. The calculated masses of unmodified C145A hAGT protein and C145A hAGT protein containing a single HEB monoadduct are 23,016 and 23102 Da, respectively.
Cys\textsuperscript{145} residue and the nearby active site Cys\textsuperscript{150} residue, that were capable of forming cross-links to DNA \textit{in vitro} in the presence of DEB.\textsuperscript{94}

The ion trap mass spectrometer was operated in ESI\textsuperscript{+} mode, and auto MS\textsuperscript{2} was used to select and fragment doubly charged ions corresponding to the \textit{m/z} of the unmodified G\textsuperscript{136}NPVPILIPCHR\textsuperscript{147} and V\textsuperscript{148}VCSSGAVGNYSGLAVK\textsuperscript{165} peptides, as well as the \textit{m/z} corresponding to both peptides containing the DEB-induced HEB and THB adducts. Following a 1 h treatment with DEB, HPLC-ESI\textsuperscript{+}-MS/MS analysis of the tryptic digest of hAGT revealed the presence of the V\textsuperscript{148}VCSSGAVGNYSGLAVK\textsuperscript{165} peptide containing a single HEB adduct (\textit{m/z} 877.1 [M+2H]\textsuperscript{2+}, calculated \(M = 1752.2, \Delta M = +86\)) (\textit{Figure 4.3}, panel A). CID of this peptide in an ion trap mass spectrometer yielded an MS/MS spectrum showing b- and y-series ions consistent with the presence of an HEB adduct on one of the three N-terminal amino acids. The masses of the \(y_3 - y_{10}\) ions were consistent with the theoretical values expected for the unmodified peptide, whereas the masses of the \(b_3\) and \(b_5 - b_9\) ions exhibited a mass increase of +86 Da, consistent with alkylation near the N-terminus (VVC\textsuperscript{*}). Previous studies of AGT-DNA cross-linking by DEB suggest that alkylation occurs at the Cys\textsuperscript{150} position.\textsuperscript{94} This HEB-containing peptide was also detected following a 3 h treatment with DEB prior to tryptic digestion (results not shown). In contrast, the only HPLC-ESI\textsuperscript{+}-MS/MS signal detected for DEB-modified G\textsuperscript{136}NPVPILIPCHR\textsuperscript{147} peptide immediately following a 1 h treatment contained the hydrolyzed THB adduct (\textit{Figure 4.3}, panel B). These data are consistent with the findings of Kalapila \textit{et al.} that HEB adducts at Cys\textsuperscript{150} is significantly more stable to hydrolysis in an aqueous environment than those at Cys\textsuperscript{145}.\textsuperscript{230}

In analogous experiments, Cys\textsuperscript{150} AGT monoepoxide was prepared by incubation of the C145A hAGT protein with DEB for 2 h and subjected to tryptic digestion and HPLC-ESI\textsuperscript{+}-MS/MS analysis as described above. As expected, the V\textsuperscript{148}VCSSGAVGNYSGLAVK\textsuperscript{165} peptide containing an HEB adduct on one of the three N-terminal residues was observed (\textit{Figure 4.4}, 120
Figure 4.3 HPLC-ESI\(^+\)-MS/MS analysis of tryptic peptides from wild type AGT monoepoxide. AGT monoepoxide was prepared by incubating recombinant hAGT with 25 molar equivalents DEB for 2 h at 37º. (A). HPLC-ESI\(^+\)-MS/MS analysis of hAGT tryptic peptide V\(^{148}\)C\(^{50}\)G\(^{51}\)N\(^{52}\)S\(^{53}\)Y\(^{54}\)S\(^{55}\)G\(^{56}\)L\(^{57}\)G\(^{58}\)L\(^{59}\)A\(^{60}\)V\(^{61}\)K\(^{62}\) containing a HEB monoadduct on Cys\(^{150}\) following treatment of the wild type protein with excess DEB. (top panel) Extracted ion chromatogram of \(m/z\ 877.1\ [M+2H]^2\); (bottom panel) MS/MS spectrum, with modified fragment ions indicated by “*”.

(B). HPLC-ESI\(^+\)-MS/MS analysis of hAGT tryptic peptide G\(^{136}\)N\(^{137}\)P\(^{138}\)V\(^{139}\)P\(^{140}\)I\(^{141}\)P\(^{142}\)C\(^{143}\)H\(^{144}\) containing a THB monoadduct on Cys\(^{145}\) following treatment of the wild type protein with excess DEB. (top panel) Extracted ion chromatogram of \(m/z\ 710.3\ [M+2H]^2\); (bottom panel) MS/MS spectrum, with modified fragment ions indicated by “*”.
**Figure 4.4** HPLC-ESI-MS/MS analysis of tryptic peptides from Cys\(^{150}\) AGT monoepoxide. Cys\(^{150}\) AGT monoepoxide was prepared by incubating recombinant C145A hAGT with 25 molar equivalents DEB for 2 h at 37º. (A). HPLC-ESI-MS/MS analysis of tryptic peptide V\(^{148}\)VCSSGAVGNYSGGLAVK\(^{165}\) containing a HEB monoadduct on Cys\(^{150}\) following treatment of the C145A protein with excess DEB. (top panel) Extracted ion chromatogram of \(m/z\) 877.1 [M+2H]\(^{2+}\); (bottom panel) MS/MS spectrum, with modified fragment ions indicated by “*”.

(B). HPLC-ESI-MS/MS analysis of C145A hAGT tryptic peptide V\(^{148}\)VCSSGAVGNYSGGLAVK\(^{165}\) containing a THB monoadduct on Cys\(^{150}\) following treatment of the C145A protein with excess DEB. (top panel) Extracted ion chromatogram of \(m/z\) 886.6 [M+2H]\(^{2+}\); (bottom panel) MS/MS spectrum, with modified fragment ions indicated by “*”.

---

**A**

![HPLC-ESI-MS/MS analysis of tryptic peptide](image)

**B**

![HPLC-ESI-MS/MS analysis of tryptic peptide](image)
panel A), confirming that DEB reacts with the Cys<sup>150</sup> of C145A hAGT to form a reactive HEB monoadduct. In addition, we observed the doubly charged peptide corresponding to the THB adduct-containing V<sup>148</sup>VCSSGAVGNYSGGLAVK<sup>165</sup> peptide (m/z 886.6 [M+2H]<sup>2+</sup>, calculated M = 1771.2, ΔM = + 86), suggesting that some of the epoxide-containing C145A AGT protein becomes hydrolyzed (Figure 4.4, panel B). Importantly, the masses of the y<sub>3</sub>-y<sub>15</sub> fragment ions observed in the MS/MS spectrum for this peptide were consistent with the theoretical values for the unmodified peptide, while the b<sub>3</sub> and y<sub>16</sub> fragment ions were shifted by +86 Da, providing direct evidence for DEB alkylation at Cys<sup>150</sup>.

4.3.4 SDS-PAGE Analysis of AGT Monoepoxide-Induced AGT-DNA Cross-links

To investigate the ability of the AGT monoepoxide to react with DNA in vitro to form an AGT-DNA cross-link, Cys<sup>150</sup> AGT monoepoxide was prepared by incubating recombinant C145A hAGT with excess DEB for various time periods (15, 30, or 60 min). The C145A AGT mutant protein was selected for these experiments after initial experiments with AGT monoepoxide prepared from wild type hAGT yielded poor results due to the insolubility of the DEB-alkylated wild type protein (results not shown). Following the alkylation reactions, the sample mixtures were subjected to size exclusion chromatography to remove residual unreacted DEB, and the Cys<sup>150</sup> AGT monoepoxide was combined with <sup>32</sup>P-labeled double-stranded oligonucleotides to induce cross-linking. As shown in Figure 4.5 (panel A), SDS-PAGE analysis of the resulting cross-linking reactions provided evidence for the ability of Cys<sup>150</sup> AGT monoepoxide to react with DNA, as indicated by the appearance of a DNA band with reduced mobility in the gel relative to the free DNA oligomer (lanes 5-7 in Figure 4.5, panel A). This slow-moving band corresponds to a covalent AGT-DNA complex, and increased in intensity as a result of longer preincubation periods (Figure 4.5, panel B). Additional higher molecular weight bands observed in samples which were subject to preincubation periods of 30 min or longer...
Figure 4.5 SDS-PAGE analysis of the time course of Cys\textsuperscript{150} AGT monoepoxide formation. (A) SDS-PAGE analysis of \textsuperscript{32}P-endlabeled DNA duplexes (5'-GGA GCT GGT GGC GTA GGC-3' + strand) following incubation with C145A hAGT protein which was preincubated with excess DEB for 15 min (lane 5), 30 min (lane 6), or 60 min (lane 7). Unreacted DEB was removed by size exclusion prior to addition of the DNA. AGT-DNA cross-links appear as slow-moving bands on the gel. Lane 4 serves as a positive control, and lanes 1-3 serve as negative controls. (B) Densitometric analysis of protein bands to determine relative cross-linking efficiency.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Oligo</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>AGT</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>DEB</td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>-</td>
<td>15 min</td>
<td>30 min</td>
<td>1 h</td>
<td>Preincubate AGT and DEB</td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>Size exclusion of DEB of AGT and DEB</td>
</tr>
</tbody>
</table>

![Oligo-AGT complexes]

![Free oligo]

![Graph of Relative Cross-linking Intensity vs. Preincubation Time]

- Lower band: AGT-DNA cross-links
- Upper band: (AGT)\textsubscript{2}-DNA cross-links
lanes 6 and 7) are indicative of ternary covalent DNA-protein-protein complexes, suggesting that long preincubation periods (>30 min) yield greater amounts of Cys\textsuperscript{150} AGT monoepoxide available for reacting with DNA. Indeed, the 18-base pair oligonucleotide duplex employed in these studies contains a total of 13 guanine bases, rendering it likely that several molecules of Cys\textsuperscript{150} AGT monoepoxide can become cross-linked to the same duplex DNA oligomer.

Several control experiments were done to confirm that the shifted bands (lanes 5-7 in Figure 4.5, panel A) were the result of AGT monoepoxide-induced DNA-protein cross-linking. Bands corresponding to covalent AGT-DNA complexes were absent in control samples lacking AGT or DEB (lanes 1 and 2). As a negative control experiment to demonstrate that our size exclusion procedure effectively removes residual unreacted DNA from the sample mixtures, DEB was incubated for 15 min in the absence of protein under otherwise identical conditions as the AGT-DEB samples. Half of the sample was subjected to size exclusion to remove free DEB prior to incubation with \textsuperscript{32}P-labelled DNA oligomer and C145A hAGT protein to induce cross-linking by any remaining DEB. As expected, AGT-DNA complexes were formed when the DEB sample was not passed through size exclusion, indicating that after the preincubation period, free DEB in the sample mixtures is still reactive (lane 4). In contrast, no AGT-DNA complex was observed when residual DEB remaining after size exclusion was incubated with the protein and DNA, demonstrating that size exclusion effectively eliminates free DEB from sample mixtures (lane 3). These findings suggest that any DPCs formed by Cys\textsuperscript{150} AGT monoepoxide that was subjected to size exclusion is the result of the AGT-DEB intermediate rather than free DEB.

To show that DEB alkylation of AGT protein is concentration-dependent, C145A hAGT was incubated for 1 h in the presence of different concentrations of DEB (30, 100, 300, 1000, or 3000 molar equivalents). After removing unreacted DEB by size exclusion, the protein samples were reacted with \textsuperscript{32}P-labelled DNA oligo. As observed in Figure 4.6 (panel A), SDS-PAGE analysis of these reaction mixtures showed evidence for the formation of covalent AGT-DNA
Figure 4.6 SDS-PAGE analysis of the concentration-dependence of Cys\textsuperscript{150} AGT monoepoxide formation. (A) 12\% SDS-PAGE analysis of \textsuperscript{32}P-endlabeled DNA duplexes (5'-GGA GCT GGT GGC GTA GGC-3' + strand) following incubation with C145A hAGT protein which was preincubated for 1 h with 30 (lane 6), 100 (lane 7), 300 (lane 8), 1000 (lane 9), or 3000 molar equivalents DEB (lane 10). Unreacted DEB was removed by size exclusion prior to addition of the DNA. AGT-DNA cross-links appear as slow-moving bands on the gel. Lanes 4 and 5 serves as positive controls, and lanes 1-3 serve as negative controls. (B) Densitometric analysis of protein bands to determine relative cross-linking efficiency.
conjugates (lanes 6-10) that was not present in reaction mixtures lacking DEB or protein (lanes 1-3). Cross-linking was concentration-dependent (Figure 4.6, panel B), with the highest levels of cross-linking observed following preincubation of the protein with 3000 molar equivalents of DEB (lane 10 in Figure 4.6, panel B), suggesting that preincubation of AGT protein with large excesses of DEB yields maximal cross-linking efficiency.

4.3.5 Cytotoxicity of Protein Monoepoxides in Mammalian Cell Cultures

Measuring the cytotoxic effects of AGT monoepoxide-induced DNA-protein cross-linking in cell cultures requires that the epoxide-containing protein is stable enough to tolerate electroporation and retain its ability to react with chromosomal DNA within the nucleus. As described above, our mass spectrometry and SDS-PAGE data indicate that the DEB-alkylated wild type hAGT protein is more susceptible to hydrolysis and precipitation than the AGT monoepoxide prepared from the C145A mutant AGT protein. These results are consistent with previous findings that alkylation at Cys\(^{145}\) destabilizes the protein’s tertiary structure\(^{128}\) and that HEB adducts at Cys\(^{145}\) are more susceptible to hydrolysis than those at Cys\(^{150}\).\(^{230}\) Furthermore, Kalapila et al. concluded that DEB alkylation at Cys\(^{150}\) is limited when Cys\(^{145}\) is available for reaction, due to the lower reactivity of Cys\(^{150}\) compared to Cys\(^{145}\).\(^{230}\) We thus concluded that AGT monoepoxide prepared from the C145A mutant AGT protein would be optimal for electroporation studies, as this form of the protein is capable of forming HEB adducts at the Cys\(^{150}\) position, but lacks the ability to become alkylated at Cys\(^{145}\).

To assess the cytotoxic effects of DPCs formed in the absence of other types of DNA damage, AGT monoepoxide was introduced into HT1080 cells by electroporation, and cell survival was measured by comparing the number of cells surviving electroporation with the Cys\(^{150}\) AGT monoepoxide to the survival of control cells electroporated with either unalkylated
Table 4.1 Cytotoxicity and mutagenicity of Cys\textsuperscript{150} AGT monoepoxide in human fibrosarcoma (HT1080) cells. Cells (~10\textsuperscript{7}) were electroporated (in duplicate) in the presence of Cys\textsuperscript{150} AGT monoepoxide (2 µM), unalkylated protein, or eluate resulting from size exclusion of DEB. Cell survival was expressed as the number of cells surviving electroporation with Cys\textsuperscript{150} AGT monoepoxide relative to controls. Mutation frequency was assessed by measuring cell survival in the presence of 6-MP. The results are statistically significant: \textsuperscript{*}P < 0.01 (t-test); \textsuperscript{†}P < 0.02 (Chi-square test). N, number of independent experiments performed. SEM, standard error of the mean.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>% Cell Death ± SEM\textsuperscript{*}</th>
<th>hprt Mutations per 10\textsuperscript{6} cells\textsuperscript{†}</th>
</tr>
</thead>
<tbody>
<tr>
<td>C145A AGT</td>
<td>No toxicity</td>
<td>Not tested</td>
</tr>
<tr>
<td>Residual DEB (Size Exclusion)</td>
<td>No toxicity</td>
<td>1.0</td>
</tr>
<tr>
<td>Cys\textsuperscript{150} AGT Monoepoxide</td>
<td>20% ± 3 (N=11)</td>
<td>1.8</td>
</tr>
</tbody>
</table>
C145A hAGT protein or the eluate resulting from size exclusion of a DEB-only control. As shown in Table 4.1, the Cys\textsuperscript{150} AGT monoepoxide was cytotoxic to HT1080 cells, with 20% ± 3 cell death following electroporation with 2 µM Cys\textsuperscript{150} AGT monoepoxide, relative to unalkylated C145A AGT or residual DEB controls. These results suggest that C145A AGT protein containing a highly reactive HEB monoadduct can exert strong toxic effects, likely due to reaction with chromosomal DNA to form DPCs. There was no significant difference in cell death when the control was unalkylated protein versus residual DEB (Table 4.1), indicating that the C145A protein is not inherently cytotoxic and that residual DEB remaining after size exclusion does not cause quantifiable levels of cell death.

To provide further evidence for attributing the observed cell death to DPC formation, these experiments were repeated in CHL cell lines. The three CHL cell lines employed differ in their repair capabilities, with V-79 CHL cells representing the wild type phenotype. The V-B11 CHL cells are known to be UV-sensitive due to defects in NER capabilities\textsuperscript{232}, while the V-H4 CHL cells have a Fanconi Anemia complementation group A-like repair deficiency which makes them sensitive to cross-linking agents\textsuperscript{233}. We hypothesized that cells with defects in repair capabilities would exhibit altered sensitivity to the cytotoxic effects of AGT monoepoxide.

The CHL cells were electroporated with Cys\textsuperscript{150} AGT monoepoxide, and cell survival was assessed by comparing cell counts in cells electroporated with alkylated protein to cell counts in cells electroporated with protein-free buffer containing any residual DEB which remained after size exclusion. When wild type V-79 CHL cells were electroporated with Cys\textsuperscript{150} AGT monoepoxide, cell death was 17 ± 3% relative to controls (Figure 4.7). Interestingly, the Fanconi anemia-like V-H4 cells were hypersensitive to death induced by Cys\textsuperscript{150} AGT monoepoxide, with 22 ± 4% cell death relative to controls. These findings are consistent with our hypothesis that the C145A AGT protein containing a highly reactive HEB-monoadduct at Cys\textsuperscript{150} is capable of inducing cell death \textit{via} a DNA damage mechanism. In contrast, the NER-deficient V-B11 cells
Figure 4.7 Cytotoxicity of Cys\textsuperscript{150} AGT monooxepoxide in CHL cells. V-79 (wild-type), V-B11 (NER-deficient), or V-H4 (Fanconi anemia-like) cells (~10\textsuperscript{7}) were electroporated (in duplicate) in the presence of 2 µM Cys\textsuperscript{150} AGT monooxepoxide, immediately plated in pre-warmed media, and allowed to recover overnight before being harvested and counted. Control cells were electroporated in the presence of buffer containing residual DEB remaining after size exclusion. Cell survival was determined by comparing the number of cells surviving electroporation with Cys\textsuperscript{150} AGT monooxepoxide to controls. Error bars represent the standard error from three (V-79) or two (V-B11 and V-H4) independent experiments.
exhibited only 10 ± 1% cell death relative to controls (Figure 4.7), suggesting that these cells are resistant to Cys\textsuperscript{150} AGT monoepoxide-mediated cell death.

To determine whether a monoepoxide-containing protein substrate generated from a protein without known DNA-binding capabilities could also induce DPCs \textit{in vivo}, thereby resulting in cell death, BSA monoepoxide was prepared by incubating BSA protein with excess DEB. BSA contains 35 cysteine residues, 34 of which participate in intramolecular disulfide linkages and are therefore not available for alkylation by DEB.\textsuperscript{234} It likely that the DEB-induced HEB adduct is formed at Cys\textsuperscript{34}, although DEB alkylation at lysine, arginine, and histidine residues is also possible. When HT1080 cells were electroporated with 10 µM BSA monoepoxide, cell survival was 80 ± 3% relative to residual DEB controls (Figure 4.8), suggesting that BSA monoepoxide is indeed capable of inducing cell death, presumably through the formation of covalent BSA-DNA cross-links.

\textbf{4.3.6 Assay for AGT Monoepoxide-Induced Mutagenicity}

In order to assess the mutagenic effects of DNA-protein cross-linking in HT1080 cells, we measured the frequency of AGT monoepoxide-induced mutations within the hypoxanthine guanine phosphoribosyltransferase (\textit{hprt}) gene. HPRT, the gene product of \textit{hprt}, is a nucleotide biosynthetic enzyme that functions in the purine salvage pathway.\textsuperscript{235} Additionally, the HPRT protein can bioactivate purine analogs, such as 6-MP, to form ribonucleotide phosphate derivatives which are toxic if incorporated into chromosomal DNA.\textsuperscript{235,236} Mutations in the \textit{hprt} gene render the HPRT protein nonfunctional, thereby conferring cellular resistance to 6-MP.

To determine the mutagenic effects associated with DNA-protein cross-linking, AGT monoepoxide-induced mutations at the \textit{hprt} gene were measured by analyzing cell survival in the presence of 6-MP. The frequency of \textit{hprt} mutations was increased nearly two-fold in HT1080 cells which were electroporated in the presence of Cys\textsuperscript{150} AGT monoepoxide as compared with
Figure 4.8  Cytotoxicity of BSA monoepoxide in HT1080 cells. Cells (~10^7) were electroporated (in duplicate) in the presence of 10 µM BSA monoepoxide, immediately plated in pre-warmed serum-containing media, and allowed to recover overnight before being harvested and counted. Control cells were electroporated in the presence of buffer containing residual DEB remaining after size exclusion. Cell survival was determined by comparing the number of cells surviving electroporation with BSA monoepoxide to controls. Error bars represent the standard error from twelve independent experiments.
control cells electroporated in the presence of residual DEB which remained after size exclusion (Table 4.1). Statistical analysis revealed this difference to be significant. These results suggest that the DEB-alkylated AGT protein can become covalently cross-linked to chromosomal DNA, and that the resulting DPC lesions induce mutagenesis.

4.4 Discussion

We have developed a novel approach for the selective induction of DPCs in mammalian cell cultures. This approach employs recombinant protein reagents that contain epoxide functionalities capable of reacting with the cell’s chromosomal DNA to form covalent DPC lesions (Scheme 4.3). Protein reagents derived from human C145A AGT protein formed covalent DPCs in vitro (Figures 4.5 and Figure 4.6), and were cytotoxic to human fibrosarcoma HT1080 cells and three CHL fibroblast cell lines with differential repair capabilities (Table 4.1 and Figure 4.7). Furthermore, Cys\textsuperscript{150} AGT monoepoxide induced inactivating mutations within the hprt gene in HT1080 cells (Table 4.1). Taken together, these findings suggest that electroporation of cells with epoxide-containing proteins results in the formation of covalent DPC lesions capable of causing cell death and mutagenic effects.

CHL cells with deficiencies in DNA repair capabilities displayed altered sensitivity to electroporation with Cys\textsuperscript{150} AGT monoepoxide relative to the parental wild type CHL cells (Figure 4.7). The V-H4 mutant cell line is homologous to Fanconi anemia complementation group A cells.\textsuperscript{233} Fanconi anemia is a recessive genetic disorder characterized by bone marrow failure which may result in leukemia and a variety of other conditions.\textsuperscript{237} These cells are hypersensitive to cell death mediated by agents which cause interstrand DNA-DNA cross-linking due to their inability to repair this type of DNA damage.\textsuperscript{237} Although the exact repair mechanism which is non-functional in Fanconi anemia cells is unclear, it is logical to conclude that a defect
in this repair mechanism may also confer hypersensitivity to DPC-inducing agents, such as that observed for Cys\textsuperscript{150} AGT monoepoxide (Figure 4.7).

In contrast with the cytotoxicity data from V-H4 cells, the V-B11 cells, which contain a defect in the NER complementation group 7 gene,\textsuperscript{232} were resistant to Cys\textsuperscript{150} AGT monoepoxide (Figure 4.7). These results are consistent with an earlier observation that certain NER-deficient cell lines displayed reduced sensitivity another DPC-inducing agent, aza-dC.\textsuperscript{69} One explanation for the decreased sensitivity of NER-deficient cell lines towards DPC-forming agents is that NER-mediated repair of DPC lesions may generate lethal mutations \textit{via} an error-prone repair pathway. For example, many cancer cell lines over-express DNA pol β, which can compete with error-free NER-mediated by pol δ and pol ε, decreasing the fidelity of repair synthesis.\textsuperscript{238} Therefore, inactivation of NER may enhance the cell’s ability to process bulky DPC adducts through an alternative repair pathway which is less susceptible to errors. Nakano and colleagues studied DPC repair in bacteria and mammalian cells, and found that the majority of DPCs are removed by recombinational repair, rather than NER.\textsuperscript{69,70} Further studies are needed to more fully establish the mechanisms by which cells tolerate and repair DPC lesions.

It is noteworthy that approximately 20\% of HT1080 cells were killed as a result of electroporation with 10 µM BSA monoepoxide (Figure 4.8), but similar levels of cytotoxicity were achieved with five-fold lower concentrations of epoxide-containing C145A AGT protein (Table 4.1). These findings suggest that BSA monoepoxide is less cytotoxic than its AGT counterpart. These results are not surprising given that AGT, but not BSA, is a DNA repair protein with known DNA-binding capabilities.\textsuperscript{120} Interaction of the monoepoxide-containing protein with chromosomal DNA would place the reactive epoxide in close proximity to DNA, better enabling DPC formation and DPC-mediated cell death. A comprehensive characterization of DPC formation by BSA monoepoxide is required to more fully understand the biological effects of BSA-DNA cross-linking.
Previous attempts to study the biological effects of DNA-protein cross-linking have been hampered by the ability of DPC-inducing agents, such as DEB, to form genotoxic DNA-DNA cross-links and other types of DNA lesions in addition to the DPC lesions of interest (Scheme 4.1). In contrast, the HEB-containing proteins developed in our studies are not capable of inducing any type of DNA damage other than DPCs. The altered sensitivity to Cys^{150} AGT monoepoxide exhibited in CHL cell lines deficient in DNA repair capabilities (Figure 4.7), as well as the monoepoxide-containing protein’s ability to induce mutations in HT1080 cells, supports our hypothesis that the observed cytotoxicity is mediated by alkylation of the chromosomal DNA. This hypothesis would be further supported by direct evidence for the ability of the protein monoepoxides to alkylate chromosomal DNA in vivo. Studies are currently underway to detect specific amino acid-nucleobase conjugates, such as the Cys-N7G-BD conjugates observed in previous studies of AGT-DNA cross-linking, in mammalian cells electroporated with Cys^{150} AGT monoepoxide.

Although our cytotoxicity and mutagenicity data presented above are consistent with the hypothesis that epoxide-containing proteins alkylate chromosomal DNA to form DPC adducts, our experimental design does not preclude the formation of protein-protein and RNA-protein cross-links which could potentially contribute to cell death. Therefore, the cytotoxicity and mutagenicity observed in cells electroporated with BSA or Cys^{150} AGT monoepoxide cannot be directly attributed to DPCs. To test this, future studies will employ DEB-alkylated protein reagents derived from proteins which lack DNA-binding capabilities, such as the R128A mutant AGT protein. R128A AGT monoepoxide is expected to have a diminished ability to interact with and become cross-linked to chromosomal DNA, and will therefore be less cytotoxic and mutagenic to the cells. Even more pronounced effects are anticipated with AGT monoepoxide derived from the K125L mutant AGT, which is excluded from the nucleus. It is expected that
K125L AGT monoepoxide will be unable to accumulate in the nucleus and become cross-linked to chromosomal DNA, and as such will be neither cytotoxic nor mutagenic to the cells.

In conclusion, our study demonstrates that monoepoxide-containing proteins which are capable of inducing DPCs in vitro are cytotoxic and mutagenic to mammalian cells. These protein reagents are incapable of producing any type of DNA damage other than DPCs, suggesting that cell death and mutations resulting from electroporation with these agents are due to the formation of covalent DPC lesions on the chromosomal DNA. These findings are important because they provide the first direct evidence for DPC-mediated cell death. Experiments are currently underway to more fully characterize the mechanisms of formation and repair of DPCs in mammalian cells.
V. SUMMARY AND CONCLUSIONS

The covalent cross-linking of proteins to chromosomal DNA results in the formation of super-bulky, helix-distorting DNA lesions which are likely to block the progression of protein complexes necessary for critical DNA metabolic processes. DPCs can form in cells following exposure to a variety of endogenous and exogenous agents, but their role of DPCs in the observed genotoxic effects of these agents is not as well understood. In the present work, a combination of mass spectrometry, immunological detection methods, and cytotoxicity/mutagenicity assays was employed to study DNA-protein cross-linking in vitro and in cultured mammalian cells by two representative bis-electrophiles, DEB and mechlorethamine.

Previous studies have identified several proteins that participate in DPC formation in the presence of bis-electrophiles, including AGT, GAPDH, and histones H2b and H3. Interestingly, while the cytotoxic and mutagenic effects of DEB are enhanced in bacteria which over-express human AGT protein via formation of covalent AGT-DNA cross-links, no enhanced mutagenesis is observed in cells which over-express GAPDH or histone H3, suggesting that the biological outcomes of DNA-protein cross-linking are influenced by the identity of the cross-linked protein. We therefore sought to identify other nuclear proteins which form DPCs in the presence of DEB. Our laboratory previously developed a novel approach which coupled affinity capture enrichment with mass spectrometry-based proteomics to identify nuclear extract proteins which became cross-linked to biotinylated DNA oligomers in the presence of bis-electrophiles, and employed this approach to examine DPC formation by the antitumor nitrogen mustard mechlorethamine. Using this methodology (Chapter II, Scheme 2.2), we have identified 41 nuclear proteins from human cervical carcinoma (HeLa) cells which formed DPCs in vitro in the presence of DEB (Chapter II, Table 2.1 and Figure 2.5). Relative to their cellular
abundance, a disproportionate number of these proteins were known nucleic acid binding proteins involved in DNA replication, transcriptional regulation, and DNA repair (Chapter II, Figure 2.4), suggesting that proximity to DNA and the ability to interact with DNA contributes significantly to the likelihood of a particular protein becoming cross-linked to DNA. Other factors which may contribute to DPC formation include protein abundance and sub-cellular localization. Importantly, MS/MS analysis of total proteolytic digests of DPCs revealed the presence of Cys-N7G-BD conjugates (Chapter II, Figure 2.7), consistent with our previous findings that DEB-mediated DPC formation involves the N7-position of guanine bases within DNA and cysteine thiols within proteins. All of the proteins identified in our screen contain multiple cysteine residues (Chapter II, Table 2.1). Taken together, the experiments described in Chapter II established that the genotoxic bis-epoxide DEB can induce covalent cross-linking between a number of nuclear proteins and DNA, suggesting that DPCs may contribute to the cytotoxic and mutagenic effects associated with DEB exposure.

In order to characterize DNA-protein cross-linking by bis-electrophiles in vivo, we developed a simple strategy for the isolation of DPC lesions from cells. Our method involves a modified phenol/chloroform extraction protocol for the isolation of chromosomal DNA in the presence of protease inhibitors (Chapter III, Scheme 3.2). DPCs that accumulate on the interface of the organic and aqueous layers are collected and re-extracted to remove non-cross-linked proteins, while RNA is digested with a ribonuclease. This methodology was employed to characterize DNA protein cross-linking in human fibrosarcoma (HT1080) cells following treatment with mechloethamine. Mass spectrometric analysis of proteins released by thermal hydrolysis from the DNA backbone in the form of N7-guanine adducts resulted in the identification of 34 nuclear proteins (Chapter III, Table 3.1). Four additional proteins were identified by western blotting (Chapter III, Figure 3.7). Over two-thirds of the identified proteins were classified in the GO database as nucleic acid-binding proteins (Chapter III, Figure 3.6, 138
panel A), compared to only 10% of total nuclear proteins in human fibroblasts. Additionally, approximately 40% of the proteins are involved in some aspect of transcriptional regulation, with additional proteins playing a role in chromatin organization, DNA damage response, and DNA repair (Chapter III, Figure 3.6, panel C). In contrast, only a small number of proteins were associated with DNA in untreated cells (Chapter III, Figure 3.4, panel A). Taken together, these findings suggest that our extraction procedure is selective for DPCs. Although a large fraction of the proteins are classified as RNA-binding (Chapter III, Figure 3.6, panel B), HPLC analysis of nucleoside digest mixtures exhibited no evidence of RNA contamination (Chapter III, Figure 3.2). Similar results were observed when DPCs were formed using a synthetic DNA duplex containing a biotin tag. These discrepancies can be explained if the proteins listed as associated with RNA have additional cellular roles that involve association with chromosomal DNA, and that these secondary functions are not fully accounted for in the current GO database. Consistent with our previous in vitro analysis, MS/MS analysis of total proteolytic digests of DPCs revealed the concentration-dependent formation of Cys-N7G-EMA conjugates in mechlorethamine-treated HT1080 cells (Chapter III, Figures 3.9 and 3.10). Further experiments are needed to establish that these conjugates are the result of covalent DPCs rather than DNA-GSH cross-links.

In the next part of this work, we sought to establish the biological outcomes of DPC lesions. Although numerous studies have established that bis-electrophiles capable of inducing DPCs exert cytotoxic and mutagenic effects on cells, the contribution of DPCs to the observed effects is difficult to determine because many of these agents produce numerous types of DNA lesions in addition to DPCs (Chapter I, Scheme 1.1). For example, DEB forms large numbers of interstrand and intrastrand DNA-DNA cross-links and DNA monoadducts which are likely to contribute to the genotoxic effects of this compound (Chapter IV, Scheme 4.1). In order to effectively study the biological consequences resulting specifically from DPC lesions, we
sought to develop reagents capable of selectively inducing DPCs in cells without producing other
types of DNA damage. Recombinant C145A hAGT protein was modified with DEB to form an
HEB-monoadduct on Cys\textsuperscript{150} (Chapter IV, Figures 4.2 and 4.4) capable of reacting with DNA \textit{in vitro} to form covalent AGT-DNA cross-links (Chapter IV, Figures 4.5 and 4.6). Introduction of
the Cys\textsuperscript{150} AGT monoepoxide into cultured mammalian cells \textit{via} electroporation resulted in cell
death and mutagenesis (Chapter IV, Table 4.1 and Figure 4.7), presumably through cross-linking
of the AGT protein to the chromosomal DNA. DNA repair-deficient Fanconi anemia-like V-H4
CHL cells were hypersensitive to Cys\textsuperscript{150} AGT monoepoxide relative to the parental wild-type cell
line, whereas NER-deficient V-B11 cells were resistant to Cys\textsuperscript{150} AGT monoepoxide-mediated
 cell death (Chapter IV, Figure 4.7), suggesting that the observed cell death is a result of DPCs
rather than RNA-protein or protein-protein cross-links. Furthermore, these findings indicate that
cellular DNA repair pathways contribute differentially to DPC repair, consistent with the results
of previous findings regarding the roles of NER and HR in cellular repair of DPC lesions.\textsuperscript{59,70,76}
Additionally, HEB-containing BSA protein was also cytotoxic to mammalian cells (Chapter IV,
Figure 4.8), suggesting that DPC-associated toxicity was not limited to the AGT protein.

In conclusion, we have shown that two representative \textit{bis}-alkylating agents, DEB and
mechlorethamine, are capable of forming covalent DPCs, both \textit{in vitro} and in mammalian cell
cultures, which involve a number of cellular proteins encompassing a variety of cellular
functions. Based on our experiments with AGT monoepoxides that induce DPCs upon
introduction in mammalian cells, DPC lesions induce mutations and cell death. Taken together,
our results suggest that bulky DPC lesions contribute to the biological effects of common \textit{bis}-
electrophiles and other cross-linking agents.
VI. FUTURE DIRECTIONS

6.1 Quantitation of DPCs and Other Bifunctional DNA Adducts In Vivo

The vast majority of DPC-inducing agents induce several types of DNA damage in addition to DPCs. For example, mechlorethamine forms \( N-(2\text{-hydroxyethyl})-N-[2-(\text{guan-7-yl})\text{ethyl}]-\text{methylamine} \) monoadducts and DNA-DNA cross-links involving guanine and adenine bases, such as \( N,N\text{-bis}[2-(\text{N7-guanyl})\text{ethyl}]-\text{methylamine} \) and \( N-[2-(\text{N3-adenyl})\text{ethyl}]-N-[2-(\text{N7-guanyl})\text{ethyl}]-\text{methylamine} \). Likewise, DEB forms DNA monoadducts and DNA-DNA cross-links. Many of these cross-links are known to contribute significantly to the cytotoxic and mutagenic effects of DEB and mechlorethamine. In order to fully understand the role of DPCs in the genotoxicity of these \( \text{bis} \)-electrophiles, it is important to devise methodology to establish the relative levels of DNA-DNA cross-links, exocyclic adducts, and DPCs formed in cell cultures or animal tissues following exposure to these agents.

Our laboratory has developed quantitative HPLC-ESI+-MS/MS methods for the analysis of bifunctional DEB-induced DNA adducts, including \( 1,4\text{-bis(\text{guan-7-yl})-2,3-butanediol} \) (\( \text{bis-N7G-BD} \)), \( 1-(\text{guan-7-yl})-4-(\text{aden-1-yl})-2,3\text{-butanediol} \) (N7G-N1A-BD), and \( 1,N^6\text{-}(1\text{-hydroxymethyl-2-hydroxypropan-1,3-diyl})-2'-\text{deoxyadenosine} \) (\( 1,N^6\text{-HMHP-dA} \)) (Scheme 6.1). We will employ this methodology to quantify DEB-induced DPCs in cells relative to other types of DEB-induced DNA adducts. Briefly, cells will be treated with cytotoxic concentrations of DEB, and chromosomal DNA will be isolated by phenol/chloroform extraction and hydrolyzed to release adducts. Isotope dilution HPLC-ESI+-MS/MS will be used to quantify \( \text{bis-N7G-BD} \), N7G-N1A-BD, \( 1,N^6\text{-HMHP-dA} \), and Cys-N7G-BD (Scheme 6.1). Quantitative analysis of Cys-N7G-BD along with other DEB-DNA adducts will provide insight into the relative contribution of DPCs in...
Scheme 6.1 HPLC-ESI$^+$-MS/MS analysis of bifunctional DEB-DNA adducts
DEB-induced DNA damage. Analogous experiments will be conducted to quantify adduct levels in the tissues of laboratory mice and rats following inhalation exposure to BD.

6.2 Identification of Cross-linking Sites within Proteins

Our studies presented in Chapters II and III have identified multiple nuclear proteins that form DPCs in the presence of mechlorethamine and DEB. However, since protein identification was based on the MS/MS spectra of unmodified peptides, the cross-linking sites within each protein have yet to be established. It is therefore necessary to develop methodology for the detection of peptides containing covalent cross-links to DNA nucleobases.

Mass spectrometry has been employed by many groups, including ours, to map modification sites within proteins. One approach to identify the cross-linking sites from in vitro cross-linking reactions would be to conduct MS/MS sequencing of cross-linked peptides isolated via affinity capture (Scheme 6.2). Following incubation of nuclear extract proteins with biotinylated DNA duplexed in the presence of cross-linking agents, proteins will be subjected to tryptic digestion followed by affinity capture of DNA containing cross-linked peptides on streptavidin beads. After eluting the DNA and DNA-peptide conjugates from the beads by heating under acidic conditions, the peptides will be released from the DNA backbone by thermal hydrolysis, yielding peptides containing cross-links to nucleobases. The methodology will be modified for cross-linking agents, such as cisplatin, that form thermally stable DNA adducts. In this case, DNA will be enzymatically digested to nucleosides following affinity capture. This approach allows for selective enrichment of the cross-linked peptides, which can then be sequenced by MS/MS. This will enable protein identification in addition to assigning the exact cross-linking sites within each protein.
Scheme 6.2 Experimental scheme for biotin capture enrichment of modified peptides from nuclear extract proteins cross-linked to DNA in the presence of bis-electrophiles.
6.3 Quantitative Proteomic Analysis of DPCs in Nuclear Protein Extracts and Cell Cultures

Our proteomics studies have identified multiple proteins that participate in DPC formation in human fibrosarcoma cells (Chapter III). HPLC-ESI+-MS/MS analysis of amino acid-nucleobase conjugates (Chapter III, Figures 3.9 and 3.10) can estimate the total number of DPC lesions formed in the entire proteome. However, this approach does not permit quantitative analysis of DPCs involving individual proteins. Because the biological outcomes of DNA-protein cross-linking are influenced by the identity of the cross-linked protein, it is important to quantify DPCs involving specific proteins rather than averaging DPC formation over the entire proteome.

The quantitation of individual DPCs in nuclear protein extracts cell cultures could be accomplished by a quantitative proteomics technology such as iTRAQ (isobaric tag for relative and absolute quantitation) (Scheme 6.3). iTRAQ experiments employ a multiplexed set of reagents that place isobaric mass labels on the N-termini and the lysine side chains of peptides present in digest mixtures, such as those resulting from in-gel tryptic digestion of SDS-PAGE separated DPCs. The tagged peptides are isobaric and chromatographically indistinguishable, yet CID in a mass spectrometer yields unique reporter ions. Therefore, peptides from several samples can be pooled together for simultaneous MS/MS analysis (Scheme 6.3). Absolute quantitation of DPCs involving specific proteins can be achieved by spiking peptide sample mixtures with iTRAQ-tagged synthetic peptide standards prior to MS/MS analysis. The use of 4- or 8-plex iTRAQ reagents will enable the relative quantitation of DPCs involving specific target proteins from multiple samples, e.g., concentration dependence of DPC formation by one drug, or comparison of DPC formation involving specific proteins by several different cross-linking agents.
Scheme 6.3 Experimental scheme for a 4-plex iTRAQ quantitative proteomics experiment

---

1. Isolate DNA and DPCCs
2. Digest to Peptides
3. Label peptides with iTRAQ reagents
4. Mix
5. MS
6. MS/MS
7. Reporter-Balance-Peptide INTACT
   All 4 samples have identical m/z

---

Peptide b- and y-series fragments EQUAL

---

Reporter ions DIFFERENT
VII. BIBLIOGRAPHY


141. Melnick, R. L.; Huff, J.; Chou, B. J.; and Miller, R. A. Carcinogenicity of 1,3-butadiene in C57BL/6 x C3H F1 mice at low exposure concentrations. *Cancer Res.* 1990, 50, 6592-6599.


154. Seaton, M. J.; Follansbee, M. H.; and Bond, J. A. Oxidation of 1,2-epoxy-3-butene to 1,2:3,4-diepoxybutane by cDNA-expressed human cytochromes P450 2E1 and 3A4 and human, mouse and rat liver microsomes. *Carcinogenesis* **1995**, *16*, 2287-2293.


161. Goggin, M.; Swenberg, J. A.; Walker, V. E.; and Tretyakova, N. Molecular dosimetry of 1,2,3,4-diepoxybutane-induced DNA-DNA cross-links in B6C3F1 mice and F344 rats exposed to 1,3-butadiene by inhalation. *Cancer Res.* **2009**, *69*, 2479-2486.


