MOLECULAR BINDING OF FORMALDEHYDE TO DNA AND PROTEINS

Kun Lu

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Applied Science and Engineering.

Chapel Hill

2009

Approved by:

Dr. James A Swenberg

Dr. Avram Gold

Dr. Louise M. Ball

Dr. LuChang Qin

Dr. Melvin Andersen

©2009

Kun Lu

ALL RIGHTS RESERVED

ABSTRACT

KUN LU: Molecular Binding of Formaldehyde to DNA and Proteins
(Under the direction of Dr. James A Swenberg)

Formaldehyde is produced worldwide on a large scale (21 million tons in 2000) and used in a wide spectrum of applications. Its toxicity and carcinogenic effects have evoked numerous public health concerns. According to the International Agency on Research on Cancer (IARC), formaldehyde is classified as a known animal and human carcinogen, causing nasal cancer. More limited epidemiologic evidence suggests that formaldehyde can also induce leukemia in humans, however, this is controversial. In this dissertation, we have designed an integrated bottom-up approach to address critical issues to better understand formaldehyde's carcinogenic potential.

Specifically, the N-terminus of histone and lysine residues located in both the histone N-terminal tail and the globular fold domain were identified as binding sites for formaldehyde in the current study. We also found that formaldehyde-induced lysine adducts could inhibit the formation of post translational modifications on histone, raising the possibility that formaldehyde might alter epigenetic regulation. We have also elucidated the structures of DNA-protein cross-links induced by formaldehyde. Detailed characterization of the formaldehyde-derived linkage of single amino acids with nucleosides by NMR and mass

spectrometry established that these amino acids all form cross-links involving formation of a formaldehyde-derived methylene bridge. Our results also demonstrated that Lys-dG cross-links are the most common DNA-protein cross-links induced by formaldehyde, however, they are very labile. The finding that Cys-CH_2-dG cross-links could be initiated by the *S*-hydroxymethyl group of cysteine residue lead to the identification of a novel dG-CH_2-GSH adduct. This adduct is unique because of the involvement of *S*-hydroxymethylglutathione, a key player in the detoxification of formaldehyde. After our extensive work on biomarker discovery and validation involving DNA monoadducts and DNA-DNA cross-links, we applied these methods to analyze DNA samples from rats exposed to [$^{13}CD_2$]-formaldehyde for 1 day and 5 days. The results show that exogenous formaldehyde induced N^2-hydroxymethyl-dG monoadducts and dG-dG cross-links in DNA from rat nasal mucosa, but did not form [$^{13}CD_2$]-adducts in distant tissues despite analyzing 5 times more DNA than for nasal epithelium. These data provide strong evidence supporting a genotoxic and cytotoxic mode of action for inhaled formaldehyde in the target tissue for carcinogenesis, but do not support the biological plausibility that inhaled formaldehyde causes leukemia in rats.

ACKNOWLEDGEMENTS

First of all, I would like to thank Dr Swenberg for his extraordinary support and encouragement. I had to change my lab three years ago because my former advisor took a position in Canada. Dr Swenberg agreed to serve as my advisor, which offered me a valuable opportunity to continue my PhD in his lab. I had to start over my project, but we targeted to get me finished in 3 years. I was anxious about my future in the lab at that time, but I survived eventually due to Dr Swenberg's support and inspiration. I thank Dr Swenberg for allowing me the flexibility to pursue different directions in the frame of our project. I thank Dr Swenberg for his valuable discussion and big vision on my research. I thank Dr Swenberg for everything he has done for me, from which I will certainly benefit a lot in my future career.

Secondly, I would like to acknowledge my committee for their guidance. I thank Dr Qin for agreeing to serve as my committee chair. It has been a hard time for me to get this degree. Whenever I was in trouble, Dr Qin always tried his best to help me out. I would like to specially thank Dr Gold and Dr Ball for their indispensible help on my projects. I also thank Dr Andersen for serving in my committee and his inspiring questions. Also, I offer sincere gratitude to my friends, past and present members of the Swenberg Lab.

Finally, my heartfelt thanks goes to my family, especially my parents and wife, Hongyu, for their supporting me through the ups and downs of my life. Without their encouragement and confidence, I would not have finished this degree in a timely manner under such a challenge. It has been a long journey for me to pursue my PhD. I am happy that I eventually have the chance to acknowledge persons who make this possible.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF SCHEMES

Scheme

LIST OF CHARTS

Chart

ABBREVIATIONS

IARC	International Agency on Research on Cancer
NCI	National Cancer Institute
NIOSH	National Institute for Occupational Safety and Health
FEMA	Federal Emergency Management Agency
CA	Chromosomal aberrations
SCEs	Sister chromatid exchanges
MN	Micronuclei
DPC	DNA-protein cross-links
H4	Histone 4
H3	Histone 3
H2A	Histone 2A
H2B	Histone 2B
PTM	Post-translational modifications
AcetylCoA	Acetyl coenzyme A trilithium salt
HAT	Histone acetylation
MWCO	Molecular weight cutoff
ESI	Electrospray ionization
MS	Mass spectrometry
MALDI	Matrix-Assisted Laser Desorption/Ionization
TOF	Time-of-Flight
NDMA	N-nitrosodimethylamine
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
dA	Deoxyadenosine

dT	Thymidine
dC	Deoxycytidine
dG	Deoxyguanosine
NMR	Nuclear magnetic resonance
LC	Liquid Chromatography
HPLC	High performance liquid chromatography
SIM	Selected ion monitoring
SRM	Selected reaction monitoring
Q-TOF	Quadrupole time-of-flight
FTICR	Fourier-transform ion cyclotron resonance
TPHA-1	2-Amino-6-(10-oxo-triazino[1,2-a]purin-7-yl)hexanoic acid
TPHA-2	2-Amino-6-(5-hydroxymethyl-10-oxo-triazino[1,2-a]purin-7-yl)hexanoic acid
HMBC	Heteronuclear Multiple Bond Coherence
HSQC	Heteronuclear Single Quantum Coherence
COSY	Homonuclear correlation spectroscopy
ROESY	Rotational Frame Nuclear Overhauser Effect Spectroscopy
GSH	Glutathione

CHAPTER 1

INTRODUCTION

Formaldehyde is the simplest aldehyde, with the formula CH_2O. Formaldehyde is produced worldwide on a large scale (21 million tons in 2000) and a chemical with a wide spectrum of applications and important economic impact. However, formaldehyde's toxicity and carcinogenic effects have evoked numerous health concerns. Recently, associations between residents' sickness and formaldehyde in FEMA trailers raised additional public concern for adverse effects of formaldehyde exposure (http://www.cdc.gov/nceh/ehhe/trailerstudy). Formaldehyde is classified as a known animal and human carcinogen according to International Agency on Research on Cancer (IARC), causing nasal cancer *(1)*. IARC also considered there was "*strong but not sufficient evidence for a causal association between leukemia and occupational exposure to formaldehyde*". In this chapter, we will briefly introduce its production and use, occurrence, toxicity and carcinogenesis, with the emphasis on the association between formaldehyde inhalation exposure and the induction of nasal cancer and leukemia. At the end of this chapter, we will raise the critical issues to be addressed in this dissertation to better understand formaldehyde toxicity and carcinogenesis.

1.1 Production and use

Formaldehyde has been produced since 1889 through the catalytic oxidation of methanol. Among many available methods, the silver catalyst process and the metal oxide catalyst process are currently widely used in the formaldehyde industry*(1)*. The commercially available product is primarily 37% formaldehyde solution and its worldwide annual production and consumption is approximately 21 million tons in 2000*(1)*.

Formaldehyde is used in a variety of industries. Formaldehyde is extensively used in the production of resins with urea, phenol and melamine and other resins. These formaldehyde-resins have diverse downstream applications in the production of wood product, pulp and paper, synthetic fiber, plastics, coatings and textile. Formaldehyde is also widely used as an important intermediate chemical for the synthesis of many other compounds. Formaldehyde is also used as a disinfect and to preserve biological samples in medical laboratories and facilities. It is also common to find formaldehyde present in cosmetics products, acting as an antimicrobial agent.

Given the large quantity of formaldehyde production and its wide applications, occupational exposure to formaldehyde occurs in many occupations and related industries. This list includes: manufacture of formaldehyde, formaldehyde-based resins and other chemical products; histopathology and disinfection in hospitals; embalming and anatomy laboratories; manufacture of wood products and paper; building and construction industry; manufacture of textiles and garments; foundries; synthetic fiber production; plastics production; firefighters; automobile and engine

exhausts; offices and public buildings. It is estimated that more than 2 million U.S. workers are occupationally exposed to formaldehyde *(2)*.

1.2 Occurrence

Formaldehyde is a ubiquitous environmental pollutant from many outdoor and indoor sources. Outdoor sources of formaldehyde primarily come from automobile exhaust emissions, plants and manufacturing facilities, combustion, while indoor sources of formaldehyde are primarily released from building materials, consumer products and tobacco smoke. Formaldehyde is also found in various foods, which could be the consequence of both the natural occurrence of formaldehyde and pollution*(3)*. The formaldehyde contamination in foods is emerging as an important issue in some developing countries, which may be another exposure route under certain circumstances.

In addition to the exogenous formaldehyde coming from the environment, formaldehyde is an essential intermediate in all living cells. It is endogenously generated from serine, glycine, methionine and choline and also produced from some metabolites and proteins by demethylation*(1)*. The endogenous concentration of formaldehyde in the blood of human subjects is about 0.1mM*(4;5)*.

Formaldehyde can also be formed in our bodies after metabolizing various compounds, such as nitrosamines*(6-8)*, and certain types of drugs*(9;10)* following metabolism. Therefore, both endogenous and exogenous formaldehyde is present in all human cells.

1.3 Fate and metabolism

As mentioned before, both endogenous and exogenous formaldehyde can enter or form in our body (also shown in Figure 1.1). The major detoxification of formaldehyde is through glutathione-dependent enzyme systems, with *ADH3* as a primary enzyme. Formaldehyde rapidly reacts with glutathione, forming *S*-hydroxymethyl-glutathione. *S*-hydroxymethyl-glutathione is further converted into *S*-formylglutathione and eventually formate, regenerating glutathione. Other minor pathways include the metabolism by *ALDH1A1* and *ALDH2*. In addition to the spontaneous reaction between formaldehyde and glutathione, as well as other minor pathways, formaldehyde induces diverse adducts with DNA and proteins, including protein adducts, DNA-protein cross-links (DPC), DNA adducts and DNA-DNA cross-links. We will discuss all of these formaldehyde-induced damage in the following chapters of this dissertation.

1.4 Nasal cancer and formaldehyde exposure

1.4.1 Animal data

The induction of nasal squamous-cell carcinomas in rats was first found after inhalation of at least 6 ppm formaldehyde for up to two years*(11)*. In the interim report of this study, 120 male and 120 female rats were exposed by inhalation to 0, 2, 5.6, or 14.3 ppm formaldehyde (6 hr/day, 5 days/week) for 18 months of a 24-month study. Squamous cell carcinomas were observed in the nasal cavities of 36 rats

exposed to 14.3 ppm formaldehyde after 18 months. The exposure period was followed by up to 6 months of nonexposure. Squamous cell carcinomas were observed in the nasal cavities of 103 rats (52 females and 51 males), but only in 2 rats (one male and one female) exposed to 5.6 ppm of formaldehyde*(12)*, indicating the incidence of nasal cancer was highly concentration dependent.

1.4.2 Human data

Formaldehyde was classified as a human carcinogen by IARC based on "*sufficient epidemiological evidence that formaldehyde causes nasopharyngeal cancer in humans*", which arises from six major cohort studies of industrial workers and seven case–control studies of nasopharyngeal cancer*(13)*. IARC especially emphasized the statistically significant excess of deaths from nasopharyngeal cancer observed in the National Cancer Institute (NCI) cohort*(14)*. A statistically significant exposure-response relationship for both peak and cumulative exposure was identified in this study. The NCI cohort was the largest cohort study of industrial works and considered most informative by IARC. In addition, an excess of deaths due to nasopharyngeal cancer was also found in a proportionate mortality analysis of the largest U.S. cohort of embalmers*(15)*, and in a Danish study of proportionate cancer incidence among workers*(16)*. Five of seven case-control studies found elevated risk for overall exposure to formaldehyde or in higher exposure categories*(17-24)*. The IARC working group considered that the positive finding in the NCI cohort study of industrial works, coupled with largely positive results from other studies, supported

that formaldehyde caused nasopharyngeal cancer in humans with sufficient epidemiological evidence.

However, the fact that 6 of 10 nasopharyngeal cancer deaths observed in the NCI study occurred in only one plant (Plant 1) and the remaining four cases occurred individually in the other nine plants studied casts a question about the validity of NCI's suggested causal association*(25-29)*. The reanalysis provided little evidence to support NCI's suggestion of a causal association between formaldehyde exposure and mortality from nasopharyngeal cancer*(30)*. NCI's conclusion of a possible causal association was driven heavily by anomalous findings in one study plant (Plant 1). An independent nested case–control study of 7345 workers employed at a plastics-producing plant (Plant 1 in NCI cohort) suggested that the influence of external employment in the ferrous and non-ferrous metal industry, instead of formaldehyde exposure may account for the large nasopharyngeal cancer mortality excess found in Plant 1 in NCI cohort*(31)*. Possible suspected risk factors for nasal cancer in the local area included sulfuric acid mists, mineral acid, metal dusts and heat. Thesse could be responsible for the higher nasopharyngeal cancer mortality and may explain why the NCI cohort had unique pattern nasopharyngeal cancer in Plant 1.

1.5 Leukemia and formaldehyde exposure

1.5.1 Epidemiology data

Compared with the incidence and mortality rates of nasopharyngeal cancer (0.7 and 0.2 per 100,000), leukemia has much higher incidence and mortality rates at

12.3 and 7.5 per 100,000 in U.S. population, respectively*(2)*. Thus, if formaldehyde causes leukemia, the number of cancers caused by formaldehyde should increase due to this more common and lethal cancer. Therefore, epidemiology studies may have better power and higher chance to detect an excess of deaths from leukemia. In the past 5 years, studies on this possible association and the mechanisms underlying this association have attracted more and more attention from academic, industry and regulatory agencies due to the potential significance on this high volume chemical on public health. So far, however, IARC considered the data "*strong but not sufficient evidence for a causal association between leukemia and occupational exposure to formaldehyde*".

The positive association between leukemia and formaldehyde exposure was based on epidemiology studies including case–control study in the general population, proportionate mortality studies of professionals such as embalmers, funeral workers, pathologists and anatomists, and cohort studies of industrial workers *(1)*. Among three major types of studies, cohorts in industrial works have provided the largest sample size and best quality for assessing causation between the induction of leukemia and formaldehyde exposure. Two of the three cohorts in industrial works found positive associations between formaldehyde exposure and leukemia, but one failed to find any association.

The first large cohort in industrial workers was performed in the United Kingdom *(32)*. It included 14,014 workers in chemical factories associated with either use or production of formaldehyde. Comparisons with a local populations, no

association was observed between leukemia and formaldehyde exposure (in either employment in the factories, or work in the high exposure group (> 2ppm)). IARC considered this British cohort a high-quality study due to adequate size and sufficiently long follow-up. Therefore, this study should have reasonable chance to detect an excess of deaths from leukemia if there is any. This study did not examine the response to peak exposure, however, previous studies of garments and professionals in USA found positive findings with the same absence of evaluation of peak exposure. In addition, this study did not examine the specific risk of myeloid leukemia, which was shown to be the primary type of leukemia in other cohort studies.

The second industrial cohort was conducted by National Institute for Occupational Safety and Health (NIOSH) to evaluate the mortality of 11,039 workers exposed to formaldehyde for three months or more in three garment plants *(33)*.The mean time weighted average formaldehyde exposure at the plants in the early 1980s was 0.15 ppm according to an exposure assessment involving 594 randomly selected employees, but past exposures could have been substantially higher. In this study, mortality from myeloid leukemia was greatest among workers first exposed when exposures were presumably higher although the real exposure was unknown. Mortality for myeloid leukemia was also high among workers with 10 or more years of exposure. The result from this cohort supported a possible association between formaldehyde exposure and mortality arising from myeloid leukemia.

The third study was conducted by the NCI and included 25,619 industrial workers *(34)*. They used different exposure metrics including peak exposure, average exposure intensity, cumulative exposure, and duration of exposure. The study identified positive associations between formaldehyde exposure and leukemia overall, but the stronger associations were seen for myeloid leukemia. Also, this strong association was only observed when peak exposure was used as a metric. A recent update of this cohort still found similar positive association between leukemia and formaldehyde exposure *(35)*.

1.6 Mechanistic considerations

1.6.1 Carcinogenesis of formaldehyde

Numerous previous studies show that formaldehyde is genotoxic, inducing both DNA damage such as DNA adducts and DNA-protein crosslinks (DPC) and chromosome changes including chromosomal aberrations (CA), sister chromatid exchanges (SCEs), and micronuclei (MN). These alterations have been demonstrated by a large number of studies using *in vitro* systems or exposed animals, but the results are inconsistent from study to study.*(36-58)*.

Currently available data indicate that both genotoxicity and cytoxicity play important role in formaldehyde's carcinogenesis at the site of contact. Formation of DPC is considered a primary genotoxic effect. Earlier studies of Heck and Casanova demonstrated a linear response between 6, 10 and 15 ppm formaldehyde exposures for DPC, with a reduced number of DPC per ppm formaldehyde at 2 and 0.7 ppm due

to depletion of glutathione in the nasal epithelium at 6 ppm and higher exposures to formaldehyde*(59-61)*. They also demonstrated that there was no accumulation of DPC with exposure, which was attributed to rapid repair of DPC*(62)*.

Marked and sustained increases in cell proliferation in the noses of rats has been observed after exposure to more than 6 ppm (10 ppm or higher) formaldehyde *(63;64)*, which is thought to greatly amplify the genotoxic effects of formaldehyde at the site of contact.

In humans, elevated DPC in peripheral mononuclear cells of formaldehyde-exposed workers was reported *(65;66)*, however, this study has been questioned due to an abnormally high background of DPC in the controls*(67)*. Recently, the Hecht Lab reported clear differences in formaldehyde-DNA adducts between smokers and non-smokers in leukocyte DNA samples*(68)*. This study provided the first evidence of formaldehyde-DNA adducts in humans, however, it may not be directly related to the effects of inhaled formaldehyde, since formaldehyde could be released during metabolism of various compounds present in tobacco.

1.6.2 Leukemia

The limited evidence for causation of leukemia concluded by IARC, is at least, partially based on the inability of the IRAC working group to identify a mechanism for the induction of leukemia in humans *(1;69-71)*. Generally, leukemia arises from damage to early stem or progenitor cells in the bone marrow. Hematotoxicty or

genotoxicity occurs following exposure to all established chemical leukemogens*(72-81)*. All known leukemogens act on the bone marrow, causing marrow toxicity.

Zhang et al. have hypothesized three mechanisms for leukemia*(2;82)*: (1) Targeting bone marrow hematopoietic stem cells. This could be achieved by converting inhaled formaldehyde gas to hydrated methanediol which may reach bone marrow and cause damage; (2) Targeting blood stem cells and progenitors. In this case, formaldehyde needs to enter into the blood, after reaching the nasal/oral passages and lung, and induce mutations or pre-mutagenic lesions in circulating hematopoietic stem cells. Then, the damaged cells are postulated to return to the bone marrow and undergo mutational changes leading to leukemia. (3) Targeting pluripotent nasal/oral stem cells. Under this scenario, formaldehyde directly induces mutations or pre-mutagenic lesions in primitive pluripotent stem cells residing in the oral or nasal passages. Then, the damaged stem cells are released from the nasal passages by either normal trafficking or enhanced trafficking associated with formaldehyde-induced cytotoxicity, circulate through the blood, and migrate to the bone marrow where they divide, develop a mutation and progress toward leukemia.

Although these proposed mechanisms are possible, they remain speculative since there are no data supporting the second and third mechanisms by any known chemicals. Therefore, these mechanisms must be backed up with scientific data. Based on established mechanisms by other known leukemogens and the first proposed mechanism by Zhang et al, formaldehyde needs to travel to the bone marrow to cause damage if it is to cause leukemia. According to the second and third possible

mechanisms of Zhang et al, stem cells damaged in blood or nasal/oral passage need to migrate to bone marrow, undergo cell division and mutate. Under these situations, formaldehyde-induced damage needs to have a relatively long half-life to undergo such processes. In any of the cases mentioned above, there is a critical need to provide mechanistic data on the ability of formaldehyde to induce leukemia.

1.7 Remaining challenges

Overall, after almost 30 years of extensive study on formaldehyde toxicity and carcinogenesis, the results continue to raise additional questions. Understanding these issues is confusing and frustrating, but also offers unique research opportunities. Several of these critical challenges will be addressed step by step in this dissertation.

1.7.1 Challenge 1: To generate structural information on DPC

DPC are considered a primary genotoxic effect and have been used as a biomarker to evaluate formaldehyde exposure in animals. However, before this study, there was little structural information about formaldehyde-induced DPC, even though the formation of DPC had been known for decades. The absence of structural information greatly hinders our understanding on formaldehyde toxicity and carcinogenesis. It is not clear whether DPC are initiated by formaldehyde-induced protein damage or DNA lesions. It is not clear which residues of protein are involved in the formation of DPC. It is not clear which specific DNA bases are cross-linked with protein. It is not clear what the stability of DPC are and what the corresponding

half-lives are. It is not clear if chemical-specific biomarkers based on DPC can be established. All these unknown, yet intriguing questions rely on strict structural elucidation and relevant studies, which will be discussed in Chapters 2 and 3.

1.7.2 Challenge 2: No specific DNA biomarkers are available for formaldehyde inhalation exposure

DPC were used as biomarkers to evaluate formaldehyde exposure in animals and to establish exposure-responses following inhalation exposure. However, all available methodologies on DPC are based on physical chemistry, which use either the different solvbility of DNA, protein and DPC in organic/aqueous phase (Interfacial method used by Heck et al.) or the precipitation of DPC in potassium solution (SDS/K method developed by Zhitkovich et al)*(83;84)*. None of these methods represent a chemical-specific approach that needs to be established using rigorous structure elucidation and unambiguous signal assignment. If additional protein damage or DNA lesions occurred at other positions in addition to cross-linking sites, the general resolution of these approaches does not allow the differentiation between them and DPC. Therefore, the quantitative data based on these non-specific methods are not rigorous and do not represent ideal data sets for use in risk assessment. To date, no formaldehyde DNA adducts have been identified in animals exposed by inhalation. Recently, the Hecht laboratory reported increased amounts of formaldehyde-induced N^6-$HOCH_2$-dA in multiple tissues of rats treated with *N*-nitrosodimethylamine (NDMA) or 4-(methylnitrosamino)-1-(3-pyridyl)-1-

butanone (NNK), which was attributed to the production of formaldehyde during the metabolism of NDMA and NNK . However, it is still not clear whether or not this DNA biomarker could be used to evaluate formaldehyde exposure through inhalation. Therefore, additional biomarker discovery research and validation needs to be done. The identification and application of formaldehyde DNA biomarkers will be found in Chapters 4 and 5.

1.7.3 Challenge 3: Differentiating endogenous and exogenous formaldehyde

As we mentioned before, formaldehyde is a ubiquitous environment pollutant. Of equal importance, formaldehyde is an essential metabolic intermediate in all human cells. The endogenous concentration of formaldehyde in the blood of human subjects is about 0.1mM. Thus, the ability to distinguish DNA damage caused by exogenous formaldehyde from substantial background endogenous formaldehyde-induced lesions is crucial. Accurate molecular dosimetry of both endogenous and inhaled formaldehyde is essential for the science-based risk assessment of formaldehyde exposure. It is also critical to determine the plausibility of inhaled formaldehyde to induce leukemia. Currently, there are no available studies providing molecular dosimetry for both endogenous and exogenous formaldehyde. We will address this paramountly important issue by using highly sensitive mass spectrometry, coupled with the utilization of [$^{13}CD_2$]-formaldehyde in animal exposures. This unique approach allows us to differentiate DNA damage originating from both endogenous and inhaled [$^{13}CD_2$]-formaldehyde and to quantify corresponding DNA

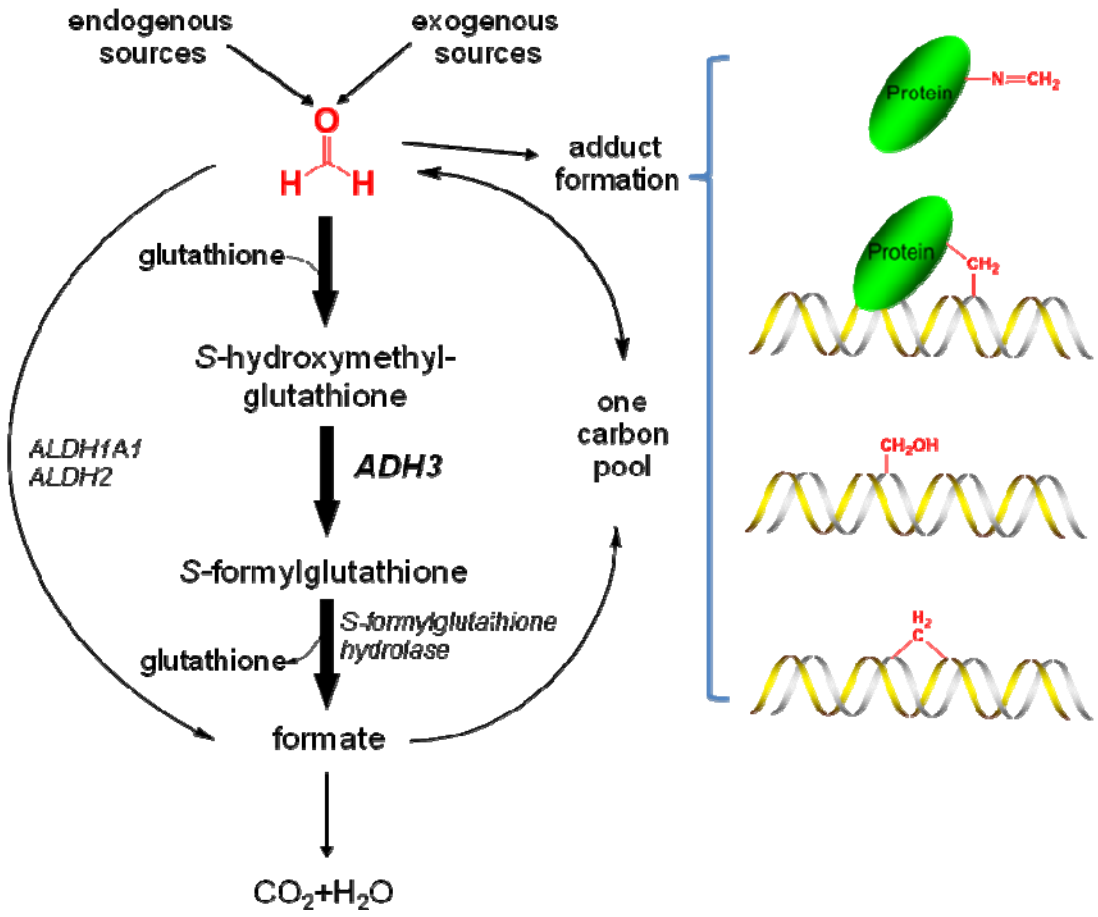
adducts. The application of this approach and results will be demonstrated in Chapter 5.

1.7.4 Challenge 4: Whether or not formaldehyde has distant genotoxic effects

The critical issue concerning the ability of formaldehyde to induce leukemia is whether or not formaldehyde has distant genotoxic effects. All established leukemogens act on the bone marrow, causing marrow toxicity. Formation of DNA adducts is thought to be a key event of significance in initiating cancer. Therefore, valuable information on this question will be provided by examining the formation of formaldehyde-induced DNA adducts at site of contact and distant tissues, such as bone marrow. In addition, the use of [$^{13}CD_2$]-formaldehyde in animal exposures coupled with mass spectrometry will unambiguously identify DNA damage arising from endogenous and inhaled sources. The molecular dosimetry data generated in this study will provide critical evidence for science-based risk assessment of formaldehyde exposure. More importantly, the tissue distribution of exogenous formaldehyde-induced DNA damage will provide strong evidence about the biological plausibility of formaldehyde to induce leukemia. The data toward this end are demonstrated in Chapter 5.

Figures

Figure 1.1. Fate and metabolism of formaldehyde



References

(1) IARC monographs on the evaluation of carcinogenic risks to humans. International Agency for Research on Cancer (2006) *IARC Monogr Eval. Carcinog. Risks Hum.* **88**, 1-287.

(2) Zhang, L., Steinmaus, C., Eastmond, D. A., Xin, X. K., and Smith, M. T. (2009) Formaldehyde exposure and leukemia: a new meta-analysis and potential mechanisms. *Mutat. Res.* **681**(2-3), 150-168.

(3) Tang, X., Bai, Y., Duong, A., Smith, M. T., Li, L., and Zhang, L. (2009) Formaldehyde in China: production, consumption, exposure levels, and health effects. *Environ. Int.* **35**(8), 1210-1224.

(4) Casanova, M., Heck, H. D., Everitt, J. I., Harrington, W. W., Jr., and Popp, J. A. (1988) Formaldehyde concentrations in the blood of rhesus monkeys after inhalation exposure. *Food Chem. Toxicol.* **26**(8), 715-716.

(5) Heck, H. D., Casanova-Schmitz, M., Dodd, P. B., Schachter, E. N., Witek, T. J., and Tosun, T. (1985) Formaldehyde (CH2O) concentrations in the blood of humans and Fischer-344 rats exposed to CH2O under controlled conditions. *Am. Ind. Hyg. Assoc. J.* **46**(1), 1-3.

(6) Cheng, G., Wang, M., Upadhyaya, P., Villalta, P. W., and Hecht, S. S. (2008) Formation of formaldehyde adducts in the reactions of DNA and deoxyribonucleosides with alpha-acetates of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and N-nitrosodimethylamine (NDMA). *Chem. Res. Toxicol.* **21**(3), 746-751.

(7) Wang, M., Cheng, G., Villalta, P. W., and Hecht, S. S. (2007) Development of liquid chromatography electrospray ionization tandem mass spectrometry methods for analysis of DNA adducts of formaldehyde and their application to rats treated with N-nitrosodimethylamine or 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Chem. Res. Toxicol.* **20**(8), 1141-1148.

(8) Chung, F. L., Wang, M. Y., and Hecht, S. S. (1985) Effects of dietary indoles and isothiocyanates on N-nitrosodimethylamine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone alpha-hydroxylation and DNA methylation in rat liver. *Carcinogenesis* **6**(4), 539-543.

(9) Kato, S., Burke, P. J., Koch, T. H., and Bierbaum, V. M. (2001) Formaldehyde in human cancer cells: detection by preconcentration-chemical ionization mass spectrometry. *Anal. Chem.* **73**(13), 2992-2997.

(10) Kato, S., Burke, P. J., Fenick, D. J., Taatjes, D. J., Bierbaum, V. M., and Koch, T. H. (2000) Mass spectrometric measurement of formaldehyde generated in breast cancer cells upon treatment with anthracycline antitumor drugs. *Chem. Res. Toxicol.* **13**(6), 509-516.

(11) Swenberg, J. A., Kerns, W. D., Mitchell, R. I., Gralla, E. J., and Pavkov, K. L. (1980) Induction of squamous cell carcinomas of the rat nasal cavity by inhalation exposure to formaldehyde vapor. *Cancer Res.* **40**(9), 3398-3402.

(12) Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J., and Swenberg, J. A. (1983) Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Res.* **43**(9), 4382-4392.

(13) Cogliano, V. J., Grosse, Y., Baan, R. A., Straif, K., Secretan, M. B., and El, G. F. (2005) Meeting report: summary of IARC monographs on formaldehyde, 2-butoxyethanol, and 1-tert-butoxy-2-propanol. *Environ. Health Perspect.* **113**(9), 1205-1208.

(14) Hauptmann, M., Lubin, J. H., Stewart, P. A., Hayes, R. B., and Blair, A. (2004) Mortality from solid cancers among workers in formaldehyde industries. *Am. J. Epidemiol.* **159**(12), 1117-1130.

(15) Hayes, R. B., Blair, A., Stewart, P. A., Herrick, R. F., and Mahar, H. (1990) Mortality of U.S. embalmers and funeral directors. *Am. J. Ind. Med.* **18**(6), 641-652.

(16) Hansen, J., and Olsen, J. H. (1995) Formaldehyde and cancer morbidity among male employees in Denmark. *Cancer Causes Control* **6**(4), 354-360.

(17) Armstrong, R. W., Imrey, P. B., Lye, M. S., Armstrong, M. J., Yu, M. C., and Sani, S. (2000) Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles, formaldehyde and heat. *Int. J. Epidemiol.* **29**(6), 991-998.

(18) Hildesheim, A., Dosemeci, M., Chan, C. C., Chen, C. J., Cheng, Y. J., Hsu, M. M., Chen, I. H., Mittl, B. F., Sun, B., Levine, P. H., Chen, J. Y., Brinton,

L. A., and Yang, C. S. (2001) Occupational exposure to wood, formaldehyde, and solvents and risk of nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **10**(11), 1145-1153.

(19) Olsen, J. H., Jensen, S. P., Hink, M., Faurbo, K., Breum, N. O., and Jensen, O. M. (1984) Occupational formaldehyde exposure and increased nasal cancer risk in man. *Int. J. Cancer* **34**(5), 639-644.

(20) Roush, G. C., Walrath, J., Stayner, L. T., Kaplan, S. A., Flannery, J. T., and Blair, A. (1987) Nasopharyngeal cancer, sinonasal cancer, and occupations related to formaldehyde: a case-control study. *J. Natl. Cancer Inst.* **79**(6), 1221-1224.

(21) Vaughan, T. L., Strader, C., Davis, S., and Daling, J. R. (1986) Formaldehyde and cancers of the pharynx, sinus and nasal cavity: II. Residential exposures. *Int. J. Cancer* **38**(5), 685-688.

(22) Vaughan, T. L., Strader, C., Davis, S., and Daling, J. R. (1986) Formaldehyde and cancers of the pharynx, sinus and nasal cavity: I. Occupational exposures. *Int. J. Cancer* **38**(5), 677-683.

(23) West, S., Hildesheim, A., and Dosemeci, M. (1993) Non-viral risk factors for nasopharyngeal carcinoma in the Philippines: results from a case-control study. *Int. J. Cancer* **55**(5), 722-727.

(24) Vaughan, T. L., Stewart, P. A., Teschke, K., Lynch, C. F., Swanson, G. M., Lyon, J. L., and Berwick, M. (2000) Occupational exposure to formaldehyde and wood dust and nasopharyngeal carcinoma. *Occup. Environ. Med.* **57**(6), 376-384.

(25) Marsh, G. M., Youk, A. O., and Morfeld, P. (2007) Mis-specified and non-robust mortality risk models for nasopharyngeal cancer in the National Cancer Institute formaldehyde worker cohort study. *Regul. Toxicol. Pharmacol.* **47**(1), 59-67.

(26) Marsh, G. M., Youk, A. O., Buchanich, J. M., Cassidy, L. D., Lucas, L. J., Esmen, N. A., and Gathuru, I. M. (2002) Pharyngeal cancer mortality among chemical plant workers exposed to formaldehyde. *Toxicol. Ind. Health* **18**(6), 257-268.

(27) Marsh, G. M., Stone, R. A., Esmen, N. A., Henderson, V. L., and Lee, K. Y. (1996) Mortality among chemical workers in a factory where formaldehyde was used. *Occup. Environ. Med.* **53**(9), 613-627.

(28) Marsh, G. M., Stone, R. A., Esmen, N. A., and Henderson, V. L. (1994) Mortality patterns among chemical plant workers exposed to formaldehyde and other substances. *J. Natl. Cancer Inst.* **86**(5), 384-386.

(29) Marsh, G. M. (1982) Proportional mortality patterns among chemical plant workers exposed to formaldehyde. *Br. J. Ind. Med.* **39**(4), 313-322.

(30) Marsh, G. M., and Youk, A. O. (2005) Reevaluation of mortality risks from nasopharyngeal cancer in the formaldehyde cohort study of the National Cancer Institute. *Regul. Toxicol. Pharmacol.* **42**(3), 275-283.

(31) Marsh, G. M., Youk, A. O., Buchanich, J. M., Erdal, S., and Esmen, N. A. (2007) Work in the metal industry and nasopharyngeal cancer mortality among formaldehyde-exposed workers. *Regul. Toxicol. Pharmacol.* **48**(3), 308-319.

(32) Coggon, D., Harris, E. C., Poole, J., and Palmer, K. T. (2003) Extended follow-up of a cohort of british chemical workers exposed to formaldehyde. *J. Natl. Cancer Inst.* **95**(21), 1608-1615.

(33) Pinkerton, L. E., Hein, M. J., and Stayner, L. T. (2004) Mortality among a cohort of garment workers exposed to formaldehyde: an update. *Occup. Environ. Med.* **61**(3), 193-200.

(34) Hauptmann, M., Lubin, J. H., Stewart, P. A., Hayes, R. B., and Blair, A. (2003) Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries. *J. Natl. Cancer Inst.* **95**(21), 1615-1623.

(35) Beane Freeman, L. E., Blair, A., Lubin, J. H., Stewart, P. A., Hayes, R. B., Hoover, R. N., and Hauptmann, M. (2009) Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the National Cancer Institute Cohort. *J. Natl. Cancer Inst.* **101**(10), 751-761.

(36) Speit, G., Neuss, S., and Schmid, O. (2009) The human lung cell line A549 does not develop adaptive protection against the DNA-damaging action of formaldehyde. *Environ. Mol. Mutagen.*

(37) Speit, G., Zeller, J., Schmid, O., Elhajouji, A., Ma-Hock, L., and Neuss, S. (2009) Inhalation of formaldehyde does not induce systemic genotoxic effects in rats. *Mutat. Res.* **677**(1-2), 76-85.

(38) Speit, G., Schmid, O., Neuss, S., and Schutz, P. (2008) Genotoxic effects of formaldehyde in the human lung cell line A549 and in primary human nasal epithelial cells. *Environ. Mol. Mutagen.* **49**(4), 300-307.

(39) Speit, G., Neuss, S., Schutz, P., Frohler-Keller, M., and Schmid, O. (2008) The genotoxic potential of glutaraldehyde in mammalian cells in vitro in comparison with formaldehyde. *Mutat. Res.* **649**(1-2), 146-154.

(40) Speit, G., Schutz, P., Hogel, J., and Schmid, O. (2007) Characterization of the genotoxic potential of formaldehyde in V79 cells. *Mutagenesis* **22**(6), 387-394.

(41) Speit, G., Schmid, O., Frohler-Keller, M., Lang, I., and Triebig, G. (2007) Assessment of local genotoxic effects of formaldehyde in humans measured by the micronucleus test with exfoliated buccal mucosa cells. *Mutat. Res.* **627**(2), 129-135.

(42) Speit, G. (2006) The implausibility of systemic genotoxic effects measured by the comet assay in rats exposed to formaldehyde. *J. Proteome. Res.* **5**(10), 2523-2524.

(43) Speit, G., and Schmid, O. (2006) Local genotoxic effects of formaldehyde in humans measured by the micronucleus test with exfoliated epithelial cells. *Mutat. Res.* **613**(1), 1-9.

(44) Speit, G., and Merk, O. (2002) Evaluation of mutagenic effects of formaldehyde in vitro: detection of crosslinks and mutations in mouse lymphoma cells. *Mutagenesis* **17**(3), 183-187.

(45) Merk, O., Reiser, K., and Speit, G. (2000) Analysis of chromate-induced DNA-protein crosslinks with the comet assay. *Mutat. Res.* **471**(1-2), 71-80.

(46) Speit, G., Schutz, P., and Merk, O. (2000) Induction and repair of formaldehyde-induced DNA-protein crosslinks in repair-deficient human cell lines. *Mutagenesis* **15**(1), 85-90.

(47) Merk, O., and Speit, G. (1999) Detection of crosslinks with the comet assay in relationship to genotoxicity and cytotoxicity. *Environ. Mol. Mutagen.* **33**(2), 167-172.

(48) Merk, O., and Speit, G. (1998) Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ. Mol. Mutagen.* **32**(3), 260-268.

(49) Quievryn, G., and Zhitkovich, A. (2000) Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* **21**(8), 1573-1580.

(50) Yu, L. Q., Jiang, S. F., Leng, S. G., He, F. S., and Zheng, Y. X. (2005) [Early genetic effects on workers occupationally exposed to formaldehyde]. *Zhonghua Yu Fang Yi. Xue. Za Zhi.* **39**(6), 392-395.

(51) Orsiere, T., Sari-Minodier, I., Iarmarcovai, G., and Botta, A. (2006) Genotoxic risk assessment of pathology and anatomy laboratory workers exposed to formaldehyde by use of personal air sampling and analysis of DNA damage in peripheral lymphocytes. *Mutat. Res.* **605**(1-2), 30-41.

(52) Iarmarcovai, G., Bonassi, S., Sari-Minodier, I., Baciuchka-Palmaro, M., Botta, A., and Orsiere, T. (2007) Exposure to genotoxic agents, host factors, and lifestyle influence the number of centromeric signals in micronuclei: a pooled re-analysis. *Mutat. Res.* **615**(1-2), 18-27.

(53) Recio, L., Sisk, S., Pluta, L., Bermudez, E., Gross, E. A., Chen, Z., Morgan, K., and Walker, C. (1992) p53 mutations in formaldehyde-induced nasal squamous cell carcinomas in rats. *Cancer Res.* **52**(21), 6113-6116.

(54) Zhong, W., and Hee, S. S. (2005) Comparison of UV, fluorescence, and electrochemical detectors for the analysis of formaldehyde-induced DNA adducts. *J. Anal. Toxicol.* **29**(3), 182-187.

(55) Zhong, W., and Hee, S. Q. (2004) Quantitation of normal and formaldehyde-modified deoxynucleosides by high-performance liquid chromatography/UV detection. *Biomed. Chromatogr.* **18**(7), 462-469.

(56) Zhong, W., and Que Hee, S. S. (2004) Formaldehyde-induced DNA adducts as biomarkers of in vitro human nasal epithelial cell exposure to formaldehyde. *Mutat. Res.* **563**(1), 13-24.

(57) Shaham, J., Bomstein, Y., Gurvich, R., Rashkovsky, M., and Kaufman, Z. (2003) DNA-protein crosslinks and p53 protein expression in relation to occupational exposure to formaldehyde. *Occup. Environ. Med.* **60**(6), 403-409.

(58) Shaham, J., Bomstein, Y., Meltzer, A., Kaufman, Z., Palma, E., and Ribak, J. (1996) DNA--protein crosslinks, a biomarker of exposure to formaldehyde--in vitro and in vivo studies. *Carcinogenesis* **17**(1), 121-125.

(59) Casanova-Schmitz, M., Starr, T. B., and Heck, H. D. (1984) Differentiation between metabolic incorporation and covalent binding in the labeling of macromolecules in the rat nasal mucosa and bone marrow by inhaled [14C]- and [3H]formaldehyde. *Toxicol. Appl. Pharmacol.* **76**(1), 26-44.

(60) Swenberg, J. A., Barrow, C. S., Boreiko, C. J., Heck, H. D., Levine, R. J., Morgan, K. T., and Starr, T. B. (1983) Non-linear biological responses to formaldehyde and their implications for carcinogenic risk assessment. *Carcinogenesis* **4**(8), 945-952.

(61) Casanova-Schmitz, M., and Heck, H. D. (1983) Effects of formaldehyde exposure on the extractability of DNA from proteins in the rat nasal mucosa. *Toxicol. Appl. Pharmacol.* **70**(1), 121-132.

(62) Casanova, M., Morgan, K. T., Gross, E. A., Moss, O. R., and Heck, H. A. (1994) DNA-protein cross-links and cell replication at specific sites in the nose of F344 rats exposed subchronically to formaldehyde. *Fundam. Appl. Toxicol.* **23**(4), 525-536.

(63) Monticello, T. M., Swenberg, J. A., Gross, E. A., Leininger, J. R., Kimbell, J. S., Seilkop, S., Starr, T. B., Gibson, J. E., and Morgan, K. T. (1996) Correlation of regional and nonlinear formaldehyde-induced nasal cancer with proliferating populations of cells. *Cancer Res.* **56**(5), 1012-1022.

(64) Conolly, R. B., Kimbell, J. S., Janszen, D., Schlosser, P. M., Kalisak, D., Preston, J., and Miller, F. J. (2003) Biologically motivated computational modeling of formaldehyde carcinogenicity in the F344 rat. *Toxicol. Sci.* **75**(2), 432-447.

(65) Shaham, J., Gurvich, R., and Kaufman, Z. (2002) Sister chromatid exchange in pathology staff occupationally exposed to formaldehyde. *Mutat. Res.* **514**(1-2), 115-123.

(66) Shaham, J., Bomstein, Y., Melzer, A., and Ribak, J. (1997) DNA-Protein Crosslinks and Sister Chromatid Exchanges as Biomarkers of Exposure to Formaldehyde. *Int. J. Occup. Environ. Health* **3**(2), 95-104.

(67) Casanova, M., Heck, H. D., and Janszen, D. (1996) Comments on 'DNA-protein crosslinks, a biomarker of exposure to formaldehyde--in vitro and in vivo studies' by Shaham et al. *Carcinogenesis* **17**(9), 2097-2101.

(68) Wang, M., Cheng, G., Balbo, S., Carmella, S. G., Villalta, P. W., and Hecht, S. S. (2009) Clear differences in levels of a formaldehyde-DNA adduct in leukocytes of smokers and nonsmokers. *Cancer Res.* **69**(18), 7170-7174.

(69) Pyatt, D., Natelson, E., and Golden, R. (2008) Is inhalation exposure to formaldehyde a biologically plausible cause of lymphohematopoietic malignancies? *Regul. Toxicol. Pharmacol.* **51**(1), 119-133.

(70) Golden, R., Pyatt, D., and Shields, P. G. (2006) Formaldehyde as a potential human leukemogen: an assessment of biological plausibility. *Crit Rev. Toxicol.* **36**(2), 135-153.

(71) Heck, H., and Casanova, M. (2004) The implausibility of leukemia induction by formaldehyde: a critical review of the biological evidence on distant-site toxicity. *Regul. Toxicol. Pharmacol.* **40**(2), 92-106.

(72) McDevitt, M. A., Condon, M., Stamberg, J., Karp, J. E., and McDiarmid, M. (2007) Fluorescent in situ hybridization (FISH) in bone marrow and peripheral blood of leukemia patients: implications for occupational surveillance. *Mutat. Res.* **629**(1), 24-31.

(73) Hirabayashi, Y., Yoon, B. I., Li, G. X., Kanno, J., and Inoue, T. (2004) Mechanism of benzene-induced hematotoxicity and leukemogenicity:

current review with implication of microarray analyses. *Toxicol. Pathol.* **32 Suppl 2**, 12-16.

(74) Rothman, N., Smith, M. T., Hayes, R. B., Li, G. L., Irons, R. D., Dosemeci, M., Haas, R., Stillman, W. S., Linet, M., Xi, L. Q., Bechtold, W. E., Wiemels, J., Campleman, S., Zhang, L., Quintana, P. J., Titenko-Holland, N., Wang, Y. Z., Lu, W., Kolachana, P., Meyer, K. B., and Yin, S. (1996) An epidemiologic study of early biologic effects of benzene in Chinese workers. *Environ. Health Perspect.* **104 Suppl 6**, 1365-1370.

(75) Snyder, R., and Kalf, G. F. (1994) A perspective on benzene leukemogenesis. *Crit Rev. Toxicol.* **24**(3), 177-209.

(76) Carbonell, F., Seidel, H. J., Saks, S., and Kreja, L. (1982) Chromosome changes in butylnitrosourea (BNU)-induced mouse leukemia. *Int. J. Cancer* **30**(4), 511-516.

(77) McHale, C. M., Lan, Q., Corso, C., Li, G., Zhang, L., Vermeulen, R., Curry, J. D., Shen, M., Turakulov, R., Higuchi, R., Germer, S., Yin, S., Rothman, N., and Smith, M. T. (2008) Chromosome translocations in workers exposed to benzene. *J. Natl. Cancer Inst. Monogr*(39), 74-77.

(78) Zhang, L., Rothman, N., Li, G., Guo, W., Yang, W., Hubbard, A. E., Hayes, R. B., Yin, S., Lu, W., and Smith, M. T. (2007) Aberrations in chromosomes associated with lymphoma and therapy-related leukemia in benzene-exposed workers. *Environ. Mol. Mutagen.* **48**(6), 467-474.

(79) Smith, M. T., McHale, C. M., Wiemels, J. L., Zhang, L., Wiencke, J. K., Zheng, S., Gunn, L., Skibola, C. F., Ma, X., and Buffler, P. A. (2005) Molecular biomarkers for the study of childhood leukemia. *Toxicol. Appl. Pharmacol.* **206**(2), 237-245.

(80) Zhang, L., Rothman, N., Wang, Y., Hayes, R. B., Li, G., Dosemeci, M., Yin, S., Kolachana, P., Titenko-Holland, N., and Smith, M. T. (1998) Increased aneusomy and long arm deletion of chromosomes 5 and 7 in the lymphocytes of Chinese workers exposed to benzene. *Carcinogenesis* **19**(11), 1955-1961.

(81) Smith, M. T., and Zhang, L. (1998) Biomarkers of leukemia risk: benzene as a model. *Environ. Health Perspect.* **106 Suppl 4**, 937-946.

(82) Zhang, L., Freeman, L. E., Nakamura, J., Hecht, S. S., Vandenberg, J. J., Smith, M. T., and Sonawane, B. R. (2009) Formaldehyde and leukemia: Epidemiology, potential mechanisms, and implications for risk assessment. *Environ. Mol. Mutagen.*

(83) Costa, M., Zhitkovich, A., Harris, M., Paustenbach, D., and Gargas, M. (1997) DNA-protein cross-links produced by various chemicals in cultured human lymphoma cells. *J. Toxicol. Environ. Health* **50**(5), 433-449.

(84) Zhitkovich, A., and Costa, M. (1992) A simple, sensitive assay to detect DNA-protein crosslinks in intact cells and in vivo. *Carcinogenesis* **13**(8), 1485-1489.

CHAPTER 2

FORMALDEHYDE-INDUCED HISTONE MODIFICATIONS *IN VITRO*

This paper has been published and therefore is reproduced with permission from [Kun Lu, Gunnar Boysen, Lina Gao, Leonard B. Collins and James A. Swenberg. Formaldehyde-Induced Histone Modifications in Vitro. *Chem. Res. Toxicol*., 2008, 21 (8), pp 1586–1593] Copyright [2008] American Chemical Society.

2.1 Abstract

Numerous experiments have demonstrated the genotoxic and mutagenic effects of formaldehyde, including DNA-protein crosslinks (DPC). Histone was reported to be involved in the formation of DPC in which the epsilon-amino groups of lysine and exocyclic amino groups of DNA were thought to be cross-linked through multiple step reactions. Using mass spectrometry, the N-terminus of histone and lysine residues located in both the histone N-terminal tail and the globular fold domain were identified as binding sites for formaldehyde in the current study. The observation that only lysine residues without post-translational modification (PTM) can be attacked by formaldehyde indicates that PTM blocks the reaction between lysine and formaldehyde. Additionally, we found that formaldehyde-induced Schiff

based on lysine residues could inhibit the formation of PTM on histone, raising the possibility that formaldehyde might alter epigenetic regulation.

2.2 Introduction

Formaldehyde is produced worldwide on a large scale and is widely used in the manufacture of resins, particle board, plywood, leather goods, paper, pharmaceuticals and other products (*1*). Therefore, occupational and environmental exposures to formaldehyde are quite common. Previous research has demonstrated that formaldehyde is genotoxic and mutagenic to mammalian cells, as well as to bacteria and lower eukaryotes (*1-6*). Exposure to formaldehyde results in the formation of DNA-protein crosslinks (DPC) as a primary genotoxic effect (*2-7*), and histones are reported to be cross-linked to DNA (*7, 8*). The formation of DPC, in which the epsilon-amino groups of lysine and exocyclic amino groups of DNA are thought to be involved (*9, 10*), proceeds in multiple-step reactions, with formaldehyde initiating an attack on the amino groups of protein, followed by DPC formation.

Core histones are located in the nucleosome, the fundamental repeating unit of chromatin, and contain a histone 3 (H3) /histone 4 (H4) tetramer flanked by 2 histone 2A (H2A) /histone 2B (H2B) dimers, around which 147 base pairs of DNA are wrapped (11). The highly conserved structure of histone includes a folded core and an unstructured tail. The histone core is a globular fold domain with a helix-loop-helix-loop-helix motif, which facilitates dimerization through a hand-shaking motion (*12*). However, the crystal structure reveals that histone tails do not have any defined

conformation. Many conserved amino acid residues such as lysine are subject to a broad spectrum of post-translational modifications (PTM) including methylation and acetylation. These modifications impact biological actions such as gene expression by acting as markers for the specific recruitment of regulatory complexes and by changing the structure of the chromatin between heterochromatin and euchromatin, which is usually referred to as epigenetic regulation (*13*).

It has been shown that the lysine residues of histones are probably involved in DPC. However, the accessibility of amino acids, which is controlled not only by reactive ability but also the conformation states of the residues (*14*), has not been evaluated. In addition, a wealth of post-translational modifications on histone raises the question whether the PTM on histone have any influence on the reaction with formaldehyde. Another very important question is whether formaldehyde-induced modifications impact the formation of PTM on histone lysine residues. In this investigation, histone 4 was chosen as our model histone protein due to its indispensable role in epigenetic regulation. Using mass spectrometry, we have now identified the residues of histone 4 which are accessible to formaldehyde binding. We have also demonstrated that PTM on lysine residues block the reaction between formaldehyde and the epsilon-amino groups of lysine residues. In addition, we found that formaldehyde-induced Schiff bases on lysine can inhibit the formation of PTM.

2.3 Materials and Methods

2.3.1 Chemicals and Reagents

Potassium phosphate, ammonium bicarbonate, glycine, trifluoroacetic acid, formic acid, acetonitrile, sodium cyanoborohydride ($NaCNBH_3$), acetyl coenzyme A trilithium salt (acetylCoA) and endoproteinase Asp-N were purchased from Sigma (St. Louis, MO). Modified sequencing grade trypsin and chymotrypsin were from Promega (San Luis Obispo, CA) and PrinSep (Adelphia, NJ), respectively. The histone 4 isolated from calf thymus was purchased from Roche Applied Science (Indianapolis, IN) and human recombinant histone 4 expressed in *E.coli*, active PCAF enzyme and 5X histone acetylation (HAT) assay buffer (250 mM Tris-base, pH=8.0, 50% glycerol, 0.5 mM EDTA and 5 mM dithiothreitol) were obtained from Millipore (Billerica, MA). 20% formaldehyde in water was procured from Tousimis (Rockville, MD), while stable isotope-labeled formaldehyde, $^{13}CH_2O$, was obtained from Cambridge Isotope Laboratories (Andover, MA). The peptides were synthesized by Genscript Corporation (Piscatway, NJ). All chemicals were used as received unless otherwise stated.

2.3.2 Experimental Methods

The analytical approach used in this study is illustrated in Scheme 2.1. First, the reaction of histone 4 with formaldehyde was performed, followed by reduction with $NaCNBH_3$. 20 μg *of* histone 4 was dissolved in 50 μL 10 mM potassium phosphate buffer (pH=7.2). Formaldehyde then was added to a final concentration of

either 5 mM or 50μM. After a 10 min of reaction between formaldehyde and histone, reduction was performed with $NaCNBH_3$ at a final concentration of 5 mM. The reaction solution was further incubated for 3 hours. Excessive formaldehyde and other small chemicals were removed using Millipore Microcon YM-3 spin columns with a 3000 molecular weight cutoff (MWCO). Two identical reactions were prepared using either native formaldehyde ($^{12}CH_2O$) or ^{13}C labeled formaldehyde ($^{13}CH_2O$). After proteolysis, equimolar amounts of $^{12}CH_2O$-treated and $^{13}CH_2O$-treated histone 4 desalted digestion solutions were mixed for further mass spectrometry analyses. In addition, the resultant histone 4 from the formaldehyde and $NaCNBH_3$ treatment was further treated separately with formaldehyde or formaldehyde plus equimolar glycine. In these reactions, the final concentration of formaldehyde or glycine varied from 5 mM to 100 mM and the reaction time ranged from 3 hours to 2 weeks.

2.3.3 Digestion by Enzymes

Untreated histone 4 and formaldehyde-treated histone 4 were digested with trypsin, chymotrypsin, and endoproteinase Asp-N, individually. For the trypsin digestion, 20 *μg of* histone was dissolved in 50 mM NH_4HCO_3 (pH=7.8), and 0.4 *μg of* trypsin was added. The solution was incubated for 3 hours at 37 °C. For chymotrypsin cleavage, 0.4 *μg of* chymotrypsin was mixed with 20 *μg of* histone dissolved in 50 mM Tris HCl (pH=8.0) and 1 mM $CaCl_2$. The digestion mixture was incubated for 3 hours at 30 °C. Endoproteinase Asp-N digestion was performed in

100 mM NH_4HCO_3 (pH=8.5) at a 1:50 enzyme to substrate ratio for 18 hours at 37 °C.

2.3.4 Histone Acetylation Assay

5 μg of synthetic histone 4 N-terminal peptide (amino acid 1-23) and 500 ng of active PCAF enzyme were dissolved in 1X HAT assay buffer. AcetylCoA was added to a final concentration of 0.2 mM and the solutions were incubated for 1 hour at 30 °C with shaking. The resultant reaction mixtures were analyzed by liquid chromatography-mass spectrometry. In parallel, 10 μg of histone 4 synthetic peptide was treated with 50 mM formaldehyde at room temperature for 3 hours. The formaldehyde-modified peptide was dried using a speed vacuum. Then the formaldehyde-modified peptide was incubated with active PCAF enzyme for the HAT assay, followed by mass spectrometry analysis.

2.3.5 Liquid Chromatography-Mass Spectrometry (LC-MS)

LC-MS analyses were performed on an ion trap mass spectrometer LCQ-Deca (Thermo Electron, Waltham, MA) operating in full scan as well as dependent scan mode. Analytes were separated by reverse phase chromatography using a 250 mm × 2.5mm analytical column from Grace Vydac (Hesperia, CA). The mobile phase consisted of 0.1% formic acid in water (solvent A) and acetonitrile (solvent B). A linear gradient was started from 5% acetonitrile to 55% in 15 min. The flow rate of the mobile phase was set as 200 μL/min. An electrospray ionization (ESI) source was

used to analyze peptide samples. For the fragmentation of precursor ions, the normalized collision energy varied from 30% to 35%, depending on the structures of peptides. The activation time was set at 30 msec.

2.3.6 Matrix-Assisted Laser Desorption/Ionization Tandem Time-of-Flight Mass Spectrometry (MALDI-TOF MS)

MALDI-TOF mass spectrometric analyses were performed on a 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA) operating in reflectron mode. Most of the tandem mass spectrometric experiments were performed on the MALDI-TOF/TOF mass spectrometer using air as the collision gas at a medium pressure setting (4e-007 torr) and a laser intensity of 5400 ABI units (Nd-Yag laser, 355 nm wavelength, 3–7 ns pulse, >12 μJ pulse energy.). The matrix applied for peptide analysis was α-cyano-4-hydroxy-cinnamic acid. The MS-Digest or MS-Fit program (Protein Prospector, University of California at San Francisco) was used to automatically assign the peptides. The MS/MS fragment ion spectra were manually matched to the predicted peptide fragmentation generated by Data Explorer software.

2.4 Results

2.4.1 Characterization of Histone 4

Two different types of histone 4 were used in this study. Histone 4 with PTM was isolated from calf thymus tissues, while unmodified human recombinant H4 was purified after expression in *E.coli* cells. The sequences of the both proteins were

identical. Prior to the formaldehyde reaction, the purity and molecular weight of histone 4 were examined by LC-MS. Histone 4 was digested with trypsin and chymotrypsin to improve the coverage for peptide mapping, yielding 6 and 11 detectable fragments for trypsin and chymotrypsin proteolysis, respectively, as listed in Table 1.

By combining the peptides from independent trypsin and chymotrypsin digestion, all the residues located in the histone fold domain could be covered. However, neither trypsin nor chymotrypsin was able to provide any information about the histone N-terminal tail. To obtain this data, endoproteinase Asp-N was utilized to produce N-terminal peptides. This protease cleaves proteins at the amino side of aspartic acid. In theory, this will result in 4 non-overlapping peptides in complete digestion. The largest predicted fragment at 5004.91 Da is over the cutoff of the MALDI-TOF mass spectrometer operating at reflectron mode, so only 3 peptides were observed, as listed in Table 1. In addition, after the Asp-N digestion of histone 4 with PTM, two abundant peptides with PTM were detected, as shown in Table 1.

2.4.2 Evaluation of the Accessibility of Lysine Residues

The reaction between formaldehyde and amine involves a nucleophilic attack of amine on the carbonyl group of formaldehyde, followed by rapid proton transfer resulting in methylol groups. Subsequently, labile Schiff bases are produced after dehydration of methylol. These two reactions are both reversible, thus they can not be easily detected. Therefore, a reduction approach with sodium cyanoborohydride was

chosen to reduce the Schiff base to a detectable dimethylamino structure. The reaction mechanism is shown in Scheme 2.2. First, a Schiff base formed from formaldehyde attack is quickly reduced to a secondary amine, which is relatively more reactive than a primary amine. The secondary amine then reacts with another formaldehyde molecule and is further reduced to form a dimethylamino group (*15*).

In addition to $^{12}CH_2O$, stable isotope-labeled formaldehyde, $^{13}CH_2O$, was used to treat histone 4. Mixing equal ratios of $^{12}CH_2O$- and $^{13}CH_2O$- peptides results in the formation of doublets in a mass spectrum. The mass difference between a doublet allows us to calculate how many formaldehyde molecules were incorporated into each peptide. For instance, if the mass difference between the peptide pair equals 2, this indicates that 2 formaldehyde units reacted with the residue and, consequently, 1 lysine residue is reactive with formaldehyde. Thus the general formula is $n=\Delta m/2$, where n is the number of reactive lysines and Δm corresponds to the mass difference between the doublet.

The accessibility of each lysine residue by formaldehyde was confirmed by either MALDI-TOF or LC-MS. A good match of mass increase after formaldehyde treatment or MS/MS fragmentation pattern of formaldehyde-modified peptides permitted the identification of the exact reactive sites with high confidence. The formaldehyde-modified histones were digested with 3 different enzymes, and the resulting peptides were analyzed by MALDI-TOF mass spectrometry to determine the mass increases of individual peptide sequences upon formaldehyde treatment (Table 2).

There were 5 detectable tryptic peptides by MALDI-TOF mass spectrometry, with three of them having the corresponding doublets when ^{13}C labeled formaldehyde was used, as listed in Table 2. Figure 2.1A shows a typical doublet (Δm=2 Da) arising from the application of $^{12}CH_2O$ and $^{13}CH_2O$. It is very straightforward to assign the peptide at 1353.85 Da as the dimethylated product of original tryptic peptide at 1325.77 Da ($_{24}$DNIQGITKPAIR$_{35}$) after observing the 28 Da mass increase and a doublet separated by 2 Da. The specific sequence and exact reaction sites were confirmed by MS/MS sequencing with MALDI TOF/TOF and the MS/MS fragmentation pattern of the precursor ion of 1353.82 Da is given in Figure 2.1B.

According to the doublets resulting from ^{13}C labeled formaldehyde, the fragments at 1414.96 and 1651.07 Da were identified as having 1 and 2 lysine residues accessible to formaldehyde. Their structures were determined to be $_{56}$GVL**K**$_{2me}$VFLENVIR$_{67}$ and $_{79}$**K**$_{2me}$TVTAMDVVYAL**K**$_{2me}$R$_{92}$, respectively. It is interesting to note that two expected tryptic peptides at 989.57 and 1310.70 Da disappeared after the reaction with formaldehyde. However, no direct formaldehyde-induced precursor ions could be attributed to them. Increased cleavage resistance from the dimethylation at K59 and K91 after $NaCNBH_3$ reduction could be responsible for their absence.

Two unique fragments at 1297.87 Da and 1905.10 Da (shown in Figure 2.1C) were identified as $_{38}$ARRGGV**K**$_{2me}$RISGL$_{49}$ and $_{73}$TEHA**K**$_{2me}$R**K**$_{2me}$TVTAMDVVY$_{88}$ after chymotrypsin proteolysis, suggesting that

K44 and K77 are two additional residues in the fold domain that can be attacked by formaldehyde.

The accessibility of lysine residues along the N-terminal tail was evaluated after endoproteinase Asp-N cleavage, which liberated the N-terminal peptide consisting of the first 23 amino acids. According to the mass difference (Δm=12 Da) between the resulting doublet, 12 formaldehyde units were incorporated into the 5 lysine residues and the N-terminus, resulting in 168 Da mass increases (ΔM=168 Da) compared with the untreated N-terminal peptide (shown in Figure 2.1D). Therefore, all the lysine residues located in the globular fold domain and the N-terminal tail coupled with the end amino group of the N-terminus are accessible to formaldehyde**,** providing a total of 12 potential reactive sites. The reactive sites identified by mass spectrometry are visualized in a structural model, as shown in Figure 2.1E.

In addition to the high concentration formaldehyde used in this study, a physiological level of formaldehyde (50 μM) was also used to treat histone 4. The binding sites we identified were consistent with those found based on the experiment using the higher formaldehyde treatment (5mM) (data not shown here). The concentration of formaldehyde only influenced the percentage of modified histone, as shown in Figure 2.2. Figure 2.2A shows the molecular weight of unmodified histone 4 (11236 Da), which was determined after deconvoluting the peaks from ESI-MS. Figure 2.2B gives the deconvoluted spectrum for the 5mM formaldehyde-treated histone sample. The most intense peak was attributed to modified histone and its

molecular weight was determined to be 11574 Da. The mass increase (about 338 Da) resulted from the formation of dimethyl groups on 11 lysine residues as well as the N-terminal amino group. Figure 2.2C shows the deconvoluted mass spectrum of histone treated with 50 μM formaldehyde. Two major peaks were found and assigned as unmodified histone 4 (11236 Da) and formaldehyde-modified histone 4 (11574 Da). Compared with Figure 2.2B, the percentage of modified histone is much lower than that of the higher formaldehyde-treated sample.

In theory, other amino acids including arginine, glutamine, tyrosine, histidine, cystine, and asparagines could also potentially react with formaldehyde (*14*). However neither the formaldehyde, nor the formaldehyde-glycine treatment described in the experimental section yielded any detectable modifications on these residues using the current analytical approach. Remarkably, we found that a 1 hour reaction at 37 °C with 5 mM formaldehyde can induce obvious intermolecular crosslinks, but no such crosslinks could be observed even with a 2 week, 100 mM formaldehyde treatment once these lysine residues of histone were blocked by dimethylation (data not shown). These results may further highlight the higher reactivity and importance of lysine in forming intermolecular crosslinks.

2.4.3 Influence of PTM on the Reaction

Prior to formaldehyde treatment, H4 with PTM was cleaved by endoproteinase Asp-N, and the resultant abundant N-terminal peptides are shown in Figure 2.3 (Control panel). The peptide fragment at 2430.54 Da was assigned to the peptide

$_{1}$SGRGKGGKGLGKGGAKRHRKVLR$_{23}$ with dimethylation on K20 and one acetylation on the N-terminus, while the peptide observed at 2472.55 Da had an additional acetylation on lysine, resulting in 42 Da mass increases compared with the fragment at 2430.54 Da. These assignments are consistent with previous identifications of histone acetylation isomers (*16*).

Figure 2.3 (Treatment panel) shows the mass spectrum of formaldehyde-treated histone 4 after Asp-N digestion. Two relatively abundant fragments were observed at 2542.71 Da and 2556.68 Da corresponding to the formaldehyde-modified products of untreated peptides at 2430.54 Da and 2472.55 Da (ΔM equals 112 and 84 Da, respectively). Although unmodified histone offers 5 lysine residues and the N-terminus as potential reactive sites along the N-terminal tail, the number of reactive lysine residues was determined to be 4 and 3 for the fragments at 2542.71 Da and 2556.68 Da, respectively. The observation that only the lysine residues void of PTM could be attacked by formaldehyde clearly shows that PTM inhibits the reaction between formaldehyde and lysine.

2.4.4 Impact of Formaldehyde-Induced Modifications on PTM Formation

Figure 2.4A shows the ESI spectrum of a synthetic H4 N-terminal peptide, which was used as the substrate of histone acetyltransferases for the histone acetylation assay. The precursor ion at 1181.45 m/z is the doubly charged form of H4 N-terminal peptide. Figure 2.4B shows the ESI mass spectrum of histone acetyltransferase (PCAF)-treated H4 N-terminal peptide, which was obtained after

incubating H4 N-terminal peptide with PCAF enzyme in the presence of acetylCoA for 1 hour at 30 °C. Doubly charged ions at 1202.24 m/z and 1223.84 m/z are monoacetylated (m/z increase=21 Da) and diacetylated (m/z increase=42 Da) H4 N-terminal peptides.

Figure 2.4C shows the ESI mass spectrum of a formaldehyde-modified H4 N-terminal tail. The m/z shift is about 36 Da for a doubly charged ion, which is attributed to the formation of 6 Schiff bases on lysine residues as well as the N-terminus. Figure 2.4D shows the ESI mass spectrum of a formaldehyde-modified H4 N-terminal peptide after a histone acetyltransferase treatment. It was generated by incubating formaldehyde-modified H4 N-terminal peptide with PCAF enzyme in the presence of acetylCoA for 1 hour at 30 °C. Compared with the formaldehyde-modified H4 N-terminal peptide shown in Figure 2.4C, no mass increase upon further histone acetyltransferase treatment could be observed, which clearly shows that formaldehyde-induced Schiff bases inhibit the formation of PTM *in vitro.*

2.5 Discussion

One of the goals of this study was to evaluate the reactivity of histone residues towards formaldehyde using mass spectrometry. We have unambiguously demonstrated that all the lysine residues located in both the globular fold domain and the unstructured N-terminal tail are accessible to formaldehyde. We also found that PTM on histone residues inhibited the attack of formaldehyde. Additionally, we

demonstrated that formaldehyde-induced lysine modifications inhibited the formation of PTM *in vitro*.

One question may be raised by the observation that PTM on residues inhibits the reaction between formaldehyde and lysine residues. Histones are subject to a variety of post- translational modifications, including methylation, acetylation and phosphorylation. The accessibility of DNA for many important biological events such as transcription, replication, recombination and repair is regulated by these modifications through forming either open or condensed chromatin (*17*). As far as the histone 4 tail is concerned, K5, K8, K12, K16 and the end amino group of S1 can be acetylated, whereas methylation occurs on K20, which can be mono-, di- or trimethylated on the side chain (*16*). Are the lysine residues still accessible to formaldehyde considering the widespread occurrence of PTM on histone? It is well known that histone acetylation is catalyzed by histone acetyltransferases, and histone deacetylation is realized by histone deacetylases, making this process completely reversible (*18*). Lysine methylation was once thought to be an irreversible process, however enzymes identified recently are capable of demethylating histone at specific sites (*19*). Although lysine residues are protected by PTM, formaldehyde could still attack lysine residues provided the PTM is removed. Our finding highlights the importance of less modified states of histone or nascent histone, which usually is only diacetylated in the cytoplasm at K5 and K12 by B-type histone acetyltransferase (*16*).

The second question is how the modifications caused by formaldehyde evolve. There are three different structures for such modifications. The reaction between

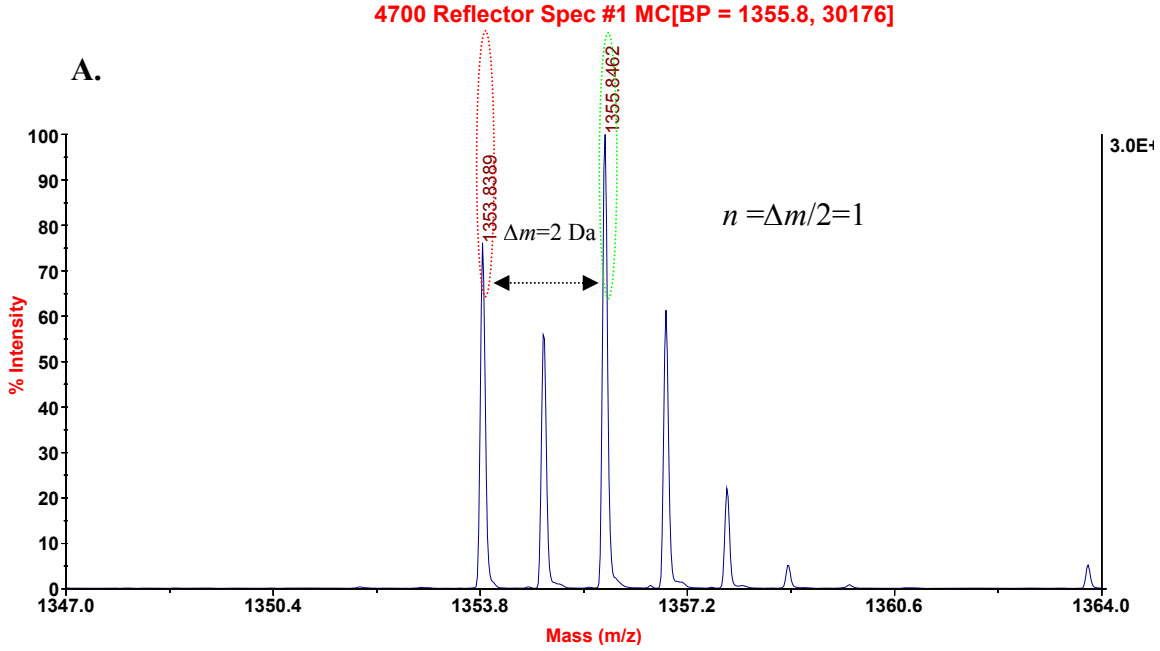
formaldehyde and lysine includes the quick formation of methylol groups, followed by the formation of Schiff bases from partial dehydration of methylol groups. The final step may yield intra-molecular or inter-molecular crosslinks, depending on the local physical and chemical environment of reactive groups. Although methylol and Schiff bases are involved in reversible reactions, the existence of all three structures was confirmed in previous research using mass spectrometry or NMR (*14*). Therefore, these distinct structures have potential biological impact. As we have shown, Schiff bases induced by formaldehyde inhibit the formation of PTM on lysine *in vitro*. Clearly, additional experiments need to be done in order to track the exact structure occurring on individual lysine molecules and the corresponding biological influence *in vivo*.

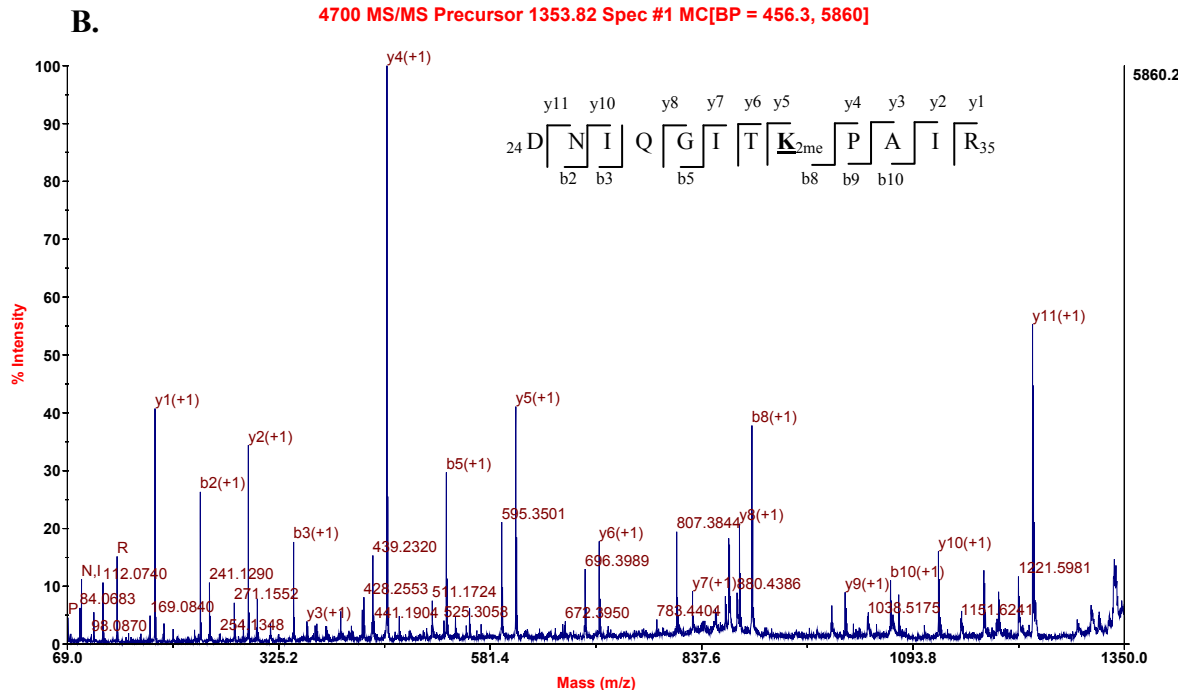
Although previous studies have demonstrated the genotoxic and mutagenic effects of formaldehyde, the fact that all lysine residues are the targets of formaldehyde raises the possibility of an additional mode of action for its toxicity. Formaldehyde could alter epigenetic regulation in which histone modifications occurring on lysine play a central role. We have clearly shown that formaldehyde-induced Schiff bases inhibit the formation of PTM on lysine. Therefore, these formaldehyde-induced lysine adducts on histone may impair the PTM pattern and possibly disturb the subsequent recruitment of specific proteins highly associated with the PTM pattern, triggering a series of abnormal cascade effects. Additionally, the balance between histone acetylation and deacetylation could be disturbed by the attachment of formaldehyde on lysine residues. The balance between histone

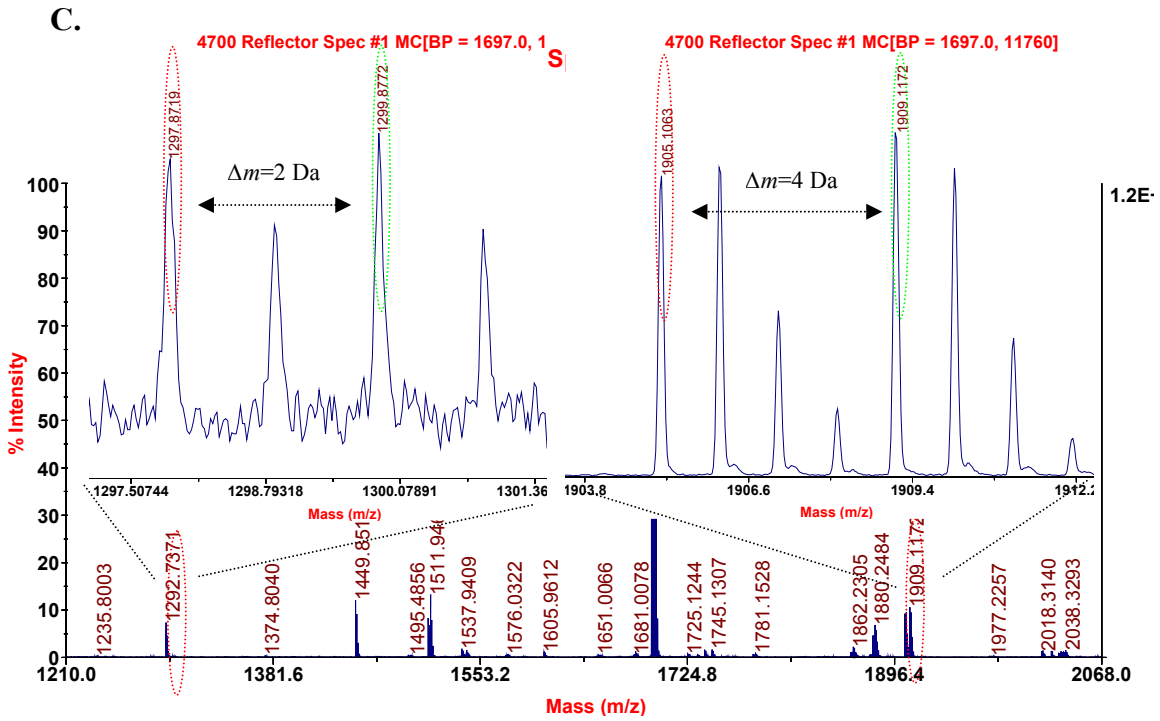
acetylation and deacetylation is important for normal cell growth, and an imbalance of acetylation in promoter regions may induce the deregulation of gene expression. Accumulating evidence has linked imbalances between acetylation and deacetylation to carcinogenesis and cancer progression (*20-22*). However, our finding that formaldehyde-induced Schiff bases inhibit the formation of PTM is based on a simplified *in vitro* model. Furthermore, the Schiff bases are generally reversible in nature. Therefore, additional experiments in cells or tissues will be needed to demonstrate that such effects occur in biology.

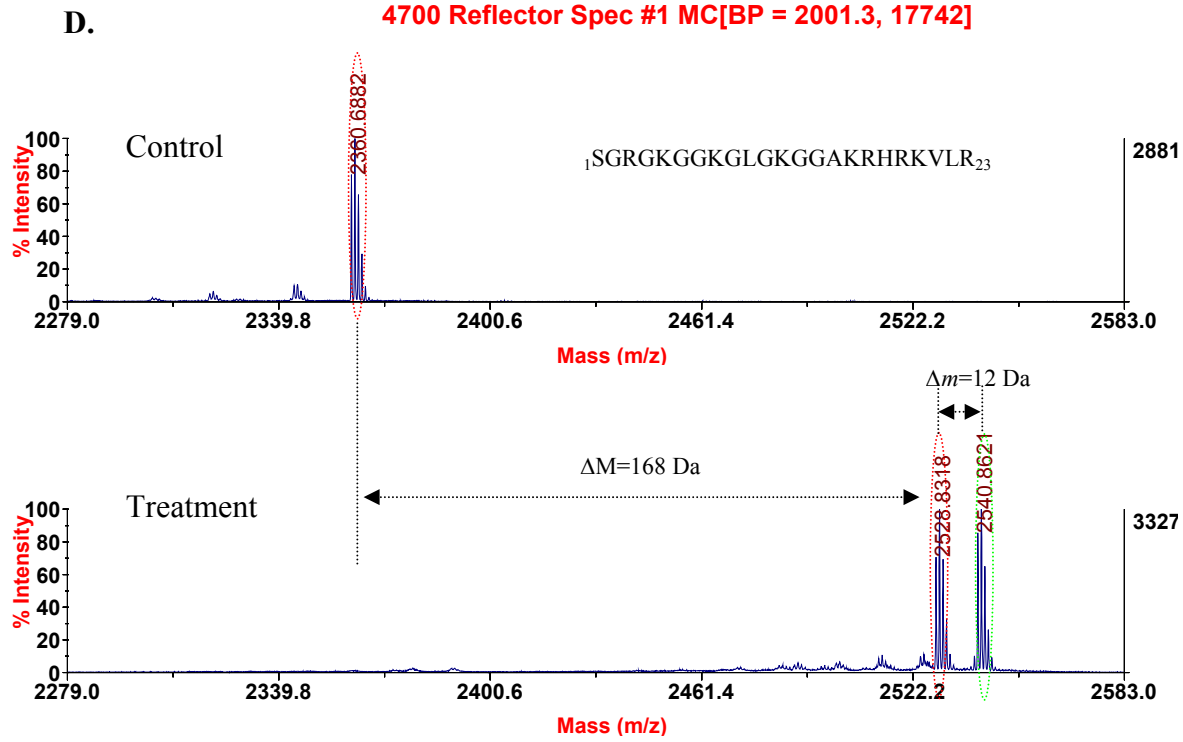
In conclusion, identification of reactive sites on histone is an initial step in understanding the mechanisms of formaldehyde toxicity and carcinogenicity, with many questions remaining to be elucidated. What is the structure and fate of formaldehyde-induced lysine adducts? What is the biological impact of formaldehyde-induced modification on each lysine residue? Where are the exact sites for the formation of DPC? What are the toxicological mechanisms of DPC? How are DPC repaired? Further *in vitro* and *in vivo* experiments should be carried out in order to shed light on these intriguing questions.

Figures


A.
4700 Reflector Spec #1 MC[BP = 1355.8, 30176]
1353.8389
1355.8462
Δm=2 Da
n =Δm/2=1
% Intensity
Mass (m/z)
3.0E4


B.
4700 MS/MS Precursor 1353.82 Spec #1 MC[BP = 456.3, 5860]
y4(+1)
y11 y10 y8 y7 y6 y5 y4 y3 y2 y1
24 D N I Q G I T K2me P A I R35
b2 b3 b5 b8 b9 b10
y11(+1)
y1(+1)
y5(+1)
y2(+1)
b8(+1)
b5(+1)
b2(+1)
595.3501
807.3844
y8(+1)
b3(+1)
y6(+1)
696.3989
y10(+1)
1221.5981
b10(+1)
y7(+1)880.4386
y9(+1)
1038.5175
1151.6241
672.3950
783.4404
439.2320
428.2553
511.1724
441.1904
525.3058
y3(+1)
241.1290
271.1552
254.1348
169.0840
112.0740
84.0683
98.0870
R
N,I
P
% Intensity
Mass (m/z)
5860.2


C.
4700 Reflector Spec #1 MC[BP = 1697.0, 1
4700 Reflector Spec #1 MC[BP = 1697.0, 11760]
1297.8719
1299.8772
1905.1063
1909.1172
Δm=2 Da
Δm=4 Da
% Intensity
Mass (m/z)
1235.8003
1292.7371
1374.8040
1449.851
1495.4856
1511.94
1537.9409
1576.0322
1605.9612
1651.0066
1681.0078
1725.1244
1745.1307
1781.1528
1862.2305
1880.2484
1909.1172
1977.2257
2018.3140
2038.3293
1210.0
1381.6
1553.2
1724.8
1896.4
2068.0


D.
4700 Reflector Spec #1 MC[BP = 2001.3, 17742]
Control
2360.6882
1SGRGKGGKGLGKGGAKRHRKVLR23
2881
Δm=12 Da
ΔM=168 Da
Treatment
2528.8318
2540.8621
3327
% Intensity
Mass (m/z)
2279.0
2339.8
2400.6
2461.4
2522.2
2583.0

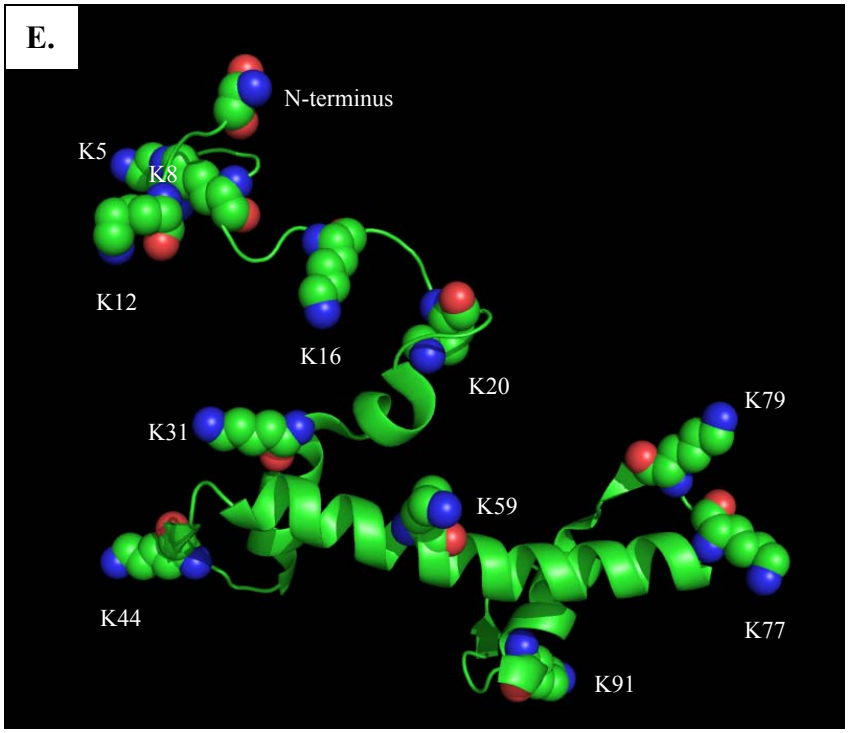


Figure 2.1. Mass spectra of formaldehyde-modified histone 4 peptides. (A) The formation of a doublet separated by 2 Da after native formaldehyde and 13C labeled formaldehyde treatment. (B) MALDI-TOF/TOF MS/MS spectrum of the precursor ion at 1353.82 Da with the identified structure as $_{24}$DNIQGITK$_{2me}$PAIR$_{35}$. (C) MALDI-TOF MS spectrum of two doublets, 1297.87-1299.87 and 1905.10-1909.11, obtained after chymotrypsin proteolysis. (D) MALDI-TOF Mass spectrum of unmodified N-terminal peptide observed at 2360.68 Da after Asp-N cleavage (Control panel) and formaldehyde-modified N-terminal doublet separated by 12 Da after Asp-N digestion (Treatment panel). (E) The representation of reactive sites including lysine and the N-terminus of histone 4. The potential reactive residues are shown with a sphere model, constructed based on a 1.9 Å crystal structure (PDB Code:1KX5) and rendered with PyMOL (DeLano Scientific LLC, Palo Alto, CA)

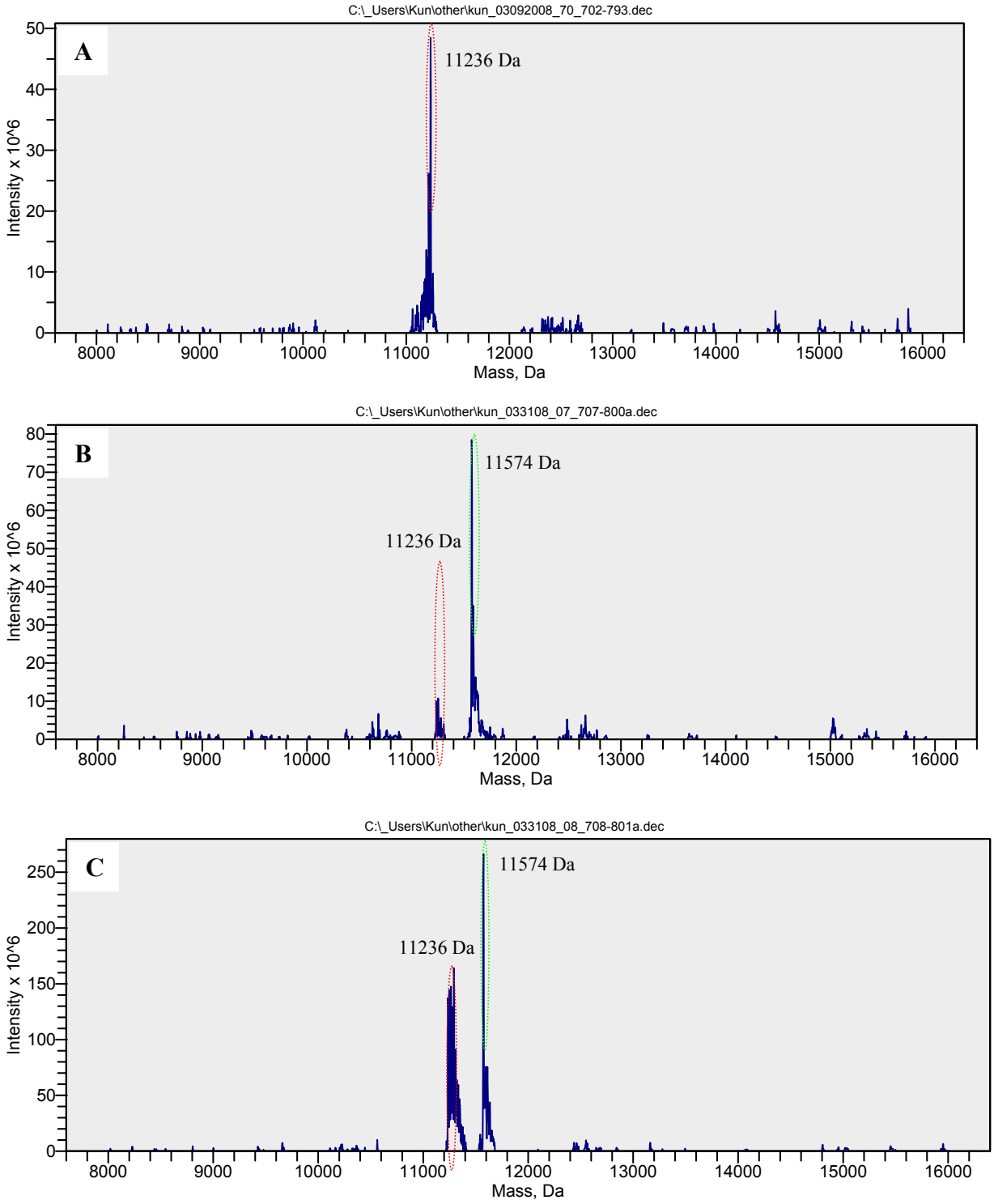


Figure 2.2. The molecular weight of untreated histone 4 and formaldehyde-treated histone 4. (A) The mass spectrum of histone 4 after deconvoluting major peaks from ESI-MS, showing the molecular weight of histone 4 equal to 11236 Da. (B) The deconvoluted mass spectrum for the 5mM formaldehyde-treated histone 4 sample. (C) The deconvoluted mass spectrum for the 50 μM formaldehyde-treated histone 4 sample.

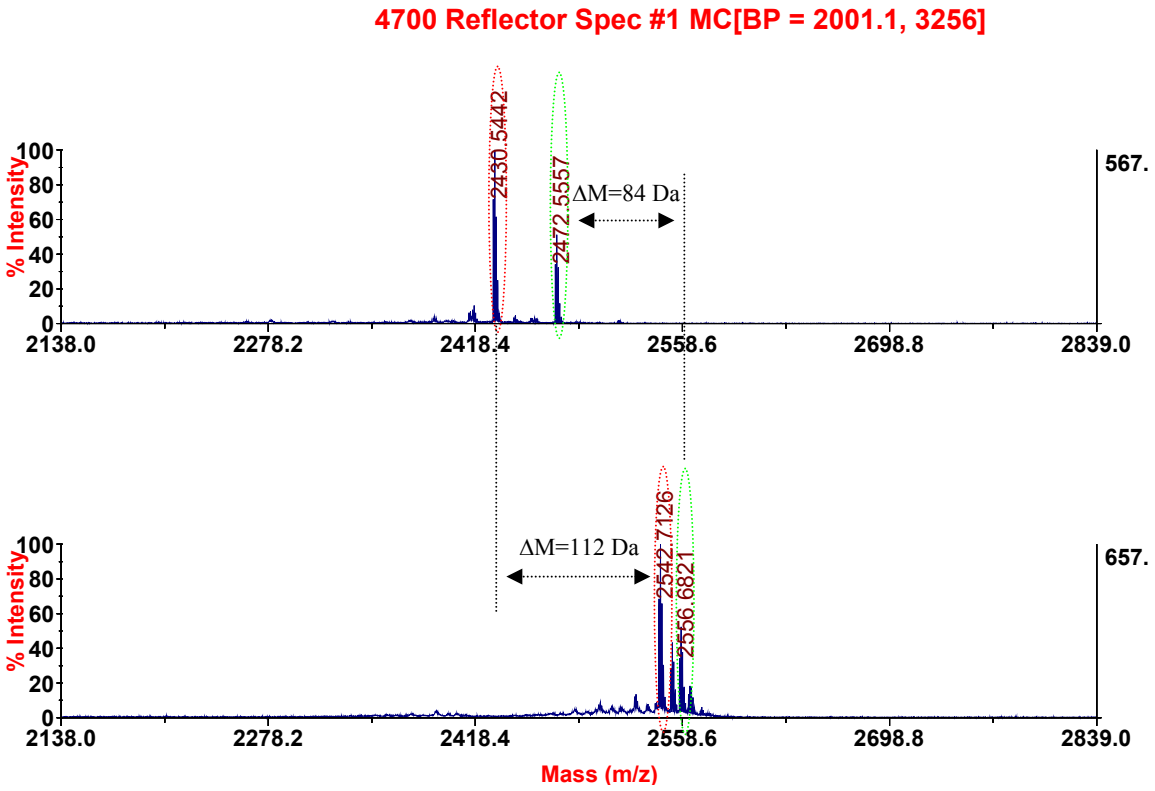
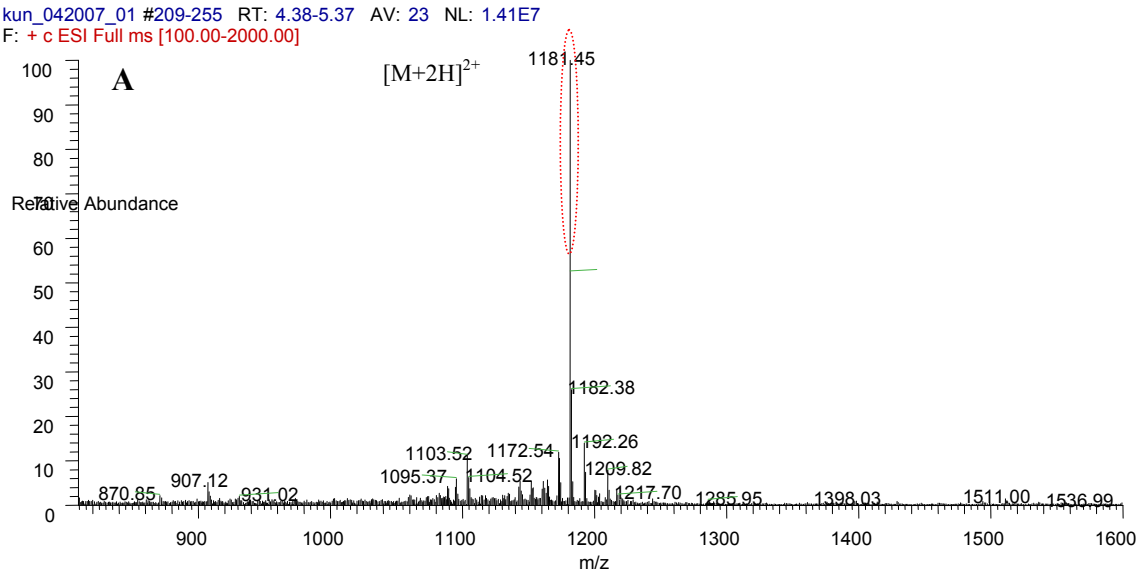
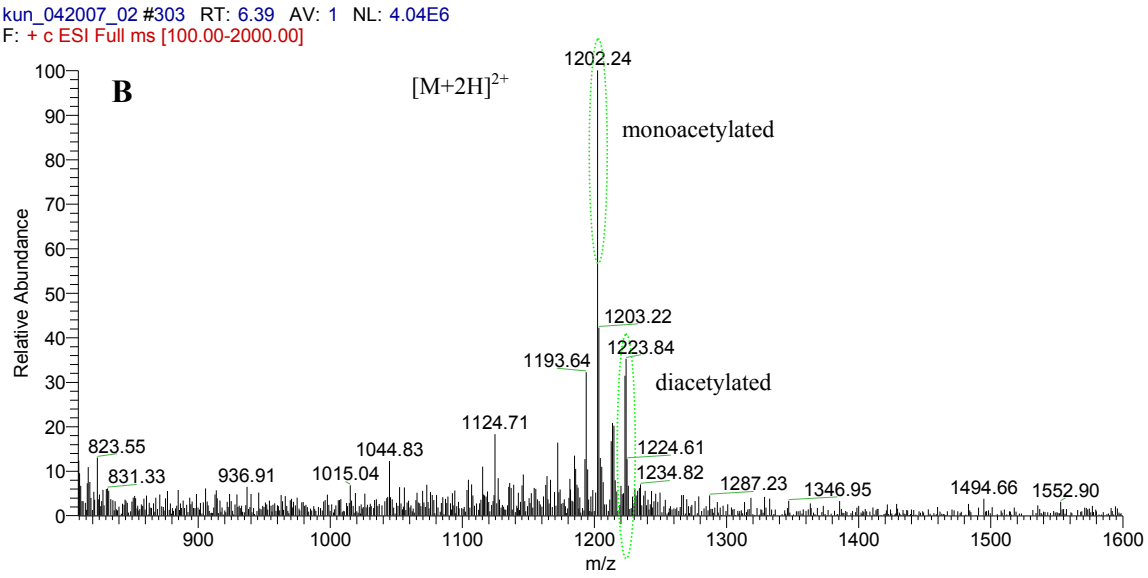


Figure 2.3. MALDI-TOF MS spectrum of two abundant histone 4 N-terminal peptides with PTM observed at 2430.54 and 2472.55 Da after Asp-N cleavage prior to formaldehyde treatment (Control panel) and MS spectrum of two formaldehyde-modified N-terminal peptides observed at 2542.71 and 2556.68 Da after Asp-N digestion (Treatment panel).


kun_042007_01 #209-255 RT: 4.38-5.37 AV: 23 NL: 1.41E7
F: + c ESI Full ms [100.00-2000.00]
A
[M+2H]2+
1181.45
1182.38
1192.26
1209.82
1217.70
1172.54
1103.52
1104.52
1095.37
907.12
870.85
931.02
1285.95
1398.03
1511.00
1536.99
Relative Abundance
m/z


kun_042007_02 #303 RT: 6.39 AV: 1 NL: 4.04E6
F: + c ESI Full ms [100.00-2000.00]
B
[M+2H]2+
1202.24
monoacetylated
1203.22
1223.84
diacetylated
1193.64
1124.71
1224.61
1234.82
1044.83
1015.04
823.55
831.33
936.91
1287.23
1346.95
1494.66
1552.90
Relative Abundance
m/z

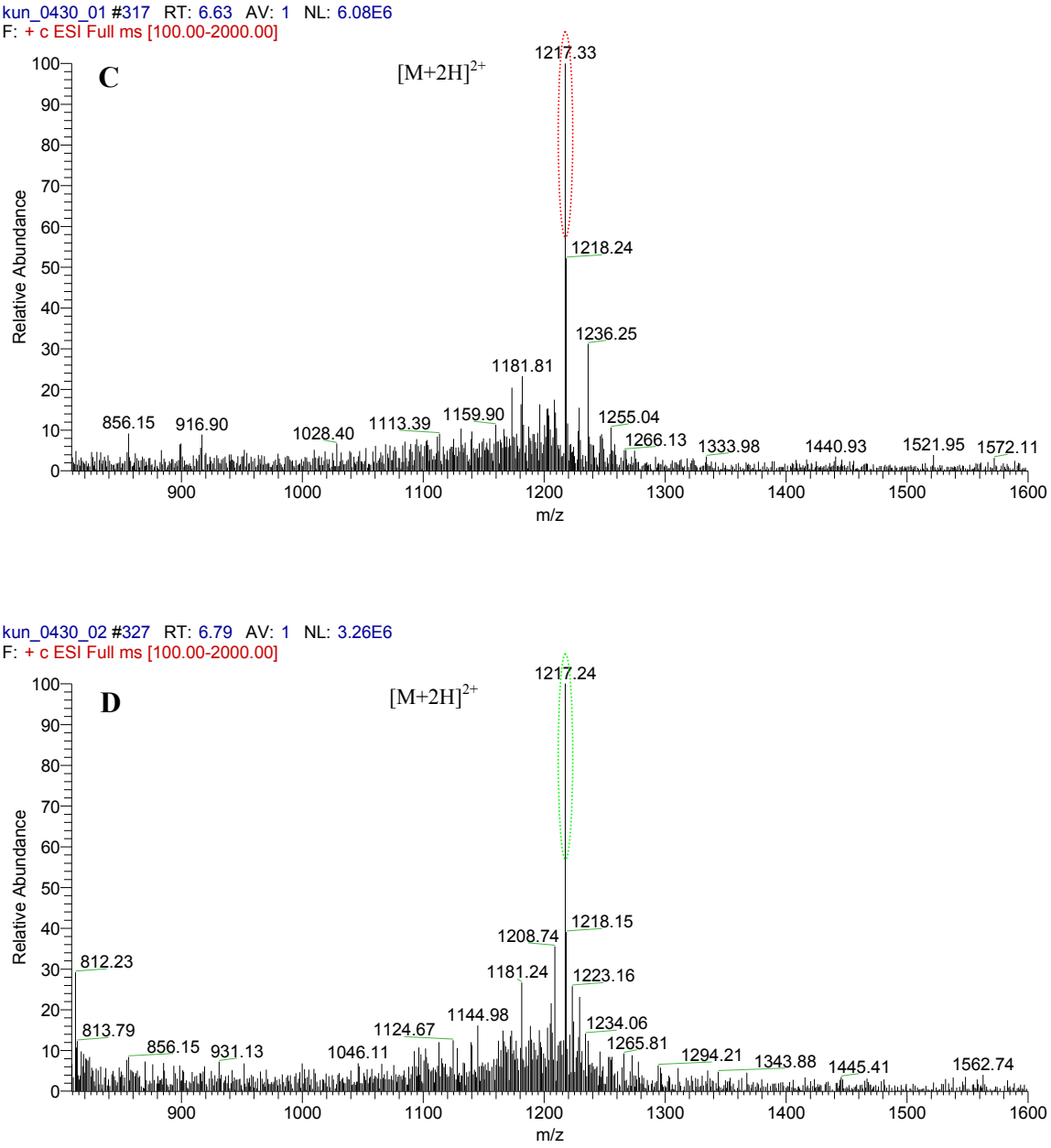
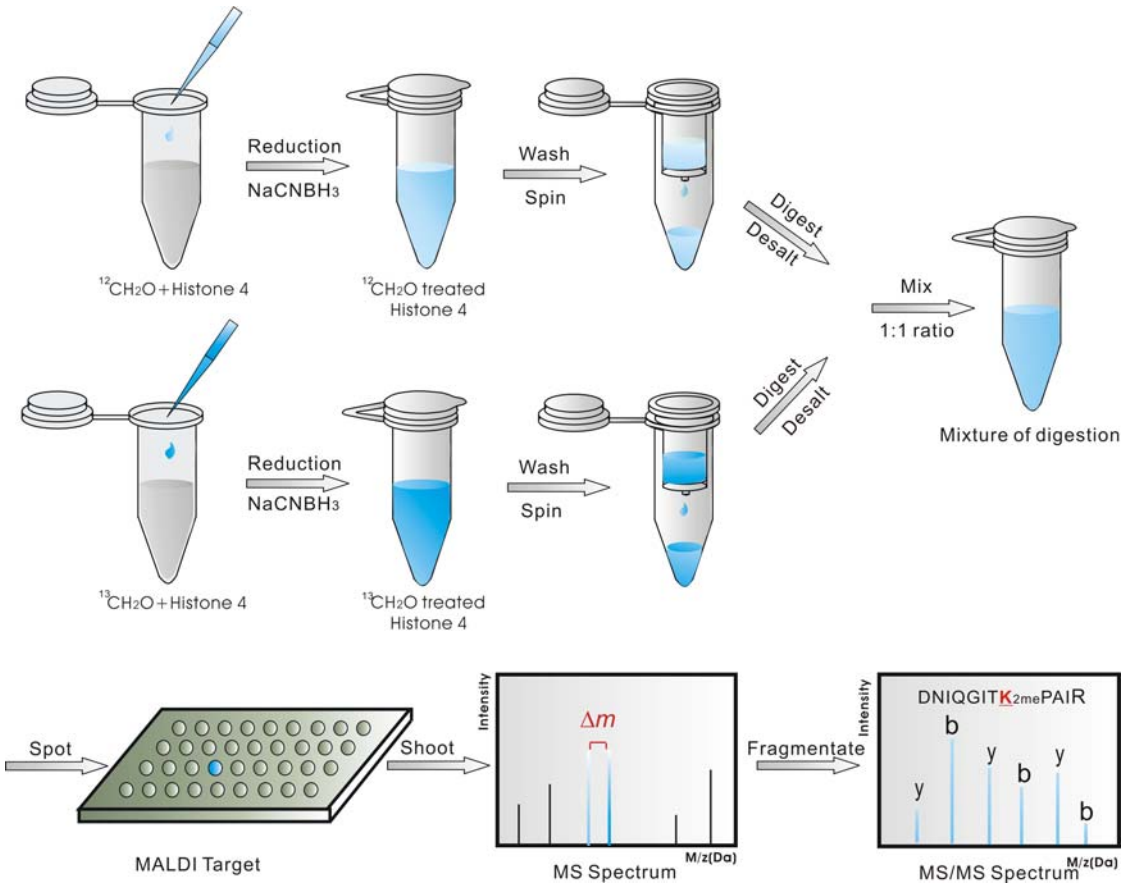


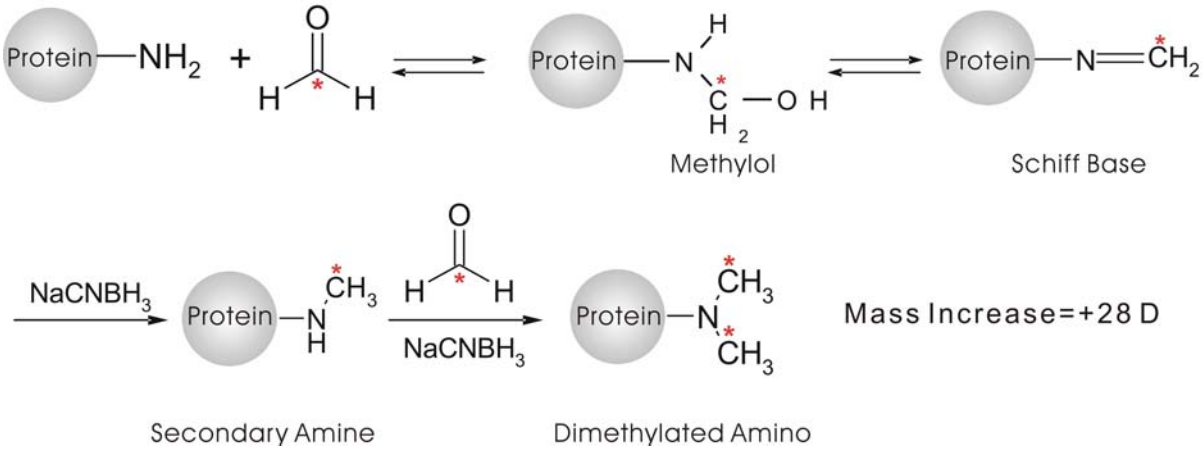
Figure 2.4. Impact of formaldehyde-induced modification on PTM formation. (A) ESI mass spectrum of a synthetic histone 4 N-terminal peptide (amino acid 1-23). A doubly charged ion is shown in the spectrum. (B) ESI mass spectrum of the PCAF enzyme-treated histone 4 N-terminal peptide. The fragments at 1202.24 m/z and 1223.84 m/z are monoacetylated and diacetylated histone 4 N-terminal peptides. (C) ESI mass spectrum of the formaldehyde-modified histone 4 N-terminal peptide. The 36 m/z increase is attributed to the formation of 6 Schiff bases on lysine and the N-terminus. (D) ESI mass spectrum of the formaldehyde-modified histone 4 N-terminal peptide treated by PCAF enzyme. The precursor ion at 1217.31 m/z presents the same mass with formaldehyde-modified histone 4 N-terminal peptide.

Schemes

Scheme 2.1. The analytical approach used in this study.



Scheme 2.2. The formation of a Schiff base and its reduction to a dimethylated group by NaCNBH3.



Tables

Table 2.1. Peptide fragments of histone 4 observed by mass spectrometry

No	Peptide fragments	Calculated (Da)	Measured (Da)	Δ(ppm)
1[a]	$_{96}$TLYGFGG$_{102}$	714.34	714.35	19.4
2	$_{60}$VFLENVIR$_{67}$	989.57	989.58	10.3
3	$_{46}$ISGLIYEETR$_{55}$	1180.62	1180.66	36.2
4	$_{68}$DAVTYTEHAK$_{77}$	1134.54	1134.55	9.07
5	$_{80}$TVTAMDVVYALK$_{91}$	1310.70	1310.71	6.71
6	$_{24}$DNIQGITKPAIR$_{35}$	1325.75	1325.77	14.5
7[b]	$_{91}$KRQGRTLV$_{98}$	1021.59	1021.63	40.5
8	$_{89}$ALKRQGRTL$_{97}$	1042.64	1042.69	42.4
9	$_{50}$IYEETRGVL$_{58}$	1079.57	1079.61	42.7
10	$_{89}$ALKRQGRTLY$_{98}$	1205.71	1205.76	43.7
11	$_{38}$ARRGGVKRISGL$_{49}$	1269.78	1269.84	45.9
12	$_{62}$LENVIRDAVTY$_{72}$	1292.68	1292.74	46.4
13	$_{38}$ARRGGVKRISGLIY$_{51}$	1545.93	1546.01	53.4
14	$_{59}$KVFLENVIRDAVTY$_{72}$	1666.91	1667.01	57.8
15	$_{23}$RDNIQGITKPAIRRL$_{37}$	1751.03	1751.12	48.9
16	$_{73}$TEHAKRKTVTAMDVVY$_{88}$	1848.96	1849.07	59.7
17	$_{38}$ARRGGVKRISGLIYEETRG$_{58}$	2330.34	2330.50	71.1
18[c]	$_{68}$DAVTYTEHAKRKTVTAM$_{84}$	1921.98	1922.03	28.7
19	$_{85}$DVVYALKRQGRTLYGFGG$_{102}$	2000.07	2000.11	19.7
20	$_{1}$SGRGKGGKGLGKGGAKRHRKVLR$_{23}$	2360.43	2360.43	15.5
21[d]	$_{1}$S$_{ac}$GRGKGGKGLGKGGAKRHRK$_{2me}$VLR$_{23}$	2430.47	2430.54	28.8
22	$_{1}$S$_{ac}$GRGKGGKGLGKGGAK$_{ac}$RHRK$_{2me}$VLR$_{23}$	2472.48	2472.55	28.3

a: Peptides 1-6 were obtained after trypsin digestion.

b: Peptides 7-17 were yielded by chymotrypsin cleavage.

c: Peptides 18-22 were observed after Asp-N proteolysis.

d: Peptides 21-22 were obtained from histone 4 with PTM

Table 2.2. Formaldehyde-modified histone 4 peptide fragments observed by MALDI-TOF mass spectrometry

No.	Observed Doublets		Sequence	ΔM (Da)	Δ*m* (Da)	*n*
1 [a]	1353.85	1355.86	$_{24}$DNIQGIT**K**$_{2me}$PAIR$_{35}$	+28	2	1
2	1414.96	1416.96	$_{56}$GVL**K**$_{2me}$VFLENVIR$_{67}$	+28	2	1
3	1651.07	1655.08	$_{79}$**K**$_{2me}$TVTAMDVVYAL**K**$_{2me}$R$_{92}$	+56	4	2
4 [b]	1297.87	1299.87	$_{38}$ARRGGV**K**$_{2me}$RISGL$_{49}$	+28	2	1
5	1905.10	1909.11	$_{73}$TEHA**K**$_{2me}$R**K**$_{2me}$TVTAMDVVY$_{88}$	+56	4	2
6 [c]	2528.83	2540.86	$_{1}$S$_{2me}$GRG**K**$_{2me}$GG**K**$_{2me}$GLG**K**$_{2me}$GGA**K**$_{2me}$RHR**K**$_{2me}$VLR$_{23}$	+168	12	6

ΔM: mass increase of peptide upon formaldehyde treatment

Δ*m:* mass difference between the doublets

n: the number of reactive lysine residue and N-terminal amino group

a: Peptides 1-3 were obtained after trypsin digestion.

b: Peptides 4-5 were yielded by chymotrypsin cleavage.

c: Peptide 6 were observed after Asp-N proteolysis.

54

References

(1) International Agency for Research on Cancer. (1995) IARC Monographs on the evaluation of carcinogenic risks to humans.

(2) Speit, G., Schutz, P. and Merk,O. (2000) Induction and repair of formaldehyde-induced DNA-protein crosslinks in repair-deficient human cell lines. *Mutagenesis*, *15*, 85-90.

(3) Craft,T.R., Bermundez,E. and Skopek,T.R. (1987) Formaldehyde mutagenesis and formation of DNA-protein crosslinks in human lymphoblasts in vitro. *Muta. Res.*, *176*, 147-155

(4) Merk,O. and Speit,G. (1998) Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ. Mol. Mutagen*, *32*, 260-268

(5) Shaham,J., Bomstein,Y., Meltzer, A., Kaufman,Z., Palma,E. and Ribak,J. (1996) DNA-protein crosslinks, a biomarker of exposure to formaldehyde—*in vitro* and *in vivo* studies. *Carcinogenesis*, *17*, 121-125

(6) Shaham,J., Bomstein,Y., Gurvich,R., Rashkovsky,M. and Kaufman,Z. (2003) DNA-protein crosslinks and p53 protein expression in relation to occupational exposure to formaldehyde. *Occup. Environ. Med.*, *60*, 403-409

(7) Quievryn,G. and Zhitkovich,A. (2000) Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis*, *21*, 1573-1580.

(8) Solomon,M.J. and Varshavsky,A. (1985) Formaldehyde-mediated DNA-protein crosslinks: A probe for in vivo chromatin structures. *Proc. Natl. Acca. Sci. U.S.A*, *82*, 6470-6474

(9) Siomin,Y.A., Simonov,V.V. and Poverenny,A.M. (1973) The reaction of formaldehyde with deoxynucleotides and DNA in the presence of amino acids and lysine-rich histone. *Biochim. Biophys. Acta, 331*, 27-32

(10) Jackson,V. (1999) Formaldehyde crosslinking for studying nucleosomal dynamics. *Methods, 17*, 125-139

(11) Davery,C.A., Sargent,D.F., Lunger,K., Maeder, A.W. and Richmond,T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol*., *319*, 1097-1113

(12) Luger,K., Mader,A. W., Richmond,R. K., Sargent, D. F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particles at 2.8 Å resolution. *Nature*, *389*, 251-260.

(13) Rosa,H.S. and Caldas,C. (2005) Chromatin modifier enzymes, the histone code and cancer. *Eur. J. Cancer, 41*, 2381-2402

(14) Metz,B., Kersten,G.F.A., Hoogerhout,P., Brugghe,H.F., Timmermans,H.A.M., de Jong,A., Meiring, H., Hove,J. T. and Hennink,W. E. (2004) Identification of formaldehyde-induced modifications in proteins: reaction with model peptides. *J. Biol. Chem*., *279*, 6235-6243

(15) Means,C.E. and Feeney,R.E. (1995) Reductive alkylation of protein. *Anal. Biochem.*, *224*, 1-16

(16) Zhang,K., Williams,K.E., Huang, L., Yau,P., Siino,J.S., Bradbury,E.M., Jones,P.R., Minch,M.J. and Burlingames,A.L. (2002) Histone acetylation and deacetylation: identification of acetylation and methylation sites of Hela histone 4 by mass spectrometry. *Mol. & Cell. Proteomics*, *1*, 500-508

(17) Cosgrove,M.S. and Wolberger, C. (2005) How does the histone code work?. Biochem. *Cell Biol*., *83***,** 468-476

(18) Peterson,C.L. and Laniel,M. A. (2004) Histone and histone modifications, *Curr. Biol.*, *14*, 546-551

(19) Tsukada,Y., Fang,J., Erdjument-Bromage,H., Warren,M.E., Borchers,C.H., Tempst,P. and Zhang, Y. (2006) Histone demethylation be a family of JmjC domain-containing proteins. *Nature*, *439*, 811-816

(20) Santos-Reboucas,C.B. and Pimentel,M.M. (2007) Implication of abnormal epigenetic patterns for human diseases. *Eur J Hum Genet*, *15*, 10-17

(21) Ducasse,M. and Brown,M. A. (2006) Epigentic aberrations and cancer. *Mol. Cancer*, *5*, 1-10

(22) Agalioti,T., Chen G. and Thanos,D. (2005) Deciphering the transcriptional histone acetylation code for a human gene. *Cell*, *121*, 381-392.

CHAPTER 3

CHARACTERIZATION OF FORMALDEHYDE-INDUCED DNA-PROTEIN CROSS-LINKS

3.1 Introduction

Formaldehyde is a ubiquitous environmental contaminant, with human exposures occurring during use in industrial processes such as the manufacture of resins, particle board, plywood, leather goods, paper, and pharmaceuticals and through emission as a vapor over the lifetimes of these products*(1)*. Formaldehyde also occurs endogenously as a normal metabolic intermediate in human cells as well as under certain pathological conditions such as oxidative stress. While formaldehyde is a known carcinogen and mutagen and has thus evoked serious health concerns*(2-6)*, its mode of action is not well understood. Among the interactions of formaldehyde considered to have biological relevance is the induction of DNA adducts*(7-13)*, protein modifications*(14-16)*, inter-strand DNA cross-links*(17-19)*, and DNA-protein cross-links (DPCs)*(20;21)*, with the formation of DPCs being considered as the primary genotoxic effect following exposure to formaldehyde*(2-6)*.The two routes shown in Scheme 3.1 may be responsible for the formation of DPCs. In pathway A, formaldehyde addition at a nucleophilic site on a protein is followed by cross linking between the resulting protein methylol adduct and a nucleophilic site on DNA, while in pathway B, addition at a nucleophilic site on a DNA base is followed by cross-linking between the DNA methylol adduct and a protein

residue. The formation of DPCs is favored by intimate interactions between DNA and proteins*(22)* and the lysine-rich DNA-binding histones have been reported to crosslink with DNA*(20;21)* in the presence of formaldehyde.

Key to understanding the toxicity of formaldehyde is the detection and quantitation of the DPCs as biomarkers for formaldehyde exposure, both at sites of contact and at sites removed from contact. Of techniques available for this purpose, those based on liquid chromatography coupled with mass spectrometry have shown great promise in recent years*(23)*. Selected ion monitoring (SIM), selected reaction monitoring (SRM), or multiple reaction monitoring (MRM) provide ideal sensitivity and specificity for the quantitation of DNA adducts. To take advantage of these techniques, the molecular structures of the DPCs must be characterized and standards generated. In a previous communication, we described a formaldehyde-derived cross-link formed between the Cys sulfhydryl group of glutathione and dG, and suggested that this reaction may be relevant to systemic effects of formaldehyde transported as an *S*-hydroxymethylene conjugate of glutathione. In the present work, we have established that cross-linking reactions occur between the amino acids Lys, Cys, His and Trp and the nucleosides dG, dA or dC. We have characterized the nature of the cross-links formed in the presence of formaldehyde between the reactive amino acids and trinucleotides d($T_1B_2T_3$) where B_2 is the target base G, A or C. We have also examined the cross links formed in the presence of formaldehyde between dG, dA and dC and *N*-terminal protected 8-mer peptides containing a single target residue at position 5. Finally, we have determined by NMR and mass spectrometric studies the structures of the cross-links formed between potentially cross-linking amino acids and dG, dA and dC in the presence of formaldehyde as shown in Chart 3.1.

3.2 Experimental Section

3.2.1 Chemicals

Potassium phosphate, ammonium bicarbonate, trifluoroacetic acid, formic acid, acetonitrile, methanol, N^{α}-(*tert*-butoxycarbonyl)-L-lysine, *tert*-butoxycarbonyl-L-cysteine, N^{α}-(*tert*-butoxycarbonyl)-L-histidine, N^{α}-(*tert*-butoxycarbonyl)-L-tryptophan, N^{α}-(*tert*-butoxycarbonyl)-L-arginine, N^{α}-(*tert*-butoxycarbonyl)-L-asparagine, N^{α}-(*tert*-butoxycarbonyl)-L-glutamine, N^{α}-(*tert*-butoxycarbonyl)-L-tyrosine, deoxyadenosine, deoxythymidine, deoxycytidine and deoxyguanosine were purchased from Sigma (St. Louis, MO). The *N*-terminal acetylated 8-mer peptides Ac-VEGGRGAA, Ac-VEGGQGAA, Ac-GEGGCGAA, Ac-GEGGYGAA, Ac-VEGGKGAA, Ac-VEGGNGAA, Ac-GEGGHGAA and Ac-GEGGWGAA were synthesized by GenScript Corporation (Piscataway, NJ). 20% Formaldehyde in water was purchased from Tousimis (Rockville, MD). $^{13}CH_2O$ was purchased from Cambridge Isotope Laboratories (Andover, MA). All chemicals were used as received unless otherwise stated.

3.2.2 Instrumental methods

Nuclear magnetic resonance (NMR) analysis. NMR Spectra were recorded on a Varian INOVA NMR spectrometer (Varian, Inc., Palo Alto, CA) at 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C spectra with the Varian cold probe.

Liquid Chromatography-Mass Spectrometry (LC-MS). LC-MS analyses were performed on a LCQ-Deca ion trap mass spectrometer (Thermo Electron, Waltham, MA) or a TSQ Quantum Ultra triple quadrupole mass spectrometer (Thermo Electron, Waltham, MA). The mass spectrometers were equipped with electrospray ionization

(ESI) sources. Analytes were separated by reverse phase HPLC using a 250 mm × 2.5 mm C18 analytical column from Grace Vydac (Hesperia, CA) eluted at 200 µL/min with a linear gradient from 2% to 60% solvent A in B over 15 min. Solvent A consisted of 0.1% formic acid in water and solvent B was methanol. The ion trap mass spectrometer was operated in full scan as well as dependent scan modes. For the fragmentation of precursor ions, the normalized collision energy of the ion trap mass spectrometer varied from 30% to 35% depending on the adduct. The activation time was set at 30 msec. The triple quadrupole mass spectrometer was operated in full scan, parent mode scan, MS/MS scan, selected ion monitoring (SIM) or selected reaction monitoring (SRM) mode. The collision energy was set at 17 V for most fragmentation experiments. High resolution mass spectra were obtained on an Agilent 6250 Accurate Mass Q-TOF LC/MS (Agilent Technologies, Santa Clara, CA) equipped with a dual spray ESI source. For liquid chromatography, a Hypersil Gold column (Thermo Scientific, Waltham, MA) (150 x 2.1 mm, 3 µm particle size) was used with a linear gradient from 2% acetonitrile in 0.1% formic acid to 80% acetonitrile over 15 min, eluted at 200 µL/min. The ESI source was set as follows: gas temperature, 350 ^{o}C; drying gas, 10 L/min; Vcap, 4000 V; nebulizer, 35 psig; fragmentor, 100 V; skimmer, 65 V. Fourier-transform ion cyclotron resonance mass spectra (FTICR-MS) (10 scans) were acquired on a hybrid Qe-Fourier transform ion cyclotron resonance mass spectrometer equipped with a 12 T actively shielded magnet (Apex Qe-FTICR-MS, 12.0 T AS, Bruker Daltonics, Billerica, MA, USA), and an Apollo II microelectrospray (μESI) source. The voltages on the µESI sprayer, interface plate, heated capillary exit, deflector, ion funnel and skimmer were set at 4.2 kV, 3.9 kV, 300 V, 250 V, 175 V and 30 V, respectively. The μESI source temperature was maintained at 180 ^{o}C. Desolvation was carried out by using a nebulization gas flow (2 bar) and a

countercurrent drying gas flow. Before transfer, ion packets were accumulated inside the collision cell for a duration of 0.02 second. Using a syringe pump (Cole Parmer, Vernon Hills, IL, USA), sample solutions were infused with a 250 μL syringe (Hamilton, Reno, NV, USA) at 90 μL/hour.

3.2.3 Formation of formaldehyde-induced cross links

Typical reaction conditions were as follows: 10 μL each of 5 mM amino acid and deoxynucleoside solution were mixed in 50 μL of 10 mM potassium phosphate buffer (pH 7.2). Formaldehyde was added to a final concentration of 50 mM. After 48 h, the reaction mixtures were either separated by reverse phase chromatography or analyzed by LC-MS. Lys-dG coupling reactions were carried out with 5, 50 and 100 mM formaldehyde. Reactions using 50 mM formaldehyde were run for 48, 60, 72, and 84 h. The coupling reaction was repeated with 50 mM ^{13}C-formaldehyde for 48 h.

Formaldehyde-induced coupling between peptides and deoxynucleosides was done in the same manner in 50 μL reaction volumes with final concentrations of 1.5 mM peptide and 5 mM deoxynucleoside. For coupling between amino acids and trinucleotides in 50 μL reaction volumes, final concentrations were 5 mM amino acid and 1 mM trinucleotide.

Large scale reactions between amino acids and deoxynucleosides for structural characterization by NMR were run with 20 mg of deoxynucleoside and 40 mg of N^{α}-Boc-amino acid in 5 mL 10 mM potassium phosphate buffer (pH 7.2) and 100mM formaldehyde for 12 hours to 1 week, monitored by HPLC until the chromatographic trace remained constant. The reaction mixtures were separated by semi-preparative HPLC and collected products were characterized by mass spectrometry (Table 3.2) and

NMR. For the Lys-dG reactions, HPLC fractions were collected on dry ice, lyophilized and then stored at – 80 ^{o}C until analysis. Exact masses of cross-linked products are tabulated in the text (Table 3.4). MS/MS data are presented in full as Supporting Information. ^{1}H and ^{13}C shifts are tabulated below.

3.2.4 NMR data of cross-links

2-Amino-6-(10-oxo-triazino[1,2-*a*]purin-7-yl)hexanoic acid (TPHA-1): ^{1}H NMR (500 MHz, DMSO-d_6) δ 7.91 (s,1H, H2), 7.87 (s, 1H, *N5H*, *J* = 2.1 Hz), 6.74 (d, 1H, Boc-$N^{\alpha}H$-, *J* = 4.7 Hz), 6.09 (dd,1H, H1′, *J* = 7 .8, 5.9 Hz), 4.89 (s, 2H, C8H_2), 4.33 (m, 1H, H3′), 4.24 (bs, 2H, C6H_2), 3.83 (m,1H, H4′), 3.75 (dt, 1, C$^{\alpha}H$, *J* = 8.1, 8.1, 4.7 Hz), 3.54 (dd, 2H, H5′, *J* = 11.7, 4.5 Hz), 3.47 (dd, 2H, H5″, *J* = 11.7, 4.5 Hz), ~2.47 (C$^{\varepsilon}H_2$ (overlaps DMSO-d_6)), ~2.50 (H2′ (overlaps DMSO-d_6)), 2.18 (ddd, 1H, H2″-TPHA-1, *J*=13.1, 5.9, 3.0 Hz), 1.34-1.48 (m, 2H, C$^{\delta}H_2$), 1.24-1.37 (m, 2H, C$^{\gamma}H_2$), 1.44-1.58 (m, 2H, C$^{\beta}H_2$), 1.35 (s, 9H, CH_3-Boc). (^{13}C NMR, 125 MHz, DMSO-d_6) δ 173.9 *C*OOH, 155.8 C10, 155.2 *C*OO-Boc, 150.2 C4a, 149.0 C3a, 134.9 C2, 115.5 C10a, 87.3 C4′, 82.1 C1′, 77.4 *C*-Boc 70.5 C3′, 61.4 C5′, 60.5 C8, 59.5 C6, 53.8 C$^{\alpha}$, 49.2 C$^{\varepsilon}$, 39.3 C2′, 30.8 C$^{\beta}$, 27.9 CH_3-Boc, 26.6 C$^{\delta}$, 22.8 C$^{\gamma}$.

2-Amino-6-(5-hydroxymethyl-10-oxo-triazino[1,2-*a*]purin-7-yl)hexanoic acid (TPHA-2): (^{1}H NMR, 500 MHz, DMSO-d_6) δ 7.96 (s,1H, H2), 6.93-6.99 (m,1H, *N*H-Lys), 6.19 (dd, 1H, H1′, *J* = 7.8, 5.9 Hz), 4.99 (d, 2H, *N*5CH_2OH, *J* = 1.3 Hz), 4.95 (s, 2H, C$H_2$8), 4.45 (s, 2H, C6H_2), 4.35 (m, 1H, H3′), 3.78-3.83 (m, 1H, H4′), 3.77-3.84 (m, 1H, C$^{\alpha}H$), 3.55 (dd, 1H, H5′, *J* = 11.6, 4.8 Hz), 3.49 (dd, 1H, H5, *J* = 11.6, 4.7 Hz), 2.58 (ddd, 1H, H2′, *J* = 13.2, 7.7, 5.9 Hz), 2.49 (m, 2H, C$^{\varepsilon}H_2$), 2.21 (ddd,1H, H2″, *J* = 13.2, 7.8, 3.1 Hz), 1.50-1.69 (m, 2H, C$^{\beta}H_2$), 1.36-1.50 (m, 2H, C$^{\delta}H_2$), 1.36 (s, 9H, CH_3-*t*-Boc),

1.28-1.34 (m, 2H, C$^{\gamma}H_2$). (^{13}C NMR, 125 MHz, DMSO-d_6) δ 173.9 *C*OOH-Lys, 156.0 C10, 155.4 *C*OO- *t*-Boc, 149.4 C4a, 148.2 C3a, 136.0 C2, 115.9 C1a, 87.4 C4′, 82.3 C1′, 77.6 C-*t*-Boc, 70.7 C3′, 68.5 *N*5*C*H$_2$OH, 63.9 C6, 61.5 C5′, 61.2 C8, 53.2 C$^{\alpha}$, 49.0 C$^{\varepsilon}$, 39.2 C2′, 30.4 C$^{\beta}$, 27.9 CH$_3$- *t*-Boc, 26.6 C$^{\delta}$, 22.9 C$^{\gamma}$.

Cys-CH$_2$-dG: ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.73 (s, 1H, COOH-Cys), 10.82 (s, 1H, N1*H*-dG), 7.99 (s, 1, H8), 7.11 (bs, 1H, *N*$_2$*H*-dG), 6.99-7.08 (m, 1H, *NH*-Cys), 6.16 (ψt, 1H, H1', *J* = 6.8 Hz), 4.48-4.57 (m, 2H, CH$_2$-linker, *J* = 10 Hz), 4.35 (td, 1, H3′, *J* = 3.2, 3.2, 6.1 Hz), 4.06-4.12 (m, 1H, C$^{\alpha}$*H*-Cys), 3.78-3.82 (m, 1H, H4), 3.55 (dd, 1H, H5′, *J* = 11.6, 4.8), 3.48 (dd, 1H, H5″, *J* = 11.62, 4.8 Hz), 3.03 (dd, 1H, C$^{\beta}H_{2a}$, *J* = 13.5, 4.47 Hz), 2.84 (dd, 1H, C$^{\beta}H_{2b}$, *J* = 13.5, 9.19 Hz), 2.62-2.65 (m, 1H, H2′), 2.21 (ddd, 1H, H2″, *J* = 13.1, 6.8, 3.2 Hz), 1.36 (s, 9H, C*H*$_3$-*t*-Boc. ^{13}C NMR (125 MHz, DMSO-d_6) δ 172.6 *C*OOH-Cys, 156.7 C6, 155.3 *C*OO-*t*-Boc, 151.9 C2, 150.0 C4, 136.1 C8, 120.0 C5, 87.8 C4′, 82.9 C1′, 78.2 *C*-*t*-Boc, 70.5 C3′, 61.5 C5′, 53.9 C$^{\alpha}$, 43.2 *C*H$_2$-linker, 39.3 C2′, 31.8 C$^{\beta}$, 28.0 *C*H$_3$-*t*-Boc.

Cys-CH$_2$-dA: (^{1}H NMR, 500 MHz, DMSO-d_6) δ 8.42 (bs, 1H, *N*^{6}*H*), 8.37 (s, 1H, H8), 8.26 (s, 1H, H2), 6.35 (ψt, 1H, H1′, *J* = 6.8 Hz), 6.01 (d, 1H, *NH*-Cys, *J* = 5.26 Hz), 4.66-4.77 (m, 1H, C*H*$_{2a}$-linker), 4.50-4.61 (m, 1H, C*H*$_{2b}$-linker), 4.40-4.44 (m, 1H, H3′), 3.85-3.89 (m, 1H, H4′), 3.61 (dd, 1H, H5′, *J* = 11.9, 4.3 Hz), 3.66-3.74 (m, 1H, C*H*$^{\alpha}$), 3.53 (dd, 1H, H5″, *J* = 11.9, 4.3 Hz), 2.96-3.07 (m, 2H, C$^{\beta}H_2$), 2.66-2.74 (m, 1H, H2′), 2.27 (ddd, 1H, H2″, *J* = 13.1, 6.8, 3.2 Hz), 1.33 (s, 9H, C*H*$_3$-*t*-Boc).). (^{13}C NMR, 125 MHz, DMSO-d_6) δ 172.2 *C*OOH-Cys, 154.6 *C*OO-*t*-Boc, 153.6 C6, 150.5 C2, 148.5 C4, 139.9 C8, 120.2 C5, 87.3 C4′, 82.7 C1′, 78.0 *C*-*t*-Boc, 70.4 C3′, 61.5 C5′, 55.2 C$^{\alpha}$, 43.0 *C*H$_2$-linker, 39.3 C2′, 34.7 C$^{\beta}$, 27.8 *C*H$_3$-*t*-Boc.

Cys-CH$_2$-dC: (^{1}H NMR, 500 MHz, DMSO-d_6) δ 8.59 (bs, 1H, $N^4\underline{H}$), 7.83 (d, 1H, H6, J = 7.5 Hz), 6.11 (ψt, 1H, H1′ J = 6.9 Hz), 5.986 (d, 1H, $N\underline{H}$-Cys, J = 4.2 Hz), 5.74 (d, 1H, H5, J = 7.4Hz), 4.48 (dd, 1H, CH_{2a}-linker, J = 13.4, 6.6 Hz), 4.26 (dd, 1H, CH_{2b}-linker, J = 13.4 6.5 Hz), 4.20-4.24 (m, 1H, H3′), 3.71-3.74 (m, 1H, H4′), 3.64-3.69 (m, 1H, CH^{α}), 3.53-3.60 (m, 2H, H5′,H5″), 2.98 (ddd, 2H, C$^{\beta}H_2$, J = 42.10, 13.5 4.3 Hz), 2.11 (ddd, 1H, H2′, J = 12.9, 6.9, 4.4; Hz), 1.93 (td, 1H, H2″, J = 12.9, 6.9, 6.4 Hz), 1.35 (s, 9H, CH_3-*t*-Boc). (^{13}C NMR, 125 MHz, DMSO-d_6) δ 171.1 *C*OOH-Cys, 162.5 C4, 154.7 C2, 154.3 *C*OO-*t*-Boc, 140.0 C6, 94.7 C5, 87.1 C4′, 84.5 C1′, 77.2 *C*-*t*-Boc, 69.5 C3′, 60.8 C5′, 55.2 C$^{\alpha}$, 42.3 *C*H$_2$-linker, 40.1 C2′, 35.1 C$^{\beta}$, 27.8 *C*H$_3$-*t*-Boc.

His-CH$_2$-dA: (^{1}H NMR, 500 MHz, DMSO-d_6) δ 8.83 (bs, 1H, $N\underline{H}$-His), 8.42 (s, 1H, H8), 8.34 (s, 1H, H2), 7.61 (s, 1H, H$^{\varepsilon 2}$-His), 6.96 (s, 1H, H$^{\delta 1}$-His), 6.4 (bs, 1, $N^6\underline{H}$), 6.36 (ψt, 1H, H1′, J = 6.7 Hz), 5.59 (bs, 2H, CH_2-linker), 5.32 (bs, 1H, OH3′), 5.12 (bs, 1H, OH5′), 4.39-4.43 (m, 1H, H3′), 3.84-3.89 (m, 1H, H4′), 3.83-3.87 (m, 1H, H$^{\alpha}$-His), 3.48-3.65 (m, 2H, H5′,H5″), 2.69-2.81 (m, 2H, C$^{\beta}H_2$-His), 2.70-2.75 (m, 1H, H2′), 2.27 (ddd, 1H, H2″, J = 12.8, 6.7, 2.8 Hz), 1.27 (s, 9H, CH_3-*t*-Boc). (^{13}C NMR, 125 MHz, DMSO-d_6) δ 173.4 *C*OOH-His, , 154.6 *C*OO-*t*-Boc, 148.7 C4, 140.23 C8, 139.3 C2, 138.2 C$^{\gamma}$-His, 136.1 C$^{\varepsilon 2}$-His, 119.5 C5, 115.5 C$^{\delta 1}$-His, 87.8 C4′, 83.7 C1′, 77.1 *C*-*t*-Boc, 70.6 C3′, 61.5 C5′, 54.0 C$^{\alpha}$-His, 49.9 *C*H$_2$-linker, 39.1 C2′, 30.2 C$^{\beta}$-His, 27.9 *C*H$_3$-*t*-Boc.

Trp-CH$_2$-dG: (^{1}H NMR, 500 MHz, DMSO-d_6) δ 7.90 (s, 1H, H8), 7.06 (d, 1H, H$^{\varepsilon 3}$-Trp, J=7.14 Hz), 6.91-6.98 (m, 1H, H$^{\eta 2}$-Trp), 6.50-6.58 (m, 2H, H$^{\zeta 2}$, H$^{\zeta 3}$-Trp), 6.14 (ψt, 1H, H1′, J=6.8 Hz), 6.24-6.06 (m, 1H, $N^2\underline{H}$), 5.28-5.33 (bs, 2H, CH_2-linker), 5.00-4.72 (bs, OH3′ + OH5′), 4.36-4.32 (m, 1H, H3′), 4.29-4.21 (m, 1H, CH^{α}), 3.79-3.83 (m, 1H, H4′), 3.68-3.66 (m, 1H, C$^{\beta}H_{2a}$), 3.58-3.44 (m, 3H, CβH_{2b}, + H5′,H5″), 2.51-2.61 (overlapping DMSO-d_6), 2.17-2.25 (m, 1H, H2″), 1.40 (s, 4H, CH_3-*t*-Boc1), 1.34 (s, 5H,

CH_3-*t*-Boc2). (^{13}C NMR, 125 MHz, DMSO-d_6) δ 156.3 C6, 152.8 *C*O_2H, 150.0 C4, 150.0 $C^{\delta 2}$, 135.5 C8, 129.6 $C^{\varepsilon 2}$, 128.1 $C^{\eta 2}$, 122.8 $C^{\varepsilon 3}$, 116.9 $C^{\zeta 3}$, 116.8 C5, 108.4 $C^{\zeta 2}$, 87.5 C4′, 82.8 C1′, 79.8 *C*H_2-linker, 79.0 *C*-*t*-Boc, 70.8 C3′, 61.5 C5′, 56.4 C^{α}, 45.3 C^{β}, 39.3 C2′, 27.9 *C*H_3-*t*-Boc.

3.3 Results and discussion

Eight amino acids previously reported to form stable adducts with formaldehyde[15] were investigated (as N^{α}-Boc derivatives) in coupling reactions with all four nucleosides to determine which would be of interest for characterization of reactions with oligonucleotides and oligopeptides. Results are summarized in Table 3.1. No cross-links could be detected with Arg, Gln, Tyr or Asn and consistent with previous studies *(10-13)*, the endocyclic nitrogen of dT did not form a coupling product with any of the amino acids.

The high level of cross-links formed by Lys is of particular interest because this residue is involved in extensive DNA-protein contacts and may therefore be considered highly likely to form cross-links *in vivo*. The second most abundant cross-link, which was between Cys and dG, may also have relevance for the active site of alkylguanine alkyltransferases*(24)* where a Cys residue can come into proximity with a formaldehyde adduct of guanine.

3.3.1 Trinucleotides cross-linked to N^{α}-Boc-protected amino acids.

Since dT did not form adducts (Table 3.1), we selected trinucleotides having G, A or C flanked by T as targets in cross-linking investigations. Elemental compositions of the deprotonated molecules $[M - H]^-$ of cross-linked products of the trinucleotides and

N^{α}-Boc-protected amino acids were determined by high resolution ESI-QTOF mass spectrometry and are given in Table 3.2. All trinucleotides yielded deprotonated molecules with summed masses expected for the trinucleotide + N^{α}-Boc protected amino acid + 12 mass units, consistent with formation of a methylene link between the trinucleotide and amino acid. The reaction of TGT with Lys and formaldehyde yielded two additional products: [T(TPHA-1)T], having a composition expected for the formation of two methylene linkages and [T(TPHA-2)T], having a composition expected for two methylene linkages and a hydroxymethylene adduct.

As discussed below, NMR analysis of the products with multiple-methylene linkages establishes T(TPHA-1)T as T(10-oxo-triazino[1,2-*a*]purin-7-yl)T−substituted 2-aminohexanoic acid and T(TPHA-2)T as the corresponding *N*5-hydroxymethyl adduct. Formation of triazinane rings has precedent in the intramolecular condensation of terminal amino nitrogens from two peptide residues with formaldehyde as well as from glycine-formaldehyde condensation with the guanidino moiety of Arg[15]. These products would be unlikely to form *in vivo*, due to much lower concentrations of formaldehyde.

Fragmentation of the reaction products between trinucleotides and amino acids was investigated by high resolution QTOF MS/MS to confirm that the target base in the second position was indeed the site of the cross-linking reaction. The high resolution data provide elemental compositions to support structural assignments of product ions. The product ion nomenclature applied for describing backbone fragmentation patterns of the trinucleotides follows the widely used convention given in Chart 3.2.

MS/MS spectra, acquired on an ion trap mass spectrometer, have been reported for all 64 possible unmodified trinucleotides[25]. While extensive sequential decomposition would be more likely in the ion trap than in the QTOF used in our work, the

fragmentation patterns observed for the modified trinucleotides show reaction sequences similar to those reported previously*(25)*. Without exception, the cross-linked trinucleotides formed singly charged anions. The MS/MS spectra of the cross-linked trinucleotides were characterized by initial cleavage of the coupling linkage, with the resulting product ions undergoing backbone fragmentation. With the exception of TAT cross-linked with His, none of the ions containing an intact base−methylene−amino acid linkage was the source of backbone fragmentations. Full MS-MS spectra and assignments of product ions for all cross-linked trinucleotides are presented as Supporting Information (Figures 3.1 – 3.8).

TGT derivatives. In addition to the w_1^- ion (dT-5′-P^-; *m/z* 321), a major ion in the MS/MS of all of the TGT derivatives was observed at nominal mass *m/z* 866 [(TGT – H + 12)$^-$]. This ion corresponds in composition to the Schiff base derivative of the trinucleotide at G, which is possible only at the exocyclic N^2 and confirms that the formaldehyde-induced cross-linking reactions of TGT involve the target G. The non sequence ion from loss of neutral TH from the Schiff base adduct of the trinucleotide is also common to all the MS/MS spectra of cross-linked TGT products. In the MS/MS spectrum of TGT—CH_2—Lys, bonds on either side of the methylene linkage cleave, leading to a prominent ion at *m/z* 874 resulting from the loss of the Schiff base adduct of Lys in addition to the ion at *m/z* 886. Backbone cleavages of the product ions at *m/z* 874 and 886 give rise to parallel series of sequence ions w_2^-, x_2^-, y_2^- and z_2^- separated by 12 mass units, as expected for source ions TGT^- and its Schiff base adduct at G. The MS/MS of TGT—CH_2—Lys in Figure 3.9 is illustrative of the data obtained from QTOF analysis.

The triazino ring of T(TPHA-1)T fragments to yield the product ions TGT^- and Schiff base adduct of the trinucleotide from which are derived the same parallel series of backbone fragmentations observed in the MS/MS spectrum of the singly bridged TGT-Lys product (Figure 3.1; formation of sodium adducts of the major ions in the MS/MS spectrum of T(TPHA-1)T should be noted). In the MS/MS of T(TPHA-2)T, the cross-linking structure fragments sequentially to give prominent ions corresponding to [M − H − hydroxymethylene]$^-$ (nominal *m/z* 1144) and [M − H − *N*-Boc − hydroxymethylene]$^-$ (nominal *m/z* 1044) in addition to the Schiff base adduct of TGT at *m/z* 866, which is progenitor of the sole series of observed sequence ions w_2^-, x_2^-, y_2^- and z_2^-.

The MS/MS spectrum of TGT—CH_2—Cys shows a single fragmentation pathway consistent with loss of Cys to give the Schiff base adduct of TGT at *m/z* 866 and formation of the sequence ions w_2^-, x_2^- y_2^- and z_2^- from the expected backbone fragmentations. The MS/MS spectrum of TGT—CH_2—Trp yielded product ions in low abundance. The Schiff base adduct of TGT and backbone cleavage ions w_2^-, x_2^- y_2^- and z_2^- were present at the expected nominal mass-to-charge ratios, however the accuracy of mass measurements was low.

TAT and TCT derivatives. Loss of N^{α}-Boc (*m/z* 1025) and (N^{α}-Boc + His) (*m/z* 870) were prominent ions in the MS/MS spectrum of TAT−CH_2−His. The ion at *m/z* 1025 gave the non sequence ion (M − H − TH)$^-$ and sequence ions w_2^- and x_2^- ions from backbone fragmentation in which the cross-link is intact (the only instance of backbone fragmentations with an intact cross-link). The Schiff base derivative from loss of His (*m/z* 870) yielded the parallel series of sequence ions w_2^- and x_2^-, fixing the point of cross link attachment at exocyclic N^6 of adenine. The MS/MS spectrum of the deprotonated molecule TAT−CH_2−Cys features $(y_2 - B_2)^-$, w_1^- and T^- (base peak) ions along with a

prominent ion at nominal mass *m/z* 545. However, no ion containing the Schiff base adduct of A was detected. The nominal mass *m/z* 545 corresponds to a 2′,3′-dideoxy-2′,3′-dehydro-AMP or 2′-deoxyadenosine-5′-phosphenate linked through a methylene bridge to N^{α}-Boc-His. The exact mass differs by +10.8 ppm from that calculated for the composition of the proposed structures, within sufficient tolerance to support the methylene bridge between Cys and the target N^6 of A. However, a backbone fragmentation pathway yielding ions of either of the proposed structures was not reported for any of the 64 possible unmodified trinucleotides or observed for any of the other cross-linked trinucleotides in this study. Thus, support for the proposed structure of the cross-link relies on the NMR studies of the formaldehyde-derived cross-link between dA and Cys discussed below. Although product ions in MS/MS spectrum of TCT−CH_2−Cys were in low abundance, the exact mass of the base peak lies within an acceptable tolerance (4.1 ppm) for a composition corresponding to the Schiff base adduct of trinucleotide which is consistent with a methylene bridge between the C-N^4 and the Cys sulfhydryl group. Sequence ions w_2^-, x_2^- and w_1^- were also present at the expected nominal masses, in addition to the non-sequence ion $(M - H - TH)^-$.

3.3.2 Peptides cross-linked to deoxynucleosides.

We reacted dG, dA and dC with formaldehyde and *N*-terminal acetylated 8-mers in which position 5 contained one of the four residues which had been established as targets for cross-linking in the screening reactions. The exact masses of the protonated molecules are given in Table 3.3. As in the case of the trinucleotides, all of the 8-mers yielded singly charged ions with compositions corresponding to formation of a methylene

cross-link, and the Lys-containing 8-mers yielded two additional products consistent with formation of the tricyclic cross linked structures TPHA-1 and TPHA-2.

By MS/MS, the protonated molecules undergo initial fragmentation of the glycosidic bond or the bridging structure, and the resulting product ions undergo backbone fragmentations similar to those observed for unmodified peptides, giving predominantly y_n, a_n and b_n ions*(26)*. Analysis of MS/MS spectra of the protonated molecules definitively established the site of formaldehyde-induced cross linking for all of the peptides with the exception of the His-containing 8-mer cross-linked with dA.

Backbone fragmentations in the MS/MS spectra of the Lys- and Trp-containing 8-mers with a single methylene link to dG are derived from the intermediate product ions in which the cross-link has broken to yield a peptide sequence containing the Schiff base of Lys (Scheme 3.2) or the 2-methylene indole derivative of Trp. Figure 3.10 illustrates the MS/MS of the singly-bridged Lys-containing octapeptide. MS/MS spectra of the cross-linked peptides along with identification of the fragment ions are given in Figures 3.11 − 3.18). The mass difference between the b_5 and b_4 ions corresponds to the Schiff base adduct of the target Lys (or 2-methylene-substituted Trp) at position 5, which establishes the point of attachment of the methylene bridge at the predicted target residues. The modification of the Lys and Trp residues is further confirmed by the identification of y_4 ions, and the determination that the mass differences between (b_5 + y_4) and the intermediate product ions yielding the backbone fragmentation series (m/z 742 for the Lys 8-mer and 758 for the Trp 8-mer) correspond to the mass of the modified residue. The product ions representing loss of the Schiff base adduct of dG from the cross-linked peptides establishes the exocyclic N^2 of dG as the second point of attachment of the methylene bridge, as discussed above for the cross-linked trinucleotides.

In the MS/MS spectrum of the Cys-containing 8-mer cross linked to dC, a series of low-abundance b_n ions originates from the intermediate product ion (*m/z* 697) after loss of the deoxyribose from the protonated molecule. The difference in mass between the b_5 and b_4 ions of this b_n series corresponds to the mass of the Cys−CH_2−C unit, confirming that the cross link is attached to the peptide at Cys. Additional support for the cross link to Cys is a second series of low abundance b_5, b_6 and b_7 ions having compositions consistent with backbone fragmentation of the product ion from cleavage of the C−S bond of methylene bridge which transforms Cys to an α-amidoacrylic acid residue. As required by this scheme, the difference in mass between the b_5 (*m/z* 412) and b_4 (*m/z* 343), which does not contain a modified residue, is 69 mass units, corresponding to α-amidoacrylic acid at position 5. Major backbone fragmentation series for the Cys-containing 8-mers cross-linked with dA and dG originate from the product ions subsequent to the loss of the Schiff base derivatives of the nucleosides, and therefore are not informative with respect to the site of attachment of the methylene bridge to the peptide. However, the b_n series from the intermediate product ions containing α-amidoacrylic acid at position 5 are present in the MS/MS spectra of both cross-linked 8-mers, fixing the methylene bridge attachment at Cys.

The MS/MS spectrum of the Lys-containing 8-mer linked by two methylene groups through formation of a fused triazino ring is consistent with initial fragmentation of the triazino ring followed by proton transfer to give Schiff base adducts of both dG and the 8-mer (nominal mass, *m/z* 742) according to Scheme 3.3, with the charge residing predominantly on the 8-mer fragment. The Schiff base adduct of dG loses deoxyribose to give an ion at *m/z* 164 in low abundance. The absence of an ion corresponding to $[MH - dG]^+$ (nominal mass *m/z* 730) accompanied by product ions from its backbone

fragmentation, such as observed above for a single cross-linking methylene, supports the triazino structure. The difference in mass between b_5 and b_4 ions corresponds to the Schiff base adduct at the terminal $-NH_2$ of the Lys residue as required by Scheme 3.3.

The MS/MS spectrum of the product the containing the tricyclic (TPHA-2) cross-linking structure with an *N*5 hydroxymethyl-substituted triazinopurine is dominated by two series of ions derived from peptide backbone fragmentations of the intermediate product ions shown in Scheme 3.4. The y_4, b_4 and b_5 ions can be identified for both series, and the compositions establish that the terminal $-NH_2$ of Lys has been incorporated into the linking structure. The sequential fragmentations of the protonated molecule which lead to the intermediate products in Scheme 3.4 are suggested by the presence of product ions both at *m/z* 164 for the protonated N^2-Schiff base adduct of G and at *m/z* 178 compatible with the protonated tricyclic base 1,N^2-ethano-G.

In the MS/MS spectrum of the His-containing 8-mer cross-linked to dA, a single series of backbone fragmentations arises from an intermediate product ion via loss of the Schiff base adduct of dA. This fragmentation pattern is not informative regarding the site of methylene attachment to the 8-mer, and the structure of the cross-link is inferred from the structures determined by NMR studies for cross-linked monomeric units dA$-CH_2-$His.

3.3.3 N^{α}-Boc-protected amino acids cross-linked to deoxynucleosides.

Formaldehyde-induced coupling reactions between the N^{α}-Boc-protected amino acids and nucleoside monomers were investigated on a scale which allowed detailed structural determination of the linkages by NMR spectrometry. Table 3.4 gives the exact masses and corresponding elemental compositions of the cross-linked products, MS/MS

spectra are given in Figures 3.19 – 3.26. As observed for the trinucleotide and 8-mer, coupling products of Lys with dG incorporated one or two methylene groups or two methylenes with a hydroxymethyl adduct. All of the Lys coupling products were highly labile and were isolated and stored at sub-ambient temperature. The lability of the Lys coupling products is consistent with a recent report that formaldehyde-induced DPCs involving the Lys- and Arg-rich major histones are hydrolytically unstable *(20)*. This observation may have a bearing on the surprising lack of Arg coupling products in this study. Although Arg, like Lys, is involved in a large proportion of DNA-protein contacts, the cross linked products may be too unstable towards hydrolysis to survive analytical procedures.

3.3.4 Characterization of formaldehyde-induced cross-links.

The structures of formaldehyde-induced cross-links formed between the amino acids and nucleosides are given in Chart 3.1.

Lys cross-links with dG. In the presence of formaldehyde, Lys and dG formed three coupling products with the structures given in Chart 3.1. All three structures were detected in coupling reactions run at the lowest formaldehyde concentration. The proportion of tricyclic and *N*-hydroxymethyl-substituted tricyclic nucleosides increased with increasing formaldehyde concentration and reaction time (confirmed with ^{13}C-formaldehyde), consistent with progressive incorporation of formaldehyde molecules into the initial coupling product (Figure 3.27). As described above, the products were all labile at ambient temperature and required care in isolation and characterization. The tricyclic structures TPHA-1 and TPHA-2 were definitively established by mass spectrometry and NMR. Lys-CH_2-dG appears to be too labile for characterization by

NMR and its assignment as a product of cross-linking is based on indirect evidence from the analyses below.

Exact mass measurements of major ions at *m/z* 526, 538 and 568 in the coupling reaction mixture by ESI-MS in the positive ion mode were compatible with the protonated molecules Lys- CH_2-dG, TPHA-1 and TPHA-2, respectively (Table 3.4). When ^{13}C-formaldehyde was used in the coupling reaction, the protonated molecules are observed at *m/z* 527, 540 and 571, confirming formaldehyde as the origin of the methylene linkers and the hydroxymethylene group.

Well-resolved peaks were collected by semi-preparative HPLC (Figure 3.28) and characterized by NMR. Despite the absence of any features such as tailing in the chromatographic trace indicative of on-column decomposition, the ^1H NMR spectra indicated that the fractions collected at 17.2 and 26.5 min were mixtures.

The peak eluting at 17.2 min contained a 1:1:1 dG/Lys/TPHA-1 mixture, while the peak at 26.5 min contained a 0.4:1 TPHA-1/ TPHA-2 mixture. In the early-eluting fraction, Lys and dG were readily identified by comparison of well resolved signals in critical regions of the spectrum with authentic standards and in the late-eluting fraction, signals of TPHA-1 and TPHA-2 were readily distinguished because of the difference in signal intensities and peak integrals. The absence of signals attributable to Lys or dG in the product collected at 26.5 min indicates that the triazino ring is stable under conditions of NMR data acquisition and thus the Lys and dG observed in the 17.2 min fraction arises from uncoupling of Lys-CH_2-dG. Thus, the composition of the fractions may be explained by co-chromatography of TPHA-1 and Lys-CH_2-dG at 17.2 min followed by uncoupling of Lys-CH_2-dG post-chromatography, and by the partial loss of hydroxymethylene from TPHA-2 in the 26.5 min fraction. Uncoupling in the 17.2 min

fraction is supported by a strong signal in the ^{1}H NMR spectrum attributable to formaldehyde or formaldehyde hydrate corresponding to ~ 0.5 mole formaldehyde/mole dG (Figure 3.29) from hydrolysis of the Lys-CH_2-dG cross-link in accord with reported reversibility of formaldehyde-induced DPCs. Similarly, a formaldehyde signal of comparable magnitude in the 26.5 min fraction (Figure 3.32) is consistent with hydrolytic cleavage of the N5-hydroxymethylene group of TPHA-2. The formaldehyde/formaldehyde hydrate assignment is based on the appearance of a proton singlet below 8 ppm having unsuppressed 1-bond coupling with a carbon signal at 164.4 ppm that has no connectivity with any component in the HMBC spectra of the mixture (Figures 3.29, 3.32).

In each fraction, the signals of the components of the mixtures could be resolved, allowing the structures of TPHA-1 and TPHA-2 to be unambiguously assigned. Critical in establishing the triazine ring is the presence of two formaldehyde-derived methylenes and identification of the connectivities between the methylene groups, guanine and Lys moieties. Figures 3.33 and 3.34 show expansions of the HMBC spectra that are key to establishing the fused triazine linkages. In the HMBC spectrum of TPHA-1 (Figure 3.33), a broad two-proton singlet at 4.89 with three bond coupling to a ^{13}C signal at 59.5 ppm and a second two-proton singlet at 4.24 ppm with $^{3}J_{C\text{-}H}$ coupling to a ^{13}C signal at 60.5 ppm are assigned to methylene groups at positions 8 and 6, respectively, of the triazino[1,2-*a*]purine framework. Unsuppressed one-bond couplings, confirmed by the corresponding C/H cross peaks in the HSQC spectrum (Figure 3.30), are observed for both methylene signals and allow assignment of the methylene carbon shifts. The protons attached to the methylene carbon assigned to C8 show the expected connectivities within the tricyclic framework between C4a at 150.2 ppm and C10 at 155.8 ppm, while the

methylene protons at position 6 couples only with the carbon at C4a. Attachment of the hexanoic acid moiety at N7 is confirmed by cross peaks between the methylene protons at positions 6 and 8 and a carbon signal at 49.2 ppm, which can be assigned to hexanoic acid C^{ε} by virtue of unsuppressed one-bond coupling in the HMBC spectrum and the corresponding C/H cross peak in the HSQC spectrum. The complementary three-bond coupling between the methylene protons attached to C^{ε} (overlapping with the DMSO signal and sugar H2″ signals at ~ 2.5 ppm) and C6 and C8 is also observed.

By a similar analysis of the HMBC spectrum in Figure 3.34, connectivities can be identified to confirm the triazino[1,2*a*]purine framework of TPHA-2, the major component of the late-eluting peak. The critical cross peaks are between C8H_2 (4.95 ppm) and C6 (63.9 ppm), C4a (149.4 ppm), C10 (156.0 ppm) and C^{ε} (49.0 ppm); between C6H_2 at 4.45 ppm and C8 (61.5 ppm), C5a (149.8 ppm) and C^{ε} (49.2 ppm) and between $C^{\varepsilon}H_2$ (2.5 ppm, overlapping with DMSO) and C6 and C8. A methylene proton signal at 4.99 ppm can be assigned to the hydroxymethylene group with the position of attachment fixed at N5 by virtue of cross peaks with C6 and C4a. In the ^{1}H NMR spectrum (Figure 3.32), a singlet, assignable to formaldehyde by the rationale described above, strongly supports the conclusion that the presence of THPA-1 in the late-eluting fraction results from the hydrolytic elimination of the *N*5-hydroxymethylene substituent of TPHA-2.

The mechanism proposed for the dynamic nature of formaldehyde-induced Lys-dG cross links in Scheme 3.5 is supported by the lability of the products and the presence of formaldehyde in the NMR spectra of TPHA-1 and TPHA-2 which indicate that the condensations are reversible.

Cys cross links with dG, dA, dC. A single product was isolated from the cross-linking of Cys and dG with formaldehyde. In contrast to the cross-linked products of Lys, the product with Cys was stable and readily isolated and purified. The exact mass corresponded in elemental composition to addition of one methylene group. ESI-MS/MS of the protonated molecule (Figure 3.22) yielded product ions from loss of deoxyribose and *t*-Boc. The base peak of the MS/MS spectrum appears at *m/z* 164 (Figure 3.22), in accord with fragmentation to a Schiff base derivative of guanine, which indicates that the methylene cross-link bridges N^2 of guanine and the Cys sulfhydryl group.

NMR spectrometry confirmed guanine N^2 and Cys SH as the points of attachment of the methylene cross-link. The carbon and proton signals of formaldehyde-derived methylene cross-link are assigned from the HMBC spectrum on the basis of unsuppressed one-bond coupling between the carbon signal at 43.2 ppm and a 2-proton methylene multiplet centered at 4.52 ppm. Attachment at the Cys sulfhydryl is then fixed by coupling between the diastereotopic $C^{\beta}\underline{H_2}$ protons of Cys at 2.84 and 3.03 ppm and the formaldehyde-derived methylene bridging carbon, while attachment at guanine N^2 is fixed by coupling between the bridging methylene protons and C2 of guanine at 151.9 pm. The presence of the guanine imino N1$\underline{H}$ at 10.82 ppm in the ^{1}H NMR spectrum, unambiguously assigned from coupling with guanine C5 at 117 ppm in the HMBC spectrum (Figure 3.37), rules out attachment of the cross link at guanine N1. Attachment to Cys at N^{α} is ruled out by identification of Cys-$N^{\alpha}\underline{H}$ as one of two incompletely resolved proton signals at ~7.10 ppm based on cross peaks with Cys C^{α} (53.9 ppm) and C^{β} (31.8 ppm).

Cys formed stable cross-links with dA and dC in minor amounts. By high resolution mass spectrometry, both products had empirical compositions consistent with a

methylene bridge linking the nucleoside and Cys. For both products, critical C/H connectivities in the HMBC spectra (Figures 3.39, 3.41) were observed between the proton signals of the bridging methylene and a carbon signal of the base, and between the bridging methylene and Cys $^{\beta}CH_2$. In the case of Cys-CH_2-dA, attachment to dA at the exocyclic amino group was established by coupling between the diastereotopic methylene protons of the bridging group at 4.56 and 4.73 ppm and C6 of dA at 153.7 ppm (Figure 3.39). Attachment to the Cys sulfhydryl group was established by coupling between the diastereotopic methylene bridge protons and the C^{β} signal of Cys at 34.7 ppm. Identification of a one-proton $N^6\underline{H}$ signal confirmed the substitution at the exocyclic amino group of dA, while presence of the Boc-$N^{\alpha}\underline{H}$ proton signal ruled out attachment at the α-amino group of Cys. The cross link between dC and Cys was similarly determined to be between the exocyclic amino group of dC and the Cys sulfhydryl because of the observed coupling between the bridging methylene protons at 4.48 and 4.26 ppm and C4 of dC at 162.5 ppm and between the bridging methylene protons and C^{β} of Cys at 35.1 ppm (Figure 3.41). As in the case of Cys-CH_2-dA, the position of the cross link determined from the HMBC spectrum is supported one-proton signals attributable to $N^4\underline{H}$ of dA and Boc-$N^{\alpha}\underline{H}$.

His cross-link with dA. A low yield of stable cross-linked product was isolated from the reaction of His and dA with formaldehyde incorporating one methylene linkage identified by accurate mass measurement. Despite the absence of C,H cross-peaks between the linker methylene and His, attachment at $N^{\varepsilon 1}$ could readily be established by the presence of imidazole ring C-H proton signals $H^{\delta 1}$ at 6.96 and $H^{\varepsilon 2}$ at 7.61 ppm, which were definitively assigned by unsuppressed one-bond coupling to carbon signals at 115.5 and 136.1 ppm, respectively, in the HMBC spectrum (Figure 3.43). Site of attachment to

dA rests on the presence of a signal having an integral value of one proton assigned to $N^6\underline{H}$. A broad two-proton signal at 5.59 ppm in the ^{1}H NMR spectrum is replaced by a two-proton doublet centered at 5.59 ppm, $^1J_{C\text{-}H}$ = 153.2 Hz, when the cross-link is derived from ^{13}C-formaldehyde (Figure 3.44), and can be assigned to the formaldehyde-derived methylene bridge. This signal has the expected NOESY interactions with the imidazole protons in the ROESY spectrum (Figure 3.45).

Trp cross-link with dG. The Trp coupling product incorporated one methylene cross-link identified by accurate mass measurement. The site of attachment to Trp can be definitively fixed at C2 of the indole ring, and analogous to the reported formation of tetrahydro-β-carboline from Trp via intramolecular incorporation of a formaldehyde-derived methylene when the α-amino group is unprotected*(27;28)*. The proton signal of the formaldehyde-derived methylene forming the cross-link to dG is identified by a broad two-proton resonance at 5.30 ppm in the ^{1}H NMR spectrum attached to a carbon having a signal at 79.4 ppm in the HSQC spectrum (Figure 3.46) which is also detected by unsuppressed one-bond coupling ($^1J_{C\text{-}H}$ = 164.4 Hz; Figure 3.47) in the HMBC spectrum. Attachment of the methylene at indole C2 is supported by the detection of indole benzo-ring carbons as the only indole carbons having attached protons in the HSQC spectrum and by NOESY interactions between the bridge methylene and Trp$C^{\beta}H_2$ proton signals in the ROESY spectrum (Figure 3.48), which would not be observed if attachment of the methylene were at the indole nitrogen. While the proton-bearing carbons of the Trp indole ring were readily identified in the HSQC spectrum, overlapping of the quaternary carbon signals of both indole and dG moieties precludes establishing connectivity between the methylene bridge and dG from the HMBC spectrum. Attachment of the methylene bridge at the exocyclic amino group is inferred from the absence of a 2-proton

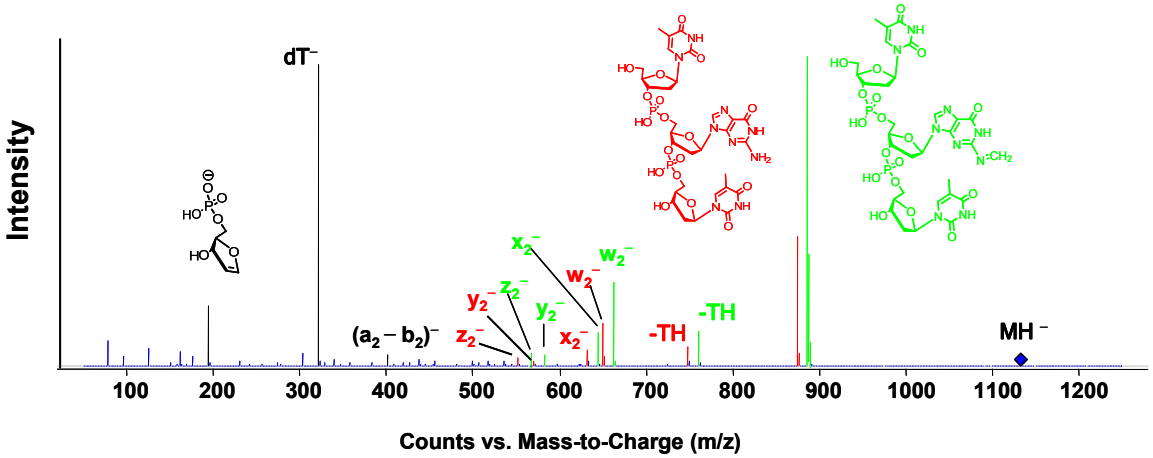
signal assignable to an unsubstituted exocyclic amino group of guanine and the fragmentation of the protonated molecule to yield the Schiff base of dG as the base peak in the MS/MS spectrum (Figure 3.26).

3.4 Conclusion

We have investigated the formation of formaldehyde-induced cross-links between nucleosides and amino acids and provide the first rigorous structural characterizations of DNA-protein cross-links induced by coupling with formaldehyde. Eight cross-linked structures have been characterized. In five of these structures, the cross-linking arose via a methylene bridge connecting the exocyclic amino group of a nucleoside to a nucleophilic nitrogen or sulfur of an amino acid side chain. In the case of Trp, the points of attachment were the N^2 exocyclic amino group of dG and C2 of the indole ring. Lys, in addition to forming an expected dG-N^2-CH_2-N^{ε}-Lys cross-link, formed two tricyclic adducts with dG, a 10-oxo-triazino[1,2-*a*]purinyl derivative incorporating two formaldehyde molecules and a 5-formyl-10-oxo-triazino[1,2-*a*]purinyl adduct incorporating three formaldehyde molecules. While the formation of the tricyclic adducts was favored at formaldehyde levels of 50 and 100 mM, all three adducts were identified at the lowest levels tested (5 mM). The three Lys cross-linked products were labile in solution, and the cross-link formed by a single methylene bridge was too unstable for characterization by NMR spectrometry. The lability of the Lys-dG cross-links support the reported reversibility of formaldehyde-induced cross-links formed *in vivo* between histones and DNA[20]. Our inability to detect products with cross-links between the nucleobases and Arg, which along with Lys is responsible for DNA–histone contacts, may reflect more pronounced dynamic reversibility of cross-links involving Arg. The

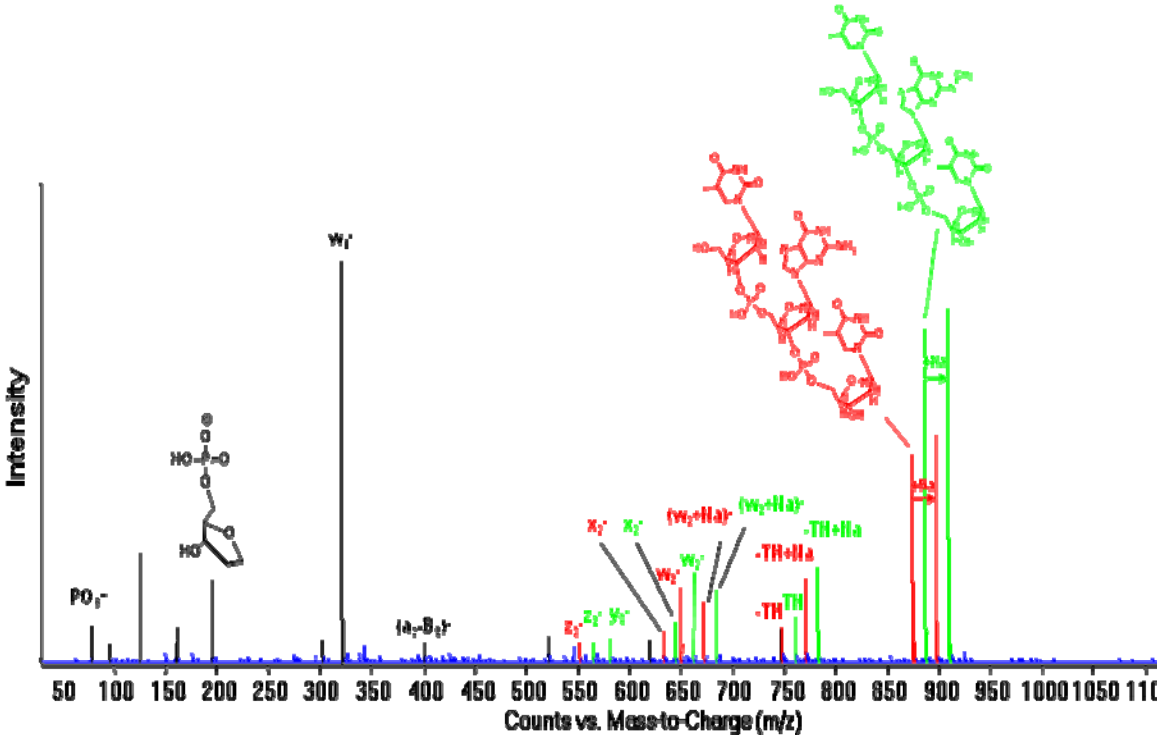
cross-links Cys-CH_2-dG, Cys-CH_2-dA and Cys-CH_2-dC were stable and readily isolated and characterized. The cross-links characterized in this work will contribute to a better understanding of DPC formation induced by formaldehyde and of the mode of action of this known human carcinogen. The stable structures identified in this study have potential as biomarkers for the occurrence of DPCs following formaldehyde exposure. Furthermore, by using [$^{13}CD_2$]-formaldehyde, specific exposure related exogenous cross-links can be differentiated from endogenous cross-links. Such sophisticated methods will be necessary to carefully examine site of contact versus distant site adduction of formaldehyde to determine the plausibility of inhaled formaldehyde as a causative agent for leukemia.

Figures



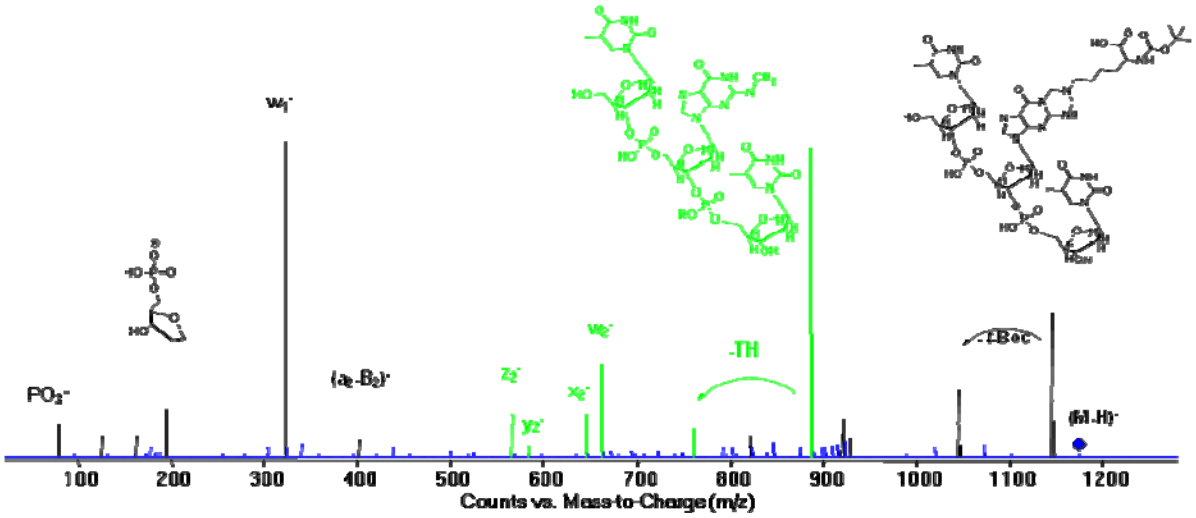
m/z obs	*m/z* calc	ion
886.1785	886.1816	M – H –(*N*-Boc) – Lys
874.1783	874.1816	TGT
760.1349	760.1386	–TH
748.1359	748.1386	–TH
662.0988	662.1018	w_2
650.0986	650.1018	w_2
644.0880	644.0913	x_2
632.0879	632.0913	x_2
582.1341	582.1355	y_2
570.1309	570.1355	y_2
564.1227	564.1250	z_2
552.1231	552.1250	z_2
401.0755	401.0755	$a_2 – B_2$
321.0477	321.0493	w_1
195.0046	195.0064	deoxyribose-5'-phosphate
78.9590	78.9591	PO_3^-

Figure 3.1. MS/MS spectrum of TGT—CH2—Lys.



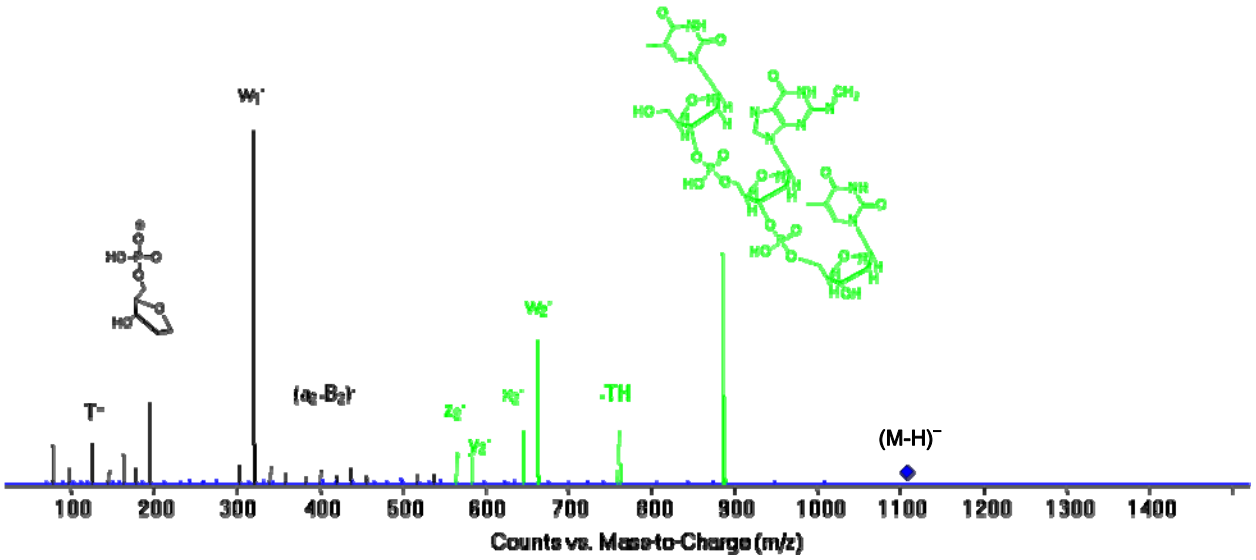
m/z obs	*m/z* calc	ion
908.1605	908.1635	M – H –(*N*-Boc) – Lys + Na
896.1605	896.1635	TGT + Na
886.1779	886.1816	M – H –(*N*-Boc) – Lys
874.1768	874.1816	TGT
782.1134	782.1206	– TH + Na
770.1195	770.1206	– TH + Na
760.1364	760.1386	– TH
748.1353	748.1386	– TH
684.0865	684.0838	w_2 + Na
672.0768	672.0838	w_2 + Na
662.0991	662.1018	w_2
650.0993	650.1018	w_2
644.0821	644.0913	x_2
632.0860	632.0913	x_2
582.1380	582.1355	y_2
564.1249	564.1250	z_2
552.1234	552.1250	z_2
401.0743	401.0755	$a_2 – B_2$
321.0427	321.0493	w1
195.0051	195.0064	deoxyribose-5'-phosphate
78.9582	78.9591	PO_3^-

Figure 3.2. MS/MS spectrum of T(TPHA-1)T.



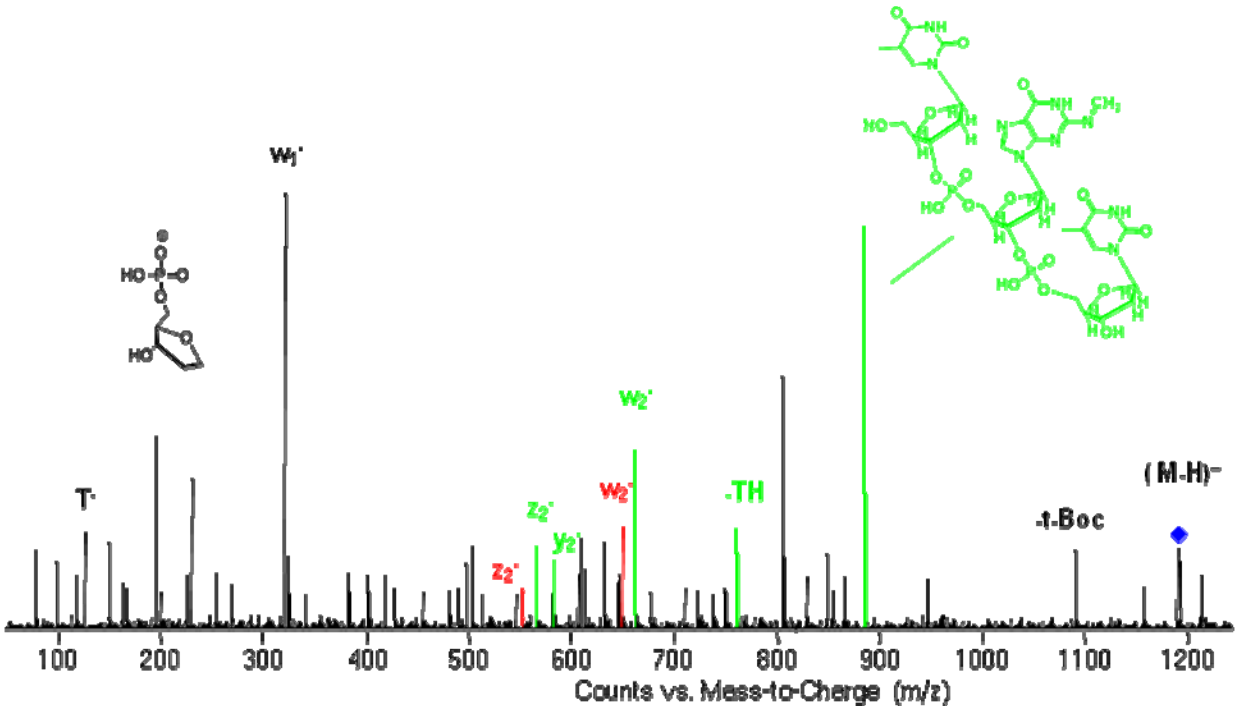
m/z obs	m/z calc	Ion
1144.3324	1144.3395	M – H – CH_2OH
1044.2822	1044.2871	M – H – CH_2OH – (*N*-Boc)
886.1788	886.1816	M – H – CH_2OH – (*N*-Boc-Lys)
760.1323	760.1386	– TH
662.0981	662.1018	w_2
644.0918	644.0913	x_2
582.1356	582.1355	y_2
564.1261	564.1250	z_2
401.0693	401.0755	$a_2 – B_2$
321.0476	321.0493	w_1
195.0040	195.0064	deoxyribose-5'-phosphate
78.9593	78.9591	PO_3^-

Figure 3.3. MS/MS spectrum of T(TPHA-2)T



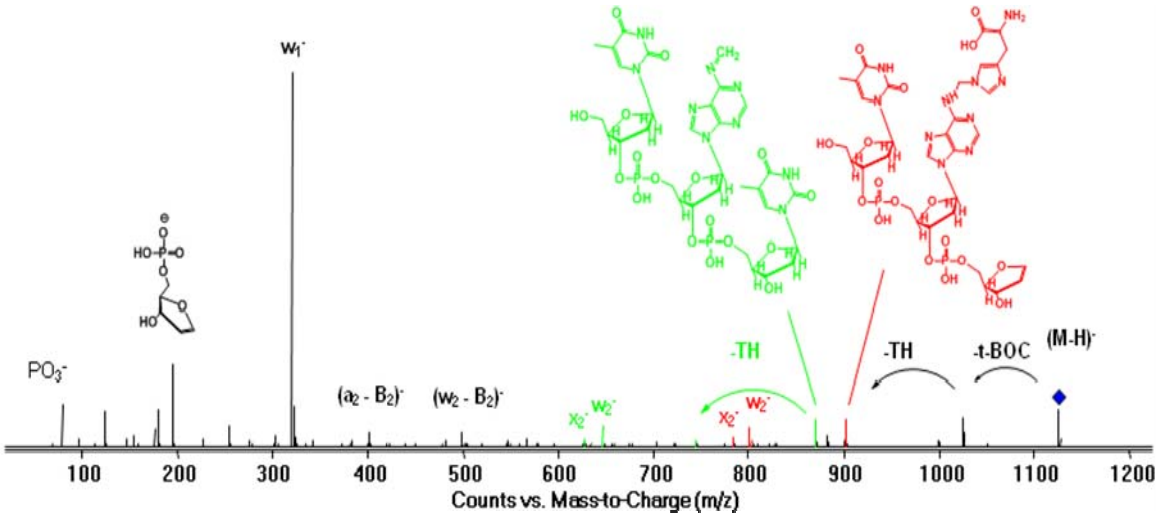
m/z obs	*m/z* calcd	ion
886.1777	886.1816	M – H – (*N*-Boc-Cys)
760.1363	760.1386	– TH
662.1004	662.1018	w_2
644.0850	644.0913	x_2
582.1424	582.1355	y_2
564.1232	564.1250	z_2
401.0694	401.0755	$a_2 – B_2$
321.0488	321.0493	w_1
195.0059	195.0064	deoxyribose-5'-phosphate
125.0352	125.0357	T^-

Figure 3.4. MS/MS spectrum of TGT–CH2–Cys.



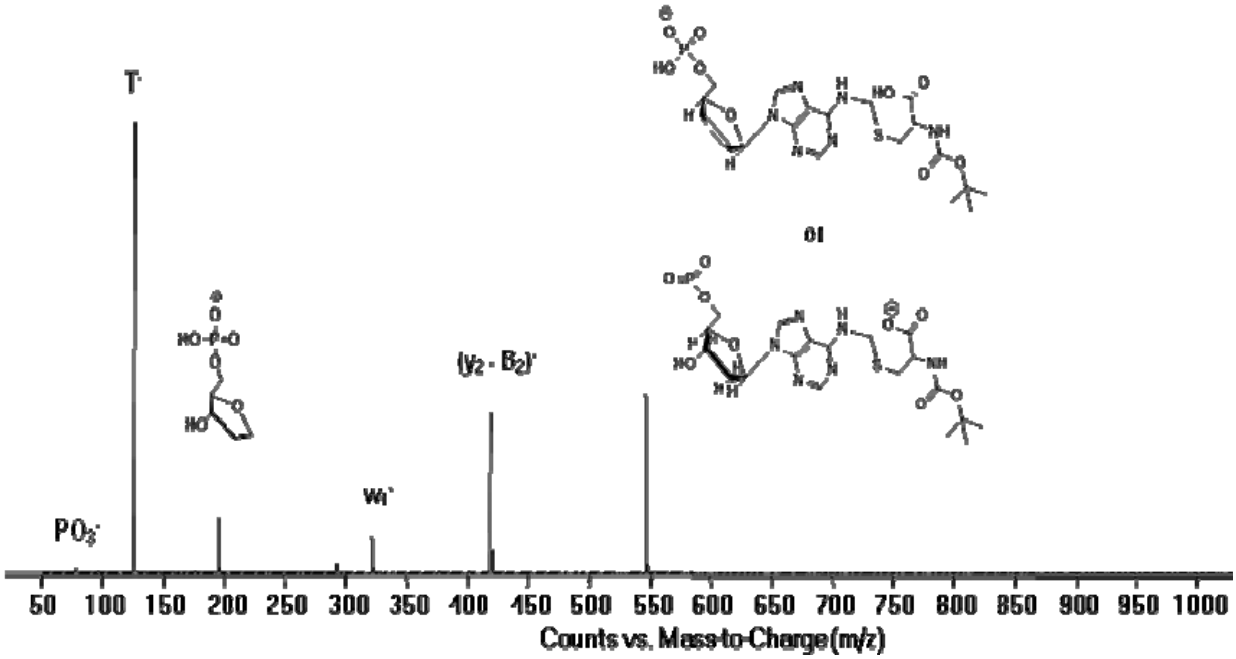
***m/z* obs**	***m/z* calc**	**ion**
1090.2667	1090.2714	M – H – Boc
886.1723	886.1816	M – H – *N*-Boc-Trp
760.1315	760.1386	–TH
662.1108	662.1018	w_2
650.0996	650.1018	w_2 (origin TGT)
644.0702	644.0913	x_2
582.1255	582.1355	y_2
564.1138	564.1250	z_2
552.1118	552.1250	z_2
321.0480	321.0493	w_1
195.0073	195.0064	deoxyribose-5'-phosphate
125.0385	125.0357	T^-

Figure 3.5. MS/MS spectrum of TGT–CH2–Trp.



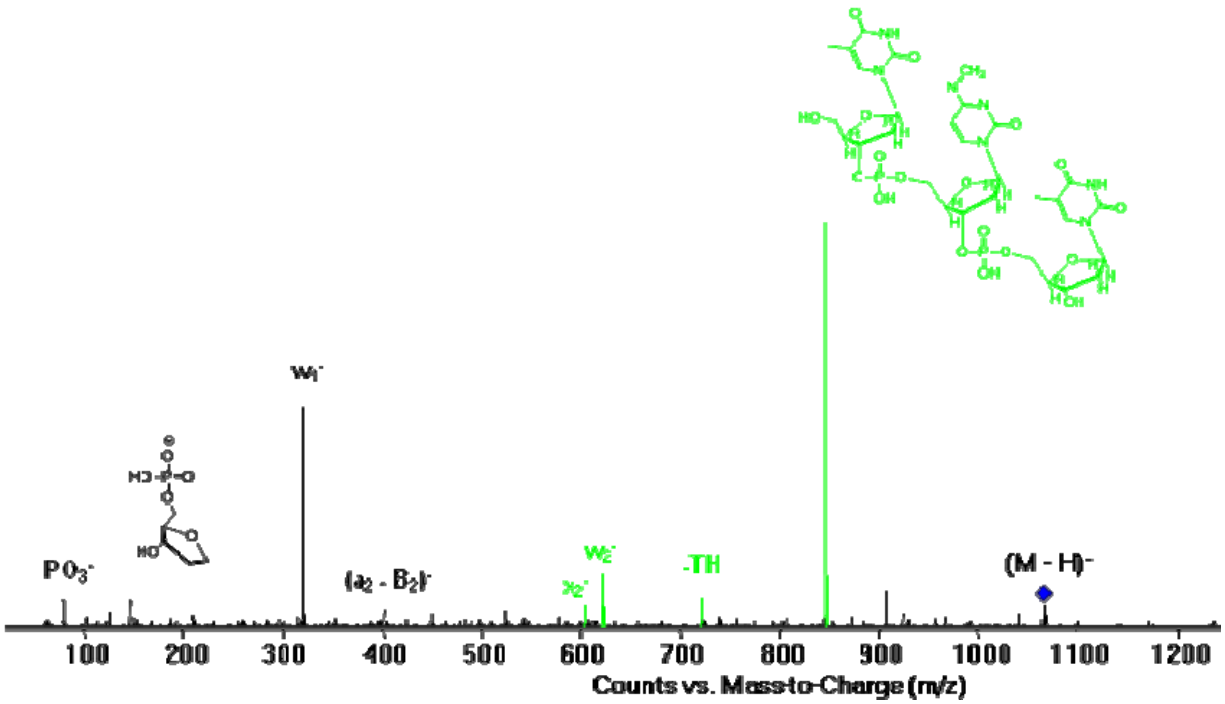
***m/z* obs**	***m/z* calc**	**ion**
1025.2529	1025.2561	M – H – Boc
901.2276	901.2288	M – H – Boc – TH
870.1830	870.1866	M – H – (*N*-Boc-His)
801.1668	801.1764	w_2
783.1625	783.1658	x_2
744.1440	744.1437	M – H – (*N*-Boc-His) – TH
646.1034	646.1069	w_2
628.0949	628.0964	x_2
499.0514	499.0524	$w_2 – B_2$
401.0619	401.0755	$a_2 – B_2$
321.0480	321.0493	w_1
195.0055	195.0064	deoxyribose-5'-phosphate
125.0351	125.0357	T^-

Figure 3.6. MS/MS spectrum of TAT–CH2–His.



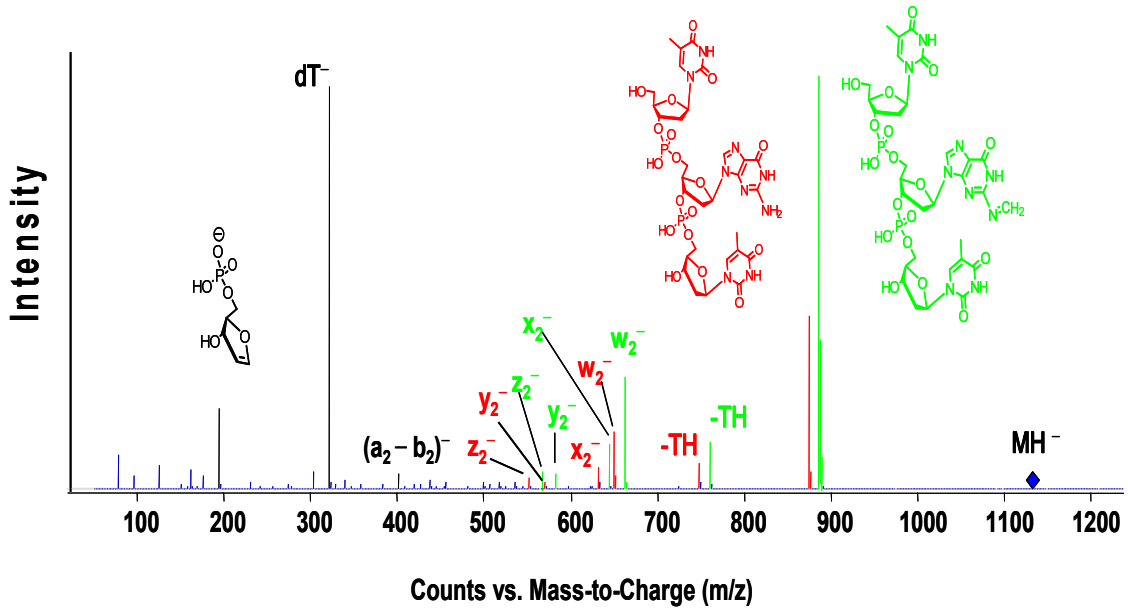
m/z obs	m/z calc	ion
545.1284	545.1225	2',3'-dideoxy-2',3'-dehydro-AMP-CH_2-(*N*-Boc-Cys) or dA-5'-phosphenate- CH_2-(*N*-Boc-Cys)
419.0853	419.0861	$y_2 - B_2$
321.0485	321.0493	w_1
195.0056	195.0064	deoxyribose-5'-phosphate
125.0537	125.0357	T^-
78.9592	78.9591	PO_3^-

Figure 3.7. MS/MS spectrum of TAT–CH2–Cys.



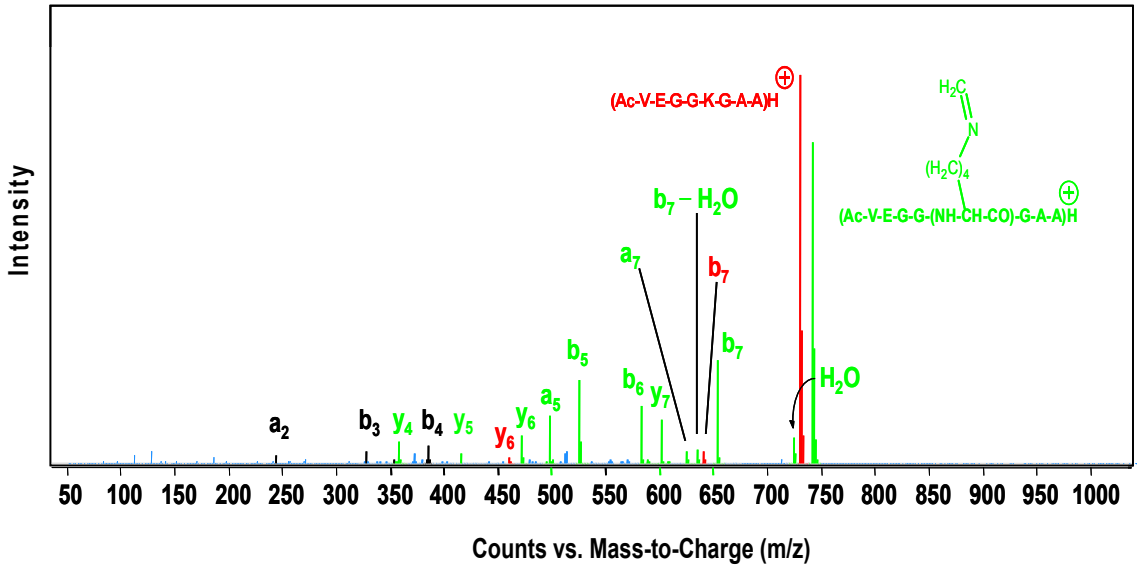
***m/z* obs**	***m/z* calc**	**ion**
846.1718	846.1754	M – H – (*N*-Boc-Cys)
720.1192	720.1325	M – H – (*N*-Boc-Cys) – TH
622.0944	622.0957	w_2
604.0830	604.0851	x_2
401.0696	401.0755	$a_2 – B_2$
321.0488	321.0493	w_1
195.0081	195.0064	deoxyribose-5'-phosphate
78.9584	78.9591	PO_3^-

Figure 3.8. MS/MS spectrum of TCT–CH2–Cys.



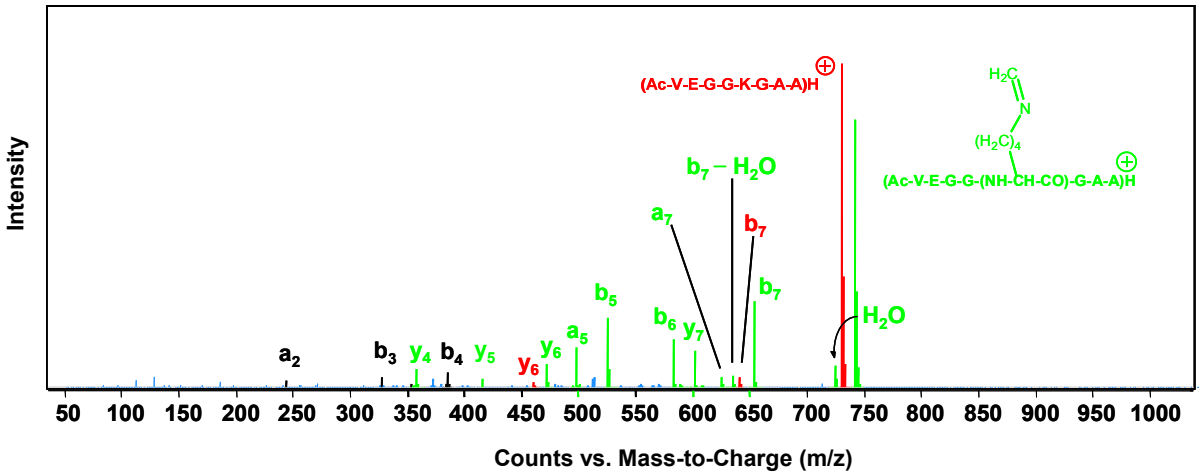
— ions containing Schiff base guanine; — ions containing unmodified guanine; — ions that do not contain guanine.

Figure 3.9. ESI-QTOF-MS/MS of the deprotonated molecule TGT—CH_2—Lys.



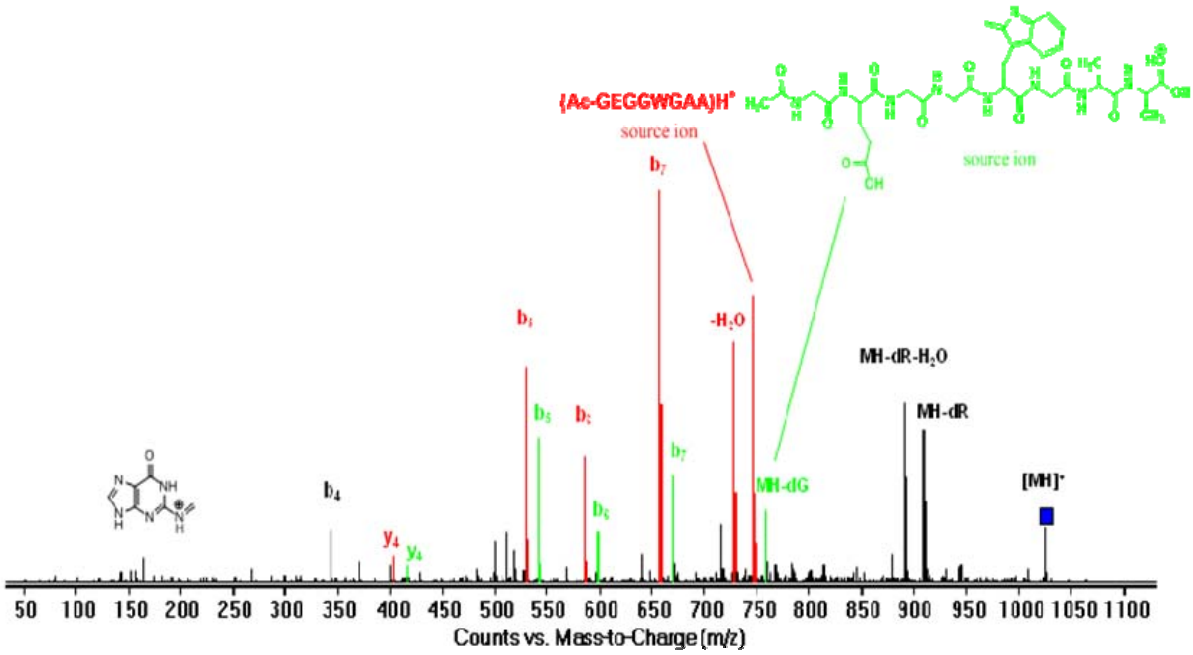
Legend: — ions containing Schiff base derivative of Lys;
— ions containing unmodified Lys;
— ions that do not contain residue 5.

Figure 3.10. ESI-QTOF-MS/MS spectrum of the protonated molecule Acetyl-VEGG(K-CH2-dG)GAA



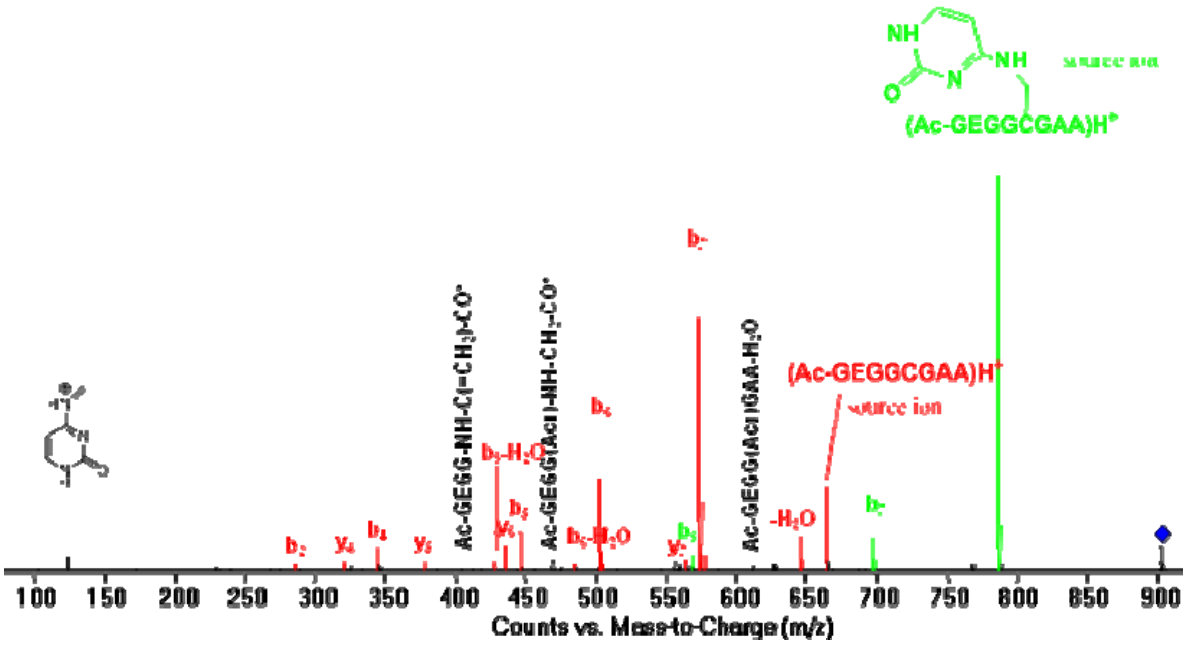
m/z obs	*m/z* calc	Ion
742.3737	742.3730	Ac-NH-VEGG-(NH-CH-CH_2)$_4$-CH=NH)-GAA$^+$ (Schiff base of peptide)
730.3738	730.3730	(Ac-NH-VEGGKGAA)$^+$
724.3626	724.3624	– H_2O
653	653.3253	b_7
641.3248	641.3253	b_7
635.3145	635.3148	b_7 – H_2O
625.3298	625.3304	a_7
601.2940	601.2940	y_7
582.2885	582.2882	b_6
525.2672	525.2667	b_5
513.2641	513.2667	b_5
497.2722	497.2718	a_5
472.2512	472.2514	y_6
460.2513	460.2514	y_6
385.1719	385.1718	b_4
358.2082	358.2085	y_4
346.2084	346.2085	y_4
328.1498	328.1503	b_3

Figure 3.11. MS/MS spectrum of Acetyl-VEGG(K-CH2-dG)GAA.



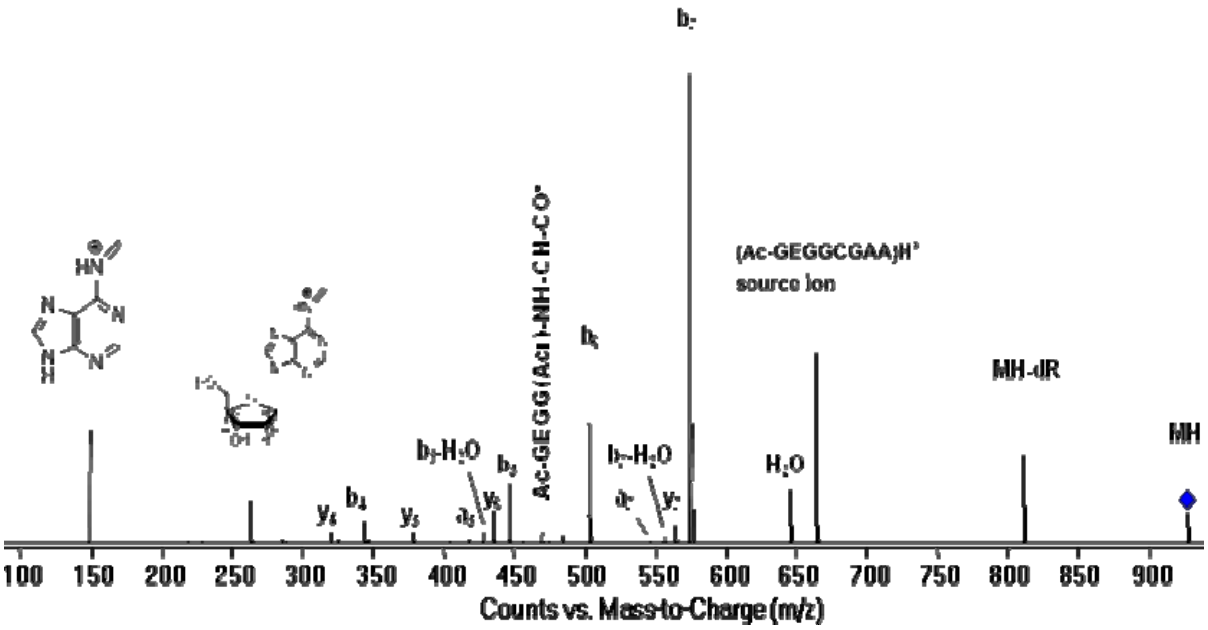
***m/z* obs**	***m/z* calc**	**Ion**
909.3567	909.3598	MH – dR
891.3452	891.3492	MH – dR– H_2O
758.3072	758.3104	MH – dG
746.3097	746.3104	$(Ac\text{-}GEGGWGAA)H^+$
728.3005	728.2998	$- H_2O$
669.2593	669.2627	b_7
657.2643	657.2627	b_7
598.2225	598.2256	b_6
586.2255	586.2256	b_6
541.2022	541.2041	b_5
529.2024	529.2041	b_5
416.1883	416.1928	y_4
404.1950	404.1928	y_4
343.1229	343.1248	b_4
164.0503	164.0567	Schiff base of Gua

Figure 3.12. MS/MS spectrum of Acetyl-GEGG(W-CH2-dA)GAA.



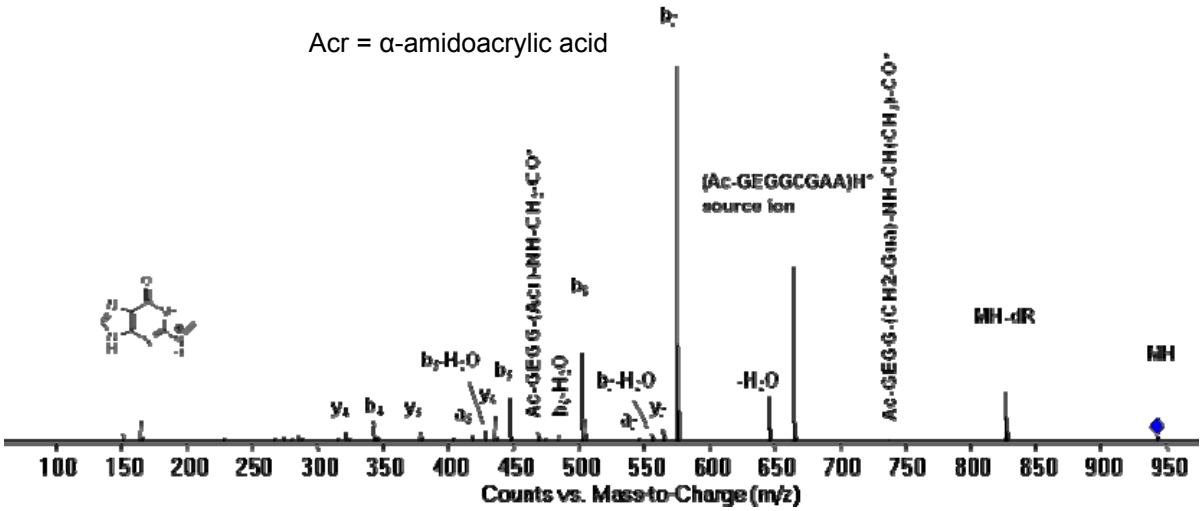
m/z obs	*m/z* calc	ion
786.2820	786.2835	MH – dR
697.2344	697.2358	b_7
663.2392	663.2403	MH – dC=CH_2
645.2274	645.2297	– H_2O
626.1975	626.1987	b_6
611.2417	611.2420	AcGEGG-(Acr)-GAA – H_2O
574.1919	574.1926	b_7
569.1759	569.1773	b_5
564.2075	564.2082	y_7
540.1943	540.2049	AcGEGG-(Acr)-G-NH-CH(CH_3)CO^+
503.1544	503.1555	b_6
485.1415	485.1449	b_6 – H_2O
469.1649	469.1678	AcGEGG-(Acr)-NH-CH_2-CO^+
446.1331	446.1340	b_5
435.1646	435.1656	y_6
428.1229	428.1234	b_5 – H_2O
412.1440	412.1463	AcGEGG-NH-C(=CH_2)-CO^+
378.1426	378.1442	y_5
343.1244	343.1248	b_4
321.1218	321.1227	y_4
286.1007	286.1039	b_2
124.0507	124.0505	Schiff base of CH

Figure 3.13. MS/MS spectrum of Acetyl-GEGG(C-CH2-dC)GAA.



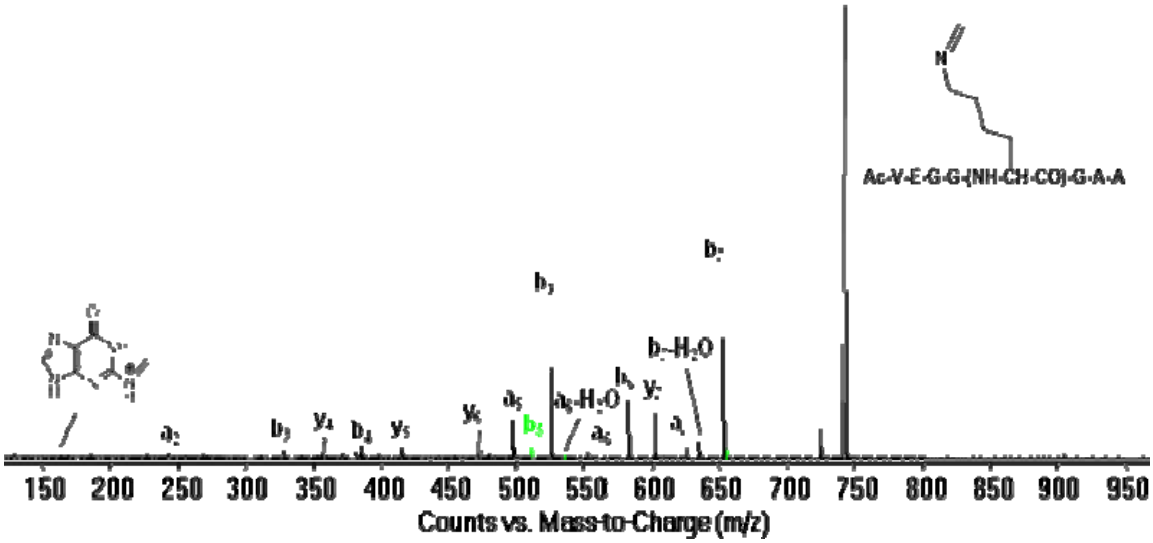
m/z obs	*m/z* calc	Ion
810.2938	810.2948	MH - dR
663.2400	663.2403	$(Ac\text{-}GEGGCGAA)H^+$
645.2287	645.2297	$-H_2O$
574.1926	574.1926	b_7
564.2062	564.2082	y_7
556.1818	556.1820	$b_7 - H_2O$
546.1971	546.1977	a_7
503.1551	503.1555	b_6
469.1674	469.1678	$Ac\text{-}GEGG(Acr)\text{-}NH\text{-}CH\text{-}CO^+$
446.1337	446.1340	b_5
435.1652	435.1656	y_6
428.1231	428.1234	$b_5 - H_2O$
418.1399	418.1391	a_5
412.1431	412.1463	$Ac\text{-}GEGG\text{-}NH\text{-}C(=CH_2)\text{-}CO^+$
378.1442	378.1442	y_5
343.1250	343.1248	b_4
321.1223	321.1227	y_4
264.1091	264.1091	Schiff base of dA
148.0613	148.0618	Schiff base of A

Figure 3.14. MS/MS spectrum of Acetyl-GEGG(C-CH2-dA)GAA.



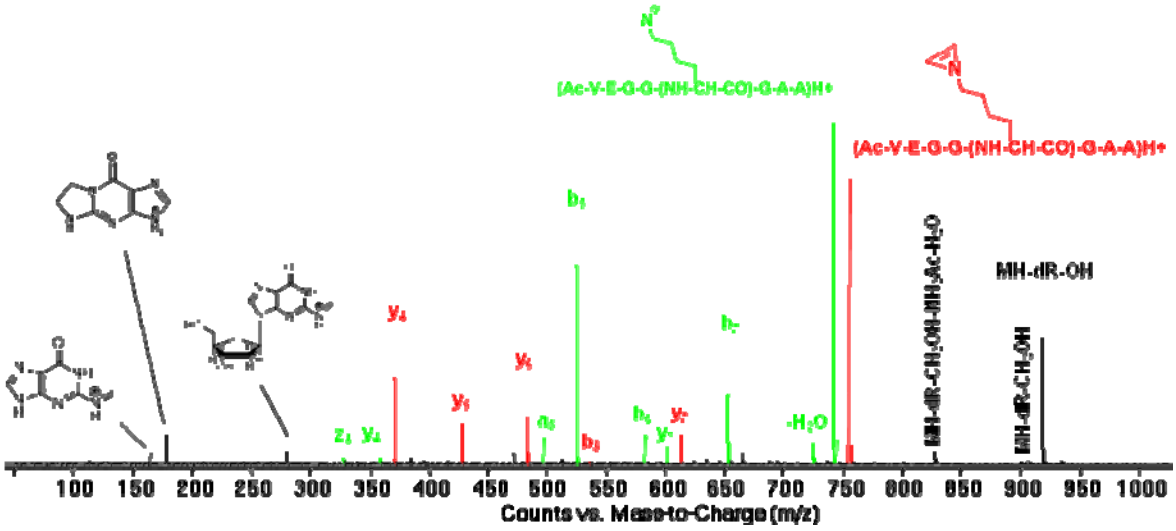
m/z obs	*m/z* calc	Ion
826.2883	826.2897	MH − dR
737.2441	737.2420	Ac-GEGGC(-CH_2-Gua)G-NH-CH(CH_3)CO^+
663.2390	663.2403	(Ac-GEGGCGAA)H^+
645.2283	645.2297	− H_2O
574.1921	574.1926	b_7
564.2083	564.2082	y_7
556.1806	556.1820	$b_7 – H_2O$
546.1964	546.1977	a_7
503.1556	503.1555	b_6
485.1453	485.1449	$b_6 – H_2O$
469.1670	469.1678	Ac-GEGG-(Acr)-NH-CH_2-CO^+
446.1338	446.1340	b_5
435.1654	435.1656	y_6
428.1233	428.1234	$b_5 – H_2O$
418.1388	418.1391	a_5
412.1431	412.1463	Ac-GEGG-NH-C(=CH_2)-CO^+
378.1429	378.1442	y_5
343.1239	343.1248	b_4
321.1230	321.1227	y_4
164.0557	164.0567	Schiff base of G
152.0558	152.0567	G

Figure 3.15. MS/MS spectrum of Acetyl-GEGG(C-CH2-dG)GAA.



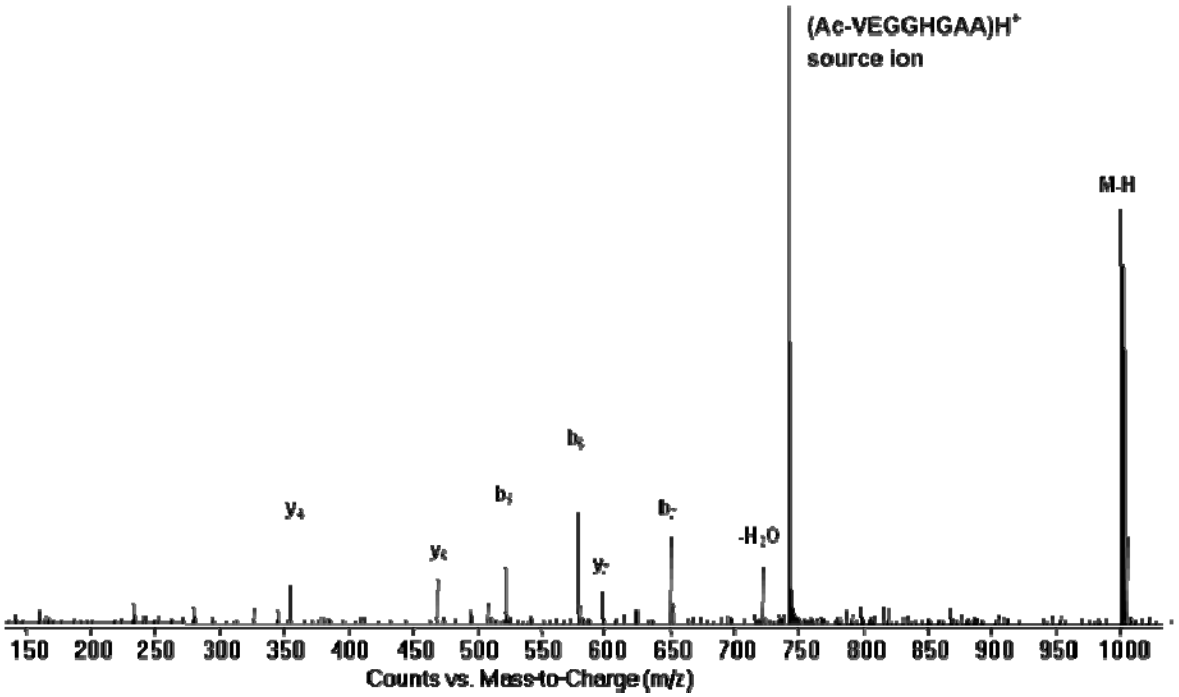
m/z obs	*m/z* calc	Ion
742.3735	742.3730	(Ac- VEGG-(NH-CH-CH_2)$_4$-CH=NH)-GAA)$^+$ (Schiff base of peptide)
724.3620	724.3624	− H_2O
653.3255	653.3253	b_7
635.3141	635.3148	b_7 − H_2O
625.3296	625.3304	a_7
601.2939	601.2940	y_7
582.2885	582.2882	b_6
554.2932	554.2933	a_6
536.2823	536.2827	a_6 − H_2O
525.2671	525.2667	b_5
513.2639	513.2667	b_5 (no Schiff base)
97.2722	497.2718	a_5
472.2512	472.2514	y_6
415.2295	415.2300	y_5
385.1716	385.1718	b_4
358.2085	358.2085	y_4
328.1516	328.1503	b_3
243.1445	243.1339	a_2
164.0561	164.0567	N^2-Schiff base of G

Figure 3.16. MS/MS spectrum of Acetyl-VEGG(TPHA-1)GAA.



m/z obs	*m/z* calcd	ion
917.4220	917.4224	MH – dR − OH
905.4206	905.4224	MH – dR – CH_2OH
828.3736	828.3747	MH – dR – CH_2OH – NH_2Ac – H_2O
754.3727	754.3730	source ion
742.3727	742.3730	source ion
724.3622	724.3624	− H_2O
653.3252	653.3253	b_7
613.2940	613.2940	y_7
601.2937	601.2940	y_7
582.2882	582.2882	b_6
537.2672	537.2667	b_5
525.2670	525.2667	b_5
497.2716	497.2718	a_5
484.2515	484.2514	y_6
427.2298	427.2300	y_5
370.2082	370.2085	y_4
358.2080	358.2082	y_4
327.1656	327.1663	z_4
280.1030	280.1040	Schiff base of dG
178.0717	178.0723	8,9-dihydro-εG
164.0560	164.0567	Schiff base of G

Figure 3.17. MS/MS spectrum of Acetyl-VEGG(TPHA-2)GAA.



m/z obs	*m/z* calcd	Ion
739.3361	739.3369	$(Ac\text{-}VEGGHGAA)H^+$
721.3254	721.3264	$-H_2O$
650.2905	650.2893	b_7
598.2542	598.2580	y_7
579.2424	579.2522	b_6
522.2354	522.2307	b_5
469.2102	469.2154	y_6
355.1724	355.1724	y_4

Figure 3.18. MS/MS spectrum of Acetyl-VEGG(H-CH_2-dA)GAA.

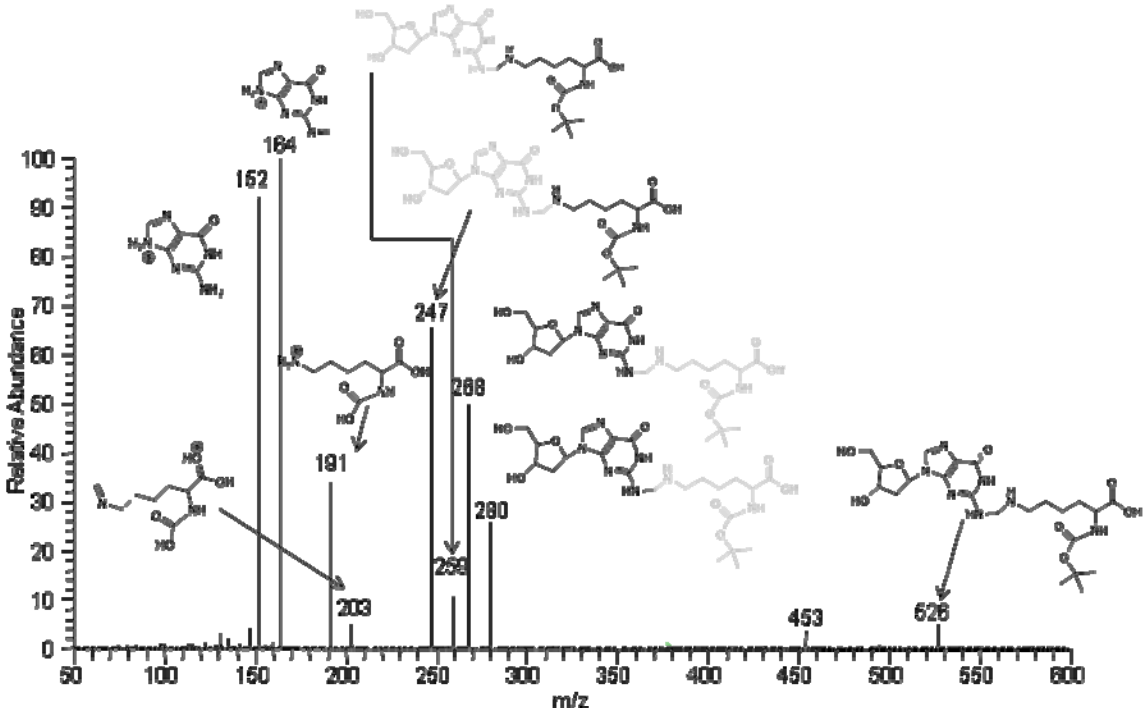


Figure 3.19. MS-MS spectrum of Lys-CH_2-Dg

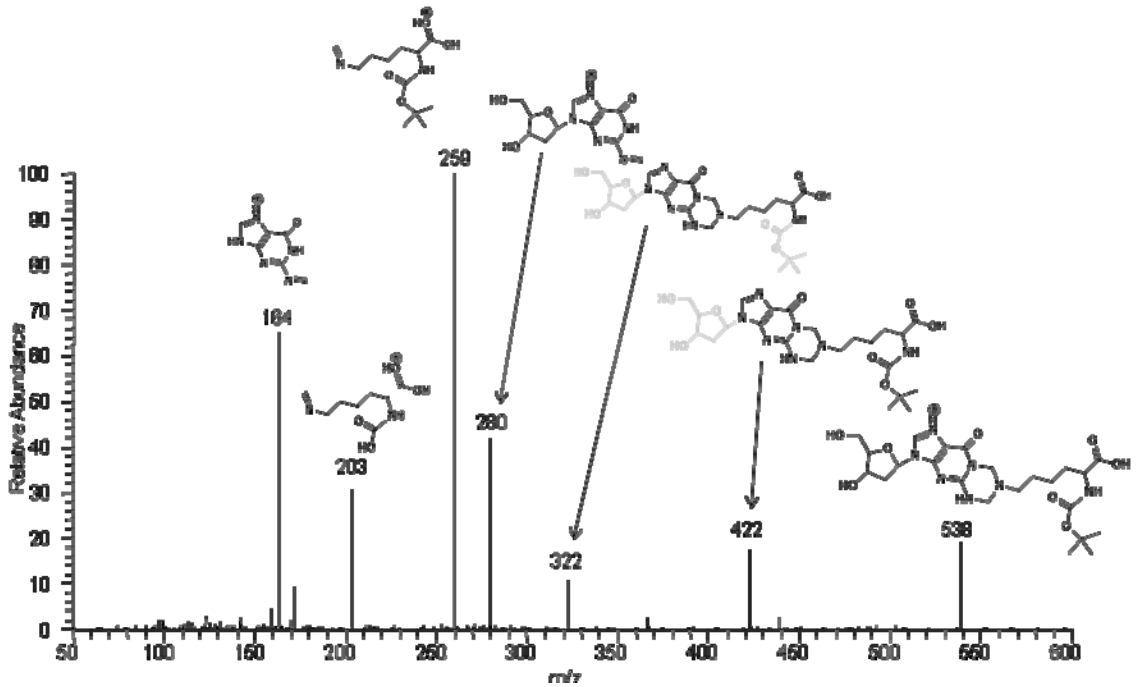


Figure 3.20. MS-MS spectrum of TPHA-1

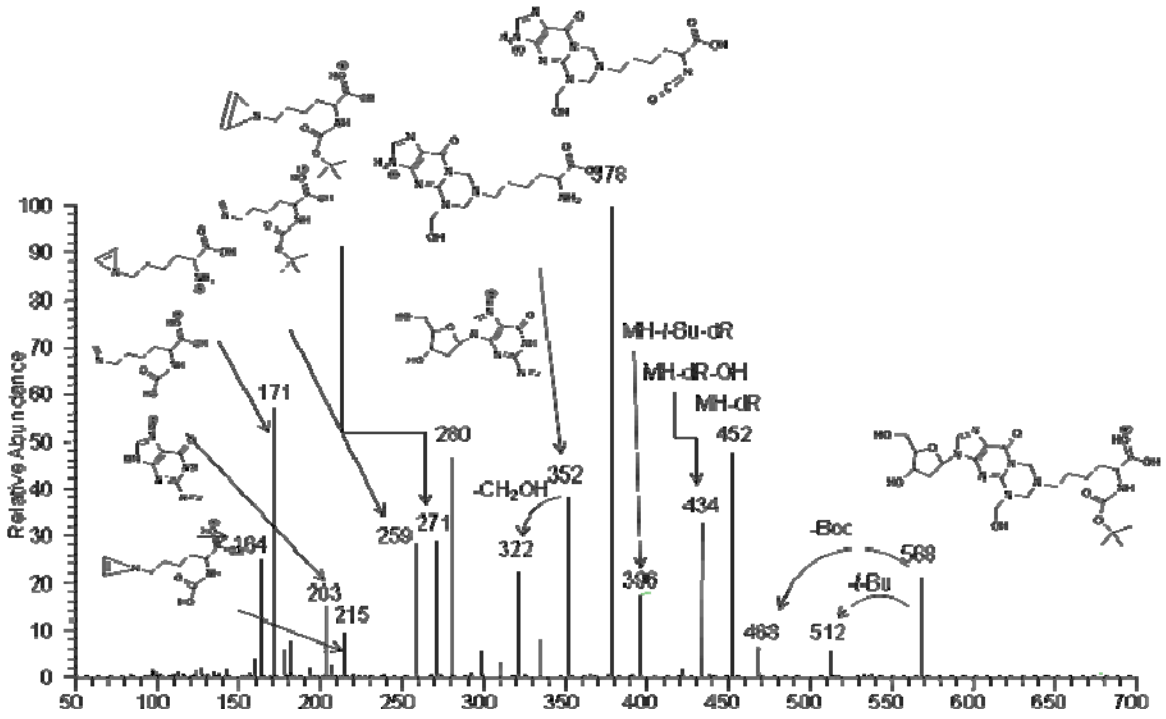


Figure 3.21. MS-MS spectrum of TPHA-2

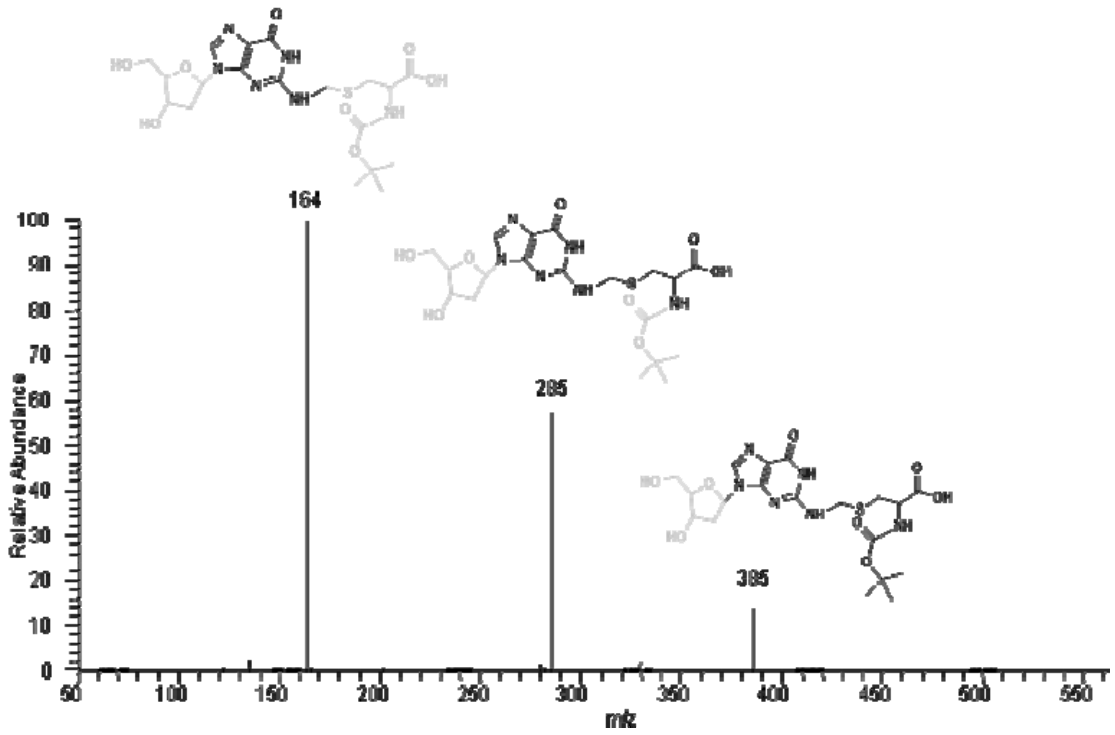


Figure 3.22. MS-MS spectrum of Cys-CH_2-dG

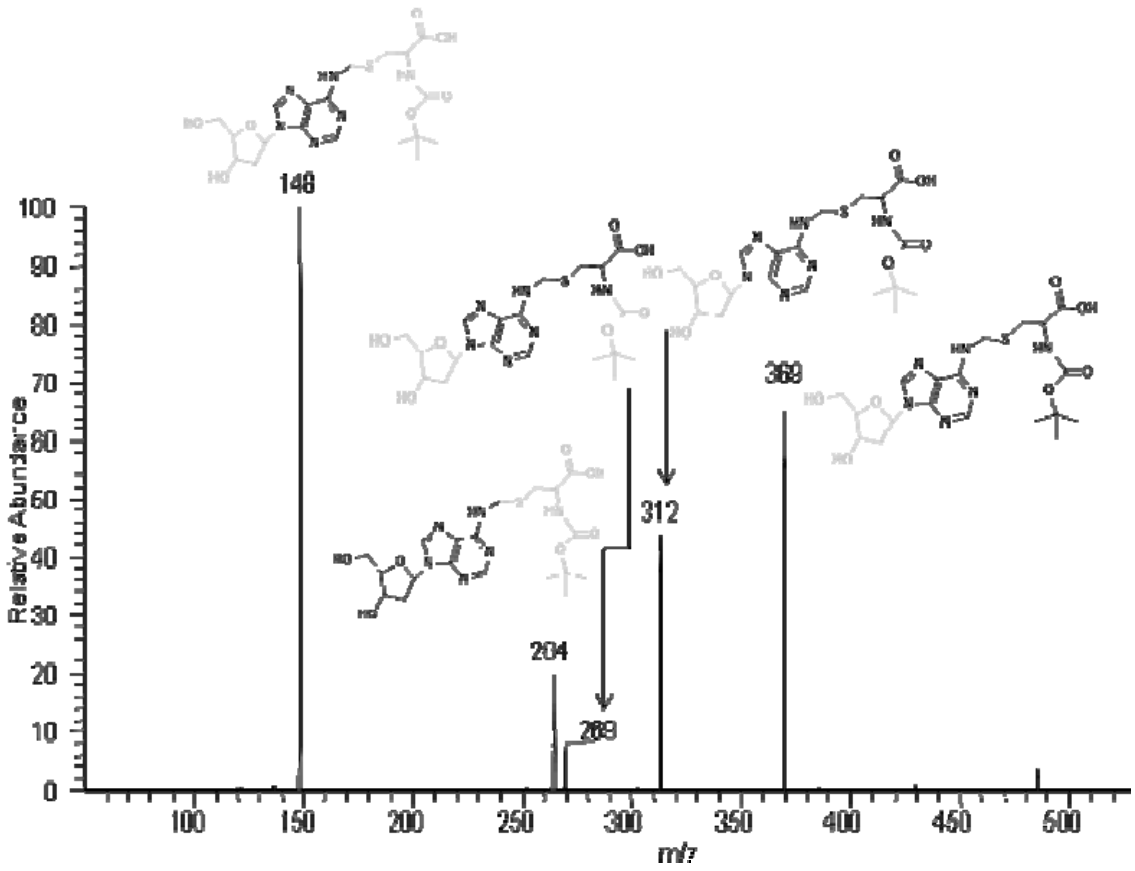


Figure 3.23. MS-MS spectrum of Cys-CH_2-dA

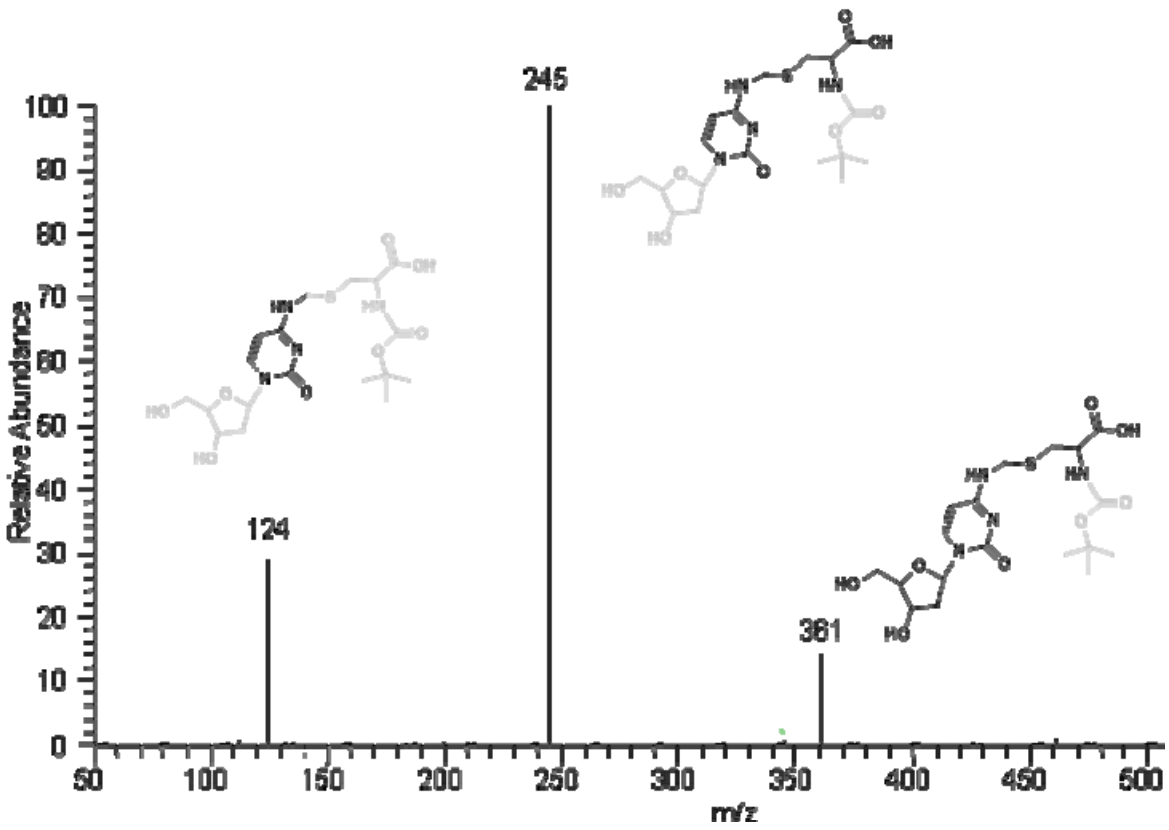


Figure 3.24. MS-MS spectrum of Cys-CH_2-dC

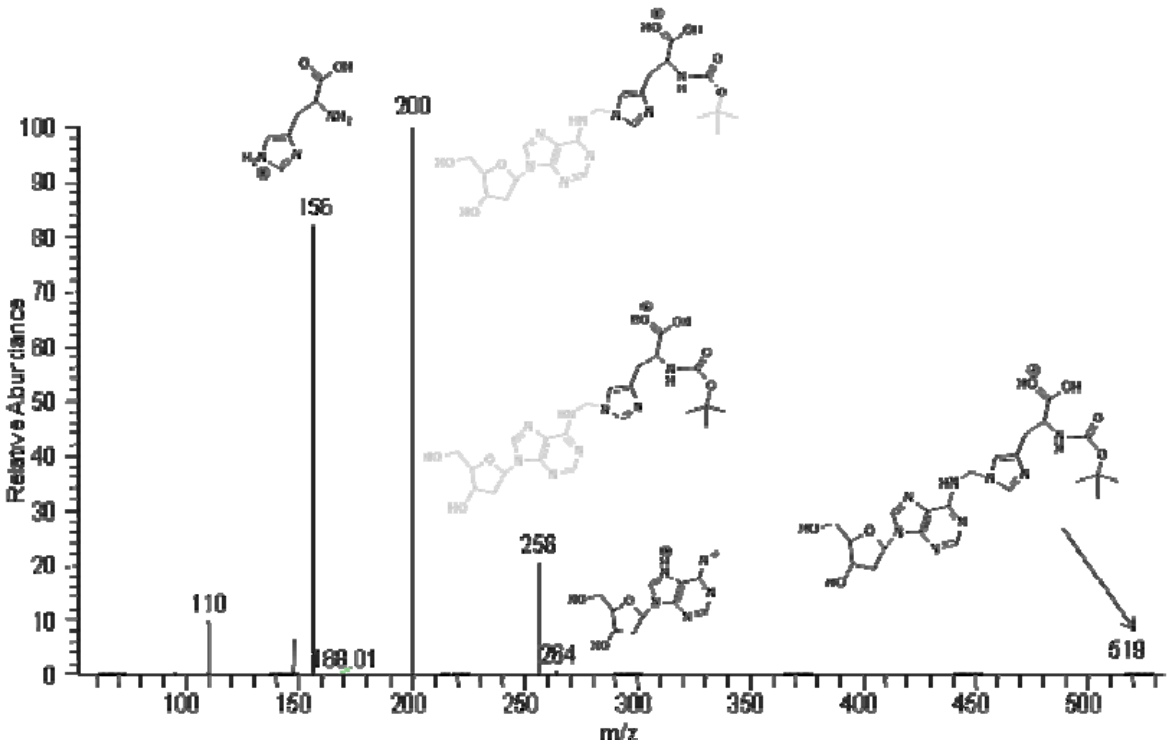


Figure 3.25. MS-MS spectrum of His-CH_2-dA

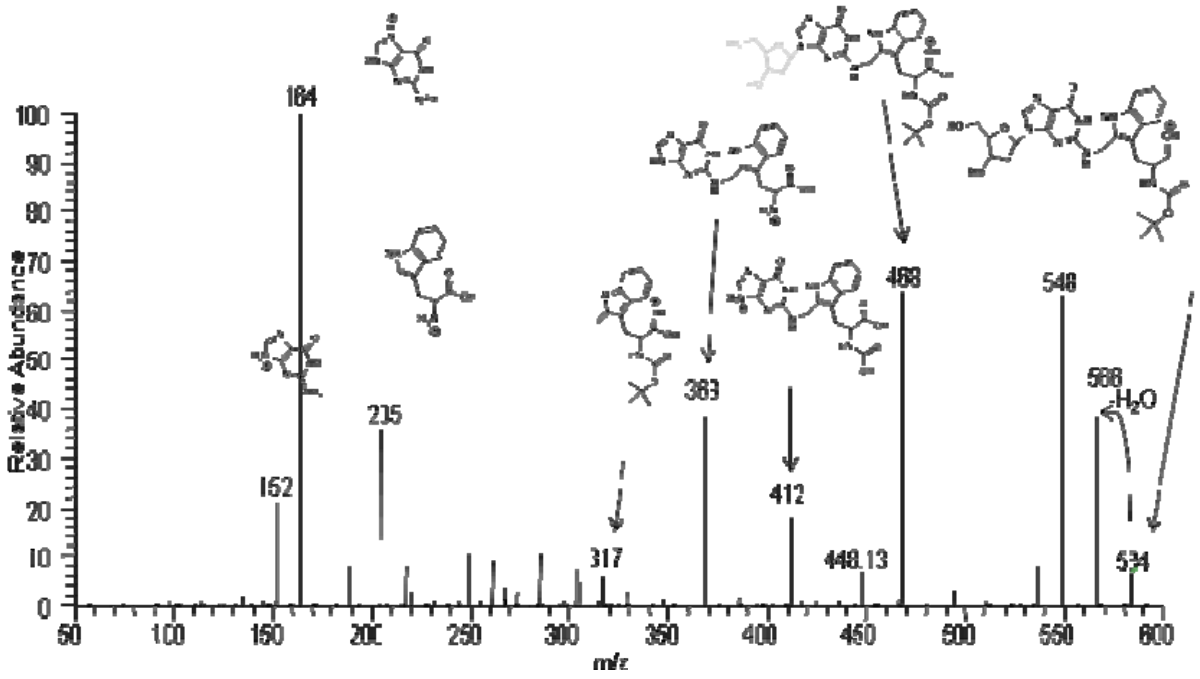


Figure 3.26. MS-MS spectrum of Trp-CH2-dG

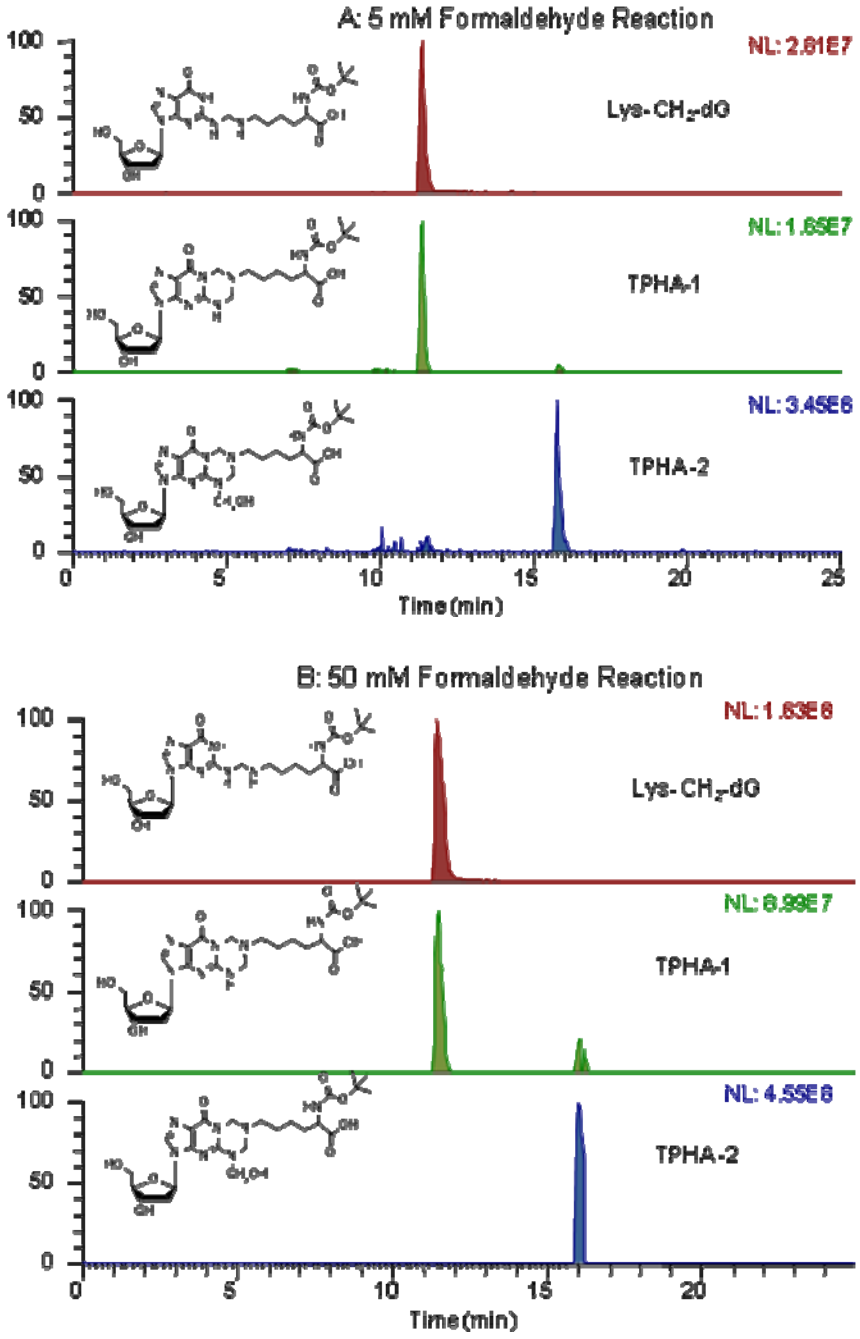


Figure 3.27. SIM of Lys-dG coupling induced by formaldehyde over 48 h: (A: 5 mM formaldehyde; B: 50 mM formaldehyde)

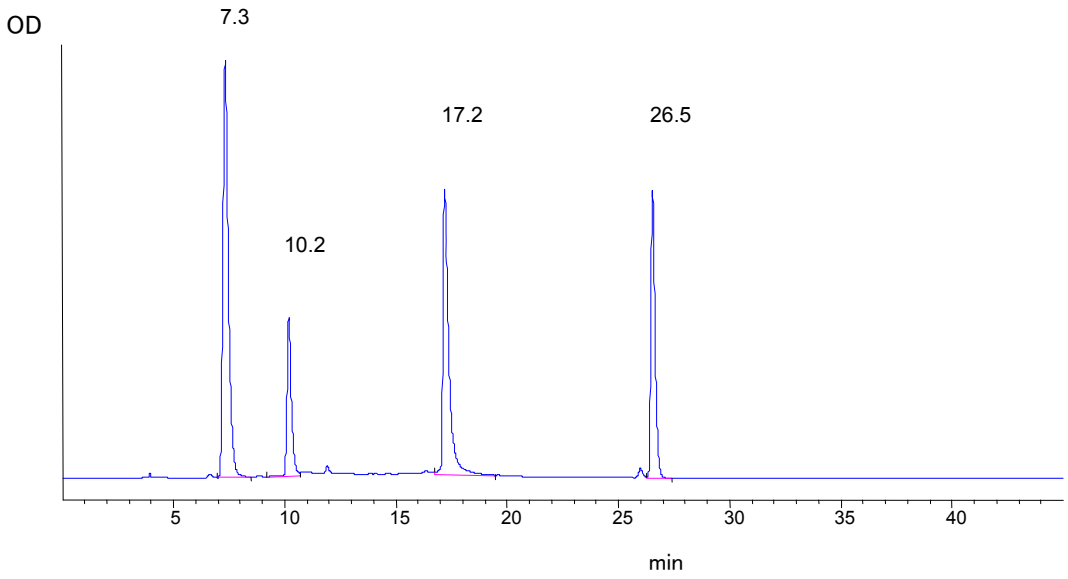


Figure 3.28. Trace (254 nm) of semi-preparative HPLC of the reaction mixture from the formaldehyde-induced coupling of Lys and dG. The peaks at 17.2 and 26.5 min were characterized by NMR. Peaks at 7.3 and 10.2 min are dG and N^2-CH_2OH-dG, respectively.

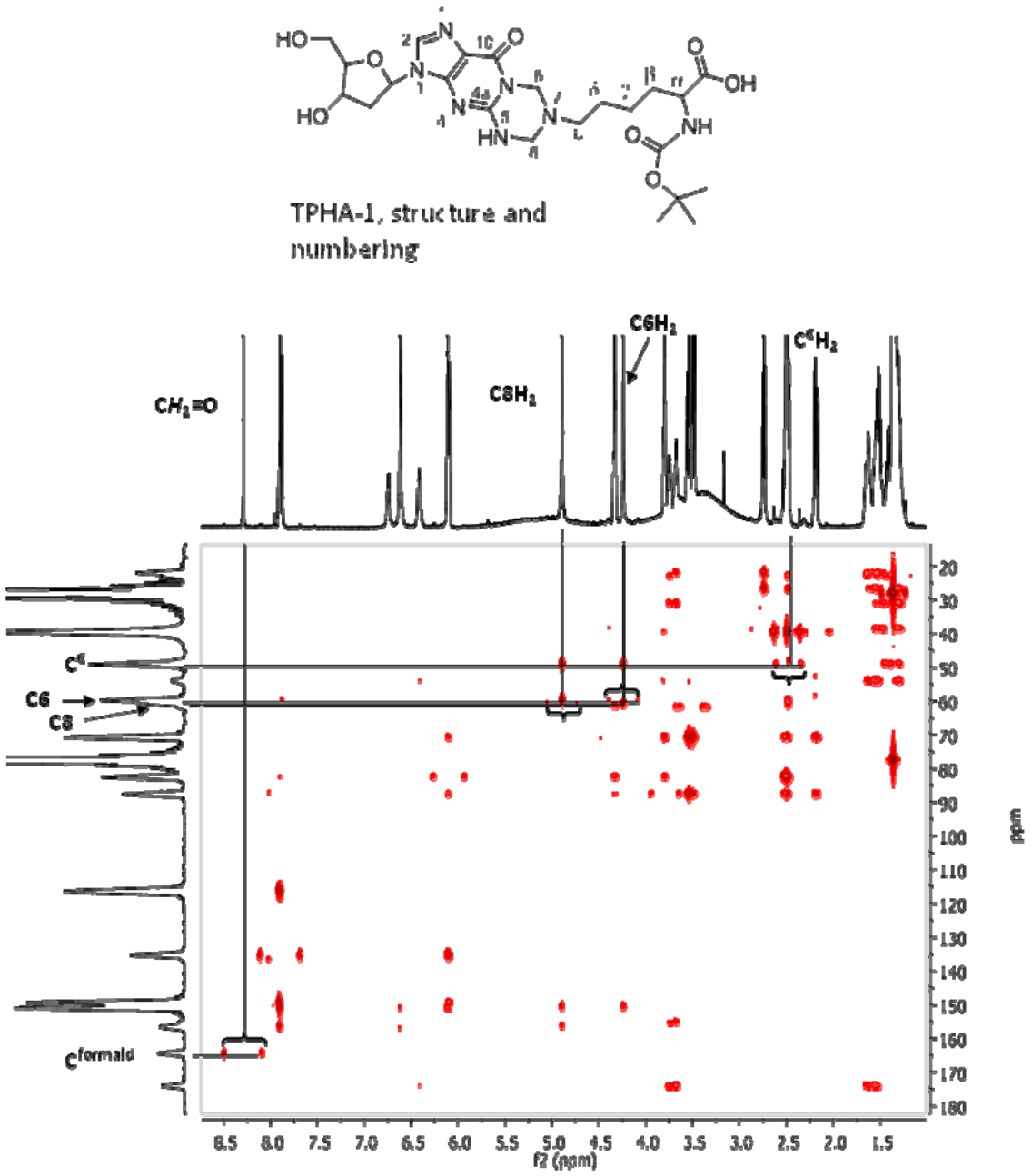


Figure 3.29. HMBC spectrum (DMSO-d_6) of TPHA-1 mixture, indicating C,H connectivities establishing structure of triazino linkage and C,H signals tentatively assigned to formaldehyde. Unsuppressed 1-bond couplings are indicated by brackets.

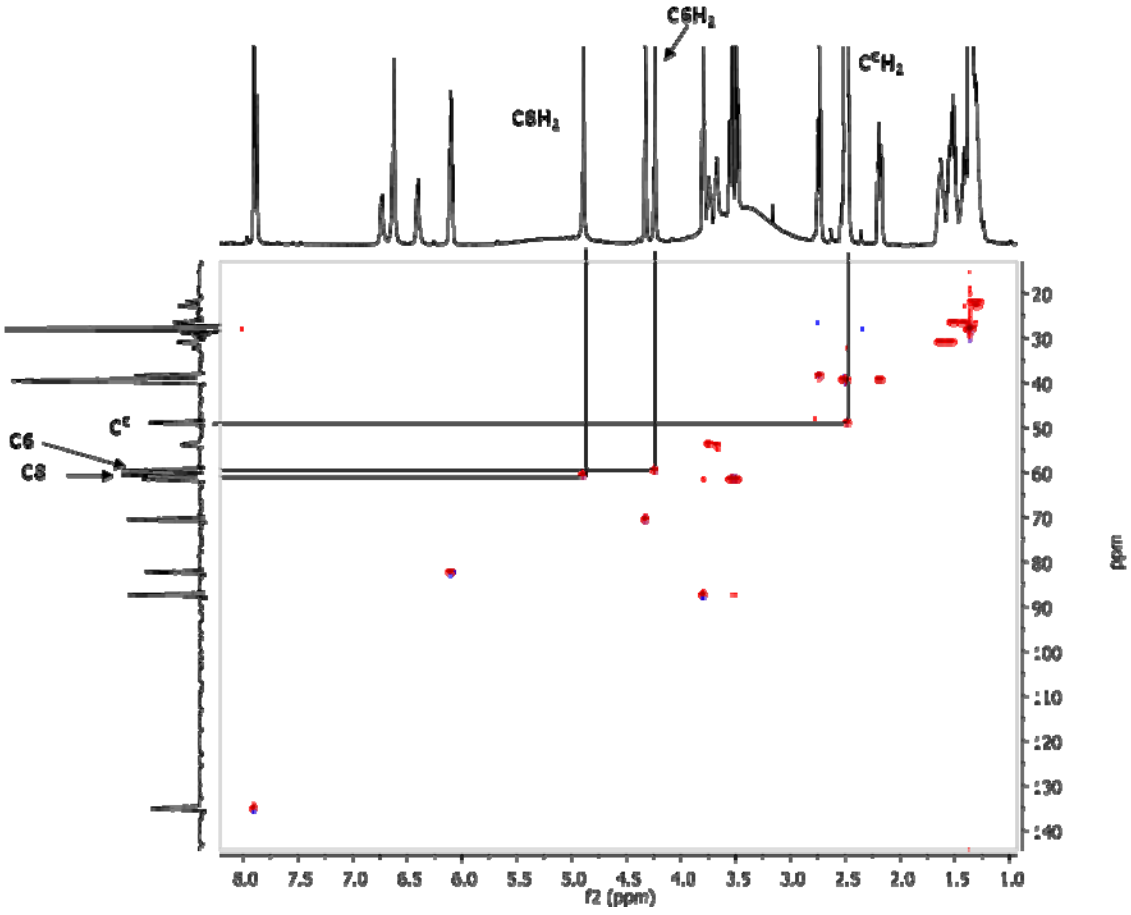


Figure 3.30. HSQC spectrum (DMSO-d_6) of the mixture containing TPHA-1, identifying the C,H cross-peaks discussed in the text.

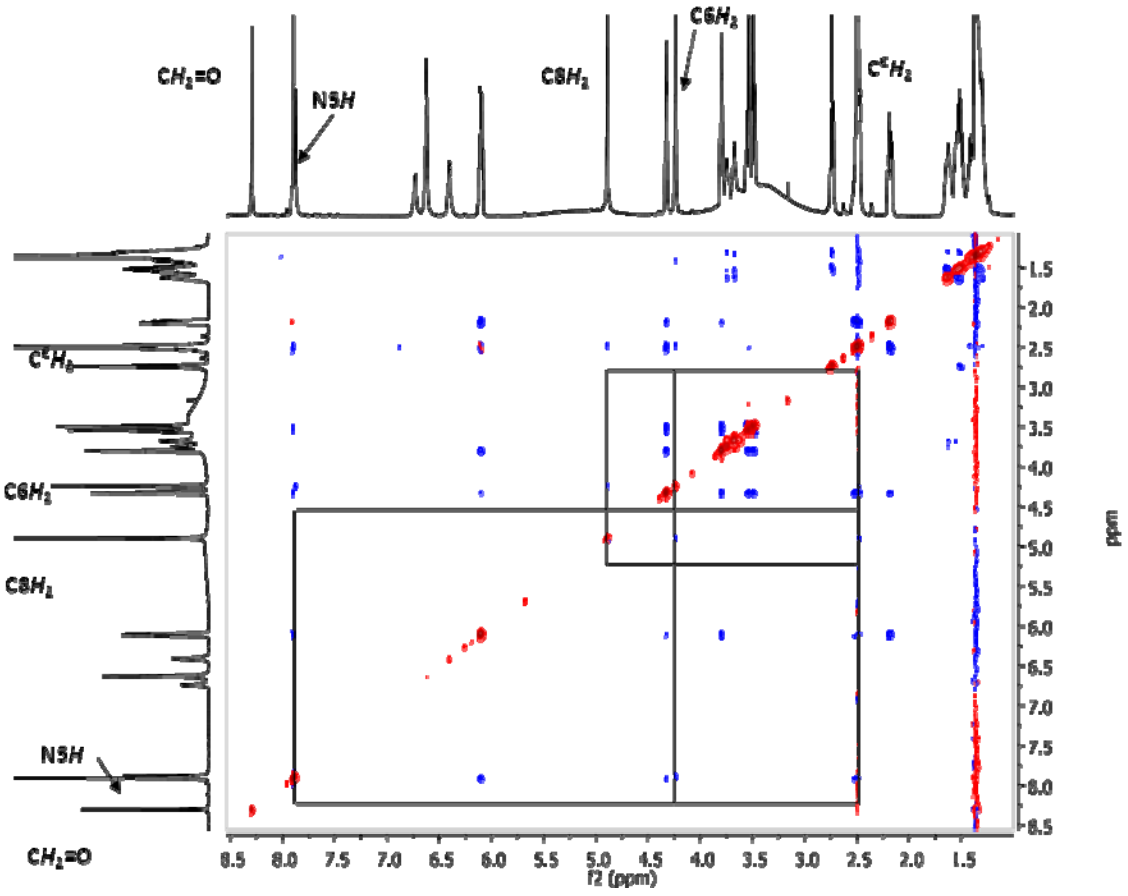


Figure 3.31. ROESY spectrum (500 MHz, DMSO-d_6) of TPHA-1. NOESY cross-peaks between methylene protons at C6 and C8 of the triazino ring and C^{ε} hexanoic acid and between the methylene protons at C6 and N5H are consistent with the HMBC analysis and support the triazino structure for the linkage. Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

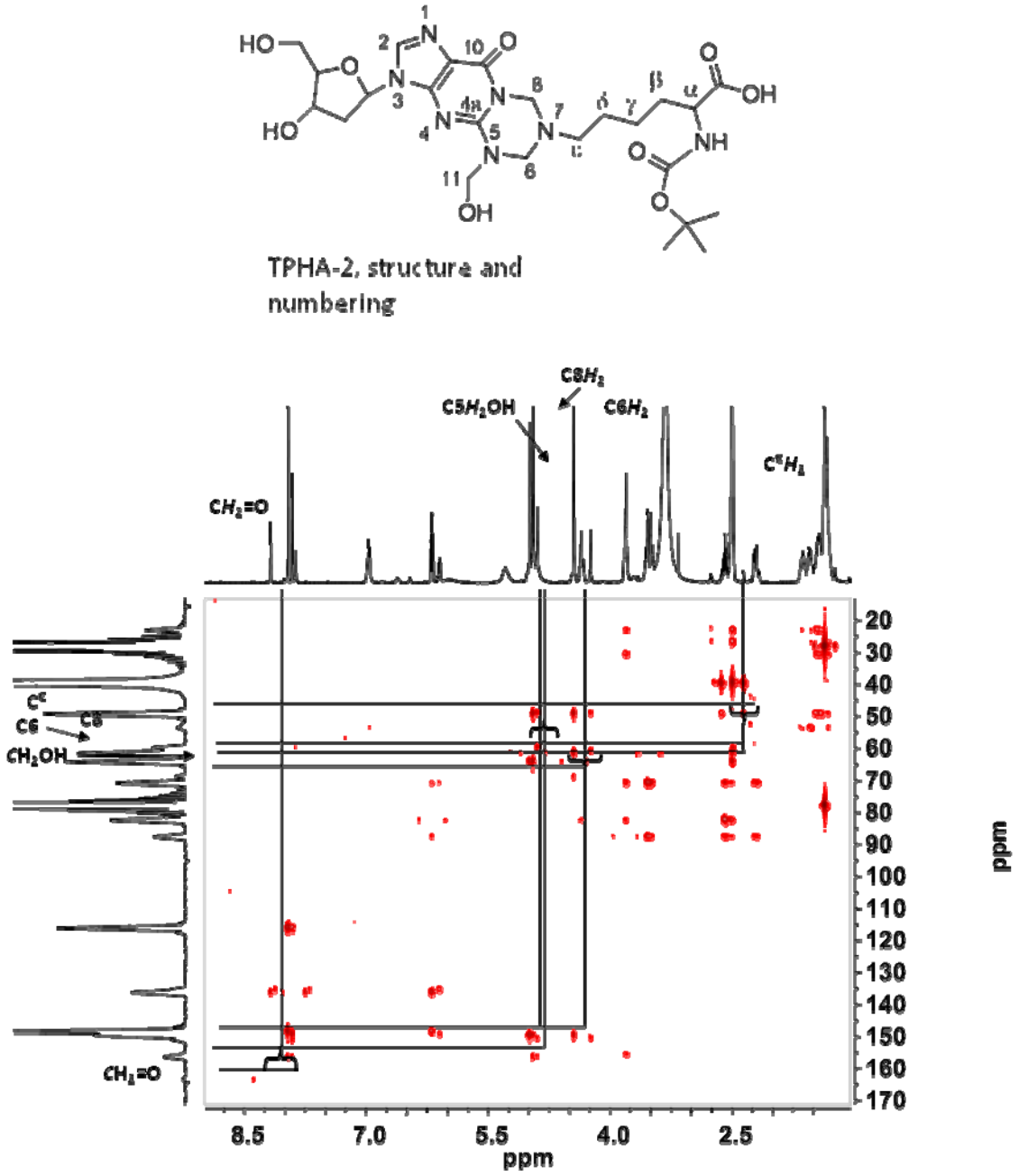


Figure 3.32. HMBC spectrum (DMSO-d_6) of TPHA-2 mixture, indicating C,H connectivities establishing structure of triazino linkage and C,H signals tentatively assigned to formaldehyde. Unsuppressed 1-bond couplings are indicated by brackets.

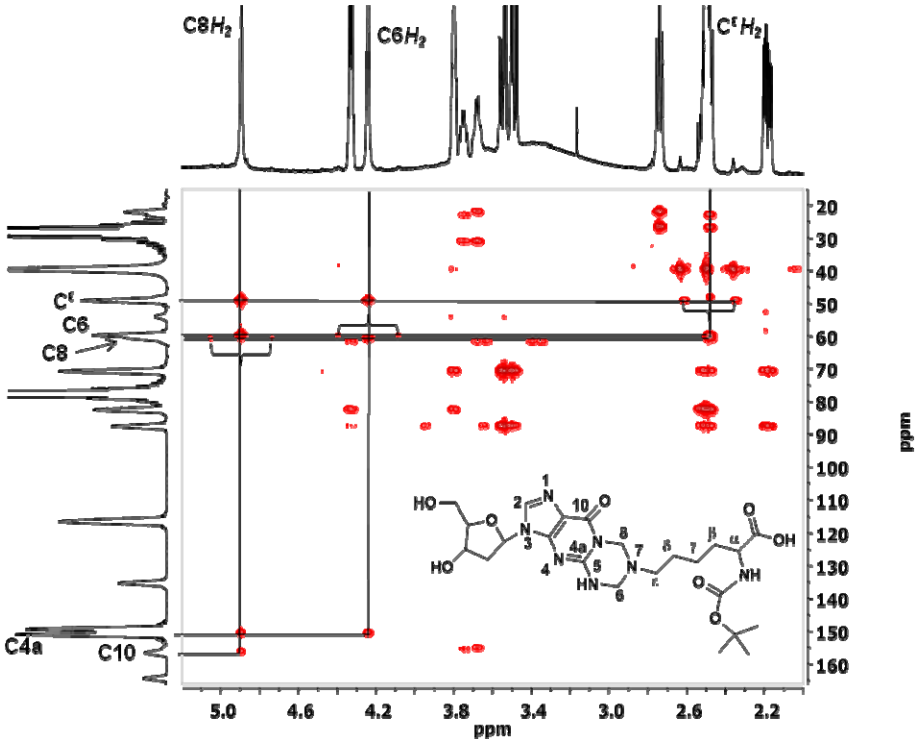


Figure 3.33. Expansion of HMBC spectrum of TPHA-1between 2.0 and 5.2 ppm on the ^{1}H-axis and 15 and 166 ppm on the ^{13}C-axis . Key signals are identified on the marginal ^{1}H and ^{13}C traces. Unsuppressed 1-bond couplings are indicated by brackets.

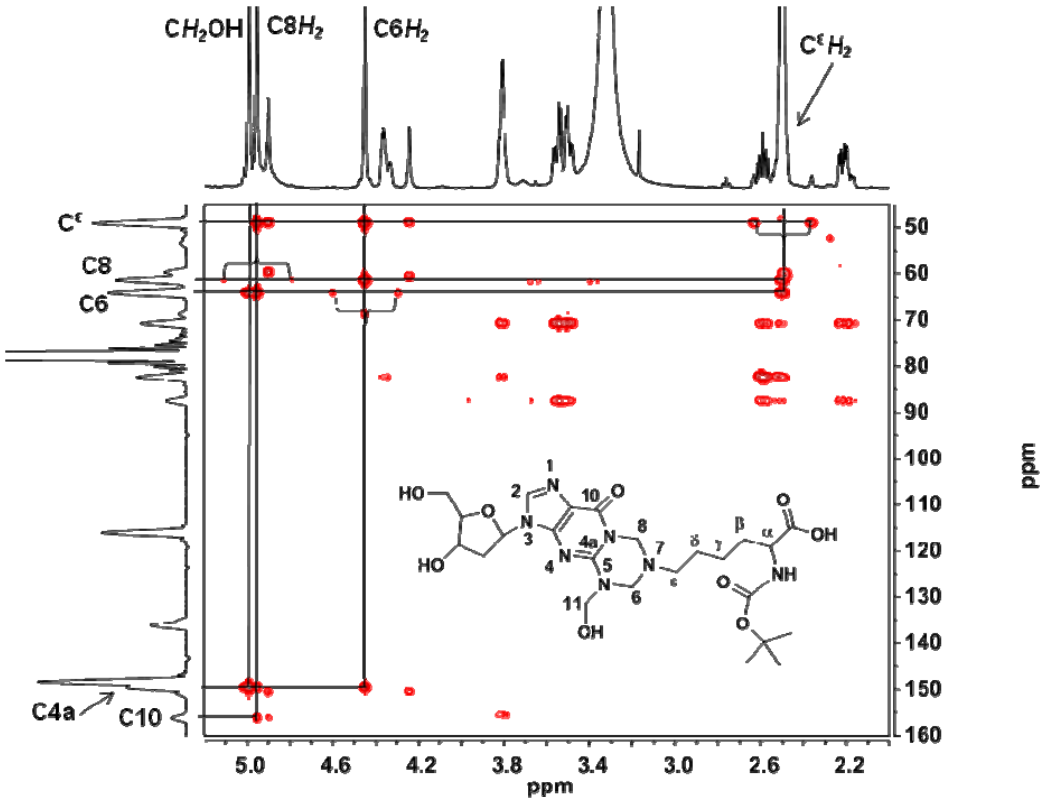


Figure 3.34. Expansion of HMBC spectrum of TPHA-2 between 2.0 and 5.2 ppm on the ^{1}H-axis and 40 and 160 ppm on the ^{13}C-axis . Key signals are identified on the marginal ^{1}H and ^{13}C traces. Unsuppressed 1-bond couplings are indicated by brackets.

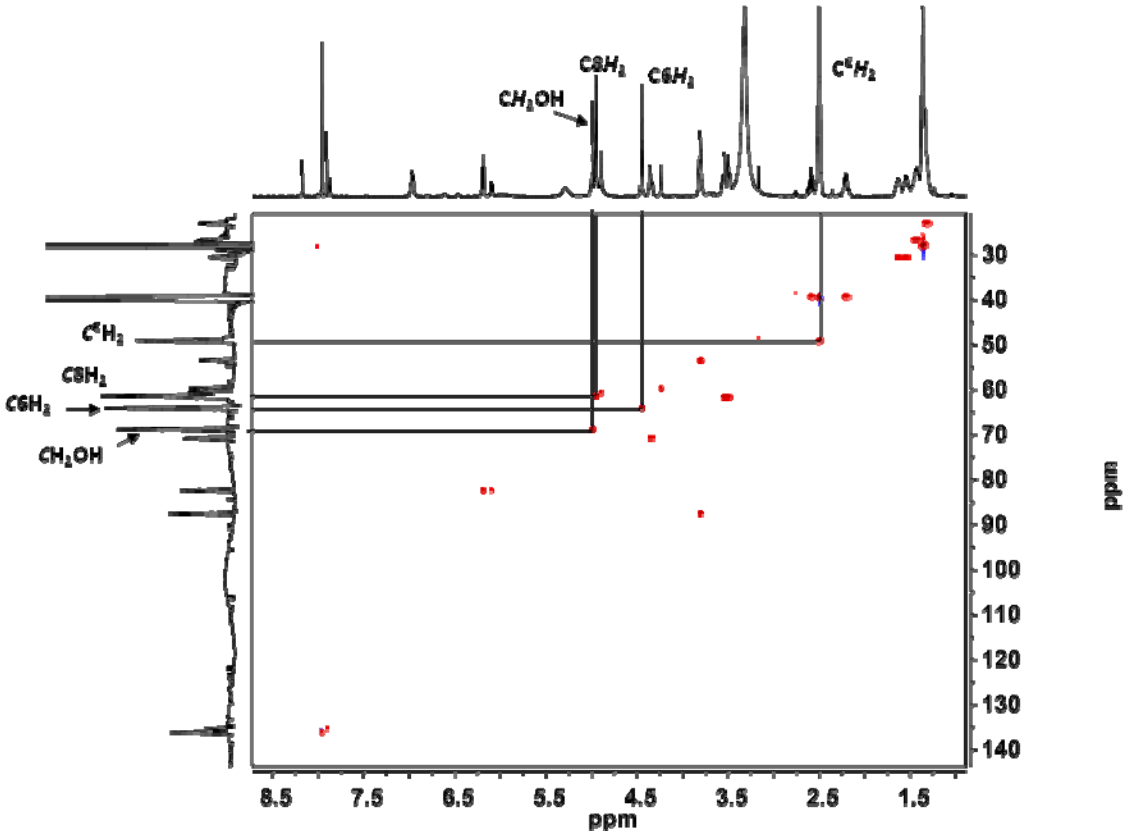


Figure 3.35. HSQC spectrum (DMSO-d_6) of the mixture containing TPHA-2, identifying the C,H cross-peaks discussed in the text.

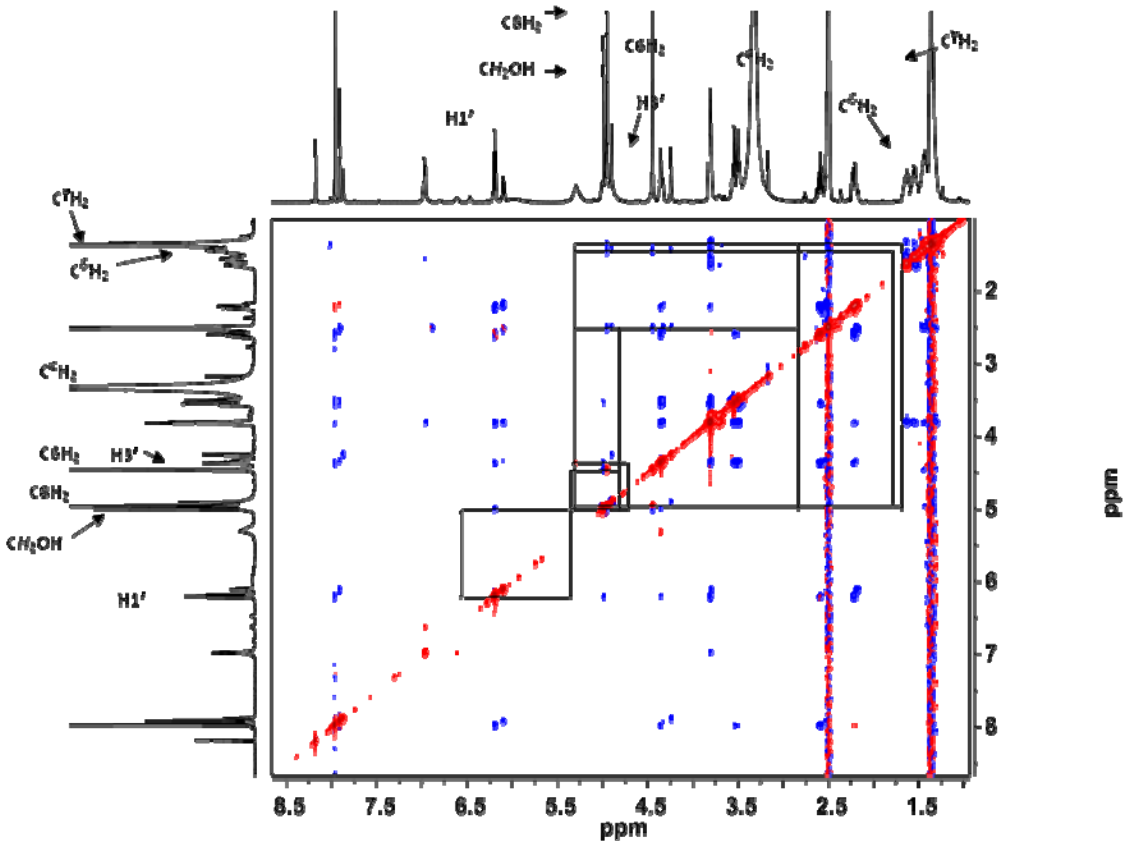


Figure 3.36. ROESY spectrum (500 MHz, DMSO-d_6) of TPHA-2. NOESY cross-peaks are present between protons of the three formaldehyde-derived methylene groups and Cγ, Cδ and Cε of hexanoic acid. The methylene at N5 also has NOESY interactions with H1′ and H3′. The NOESY interactions are consistent with the proposed tricyclic structure. Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

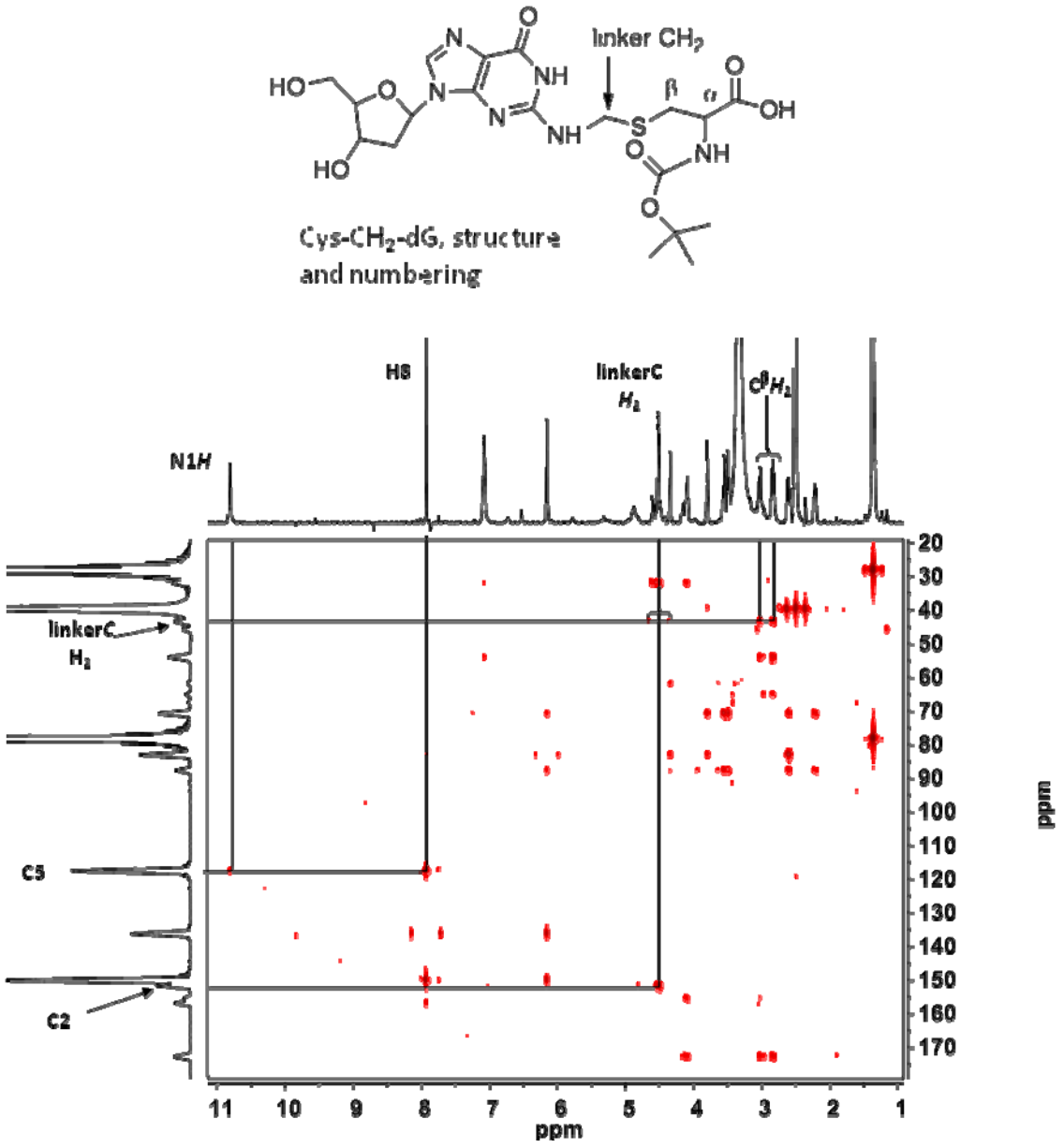


Figure 3.37. HMBC spectrum (DMSO-d_6) of Cys-CH_2-dG, indicating C,H connectivities establishing structure of the linkage. Unsuppressed 1-bond coupling between the linker methylene carbon and attached protons is indicated by a bracket.

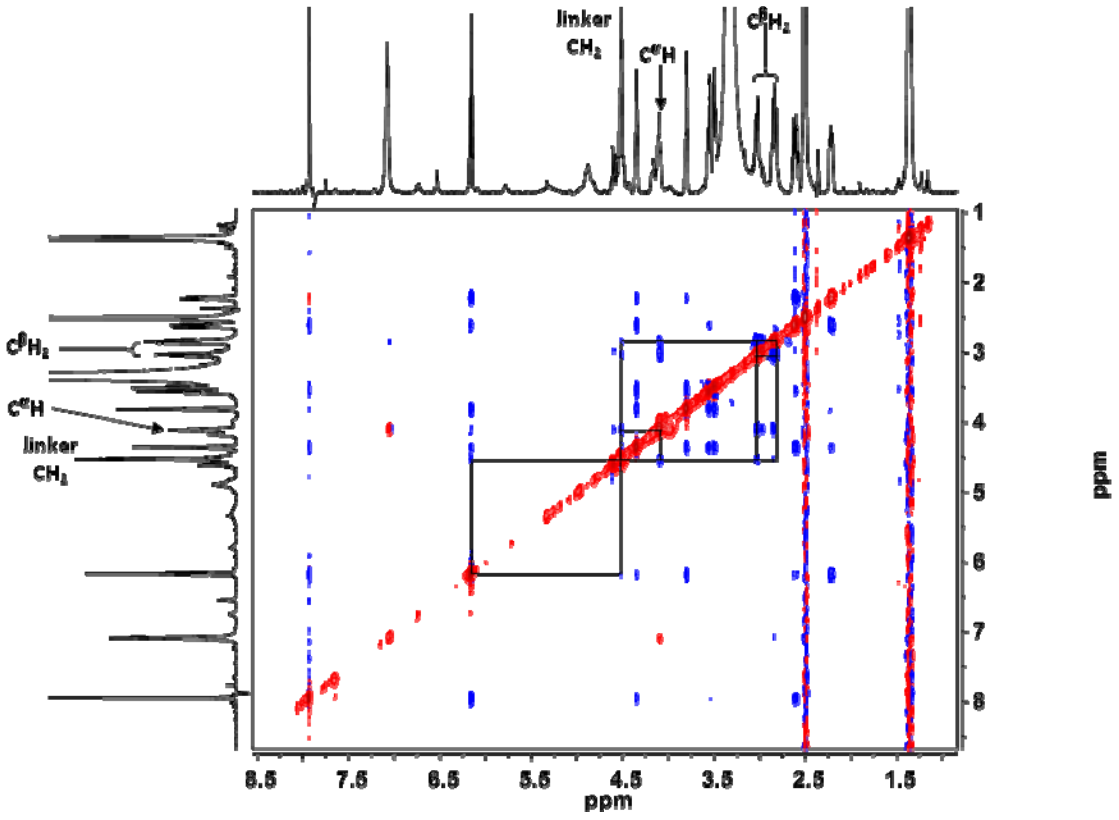


Figure 3.38. ROESY spectrum (500 MHz, DMSO-d_6) of Cys−CH_2−dG. NOESY connectivity is observed between linker methylene protons, Cys β-methylene protons and between linker methylene protons and H1′. Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

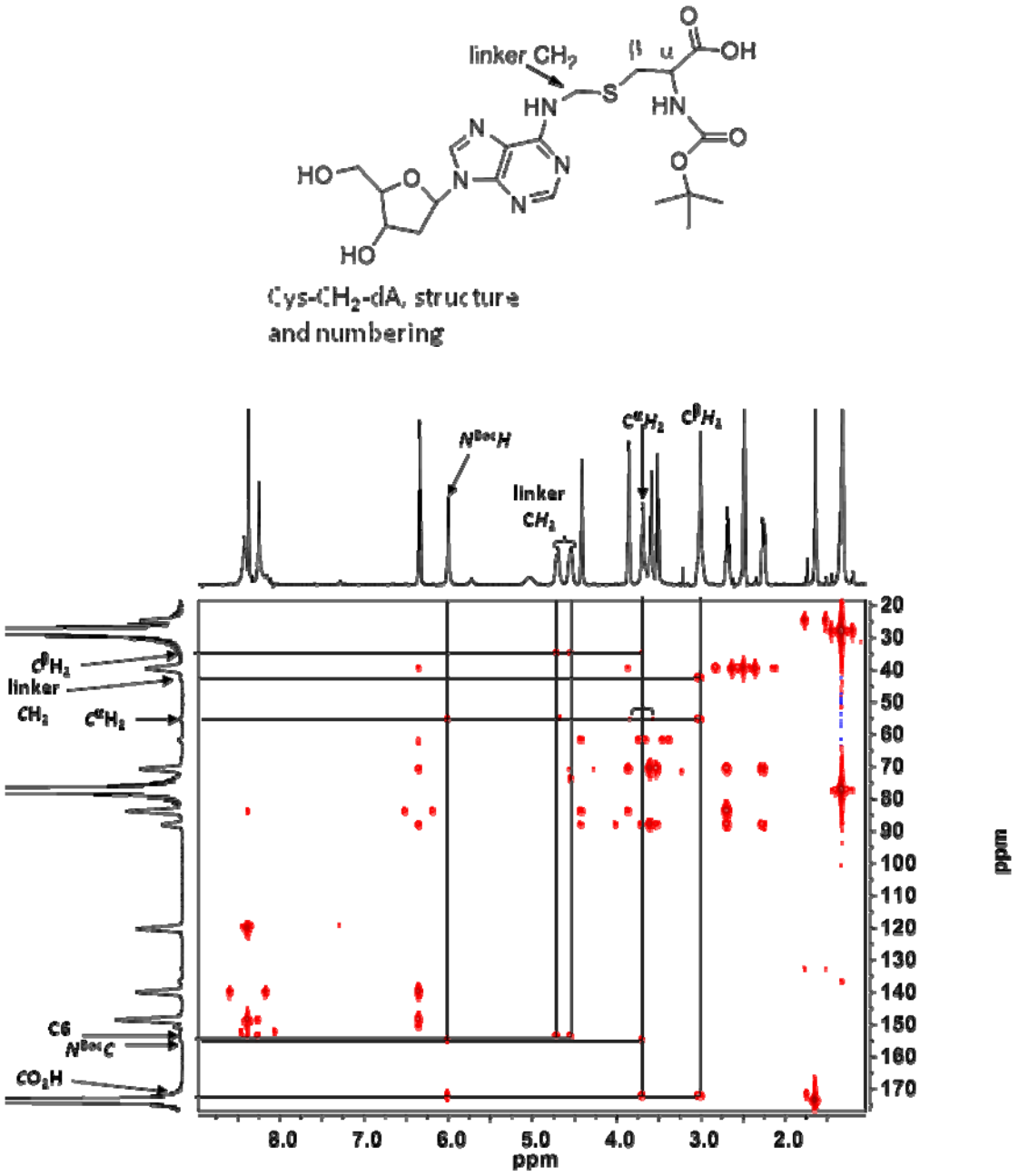


Figure 3.39. HMBC spectrum (DMSO-d_6) of dA−CH_2−Cys, indicating C−H connectivities between the linker methylene group and CβH2 of Cys and C6 of dA. Unsuppressed one-bond coupling is indicated by bracket.

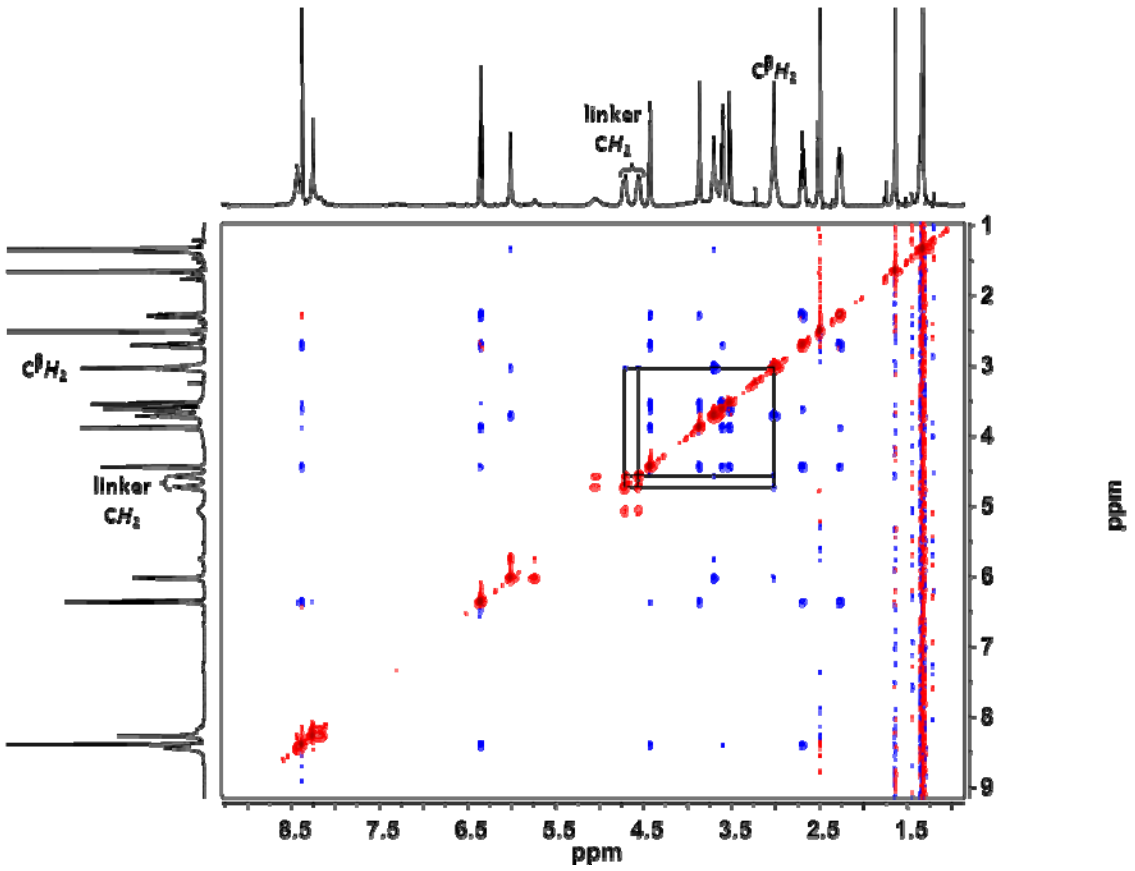


Figure 3.40. ROESY spectrum (500 MHz, DMSO-d_6) of dA−CH_2−Cys, showing NOESY interaction between the linker methylene protons and Cys CβH2. The positive phase of the cross-peaks between the linker methylene protons represents COSY interactions. Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

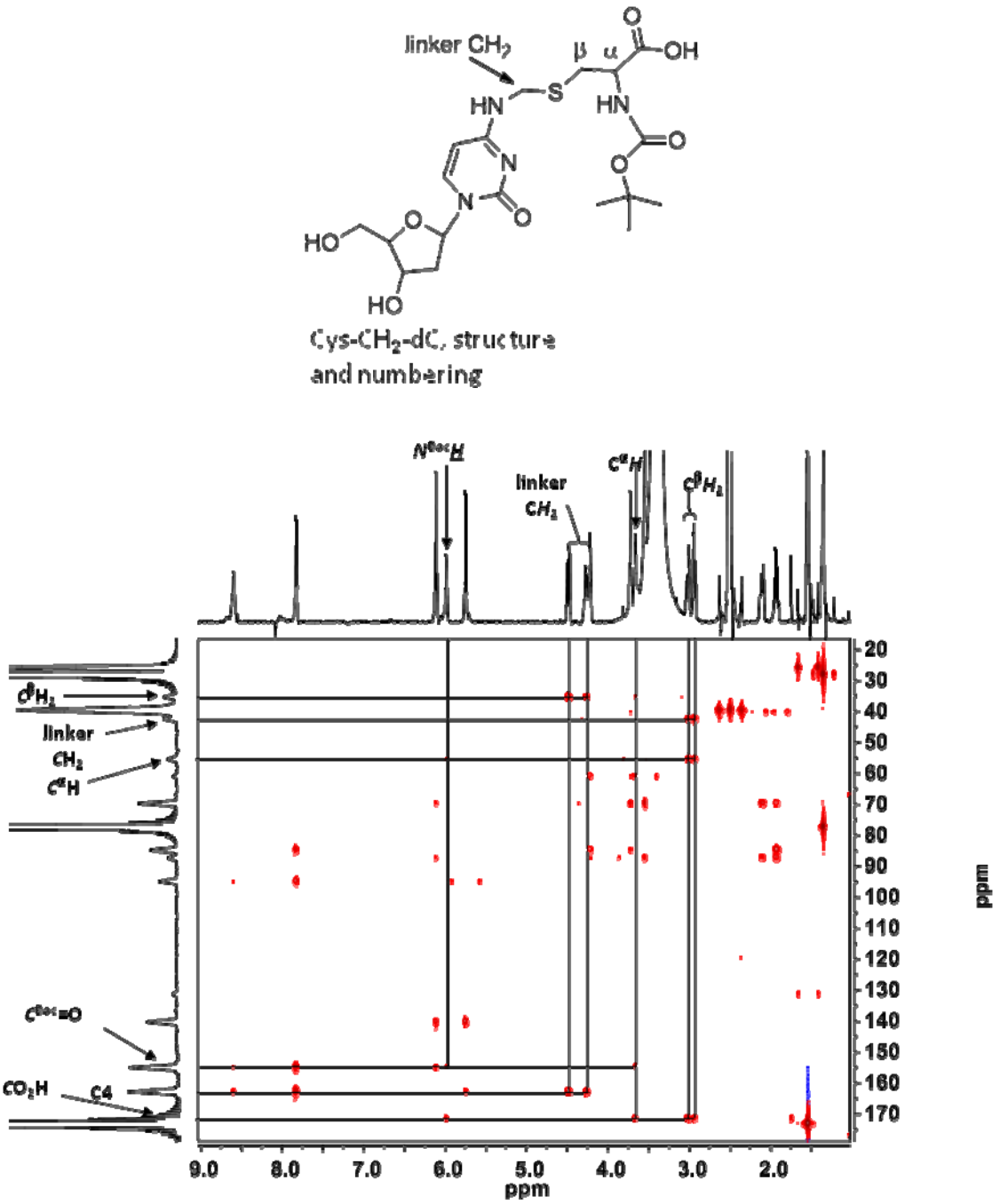


Figure 3.41. HMBC spectrum of dC−CH_2−Cys indicating C−H connectivities between the linker methylene, Cys CβH2 and C4 of dC.

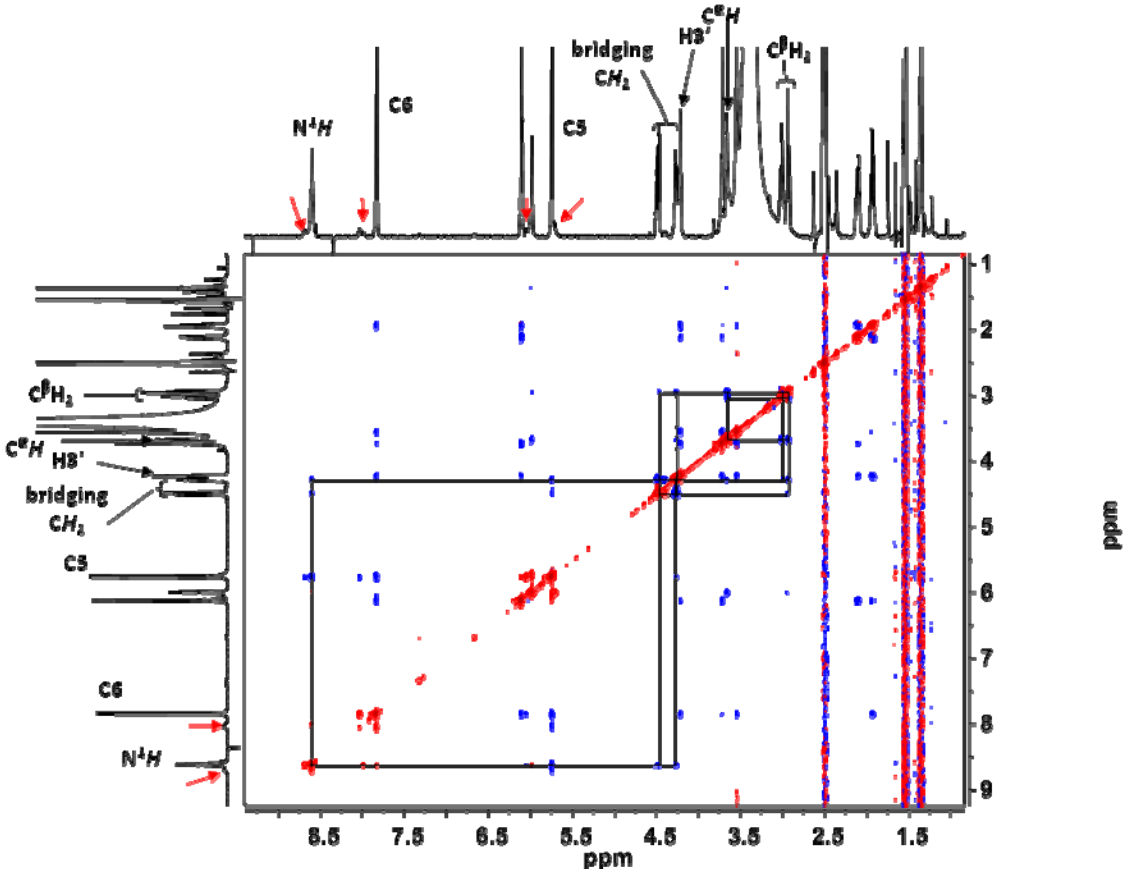


Figure 3.42. ROESY spectrum (500 MHz, DMSO-d_6) of Cys−CH_2−dC, indicating NOESY interactions between the bridging methylene protons, N4H, and CH_2 of Cys. Arrows in red indicate signals from a minor species related by exchange (red cross peaks). Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

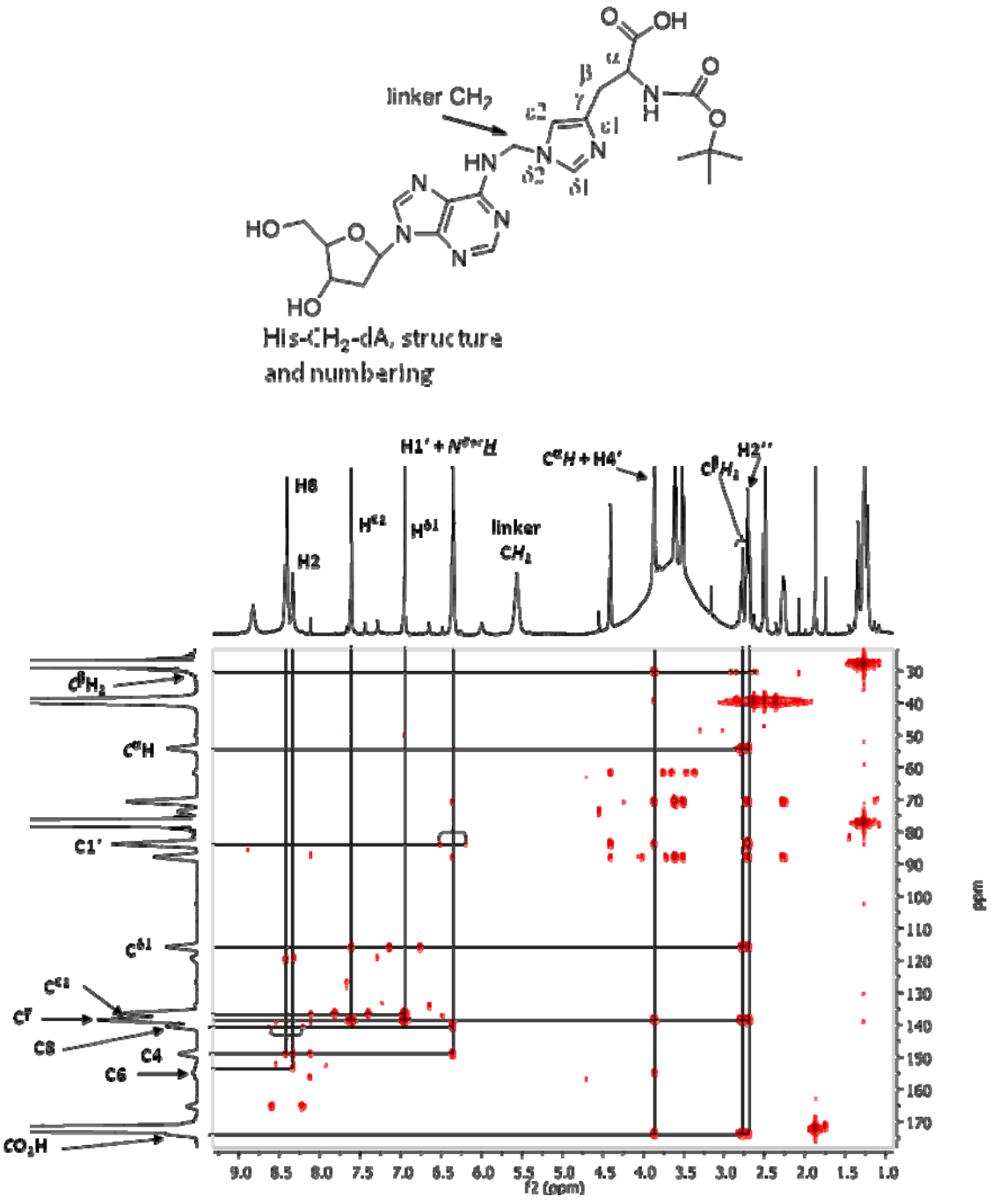


Figure 3.43. HMBC spectrum of His−CH_2−dA, shows C,H connectivities within nucleoside and His, but no connectivity through the linker.

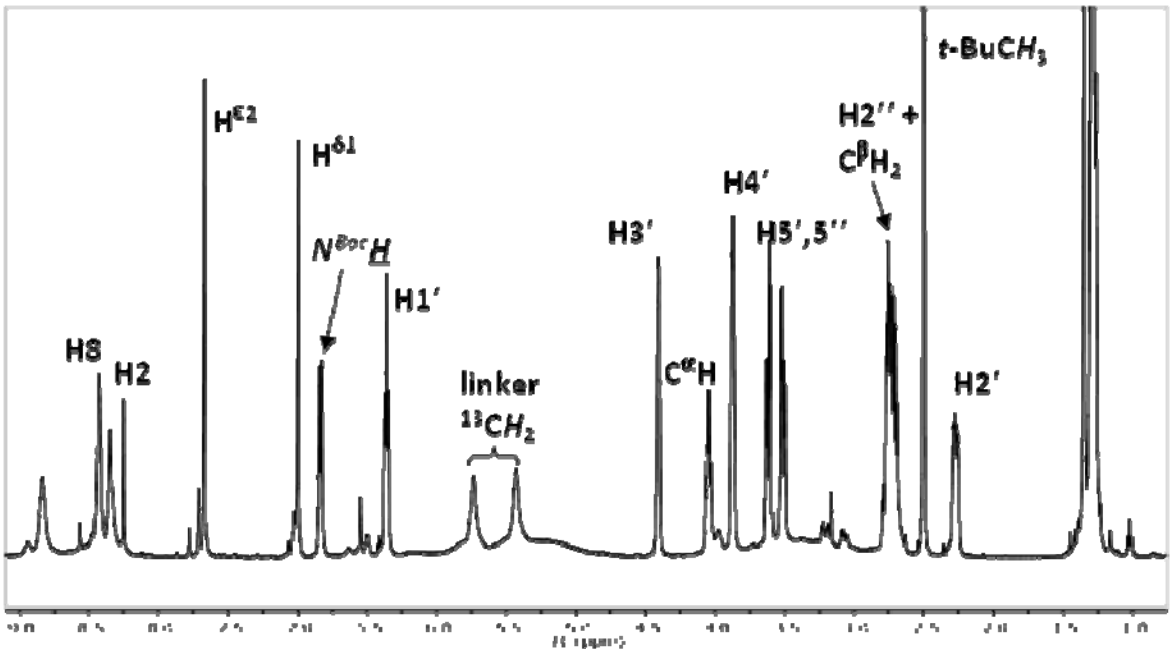


Figure 3.44. ^{1}H NMR (500 MHz, DMSO-d$_6$) of His−^{13}CH$_2$−dA, identifying the linker methylene signal at 5.59 ppm by ^{13}C splitting ($^{1}J_{C-H}$ = 153.2 Hz). In this trace, NBocH is resolved from H1′ and C$^{\alpha}$H from H4′.

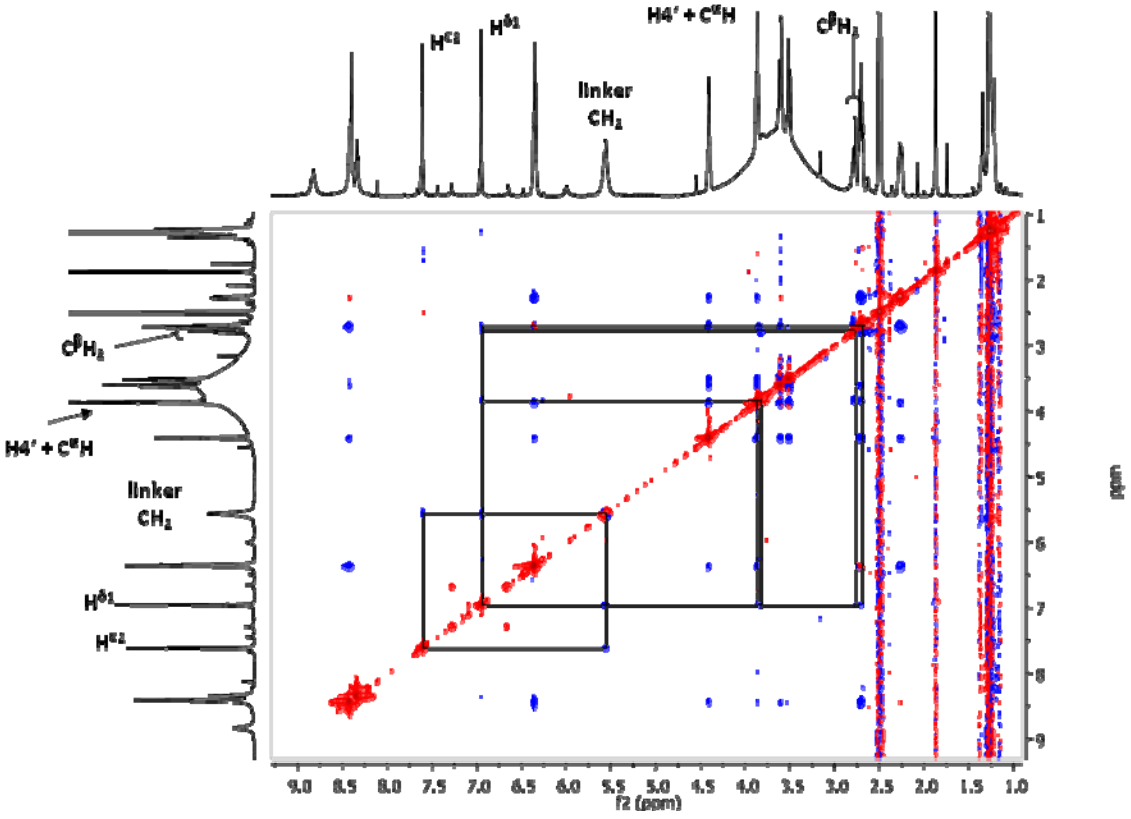


Figure 3.45. ROESY spectrum (500 MHz, DMSO-d_6) of His−CH_2−dA. The linker CH_2 exhibits NOESY cross-peaks only with the imidazole C−H protons. Connectivity between the nucleoside and His is indicated by a NOESY cross-peak between the His Hδ1 and H4′ of dA. Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

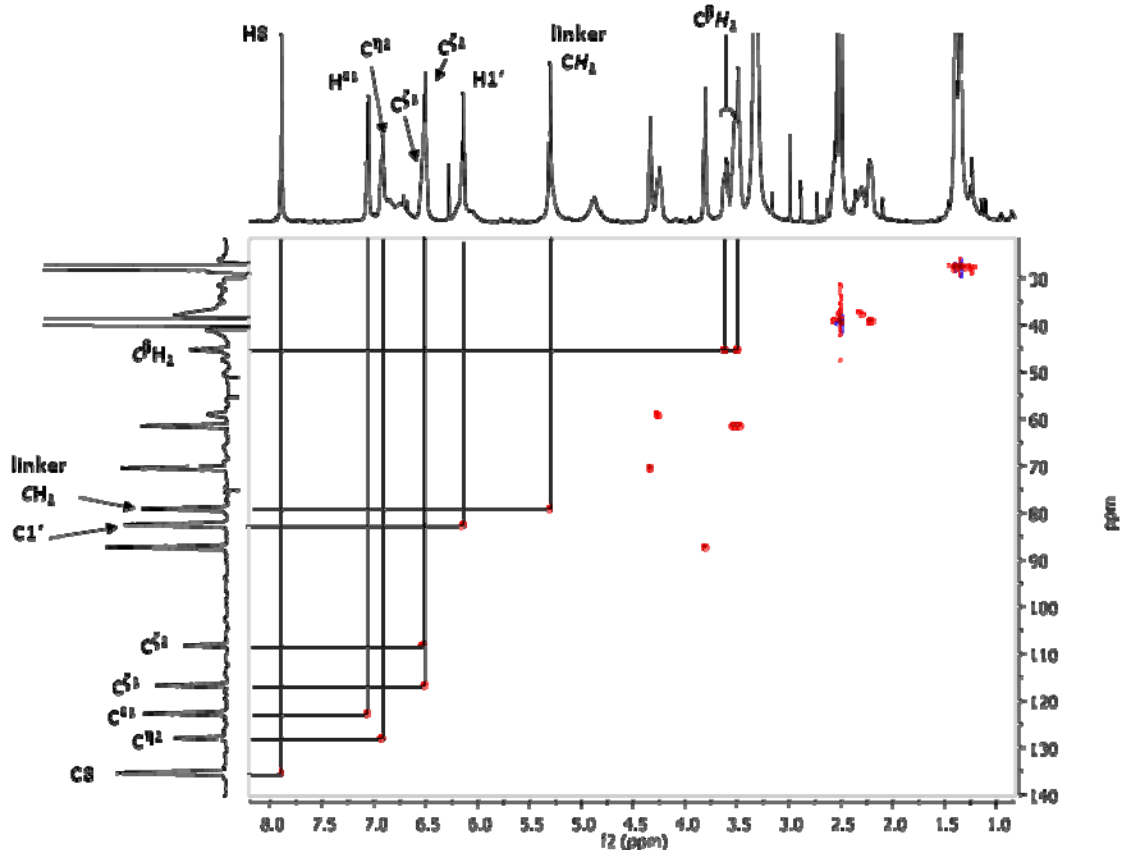


Figure 3.46. HSQC spectrum (DMSO-d_6) of Trp−CH_2−dG identifying the ^{1}H and ^{13}C signals of the bridging CH_2 and Trp $C^{\beta}H_2$ and demonstrating that four of the indole carbons of Trp have attached hydrogens.

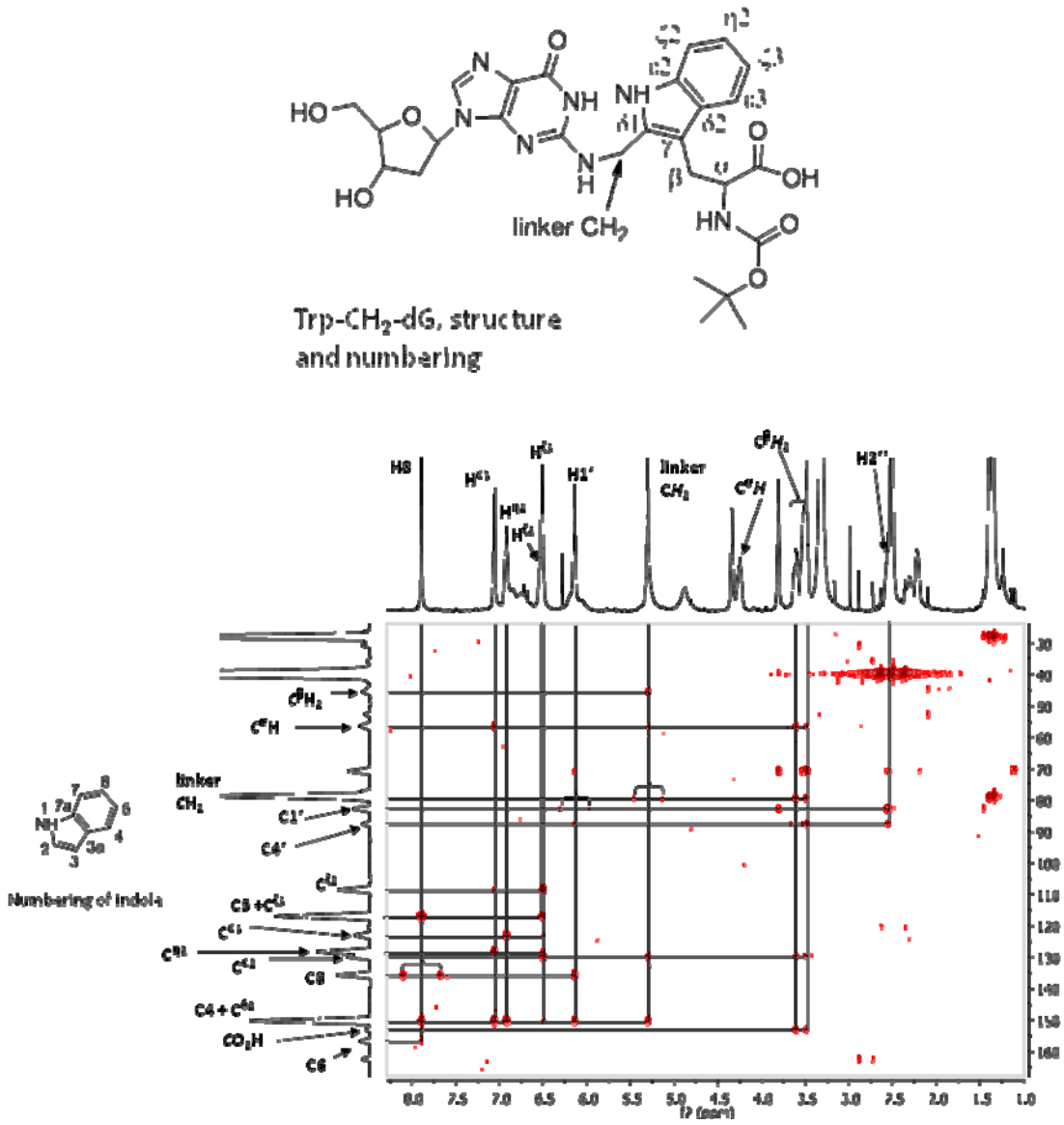


Figure 3.47. HMBC spectrum of Trp−CH_2−dG (DMSO-d_6), establishing C,H connectivity between linker CH_2 and Trp. Additional C,H connectivities are indicated on the spectrum. Unsuppressed 1-bond couplings are indicated on data by brackets. Numbering convention is according to citation [*].

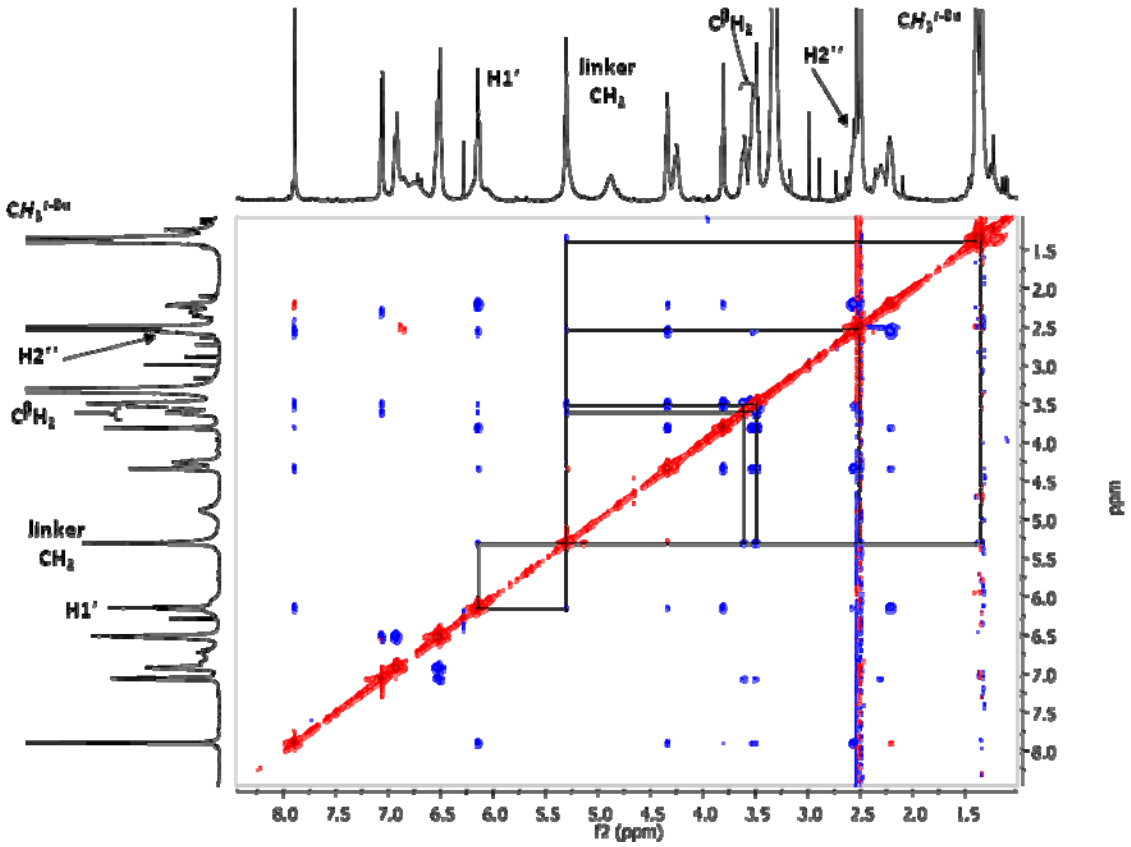


Figure 3.48. ROESY spectrum (500 MHz, DMSO-d_6) of Trp−CH_2−dG, indicating NOESY interactions of the linker CH_2 with Trp (and the deoxyribose). Red cross-peaks are positively phased and blue cross-peaks are negatively phased.

Tables

Table 3.1. Relative yielda of formaldehyde-induced cross-links formed over 48 h by reaction between 5 mM amino acid and 50 mM formaldehyde.

	dA	**dT**	**dC**	**dG**
Lys	-	-	-	**1.0**[b]
Cys	**0.05**	-	**0.02**	**0.53**
His	**0.04**	-	-	-
Trp	-	-	-	**0.03**

[a]Relative yields based on integration of HPLC peaks in UV traces acquired at a detector wavelength of 254 nm.
[b]Sum of three cross-linked products.

Table 3.2. Exact masses for the reaction products of trinucleotides and Nα-Boc-protected amino acids cross-linked by formaldehyde, determined by negative ion ESI-QTOF-MS.

Cross-links	Experimental Mass ($[M-H]^-$)	Calculated Mass ($[M-H]^-$)	Composition	Δ (ppm)
TGT—CH_2—Lys	1132.3360	1132.3395	$C_{42}H_{60}N_{11}O_{22}P_2^-$	–3.1
T(TPHA-1)T	1144.3373	1144.3395	$C_{43}H_{60}N_{11}O_{22}P_2^-$	–1.9
T(TPHA-2)T	1174.3494	1174.3501	$C_{44}H_{62}N_{11}O_{23}P_2^-$	–0.6
TGT—CH_2—Trp	1190.3264	1190.3239	$C_{47}H_{58}N_{11}O_{22}P_2^-$	2.2
TGT—CH_2—Cys	1107.2526	1107.2537	$C_{39}H_{53}N_{10}O_{22}P_2S^-$	–1.0
TAT—CH_2—His	1125.3048	1125.3085	$C_{42}H_{55}N_{12}O_{21}P_2^-$	3.3
TAT—CH_2—Cys	1091.2556	1091.2588	$C_{39}H_{53}N_{10}O_{21}P_2S^-$	–2.9
TCT—CH_2—Cys	1067.2464	1067.2476	$C_{38}H_{53}N_8O_{22}P_2S^-$	–1.1

Table 3.3. Exact masses of protonated 8-mers cross-linked to deoxynucleosides by formaldehyde determined by positive ion ESI-QTOF-MS.

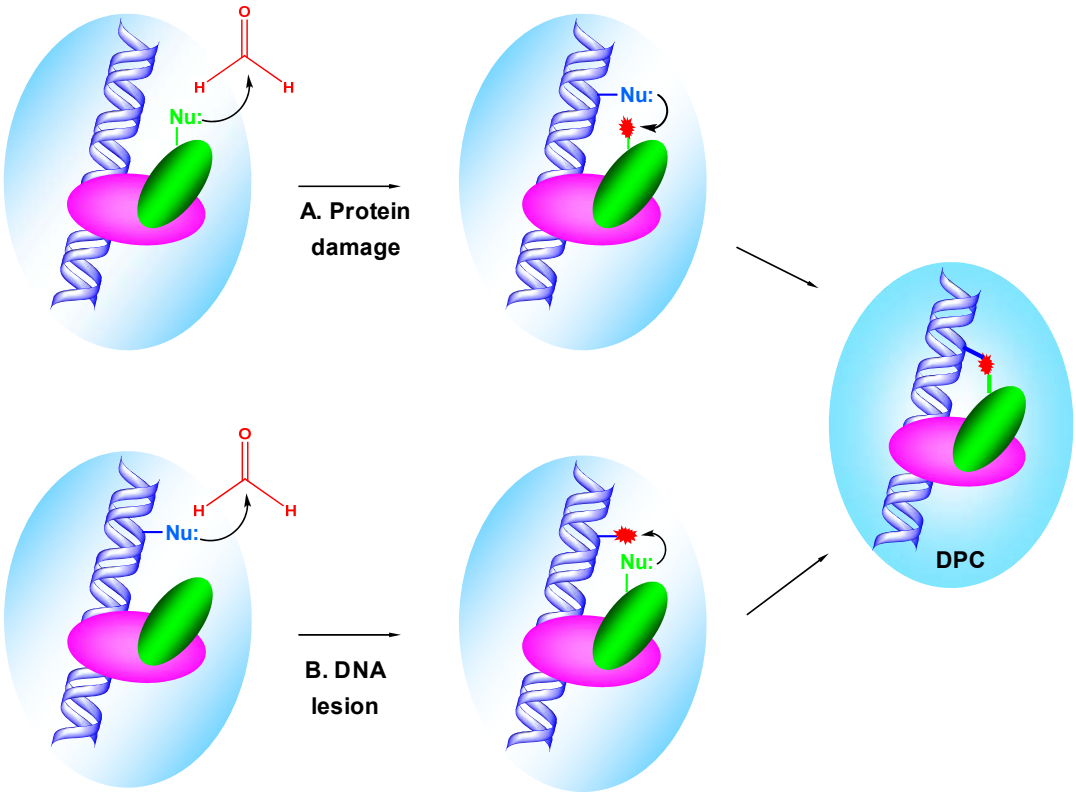
Cross-links	Experimental Mass ($[MH]^+$)	Calculated Mass ($[MH]^+$)	Composition	Δ ppm
Acetyl-VEGG(K-CH_2-dG)GAA	1009.4687	1009.4697	$C_{41}H_{65}N_{14}O_{16}^+$	–1.0
Acetyl-VEGG(TPHA-1) GAA	1021.4702	1021.4697	$C_{42}H_{65}N_{14}O_{16}^+$	0.5
Acetyl-VEGG(TPHA-2)GAA	1051.4805	1051.4803	$C_{43}H_{67}N_{14}O_{17}^+$	0.2
Acetyl-GEGG(W-CH_2-dG)GAA	1025.4048	1025.4071	$C_{43}H_{57}N_{14}O_{16}^+$	–2.2
Acetyl-GEGG(C-CH_2-dG)GAA	942.3372	942.3370	$C_{35}H_{52}N_{13}O_{16}S^+$	0.2
Acetyl-VEGG(H- CH_2-dA)GAA	1002.4383	1002.4388	$C_{41}H_{60}N_{15}O_{15}^+$	–0.5
Acetyl-GEGG(C-CH_2-dC)GAA	902.3304	902.3309	$C_{34}H_{52}N_{11}O_{16}S^+$	–0.6
Acetyl-GEGG(C-CH_2-dA)GAA	926.3427	926.3421	$C_{35}H_{52}N_{13}O_{15}S^+$	0.6

Table 3.4. Exact masses of formaldehyde-induced deoxynucleoside-amino acid cross-links by FTICR-MS.

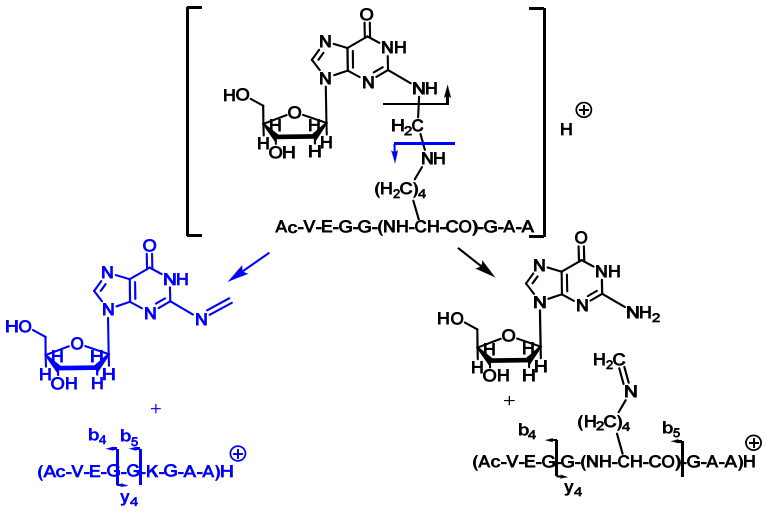
Cross-links	Experimental Mass ($[MH]^+$)	Calculated Mass ($[MH]^+$)	Composition	Δ (ppm)
Lys- CH_2-dG	526.2622	526.2619	$C_{22}H_{35}N_7O_8$	0.6
TPHA-1	538.2623	538.2619	$C_{23}H_{35}N_7O_8$	0.7
TPHA -2	568.2726	568.2725	$C_{24}H_{37}N_7O_9$	0.2
Trp- CH_2-dG	584.2465	584.2463	$C_{27}H_{33}N_7O_8$	0.3
Cys-CH_2-dG	501.1763	501.1762	$C_{19}H_{28}N_6O_8S$	0.2
His- CH_2-dA	519.2312	519.2310	$C_{22}H_{30}N_8O_7$	0.4
Cys- CH_2-dA	485.1813	485.1813	$C_{19}H_{28}N_6O_7S$	0
Cys- CH_2-dC	461.1701	461.1700	$C_{19}H_{28}N_6O_8S$	0.2

Schemes

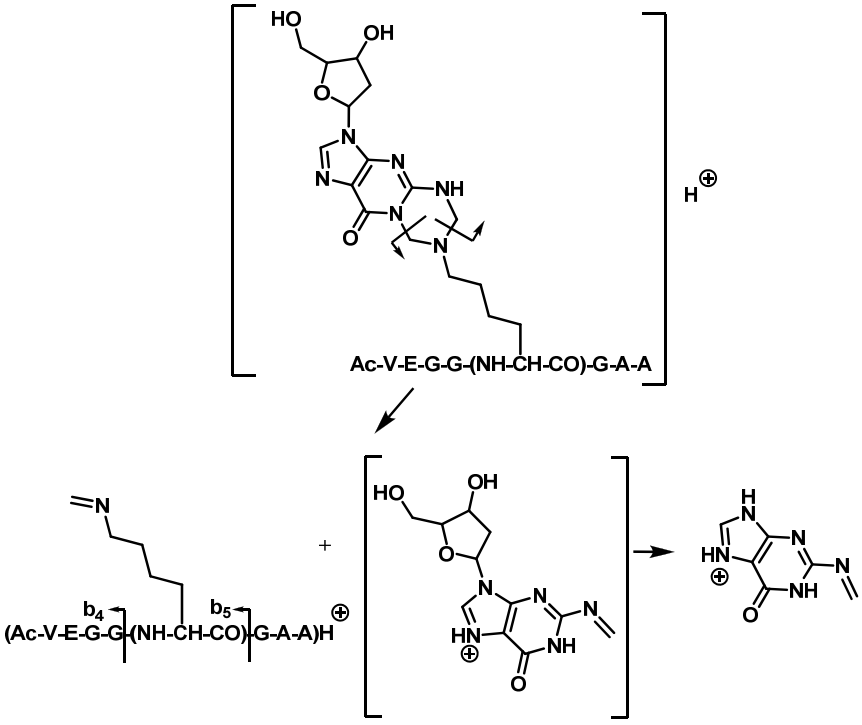
Scheme 3.1. The formation of formaldehyde-induced DPCs originating from the initial attack of formaldehyde on protein residues (A) and from the initial attack of formaldehyde on DNA (B).



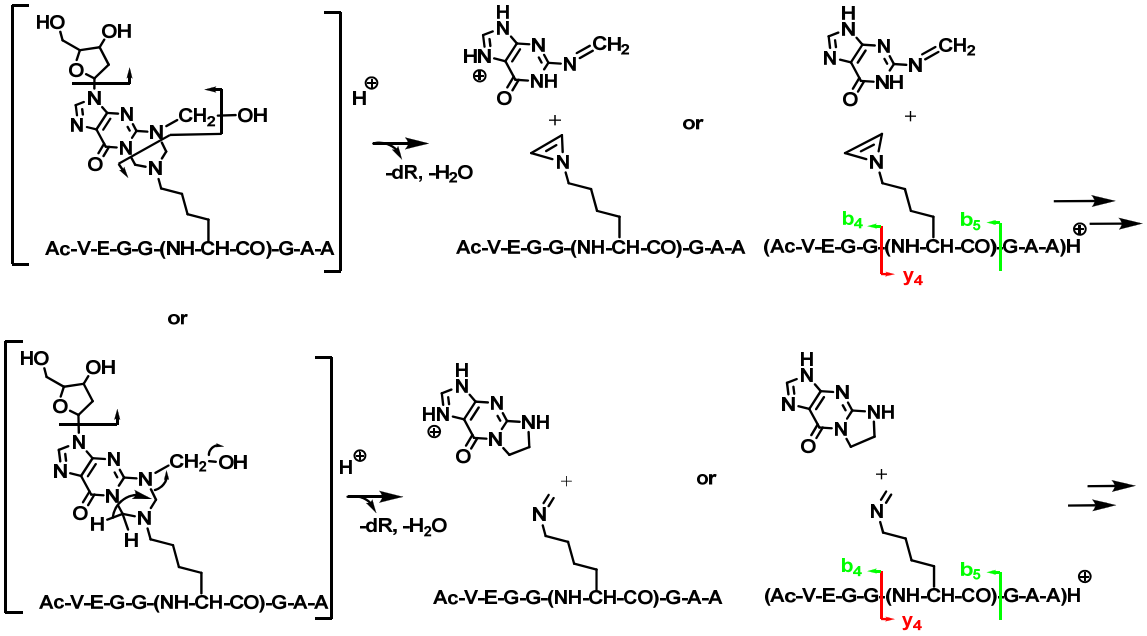
Scheme 3.2. Fragmentation of AcVEGGC(-CH_2-dG)GAA.



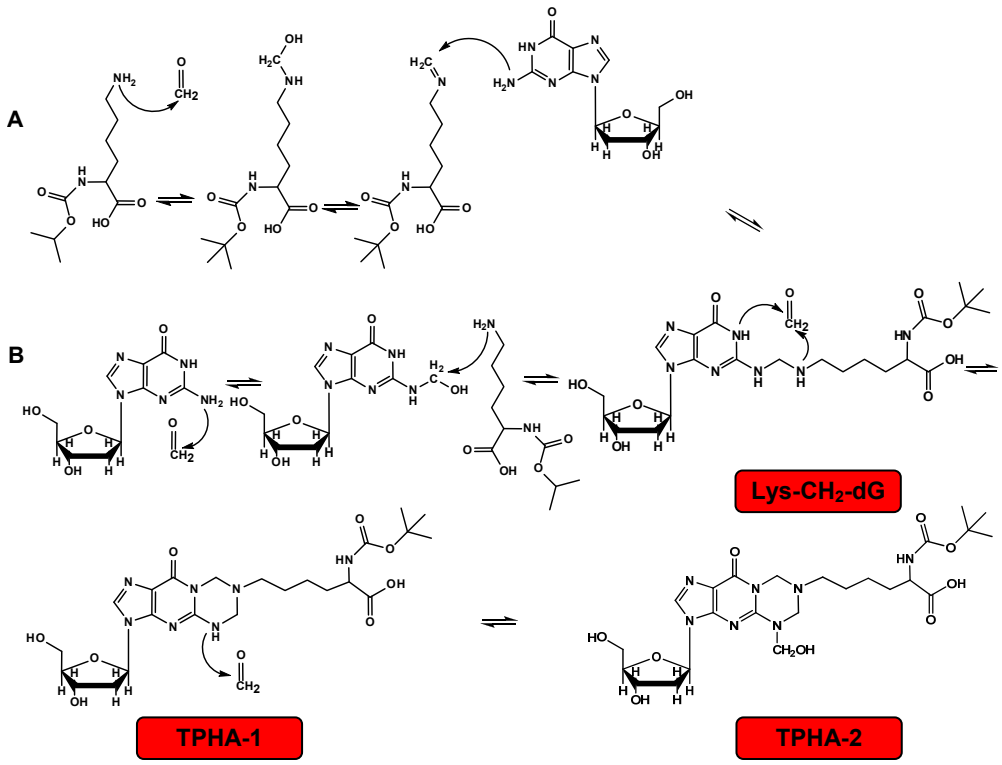
Scheme 3.3. Fragmentation of AcVEGG(TPHA-1)GAA.



Scheme 3.4. Fragmentation of AcVEGG(TPHA-2)GAA.



Scheme 3.5. Mechanism proposed for formation of cross-linked formaldehyde adducts. A: initial attack by the amino acid-formaldehyde adduct; B: initial attack by the nucleoside adduct.



Charts

Chart 3.1. Structures of cross-linked adducts between amino acids and nucleosides identified in this study. Formaldehyde-derived linkages are shown in red.

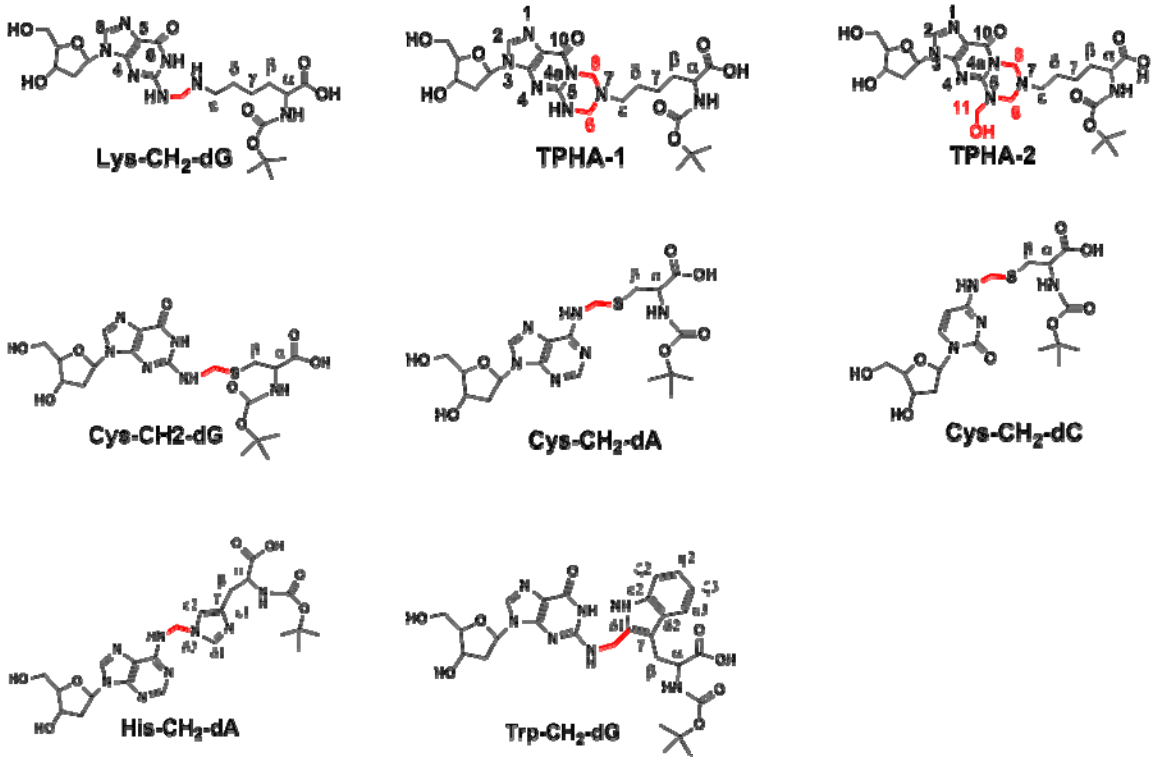
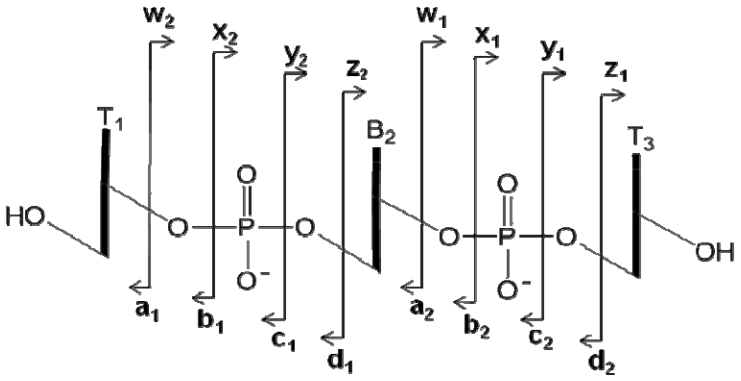


Chart 3.2. Nomenclature of ions after the fragmentation of trinucleotides



References

(1) IARC monographs on the evaluation of carcinogenic risks to humans. International Agency for Research on Cancer (2006) *IARC Monogr Eval. Carcinog. Risks Hum.* **88**, 1-287.

(2) Swenberg, J. A., Richardson, F. C., Boucheron, J. A., Deal, F. H., Belinsky, S. A., Charbonneau, M., and Short, B. G. (1987) High- to low-dose extrapolation: critical determinants involved in the dose response of carcinogenic substances. *Environ. Health Perspect.* **76**, 57-63.

(3) Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J., and Swenberg, J. A. (1983) Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Res.* **43**(9), 4382-4392.

(4) Monticello, T. M., Swenberg, J. A., Gross, E. A., Leininger, J. R., Kimbell, J. S., Seilkop, S., Starr, T. B., Gibson, J. E., and Morgan, K. T. (1996) Correlation of regional and nonlinear formaldehyde-induced nasal cancer with proliferating populations of cells. *Cancer Res.* **56**(5), 1012-1022.

(5) Speit, G., Schutz, P., and Merk, O. (2000) Induction and repair of formaldehyde-induced DNA-protein crosslinks in repair-deficient human cell lines. *Mutagenesis* **15**(1), 85-90.

(6) Merk, O., and Speit, G. (1998) Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ. Mol. Mutagen.* **32**(3), 260-268.

(7) Cheng, G., Wang, M., Upadhyaya, P., Villalta, P. W., and Hecht, S. S. (2008) Formation of formaldehyde adducts in the reactions of DNA and deoxyribonucleosides with alpha-acetates of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and N-nitrosodimethylamine (NDMA). *Chem. Res. Toxicol.* **21**(3), 746-751.

(8) Lu, K., Ye, W., Gold, A., Ball, L. M., and Swenberg, J. A. (2009) Formation of S-[1-(N2-deoxyguanosinyl)methyl]glutathione between glutathione and DNA induced by formaldehyde. *J. Am. Chem. Soc.* **131**(10), 3414-3415.

(9) Siomin, Y. A., Simonov, V. V., and Poverenny, A. M. (1973) The reaction of formaldehyde with deoxynucleotides and DNA in the presence of amino acids and lysine-rich histone. *Biochim. Biophys. Acta* **331**(1), 27-32.

(10) McGhee, J. D., and von Hippel, P. H. (1977) Formaldehyde as a probe of DNA structure. 3. Equilibrium denaturation of DNA and synthetic polynucleotides. *Biochemistry* **16**(15), 3267-3276.

(11) McGhee, J. D., and von Hippel, P. H. (1977) Formaldehyde as a probe of DNA structure. r. Mechanism of the initial reaction of Formaldehyde with DNA. *Biochemistry* **16**(15), 3276-3293.

(12) McGhee, J. D., and von Hippel, P. H. (1975) Formaldehyde as a probe of DNA structure. II. Reaction with endocyclic imino groups of DNA bases. *Biochemistry* **14**(6), 1297-1303.

(13) McGhee, J. D., and von Hippel, P. H. (1975) Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry* **14**(6), 1281-1296.

(14) Metz, B., Kersten, G. F., Baart, G. J., de, J. A., Meiring, H., ten, H. J., van Steenbergen, M. J., Hennink, W. E., Crommelin, D. J., and Jiskoot, W. (2006) Identification of formaldehyde-induced modifications in proteins: reactions with insulin. *Bioconjug. Chem.* **17**(3), 815-822.

(15) Metz, B., Kersten, G. F., Hoogerhout, P., Brugghe, H. F., Timmermans, H. A., de, J. A., Meiring, H., ten, H. J., Hennink, W. E., Crommelin, D. J., and Jiskoot, W. (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* **279**(8), 6235-6243.

(16) Lu, K., Boysen, G., Gao, L., Collins, L. B., and Swenberg, J. A. (2008) Formaldehyde-induced histone modifications in vitro. *Chem. Res. Toxicol.* **21**(8), 1586-1593.

(17) Chaw, Y. F., Crane, L. E., Lange, P., and Shapiro, R. (1980) Isolation and identification of cross-links from formaldehyde-treated nucleic acids. *Biochemistry* **19**(24), 5525-5531.

(18) Huang, H.F., Solomon, M.S., Hopkins, P. B. (1992) Formaldehyde preferentially interstrand cross-links duplex through deoxyadenosine residues at the sequence 5'-d(AT). *J. Am. Chem. Soc.* **114** (23), 9240-9241.

(19) Huang, H.F., Solomon, M.S., Hopkins, P. B. (1993) DNA interstrand cross-linking by formaldehyde: Nucleotide sequence preference and covalent

structure of the predominant cross-link formed in synthetic oligonucleotides. *J. Am. Chem. Soc.* **115** (21), 9402-9408.

(20) Quievryn, G., and Zhitkovich, A. (2000) Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* **21**(8), 1573-1580.

(21) Solomon, M. J., and Varshavsky, A. (1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U. S. A* **82**(19), 6470-6474.

(22) Brodolin, K. L., Studitskii, V. M., and Mirzabekov, A. D. (1993) [Study of the structure of Escherichia coli RNA polymerase and its complex with the lacUV5-promotor using protein-protein and DNA-protein crosslinks, formed by formaldehyde]. *Mol. Biol. (Mosk)* **27**(5), 1085-1093.

(23) Koc, H., and Swenberg, J. A. (2002) Applications of mass spectrometry for quantitation of DNA adducts. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **778**(1-2), 323-343.

(24) Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., and Tainer, J. A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* **11**(8), 714-720.

(25) Vrkic, A. K., and O'Hair, R. A. (2000) Fragmentation Reactions of All 64 Deprotonated Trinucleotide and 16 Mixed Base Tetranucleotide Anions via Tandem Mass Spectrometry in an Ion Trap. *Aust. J. Chem.*, 2000, **53**(4), 307–319.

(26) Paizs, B., and Suhai, S. (2004) Towards understanding the tandem mass spectra of protonated oligopeptides. 1: mechanism of amide bond cleavage. *J. Am. Soc. Mass Spectrom.* **15**(1), 103-113.

(27) Tammler, U., Quillan, J. M., Lehmann, J., Sadee, W., and Kassack, M. U. (2003) Design, synthesis, and biological evaluation of non-peptidic ligands at the Xenopus laevis skin-melanocortin receptor. *Eur. J. Med. Chem.* **38**(5), 481-493.

(28) Zhao, M., Bi, L., Wang, W., Wang, C., Baudy-Floc'h, M., Ju, J., and Peng, S. (2006) Synthesis and cytotoxic activities of beta-carboline amino acid ester conjugates. *Bioorg. Med. Chem.* **14**(20), 6998-7010.

CHAPTER 4

THE FORMATION OF S-[1-(N2-DEOXYGUANOSINYL)METHYL]GLUTATHIONE BETWEEN GLUTATHIONE AND DNA INDUCED BY FORMALDEHYDE

This paper has been published and therefore is reproduced with permission from [Kun Lu, Wenjie Ye, Avram Gold, Louise M. Ball and James A. Swenberg. Formation of S-[1-(N2-Deoxyguanosinyl)methyl]glutathione between Glutathione and DNA Induced by Formaldehyde. *J. Am. Chem. Soc., 2009, 131 (10), pp 3414–3415*] Copyright [2009] American Chemical Society.

4.1 Introduction

Formaldehyde, an essential metabolic intermediate generated endogenously from serine, glycine, methionine and choline and also produced from some metabolites and proteins by demethylation*(1)*, is present in human blood at about 0.1 mM*(2)*. Formaldehyde can also enter the body through environmental exposures. Formaldehyde forms DNA and protein adducts and DNA-protein crosslinks, and its toxicity has been the object of intensive investigation*(3-13)*. Previous studies have shown that formaldehyde is genotoxic*(14)*. Although N^6-dA, N^2-dG and N^4-dC adducts of formaldehyde are found *in vitro(15-18)*, no exogenous formaldehyde-induced DNA adducts have ever been detected in animals exposed by inhalation. This inability to detect DNA adducts may result from

rapid binding of formaldehyde by the tripeptide glutathione (GSH), which significantly decreases the chance that exogenous formaldehyde will attack DNA directly.

GSH is a major reducing thiol present in all human cells at a concentration around 5 mM. Formaldehyde (**1**; Scheme 4.1) reacts spontaneously with GSH (**2**) to form *S*-hydroxymethylglutathione (**3**). Formaldehyde dehydrogenase (*ADH3*) oxidizes **3** to *S*-formylglutathione, which is then hydrolyzed to formate by *S*-formylglutathione hydrolase, regenerating free glutathione *(1)*. The *S*-hydroxymethyl group of **3** is a reactive target for nucleophilic substitution. Recent work in our laboratory on formaldehyde-induced DNA-protein crosslinks shows that the thiol groups of cysteine residues can readily crosslink with DNA bases in the presence of formaldehyde, raising the possibility that **3** can conjugate with DNA as shown in Scheme 4.1.

4.2 Materials and Methods

4.2.1 Chemicals and Enzymes

Glutathione, deoxyguanosine, calf thymus DNA, potassium phosphate, Tris-HCl, $MgCl_2$, formic acid, methanol, acetonitrile, HPLC grade water were all purchased from Sigma (St. Louis, MO). 20% formaldehyde in water was procured from Tousimis (Rockville, MD). DNase I was purchased from New England Biolabs (Ipswich, MA). Alkaline phosphatase and phosphodiesterases were ordered from Fisher Scientific (Pittsburgh, PA). All chemicals were used as received unless otherwise stated.

4.2.2 Instrumentation

High Performance Liquid Chromatography (HPLC). The purification of *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione was carried out on an Agilent 1200 series HPLC system equipped with a diode-array detector (Santa Clara, CA). Analytes were separated by reverse phase chromatography using a 250 mm × 2.5mm 218MS52 analytical column from Grace Vydac (Hesperia, CA). A linear gradient was run from 2% methanol in 0.1% aqueous formic acid to 60% methanol over 15 min, at a flow rate of 200 μL/min and monitored at 254 nm. 25 µl of mixture from the small scale reaction was injected to determine retention times. *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione eluted at 11.8 min in this system. For the large scale reaction, 100 µl of reaction mixture was used for each injection to collect *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione for NMR characterization.

Liquid Chromatography-Mass Spectrometry (LC-MS). LC-MS analyses were performed on a triple quadrupole mass spectrometer TSQ-Quantum (Thermo Electron, Waltham, MA) operating in selected reaction monitoring (SRM) mode to detect and quantify *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione. A 150 mm × 2.5mm Hypersil Gold column (3 µm particle size) from Thermo Scientific (Pittsburgh, PA) was used. A linear gradient was run from 2% methanol in 0.1% aqueous formic acid to 60% methanol over 10 min, at 200 μL/min. The electrospray ionization (ESI) source was set as follows: spray voltage, 4.0 kV; capillary temperature, 300 °C; sheath gas pressure, 40 au; aux gas pressure, 10 au. *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione was detected at 7.10 min. The accurate mass and MS/MS were acquired on an Agilent 6500 Series Quadrupole Time-of-Flight (Q-TOF) LC/MS (Santa Clara, CA) with an ESI source. A linear gradient was run from 2% acetonitrile in 0.1% formic acid to 98% acetonitrile in 10 min at 200

μL/min. The ESI source was set as follows: gas temperature, 350 °C; drying gas, 10 L/min; Vcap, 4000 V; Nebulizer, 35 psig; fragmentor, 100 V; skimmer, 65 V. A 150 mm × 2.5mm Hypersil Gold column (3 μm particle size) was used and *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione was detected at 4.8 min. For fragmentation of *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione, the collision energy was set at 20 V.

Quantitation of S-[1-(N^2-deoxyguanosinyl)methyl]glutathione from DNA Samples. The adduct was quantified by a triple quadrupole mass spectrometer TSQ-Quantum (Thermo Electron, Waltham, MA) using SRM mode (587m/z→308 m/z). The collision energy was set at 20 V after optimization. The calibration curve for quantitation was obtained using the integrated peak area and amount of injected standard.

NMR. NMR spectra were recorded on a Varian INOVA 500 NMR spectrometer (Palo Alto, CA) at 500 MHz for ^{1}H and 125 MHz for ^{13}C spectra. HMBC data were acquired using the standard Varian program gHMBC, with a mixing time of 62.5 msec.

4.2.3 Experimental Methods

Glutathione (5 mM) was incubated with formaldehyde (2.5 mM) in 40 μl of 10 mM potassium phosphate buffer (pH=7.2) for 4 hours at 37 °C. Then, 10 μl of 10 mM deoxyguanosine was added and further incubated for 8 hours. The resultant reaction mixtures were either separated by reverse phase chromatography or analyzed by mass spectrometry. To prepare larger quantities of product for NMR characterization and for use as a standard, 92 mg of glutathione were treated with 100 mM formaldehyde in 3.5 ml of 100 mM potassium phosphate buffer (pH=7.2) for 4 hours at 37°C, followed by incubation with 8 mg of deoxyguanosine for 6 hours at 37°C. The resultant reaction

mixture was purified by HPLC using a C18 reverse phase column. HPLC fractions corresponding to *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione were collected and dried by lyphilization, followed by 1D and 2D NMR analysis. To measure adduct formation between DNA and GSH, 5 mM GSH solution was first treated with different concentrations of formaldehyde (0.1, 0.5, 1, 5, 50 mM) in 100 µl of 10 mM potassium phosphate buffer (pH=7.2) for 4 hours at 37°C, followed by incubation with 100 µg of calf thymus DNA for another 12 hours. The resultant modified DNA was washed with water 5 times using a Millipore Microcon YM-10 spin column. Then, DNA was treated by DNaseI (50U) for 30 min in the digestion buffer (80mM Tris-HCl 20mM $MgCl_2$ pH=7.2), followed by the addition of alkaline phosphatase (2U) and phosphodiesterases (1U) for additional 1 hour. Enzymes were removed by a Millipore Microcon YM-10 spin column and the resultant solution was separated by HPLC. The fraction containing *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione was dried with speed vacuum, followed by detection with mass spectrometry.

4.2.4 Stability test

S-[1-(N^2-deoxyguanosinyl)methyl]glutathione standard (100 µL, 4.14 nmol) was dissolved in 900 µL of water containing 0.1% formic acid (pH=4) or DNA digestion buffer (80mM Tris-HCl 20mM $MgCl_2$ pH=7.2). Solutions were stored at room temperature, 100 µL aliquots were analyzed at specified time points (0, 2, 4, 6, 8 and 16 hour) by HPLC. Stability was calculated from the integrated peak area for each injection, relative to time = 0 hour.

4.2.5 NMR data

^{1}H NMR (500 MHz, DMSO-d_6): δ = 8.51 (bt, 1H, NH-Gly), 8.45 (bs,1H, N^{2}*H*-dG), 8.30 (bd, 1H, *J* = 8.5 Hz, N*H*-Cys), 7.90 (s, 1H, H8-dG), 6.15 (ψt, 1H, *J* ~ 7 Hz, H1′-dG), 4.49-4.4 (m, 3H, Cys-H$^{\alpha}$, ^{formald}C*H*$_2$ overlapping), 4.35 (td,1H, *J* = 6.0, 3.2, 3.0 Hz, H3′-dG), 3.81 (m, 1H, *J* = 4.9, 4.8, 3.1 Hz, H4′-dG), 3.71 (d, 2H, *J* = 6.7 Hz, Gly-C*H*$_2$), 3.56 (dd, 1H, *J* = 11.6, 4.9 Hz, H5′ or H5″-dG), 3.49 (dd, 1H, *J* = 11.6, 4.8 Hz, H5″ or H5′-dG), 2.99 (dd,1H, *J* = 14.0, 5.0 Hz, Cys-H$^{\beta}$), 2.87 (dd,1H, *J* = 14.0, 7.4 Hz, Cys-H$^{\beta'}$), 2.63 (ddd, 1H, *J* = 13.3, 7.3, 6.0 Hz, H2′-dG), 2.36 (m, 2H, Glu-C*H*$_2^{\gamma}$), 2.22 (ddd, 1H, *J* = 13.3, 6.3, 3.2 Hz, H2″-dG), 1.95 (m, 2H, Glu-C*H*$_2^{\beta}$).

4.3 Results

To characterize potential cross-linked products, excess GSH (5 mM) was allowed to react with formaldehyde in 10 mM potassium phosphate buffer (pH=7.2) for 4 hours at 37°C, followed by incubation with deoxyguanosine (dG) (**4)** for another 8 hours. A single coupling product eluted at 11.9 min on a C18 reverse phase column, giving a UV spectrum similar to that of dG, with an absorbance maximum at 260 nm (shown in Figure 4.1). The exact mass of the protonated molecule was 587.1896 Da (Figure 4.2), consistent with elemental composition $C_{21}H_{30}N_8O_{10}S$ expected for *S*-[1-(*N*^{2}-deoxy-guanosinyl)methyl]glutathione (**5**).

The ESI-MS/MS of the protonated molecule (Figure 4.3) shows major product ions corresponding to loss of deoxyribosyl and deoxynucleoside fragments, in accord with structural assignment **5**. Definitive structural characterization was provided by 1D (Figure 4.4) and 2D NMR (Figure 4.5) analysis of product isolated from a larger-scale reaction, which also served as standard for subsequent quantitation. The formaldehyde-

derived methylene linkage between the β methylene carbon of the Cys residue and the exocyclic N^2 of dG is established by C-H connectivities in the HMBC spectrum (Figure 4.6), which shows the expected cross peaks between the diastereotopic Cys β-methylene protons and the formaldehyde-derived carbon of the methylene linker and between the protons of the methylene linker C2 of dG. Adduct **5** was stable in aqueous solution at room temperature over 16 hours at pH 4, approximately 40% loss was observed at pH 7.2 (see Figure 4.7), supporting the hypothesis that if formed, this adduct would be detectable in DNA.

In order to test this hypothesis, DNA was incubated with GSH in the presence of formaldehyde. GSH (5 mM) in 10 mM potassium phosphate buffer (pH=7.2) was treated with different concentrations of formaldehyde (0.1, 0.5, 1, 5, 50 mM) for 4 hours at 37°C, followed by incubation with 100 μg of calf thymus DNA for 12 hours. After extensive washing, DNA was digested with DNaseI (50U) for 30 min at 37°C in 80mM Tris-HCl 20mM $MgCl_2$ (pH=7.2), followed by addition of alkaline phosphatase (2U) and phosphodiesterases (1U) for an additional hour. The resultant adduct **5** was collected by HPLC and quantified by triple quadrupole mass spectrometer using selected reaction monitoring (SRM) mode (587m/z→308 m/z) and the previously-generated **5** from the large scale reaction as a standard (the calibration curve is shown in Figure 4.8). Figure 4.9 shows that production of **5** rises with increasing formaldehyde concentration from 0.1 mM to 5 mM, then declines when the formaldehyde concentration is further increased to 50 mM. It was shown previously that when formaldehyde is in excess in the reaction mixture, GSH traps two formaldehyde molecules to form a bicyclo[4.4.1]undecane structure (**6**)*(19;20)*, which is not able to react with DNA. Mass spectrometry analysis of

6 was consistent with the previously identified compound (Figure 4.10). This evidence further supports the involvement of **3** in the formation of **5**.

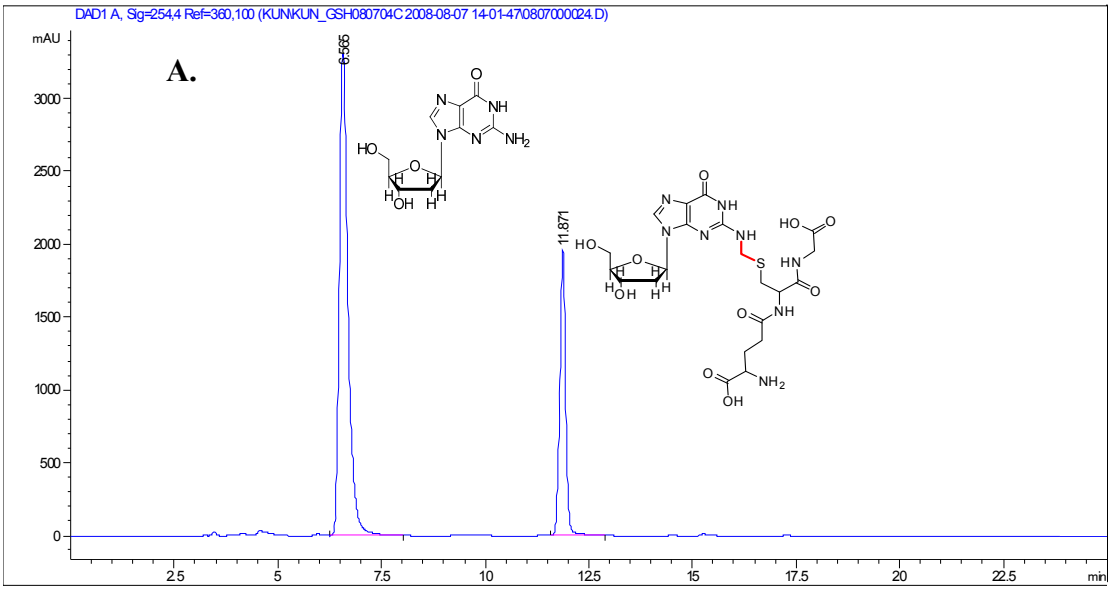
4.4 Conclusions

We have demonstrated that **5** is formed as a consequence of formaldehyde attack *in vitro*. Both formaldehyde and GSH are ubiquitous cellular components, thus, this DNA adduct is expected to form endogenously in cells. The involvement of the cysteine residue of GSH in coupling suggests that other thiols may participate in the formation of this type of DNA damage arising from formaldehyde.

The formation of **5** from exogenous formaldehyde may serve as a biomarker to evaluate formaldehyde exposure. Concentrations of formaldehyde in the blood of humans and of rats after formaldehyde exposure (1.9 ppm for 40 min and 14.4 ppm for 2 hours) were not different from the pre-exposure concentration*(2)*, consistent with extensive formation of **3**. It has also been shown that formaldehyde exposure depletes GSH levels in cells and tissues*(1)*, suggesting that GSH was not completely regenerated. Therefore, the scavenging of formaldehyde-induced **3** by ADH3 may be limited, which allows opportunity for reaction between **3** and DNA to form **5**. Also, adduct **5** is of potential importance for investigating effects of formaldehyde at distal sites. This issue remains one of the biggest challenges for understanding formaldehyde toxicity, carcinogenicity and epidemiology, which have been controversial for many years*(21-23)*. Formaldehyde exposure by inhalation resulted in decreases in cellular GSH concentration in the liver*(24)*, a remote site that inhaled formaldehyde is unlikely to reach by simple diffusion. Detection of **5** in distal tissues will shed light on the intriguing question of whether formaldehyde exhibits systemic toxicity.

In conclusion, we have demonstrated that formaldehyde can cross-link between GSH and DNA by forming *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione. This adduct may form endogenously since formaldehyde and GSH are ubiquitous in human cells. This adduct is unique because of the involvement of the reactive *S*-hydroxymethylglutathione intermediate that normally serves for formaldehyde detoxication. Since *S*-hydroxymethylglutathione is relatively abundant and highly reactive, and the adduct *S*-[1-(N^2-deoxyguanosinyl)-methyl]glutathione is reasonably stable, this adduct may serve as a biomarker to understand formaldehyde toxicity and to evaluate formaldehyde exposure if coupled with the application of isotope-labeled formaldehyde to differentiate between endogenous and exogenous formaldehyde-derived adducts.

Figures



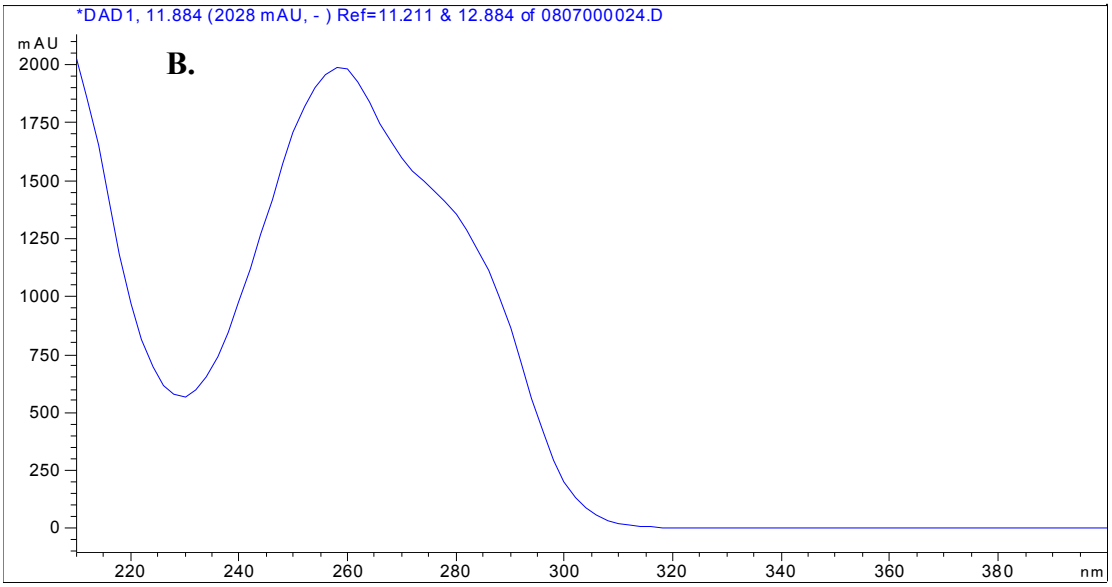


Figure 4.1. HPLC analysis of the products of reaction between glutathione, formaldehyde and deoxyguanosine. Panel A, chromatogram (254 nm) of the complete reaction mixture. Panel B, UV-Vis spectrum (diode array detector) of the peak at 11.9 minutes, identified as S-[1-(N2-deoxyguanosinyl)methyl]glutathione.

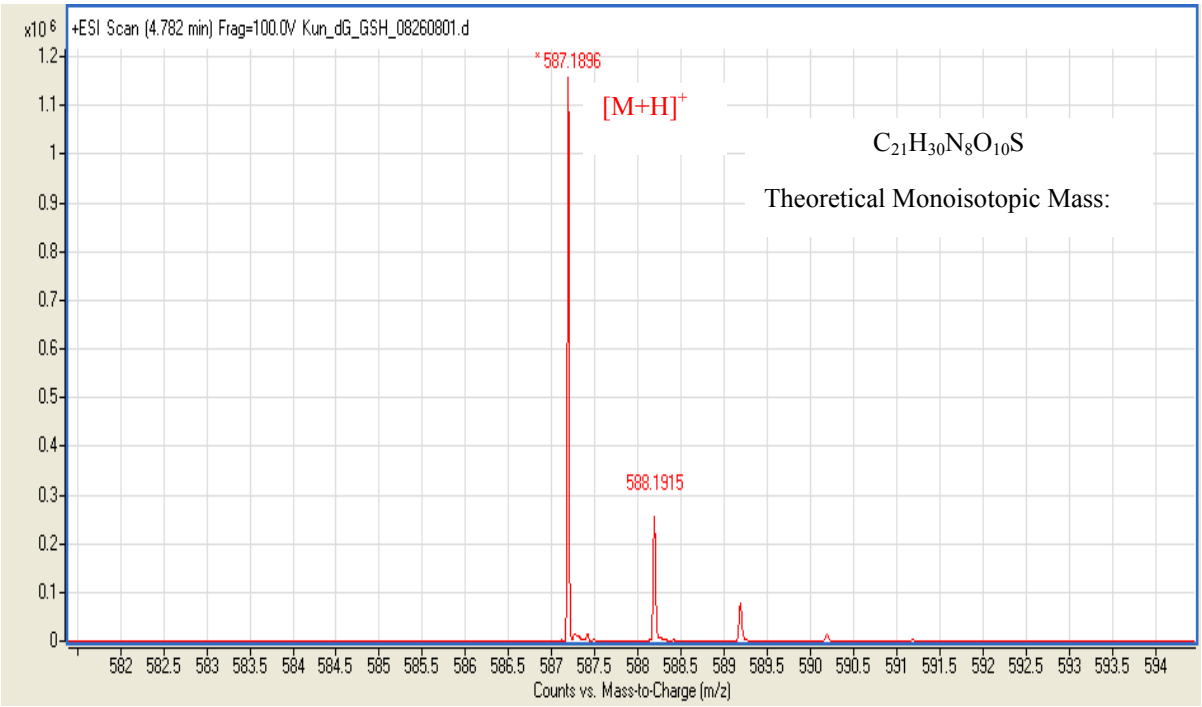


Figure 4.2. Exact mass of *S*-[1-(*N*^{2}-deoxyguanosinyl)methyl]glutathione

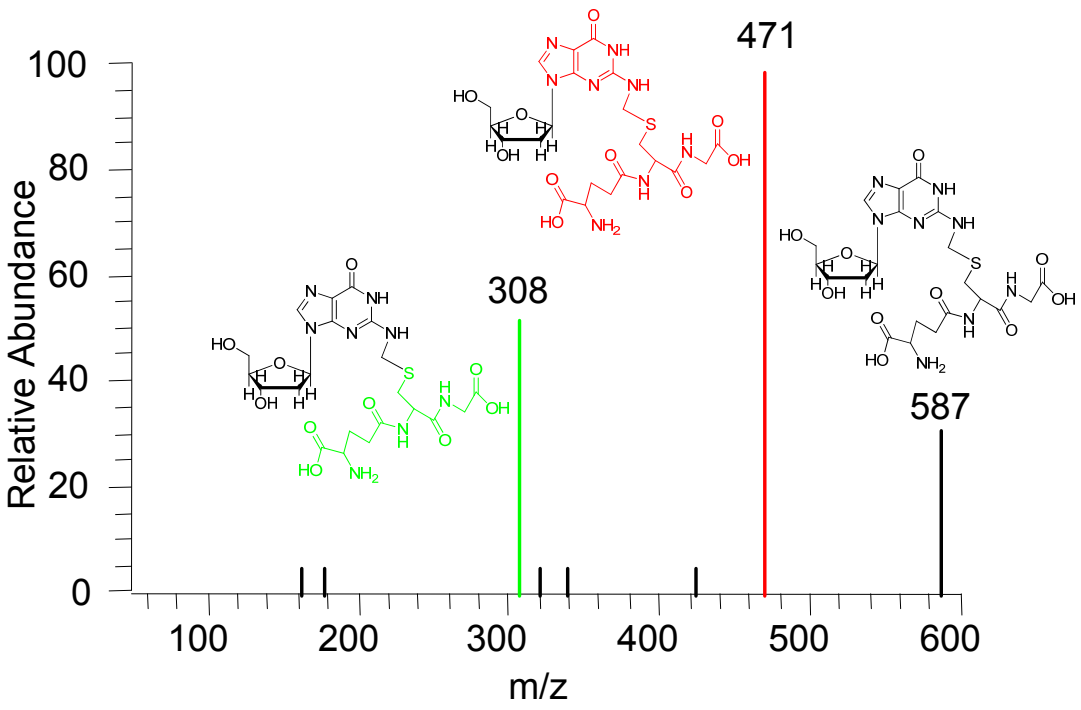


Figure 4.3. ESI-MS/MS of the protonated molecular ion of *S*-[1-(*N*^{2}-deoxyguanosinyl)methyl]glutathione.

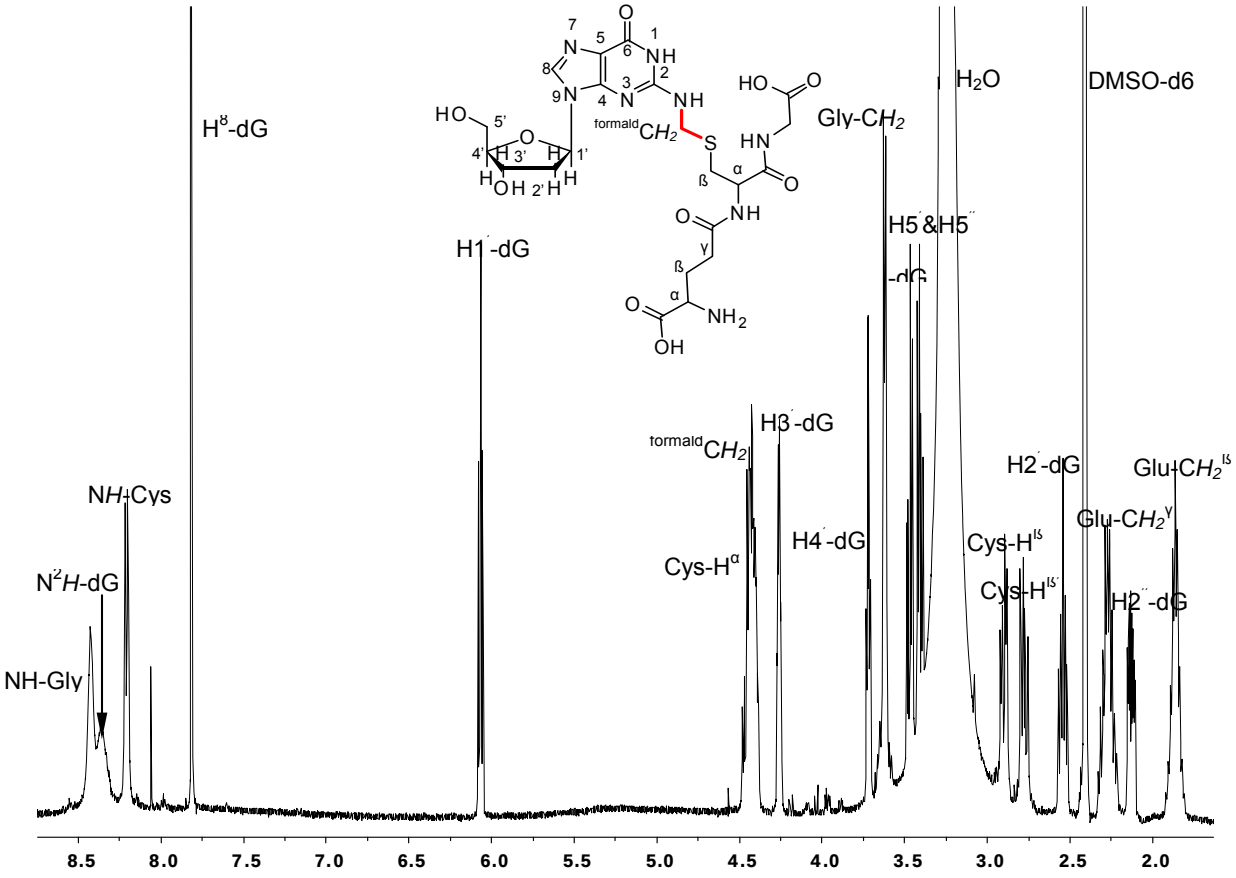


Figure 4.4. ^{1}H NMR spectrum [Varian INOVA 500 NMR spectrometer (Palo Alto, CA), at 500 MHz] of *S*-[1-(*N*^{2}-deoxyguanosinyl)methyl]glutathione

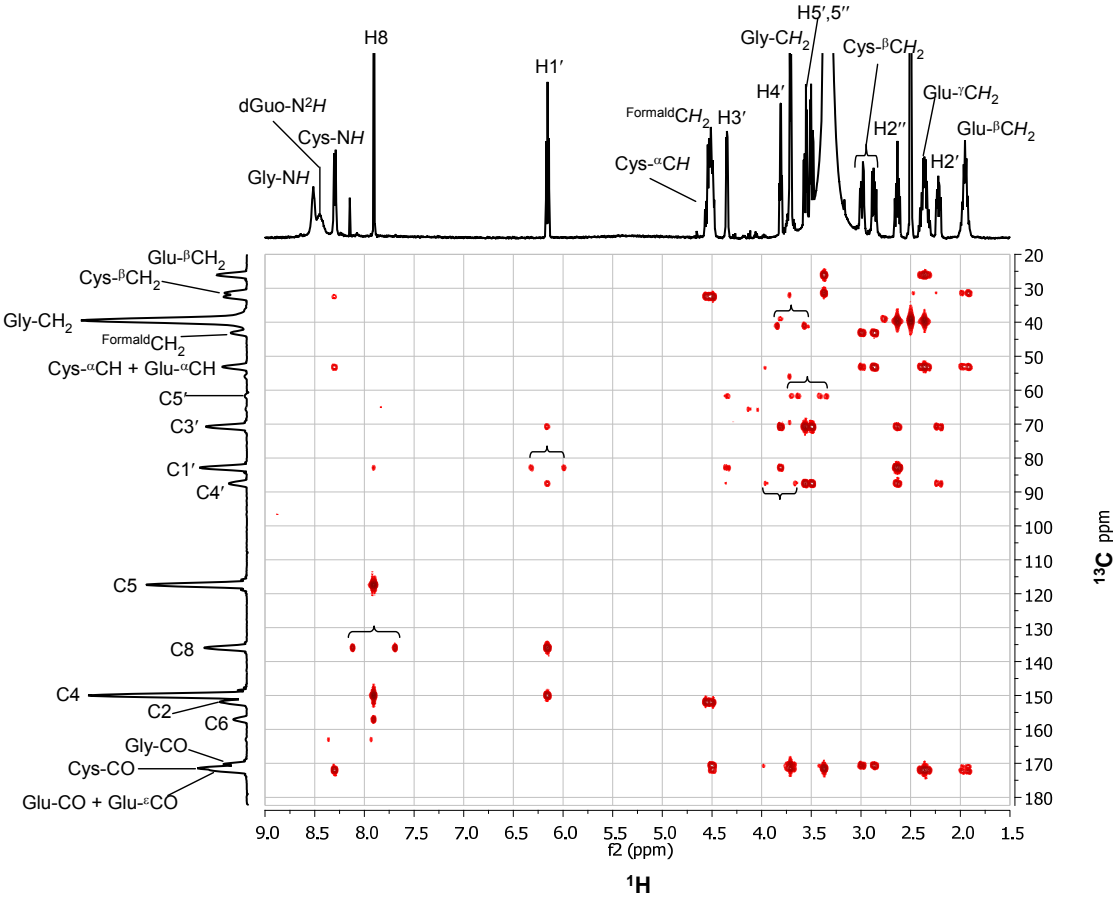


Figure 4.5. HMBC spectrum of *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione. Unsuppressed ^1J splittings are indicated on the spectrum by brackets

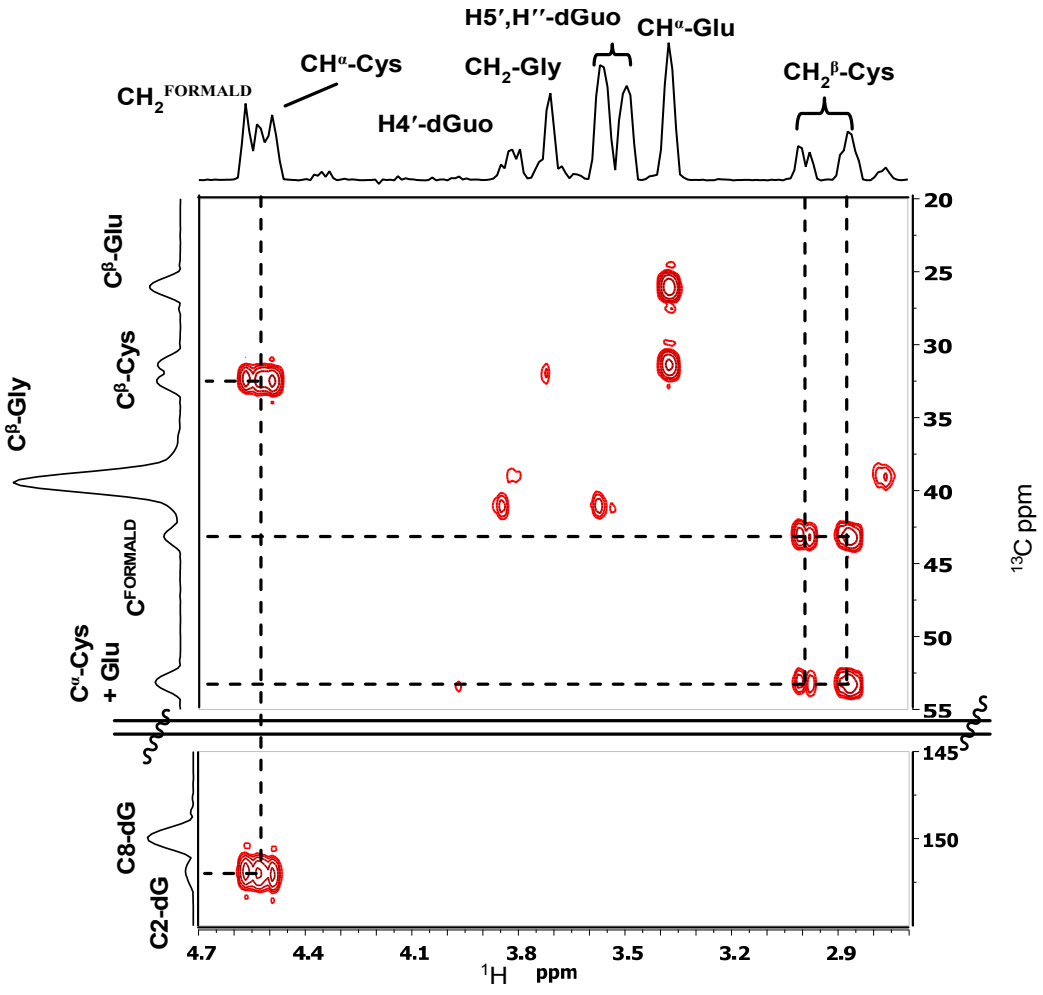


Figure 4.6. Expansion of the HMBC spectrum of 5 to show the Cys-β-methylene-formaldehyde linker-N^2-dG connectivity.

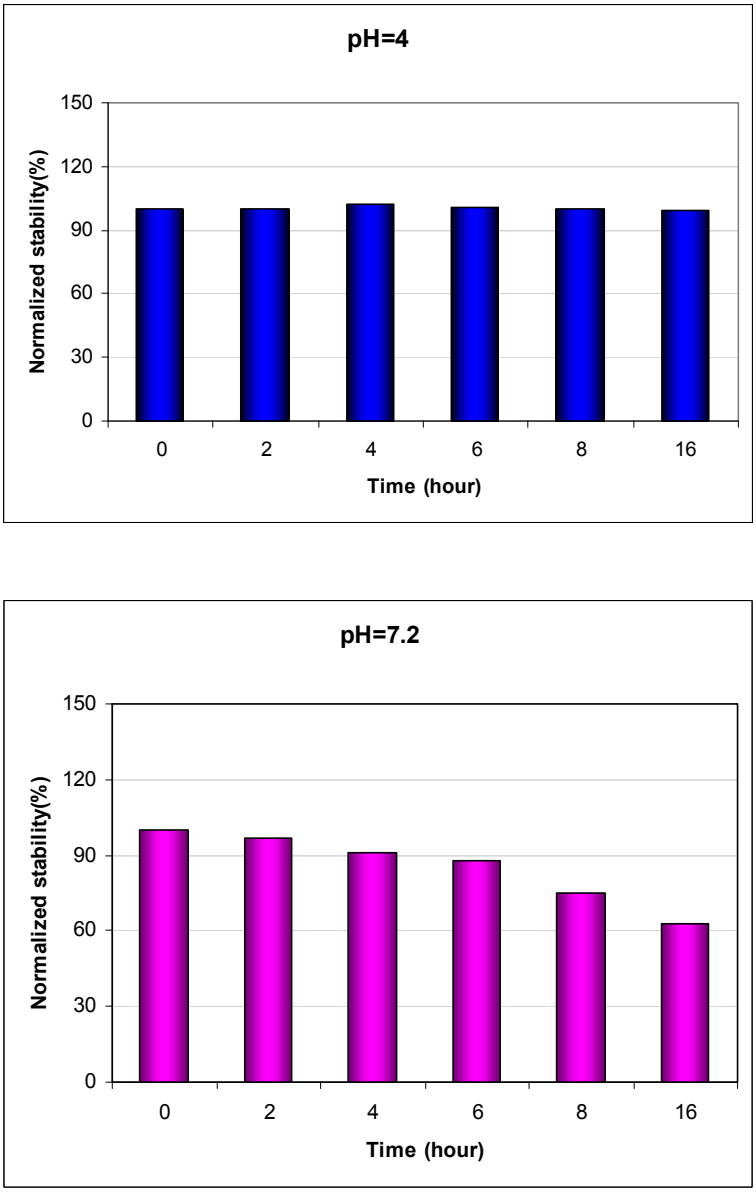


Figure 4.7. Stability of *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione in aqueous solutions (pH=4 and pH=7.2).

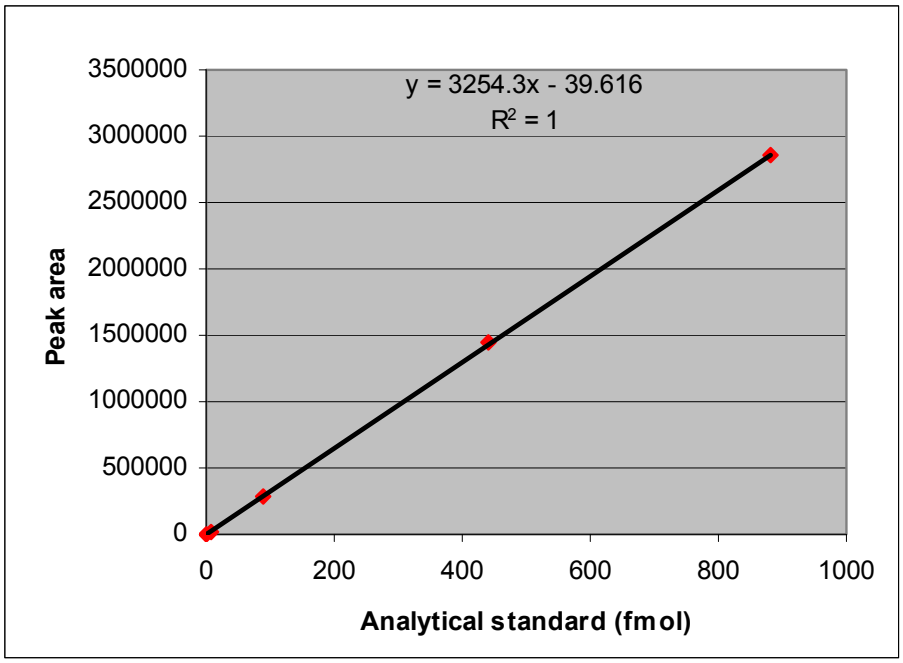


Figure 4.8. Calibration curve used for the quantitation of *S*-[1-(*N*^{2}-deoxyguanosinyl)methyl]glutathione from DNA. Adduct 5 isolated from large-scale reaction and quantitated by HPLC served as the standard.

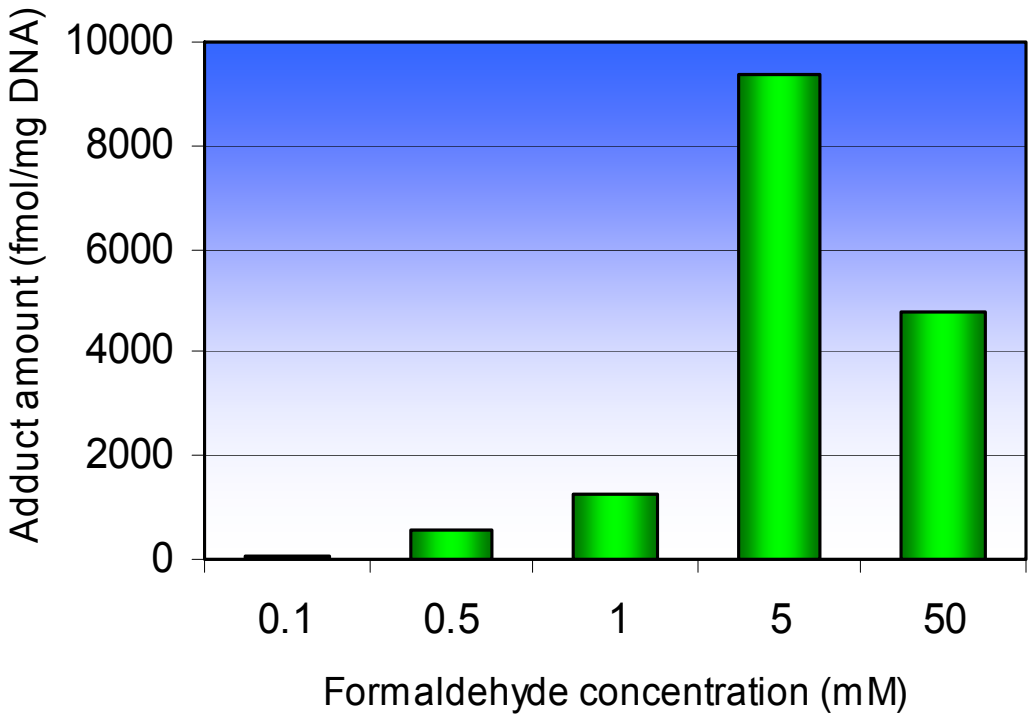


Figure 4.9. The influence of formaldehyde concentration on the formation of 5 in DNA.

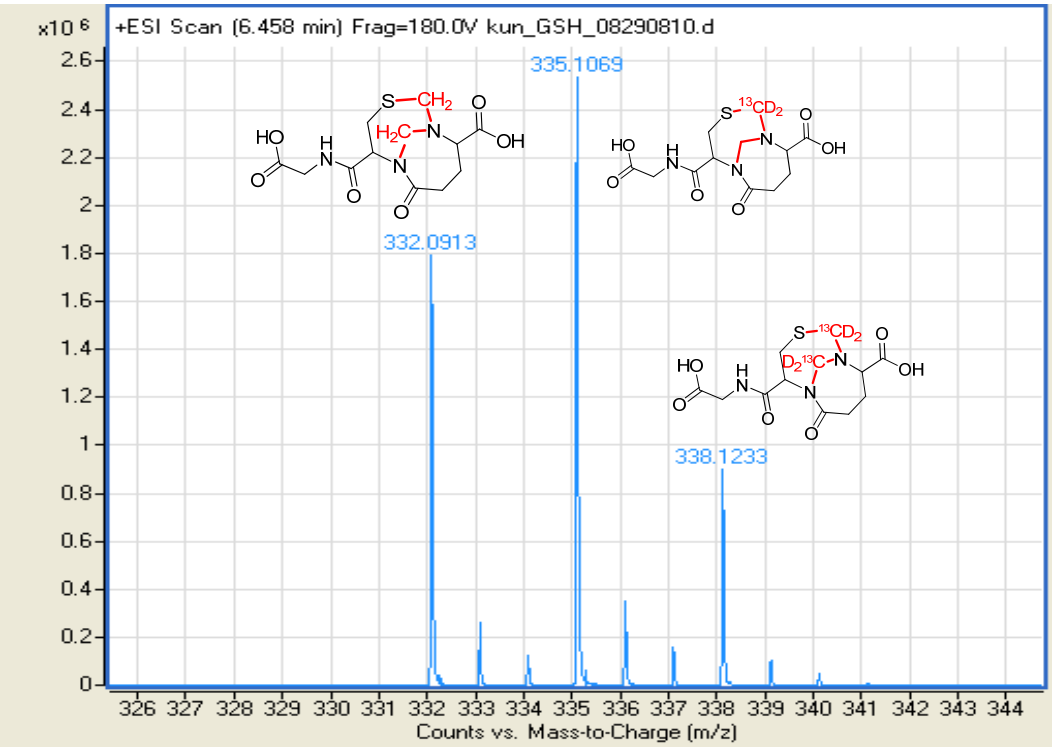
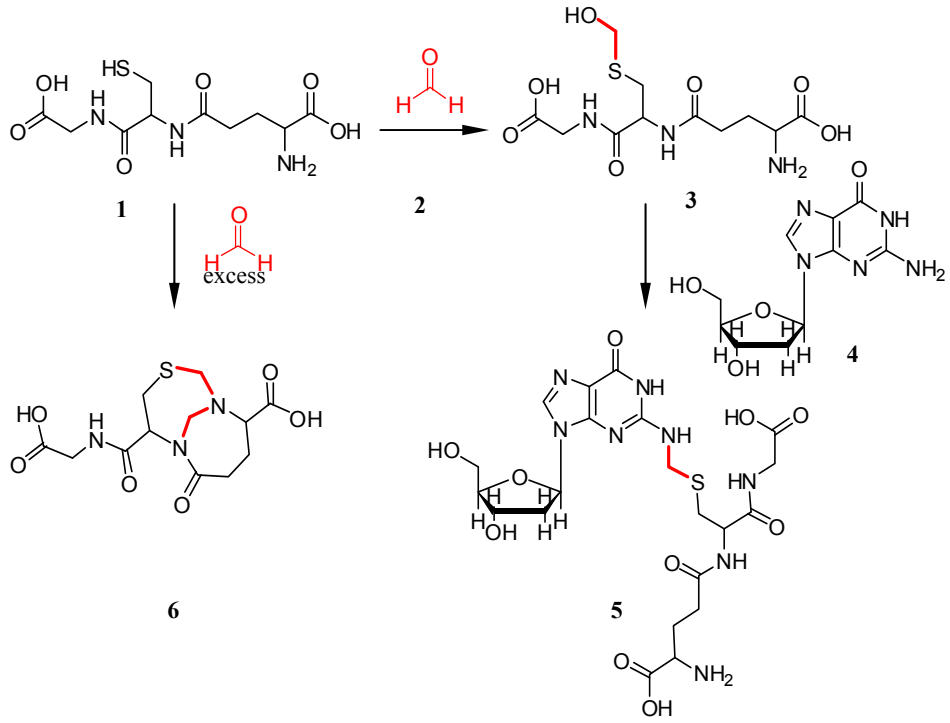


Figure 4.10. Exact mass of bicyclo[4.4.1]undecane induced by normal formaldehyde and [$^{13}CD_2$]-formaldehyde (50mM).

Schemes

Scheme 4.1. Formation of S-[1-(N2-deoxyguanosinyl)-methyl]glutathione induced by formaldehyde.



References

(1) IARC monographs on the evaluation of carcinogenic risks to humans. International Agency for Research on Cancer (2006) *IARC Monogr Eval. Carcinog. Risks Hum.* **88**, 1-287.

(2) Heck, H. D., Casanova-Schmitz, M., Dodd, P. B., Schachter, E. N., Witek, T. J., and Tosun, T. (1985) Formaldehyde (CH2O) concentrations in the blood of humans and Fischer-344 rats exposed to CH2O under controlled conditions. *Am. Ind. Hyg. Assoc. J.* **46**(1), 1-3.

(3) Casanova, M., Deyo, D. F., and Heck, H. D. (1989) Covalent binding of inhaled formaldehyde to DNA in the nasal mucosa of Fischer 344 rats: analysis of formaldehyde and DNA by high-performance liquid chromatography and provisional pharmacokinetic interpretation. *Fundam. Appl. Toxicol.* **12**(3), 397-417.

(4) Casanova-Schmitz, M., Starr, T. B., and Heck, H. D. (1984) Differentiation between metabolic incorporation and covalent binding in the labeling of macromolecules in the rat nasal mucosa and bone marrow by inhaled [14C]- and [3H]formaldehyde. *Toxicol. Appl. Pharmacol.* **76**(1), 26-44.

(5) Swenberg, J. A., Barrow, C. S., Boreiko, C. J., Heck, H. D., Levine, R. J., Morgan, K. T., and Starr, T. B. (1983) Non-linear biological responses to formaldehyde and their implications for carcinogenic risk assessment. *Carcinogenesis* **4**(8), 945-952.

(6) Casanova-Schmitz, M., and Heck, H. D. (1983) Effects of formaldehyde exposure on the extractability of DNA from proteins in the rat nasal mucosa. *Toxicol. Appl. Pharmacol.* **70**(1), 121-132.

(7) Merk, O., and Speit, G. (1998) Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ. Mol. Mutagen.* **32**(3), 260-268.

(8) Swenberg, J. A., Richardson, F. C., Boucheron, J. A., Deal, F. H., Belinsky, S. A., Charbonneau, M., and Short, B. G. (1987) High- to low-dose extrapolation: critical determinants involved in the dose response of carcinogenic substances. *Environ. Health Perspect.* **76**, 57-63.

(9) Cheng, G., Wang, M., Upadhyaya, P., Villalta, P. W., and Hecht, S. S. (2008) Formation of formaldehyde adducts in the reactions of DNA and deoxyribonucleosides with alpha-acetates of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and N-nitrosodimethylamine (NDMA). *Chem. Res. Toxicol.* **21**(3), 746-751.

(10) Wang, M., Cheng, G., Villalta, P. W., and Hecht, S. S. (2007) Development of liquid chromatography electrospray ionization tandem mass spectrometry methods for analysis of DNA adducts of formaldehyde and their application to rats treated with N-nitrosodimethylamine or 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Chem. Res. Toxicol.* **20**(8), 1141-1148.

(11) Lu, K., Boysen, G., Gao, L., Collins, L. B., and Swenberg, J. A. (2008) Formaldehyde-induced histone modifications in vitro. *Chem. Res. Toxicol.* **21**(8), 1586-1593.

(12) Ridpath, J. R., Nakamura, A., Tano, K., Luke, A. M., Sonoda, E., Arakawa, H., Buerstedde, J. M., Gillespie, D. A., Sale, J. E., Yamazoe, M., Bishop, D. K., Takata, M., Takeda, S., Watanabe, M., Swenberg, J. A., and Nakamura, J. (2007) Cells deficient in the FANC/BRCA pathway are hypersensitive to plasma levels of formaldehyde. *Cancer Res.* **67**(23), 11117-11122.

(13) Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J., and Swenberg, J. A. (1983) Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Res.* **43**(9), 4382-4392.

(14) Recio, L., Sisk, S., Pluta, L., Bermudez, E., Gross, E. A., Chen, Z., Morgan, K., and Walker, C. (1992) p53 mutations in formaldehyde-induced nasal squamous cell carcinomas in rats. *Cancer Res.* **52**(21), 6113-6116.

(15) McGhee, J. D., and von Hippel, P. H. (1975) Formaldehyde as a probe of DNA structure. II. Reaction with endocyclic imino groups of DNA bases. *Biochemistry* **14**(6), 1297-1303.

(16) McGhee, J. D., and von Hippel, P. H. (1975) Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry* **14**(6), 1281-1296.

(17) Zhong, W., and Hee, S. Q. (2004) Quantitation of normal and formaldehyde-modified deoxynucleosides by high-performance liquid chromatography/UV detection. *Biomed. Chromatogr.* **18**(7), 462-469.

(18) Zhong, W., and Que Hee, S. S. (2004) Formaldehyde-induced DNA adducts as biomarkers of in vitro human nasal epithelial cell exposure to formaldehyde. *Mutat. Res.* **563**(1), 13-24.

(19) Naylor, S., Mason, R. P., Sanders, J. K., Williams, D. H., and Moneti, G. (1988) Formaldehyde adducts of glutathione. Structure elucidation by two-dimensional n.m.r. spectroscopy and fast-atom-bombardment tandem mass spectrometry. *Biochem. J.* **249**(2), 573-579.

(20) Bateman, R., Rauh, D., and Shokat, K. M. (2007) Glutathione traps formaldehyde by formation of a bicyclo[4.4.1]undecane adduct. *Org. Biomol. Chem.* **5**(20), 3363-3367.

(21) Heck, H., and Casanova, M. (2004) The implausibility of leukemia induction by formaldehyde: a critical review of the biological evidence on distant-site toxicity. *Regul. Toxicol. Pharmacol.* **40**(2), 92-106.

(22) Zhang, L., Freeman, L. E., Nakamura, J., Hecht, S. S., Vandenberg, J. J., Smith, M. T., and Sonawane, B. R. (2009) Formaldehyde and leukemia: Epidemiology, potential mechanisms, and implications for risk assessment. *Environ. Mol. Mutagen.*

(23) Zhang, L., Steinmaus, C., Eastmond, D. A., Xin, X. K., and Smith, M. T. (2009) Formaldehyde exposure and leukemia: a new meta-analysis and potential mechanisms. *Mutat. Res.* **681**(2-3), 150-168.

(24) Söğ üt S.; Songur A.; Özen O. A.; Özyurt, H. ; Sarsı lmaz M. (2004) Dose the subacute (4 week) exposure to formaldehyde inhalation lead to oxidant/antioxidant imbalance in rat liver? *Eur. J. Gen. Med*., **1** (3), 26-32.

CHAPTER 5

DNA ADDUCTS CAUSED BY INHALED FORMALDEHYDE ARE FOUND IN NASAL EPITHELIUM, BUT NOT BONE MARROW

5.1 Introduction

Formaldehyde is a known human and animal carcinogen according to the International Agency for Research on Cancer (IARC) *(1)*, causing nasopharyngeal cancer in humans and squamous cell carcinomas in the nose of rats*(2)*. Limited evidence for an association between formaldehyde exposure and the induction of leukemia was published in several epidemiology studies on professional workers and industrial workers*(3;4)*, however, this finding was not consistent across studies. A large study of British industrial workers failed to find excess mortality among workers exposed to formaldehyde*(5)*. Overall, IARC considered that there was strong, but not sufficient evidence for a causal association between leukemia and occupational exposure to formaldehyde. A recent follow-up of the epidemiology study of industrial workers in USA also found an association between peak exposure to formaldehyde with the induction of leukemia, especially myeloid leukemia*(6)*. However, whether or not formaldehyde causes leukemia is still very debatable because there are no experimental data that provide any mechanistic insight related to how formaldehyde could induce leukemia, although several possible mechanisms have been proposed*(7)*. There is wide spread human exposure to formaldehyde that has raised public concern over its safety.

Formation of DNA adducts and increased cell proliferation secondary to cytotoxicity are well established key events that drive the induction of nasal cancer in rats by formaldehyde *(1)*. Formaldehyde readily induces DNA protein cross-links (DPC) in nasal epithelial DNA of rats*(8)*, the site of contact following inhalation exposure, however, DPC have not been demonstrated in distant tissues*(9)*. DPC measurements have utilized non-chemical-specific methods, primarily based on physical chemistry. However, no chemical-specific DNA biomarkers have been evaluated following inhalation exposure to formaldehyde, a primary route of exposure.

Formaldehyde can directly enter into the body through environmental exposure such as vehicle emissions, building materials, food, tobacco smoke and indoor air such as living in FEMA trailers (http://www.cdc.gov/nceh/ehhe/trailerstudy). Formaldehyde can also be indirectly generated through metabolism of a variety of compounds including nutrients, drugs and proteins by demethylation. Of equal importance, formaldehyde is an essential metabolic intermediate in all living cells and is endogenously produced from serine, glycine, methionine and choline. The endogenous concentration of formaldehyde in the blood of human subjects is about 0.1mM *(1)*. Therefore, the presence of both endogenous and exogenous formaldehyde makes developing chemically-specific biomarkers to monitor formaldehyde exposure and to understand its toxicity especially challenging.

Several previous *in vitro* studies demonstrated that N^6-hydroxymethyl-dA (N^6-$HOCH_2$-dA) was a primary formaldehyde-DNA monoadduct*(10;11)*. Recently, the Hecht laboratory reported increased amounts of N^6-$HOCH_2$-dA in multiple tissues of rats treated with *N*-nitrosodimethylamine (NDMA) or 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) *(12;13)*, which were attributed to the production of formaldehyde during

the metabolism of NDMA and NNK. Thus, N^6-$HOCH_2$-dA may serve as a good biomarker for formaldehyde exposure after inhalation. In preliminary cell culture experiments, we found that N^2-hydroxymethy-dG was also a suitable DNA biomarker for formaldehyde exposure. Therefore, in the present study, we measured both N^2-hydroxymethy-dG and N^6-hydroxymethyl-dA in tissues of rats exposed by inhalation to 10 ppm [$^{13}CD_2$]-formaldehyde for 6 hr or for 5 days (6h/day). The use of [$^{13}CD_2$]-formaldehyde permitted the simultaneous measurement of both endogenous and exogenous formaldehyde DNA adducts in any tissue. We also developed sensitive LC-ESI-MS/MS-SRM methods to detect and quantify formaldehyde-induced DNA adducts and DNA-DNA cross-links.

Our results show that N^2-hydroxymethyl-dG is a suitable DNA biomarker for inhalation exposure to formaldehyde. Surprisingly, there were no detectable exogenous formaldehyde induced N^6-HO$^{13}CD_2$-dA in 1 day or 5 day-exposed rat nasal DNA. In addition, we demonstrated that exogenous [$^{13}CD_2$]-formaldehyde-DNA adducts and DNA-DNA cross-links occurred only in nasal DNA, the site of contact for inhaled formaldehyde, whereas endogenous dA and dG monoadducts were present in DNA of all tissues examined. This suggests that distant site genotoxic effects of inhaled formaldehyde are highly implausible.

5.2 Experiment and methods.

5.2.1 Chemicals and Materials.

Deoxyguanosine, deoxyadenine, potassium phosphate, Tris-HCl, $MgCl_2$, formic acid, $NaCNBH_3$, methanol, acetonitrile, HPLC grade water and 10X PBS were all

purchased from Sigma (St. Louis, MO). 20% formaldehyde in water was procured from Tousimis (Rockville, MD). DNase I, alkaline phosphatase and phosphodiesterases were purchased from Sigma (St. Louis, MO). [$^{15}N_5$]-dA and [$^{13}C_{10}{}^{15}N_5$]-dG were ordered from Cambridge Isotope Lab (Cambridge, MA). N^6-CH_3-dA and N^2-CH_3-dG were obtained from Sigma (St. Louis, MO) and Berry & Associates (Dexter, MI), respectively. All chemicals were used as received unless otherwise stated.

5.2.2 High Performance Liquid Chromatography (HPLC).

The purification of formaldehyde-DNA adducts was carried out on an Agilent 1200 series HPLC system equipped with a diode-array detector (Santa Clara, CA). Analytes in 40-200 μg DNA hydrolysate were separated by reverse phase chromatography using a 150 mm × 2.5mm T3 analytical column from Waters (Milford, MA). The mobile phases were 0.1% formic acid (A) and methanol (B). A linear gradient was run from 2% methanol to 30% methanol over 30 min, at a flow rate of 200 μL/min and monitored at 254 nm. N^6-Me-dA and N^2-Me-dG eluted at 22.2 and 25.2 min on the column in this system, respectively. dG-dG cross-links eluted at 34.3 min with a linear gradient from 2% methanol to 60% methanol over 60 min, at a flow rate of 200 μL/min.

5.2.3 Liquid Chromatography-Mass Spectrometry (LC-MS).

LC-MS analyses were performed on a triple quadrupole mass spectrometer Ultra TSQ-Quantum (Thermo Electron, Waltham, MA) operating in selected reaction monitoring (SRM) mode to detect and quantify formaldehyde-DNA adducts. The mass spectrometer was interfaced with a Ultra Performance Liquid Chromatography (UPLC) system from Waters (Milford, MA). A 150 mm × 1.0 mm T3 column (3 μm particle

size) from Waters was used. A linear gradient was run from 2% methanol in 0.1% aqueous formic acid to 60% methanol over 10 min, at 40 μL/min. For the analysis of monoadducts, the electrospray ionization (ESI) source was set as follows: spray voltage, 4.0 kV; capillary temperature, 300 °C; sheath gas pressure, 40 au; aux gas pressure, 10 au. For the analysis of dG-CH_2-dG, the mass spectrometer was coupled with a Waters nano-Acquity LC system. A 100 mm × 100 μm nanoAcquity UPLC HSS T3 column was used. A linear gradient was run from 2% acetonitrile in 0.1% acetic acid to 50% in 10 min, at a flow rate at 0.6 or 1 μL/min. The ESI source was set as follows: spray voltage, 2.2-2.5 kV; capillary temperature, 280 °C.

5.2.4 Preparation of Internal Standards.

10 mM [$^{13}C_{10}{}^{15}N_5$]-dG and [$^{15}N_5$]-dA solution was treated by 100 mM formaldehyde in phosphate buffer (pH=7.2) overnight at 37 °C. The reaction mixture was separated by HPLC using a 150 mm × 2.5mm C18 T3 analytical column. N^2-hydroxymethyl-dG and N^6-hydroxymethyl-dA eluted at 20.5 and 24.3 min, respectively. N^2-hydroxymethyl-dG and N^6-hydroxymethyl-dA were collected and incubated with 50 mM $NaCNBH_3$ (pH=7.1) overnight at 37°C, followed by further separation using HPLC. [$^{13}C_{10}{}^{15}N_5$]-N^2-CH_3-dG and [$^{15}N_5$]-N^6-CH_3-dA eluted at 27.2 and 33.5 min on a 150 mm × 2.5mm T3 column using 10 mM ammonium acetate in 0.1% acetic acid and methanol as mobile phases, separately. The concentration of [$^{13}C_{10}{}^{15}N_5$]-N^2-CH_3-dG and [$^{15}N_5$]-N^6-CH_3-dA was determined by HPLC with unlabelled N^6-CH_3-dA and N^2-CH_3-dG as references. The usual conversion rate from hydroxymethyl group to methyl group was 65% to 85%. The internal standard of dG-dG cross-links was prepared by incubating 10 mM [$^{13}C_{10}{}^{15}N_5$]-dG with 100 mM formaldehyde in phosphate buffer (pH=7.2) for 96 h at

37 °C. dG-dG cross-links eluted at 34.3 min using a linear gradient from 2% methanol in 0.1% formic acid to 60% methanol over 60 min, at a flow rate of 200 µL/min.

5.2.5 Animal Exposures.

Animal use in this study was approved by the Institutional Animal Use and Care Committee of The Hamner Institutes for Health Sciences and was conducted in accordance with the National Institutes of Health guidelines for the care and use of laboratory animals. Animals were housed in fully accredited American Association for Accreditation of Laboratory Animal Care facilities. F344 rats were nose-only exposed to 10 ppm formaldehyde atmospheres generated by the thermal depolymerization of solid [$^{13}CD_2$]-paraformaldehyde for 1 or 5 days (6h/day), sacrificed within two hours of exposure for tissue collection and immediate freezing on dry ice, followed by storage at -80°C. The atmospheric concentration of formaldehyde was continuously monitored during exposure , using a calibrated infrared analyzer.

5.2.6 DNA Isolation and Digestion.

DNA was isolated from the tissues of rats using a NucleoBond DNA Isolation Kit (Bethlehem, PA), as instructed by the manufacturer with small modifications. The resultant DNA was quantified and stored at -80°C for further analysis. DNA isolated from each nose sample (approximately 30~50 µg) was incubated with 50 mM $NaCNBH_3$ at 37°C for 6 hour in phosphate buffer (pH=7.1). Then, DNA was treated by DNaseI (200U) for 10 min in the digestion buffer (80mM Tris-HCl 20mM $MgCl_2$ pH=7.2), followed by the addition of 25 µl of alkaline phosphatase and 25 µl of phosphodiesterases for additional 1 hour. Enzymes and undigested DNA were removed by a Millipore

Microcon YM-10 spin column for 45 min at 23 °C and the resultant solution was separated by HPLC to collect the fractions containing the corresponding DNA adducts. For monoadduct analysis of other tissues, 200 μg of DNA was usually used for digestion using the same methods. Typically, 10 μl of hydrolysate was used to quantify dG and dA generated from DNA enzymatic digestion.

5.2.7 Sample workup for DNA-DNA cross-links.

For the analysis of dG-dG cross-links in nose samples, 30-50 μg isolated nasal DNA was digested enzymically as described above in Tris buffer for 70 min at 37°C, followed by a cleaning procedure with a Millipore Microcon YM-10 spin column for 45 min at 23 °C. $NaCNBH_3$ was not used in DNA isolation for dG-dG cross-link analysis, as it labilized the cross-links. The hydrolysate, after removing 10 μl for dG quantitation, was frozen at -20°C and separated by HPLC after 12 h storage. A parallel control experiment was carried out, to determine the extent of artifact. We spiked an equal amount of isotope labeled [$^{13}C_{10}{}^{15}N_5$]-dG as was produced from DNA digestion (determined by HPLC with UV detector using a small aliquot of DNA solution) into the sample before hydrolysis. Thus, the signal corresponding to cross-links dG-CH_2-[$^{13}C_{10}{}^{15}N_5$]-dG accounted for ~half of the artifacts formed during sample workup and storage, which demonstrated the amount of artifactual dG-CH_2-dG formed during sample workup and storage. The extent of artifact was determined using the integrated peak area of dG-CH_2-[$^{13}C_{10}{}^{15}N_5$]-dG divided by the peak area of dG-CH_2-dG. For the analysis of dG-dG cross-links in other tissues, 200 μg DNA was digested for 70 min at 37°C, followed by a cleaning procedure with a Millipore Microcon YM-10 spin column for 60

min at 4 °C. The resultant solution was frozen on dry ice immediately, followed by the storage in -80°C.

5.2.8 Quantitation of formaldehyde-DNA adducts.

The adducts were quantified by a triple quadrupole mass spectrometer Ultra TSQ-Quantum (Thermo Electron, Waltham, MA) using SRM mode. N^2-$HOCH_2$-dG was quantified as N^2-CH_3-dG after reduction with $NaCNBH_3$ using the transition of m/z 282.2→m/z166.1 and N^2-$HO^{13}CD_2$-dG was quantified as N^2-$^{13}CD_2H$-dG with the transition of m/z 285.2→m/z 169.1. N^6-$HOCH_2$-dA was detected as N^6-CH_3-dA after treatment by $NaCNBH_3$ with the transition of m/z 266.2 to m/z 150.1. The collision energy was set at 17 V after optimization. dG-CH_2-dG and dG-$^{13}CD_2$-dG was detected and quantified using the transition of m/z 547.5 to m/z 152.1 and m/z 550.5 to m/z 152.1, with the collision energy set at 23V. The calibration curves for quantitation were obtained using the integrated peak area and amount of injected analytical standard and internal standard.

5.2.9 Statistical analysis of data.

Data represent mean ± standard deviation (SD). Unpaired Student's *t* tests were performed using the SAS statistical package (SAS Institute, Cary, NC) with the sample size ranging from 5 to 8 for monoadduct analysis. Differences were considered statistically significant if $p < 0.05$.

5.3 Results

5.3.1 Method development for formaldehyde-induced monoadducts.

A highly sensitive method for the analysis of N^2-hydroxymethyl-dG was developed. Briefly, DNA was incubated with 50 mM $NaCNBH_3$ to convert moderately stable N^2-hydroxymethyl-dG to stable N^2-methyl-dG. The resultant DNA was digested by multiple enzymes, followed by fraction collection using HPLC. The fractions containing N^2-Me-dG were dried and analyzed by capillary LC-ESI-MS/MS in SRM mode. As shown in Scheme 5.1, both endogenous and exogenous formaldehyde can contribute to the formation of formaldehyde-DNA adducts. Therefore, the mass spectrometer was set up to monitor different transitions: m/z 282.2→166.1 for N^2-CH_3-dG; m/z 285.2→m/z 169.1 for N^2-$^{13}CD_2H$-dG and m/z 297.2→m/z 176.1 for the internal standard [$^{13}C_{10}{}^{15}N_5$]-N^2-CH_3-dG. The method for N^6-$HOCH_2$-dA was essentially the same as that for N^2-$HOCH_2$-dG with appropriate transitions in SRM.

Figure 5.1 A shows a typical LC-ESI-MS/MS-SRM chromatogram of N^2-CH_3-dG and its corresponding internal standard. 0.8 fmol of N^2-CH_3-dG and 80 fmol of [$^{13}C_{10}{}^{15}N_5$]-N^2-CH_3-dG internal standard were loaded on the column, with a retention time at 7.58 min. The limit of detection was ~240 amol on the column (S/N=3). Figure 5.1 B shows a typical LC-ESI-MS/MS-SRM chromatogram of 0.15 fmol N^6-CH_3-dA and 37.5 fmol internal standard [$^{15}N_5$]-N^6-CH_3-dA with a retention time of 8.55 min. The limit of detection was ~75 amol on the column (S/N=3). The calibration curves used to quantify formaldehyde-dG and dA adducts are presented in Figure 5.1 C and Figure 5.1 D, respectively. Satisfactory accuracy and precision of this assay were shown by adding known amounts of N^2-CH_3-dG and N^6-CH_3-dA analytical standards to 50 μg of DNA, as shown in Table 5.1.

5.3.2 N^2-hydroxymethyl-dG in tissues of exposed rats.

Figure 5.2 A shows the LC-ESI-MS/MS SRM chromatogram of N^2-Me-dG in nasal DNA from a 1 day-exposed rat using the methods just described. The peak corresponding to the specific transition of m/z 282.2→ m/z166.1 and the same retention time with [$^{13}C_{10}{}^{15}N_5$]-N^2-CH_3-dG internal standard unambiguously identified the formation of N^2-$HOCH_2$-dG from endogenous formaldehyde, as shown by the peak at 7.55 min in the top panel of Figure 5.2A. In addition to the peak of endogenous adducts, a new peak corresponding to the transition of m/z 285.2→m/z 169.1 coeluted with the internal standard, which is consistent with N^2-$HO^{13}CD_2$-dG formed from exogenous [$^{13}CD_2$]-formaldehyde. Similarly, both endogenous and exogenous dG adducts could be clearly identified and quantified in 30-50 μg of nasal epithelial DNA after 5 days of exposure, as shown in Figure 5.2 B. Compared with Figure 5.2A, the amount of exogenous N^2-$HO^{13}CD_2$-dG in nasal DNA increased as a consequence of extended exposure. In all other tissues including liver, lung, thymus, bone marrow and spleen of 1 or 5 day-exposed rats, only endogenous formaldehyde-induced N^2-$HOCH_2$-dG could be observed in 200 μg DNA. As shown by the middle panels of Figure 5.2C and Figure 5.2D, no signal corresponded to the transition of m/z 285.2→m/z 169.1 of exogenous adducts. Therefore, there were no detectable amounts of N^2-$HO^{13}CD_2$-dG in distant site tissues after 1 or 5 days of exposure, despite the utilization of 5-fold more DNA for these tissues. These data strongly indicate that formaldehyde does not reach distant sites in an active form to cause detectable amounts of DNA adducts.

5.3.3 N^6-hydroxymethyl-dA in tissues of exposed rats.

N^6-hydroxymethyl-dA has been reported to be induced by formaldehyde resulting from the metabolism of nitroso compounds*(12;13)*. Therefore, we examined the formation of N^6-hydroxymethyl-dA in rats exposed to 10 ppm [$^{13}CD_2$]-formaldehyde, as shown in Figure 5.3. Figure 5.3 A and B illustrate the LC-ESI-MS/MS-SRM chromatograms of N^6-Me-dA in noses of rats exposed to [$^{13}CD_2$]-formaldehyde for 1 or 5 days. The peak corresponding to the specific transition of m/z 266.2→m/z 150.1 and the same retention time as the internal standard identified the formation of N^2-$HOCH_2$-dA from endogenous formaldehyde. However, there was no signal for N^6-$HO^{13}CD_2$-dA (m/z 269.2→m/z 153.1). This result indicates that N^2-hydroxymethyl-dA did not form in detectable amounts in the noses of rats exposed to 10 ppm formaldehyde for either 1 or 5 days. Figure 5.3C and 5.3D show the chromatograms of N^6-Me-dA in bone marrow and spleen of 5 day-exposed rats. We did not detect a peak for N^6-$HO^{13}CD_2$-dA after [$^{13}CD_2$]-formaldehyde exposure, while we clearly observed the peak of N^6-$HOCH_2$-dA from endogenous formaldehyde.

5.3.4 Monoadduct amounts in tissues of exposed rats.

Table 5.2 summarizes the number of adducts in all the tissues measured in this study. For 1 day-exposed samples, endogenous N^2-$HOCH_2$-dG ranged from 1.05-2.66 adducts/10^7 dG depending on tissues, while endogenous N^6-$HOCH_2$-dA was from 1.85 to 3.95 adducts/10^7 dA. For 5 day-exposed rats, the numbers of endogenous N^2-$HOCH_2$-dG and N^6-$HOCH_2$-dA were 1.17-3.24 adducts/10^7 dG and 2.23-3.61 adducts/10^7 dA, respectively. The exogenous [$^{13}CD_2$]-formaldehyde only induced DNA adducts with dG and this adduct was only detected in nasal epithelium from 1 or 5 day-exposed rats. The

mean amounts of exogenous N^2-HO$^{13}CD_2$-dG in 1 or 5 day-exposed nasal epithelium samples were 1.28±0.49 and 2.43±0.78 adducts/10^7 dG, respectively.

Figure 5.4 illustrates statistical analyses of the data from 1 day-exposed and 5 day-exposed nose samples. Figure 5.4 A shows that the number of exogenous N^2-hydroxymethyl-dG adducts was significantly greater for 5 day-exposed noses ($p < 0.05$), with almost double the amount in 5 day samples, suggesting that some N^2-hydroxymethyl-dG accumulated with the longer exposure. The ratio of endogenous versus exogenous N^2-hydroxymethyl-dG is shown in Figure 5.4 B. There was a significant difference between 1 day and 5 day-exposed samples ($p < 0.05$). The former had a ratio of 0.57±0.28, while the ratio of 5 day-exposed sample was 1.06±0.40. These results suggest that the exogenous formaldehyde adducts increased during exposure.

5.3.5 DNA-DNA cross-links in tissues of exposed rats.

A sensitive nano-LC-MS/MS-SRM method was also developed, with a limit of detection of ~60 amol on column, to detect dG-dG cross-links in the tissue samples (Scheme 5.1). Some typical chromatograms are shown in Figure 5.5. Similar to the monoadducts, the exogenous dG-$^{13}CD_2$-dG only could be detected in nasal DNA (Figure 5.5A and 5.5B) and not in DNA from any distant tissue (Figure 5.5C and 5.5D). These data demonstrate that no detectable amounts of exogenous dG-dG cross-links were induced at distant sites in rats by inhalation exposure to formaldehyde.

The numbers of dG-dG cross-links in individual tissues are summarized in Table 5.3. The amount of endogenous cross-links ranged from 0.09–0.21 adducts/10^7 dG across all tissues. In contrast, the amounts of exogenous dG-$^{13}CD_2$-dG were 0.14±0.06 and

0.26±0.07 adducts/10^7 dG in DNA from nasal epithelium of 1 day and 5 day-exposed rats, respectively, but were not detectable in other tissues. These numbers were roughly 10% of the amount of corresponding monoadducts. Since N^2-hydroxymethyl-dG can further react with dG, potential artifacts could form during sample workup and storage. Therefore, we carried out a parallel control experiment adding amounts of [$^{13}C_{10}{}^{15}N_5$]-dG equal to the amount of dG in the sample and determined that up to 65% of the cross-links were artifact by calculating the peak ratio under the conditions used to analyze nose samples (Figure 5.6). The data in Table 5.3 were corrected to remove such artifact.

5.4 Discussion

This study addressed several critical issues related to formaldehyde toxicity and carcinogenicity. Exogenous N^2-HO$^{13}CD_2$-dG was only detected in DNA from nasal epithelium of rats after inhalation. The same was true for exogenous dG-dG cross-links. The absence of exogenous formaldehyde-induced DNA adducts and cross-links in other tissues supports the conclusion that distant genotoxic effects of inhaled formaldehyde are highly implausible. Endogenous formaldehyde-induced N^6-HOCH$_2$-dA was detected in all tissues, but N^6-HO$^{13}CD_2$-dA was not detectable in DNA from any tissue examined in rats following inhalation exposure. It is important to note that previous methods used to measure DPC *(8;9)* and specific DNA adducts *(12;13)*could not differentiate between endogenous and exogenous DNA damage. It was only through the use of the [$^{13}CD_2$]-formaldehyde and mass spectrometry that endogenous and exogenous adducts could specifically be measured in the same tissue.

Formaldehyde was first identified as a carcinogen when rats exposed to 15 ppm for 12-18 months developed squamous cell carcinomas of the nasal passages*(2)*. This

study was continued through 24 months of exposure, with nasal carcinomas developing in 50% of the 15 ppm rats, but only 1% of the rats exposed to 6 ppm formaldehyde. No malignant tumors of the nose occurred in rats exposed to 2 ppm, or in controls*(14)* . A second carcinogenicity study was conducted that expanded the exposure response to include 0.7, 2, 6, 10 and 15 ppm formaldehyde. Formaldehyde induced nasal squamous cell carcinomas in a highly nonlinear fashion, with no neoplasms at the lowest two concentrations, and 1, 22 and 47% carcinomas at 6, 10 and 15 ppm formaldehyde. Cell proliferation was also greatly increased at 10 and 15 ppm exposures*(15)*. Exposures for the present study were modeled after the 10 ppm exposure data from this expanded cancer bioassay because it was clearly carcinogenic, induced marked increases in sustained cell proliferation, yet was less cytotoxic than15 ppm formaldehyde.

IARC recently classified formaldehyde as a known human carcinogen based on sufficient evidence for nasopharyngeal carcinoma*(1)*. Extensive research has demonstrated that both genotoxicty and cytoxicity contribute to the carcinogenesis of formaldehyde in nasal tissues*(1)*. As mentioned above, marked and sustained increases in cell proliferation in the noses of rats have been observed after exposure to ≥10 ppm formaldehyde *(15;16)*. In this study, we found that exogenous formaldehyde-induced N^2-hydroxymethyl-dG adducts and corresponding dG-dG cross-links, a generally more mutagenic species, formed in noses of rats after 10 ppm [$^{13}CD_2$]-formaldehyde exposure. These findings provide new mechanistic insight regarding how exposure to formaldehyde may cause nasal cancer. Several guanine point mutations have been identified by DNA sequence analysis of *p53* cDNA from formaldehyde-induced squamous cell carcinomas in nasal passages of rats, including $_{398}$G→T, $_{638}$G→T, $_{812}$G→A and $_{842}$G→C*(17)*. Earlier studies of Heck and Casanova demonstrated a linear response for DPC between 6, 10 and

15 ppm formaldehyde exposures associated with depletion of glutathione in the nasal epithelium at ≥6 ppm formaldehyde, but reduced numbers of DPC per ppm formaldehyde at 2 and 0.7 ppm *(9)*. The present study quantitated endogenous dG and dG-dG cross-links in numerous tissues, including the nasal epithelium. It also quantitated the number of endogenous dA adducts in these tissues. Surprisingly, the number of endogenous formaldehyde adducts was greater than the number of exogenous adducts in nasal DNA following exposures to 10 ppm [$^{13}CD_2$]-formaldehyde.

The number of endogenous formaldehyde adducts was similar across tissues. This suggests that the marked increases in cell proliferation induced by exposure to 10 and 15 ppm plays a critical role in converting both endogenous and exogenous labile, but promutagenic adducts into mutations. At low exposures (<0.1 ppm), minimal numbers of exogenous formaldehyde adducts will be formed relative to the high number of endogenous formaldehyde adducts. Likewise, no increase in cell proliferation will be associated with such low exposures. Nasal squamous cell carcinomas are rare in non exposed rats, but were very prominent in rats exposed to 10 or 15 ppm formaldehyde. This improved understanding of the mechanisms involved strengthens the role of cytotoxicity-induced cell proliferation in the mutagenesis and carcinogenesis of formaldehyde.

We have demonstrated that exogenous [$^{13}CD_2$]-formaldehyde only induced DNA lesions at the site of contact, but not in DNA of distant organs. This result offers new insight regarding the plausibility of formaldehyde causing leukemia. Limited evidence for an association between leukemia and formaldehyde exposure have been identified in several epidemiology studies, but were not present in others *(5)*. Leukemia was increased in the U.S. formaldehyde worker cohort only when the highest vs lowest peak exposure

were compared, but not when cumulative exposure was the metric *(6)*. Likewise, a meta analysis found an increase in lymphohematopoietic malignancies, with the highest relative risk for myeloid leukemia *(7)*. In contrast, Coggon found no association with leukemia in the British cohort, even though 28% of the workers had exposures >2.0 ppm *(5)*. The IARC Working Group considered the association between formaldehyde exposure and leukemia as strong, but less than causal. They also noted that it was not possible to identify a mechanism for the induction of myeloid leukemia in humans. For example, no increase has been found in the formaldehyde concentration in the blood of exposed humans and animals, there were no detectable protein adducts or DPC in the bone marrow of rats exposed to as high as 15 ppm formaldehyde, no DPC were detectable in the bone marrow of Rhesus monkeys exposed to formaldehyde as high as 6 ppm, and no chromosomal aberrations were found in the bone marrow of rats exposed to as high as 15 ppm formaldehyde *(18)*. Zhang hypothesized three possible mechanisms for the induction of leukemia by formaldehyde *(7)*. First, formaldehyde may act directly on the bone marrow. This study demonstrated that that does not occur. He then suggested that hematopoietic stem cells/early progenitor cells in the circulation, or residing in the nasal passages, could be exposed in the nose, travel to the bone marrow and be transformed into leukemia cells. There are no data supporting this mechanism for leukemia induced by any chemical. Furthermore, the exogenous formaldehyde DNA adducts are labile and would not be expected to persist for more than a day or so, based on our data for 1 versus 5 days of exposure. The cells containing labile adducts would need to travel to the bone marrow and undergo cell division shortly after being adducted. Therefore, these hypotheses remain highly speculative. The possibility that formaldehyde induces leukemia cannot be ruled out by this study, however, our finding that exogenous

N^2-HO^{13}CD$_2$-dG adducts were only formed in nasal DNA, that the adducts are labile and that high numbers of endogenous formaldehyde adducts are always present indicate that formaldehyde is unlikely to cause leukemia.

In this study, rats were exposed to formaldehyde for 1 or 5 days (6 hours/day). Compared to the 1 day group, the 5 day group nearly doubled the amount of exogenous N^2-hydroxymethyl-dG in nasal DNA. These data indicate that formaldehyde adducts can modestly accumulate with increased exposure. Previous research compared the acute yields of DPC in pre-exposed and naive rats after 3-h exposure and found that the acute yields of DPC in pre-exposed rats were about half those of naive rats at 6 and 15 ppm *(19)*. This difference was attributed to an increase in total DNA in the noses secondary to cell proliferation. In addition, cumulative yields of DPC were also compared between pre-exposed and naive rats and the yield was not higher in pre-exposed than in those naive rats, which was attributed to the rapid repair of DPC in the rat nasal mucosa *(19)*. Even though increased cell proliferation has been demonstrated in noses of rats exposed to 10 ppm formaldehyde *(15;20)*, we found significantly increased numbers of exogenous N^2-hydroxymethyl-dG in the 5 day-exposed group, suggesting that N^2-hydroxymethyl-dG is a sensitive specific DNA biomarker of formaldehyde exposure. From this limited time course, it is not clear if the exogenous N^2-hydroxymethyl-dG level had attained steady state. Additional exposures will be needed to clarify this issue.

We detected both endogenous dG and dA adducts, but only exogenous dG adducts were observed in the nasal DNA of rats exposed to 10 ppm formaldehyde. What is different between dG and dA toward exogenous formaldehyde during exposure? Since the *in vitro* reactivity of dG and dA toward formaldehyde is quite similar, the formation of exogenous DNA adducts should not be different from direct reaction between

formaldehyde and the DNA bases, so dA adducts would be expected to occur. We hypothesize that the difference could be a consequence of different involvement of dG and dA in the formation of DPC and suggest that exogenous dG adducts are actually formed through DPC intermediates. Lysine-dG cross-links are the primary DPC formed, however, they are very labile *(21)*. Decomposition of the labile lysine-dG cross-links could yield dG monoadducts. We suggest that this does not occur with dA since it is much less involved in the formation of DPC. Exogenous formaldehyde first targets neighboring proteins due to the high reactivity of lysine residues *(22)* , followed by the further condensation with DNA to form DPC. Thus, there is little chance for exogenous formaldehyde to directly react with DNA to form adducts. In contrast, endogenous dA and dG monoaducts may arise from the direct attack of higher concentrations of intracellular formaldehyde, estimated to be present in μM concentrations *(23)*. As shown by our results, the number of endogenous dA adducts is similar, but slightly higher than that of endogenous dG adducts, which is consistent with their reactivity *in vitro*.

We found that formaldehyde induces N^2-hydroxymethyl-dG as the primary DNA monoadduct in rats exposed to formaldehyde via inhalation. However, a previous study demonstrated that formaldehyde lead to increased amounts of N^6-$HOCH_2$-dA in multiple tissues from rats exposed to NDMA and NNK *(13).* Well characterized pathways support that formaldehyde is released during intracellular metabolism of carcinogenic NDMA and NNK *(13).* Thus, increased amounts of N^6-$HOCH_2$-dA were the consequence of formaldehyde from the metabolism of these compounds *(13).* It is of interest to note that endogenous N^6-$HOCH_2$-dA was present in control rat liver at 1.96±1.86 adducts/10^7 dA *(13),* which is similar to our 1 day (2.62±0.46) and 5 day (2.87±0.65) endogenous N^6-$HOCH_2$-dA rat liver DNA. However, in our study, there was no detectable N^6-$HO^{13}CD_2$-

dA in nasal epithelial DNA following exposure to 10 ppm [$^{13}CD_2$]-formaldehyde for 1 or 5 days. N^2-hydroxymethyl-dG was not measured in the previous study *(13),* but we would predict that it will also be increased. The way formaldehyde enters the tissue (from inhalation or intracellular metabolism) may influence the formation of specific DNA adducts in metabolically active tissues.

Little is known about the biological consequence of formaldehyde-induced DNA monoadducts, including N^2-hydroxymethyl-dG and N^6-hydroxymethyl-dA. Both of these adducts could interfere with base-pairing. In addition, they can serve as precursors of DNA-DNA cross-links. In rats treated with NDMA and NNK, it has been shown that the number of N^6-$HOCH_2$-dA and dA-CH_2-dA cross-links increased as a consequence of formaldehyde generated during the metabolism of NDMA and NNK *(13).* This suggests that monoadducts are the precursor of the corresponding DNA-DNA cross-links. We have detected both endogenous and exogenous dG-dG cross-links in noses, but only endogenous dG-CH_2-dG cross-links were found in distant tissues. However, the reactivity of N^2-hydroxymethyl-dG also can contribute to the formation of artifactual cross-links during sample processing and storage, as shown in our experiment with [$^{13}C_{10}{}^{15}N_5$]-dG, where up to 65% of the cross-links were artifactually formed during sample workup and storage. Although use of lower temperature (4°C) in the sample workup and storage (-80°C) can significantly reduce artifact, special caution is needed when generating or using any quantitative data on formaldehyde-induced dG-dG cross-links. Nevertheless, no exogenous dG-dG cross-links were detected in distant tissues, even though we utilized 5-fold more DNA which could amplify true cross-link formation and formation from artifact.

Anatomical differences between rat and human nasal structures could also affect the molecular dose. Previous studies have demonstrated that formaldehyde is absorbed almost completely in the nasal passage of rats, while formaldehyde is also absorbed in adjacent tissues like nasopharynx, trachea and proximal regions of the major bronchi of monkeys *(1)*. When using DPC as an endpoint, the yield of DPC in the noses of monkeys was about an order of magnitude lower than in the noses of rats. This was attributed to species differences in minute volume and quantity of exposed tissue*(24)*. A pharmacokinetic model has been developed, predicting that the number of DPC in the human nose would be lower than those in the noses of monkeys and rats *(24)*. This would predict that lower molecular doses of formaldehyde DNA adducts would be formed in humans relative to animal models, reducing the risk in human populations exposed to formaldehyde.

The use of [$^{13}CD_2$]-formaldehyde provided a unique opportunity to distinguish the DNA adducts from different sources. However, it is well documented that deuterium is not a stable isotope and can exchange with hydrogen. The unambiguous identification and accurate quantitation of specific DNA adducts may be interfered if H-D exchange occurs in the analyzed samples. We set up the mass spectrometer to monitor all the transitions, including no H-D exchange and loss of 1 or 2 deuterium atoms. Based on the ratio of integrated peaks, we found that no H-D exchange occurred in either 1 day or 5 day exposures.

In summary, the results of this study clearly demonstrated that endogenous and exogenous formaldehyde-induced DNA adducts and DNA-DNA cross-links could be unambiguously differentiated utilizing [$^{13}CD_2$]-formaldehyde exposures and mass spectrometry. This approach allowed us to quantitatively examine the molecular dose of

endogenous and exogenous formaldehyde DNA adducts in a multitude of tissues following inhalation exposure. Moreover, we demonstrated that exogenous formaldehyde-induced N^2-HO$^{13}CD_2$-dG is the primary DNA monoadduct in rats exposed to 10 ppm [$^{13}CD_2$]-formaldehyde for 1 day or 5 days. This provides clear evidence that inhaled formaldehyde induced exposure-specific DNA adducts at the target site for carcinogenesis. It also demonstrated that N^2-hydroxymethyl-dG serves as an excellent biomarker to evaluate specific formaldehyde exposures through inhalation. This approach should also be suitable for other routes of exposure. More importantly, we have shown that exogenous formaldehyde-induced DNA monoadducts and dG-dG cross-links only occur in rat nasal mucosa, providing clear evidence that distant genotoxic effects of inhaled formaldehyde are highly implausible. The data generated in this study add greatly to our understanding of the toxicity and carcinogenicity of formaldehyde, a high volume chemical of intense public health concern.

Figures

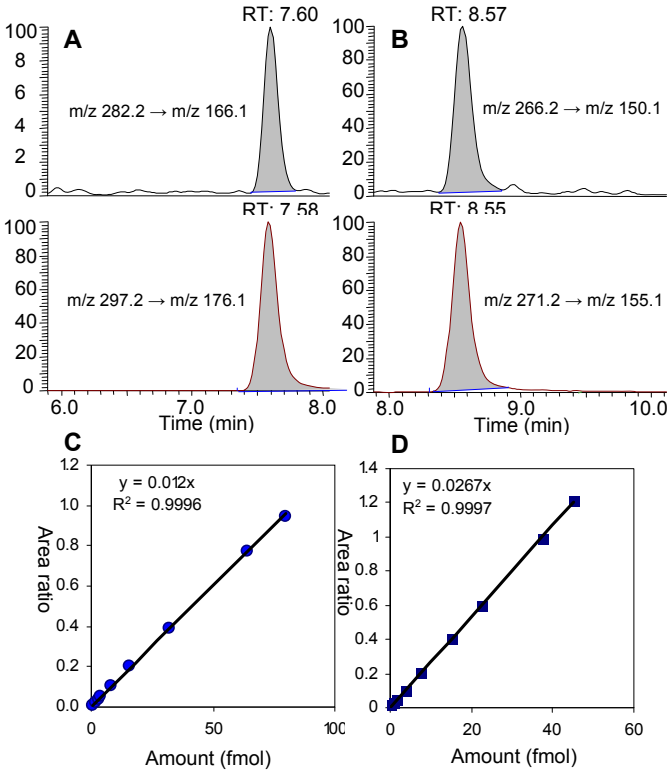


Figure 5.1. Typical LC-ESI-MS/MS SRM chromatograms standards and calibration curves. 0.8 fmol of N2-CH3-dG and 80 fmol of internal standard [$^{13}C_{10}{}^{15}N_5$]-N^2-CH_3-dG loaded on the column (A). 0.15 fmol of N^6-CH_3-dA and 37.5 fmol of internal standard [$^{15}N_5$]-N^6-CH_3-dA (B). Typical calibration curve used for the quantitation of N^2-Me-dG adducts (C). Typical calibration curves used for the quantitation of N^6-Me-dA adducts (D).

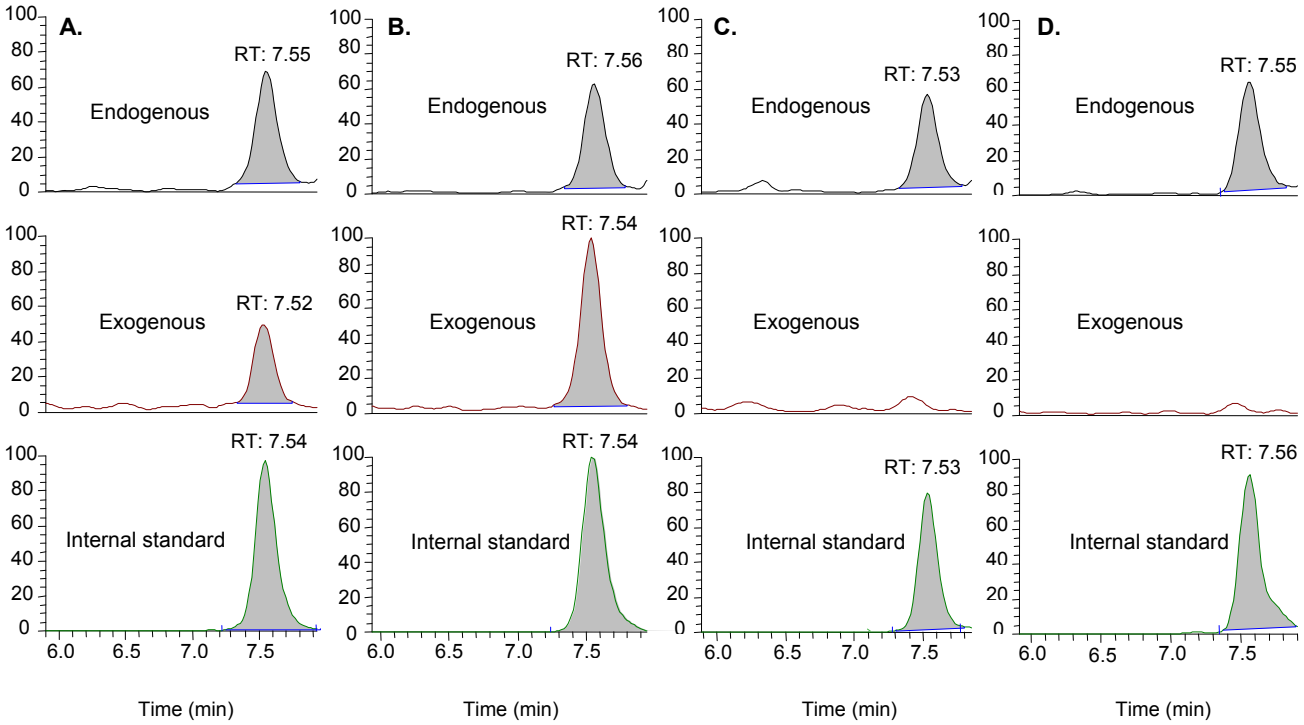


Figure 5.2. LC-ESI-MS/MS SRM chromatograms of N^2-Me-dG in typical tissues: 1 day-exposed nasal epithelium (A), 5 day-exposed nasal epithelium (B), bone marrow (C) and spleen (D).

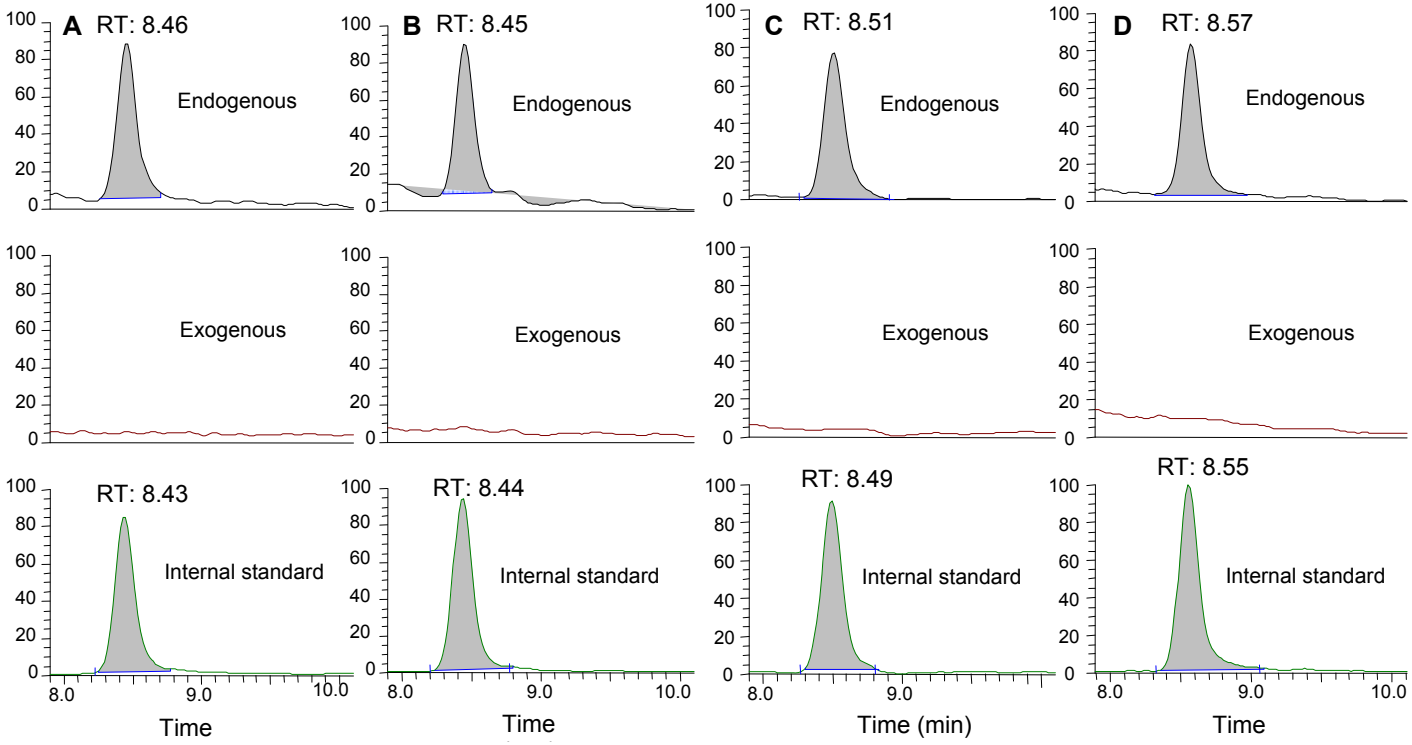


Figure 5.3. LC-ESI-MS/MS SRM chromatograms of N^6-Me-dA of typical tissues of rats: nasal epithelium of a 1 day-exposed rat (A); nasal epithelium of a 5 day-exposed rat (B) bone marrow of a 5 day-exposed rat (C); spleen of a 5 day-exposed rat (D).

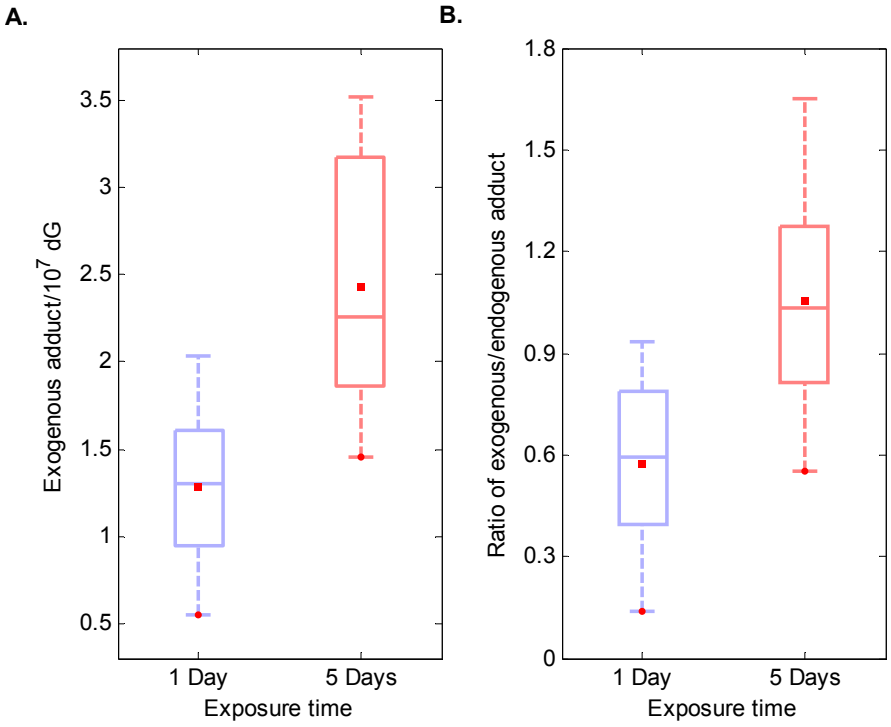


Figure 5.4. The amount of exogenous N2-HO$^{13}CD_2$-dG in nasal epithelial DNA of rats exposed to 10 ppm formaldehyde for 1 day or 5 days (A). The ratio of exogenous versus endogenous N^2-hydroxymethyl-dG for 1 day and 5 day-exposed nose samples (B).

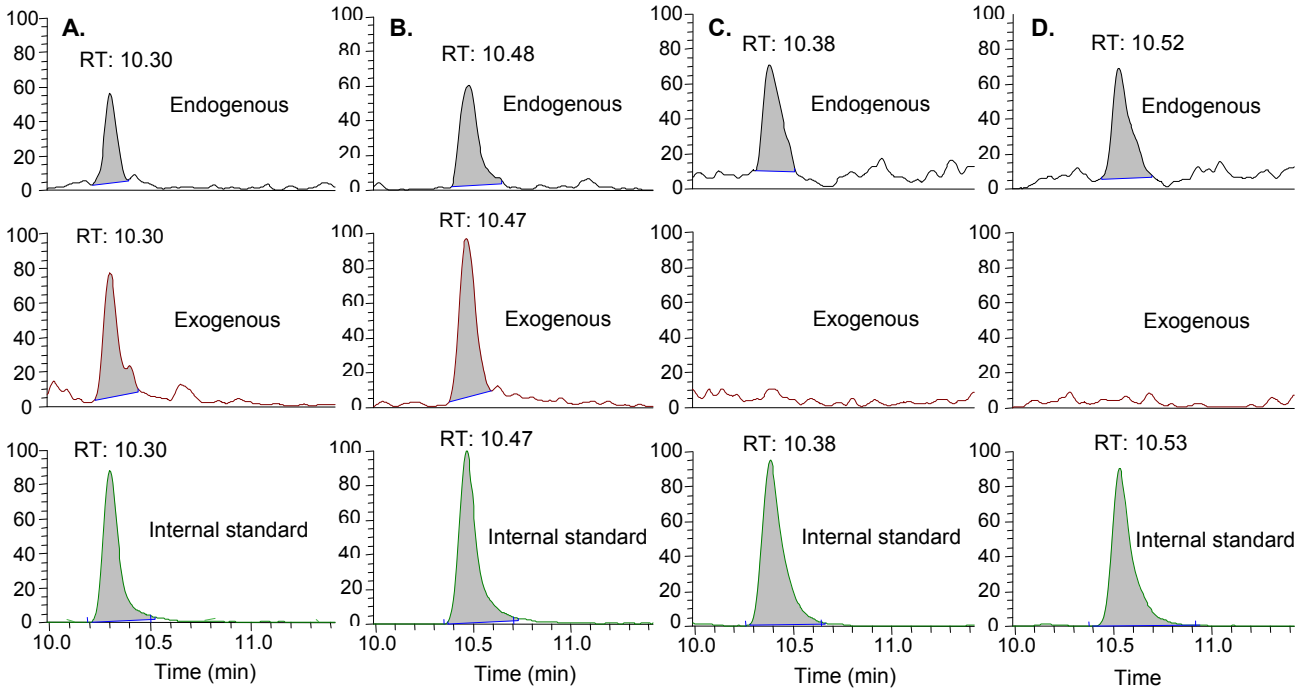


Figure 5.5. LC-ESI-MS/MS SRM chromatograms of dG-dG cross-links in typical tissues of rats: nasal epithelium of a 1 day-exposed rat (A); nasal epithelium of a 5 day-exposed rat (B); liver of a 5 day-exposed rat (C); spleen of a 5 day-exposed rat (D).

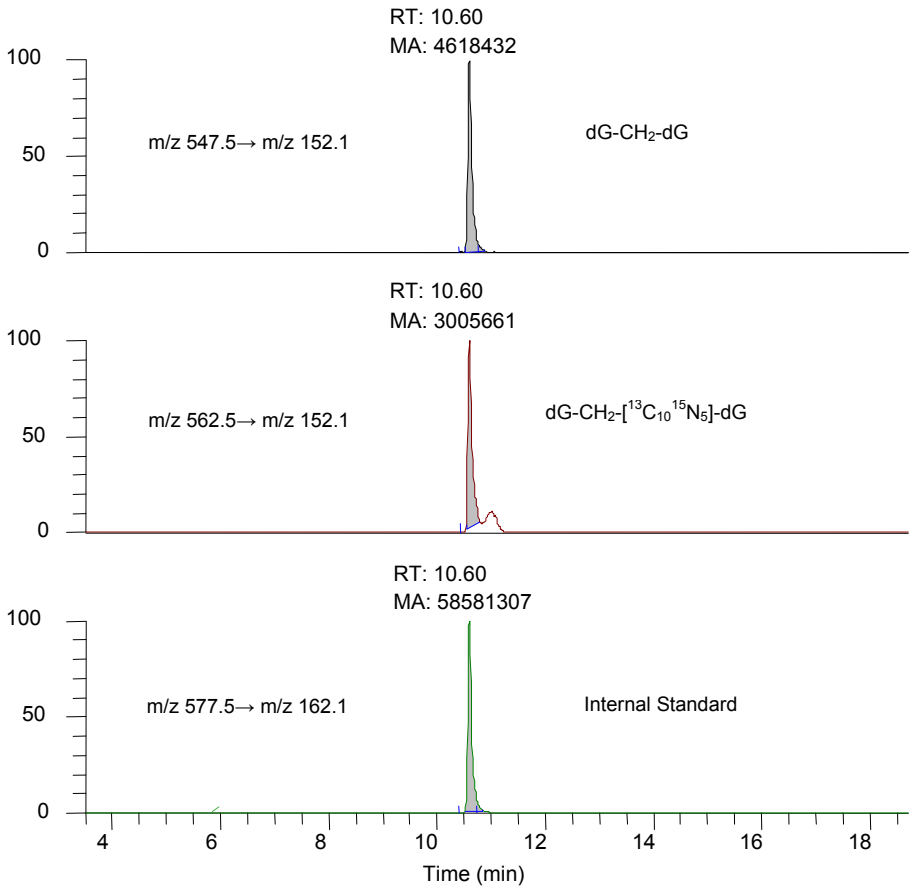


Figure 5.6. Formation of artifacts under the conditions used to analyze dG-dG cross-links in nasal DNA samples. The extent of artifacts (~65%) was determined by the area ratio of peak 562.5→152.1 (middle panel) versus peak 547.5→152.1 (top panel) in parallel control experiments by adding amounts of [$^{13}C_{10}{}^{15}N_5$]-dG equal to the amount of dG in the sample during sample workup and storage.

Tables

Table 5.1. Accuracy and precision of the LC/MS/MS-SRM analysis of monoadducts of formaldehyde*

N^2-CH_3-dG				N^6-CH_3-dA			
Added (fmol)	Detected (fmol)	Accuracy (%)	CV (%)	Added (fmol)	Detected (fmol)	Accuracy (%)	CV (%)
10	10.8±0.7[+]	108	6	10	10.2±0.5	102	5
20	22.5±0.9	112	4	20	21.6±1.2	108	5
40	42.3±0.4	105	0.1	40	40.9±1.4	102	3

[*]Rat hepatic DNA samples were spiked in triplicate with the indicated amounts of N^2-CH_3-dG or N^6-CH_3-dA

[+] Mean±SD

Table 5.2. Formaldehyde-induced monoadducts amounts in rats exposed to 10 ppm formaldehyde for 1 day and 5 days

Exposure period	Tissues	N^2-$HOCH_2$-dG		N^6-$HOCH_2$-dA	
		exogenous (adduct/10^7 dG)	endogenous (adduct/10^7 dG)	exogenous (adduct/10^7 dA)	endogenous (adduct/10^7 dA)
1 day	Nose*	1.28±0.49+	2.63±0.73	n.d.	3.95±0.26
	Lung‡	n.d.§	2.39±0.16	n.d.	2.62±0.24
	Liver	n.d.	2.66±0.53	n.d.	2.62±0.46
	Spleen	n.d.	2.35±0.31	n.d.	1.85±0.19
	Bone marrow	n.d.	1.05±0.14	n.d.	2.95±1.32
	Thymus	n.d.	2.19±0.36	n.d.	2.98±1.11
5 day	Nose	2.43±0.78	2.84±1.13	n.d.	3.61±0.95
	Lung	n.d.	2.61±0.35	n.d.	2.47±0.55
	Liver	n.d.	3.24±0.42	n.d.	2.87±0.65
	Spleen	n.d.	2.35±0.59	n.d.	2.23±0.89
	Bone marrow	n.d.	1.17±0.35	n.d.	2.99±0.08
	Thymus	n.d.	1.99±0.30	n.d.	2.48±0.11

**n=5-8 nose samples for the analysis of monoadducts in 30-50 μg of DNA; data represent mean ± SD;*

+ not detectable in 200 μg of DNA; ‡n=4-5 for distant tissues;

Table 5.3. Formaldehyde-induced dG-dG cross-links in rats exposed to 10 ppm formaldehyde for 1 day and 5 days

Exposure period	Tissues	dG-dG Cross-links	
		exogenous (adduct/10^7 dG)	endogenous (adduct/10^7 dG)
1 day	Nose*	0.14±0.06[+]	0.17±0.05
	Lung[‡]	n.d.[§]	0.20±0.04
	Liver	n.d.	0.18±0.05
	Spleen	n.d.	0.15±0.06
	Bone Marrow	n.d.	0.09±0.01
	Thymus	n.d.	0.10±0.03
5 day	Nose	0.26±0.07	0.18±0.06
	Lung	n.d.	0.20±0.03
	Liver	n.d.	0.21±0.08
	Spleen	n.d.	0.16±0.08
	Bone Marrow	n.d.	0.11±0.03
	Thymus	n.d.	0.19±0.03

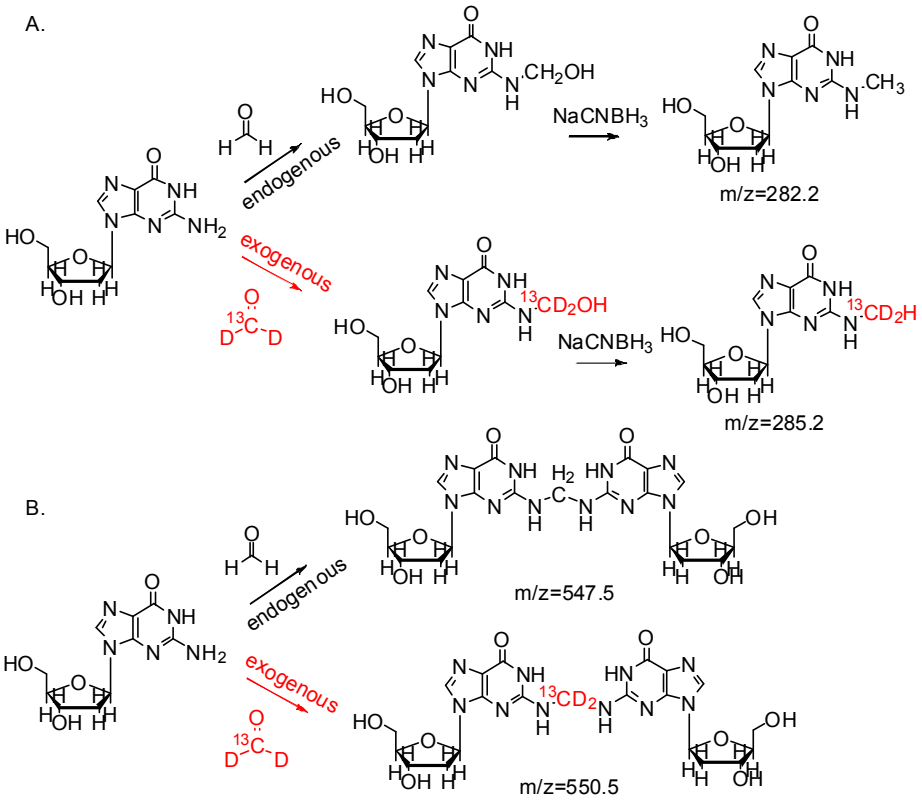
[§] *n=4-5 nose samples for the analysis of cross-links, artifacts have been subtracted from the data;*

The data on dG-dG cross-links were considered less rigorous than monoadduct data due to the issue of potential artifacts (Figure 5.6);

[¶] *n=3 for distant tissue*

Schemes

Scheme 5.1. The formation of N2-hydroxymethyl-dG (A) and dG-dG cross-links (B) originating from both endogenous and exogenous formaldehyde.



References

(1) IARC monographs on the evaluation of carcinogenic risks to humans. International Agency for Research on Cancer (2006) *IARC Monogr Eval. Carcinog. Risks Hum.* **88**, 1-287.

(2) Swenberg, J. A., Kerns, W. D., Mitchell, R. I., Gralla, E. J., and Pavkov, K. L. (1980) Induction of squamous cell carcinomas of the rat nasal cavity by inhalation exposure to formaldehyde vapor. *Cancer Res.* **40**(9), 3398-3402.

(3) Hauptmann, M., Lubin, J. H., Stewart, P. A., Hayes, R. B., and Blair, A. (2004) Mortality from solid cancers among workers in formaldehyde industries. *Am. J. Epidemiol.* **159**(12), 1117-1130.

(4) Hauptmann, M., Lubin, J. H., Stewart, P. A., Hayes, R. B., and Blair, A. (2003) Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries. *J. Natl. Cancer Inst.* **95**(21), 1615-1623.

(5) Coggon, D., Harris, E. C., Poole, J., and Palmer, K. T. (2003) Extended follow-up of a cohort of british chemical workers exposed to formaldehyde. *J. Natl. Cancer Inst.* **95**(21), 1608-1615.

(6) Beane Freeman, L. E., Blair, A., Lubin, J. H., Stewart, P. A., Hayes, R. B., Hoover, R. N., and Hauptmann, M. (2009) Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the National Cancer Institute Cohort. *J. Natl. Cancer Inst.* **101**(10), 751-761.

(7) Zhang, L., Steinmaus, C., Eastmond, D. A., Xin, X. K., and Smith, M. T. (2009) Formaldehyde exposure and leukemia: a new meta-analysis and potential mechanisms. *Mutat. Res.* **681**(2-3), 150-168.

(8) Casanova-Schmitz, M., and Heck, H. D. (1983) Effects of formaldehyde exposure on the extractability of DNA from proteins in the rat nasal mucosa. *Toxicol. Appl. Pharmacol.* **70**(1), 121-132.

(9) Casanova-Schmitz, M., Starr, T. B., and Heck, H. D. (1984) Differentiation between metabolic incorporation and covalent binding in the labeling of macromolecules in the rat nasal mucosa and bone marrow by inhaled [14C]- and [3H]formaldehyde. *Toxicol. Appl. Pharmacol.* **76**(1), 26-44.

(10) McGhee, J. D., and von Hippel, P. H. (1975) Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry* **14**(6), 1281-1296.

(11) Zhong, W., and Que Hee, S. S. (2004) Formaldehyde-induced DNA adducts as biomarkers of in vitro human nasal epithelial cell exposure to formaldehyde. *Mutat. Res.* **563**(1), 13-24.

(12) Cheng, G., Wang, M., Upadhyaya, P., Villalta, P. W., and Hecht, S. S. (2008) Formation of formaldehyde adducts in the reactions of DNA and deoxyribonucleosides with alpha-acetates of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and N-nitrosodimethylamine (NDMA). *Chem. Res. Toxicol.* **21**(3), 746-751.

(13) Wang, M., Cheng, G., Villalta, P. W., and Hecht, S. S. (2007) Development of liquid chromatography electrospray ionization tandem mass spectrometry methods for analysis of DNA adducts of formaldehyde and their application to rats treated with N-nitrosodimethylamine or 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Chem. Res. Toxicol.* **20**(8), 1141-1148.

(14) Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J., and Swenberg, J. A. (1983) Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Res.* **43**(9), 4382-4392.

(15) Monticello, T. M., Swenberg, J. A., Gross, E. A., Leininger, J. R., Kimbell, J. S., Seilkop, S., Starr, T. B., Gibson, J. E., and Morgan, K. T. (1996) Correlation of regional and nonlinear formaldehyde-induced nasal cancer with proliferating populations of cells. *Cancer Res.* **56**(5), 1012-1022.

(16) Conolly, R. B., Kimbell, J. S., Janszen, D., Schlosser, P. M., Kalisak, D., Preston, J., and Miller, F. J. (2003) Biologically motivated computational modeling of formaldehyde carcinogenicity in the F344 rat. *Toxicol. Sci.* **75**(2), 432-447.

(17) Recio, L. (1997) Oncogene and tumor suppressor gene alterations in nasal tumors. *Mutat. Res.* **380**(1-2), 27-31.

(18) Heck, H., and Casanova, M. (2004) The implausibility of leukemia induction by formaldehyde: a critical review of the biological evidence on distant-site toxicity. *Regul. Toxicol. Pharmacol.* **40**(2), 92-106.

(19) Casanova, M., Morgan, K. T., Gross, E. A., Moss, O. R., and Heck, H. A. (1994) DNA-protein cross-links and cell replication at specific sites in the nose of F344 rats exposed subchronically to formaldehyde. *Fundam. Appl. Toxicol.* **23**(4), 525-536.

(20) Chang, J. C., Gross, E. A., Swenberg, J. A., and Barrow, C. S. (1983) Nasal cavity deposition, histopathology, and cell proliferation after single or repeated formaldehyde exposures in B6C3F1 mice and F-344 rats. *Toxicol. Appl. Pharmacol.* **68**(2), 161-176.

(21) Quievryn, G., and Zhitkovich, A. (2000) Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* **21**(8), 1573-1580.

(22) Lu, K., Boysen, G., Gao, L., Collins, L. B., and Swenberg, J. A. (2008) Formaldehyde-induced histone modifications in vitro. *Chem. Res. Toxicol.* **21**(8), 1586-1593.

(23) Kato, S., Burke, P. J., Koch, T. H., and Bierbaum, V. M. (2001) Formaldehyde in human cancer cells: detection by preconcentration-chemical ionization mass spectrometry. *Anal. Chem.* **73**(13), 2992-2997.

(24) Casanova, M., Morgan, K. T., Steinhagen, W. H., Everitt, J. I., Popp, J. A., and Heck, H. D. (1991) Covalent binding of inhaled formaldehyde to DNA in the respiratory tract of rhesus monkeys: pharmacokinetics, rat-to-monkey interspecies scaling, and extrapolation to man. *Fundam. Appl. Toxicol.* **17**(2), 409-428.

CHAPTER 6

CONCLUSION AND PERSPECTIVES

In this dissertation, we have addressed several critical issues necessary to improve our understanding on formaldehyde toxicity and carcinogenesis by designing an integrated bottom-up approach. Specifically, we have shown that lysine residues of histone are highly reactive toward formaldehyde; we have elucidated the structures of DPC; we have identified a novel DNA adduct induced by formaldehyde; we have quantified endogenous and exogenous formaldehyde adducts in tissues following inhalation exposure of rats to 10 ppm [$^{13}CD_2$]-formaldehyde, and we have demonstrated that it is highly implausible that inhalation exposure of rats to formaldehyde has distant genotoxic effects. The primary conclusions from this research is summarized as follows.

6.1 Conclusions

6.1.1 Formaldehyde binding sites of histone

We started with the attempt to evaluate accessibility of histone residues to formaldehyde, since histone was reported to be involved in the formation of DPC in which the epsilon-amino groups of lysine and exocyclic amino groups of DNA were thought to be cross-linked through multiple step reactions. Using mass spectrometry, the N-terminus of histone and lysine residues located in both the histone N-terminal tail and the globular fold domain were identified as binding sites for formaldehyde in the current study. The observation that only lysine residues without post-translational modification

can be attacked by formaldehyde indicates that PTM blocks the reaction between lysine and formaldehyde. Remarkably, we found that formaldehyde-induced Schiff bases on lysine residues could inhibit the formation of PTM on histone, raising the possibility that formaldehyde might alter epigenetic regulation.

6.1.2 Structure elucidation of DPC

The identification of lysine residues on histone as binding sites of formaldehyde implies that lysine may be involved in the formation of DPC. Although DPCs are biologically important as a primary genotoxic effect and eight amino acids have been reported to form stable adducts with formaldehyde, the structures of these cross-links have not yet been elucidated. We have demonstrated that Lys, Cys, His and Trp form detectable formaldehyde-induced cross-links with dG, dA and dC. Detailed characterization of the formaldehyde-derived linkage of single amino acids with nucleosides by NMR and mass spectrometry established that these amino acids all form cross-links involving formation of a formaldehyde-derived methylene bridge. Lys yields two additional products with dG in which the linking structure is a 1,N^2-fused triazino ring. We further investigated cross-linking reactions between the four amino acids and the trinucleotides d($T_1B_2T_3$) where B_2 is the target base G, A or C. We also examined the cross-linking reactions between dG, dA and dC and 8-mer peptides containing a single reactive residue at position 5. Our results also demonstrated that Lys-dG cross-links are the primary DPC induced by formaldehyde. In addition, Lys-dG cross-links are not stable, which is consistent with previous reports on the lability of histone-DNA cross-links.

6.1.3 Identification of a novel dG-CH_2-GSH adduct

The finding that Cys-dG cross-links could be initiated by either *S*-hydroxymethyl group of cysteine residue or N^2-hydroxymethyl-dG highlighted the reactivity of *S*-hydroxymethyl group during cross-linking reaction. This finding further lead to the identification of a novel dG-CH_2-GSH adduct. Previous research has demonstrated that formaldehyde is genotoxic, causing mutations in multiple genes. However, no exogenous formaldehyde-induced DNA adducts have been detected in animals after inhalation exposure, although formaldehyde can result in N^6-deoxyadenosine, N^2-deoxyguanosine and N^4-deoxycytidine adducts *in vitro*. This can be partially attributed to the rapid metabolism of formaldehyde by glutathione-dependent enzyme systems. Among the intermediates in the pathway of formaldehyde detoxification, S-hydroxymethylgultathione is a reactive species and has the potential to further conjugate with DNA bases. We demonstrated the formation of *S*-[1-(N^2-deoxyguanosinyl)methyl]glutathione between glutathione and DNA in the presence of formaldehyde. This adduct is unique because of the involvement of *S*-hydroxymethylglutathione which is a key player during the detoxification of formaldehyde.

6.1.4 Distant genotoxic effects of formaldehyde

After our extensive work on biomarker discovery and validation involving DNA monoadducts and DNA-DNA cross-links, we applied these methods to analyze DNA samples from rats exposed to formaldehyde, trying to answer whether or not formaldehyde has distant genotoxic effects. Both genotoxicty and cytotoxicity are key

events in formaldehyde nasal carcinogenicity in rats, but no mechanistic data exist for leukemia. Highly sensitive LC-MS/MS-SRM methods were developed and [$^{13}CD_2$]-formaldehyde exposures utilized, allowing differentiation of DNA adducts and DNA-DNA cross-links originating from endogenous and inhalation-derived formaldehyde exposure. The results show that exogenous formaldehyde induced N^2-hydroxymethyl-dG monoadducts and dG-dG cross-links in DNA from rat nasal mucosa, but did not form [$^{13}CD_2$]-adducts in distant tissues. Furthermore, no N^6-HO$^{13}CD_2$-dA adducts were detected in nasal DNA, but high amounts of endogenous formaldehyde dG and dA monoadducts were present in all tissues examined. The number of exogenous N^2-HO$^{13}CD_2$-dG in 1 day and 5 day nasal DNA samples from rats exposed to 10 ppm [$^{13}CD_2$]-formaldehyde was 1.28±0.49 and 2.43±0.78 adducts/10^7 dG, respectively, while 2.63±0.73 and 2.84±1.13 N^2-HOCH$_2$-dG adducts/10^7 dG and 3.95±0.26 and 3.61±0.95 N^6-HOCH$_2$-dA endogenous adducts were present. No N^2-HO$^{13}CD_2$-dG adducts were detected in lung, liver, spleen, bone marrow or thymus, despite analyzing 5 times more DNA than for nasal epithelium, while endogenous dG and dA adducts were present in amounts similar to nasal DNA. This study provides the first quantitation of both endogenous and exogenous formaldehyde adducts. There is strong evidence supporting a genotoxic and cytotoxic mode of action for inhaled formaldehyde in the target tissue for carcinogenesis, but does not support the biological plausibility that inhaled formaldehyde causes leukemia.

6.2 Further directions

As mentioned before, we have generated data to address many of the critical challenges raised in the chapter 1. We provided mechanistic insight on the ability of

formaldehyde inducing leukemia by performing molecular dosimetry for both endogenous and inhaled formaldehyde. The results, we believe, substantially improve our knowledge on formaldehyde carcinogenesis. However, there are still some intriguing questions to be further investigated, which will bridge the current gaps and our better understanding on formaldehyde.

6.2.1 Time course experiment on the repair of formaldehyde-induced DNA adducts

We have demonstrated that inhaled formaldehyde induced N^2-hydroxymethyl-dG monoadducts and dG-dG cross-links in DNA from rat nasal mucosa, but not other distant tissues. These results support that formaldehyde is not able to travel to bone marrow to cause genotoxic effects there following exposure to rats. Tthis finding strongly refutes the first proposed mechanism by Zhang et al. We can not rule out the possibility that damaged cells can migrate to bone marrow after being exposed to inhaled formaldehyde in either blood or nasal/oral passage, although the biological plausibility of such mechanisms is slim. Under these scenarios, inhaled formaldehyde-induced DNA damage needs be stable enough to remain during the migration back to bone marrow, as well as remain there and cause further mutation during cell replication. Previous research and our current work indicate that formaldehyde-induced DNA adducts are extremely unstable *in vitro*. Therefore, the lability of these DNA lesions further casts a question on the validity of the second and third proposed mechanisms for leukemia. So far, we do not have data to demonstrate a short half-life of these adducts in animals after exposure. Therefore, time course experiment on the repair and half-life of these adducts *in vivo* should be performed to further shed light on the ability of formaldehyde inducing leukemia.

6.2.2 Dose response after exposure

Our current quantitative data on inhaled formaldehyde-induced DNA adducts are generated using rats exposed to 10 ppm formaldehyde for 1 day and 5 days. This high dose results in a 22% incidence of nasal cancer in rats after exposed for a lifetime. Presently, there are no available data on exposure-response of chemical specific adducts in rats , except previous work on DPC. Previous studies using DPC as an endpoint did show a linear response between 6, 10 and 15 ppm formaldehyde exposures for DPC, with reduced numbers of DPC per ppm formaldehyde at 2 and 0.7 ppm due to depletion of glutathione in the nasal epithelium at 6 ppm and higher exposures to formaldehyde. As we discussed before, DPC is not a chemical-specific biomarker and may not be an appropriate one for risk assessment purpose due to potential artifacts, which is a major driving force of the present study on developing specific DNA biomarker. As we demonstrated in this study, N^2-hydroxymethyl-dG is a superior specific DNA biomarker to evaluate formaldehyde exposure through inhalation. We have demonstrated the formation of this adduct in nasal epithelium of rats after exposed to 10 ppm for both 1 day and 5 days. Our next effort will focus on generating the exposure-response in rats after exposure to 0.7 ppm, 2 ppm, 6 ppm, 10 ppm and 15 ppm, [$^{13}CD_2$]-formaldehyde, which is expected to provide further critical data for science-based risk assessment of formaldehyde exposure.

6.2.3 Mechanism of the formation of dG adduct after exposure

As we have demonstrated in this study, inhaled formaldehyde induced N^2-hydroxymethyl-dG adducts as primary DNA monoadducts, while N^6-hydroxymethyl-dA

adducts were not formed as a consequence of exposure. However, both endogenous dG and dA adducts were observed in cells or tissues we examined. Why were only dG adducts induced after inhalation? Why were dA adducts formed endogenously but not exogenously. The observed difference between the formation of dA and dG adducts suggests that the underlying mechanisms through which dG and dA adducts are formed may be different. We hypothesized that endogenous dA and dG adducts could be the consequence of direct reaction between intra-cellular formaldehyde with neighboring DNA. The ratio of endogenous dA versus endogenous dG adducts in the analyzed tissues is consistent with their *in vitro* reactivity toward formaldehyde and stability (dA is a little more reactive and dA adducts are more stable). However, the observation that only dG adducts could be induced by exogenous formaldehyde indicates that dG adducts are not formed by direct reaction with exogenous formaldehyde. Otherwise, dA adducts should be also detected considering a relatively higher reactivity of dA. We hypothesize that dG addcuts are formed through certain intermediates, primarily DPC. As we showed in Chapter 3, dG instead of dA is highly involved in the formation of DPC and the primary cross-links, Lys-CH_2-dG, are extremely labile. Our preliminary *in vitro* data suggests that dG adducts could be produced after DPC decompose. Therefore, we hypothesized that exogenous dG adducts could be formed through the following steps, as illustrated in Figure 6.1: First, exogenous formaldehyde reacts with prevalent protein in cytoplasm after entering into cells; Then, formaldehyde-modified protein migrates into the nucleus and further cross-links with interacting DNA sequences to cause DPC; DPC decompose to form dG adducts.

6.2.4 Exposure routes and specific DNA biomarkers

In this study, we have clearly demonstrated N^2-hydroxymethyl-dG, rather than N^6-hydroxymethyl-dA, is the primary DNA biomarker after inhalation exposure to formaldehyde. The Hecht laboratory previously reported increased amounts of formaldehyde-induced N^6-$HOCH_2$-dA in multiple tissues of rats treated with NDMA or NNK due to the release of formaldehyde from these compounds. Thus, it is very likely dA adducts could be formed after indirect exposure to formaldehyde generated in cells, although the Hecht's study did not unambiguously demonstrate that formaldehyde was produced from the metabolism of NDMA and NNK. More recently, the Hecht Lab reported a clear difference in N^6-$HOCH_2$-dA adducts between smokers and non-smokers in leukocyte DNA samples. This clear difference (approximately 6 times higher in smokers) could be either the result of direct exposure to formaldehyde in the smoke or indirect exposure to formaldehyde produced after metabolism of various nitrosamines found in tobacco. Therefore, this study added further mystery to the critical issue whether or not specific DNA biomarkers are associated with exposure routes. We hypothesized that the formation of specific DNA biomarkers was dependent on exposure routes, as illustrated in Figure 6.2. Under direct exposure through inhalation, only dG adducts is formed with DPC being intermediates, as we explained before. However, if formaldehyde is generated in cells following exposure to certain compounds, both dA and dG are formed as the consequence of reaction between formaldehyde and DNA bases. Under this scenario, there is no evidence to support the formation of dG adducts, but we would predict the formation of dG adducts considering a similar reactivity of dA and dG toward formaldehyde. The answer to this question is of importance to rule out confounding

factors for the risk assessment of formaldehyde inhalation exposure. In addition, the finding on this issue will also add a new dimension on investigating the carcinogenesis of a variety of compounds and drugs if formaldehyde is released inside cells after metabolism.

6.2.5 Primate studies are needed to address oro-nasal breathing versus obligatory nasal breathing

Another important relevant issue is species differences between rats and humans. We did not detect the formation of dA adducts, but the Hecht Lab has demonstrated a clear increase of dA adducts in smokers. Although this adduct could be the consequence of formaldehyde released from the metabolism of nitrosamines present in tobacco, it also could arise from well known species differences in inhalation. Humans could be differentially exposed to formaldehyde through oro-nasal breathing, while rats are exposed through obligatory nasal breathing. Therefore, more studies are needed in order to develop data based on breathing patterns similar to human populations exposed to inhaled formaldehyde. This can be accomplished by exposing monkeys to 6 ppm [$^{13}CD_2$]-formaldehyde and measuring [$^{13}CD_2$]-formaldehyde DNA adducts in tissues and blood. Such studies are clearly required to establish distant site exposure.

Figures

Figure 6.1. Formation of formaldehyde-DNA adducts through endogenous and exogenous pathway

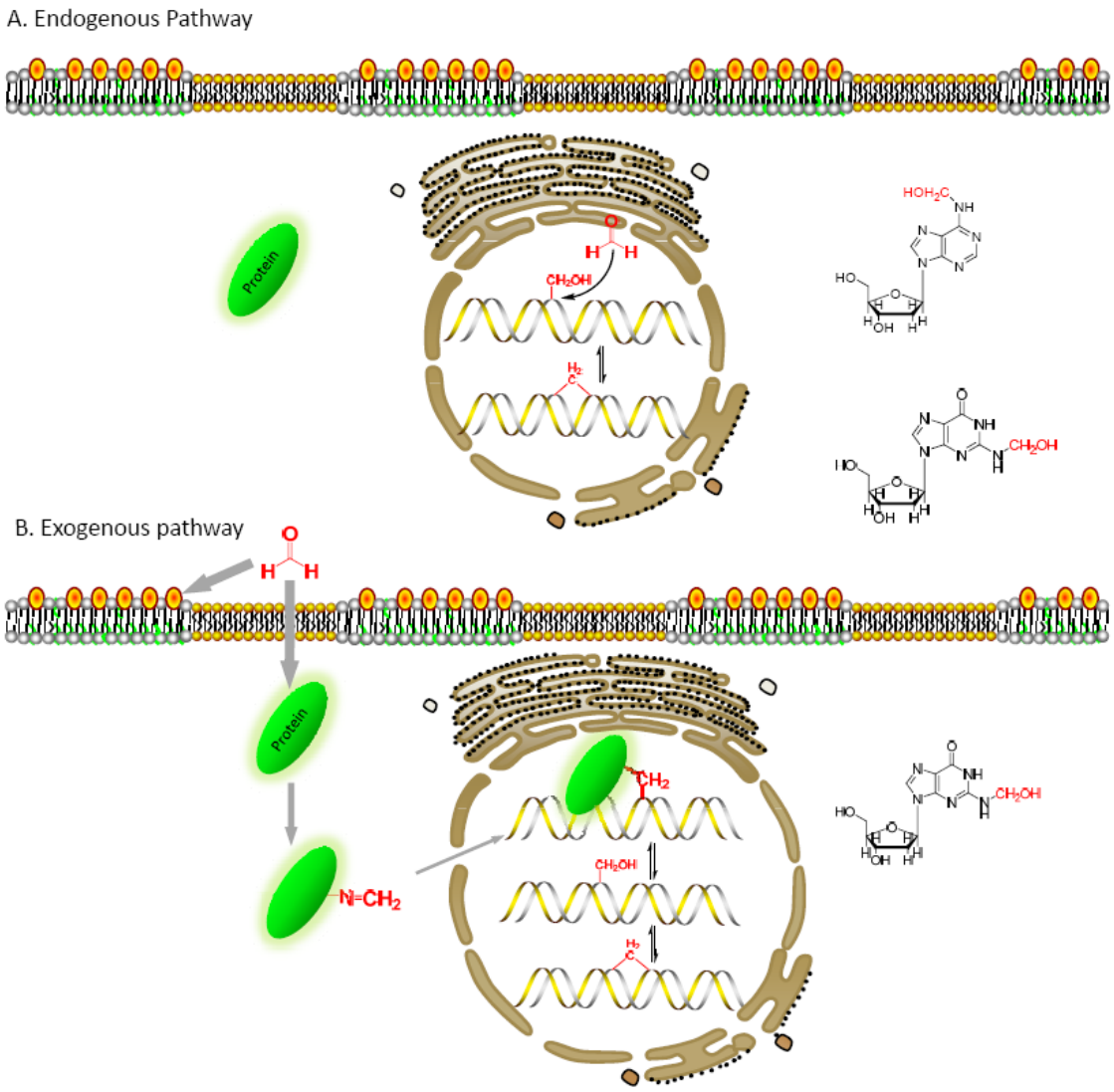


Figure 6.2. Formation of DNA adducts arising from formaldehyde generated through metabolism (indirect exposure) and exogenous formaldehyde (direct exposure).