Application of Molecular Mass Spectrometry for The Structural Characterization of a DNA-Protein Cross-Links

Jiawei Gong, Morwena J. Solivio, Edward J. Merino, Julio A. Landero-Figueroa*

Agilent Technologies/University of Cincinnati Metallomics Center of the Americas, Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221-0172, USA
*Corresponding author, email: Landerjo@ucmail.uc.edu

Abstract. To date, many different analytical methods have been used to investigate the cross-linking reaction mechanism and to obtain the chemical structure of DNA-Protein Cross-links (DPCs). Direct MS analysis of DPCs is challenging because of the ionization properties of DNA and the protein. However, peptide sequencing and mass spectrometry (MS) as analytical techniques are playing increasingly important roles for the structure determination of DPCs model. In our previous study, a novel approach was presented for purification, detection and quantification of DPCs by newly developed inductively coupled plasma mass spectrometry (ICPMS/MS), which allows sub-ppb detection of S and P, key heteroelements in DNA and proteins.

In this study, we enhanced our previously developed method and it was complemented by the use of molecular MS to allow complete characterization of a DNA-protein cross-link. First, a small molecule model is utilized to identify the adduct structure that will likely occur in an intact DNA-protein cross-link. We investigate the thermal stability of DNA-protein cross-links, both in an intact DPC and a small molecule adduct to determine feasibility of digestion/thermal degradation of DNA without the cross-link information being lost. Thermal degradation was conducted to reduce the cross-linked DNA into a single nucleoside. The remaining protein-nucleoside adduct then was proteolytically digested, generating a peptide-nucleoside adduct. The absence of the phosphate moiety allows for facile structural characterization *via* electrospray ionization mass spectrometry (ESI-MS). Additional calculations were done for peptide matching allowing us to determine the cross-link location in the protein, made possible *via* MS/MS analysis. Additionally, we show that steric effects play an important role in DPC formation.
Key words: DNA-Protein Crosslink; ICP-MS/MS; LC-MS.

Resumen. A la fecha, varios métodos analíticos se han empleado para investigar los mecanismos de reacción y obtener la estructura química de aductos ADN-Proteína (DPCs por sus siglas en inglés). El análisis por MS de los DPCs es difícil debido a las propiedades de ionización del ADN y las proteínas. Sin embargo, la secuenciación de péptidos, y el análisis por espectrometría de masas (MS) como métodos analíticos juegan un papel cada vez más importante en la determinación de estructuras de modelos de DPC. En nuestro estudio anterior, presentamos una nueva estrategia de purificación, detección y cuantificación de DPCs, basada en tecnología nueva de espectrometría de masas con plasma acoplado inductivamente (ICP-MS/MS), el cual permite detección de P y S, heteroátomos clave contenidos en ADN y proteínas, a niveles por debajo de partes por billón.

En este estudio, mejoramos el método reportado previamente, y fue complementado con el uso de espectrometría de masas molecular para permitir una completa caracterización de los aductos ADN-Proteína. Primero, un modelo basado en moléculas pequeñas fue empleado para identificar la estructura del aducto con alta probabilidad de ocurrir en ADN intacto. Investigamos la estabilidad térmica del aducto formado, en DPC intacto y en el modelo basado en moléculas pequeñas para determinar la posibilidad de degradar térmicamente el ADN sin perder la información del aducto. La degradación térmica se llevó a cabo para reducir el ADN a nucleósidos simples. El aducto resultante de nucleósido con la proteína fue digerido enzimáticamente, para liberar el aducto péptido-nucleósido. La ausencia del grupo fosfato permite la caracterización estructural simple vía espectrometría de masas con ionización por electro spray (ESI-MS). Cálculos adicionales fueron hechos para identificar el péptido, y determinar el sitio de acoplamiento del nucleósido a la proteína, a través del análisis de masas/masas (MS/MS). Adicionalmente, mostramos que impedimento estérico juega un papel importante en la formación de los DPCs.

Introduction

Oxidative DNA damage has been implicated in aging, diseases and environmental toxicity [1, 2]. Oxidative damage to DNA can be produced by reactive oxygen species (ROS), including superoxide radical anion, singlet oxygen, hydrogen peroxide and hydroxyl radical, which are present in the cell as a consequence of endogenous reactions, or from exogenous sources, including ionizing radiation [3, 4]. If reactions occur on the nucleobase, the preferential site is guanine, owing to its lowest redox potential (-1.3 V vs NHE) among the four DNA bases [5]. An important event that stems from the secondary oxidation of guanine is the formation of DNA-Protein crosslinks (DPCs). DPCs are extremely bulky lesions that are likely to impair various DNA processes including replication, transcription and repair [6, 7]. It has been estimated that the endogenous level of DPCs in human white blood cell range between 0.5 to 4.5 in 10^7 bases [8] with elevated levels linked to oxidative conditions such as exposure to heavy metals such as nickel and chromium [9-13], which are known carcinogens, and are also shown to be the major lesion associated with arsenite exposure [14]. This supports the theory that DNA-protein cross-link formation is linked to guanine oxidation. Previous work from our group and others, have demonstrated that *in vitro,* bound proteins readily form cross-links to DNA, likely forming from the secondary oxidation of 8OdG followed by addition of a protein's nucleophilic amino acid side chain, then forming a hydantoin-based lesion, the identity of which is dependent on the type of reactive oxygen involved [15-23]. For instance, cross-linking in the presence of copper-generated reactive oxygen produces adducts with guanidinohydantoin Gh [18], while singlet oxygen induced oxidative conditions generates spiroiminodihydantoin (Sp) -based adducts [17, 19]. Most of the previous reports were conducted on small molecule model systems involving the nucleoside dG and a nucleophilic amino acid such as lysine and tyrosine, given that these models are relatively easy to control and can be further analyzed [15, 16, 24]. A complete characterization of intact DNA-protein cross-links have been challenging due to adduct instability and multiple oxidative events that can occur in the protein. As a result, DNA-protein cross-links remain as one of the least understood types of DNA damage. The goal of this study is to develop a method to completely characterize DNA-protein cross-links plus identify the site of cross-linking and the type of lesion involved with the aid of ICP-MS/MS via sulfur and phosphorus detection at ultra-trace levels.

To date, many different analytical methods have been used to investigate the cross-linking reaction mechanism and to obtain the chemical structure of DPCs [24, 25]. Direct MS analysis of DPCs is challenging because of the different ionization properties of DNA and proteins. Additionally, an appropriate matrix for MALDI (matrix-assisted laser desorption/ionization MS) analysis of DPCs is not yet available. More recently, peptide sequencing and mass spectrometry (MS) as analytical techniques are playing increasingly important roles for the structure determination of DPCs [26]. In our previous study, a novel approach was presented for the sub-ppb detection and quantification of DPCs by newly developed inductively coupled plasma mass spectrometry/mass spectrometry (ICPMS/MS) [27]. This method is an effective technique for purifying DPCs. DNA-protein cross-links were readily isolated and purified *via* fraction collection of species containing both a sulfur and phosphorus signal from a HPLC-ICPMS/MS system, which is indicative of a molecule containing both protein and DNA and is likely a DNA-protein cross-link.

We aim here to enhance our developed method to allow complete characterization of DNA-protein cross-links. First, a small molecule model was utilized to identify the adduct structure that will likely occur in an intact DNA-protein cross-link. We investigate the thermal stability of DNA-protein cross-links, both in an intact DPC and a small molecule adduct to determine feasibility of digestion/thermal degradation of DNA without cross-link information being lost. Then, thermal degradation was conducted to reduce the cross-linked DNA into a single nucleoside. The remaining protein-nucleoside adduct is proteolytically digested generating a peptide-nucleoside adduct, which in the absence of the phosphate moiety allows for facile structural characterization via electrospray ionization mass spectrometric (ESI-MS) analysis. Additional calculation was done for peptide matching allowing us to determine the location of the cross-link in the protein. Importantly, using peptide MS/MS analysis, we were able to locate the specific amino acid position involved. Additionally, we show that steric effect places an important role in DNA-protein cross-link formation.

Experimental

Materials

All experiments were done in triplicate or greater. All chemicals were purchased from Fisher Scientific (Pittsburgh, PA, USA) unless otherwise noted. The 27-nucleotide DNA, 5'-GGGGCCCCGTCGTTTTACAACGTCGTG-3', and its complement 5'-CACGACGTTGTAAAACGACGGGGCCCC-3' were purchased from Eurofins Genomics Operon (Huntsville, AL, USA). Ribonuclease A, Rose Bengal, 2'-Deoxyguanosine, and N-acetyl-Lysine-Methyl ester were obtained from Sigma Aldrich (St. Louis, MO, USA). The sequence-grade modified trypsin was purchased from Promega (Madison, WI, USA).

Methods

2'-Deoxyguanosine-Lys derivative Oxidative Cross-Linking. Cross-linking experiments were carried out by mixing 2'-Deoxyguanosine (dG) and alpha-N-acetyl-Lysine-Methyl ester ($^{Ac}Lys^{OMe}$) with photo oxidant Rose Bengal [500 µL; 25 mM sodium phosphate, 4.5 mM 2'-Deoxyguanosine, 9 mM N-acetyl-Lysine-Methyl ester and 0.45 mM Rose Bengal, pH8 in a 1.5 mL centrifuge tube with the lid open and irradiated with an LED lamp providing 3 mW of radiative flux at 375nm. At these conditions, sufficient reaction yields for analysis were observed, and are easily translatable to future cell experiments without major toxicity. During the irradiation, the solution was held at a distance of 3 cm from the radiation source for 4 hours.

Purification of the dG-$^{Ac}Lys^{OMe}$ adduct by RP-HPLC. Purification of the cross-linking adduct from the reaction of 2'-deoxyguanosine and alpha-N-acetyl-lysine-methyl ester was carried as previously reported [27]. A high performance liquid chromatography system (HPLC) consisted on an Agilent 1100 series (Agilent Technologies, Santa Clara, CA, USA), equipped with a binary HPLC pump, an autosampler, a vacuum degasser system, a temperature column compartment, and a diode array detector. The UV absorbance was recorded at from 210nm to 400nm. A Beckman Coulter Ultrasphere C18 [5µm, 4.6mm x 15cm] revered phase (RP) column was used. The column was pre-equilibrated with water containing 0.1% formic acid for 20 minutes before the samples were injected. To identify different peaks on the chromatogram after the cross-linking reaction, the mixture containing all components of the reaction and the controls, which are the two reactants, were injected separately into the system with an injection volume of 2 µL, and the chromatograms overlapped as seen in Fig. 1-A. The compleate parameters for this RP-HPLC separation are given in Table. 1. The dG-$^{Ac}Lys^{OMe}$ adduct was eluted between 5.0 minute and 5.6 minute, as shown by the increase on intensity with longer reaction times, Fig. 1-B. This fraction was collected and concentrated to 100 µL by freeze-drying for further analysis.

Structure determination of the dG-$^{Ac}Lys^{OMe}$ adduct by ESI-ITMS. The identity verification of the dG-$^{Ac}Lys^{OMe}$ adduct was performed on a 6340 series MSD Ion Trap XCT Ultra system with direct infusion *via* a syringe pump at a flow rate of 18 µL/h (Agilent Technologies, Santa Clara, CA) with the micro fluidic nanospray chip cube interface, which is automatically loaded and positioned into the MS nanospray chamber. The system was operated in ultra-scan positive ion mode with the maximum accumulation time of 300 msec. The ESI-ITMS was tuned under positive ion mode with the ESI Tuning Mix for Ion Trap (Agilent Technologies, Santa Clara, CA) containing compounds at the following m/z: 118, 322, 622, 922, 1,522, and 2,122. All MS^n experiments were performed with an isolation width of 1.2 mass units and the fragmentation amplitude was 0.6 (arbitrary units). Four parent ions were selected and were isolated and fragmented per duty cycle, producing MS^2 and MS^3 spectra by collision induced dissociation with He gas equivalent to ≈15 eV. Full scan mass spectra were acquired over the *m/z* range 50–2100 in the positive ion mode. For MS/MS experiments, experimental conditions consisted of: *m/z* range: 50–2000; isolation width, 2 *m/z* units; fragmentation energy, 30–200%; fragmentation time, 40 ms.

Study of the thermal stability of the cross-linking covalent bond. The thermal stability study of the cross-linking bond was done with a voltage-controlled heater using an oil bath. Thermal decomposition was used instead of enzymatic hydrolysis for two reasons, simplicity to achieve the desired level of cleavage and to avoid the possibility of steric impediments given the bulky nature of DPCs. A 50 µL volume of the samples from the RP-HPLC peak containing the dG-$^{Ac}Lys^{OMe}$ adduct were heated to 170

ºC for 1 hour and then reconstituted in 50 µL of water. The product after heating was also analyzed by the ESI-ITMS under the conditions given above.

DNA-Ribonuclease-A oxidative cross-linking and purification. DPCs were synthesized and purified with same methods mentioned in our previous study [38]. In short, the reaction mixture was separated by size exclusion chromatography, where the eluent was directed to an ICP-MS/MS system for phosphorous and sulfur analysis. The presence of phosphorous confirmed the elution profile of the DNA, sulfur of the protein moiety while both at a large molecular mass region was selected as the formed DPC.

Thermal degradation of DNA-ribonuclease-A cross-links and purification of dG-ribonuclease. The mixture containing the purified DNA-protein cross-link was heated to 170 ºC for 1 hour and then re-dissolved in 100 µL of water. Size Exclusion Chromatography (SEC) was used to separate the components into molecular weight fractions following thermal hydrolysis. The column was calibrated using a bio-rad SEC standard containing Thyroglobulin (bovine, 670kDa), γ-globulin (bovine, 158 kDa), Ovalbumin (chicken, 44 kDa), Myoglobin (horse, 17 kDa) and Vitamin B (1.35 kDa).

Trypsin proteolysis of the dG-Ribonuclease A. After the purification of dG-RNase A from the components post-thermal hydrolysis, trypsin was added to the dG-RNAse A adduct to digest the RNAse A portion of the adduct into its peptide components. Water was used as blank and 1 mg/mL Bovine Serum Albumin (BSA) was used as standard to evaluate the extent of proteolysis. Digestion was conducted according to vendor instructions, the reaction was carried out in a pH buffered solution consisting on 15 µL of 50 mM ammonium bicarbonate, 1.5 µL of 100 mM DDT, and 10 µL of sample for digestion in a 0.5 mL tube. The mixture was adjusted to 27 µL with ultrapure water and the sample was incubated at 95ºC for 5 minutes. After cooling, a 3 µL volume of 100 mM iodoacetamide was then added into the mixture and incubated in the dark at room temperature for 20 min, followed by addition of 2 µL of sequence grade modified trypsin to a final trypsin concentration of 0.1 mg/mL. The mixture was subsequently incubated overnight at 37ºC. 1 µL volume of 88% formic acid was added to the mixture to stop the reaction. The mixture was then subjected to ultrafiltration in order to eliminate trypsin and larger digestion fragments by loading into a 0.5 mL MWCO spin filter with a 10,000 Da (nominal) cut-off membrane and centrifuged for 10 minutes at 10,000 x g. The filtrate containing the peptide fragments was injected into the HPLC-Chip-ESI-ITMS system.

Peptide separation and identification by HPLC-Chip-ESI-ITMS. LC-MS/MS analysis of the peptide mixture was performed on an Agilent 6340 series HPLC- Chip-IonTrap XCT system (Agilent Technologies, Santa Clara, CA), equipped with a Chip Cube interface with an electrospray ionization source, a well plate sampler, a nano- flow binary pump, and a capillary-flow binary pump. The Chip Cube can be operated in two modes, enrichment and analysis modes. A Large Capacity Chip II (Agilent Technologies, Santa Clara, CA) contains an enrichment 180 nL column and a 0.075x150 mm analysis column, both made of stable-bond C18 silica with 300 Å pore size packed with 5 um diameter particles. In the enrichment mode, the sample was loaded on to the enrichment column with the isocratic flow (3% B pumped by the capillary-flow pump, and then, while in analysis mode, the sample was then separated on the analytical column with the gradient elution flow pumped by the nano-flow pump. The gradient for nano-flow pump is shown in Table 1, along with other parameters for HPLC-Chip-ITMS. A 2 uL volume of digested standard BSA as quality control and the treated DPC sample solution were injected into the system separately.

MASCOT Protein Data Base Search. The MS and MS/MS spectra obtained from HPLC-Chip-ESI-ITMS were extracted and deconvoluted before they were submitted for MASCOT protein database search (Matrix Science–London, UK). They were searched under the database NCBInr. "Carbamidomethyl (C)" was selected as fixed modification, oxidation of methionine as variable modification. The MASCOT score of significance threshold was 40. The MS/MS tolerance was set as ±0.6 Da, the MS peptide tolerance as ± 1.2 Da, and peptide charges as 1+ to 3+. All the hits reported in this study have ion scores above the significance ion score thresholds.

Table 1. Operating conditions for HPLC-CHIP-ESI-ITMS

HPLC-Chip-ESI-ITMS parameters	Column	Large Capacity Chip II Zorbax 300SB-C18 (150mm x 75um, 5µm)
	Flow rate (µL/min)	3 for capillary flow pump; 0.3 for nano flow pump
	Mobile phases	A, 0.1% formic acid in water; B, 0.1 % formic acid in acetonitrile/water (90/10, v/v) mixture
	Gradient	0 min 3% B 3 min 3% B 70 min 70% B 75 min 100% B 77 min 100% B 80 min 3% B
	Nano spray needle voltage (V)	1850
	Injection volume (µL)	2
	Drying temperature (°C)	300
	Drying gas (L/min N_2)	3.0
	m/z range	50-2200

Results and Discussion

A model DPC: the dG-^{Ac}Lys^{OMe} adduct. It was previously shown that under singlet oxygen oxidative conditions, 2-deoxyguanosine is oxidized forming dGox and subsequently reacts with a nucleophile (i.e. water) to form the hyper-oxidation products such as Sp and Gh. In this work, we investigate adduct formation between the amino acid N-acetyl-Lysine-Methyl ester (^{Ac}Lys^{OMe}), containing a nucleophilic side chain with dGox to produce the dG-^{Ac}Lys^{OMe} adduct. The reaction mixture was injected and separated via HPLC. To identify the retention times of the of the reactants and products, solutions containing the individual reactants, 2-deoxguanosine and ^{Ac}Lys^{OMe} and the cross-linking reaction mixture were injected separately. Reactions were allowed to proceed for 1 hour, 2 hours or 4 hours before HPLC analysis. As shown in Fig. 1, the retention times of un-reacted ^{Ac}Lys^{OME} and dG was determined to be 3.5 and 7 minutes, respectively. Overlaying the chromatograms from the three time points, 1, 2 and 4 hours, a peak unique to the cross-linking reaction at 5.2 minutes increased in intensity with irradiation time. This peak possesses a retention time that is halfway between that of unreacted ^{Ac}Lys^{OME} and dG suggesting a combined hydrophobicity of the two components and is likely the adduct between ^{Ac}Lys^{OME} and dG. This adduct was purified via fraction collection, for further analyses.
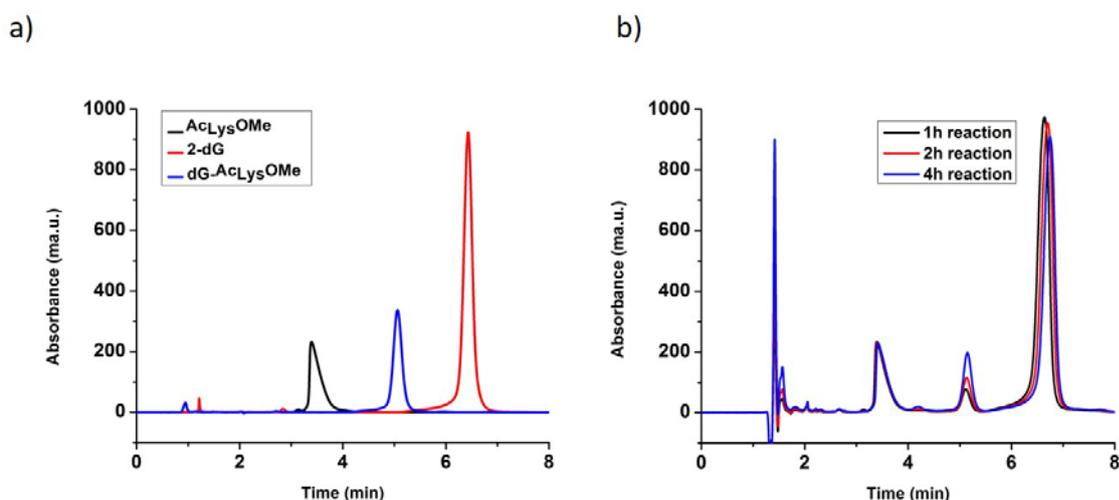
Fig. 1. a) Reversed Phase Chromatogram with UV detection of two reactants (^{Ac}Lys^{OMe}, 2-dG) and and purified dG-^{Ac}Lys^{OMe} b) Reversed Phase chromatogram with UV detection of the reaction mix after 1 hour, 2 h and 4 h reaction times

Thermolysis of the dG-^{Ac}Lys^{OMe}. The dG-^{Ac}Lys^{OMe} cross-link peak was collected and analyzed by flow Injection ESI-ITMS via syringe pump injections. Mass spectral analysis reveals a species with a m/z = 484.1, which is consistent with the of [Sp-5-^{Ac}Lys^{OMe}H]$^+$. Fig. 2a shows the resulting mass spectra and the proposed structure of the adduct. MS^2 fragmentation generates a m/z value of 386.1, which corresponds to the loss of ribose and a proton from the molecule, suggesting a nucleoside component of the adduct. The mass spectra and the proposed structure is shown in Fig. 2b.
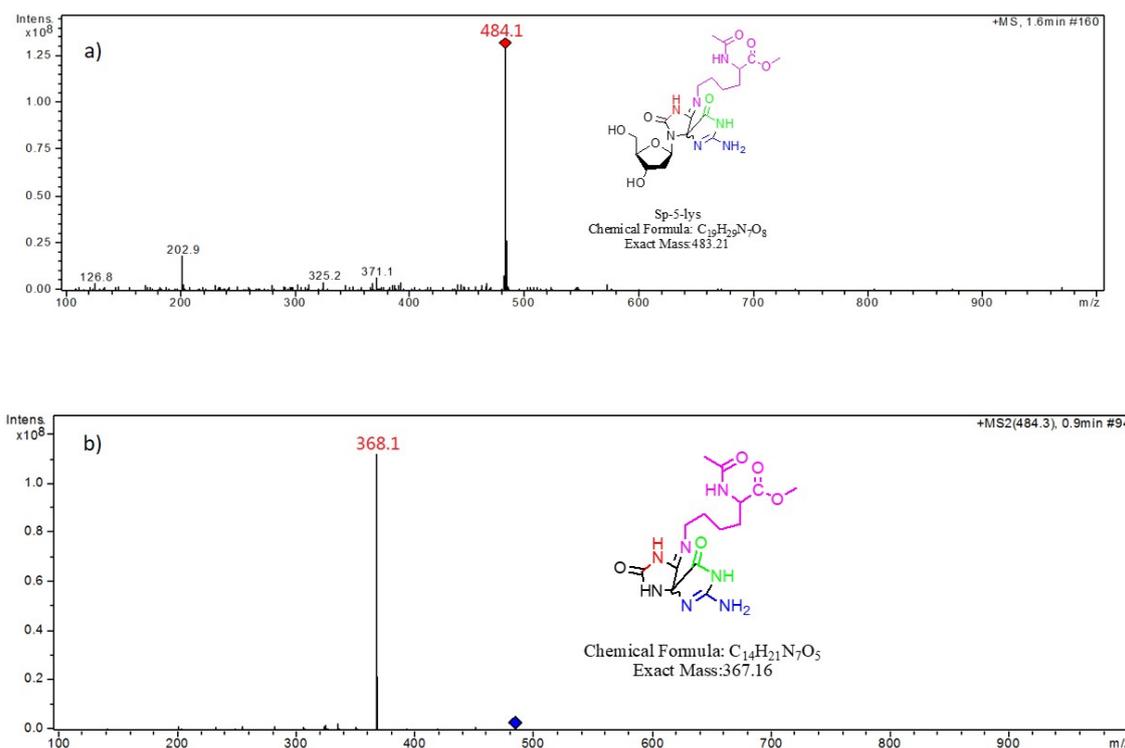
Fig. 2. a) MS of the dG-^{Ac}Lys^{OMe} cross-link compound, b) MS^2 of the dG-^{Ac}Lys^{OMe} cross-link compound, both at 15 eV CID energy.

Mass spectral analysis supports the formation of Sp-^{Ac}Lys^{OMe} as the predominant product during singlet oxygen-induced oxidation, which is consistent with previous results from our group. In order to investigate the thermal stability of the cross-link covalent bond, the purified adduct was dried and heated at 170 °C for one hour, re-suspended in 50 µL doubly distilled water and injected into the ITMS system. As shown in Fig. 3, in the resulting mass spectra, m/z 484 corresponds to the intact adduct and m/z 368.1 corresponds to the adduct minus ribose. A m/z of 203, assigned as ^{Ac}Lys^{OMe}, indicates a fragmentation at the covalent bond between ^{Ac}Lys^{OMe} and dSp. The mechanism of cross-link formation and thermal degradation is described in Fig. 4.

Article

J. Mex. Chem. Soc. **2018**, *62*(2)
Special Issue
©2018, Sociedad Química de México
ISSN-e 2594-0317
ISSN 1870-249X
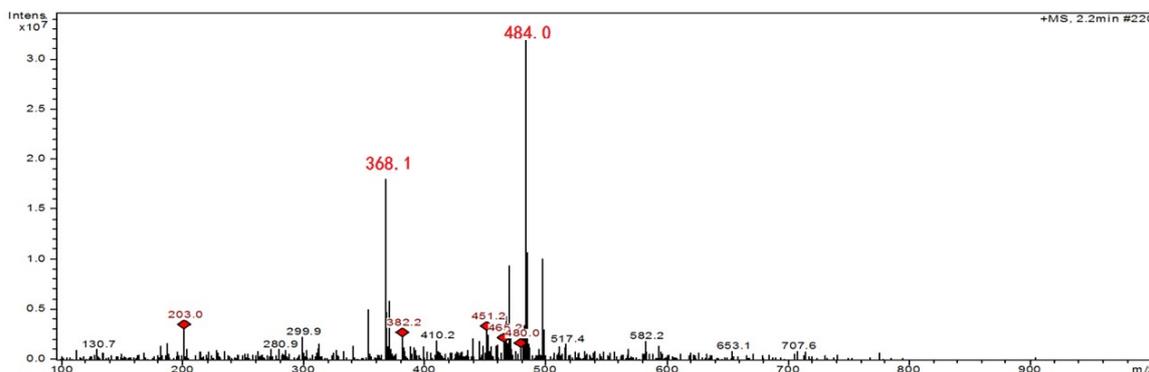
Fig. 3. Mass spectra of the dG-^{Ac}Lys^{OMe} cross-link compound after heating.
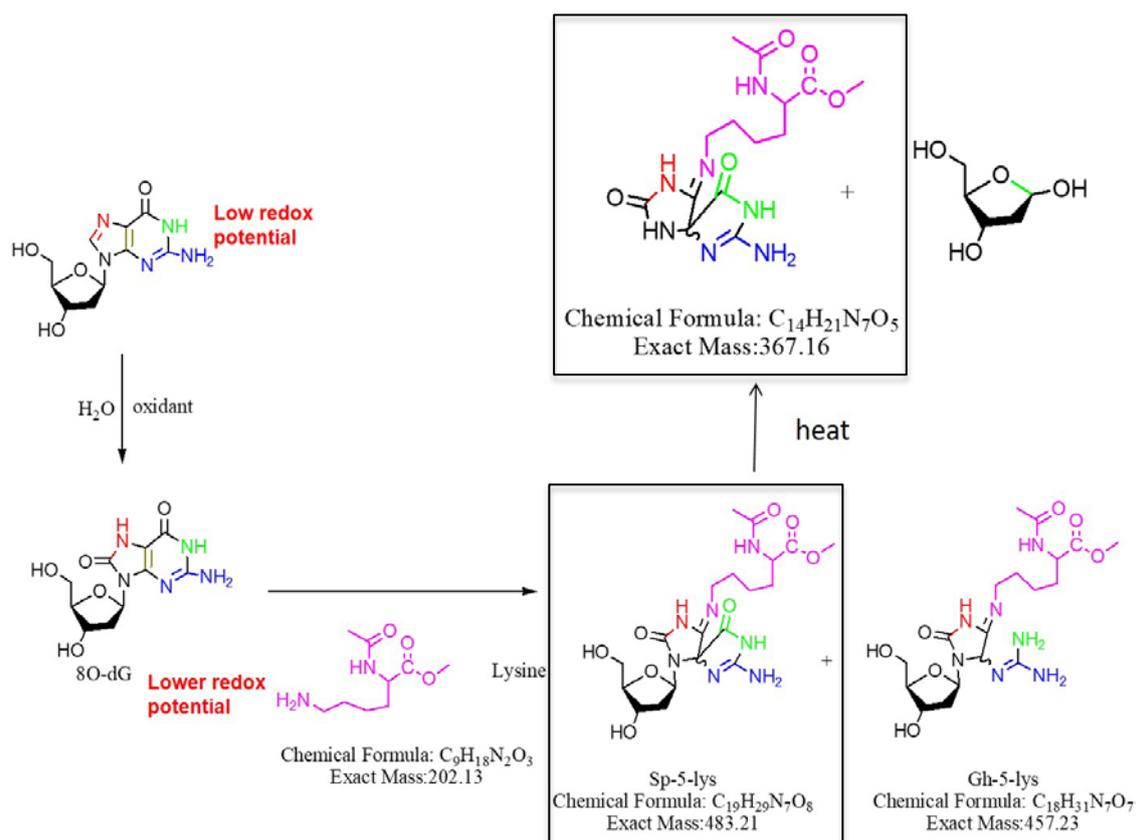
Fig. 4. Proposed reaction route for the thermal degradation of the dG-^{Ac}Lys^{OMe} adduct, the structures observed by MS analysis shown in Figure 3 are highlighted.

Identification of whole protein DPCs. Having shown that the covalent bond between ^{Ac}Lys^{OMe} and dGox is thermally stable under our experimental conditions, the purified DNA-RNase cross-linking adduct was heated and hydrolyzed under the same conditions to determine if similar stability would

Article

J. Mex. Chem. Soc. **2018**, *62*(2)
Special Issue
©2018, Sociedad Química de México
ISSN-e 2594-0317
ISSN 1870-249X

be observed. The solution after thermal treatment was separated by SEC-HPLC with UV-Vis detection. In Fig. 5, an iso-absorbance plot of the chromatogram obtained is shown. Due to the effect of its high hydrodynamic radius (highest MW), the earliest eluting peak at 260 nm was assigned to the intact DNA-RNAse A adduct. Additionally, two other peaks are observed with the absorbance maximum around 260 nm. By comparing the elution volume with the SEC standards, the peak with a retention time of 24.2 min corresponds to the DNA strand while the peaks eluting after 30 minutes are low molecular mass decomposition products of the thermal degradation. Most importantly, the fraction eluting at 26 min showed two UV absorbance bands; one at 280 nm characteristic for protein and the second at 240 nm, matching the cross-linking bond specific absorbance. Under our experimental conditions, we conclude that this fraction exhibited the unique absorbance for protein and oxidized dG is the RNAse A-dG residue produced after thermal degradation, therefore it was of interest and was isolated *via* fraction collection for further studies.
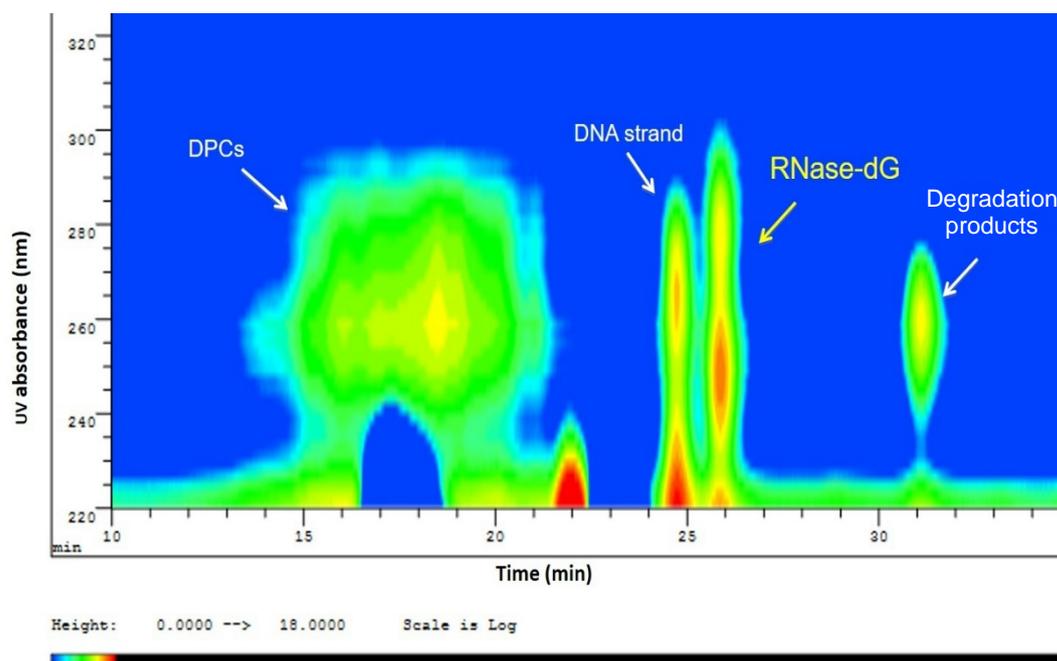
Fig. 5. Iso-absorbance plot of the SEC-UV-Vis separation after thermal degradation of DPC showing partial cleavage of the glycosidic bond, leaving a DNA strand and the Rnase-dG adduct. At the same time, some of the DPC remained stable, which eluted earlier

Identification of the dG-containing peptide-by-peptide sequence matching. The isolated dG-RNAse A adduct from the preceding section was subjected to trypsin digestion and analyzed using HPLC-Chip-ESI-ITMS. The resulting mass spectrum was used to search the MASCOT database for peptide identification. The search results are shown in Table S1. For the trypsin proteolysis of the RNase A, the maximum missed cleavage was set as 2 and the minimum peptide length was set as 5. Variable carbamidomethyl modification of Cys was accepted. The detection range was set from m/z 250 to 3000. Tryptic digest of RNAse A will result to 45 peptide fragments as predicted by the PeptideCutter search tool (http://web.expasy.org/peptide_cutter/). Lysine containing peptides are predicted to form an adduct with guanine in DNA. Addition of one dG molecule will result in Δm/z of 281.1. Because there are multiple lysine residues in the RNAse A sequence, each of these residues can potentially form adducts with dG resulting to peptide fragment cross-linking with multiple equivalents of dG. For example, if the molecular weight of one peptide is M, it is possible to observe $[M-H-ndG]^+$, $[M-H-ndG]^{++}$, and $[M-H-ndG]^{+++}$. All the possibilities for the m/z value were calculated and shown in Table S1. The method of calculation is detailed in the Electronic Supplement. The extracted ion chromatogram (EIC) for each of these possibilities was checked and both MS and MS^2 were also extracted to determine if any distinct peptide feature is observed.

Two values from Table S1, 810.36 m/z and 640.78 m/z fit the requirements for a cross-linked peptide. Their EIC showed an apparent peak and MS and MS2 indicated distinctive characteristics and peptide. The 810.36 m/z matched the m/z value of **[M-3H-1dG]**[+++], where M corresponds to the peptide, **ETGSSKYPNCAYKTTQANK**. Since there are three lysine residues in this peptide, dG can be cross-linked to any or all of these residues. Additionally, dG can form adducts with tyrosine residues in the sequence giving an m/z that is equal to that of dSp-Lysine adduct. To account for all the possible masses, a theoretical fragmentation table for each cross-linking possibility was constructed. The thirty-one most intense peaks during MS^2 are compared with the predicted values in the fragmentation table, with the highest coverage of ten of the thirty-one peaks matching the theoretical fragmentation table for the adduct, **ETGSSK(dG)YPNCAYKTTQANK.** Two cross-links had low coverage, that is, 7 in one of the cross-links and 2 in the other of the 31 most intensive peaks, show a match on the fragmentation table for the MS^2 from the corresponding EICs. Although 640.78 m/z fit the MS characteristics of KETAAK, further investigation shows that 0 of the 31 peaks in the MS^2 spectrum matched with the theoretical fragmentation table, meaning this fit is coincidental. The MS/MS spectra and its features are shown in Fig. 6.

#	b	b++	b*	b*++	b0	b0++	Sequence	y	y++	y*	y*++	y0	y0++	#
1	130.05	65.529			112.04	56.523	E							19
2	231.1	116.05			213.09	107.05	T	2300.04	1150.52	2283.01	1142.01	2282.03	1141.52	18
3	288.12	144.56	271.09	136.05	270.11	135.56	G	2198.99	1100.00	2181.96	1091.48	2180.98	1090.99	17
4	375.15	188.08	358.12	179.57	357.14	179.07	S	2141.97	1071.49	2124.94	1062.97	2123.96	1062.48	16
5	462.18	231.6	445.16	223.08	444.17	222.59	S	2054.93	1027.97	2037.91	1019.46	2036.92	1018.97	15
6	871.35	436.18	854.33	427.67	853.34	427.18	K(281.076)	1967.90	984.46	1950.88	975.94	1949.89	975.45	14
7	1034.4	517.71	1017.4	509.2	1016.4	508.71	Y	1558.73	779.87	1541.71	771.36	1540.72	770.86	13
8	1131.5	566.24	1114.4	557.73	1113.5	557.23	P	1395.67	698.34	1378.64	689.82	1377.66	689.33	12
9	1245.5	623.26	1228.5	614.75	1227.5	614.25	N	1298.62	649.81	1281.59	641.30	1280.61	640.81	11
10	1405.5	703.28	1388.5	694.76	1387.5	694.27	C(modified)	1184.57	592.79	1167.55	584.28	1166.56	583.78	10
11	1476.6	738.79	1459.6	730.28	1458.6	729.79	A	1024.54	512.77	1007.52	504.26	1006.53	503.77	9
12	1639.6	820.33	1622.6	811.81	1621.6	811.32	Y	953.51	477.26	936.48	468.74	935.49	468.25	8
13	1767.7	884.37	1750.7	875.86	1749.7	875.37	K	790.44	395.72	773.42	387.21	772.43	386.72	7
14	1868.8	934.9	1851.8	926.38	1850.8	925.89	T	662.35	331.68	645.32	323.16	644.34	322.67	6
15	1969.8	985.42	1952.8	976.91	1951.8	976.42	T	561.30	281.15	544.27	272.64	543.29	272.15	5
16	2097.89	1049.45	2080.87	1040.94	2079.88	1040.44	Q	460.25	230.63	443.22	222.12	442.24	221.62	4
17	2168.93	1084.97	2151.90	1076.46	2150.92	1075.96	A	332.19	166.60	315.17	158.09	314.18	157.59	3
18	2282.97	1141.99	2265.95	1133.48	2264.96	1132.98	N	261.16	131.08	244.13	122.57	243.15	122.08	2
19							K	147.11	74.06	130.09	65.55			1

Fig. 6. a) Mass spectrum for Identification of the dG-containing peptide by peptide sequence matching. b) De novo analysis showing the calculated theoretical MS^2 based on b and y ions. Highlighted are the matches to the experimental MS/MS spectrum of Fig. 6 a)

Cross-linking prediction. Therefore, it can be concluded that the cross-linking reaction first took place on [91]K in the peptide, **ETGSSKYPNCAYKTTQANK**, on the protein ribonuclease A. The structure of ribonuclease A is shown in Fig. 7 (generated with Visual Molecular Dynamics, University of Illinois, USA) and the peptide sequence is labeled in red chain with [91]K shown in balls. From this structure scheme it can be seen that this lysine is exposed, making it sterically possible to associate with the DNA. Secondly, this peptide appears as a loop, which is preferable to trap the DNA.
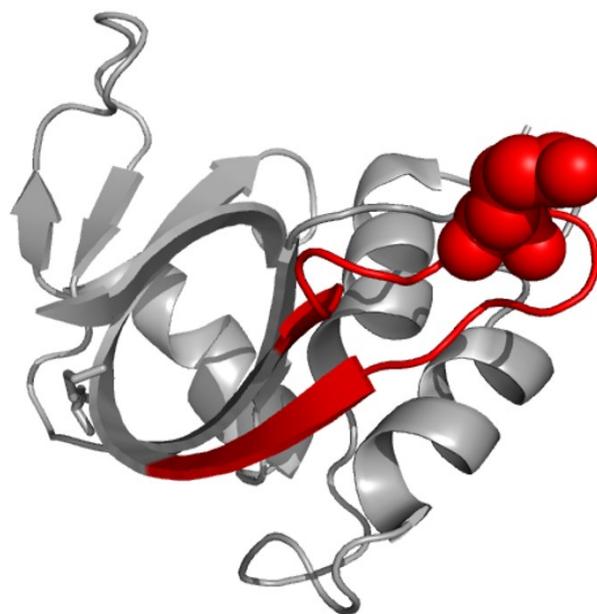
Fig. 7. 3D simulated structure of the Ribonuclease A, in red the proposed interaction site with the DNA material.

Conclusions

In this work, we demonstrate thermal stability of DNA-protein adducts, both in the nucleoside-amino acid model, and more importantly, in the DNA-protein model. These findings suggest that these lesions are stable and are not easily reversed. There are two important implications. The first one is biological in that, DNA-protein cross-link formation will have significant consequences when generated inside the cell if it remains unrepaired. It is likely to block important cellular processes such as replication (producing the highly detrimental double strand breaks that can lead to genetic instability and cancer) and can also hamper the transcription process and impede production of important proteins. The second implication of these findings is that because of the stability of the adduct, a facile method of DPC characterization (i.e. the cross-linking site, the amino acid and base involved and more importantly identifying the proteins that participate in cross-linking) can be developed. In this study, we have done the first step in intact cross-link characterization by successfully hydrolyzing the adduct between DNA and protein, removing only the un-cross-linked nucleosides. This was followed by tryptic digestion, which was initially thought to induce cross-link hydrolysis and would hinder characterization of intact DPCs by mass spectrometry. We are able to identify the cross-linking site following proteolysis of cross-linked RNAse A. This is a significant advance from the small molecule models and will be useful for total characterization of DNA-protein cross-links, protein identification and in biomarker discovery, where the cross-link identity has potential use as fingerprint for disease.

Acknowledgements

References

1. Beckman, K. B.; Ames, B. N. *The Journal of biological chemistry* **1997,** *272* (32), 19633-6.
2. Cooke, M. S.; Olinski, R.; Evans, M. D. *Clinica chimica acta; international journal of clinical chemistry* **2006,** *365* (1-2), 30-49.
3. Winterbourn, C. C. *Nature chemical biology* **2008,** *4* (5), 278-86.
4. Nathan, C.; Cunningham-Bussel, A. *Nature reviews. Immunology* **2013,** *13* (5), 349-61.
5. Steenken, S.; Jovanovic, S. V. *J Am Chem Soc* **1997,** *119* (3), 617-618.
6. Salem, A. M.; Nakano, T.; Takuwa, M.; Matoba, N.; Tsuboi, T.; Terato, H.; Yamamoto, K.; Yamada, M.; Nohmi, T.; Ide, H. *Journal of bacteriology* **2009,** *191* (18), 5657-68.
7. Ide, H.; Shoulkamy, M. I.; Nakano, T.; Miyamoto-Matsubara, M.; Salem, A. M. *Mutation research* **2011,** *711* (1-2), 113-22.
8. Voitkun, V.; Zhitkovich, A. *Mutation research* **1999,** *424* (1-2), 97-106.
9. Patierno, S. R.; Costa, M. *Chem-Biol Interact* **1985,** *55* (1-2), 75-91.
10. Patierno, S. R.; Sugiyama, M.; Basilion, J. P.; Costa, M. *Cancer Res* **1985,** *45* (11), 5787-5794.
11. Misra, M.; Olinski, R.; Dizdaroglu, M.; Kasprzak, K. S. *Chem Res Toxicol* **1993,** *6* (1), 33-37.
12. Zhitkovich, A.; Voitkun, V.; Kluz, T.; Costa, M. *Environ Health Persp* **1998,** *106*, 969-974.
13. Macfie, A.; Hagan, E.; Zhitkovich, A. *Chem Res Toxicol* **2010,** *23* (2), 341-347.
14. Bau, D. T.; Wang, T. S.; Chung, C. H.; Wang, A. S. S.; Jan, K. Y. *Environ Health Persp* **2002,** *110*, 753-756.
15. Kurbanyan, K.; Nguyen, K. L.; To, P.; Rivas, E. V.; Lueras, A. M.; Kosinski, C.; Steryo, M.; Gonzalez, A.; Mah, D. A.; Stemp, E. D. *Biochemistry* **2003,** *42* (34), 10269-81.
16. Johansen, M. E.; Muller, J. G.; Xu, X.; Burrows, C. J. *Biochemistry* **2005,** *44* (15), 5660-71.
17. Xu, X.; Muller, J. G.; Ye, Y.; Burrows, C. J. *J Am Chem Soc* **2008,** *130* (2), 703-9.
18. Solivio, M. J.; Joy, T. J.; Sallans, L.; Merino, E. J. *J Inorg Biochem* **2010**.
19. Solivio, M. J.; Nemera, D. B.; Sallans, L.; Merino, E. J. *Chem Res Toxicol* **2012,** *25* (2), 326-36.
20. Zhou, C.; Greenberg, M. M. *J Am Chem Soc* **2012,** *134* (19), 8090-3.
21. Zhou, C.; Sczepanski, J. T.; Greenberg, M. M. *J Am Chem Soc* **2012,** *134* (40), 16734-41.
22. Sczepanski, J. T.; Zhou, C.; Greenberg, M. M. *Biochemistry* **2013**.
23. Zhou, C.; Sczepanski, J. T.; Greenberg, M. M. *J Am Chem Soc* **2013,** *135* (14), 5274-7.
24. Rhee, H. S.; Pugh, B. F. *Cell* **2011,** *147* (6), 1408-1419.
25. Geyer, H.; Geyer, R.; Pingoud, V. *Nucleic Acids Res* **2004,** *32* (16).
26. Doneanu, C. E.; Gafken, P. R.; Bennett, S. E.; Barofsky, D. F. A. *Anal Chem* **2004,** *76* (19), 5667-5676.
27. Gong, J. W.; Solivio, M. J.; Merino, E. J.; Caruso, J. A.; Landero-Figueroa, J. A. *Anal Bioanal Chem* **2015,** *407* (9), 2433-2437.