

Photodissociative Cross-Linking of Diazirine-Tagged Peptides with DNA Dinucleotides in the Gas Phase

Yang Liu & František Tureček

Journal of The American Society for Mass Spectrometry

The official journal of The American Society for Mass Spectrometry

ISSN 1044-0305

Volume 30

Number 10

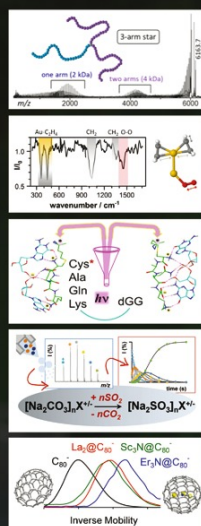
J. Am. Soc. Mass Spectrom. (2019)
30:1992–2006

DOI 10.1007/s13361-019-02189-4

Volume 30 Number 10

October 2019

Journal of the American Society for Mass Spectrometry



40617 • 30(10) 1825–2184 (2019)
ISSN 1044-0305 (Print)
ISSN 1879-1123 (Electronic)

In this Issue:

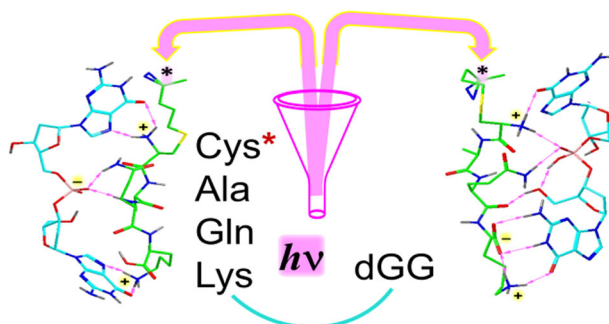
- Focus Honoring Professor Helmut Schwarz's Election to the US National Academy of Sciences
- Critical Insight on Data Analysis in Metabolomics

Your article is protected by copyright and all rights are held exclusively by American Society for Mass Spectrometry. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Photodissociative Cross-Linking of Diazirine-Tagged Peptides with DNA Dinucleotides in the Gas Phase

Yang Liu, František Tureček 

Department of Chemistry, University of Washington, Bagley Hall, Box 351700, Seattle, WA 98195-1700, USA



Abstract. Non-covalent complexes of DNA dinucleotides dAA, dAT, dGG, dGC, and dCG with diazirine-tagged Cys-Ala-Gln-Lys peptides were generated as singly charged ions in the gas phase. Laser photodissociation at 355 nm of the diazirine ring in the gas-phase complexes created carbene intermediates that underwent covalent cross-linking to the dinucleotides. The dinucleotides differed in the cross-linking yields, ranging from 27 to 36% for dAA and dAT up to 90–

98% for dGG, dGC, and dCG. Collision-induced dissociation tandem mass spectrometry (CID-MS³) of the cross-linked conjugates revealed that fragmentation occurred chiefly in the dinucleotide moieties, resulting in a loss of a nucleobase and backbone cleavages. The CID-MS³ spectra further revealed that cross-links were primarily formed in the 3'-nucleotides for the dAT, dGC, and dCG combinations. Gas-phase and solution structures of dGG complexes with S-tagged CAQK were investigated by Born-Oppenheimer molecular dynamics (BOMD) and density functional theory calculations. The low free-energy complexes had zwitterionic structures in which the peptide was protonated at the N-terminus and in the Lys residue whereas the carboxyl or dGG phosphate were deprotonated, corresponding to the respective (Cys⁺, Lys⁺, COO⁻)⁺ and (Cys⁺, Lys⁺, phosphate⁻)⁺ protomeric types. Both types preferred structures in which the peptide N-terminal cysteine carrying the S-photo-tag was aligned with the 3'-guanine moiety. BOMD trajectories at 310 K were analyzed for close contacts of the incipient peptide carbene with the positions in dGG that pointed to frequent contacts with the N-1, NH₂, and N-7 atoms of 3'-guanine, in agreement with the cross-linking results. Carbene insertion to the guanine N-1-H and NH₂ bonds was calculated by density functional and Møller-Plesset perturbational theory to be 350–380 kJ mol⁻¹ exothermic. Based on calculations, we proposed a mechanism for the carbene reaction with guanine starting with an exothermic attack at N-7 to form a dipolar intermediate that can close an aziridine ring in another exothermic reaction, forming a stable covalent cross link.

Keywords: DNA-peptide complexes, Diazirine tags, Gas-phase cross-linking, Sequence analysis, Born-Oppenheimer molecular dynamics, Close contact analysis

Received: 15 January 2019/Revised: 6 March 2019/Accepted: 6 March 2019/Published Online: 3 April 2019

Interactions of DNA with proteins are essential for life processes such as DNA replication and repair. The complexity of DNA-protein supermolecular assemblies, as revealed by X-ray crystallography [1], makes it challenging to study the

components of DNA-protein interactions at high resolution and to investigate the interaction dynamics on a time scale relevant for biochemical processes. Mass spectrometry has been used to aid identification of DNA-protein cross-links formed in solution by specific chemical reactions [2–7], as well as by ionizing radiation [8], and UV light [9–16], and the various approaches to deal with the complexity of the resulting mixtures have been reviewed [17–20]. The complexity problem can be simplified by bringing the DNA and protein components into the gas phase as non-covalent ion complexes and

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13361-019-02189-4>) contains supplementary material, which is available to authorized users.

Correspondence to: František Tureček; e-mail: turecek@chem.washington.edu

study their structure and dynamics in the absence of solvent, additives, reagents, and other intervening effects. Previous studies demonstrated the usefulness of such an approach, showing that oligonucleotide-peptide ion complexes can be generated in the gas phase [21–23], isolated by mass-to-charge ratios, and studied by tandem mass spectrometry [24–26]. However, the previous gas-phase studies were limited to using collision-induced dissociation (CID) to probe the interactions in the complexes, whereas structural details of nucleotide-peptide binding were not obtained.

We have shown recently that related peptide-peptide interactions in non-covalent ion complexes can be probed by combining site-specific photodissociative cross-linking in the gas phase with Born-Oppenheimer computational analysis of the molecular dynamics, revealing the interactions in the complexes and indicating the formation of covalent cross-links. By combining experiment and theory, we have been able to assign interactions in gas-phase complexes of the small peptide Cys-Ala-Gln-Lys (CAQK) with selected peptide motifs [27].

CAQK has recently been discovered to mitigate adverse effects of brain trauma in mice [28], possibly because of specific interactions with yet-to-be identified proteins. Because of the biological importance of CAQK, we now set out to investigate its interactions with several dinucleotides as surrogates of nuclear DNA. This small lysine-containing peptide can be viewed as a simplified surrogate of a histone interacting with DNA. Interaction with DNA and RNA of histone lysine side chains is the foundation of nucleic acids coiling in chromatin [29]. Modifications of histone lysines by alkylation and acylation are important tools for chromatin regulation [30]. However, the pattern of these modifications, the so-called histone code [31], has not been fully elucidated in spite of numerous studies [32]. In our approach, we use modified CAQK peptides which are specifically tagged with a 4,4-azipentyl group at the N-terminus (*CAQK) [33], cysteine thiol group (C*CAQK) [27], and lysine side chain (CAQK*) [33], introducing a diazirine ring that can be selectively photodissociated at 355 nm (Fig. 1). Diazirine photodissociation generates a transient carbene [34] that can undergo non-selective insertion into the proximate C–H, N–H, or O–H bonds of the partner molecule, forming a covalent bond and converting the complex into a molecular ion. The photodissociation products, both covalent and non-covalent, are further analyzed by CID tandem mass spectrometry (CID-MS³) to probe the formation and location of the covalent bonds connecting the peptide and the DNA. Since CID-MS³ often results in limited sequence coverage [35], the experimental data do not allow us to achieve atomic-level resolution in determining the cross-links and identifying all interactions in the complexes. To improve resolution, we complement experimental cross-linking data with Born-Oppenheimer molecular dynamics (BOMD) calculations that involve all valence-electron interactions along with nuclear motion at the experimental temperature. BOMD provides time-resolved trajectories that reveal contacts between the peptide and DNA atoms that may result in cross-linking in photochemically produced carbene intermediates. We now apply this

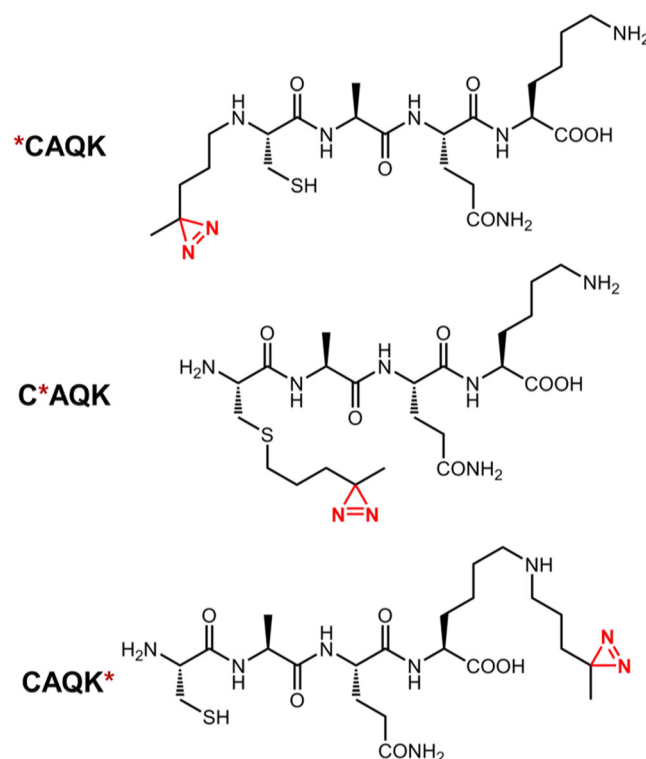


Fig. 1. Diazirine labeled Cys-Ala-Gln-Lys peptides

approach to the study of CAQK complexes with DNA dinucleotides dAA, dAT, dGG, dGC, and dCG. We wish to show that in interactions with CAQK ions, even simple dinucleotides differ in their binding efficiency and stereochemistry, and we provide structures of selected complexes obtained by electronic structure theory calculations using density functional theory (DFT).

Experimental Section

Materials and Methods

Peptides C*CAQK, *CAQK, and CAQK* were synthesized according to previously reported procedures [27]. Mass spectra were measured on an LTQ-XL-ETD (Thermo, San Jose, USA), and Bruker amaZon Speed (Bruker, Bremen, Germany) ion trap mass spectrometers equipped with an Nd-YAG laser EKSPLA laser [36]. Ions were produced by electrospray ionization of ca. 50 μ M solutions in methanol-water-acetic acid, selected by m/z and stored in the ion trap.

Calculations

Born-Oppenheimer molecular dynamics (BOMD) trajectories were run with semiempirical all-valence-electron quantum chemistry calculations using the Berendsen thermostat algorithm [37]. Temperature was set at 310 K to represent the experimental conditions in the ion trap. For each complex type, several (> 5) initial structures were constructed from PM6-optimized peptide subunits in which the diazirine tag and dGG were in different spatial orientations. These complexes

were subjected to a preliminary BOMD using PM6 [38] that was augmented by including dispersion interactions (D3H4 [39]). These calculations were run by MOPAC [40] that was coupled to the Cuby4 framework [41, 42]. Running trajectories with 1 fs steps for 20 ps furnished 20,000 snapshots per each initial structure from which 200 snapshots were extracted at 100 fs intervals. The extracted snapshot structures were fully gradient-optimized with PM6-D3H4 and sorted out by their secondary and supersecondary structural similarities to compact duplicates and reduce the size of the selection. This yielded 60–70 distinct structures whose geometries were fully optimized by DFT gradient calculations. B3LYP/6-31G(d,p) [43] optimized structures were used for frequency calculations to provide vibrational enthalpies and entropies whereas ω B97X-D/6-31+G(d,p) [44] optimized structures were used to evaluate electronic energies including dispersion interactions. The electronic, vibrational, and rotational terms were combined to yield relative free energies for the gas-phase complexes. In addition, solvent effects were included by single-point energy self-consistent reaction field calculations using the polarizable continuum model [45] with water as the dielectric and the gas-phase optimized structures. All the thermochemical calculations were performed using the Gaussian 16 (Revision A.03) [46] suite of programs. The final optimized structures with relative free energies within 40 kJ mol⁻¹ of the global energy minimum were then subjected to a full BOMD with PM6-D3H4 for 100 ps at 310 K. Cuby4 was also used to extract close contacts between the carbon atom of the diazirine ring and the dinucleotide atoms. Close contacts were limited to X–H...C distance of less than 4.5 Å on the basis of the van der Waals radii of the diazirine and X–H atoms [36].

Results and Discussion

Complex Formation and Photodissociation

Electrospray ionization of equimolar mixtures of the labeled CAQK peptides (abbreviated as M*) and dinucleotides (dXX, abbreviated as m) produced the pertinent singly charged (M* + dXX + H)⁺ complexes in ca. 1% yield relative to the dominant singly charged (M* + H)⁺ and (m + H)⁺ ions. This is illustrated by the spectra of (C*AQK + dXX) complexes in Fig. S1a–e (Supporting Information). The complexes were isolated by mass, *m/z* 1109, 1100, 1141, and 1101 for dAA, dAT, dGG, and dGC, dCG, respectively, stored in the ion trap and subjected to several laser pulses at 355 nm. This resulted in loss of N₂ and partial dissociation as shown by the UVPD-MS² mass spectra (Fig. 2a–e). With dAA and dAT complexes, UVPD resulted in substantial dissociation following the loss of N₂, producing (M* – N₂ + H)⁺ peptide ions at *m/z* 517 whereas the complementary dinucleotide (m + H)⁺ ions gave only minor peaks (*m/z* 565 and 556, Fig. 2a, b). In contrast, UVPD of the (M* + m + H)⁺ ions from dGG, dGC, and dCG, following the loss of N₂, produced (M* – N₂ + m + H)⁺ complexes with only minor dissociation to peptide and dXX monomers (Fig. 2c–e). The UVPD-MS² results for C*AQK, *CAQK, and CAQK*

complexes are summarized in Table 1. Comparing the UVPD-MS² spectra across the three photopeptides, the photolyzed C*AQK and *CAQK complexes of dGG, dGC, and dCG showed comparably high resistance to dissociation. This was expressed as a percent fraction R(MS²) of surviving (M* – N₂ + m + H)⁺ ions (Table 1) [35, 36]. The complexes of CAQK* were somewhat less stable, giving 82% survivor ions for dGG. The complexes of dAA displayed more variable stabilities, giving 36%, 27%, and 19% survivor complexes for C*AQK, *CAQK, and CAQK*, respectively.

The dissociations of the photolyzed complexes are in part driven by the highly exothermic isomerization of the carbene intermediate to an olefin ($\Delta H_{\text{isom}} \approx 200$ kJ mol⁻¹) [47] that occurs in the absence of fast X–H insertion [48–51]. The UVPD-MS² dissociations were compared with CID-MS² of (M* + m + H)⁺ ions that were not exposed to diazirine photolysis. The spectra in Fig. S2a–e (Supporting Information) indicated that among the C*AQK complexes, that with dAA was most sensitive to CID, primarily forming the (M* + H)⁺ ion at *m/z* 545. CID of the dAT, dGG, and dGC complexes produced protonated dinucleotide and peptide at comparable relative intensities. The complex of dCG was most resistant to CID, forming small but nearly equal fractions of protonated dCG and C*AQK. We note that CID did not result in elimination of N₂, the only significant dissociation besides cleavage to the monomers being a minor loss of water (Fig. S2a–e).

CID-MS³ of Photolyzed Conjugates of C*AQK

The (M* – N₂ + m + H)⁺ ions formed by photodissociation were selected by mass and subjected to CID. In the absence of covalent cross-linking, the MS³ fragmentation was expected to result in the formation of monomeric units, competing for the single available proton according to their relative basicities. In contrast, the covalently bound ions, which we call conjugates, can presumably undergo a combination of peptide and nucleotide backbone fragmentations, as well as water, ammonia, side chain, and nucleobase losses. To describe the fragment ions, we used the standard DNA ion nomenclature [52, 53], which, reading from the 5' to the 3' terminus, assigns the fragments as a, b, c, d, w, x, y, z, and those by loss of the neutral 5' and 3' base as (–B_{1H}(X_H)) and (–B_{2H}(X_H)), respectively. The protonated 5' and 3' bases are denoted as (B₁H₂⁺) and (B₂H₂⁺), respectively (Scheme 1). Backbone cleavage in the peptide was described by the standard Biemann-Roepstorff-Pohlman nomenclature for the C-terminal (x_n, y_n) and N-terminal (a_m, b_m) fragments [54, 55]. For combined dissociations in both the dXX and peptide units, we used a combined notation, e.g., w₁_B₁ refers to an ion containing the w₁ nucleobase fragment linked to the B₁ peptide fragment. A table of all theoretical fragment ion combinations was assembled and compared against the fragment ions observed in the CID-MS³ spectra. The identified fragment ions from all CID-MS³ spectra of photo-cross-linked complexes are shown in Table S1.

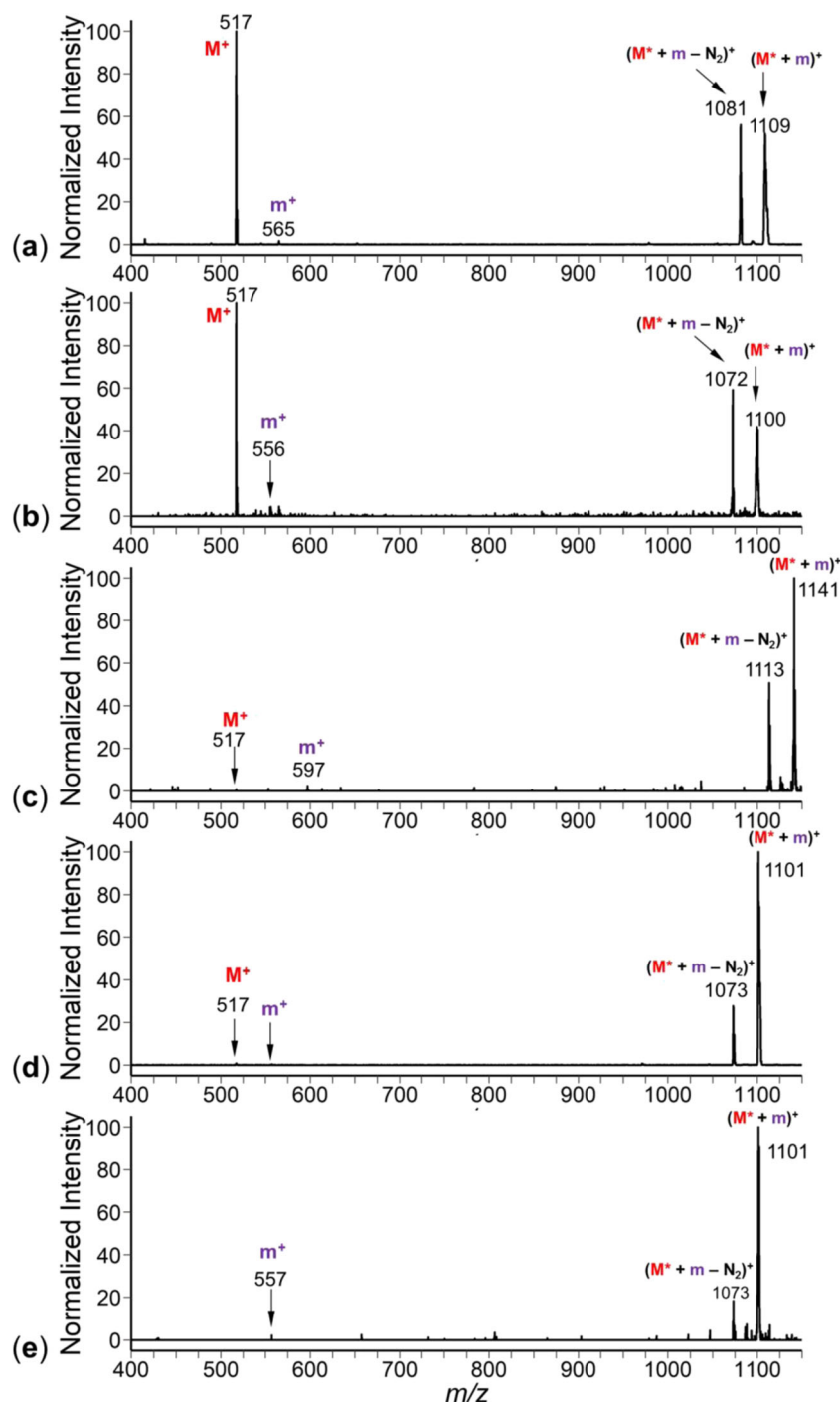


Fig. 2. UVPD-MS² spectra of C*AQK complexes with (a) dAA, (b) dAT, (c) dGG, (d) dGC, and (e) dCG

CID of the $(\text{dAA} + \text{C}^*\text{AQK} - \text{N}_2 + \text{H})^+$ ion (m/z 1081) resulted in the formation of the peptide monomer $(\text{M}^* - \text{N}_2 + \text{H})^+$ (m/z 517) and its secondary fragment $(\text{M}^* - \text{N}_2 - \text{S-tag} + \text{H})^+$ (m/z 415, Fig. 3a). In addition, the CID-MS³ spectrum showed numerous fragment ions indicating covalently linked species. We succeeded in assigning most CID-MS³ fragment ions from $(\text{dAA} + \text{C}^*\text{AQK} - \text{N}_2 + \text{H})^+$ to backbone cleavages in dAA combined with loss of a nucleobase and water. Consecutive loss of water is indicated by green arrows in Fig. 3a–e. A few fragment ions, such as m/z 659, did not correspond to logical

backbone cleavages and were unassigned. The ion series starting at m/z 946 was formed by loss of adenine ($\text{B}_{1\text{H}}$ or $\text{B}_{2\text{H}}$) and followed by successive eliminations of water (m/z 928 and m/z 910). Another prominent fragment ion series started at m/z 848 (w_1/d_1_M), indicating backbone cleavage in dAA and loss of a mononucleoside unit. In this case, as with dGG, the w_1 and d_1 fragments were isomers and thus indistinguishable by mass. Two low-intensity fragment ion series were particularly diagnostic for locating covalent cross-links. The m/z 811 ion was formed by losses of both adenine nucleobases,

Table 1. Cross-linking percent efficiencies

Photopeptide	Dinucleotide	R(MS ²) ^a	R(MS ³) ^b	R(total) ^c
C*AQK	dAA	36	58	21
	dAT	30	67	20
	dGG	98	47	46
	dGC	96	39	37
	dCG	90	38	34
*CAQK	dAA	27	58	9
	dAT	26	67	13
	dTA	36	40	14
	dGG	98	37	36
	dGC	93	27	25
	dCG	97	18	17
CAQK*	dAA	19	60	11
	dGG	82	40	33

$$^a R(\text{MS}^2) = 100 \times [M^* + m - N_2] / \{[M^+] + [m^+] + [M^* + m - N_2]\}$$

$$^b R(\text{MS}^3) = 100 \times \Sigma[\text{Cross-linked fragments}] / \{\Sigma[\text{Cross-linked fragment ions}] + [m^+] + [M^+]\}$$

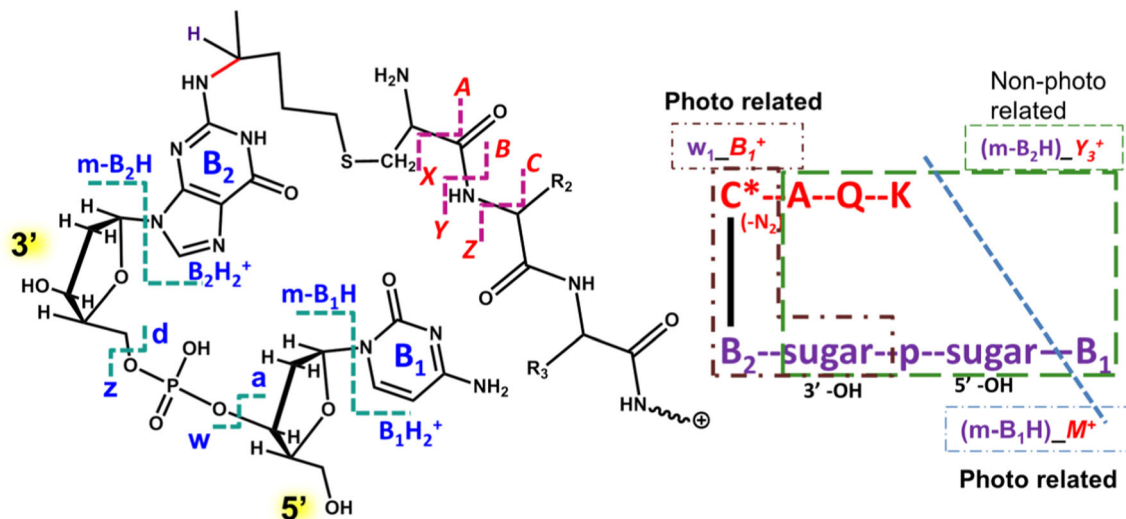
$$^c R(\text{total}) = R(\text{MS}^2) \times R(\text{MS}^3)$$

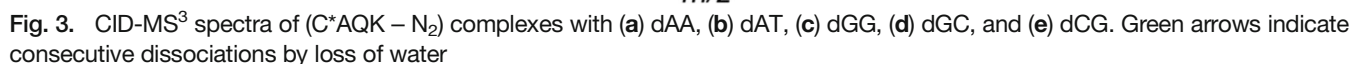
indicating peptide cross-linking to the deoxyribose-phosphate framework. Consecutive eliminations of water from the m/z 811 ion produced secondary fragment ions at m/z 793, 775, and 757. Likewise, the m/z 713 fragment ion and its secondary fragments by loss of water (m/z 695, 677, and 659) indicated loss of both nucleobases in precursors cross-linked to the deoxyribose-phosphate framework. The formation of the m/z 713 ion can be viewed as converging from multiple pathways, for example, loss of H_3PO_4 from m/z 811, loss of adenine from m/z 848, or loss of adenine-dideoxyribose (233 Da) from m/z 946. These are standard dissociations of nucleotide ions [52, 53, 56, 57]. In contrast to the above fragment ions, the m/z 652 and its loss-of-water secondary fragments were assigned to adenine-peptide cross-links. This indicated that the population of $(\text{dAA} + \text{C}^*\text{AQK} - \text{N}_2 + \text{H})^+$ ions was heterogeneous, containing isomers cross-linked to the nucleobases as well as to the deoxyribose-phosphate framework. The cross-linking efficiency for $(\text{dAA} + \text{C}^*\text{AQK} + \text{H})^+$ was $R(\text{MS}^3) = 58\%$ (Table 1), as expressed by a ratio of the intensities of the identified cross-

linked fragment ions to the sum of all fragment ion intensities in the CID-MS³ spectrum [35, 36]. Combined with the 36% fraction of $(\text{dAA} + \text{C}^*\text{AQK} - \text{N}_2 + \text{H})^+$ ions surviving UVPD-MS², the overall efficiency $R(\text{total})$ was 21%.

The CID-MS³ spectrum of the $(\text{dAT} + \text{C}^*\text{AQK} - \text{N}_2 + \text{H})^+$ ion (Fig. 3b, m/z 1072) was analyzed in a similar fashion, as briefly discussed below. A comparison of the monomer $(M^* - \text{N}_2 + \text{H})^+$ ion relative intensity with those of the cross-linked fragment ions in the m/z 530–1000 region indicated a significant fraction, $R(\text{MS}^3) = 67\%$, of covalently cross-linked conjugates of dAT. The structurally most significant fragment ions were at m/z 937 (loss of adenine) and m/z 839 (w_1_M) that both indicated peptide cross-linking to the 3'-dT moiety.

A significant fraction of cross-linked products, $R(\text{MS}^3) = 47\%$ (Table 1) was also indicated by the CID-MS³ spectrum of the $(\text{dGG} + \text{C}^*\text{AQK} - \text{N}_2 + \text{H})^+$ ion (Fig. 3c, m/z 1113). Major dissociations were a loss of guanine (m/z 962) and a backbone cleavage giving the w_1/d_1_M ion

**Scheme 1.** Nomenclature of fragment ions from dinucleotide-peptide cross-links



at m/z 864. The less intense ions at m/z 811 and m/z 713 and their secondary fragments by loss of water pointed to cross-links within the deoxyribose-phosphate framework, as discussed above for dAA. Specific cross-linking to the nucleobases was indicated by minor fragment ions at m/z 668, pertinent to G_M.

The CID-MS³ spectra of the isomeric (dGC + C*AQK - N₂ + H)⁺ and (dCG + C*AQK - N₂ + H)⁺ ions showed striking differences (Fig. 3d, e) despite similar cross-linking efficiencies, R(MS³) = 39% and 38%, respectively (Table 1). Starting with dGC (m/z 1073, Fig. 3d), the CID-MS³ spectrum displayed fragment ions by loss of both cytosine (m/z 962) and guanine (m/z 922) and their loss-of-water satellites. The loss-of-G peak was substantially more intense than that for the loss of C. Another significant fragment ion series (m/z 824, w₁_M, and its loss-of-water satellites) indicated peptide cross-linking within the 3'-dC moiety. In addition, the m/z 811 and m/z 713 fragment ions revealed the presence of isomers in which the peptide was cross-linked to the deoxyribose-phosphate framework. Peptide cross-linking to the nucleobases was indicated by minor fragment ions at m/z 668 and m/z 628 that were assigned to (G_M) and (C_M), respectively.

Contrasting the spectrum of the GC conjugate, the CID-MS³ spectrum of (dCG + C*AQK - N₂ + H)⁺ showed fragment ions by loss of cytosine (m/z 944) whereas loss of guanine (m/z 962) was undetectable (Fig. 3e). The other important fragment ions were represented by m/z 864 that was assigned as w₁_M and indicated peptide cross-linking to the 3'-dG moiety. The m/z 713 and 677 ions pointed to loss of guanine and water from the m/z 864 ion, indicating that the peptide was cross-linked to the 3'-phosphodeoxyribose moiety. However, the weaker fragment ion at m/z 668 also indicated a minor fraction of isomers where the peptide was cross-linked to the guanine residue.

Conjugates of *CAQK and CAQK*

Photodissociation of dinucleotide complexes with *CAQK gave results that were analogous to those for C*AQK. Upon photodissociative loss of N₂, the complexes of AA, AT, and TA underwent substantial dissociation to monomers, chiefly represented by the peptide ion M⁺ (m/z 517, Fig. S3a-c). The survivor (dXX + C*AQK - N₂ + H)⁺ conjugates within this group amounted to R(MS²) = 26–36% of the ion intensities (Table 2). In contrast, conjugates of GG, GC, and CG dissociated much less (Fig. S3d-f), generating 93–98% of survivor ions (Table 1). Photodissociation of complexes with CAQK* was studied for dAA and dGG only and produced results that were analogous to those for the other tagged peptides. Table 1 shows 19% and 82% survivor ion yields in the CID-MS³ spectra of conjugates with dAA and dGG, respectively. The UVPD spectra of the *CAQK complexes were compared with CID spectra that resulted in dissociation to monomers (Fig. S4a-f).

The CID-MS³ spectrum of (dAA + *CAQK - N₂ + H)⁺ (m/z 1081, Fig. S5a) showed major cross-linked fragment ions by loss of adenine (m/z 946), backbone cleavage (w₁/d₁_M, m/z 848), and B_H_M (m/z 652). The quality of this and Fig. S5b spectrum was affected by the presence of near-isobaric contaminants of ions from *CAQK disulfide that we found forming readily in solutions of peptides with free cysteine. These ions contained two diazirine groups and underwent loss of the second N₂ molecule upon CID-MS³, giving rise to fragment ions at m/z 1053 and 1044 in the respective Figs. S5a and S5b. The CID-MS³ spectrum of (dTA + *CAQK - N₂ + H)⁺ (m/z 1072, Fig. S5c) showed fragment ions by loss of adenine (m/z 937), but also an adenine-peptide adduct (m/z 652), indicating non-specific cross-linking in the 5' and 3' parts of the dinucleotide. The CID-MS³ spectrum of (dGG + *CAQK - N₂ + H)⁺ (m/z 1113, Fig. S5d) was dominated by

Table 2. Relative free energies of (dGG + C*AQK + H)⁺ complexes

Complex	$\Delta G^{\circ}_{310, g}$	Relative free energy ^a
a	7.7 ^b	1.5 ^c
b	6.3	0.6
c	0.0	0.0
d	14	-4.8
e	11	26
f	5.3	13
g	4.1	19
h	16	17
i	14	32
j	-5.8	5.0
k	1.1	36
l	-2.2	33
m	21	3.5
n	21	31
o	18	30
p	-3.8	36

^aIn kJ mol⁻¹

^bFrom ω B97X-D/6-31 + G(d,p) calculations including B3LYP/6-31G(d,p) zero-point vibrational energies, 310 K enthalpies and entropies

^cFrom ω B97X-D/6-31 + G(d,p) calculations including solvation energies in the water dielectric

fragment ions formed by loss of guanine (m/z 962), the w_1/d_1_M backbone ion (m/z 864), and their dehydrated satellites. Also in this series, the CID-MS³ spectra of the dGC and dCG complexes displayed marked differences in the fragment ion intensities (Fig. S5e, f). The dinucleotides had different propensities for competing for the proton against CAQK*–N₂, with (dCG + H)⁺ showing higher ion relative intensity than (dGC + H)⁺, implying a higher gas-phase basicity for dCG. Regarding the cross-linked fragments, loss of the 5' nucleobase was prevalent, as evidenced by the relative intensities of the m/z 922 and m/z 962 ions from the dGC and dCG conjugates, respectively. Dinucleotide backbone cleavage formed the cross-linked w_1_M ions, retaining the 3'-portion of the dinucleotide for both dGC and dCG. These results indicated that the peptide N-terminal tag in the complexes was in the vicinity of the 3'-residue regardless of whether it was guanine or cytosine.

The CID-MS³ spectrum of the (dAA + CAQK* – N₂ + H)⁺ complex was similar to that of the C*AQK complex, showing again fragment ions by loss of adenine (m/z 946), formation of w_1/d_1_M (m/z 848), m/z 811, and m/z 713 ions along with their dehydrated satellites (Fig. S6a). These results indicated that in the dAA complexes, the position of the peptide diazirine tag did not play a critical role to affect cross-linking.

Reference CID of Dinucleotide Ions

The CID-MS³ of peptide-dinucleotide conjugates revealed that most dissociations occurred in the dinucleotide moieties whereas peptide backbone dissociations were scarce or absent. To assess the effect of peptide covalent binding on the CID spectra of the conjugates, we obtained reference CID-MS² of protonated dinucleotides under the same conditions of slow heating in the ion trap, typically achieving > 80% precursor ion depletion following excitation. The reference spectra (Supplementary Figs. S7a–f) are now briefly described, and the observed fragment ions are compared with their analogues from the peptide conjugates.

CID-MS² of (dAA + H)⁺ (Fig. S7a) showed a (m-BH)⁺ > w_1/d_1^+ pattern of fragment ion intensities that was analogous to that for CID-MS³ of the C*AQK, *CAQK, and CAQK* conjugates. The [(m-BH)⁺]/[w_1/d_1^+] intensity ratio for (dAA + H)⁺ (1.4) was slightly higher than that shown for (dAA + C*AQK – N₂ + H)⁺ (1.2) in Fig. 3a. CID-MS² of (dGG + H)⁺ (Fig. S7b) showed a dominant loss of guanine but relatively much less backbone cleavage to the w_1/d_1^+ ions than was observed in CID-MS³ of the C*AQK and *CAQK conjugates. The differences for dAA and dGG may be caused by peptide cross-linking to the nucleobases, as indicated by the CID-MS³ spectra, which can be expected to hamper the loss of a free nucleobase from the conjugate. CID-MS² of (dAT + H)⁺ (Fig. S7c) chiefly produced the w_1^+ fragment ion, which was also prominent in the CID-MS³ of the dAT conjugate (Fig. 3b). The CID-MS² spectrum of (dGC + H)⁺ (Fig. S7d) displayed fragment ions by loss of G and C in a 1.1:1 ratio. This was different from the CID-MS³ spectrum of the dGC conjugate (Fig. 3d) where the loss of G was fourfold more prominent. In addition,

the conjugate showed a higher propensity for backbone cleavage forming the $w_1_M^+$ fragment ion compared to the formation of the analogous w_1^+ ion from (dGC + H)⁺. Regarding dCG, both (CG + H)⁺ (Fig. S7e) and the conjugate ion with C*AQK (Fig. 3e) preferred loss of cytosine. The main difference was in the formation of the w_1^+ fragment ion which was more prominent in the CID-MS³ spectrum of the conjugate (m/z 713, Fig. 3e). Summarizing these comparisons, covalent attachment of CAQK peptides did not change the nature of CID of the dinucleotide conjugates and indicated the prevalence of 5'-base losses and formation of w ions. The differences in the relative intensities of analogous nucleotide and conjugate fragment ions can be related to the presence of the peptide moiety, in particular in blocking the loss of free nucleobases.

Complex Ion Structures

The experimental data indicated a high overall efficiency of gas-phase cross-linking in CAQK-dXX complexes that reached 46% for (dGG + C*AQK + H)⁺ (Table 1). At the same time, the CID-MS³ data provided only limited information on the location of covalent cross-links, which, for dAA and dGG, was hampered by their near-symmetrical nature, leading to indistinguishable isomeric fragments. To improve resolution and relate the observed cross-links to the gas-phase ion structures, we carried out extensive computational analysis of the (dGG + C*AQK + H)⁺ system using Born-Oppenheimer molecular dynamics (BOMD) at the semi-empirical level of quantum theory that included all valence-electron interactions. Thermodynamic calculations were also carried out that relied on density functional theory including dispersion interactions and solvation effects in water.

Initial complex ion structures were built for two different (dGG + C*AQK + H)⁺ protomers that were protonated at the Cys N-terminus and Lys side chain, in combination with peptide carboxylate or dGG phosphate anions to produce the overall charge of + 1. Several initial structures of each protomer type were built at different relative orientations of the peptide and dGG counterparts. In total, 11 initial structures were considered as starting points for the BOMD trajectory studies (Fig. S8). BOMD was run at 410 K for 20 ps to produce a total of 11 × 20,000 = 220,000 snapshots out of which 2200 structures were extracted at 100 fs intervals, optimized by PM6-D3H4, and sorted by their secondary and supersecondary structural similarities to compact duplicates and reduce the size of the selection. In the course of BOMD runs, the conformation sampling developed the initial complexes into five groups of canonical and zwitterionic structures out of which 60–70 distinct structures were selected for full geometry optimization by DFT gradient calculations. B3LYP/6-31G(d,p) optimized structures were used for frequency calculations to provide vibrational enthalpies and entropies whereas ωB97X-D/6-31 + G(d,p) optimized structures provided electronic energies. The electronic, vibrational, and rotational terms were then combined to yield relative free energies for the complexes. In addition, solvent effects were included in single-point energy

calculations via the polarizable continuum model [45] using water as the dielectric. The calculated relative free energies are compiled in Table 2, the optimized geometries of complexes **a-d**, **m**, and **j** are given in supplementary Tables S3-S8.

The distribution of free energies for the gas-phase complexes ($\Delta G_{g,310K}$) is shown in a box-plot format (Fig. S9) for the five protomer types. The Cys-protonated canonical structures encompassed 10 conformers within a 22–237 kJ mol⁻¹ energy range and a median of 106 kJ mol⁻¹. The Lys-protonated canonical structures had 8 conformers ranging from 21 to 150 kJ mol⁻¹, with a 63 kJ mol⁻¹ median. The three zwitterionic protomer types showed distinct free-energy distributions. The (Cys⁺, Lys⁺, COO⁻)⁺ combination had 20 structures within a 5.8–170 kJ mol⁻¹ range and a 63 kJ mol⁻¹ median. The (Cys⁺, Lys⁺, phosphate⁻)⁺ protomers included ~16 structures within a narrow range of 0–69 kJ mol⁻¹ and a median of ~19 kJ mol⁻¹. The (3'-G⁺, Cys⁺, phosphate⁻)⁺ combination had 7 structures within a 35–146 kJ mol⁻¹ range with a 107 kJ mol⁻¹ median. The energies of the gas-phase complexes indicated overall preference for zwitterionic protomers of the (Cys⁺, Lys⁺, COO⁻)⁺ and (Cys⁺, Lys⁺, phosphate⁻)⁺ types.

Since the complexes were formed in solution before or during electrospray, we also strived to assess their free energies in aqueous solution. To this end, we selected structures within 66 kJ mol⁻¹ of the lowest-energy gas-phase structure for single-point energy calculations that included solvent effects. These showed that most low-energy solvated structures belonged to the zwitterionic protomer types. Solvation changed the ranking of some complexes as illustrated by the free-energy data for 16 low-energy structures **a-p** in Table 2. In particular, complex **d**, which is a (Cys⁺, Lys⁺, phosphate⁻)⁺ zwitterion, was stabilized by 19 kJ mol⁻¹ in water relative to the other complexes, becoming the global free-energy minimum. On the contrary, complexes **l** and **p** which were low free-energy zwitterions of the (Cys⁺, Lys⁺, COO⁻)⁺ type in the gas phase were relatively destabilized by 35–40 kJ mol⁻¹ in water (Table 2).

The binding between the peptide and dGG in complexes **a-p** was realized in a few general types (Fig. 4). Structures **b**, **c**, **d**, and **m** had strong hydrogen bonds between the charged phosphate and both peptide ammonium groups. In addition, the Lys NH₃⁺ developed a stabilizing hydrogen bond to the 5'-hydroxyl in these low-energy complexes (**b-d**, **m**, Fig. 4) that were particularly favored by solvation. The hydrogen bonding to 5'-OH enforced the orientation of the peptide with respect to dGG in which the N-terminal tag was in the vicinity of the 3'-guanine. Another group of (Cys⁺, Lys⁺, phosphate⁻)⁺ zwitterionic complexes, **e**, **f**, **g**, **h**, and **n**, were characterized by hydrogen bonds between the N-terminal NH₃⁺ and the 3'-guanine whereas the Lys NH₃⁺ developed hydrogen bonds to the 5'-guanine and OH group (Fig. S10). The Table 2 data indicate that this H-bonding arrangement was less energetically favorable, leading to higher free energies in the gas phase and upon solvation. The low-energy structure **a** (Fig. 4) represented a hybrid in which the N-terminal NH₃⁺

forms a hydrogen bond to the phosphate whereas the Lys NH₃⁺ bound to the 3'-guanine.

The (Cys⁺, Lys⁺, COO⁻)⁺ zwitterionic complexes, **i**, **j**, **k**, **l**, **o**, and **p** offered a variety of hydrogen-bonding patterns (Fig. 5). In some (**i**, **j**, **k**, and **o**), the COO⁻ anion participated in H-bonding to dGG, and in the others, it did not (Fig. 5). There was no obvious effect of COO⁻ hydrogen bonding on the complex free energy, as structures of both types belonged to low-energy gas-phase complexes (**j**, **k**, **l**, and **p**, Table 2). However, only complex **j** was favored by solvation to compete with **a-d** and **m**. It appears from the energy data that complexes **a-d**, **j**, and **m** were thermodynamically most favored to be formed in aqueous solution and transferred to the gas phase. Conformational changes caused by thermal motion in gas-phase ions can favor low-energy gas-phase structures (**p**) and disfavor others (**m**). These effects are discussed later in the paper in connection with BOMD trajectory analysis.

To summarize the structure analysis, the predominant interactions determining most of the low-energy dGG complex structures were due to ion-ion hydrogen bonding between the charged groups. Another important feature was hydrogen bonding of the Lys NH₃⁺ to the dGG 5'-hydroxyl that assisted the alignment of the dGG and peptide ion moieties where the N-terminus carrying the diazirine tag was close to the 3'-nucleobase. Peptide hydrogen bonding to the nucleobases, which could be considered important for DNA sequence recognition, was important in several structures **e**, **f**, **g**, **h**, **i**, **j**, **k**, **l**, **n**, **o**, and **p**. However, according to the calculated free energies, these complexes, except **j**, were disfavored by solvation with water. We presumed that these results can be generalized to the other dinucleotides and used to interpret their cross-linking.

Born-Oppenheimer Molecular Dynamics and Contact Analysis

Complexes **a-p** were analyzed for close contacts between the diazirine carbon atom (C-136) and the guanine, 2'-deoxyribose, and phosphate oxygen atoms (Fig. 6). dGG atoms carrying hydrogen (X-H) were considered as potential targets for carbene insertion, while the basic N-7 and N-3 positions in the guanines could potentially react with the transient carbene. First, we inspected close contacts within 4.5 Å in optimized structures **a-p** representing 0 K geometries. Structures **b**, **f**, **n**, **o**, and **p** displayed no 0 K contacts.

The majority of close contacts in **a**, **c**, **d**, **e**, **h**, **j**, **k**, **l**, and **m** were realized in the 3'-guanine positions N-1 (N61) and NH₂ (N62) for potential insertion to the N-H bonds, and at C-5 (C56), N-7 (N55), and C-8 (C52) for carbene addition. The 5'-nucleoside showed only two 0 K contacts, one at the NH₂ (N21 in **g**) and the other with the deoxyribose C-2 (C28 in **c**).

BOMD trajectories were run for 100 ps at 310 K for complexes **a-p**. The distribution of contacts between the incipient carbene (C136) and the guanine and deoxyribose units is

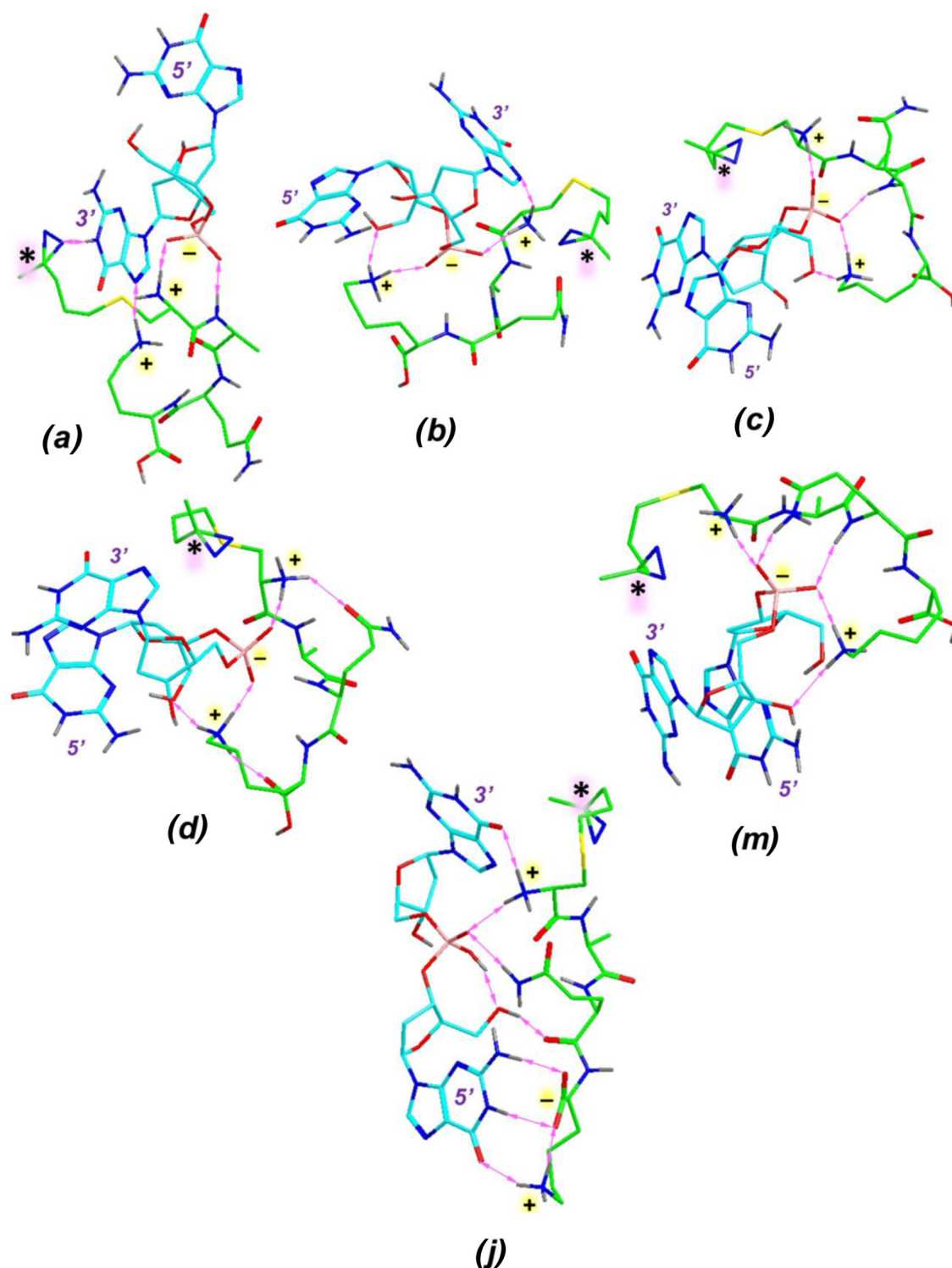


Fig. 4. ω B97X-D/6-31 + G(d,p) optimized structures of zwitterionic $(C^*AQK + dGG + H)^+$ complexes **(a)**, **(b)**, **(c)**, **(d)**, and **(m)** and **(j)**. Atom color coding is as follows: C in peptide = light green; C in dGG = cyan; O = red; N = blue; S = yellow; H = gray. Only exchangeable hydrogens are shown. Double-headed arrows indicate hydrogen bonds between the peptide and dGG. The diazine ring is labeled with an asterisk

shown in Fig. 7 and further specified for the positions in the frequently visited 3'-guanine base, such as C-8 (C52), C-4 (C53), C-5 (C56), C-2 (C59), N-7 (N55), N-3 (N57), N-1 (N61), and NH_2 (N62) that are ordered left to right in the

distribution bar graphs in Fig. 7. The overall contact frequency varied from 0.2% in complex **f** through 150% in complex **a**. The high figure exceeding 100% was due to multiple simultaneous contacts in the course of the 100,000-step trajectory. We

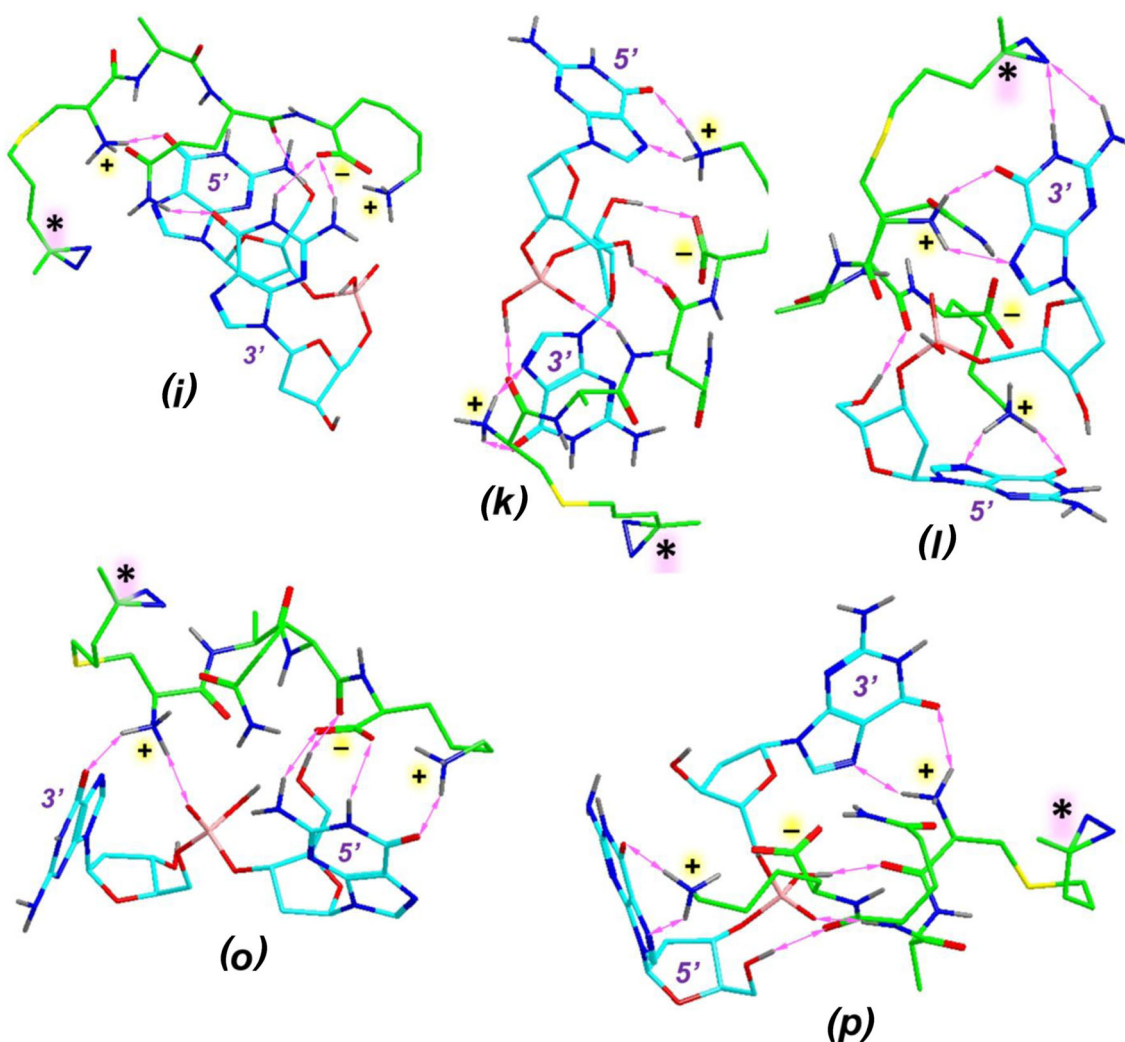


Fig. 5. ω B97X-D/6-31 + G(d,p) optimized structures of zwitterionic $(C^*AQK + dGG + H)^+$ complexes (i), (k), (l), (o), and (p) of the $(Cys^+, Lys^+, COO^-)^+$ type. Atom color coding is as in Fig. 4

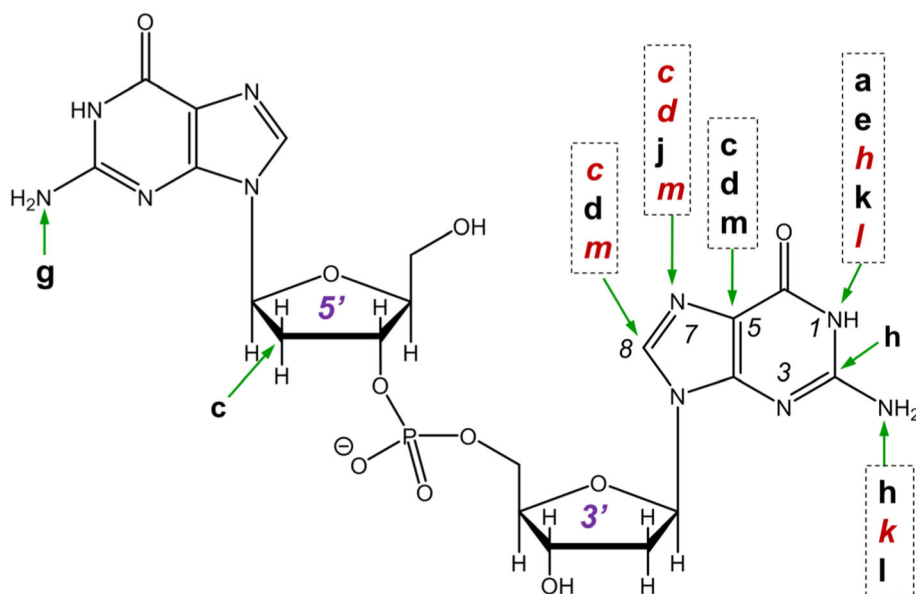


Fig. 6. Close contacts of the incipient carbene atom in optimized (0 K) structures of $(dGG + C^*AQK + H)^+$. Bold: contacts within 4.5 Å; red italics: contacts within 4.0 Å. There were no 0 K contacts in complexes (b), (l), (n), (o), and (p)

focus the discussion of the contact data on the thermodynamically most stable complexes **a-d**, **j**, **k**, **l**, **m**, and **p**. Complexes **b**, **c**, **d**, and **m** showed a high proportion of contacts at the 3'-guanine N-7 position that correlated with the proximity of N-7 and the diazine C136 in the 0 K structures (Fig. 6).

The other positions developing contacts in **b**, **c**, **d**, and **m** were C-8 and C-5 (Fig. 7). The total contact counts in **b**, **c**, **d**, and **m** were similar, ranging between 25 and 37% (Fig. 7). The contacts in complexes **c** and **m** showed similar distributions indicating similar ranges of thermal motion of the 4,4-azipentyl substituent in these complexes. Thermal motion in complexes **a**, **j**, **k**, **l**, and **p** created contacts with the N-1 and NH₂ positions in 3'-guanine that were inaccessible in **b**, **c**, **d**, and **m**. In addition, **a** was the only low-energy complex that developed contacts in the 5'-nucleoside (Fig. 7).

Reaction Mechanisms for Covalent Bond Formation

The high overall cross-linking efficiency in (dGG + C*AQK + H)⁺ (46%, Table 1) indicated that a substantial fraction of the peptide carbenes that were formed at close contact with dGG reacted by forming covalent bonds. The trajectory data indicated that the majority of contacts occurred at 3'-guanine. This result was also consistent with the cross-linking data in the GC and CG complexes that showed preferential covalent bond formation in the 3'-nucleoside. The high incidence of close contacts at the 3'-nucleobase raised the question of the reaction mechanism for the covalent bond formation. Contacts with the 3'-guanine N-1-H and NH₂ groups in complexes **j**, **k**, **l**, and **p** can proceed by the standard insertion mechanism [58]

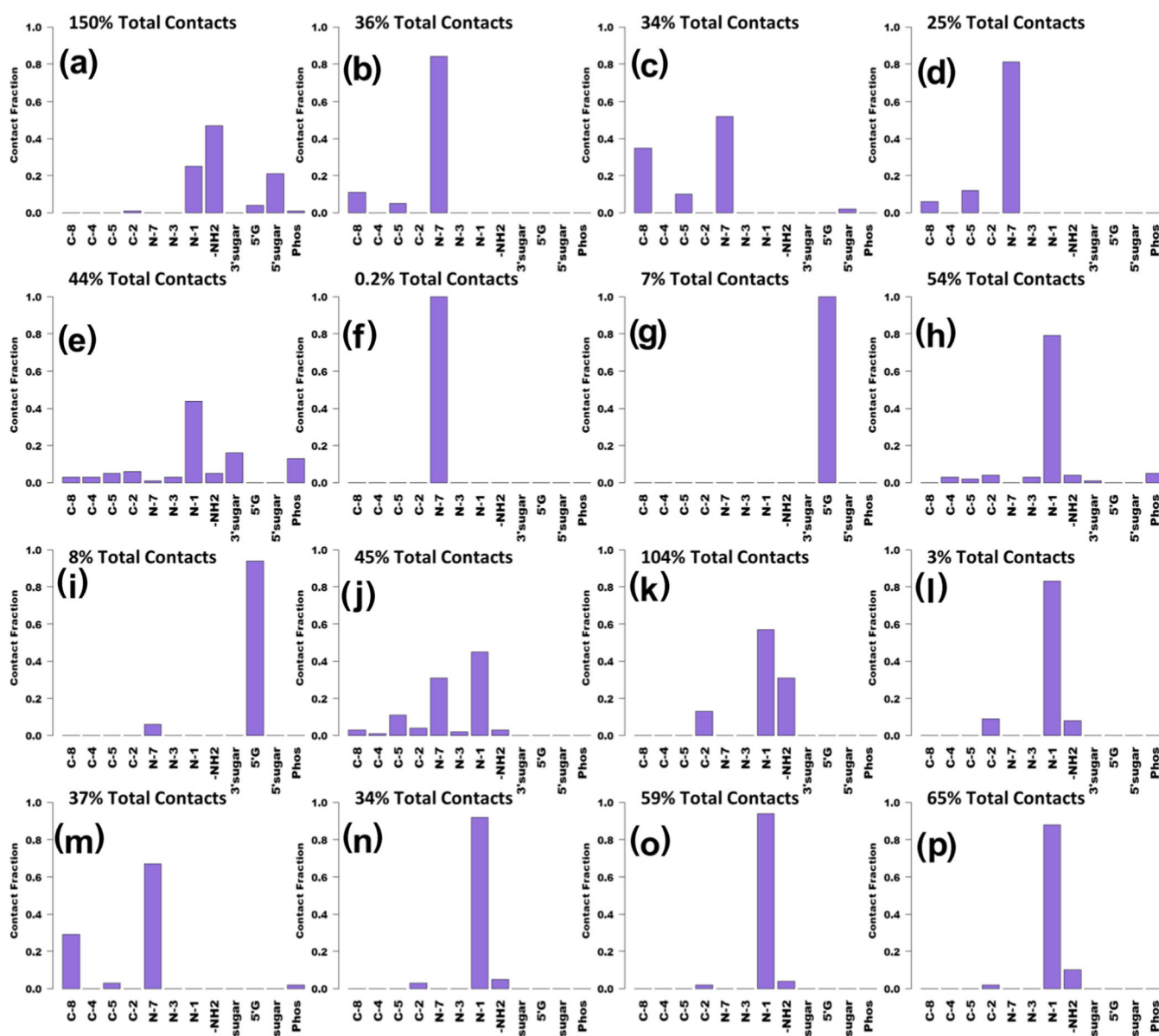


Fig. 7. Distribution of close contacts in complexes (a-p) from BOMD trajectories of (dGG + C*AQK + H)⁺ complexes. Positions C-8, C-4, C-5, C-2, N-7, N-1, and NH₂ refer to the 3'-guanine base

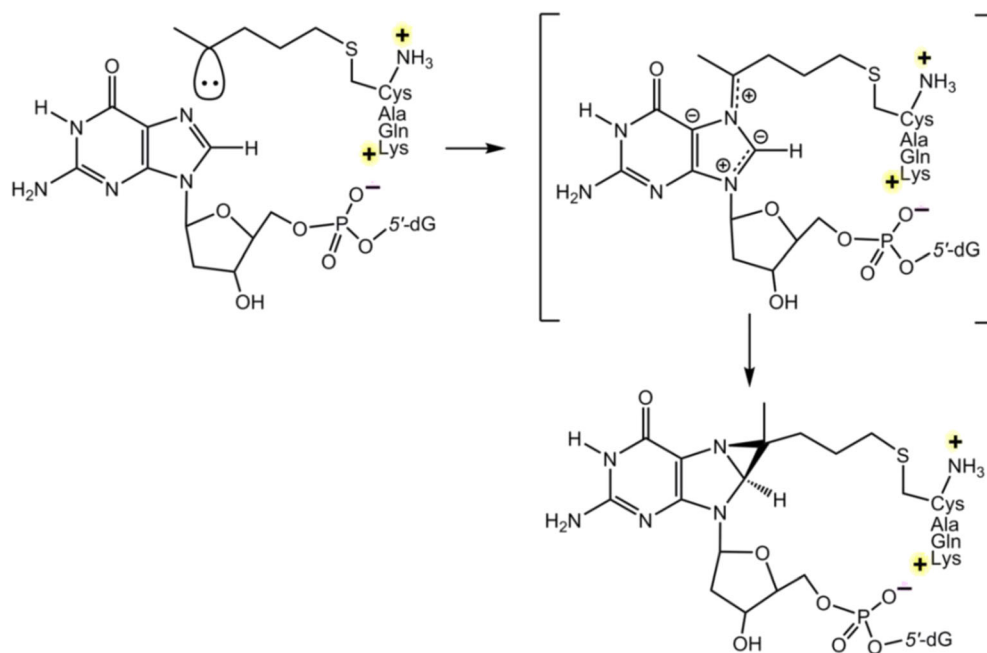
involving the N–H bonds. These insertions were expected to be 350–380 kJ mol^{−1} exothermic and thus thermodynamically driven. The exothermicity for the carbene insertion was estimated from the calculated reaction enthalpies for a model system, which was the insertion of propane-2-carbene (**2**) into the N-1-H and NH₂ bonds in 9-methyl guanine (**1**) forming 2-propylguanines **3** and **4** (Scheme S1, Supporting Information). In contrast, carbene reactions with heterocycles in general [59], and nucleosides in particular [60–62], have been studied only sparsely. Our calculations of a model propane-2-carbene addition to N-7 in 9-methylguanine revealed that the reaction was substantially exothermic, $\Delta H_{0,\text{rxn}} = -152$ to -184 kJ mol^{−1} at different levels of theory (Scheme S1, Table S2), and therefore thermodynamically favorable. The addition initially formed a dipolar intermediate in the imidazole part of the nucleobase that can be characterized as a N-7-C-8-C-9 ylide (**5**). A similar intermediate has been observed in the dichlorocarbene reaction with pyridine [59]. Intermediate **5** can further react by closing a three-membered ring. The ring closure forming a N-7-C-8 fused azirine was calculated to be 66–81 kJ mol^{−1} exothermic, leading to a stable cross-linked product **6**. The activation energy for the ring closure, 112–119 kJ mol^{−1} for **TS1** relative to **5** (Scheme S1) can be readily provided by the exothermic addition.

We envisaged an analogous mechanism for the carbene attack on guanine in the (dGG + C*AQK – N₂ + H)⁺ complex as sketched in Scheme 2. Because of the substantial exothermicity of these reactions, we did not expect the 3'- and 5'-guanine rings to display distinct reactivity toward the carbene intermediate. Rather, we argue that the preferential attack at the 3'-nucleobase was guided by the conformations of the complexes that stemmed from the hydrogen bonding interactions between the polar phosphate, ammonium, and 5'-OH groups.

At the same time, the tight hydrogen bonding in the complexes allowed for only a limited conformational motion of the diazirine-carrying side chain whereas the core of the complex was not affected. This interpretation was consistent with the contact analysis of the 0 K structures and 310 K trajectories. For example, the low-energy complex **d** showed 0 K contacts with C-5, N-7, and C-8 in 3'-guanine (Fig. 6), which developed into the most frequent contacts with 3'-N-7 at 310 K whereas the other positions were visited much less (Fig. 7).

Effects of Peptide Complexation on the Dinucleotide Structure

The optimized structures of low-energy complexes indicated strong dipolar interactions between dGG and C*AQK that were likely to also exist in complexes involving the other combinations of peptides and dinucleotides. To specify the effect of the peptide ion binding, we undertook a conformational search of dGG anions and compared their low-energy structures with those in the complexes. Three different initial monomer conformations were selected as starting points, which were of stacked, extended, and S-shape types. After conformation sorting, 13 non-degenerate structures were submitted to DFT geometry optimizations and frequency calculations to obtain the relative free energies. Based on this analysis, the lowest free-energy structures were of the stacked type (dGG1 and dGG2, Fig. S11), whereas the semi-folded S-shaped dGG3 and extended dGG4 were 28 and 31 kJ mol^{−1} higher in energy. The low-energy dGG structures displayed a stacked motif that was similar to that found in complexes **c**, **d**, and **m** (Fig. 4) where the exposed phosphate anion was solvated by the peptide dication. The hydrogen bond between the



Scheme 2. Reaction of carbene with 3'-guanine in the (dGG + C*AQK – N₂ + H)⁺ complex

phosphate and 5-OH in **dGG1** and **dGG2** was replaced by an H-bond to the lysine NH_3^+ group in complexes **c**, **d**, and **m**, which endowed them with additional stability and directed the lysine C-terminus toward the 5'-nucleobase. However, the energy data suggested that coordination with the peptide ion can also be realized with other low-energy complex structures (**a**, **b**, **j**) in which guanine stacking was absent. This led us to the conclusion that non-bonding interactions in these peptide-dinucleotide complexes could produce low free-energy protomers and conformers displaying substantial structural variability while conserving or disrupting nucleobase stacking interactions in the dinucleotides.

Conclusions

Singly charged complexes of diazirine-tagged peptides CAQK with several DNA dinucleotides were generated in the gas phase. Peptide carbene intermediates formed by photodissociation were found to undergo efficient covalent cross-linking that largely occurred in the 3'-nucleoside. Electronic structure and Born-Oppenheimer molecular dynamics calculations of the $(\text{dGG} + \text{C}^*\text{AQK} + \text{H})^+$ complex revealed a variety of low-energy conformers of zwitterionic types having doubly protonated peptides along with deprotonated dinucleotide phosphate or peptide carboxyl groups. A common feature revealed by the cross-linking data in accord with the structure analysis was that the peptide and dinucleotide preferred orientations allowing close approach of the diazirine tag at the N-terminal cysteine to the 3'-nucleoside. Thermal motion at 310 K in the complexes resulted in an extended range of close contacts between the tag and the nucleotide, but the core structure of the complex, as determined by hydrogen bonding, was not disrupted.

Acknowledgements

This research has received funding from the National Science Foundation Division of Chemistry (Grants CHE-1661815 and CHE-1624430). F.T. thanks the Klaus and Mary Ann Saegebarth Endowment for support. Y. L. thanks Dr. Emilie Viglino for generous help with peptide synthesis.

References

- von Hippel, P.H.: From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 79–105 (2007)
- Raindlova, V., Pohl, R., Hocek, M.: Synthesis of aldehyde-linked nucleotides and DNA and their bioconjugations with lysine and peptides through reductive amination. *Chem. Eur. J.* **18**, 4080–4087 (2012)
- Carette, L.L.G., Morii, T., Maddar, A.: Toxicity inspired cross-linking for probing DNA-peptide interactions. *Bioconjug. Chem.* **24**, 2008–2014 (2013)
- Flett, F.J., Walton, J.G.A., Mackay, C.G., Interthal, H.: Click chemistry generated model DNA-peptide heteroconjugates as tools for mass spectrometry. *Anal. Chem.* **87**, 9595–9599 (2015)
- Wickramaratne, S., Boldry, E.J., Buehler, C., Wang, Y.-C., Distefano, M.D., Tretyakova, N.D.: Error-prone translesion synthesis past DNA-peptide cross-links conjugated to the major groove of DNA via C5 of thymidine. *J. Biol. Chem.* **290**, 775–787 (2015)
- Ming, X., Groehler 4th, A., Michaelson-Richie, E.D., Villalta, P.W., Campbell, C., Tretyakova, N.Y.: Mass spectrometry based proteomics study of cisplatin-Induced DNA-protein cross-linking in human fibrosarcoma (HT1080) cells. *Chem. Res. Toxicol.* **30**, 980–995 (2017)
- Buxton, K.E., Kennedy-Darling, J., Shortreed, M.R., Zaidan, N.Z., Olivier, M., Scaif, M., Sridharan, R., Smith, L.M.: Elucidating protein-DNA interactions in human aliphoid chromatin via hybridization capture and mass spectrometry. *J. Proteome Res.* **16**, 3433–3442 (2017)
- Weir Lipton, M.S., Fuciarelli, A.L., Springer, D.L., Hofstadler, S.A., Edmonds, C.G.: Analysis of radiation induced nucleobase-peptide cross-links by electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **11**, 1673–1676 (1997)
- Golden, M.C., Resing, K.A., Collins, B.D., Willis, M.C., Koch, T.H.: Mass spectral characterization of a protein-nucleic acid photocross-link. *Protein Sci.* **8**, 2806–2812 (1999)
- Wong, D.L., Reich, N.O.: Identification of tyrosine 204 as the photocross-linking site in the DNA-*EcoRI* DNA methyltransferase complex by electrospray ionization mass spectrometry. *Biochemistry.* **39**, 15410–15417 (2000)
- Rieger, R.A., McTigue, M.M., Kycia, J.H., Gerchman, S.E., Grollman, A.P., Iden, C.R.: Characterization of a cross-linked DNA-endonuclease VIII repair complex by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **11**, 505–515 (2000)
- Steen, H., Petersen, J., Mann, M., Jensen, O.: Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: Tryptophans 54 and 88 of *E. coli* SSB cross-link to DNA. *Protein Sci.* **10**, 1989–2001 (2001)
- Wagenknecht, H.-A., Rajski, S.R., Pascaly, M., Stemp, E.D.A., Barton, J.K.: Direct observation of radical intermediates in protein-dependent DNA charge transport. *J. Am. Chem. Soc.* **123**(19), 4400–4407 (2001)
- Doneanu, C.E., Gafken, P.R., Bennett, S.E., Barofsky, D.F.: Mass spectrometry of UV-cross-linked protein-nucleic acid complexes: identification of amino acid residues in the single-stranded DNA-binding domain of human replication protein A. *Anal. Chem.* **76**, 5667–5676 (2004)
- Pourshahian, S., Limbach, P.A.: Application of fractional mass for the identification of peptide-oligonucleotide cross-links by mass spectrometry. *J. Mass Spectrom.* **43**, 1081–1088 (2008)
- Rosenfeld, K.K., Ziv, T., Goldin, S., Glaser, F., Manor, H.: Mapping of DNA binding sites in the tetrahymena telomerase holoenzyme proteins by UV cross-linking and mass spectrometry. *J. Mol. Biol.* **410**, 77–92 (2011)
- Steen, H., Jensen, O.N.: Analysis of protein-nucleic acid interactions by photochemical cross-linking and mass spectrometry. *Mass Spectrom. Rev.* **21**, 163–182 (2002)
- Tachenay, A., Dieu, M., Amould, T., Renard, P.: Mass spectrometry-based identification of proteins interacting with nucleic acids. *J. Proteome.* **94**, 89–109 (2013)
- Laughlin, S., Wilson, W.D.: May the best molecule win: competition ESI mass spectrometry. *Int. J. Mol. Sci.* **16**, 24506–24531 (2015)
- Groehler, A., Degner, A., Tretyakova, N.Y.: Mass spectrometry-based tools to characterize DNA-protein cross-linking by bis-electrophiles. *Basic Clin. Pharmacol. Toxicol.* **121**(S3), 63–77 (2017)
- Lin, S., Cotter, R.J., Woods, A.S.: Detection of non-covalent interaction of single and double stranded DNA with peptides by MALDI-TOF. *Proteins Struct. Funct. Genet. Suppl.* **2**, 12–21 (1998)
- Veenstra, T.D.: Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes. *Biochem. Biophys. Res. Commun.* **257**, 1–5 (1999)
- Beck, J.L., Colgrave, M.L., Ralph, S.F., Sheil, M.M.: Electrospray ionization mass spectrometry of oligonucleotide complexes with drugs, metals, and proteins. *Mass Spectrom. Rev.* **20**, 61–87 (2001)
- Alves, S., Woods, A., Tabet, J.C.: Charge state effect on the zwitterion influence on stability of non-covalent interaction of single-stranded DNA with peptides. *J. Mass Spectrom.* **42**, 1613–1622 (2007)
- Alves, S., Woods, A., Devolve, A., Tabet, J.C.: Influence of salt bridge interactions on the gas-phase stability of DNA/peptide complexes. *Int. J. Mass Spectrom.* **278**, 122–128 (2008)
- Brahim, B., Tabet, J.C., Alves, S.: Positive and negative ion mode comparison for the determination of DNA/peptide noncovalent binding sites through the formation of "three-body" noncovalent fragment ions. *Eur. J. Mass Spectrom.* **24**, 168–177 (2018)
- Liu, Y., Ramey, Z., Tureček, F.: Non-covalent interactions of a neuro-protective peptide revealed by photodissociative cross-linking in the gas phase. *Chem. Eur. J.* **24**, 9259–9263 (2018)

28. Mann, A.P., Scodeller, P., Hussain, S., Joo, J., Kwon, E., Braun, G.B., Molder, T., She, Z.-G., Kotamraju, V.R., Ranscht, B., Krajewski, S., Teesalu, T., Bhatia, S., Sailor, M.J., Ruoslahti, E.: A peptide for targeted, systemic delivery of imaging and therapeutic compounds into acute brain injuries. *Nat. Commun.* **7**, 11980 (2016)
29. Huang, R.C.C., Bonner, J.: Histone, a suppressor of chromosomal RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 1216–1222 (1962)
30. Bannister, A.J., Kouzarides, T.: Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011)
31. Strahl, B.D., Allis, C.D.: The language of covalent histone modifications. *Nature*. **403**(6765), 41–45 (2000)
32. Jenuwein, T., Allis, C.D.: Translating the histone code. *Science*. **293**(5532), 1074–1080 (2001)
33. Pepin, R., Shaffer, C.J., Tureček, F.: Position-tunable diazirine tags for peptide-peptide ion cross-linking in the gas-phase. *J. Mass Spectrom.* **52**, 557–560 (2017)
34. Das, J.: Aliphatic diazirines as photoaffinity probes for proteins: recent developments. *Chem. Rev.* **111**, 4405–4417 (2011)
35. Nguyen, H.T.H., Andrikopoulos, P.C., Rulišek, L., Shaffer, C.J., Tureček, F.: Photodissociative cross linking of noncovalent peptide-peptide ion complexes in the gas phase. *J. Am. Soc. Mass Spectrom.* **29**, 1706–1720 (2018)
36. Shaffer, C.J., Andrikopoulos, P.C., Řezáč, J., Rulišek, L., Tureček, F.: Efficient covalent bond formation in gas-phase peptide-peptide ion complexes with the photoleucine stapler. *J. Am. Soc. Mass Spectrom.* **27**, 633–645 (2016)
37. Berendsen, H.J., Postma, J.V., van Gunsteren, W.F., DiNola, A.R.H.J., Haak, J.R.: Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984)
38. Stewart, J.J.P.: Optimization of parameters for semi-empirical methods V: modification of NDDO approximations and applications to 70 elements. *J. Mol. Model.* **13**, 1173–1213 (2007)
39. Řezáč, J., Fanfrlík, J., Salahub, D., Hobza, P.: Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. *J. Chem. Theor. Comput.* **5**, 1749–1760 (2009)
40. Stewart, J.J.P.: MOPAC 16. Stewart Computational Chemistry, Colorado Springs, CO (2016)
41. Řezáč, J.: Cuby: an integrative framework for computational chemistry. *J. Comput. Chem.* **37**, 1230–1237 (2016)
42. Řezáč, J.: Cuby—ruby framework for computational chemistry, version 4, <http://cuby4.molecular.cz>
43. Becke, A.D.: Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*. **38**, 3098–3100 (1988)
44. Chai, J.D., Head-Gordon, M.: Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008)
45. Tomasi, J., Mennucci, B., Cammi, R.: Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3093 (2005)
46. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci, B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery Jr., J.A., Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N., Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M., Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K., Farkas, O., Foresman, J.B., Fox, D.J.: Gaussian 16, revision A01. Gaussian, Inc., Wallingford CT (2016)
47. Marek, A., Tureček, F.: Collision-induced dissociation of diazirine-labeled peptide ions. Evidence for Brønsted-acid assisted elimination of nitrogen. *J. Am. Soc. Mass Spectrom.* **25**, 778–789 (2014)
48. Seburg, R.A., McMahon, R.J.: Photochemistry of matrix-isolated diazoethane and methyl diazirine: ethylidene trapping? *J. Am. Chem. Soc.* **114**, 7183–7189 (1992)
49. Jackson, J.E., Soundararajan, N., White, W., Liu, M.T.H., Bonneau, R., Platz, M.S.: Measurement of the absolute rate of 1,2-hydrogen migration in benzylchlorocarbene. *J. Am. Chem. Soc.* **111**, 6874–6875 (1989)
50. Pezacki, J.P., Couture, P., Dunn, J.A., Warkentin, J.: Rate constant for 1,2-hydrogen migration in cyclohexylidene and in substituted cyclohexylidenes. *J. Org. Chem.* **64**, 4456–4464 (1999)
51. Stevens, I.D.R., Liu, M.T.H., Soundararajan, N., Paik, N.: The barrier for 1,2-hydrogen shift in dialkyl carbenes. *Tetrahedron Lett.* **30**, 481–484 (1989)
52. Murray, K.K.: DNA sequencing by mass spectrometry. *J. Mass Spectrom.* **31**, 1203–1215 (1996)
53. McLuckey, S.A., Van Berkel, G.J., Glish, G.L.: Tandem mass spectrometry of small, multiply charged oligonucleotides. *J. Am. Soc. Mass Spectrom.* **3**, 60–70 (1992)
54. Roepstorff, P., Fohlman, J.: Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **11**, 601 (1984)
55. Biemann, K.: Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* **193**, 886–887 (1990)
56. Wu, J., McLuckey, S.A.: Gas-phase fragmentation of oligonucleotide ions. *Int. J. Mass Spectrom.* **237**, 197–241 (2004)
57. Schurch, S.: Characterization of nucleic acids by tandem mass spectrometry—the second decade (2004–2013): from DNA to RNA and modified sequences. *Mass Spectrom. Rev.* **35**, 483–523 (2016)
58. Liu, M.T.H.: Chemistry of Diazirines, Vol. II, CRC Press, Boca Raton, FL (1987)
59. Tippmann, E.M., Platz, M.S., Svir, I.B., Klymenko, O.V.: Evidence for specific solvation of two halocarbene amides. *J. Am. Chem. Soc.* **126**, 5750–5762 (2004)
60. Mohan, U., Burai, R., McNaughton, B.R.: Reactivity between acetone and single-stranded DNA containing a 5'-capped 2'-fluoro-N7-methyl guanine. *Tetrahedron Lett.* **55**, 3358–3360 (2014)
61. Nowak, I., Robins, M.J.: Synthesis of 3'-deoxynucleosides with 2-oxabicyclo[3.1.0]hexane sugar moieties: addition of difluorocarbene to a 3',4'-unsaturated uridine derivative and 1,2-dihydrofurans derived from D- and L-xylose. *J. Org. Chem.* **72**, 3319–3325 (2007)
62. Jerbi, J., Springborg, M.: Reactivity descriptors for DNA bases and the methylation of cytosine. *Int. J. Quantum Chem.* **118**, e25538 (2018)