

Molecular Recognition

Delineation of G-Quadruplex Alkylation Sites Mediated by 3,6-Bis(1-methyl-4-vinylpyridinium iodide)carbazole-Aniline Mustard Conjugates

Chien-Han Chen,^[a] Tsung-Hao Hu,^[a] Tzu-Chiao Huang,^[a] Ying-Lan Chen,^[c, d] Yet-Ran Chen,^[c] Chien-Chung Cheng,^{*,[b]} and Chao-Tsen Chen^{*,[a]}

Abstract: A new G-quadruplex (G-4)-directing alkylating agent BMVC-C3M was designed and synthesized to integrate 3,6-bis(1-methyl-4-vinylpyridinium iodide)carbazole (BMVC) with aniline mustard. Various telomeric G-4 structures (hybrid-2 type and antiparallel) and an oncogene promoter, c-MYC (parallel), were constructed to react with BMVC-C3M, yielding 35% alkylation yield toward G-4 DNA over other DNA categories (<6%) and high specificity under competition conditions. Analysis of the intact alkylation adducts by electrospray ionization mass spectroscopy (ESI-MS) revealed the stepwise DNA alkylation mechanism of aniline mustard for the first time. Furthermore, the monoalkylation sites and

intrastrand cross-linking sites were determined and found to be dependent on G-4 topology based on the results of footprinting analysis in combination with mass spectroscopic techniques and in silico modeling. The results indicated that BMVC-C3M preferentially alkylated at A15 (H26), G12 (H24), and G2 (c-MYC), respectively, as monoalkylated adducts and formed A15-C3M-A21 (H26), G12-C3M-G4 (H24), and G2-C3M-G4/G17 (c-MYC), respectively, as cross-linked dialkylated adducts. Collectively, the stability and site-selective cross-linking capacity of BMVC-C3M provides a credible tool for the structural and functional characterization of G-4 DNAs in biological systems.

Introduction

Guanine-rich nucleic acid sequences are capable of folding into polymorphic G-quadruplex (G-4) structures with four-stranded supramolecular architectures stabilized by Hoogsteen hydrogen-bonding of G-tracts and coordination with monovalent cations such as potassium or sodium.^[1] The quadruplex-forming sequences can fold into different structures, depend-

ing on strand orientation and glycosidic conformation. The equilibrium of these various structures might regulate cell cycle-synchronized DNA replication and environmentally regulated gene expression.^[2] G-4 DNAs are involved in telomere maintenance,^[3] DNA recombination,^[4] and splicing,^[5] and have thus been the focus of intense research into their function as regulatory elements and therapeutic targets in oncology.^[6]

A variety of small synthetic molecules have been designed and synthesized to stabilize secondary structures of G-4 DNA through noncovalent interactions. Many of them possess high affinity and good selectivity for G-4 structures and have potent antiproliferative effects on various immortalized cell lines.^[6b,7] Although these small molecules hold great promise for the development of new anticancer therapies, reversible binding to G-4 DNA makes it possible for them to elude targeted therapeutics. To augment G-4 binding affinity, it is possible to covalently tether the G-4 binders to G-4 DNA. Various reactive agents equipped with G-4-directing ligands synchronously improve DNA cross-linking specificity and promote antiproliferative activity. For example, a series of G-4-targeting agents tethered with platinum complexes were developed to metalate the purine N7 in d[AG₃(T₂AG₃)₃] (H22)-quadruplexes.^[8] Naphthalene diimides equipped with either activatable quinone-methide precursors or reactive oxirane were shown to alkylate H22 with 10 and 16% yields, respectively.^[9] The alkylation sites were identified at the adenines on the loop with an oxirane warhead. A photoinduced cross-linking agent was recently developed by conjugating photoreactive groups to a G-4-binding

[a] C.-H. Chen, T.-H. Hu, T.-C. Huang, Prof. C.-T. Chen
Department of Chemistry, National Taiwan University
No. 1, Sec. 4, Roosevelt Road
Taipei, 10617 Taiwan (R.O.C.)
Fax: (+886) 2-23636359
E-mail: chenct@ntu.edu.tw

[b] Prof. C.-C. Cheng
Department of Applied Chemistry
Chia-Yi University, No. 300, Xuefu Road, Chiayi City
60004 Taiwan (R.O.C.)
Fax: (+886) 5-2717901
E-mail: cccheng@mail.ncyu.edu.tw

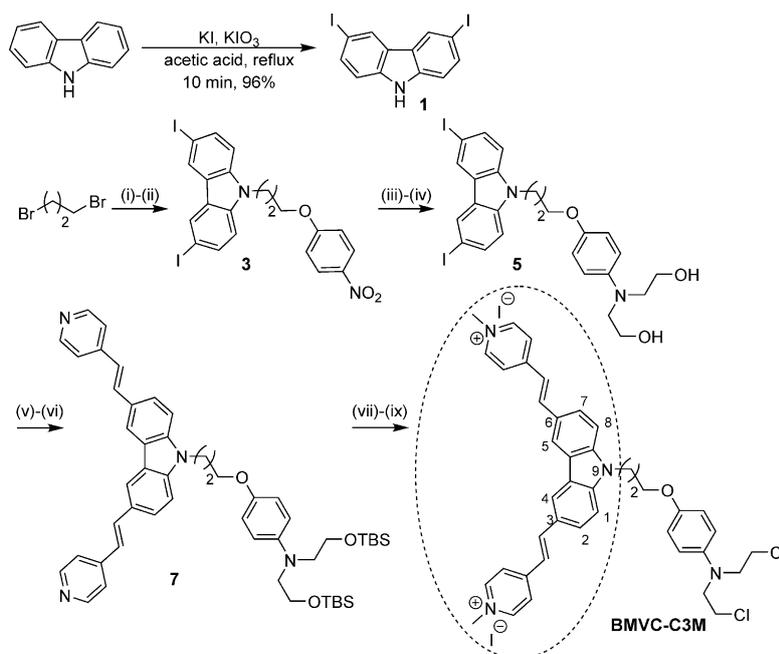
[c] Y.-L. Chen, Prof. Y.-R. Chen
Agricultural Biotechnology Research Center
Academia Sinica, No. 128, Sec. 2, Academia Road
Taipei, 11529 Taiwan (R.O.C.)

[d] Y.-L. Chen
Institute of Plant Biology and Department of Life Science
National Taiwan University, No. 1, Sec. 4, Roosevelt Road
Taipei, 10617 Taiwan (R.O.C.)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201502595>.

motif to provide G-4-topology-dependent photoreactions and high cross-linking efficiency.^[10] To protect DNA from UVA damage,^[11] a visible light stimulus was used to activate G-4 selective alkylation.^[12] A green light stimulus-triggered G-4 ligand displayed an alkylating efficiency of up to 64% and selectively targeted loop T residues in H22.^[12c] In an attempt to overcome the low reactivity and generate highly stable hybridized adducts under mild conditions, a novel G-4 alkylator was synthesized by integrating pyridostatin (PDS) with a therapeutic alkylating warhead, chlorambucil (Chl), and was so-named PDS-Chl.^[13] PDS-Chl could alkylate both h-Telo (H22) and c-MYC, located at the promoter region responsible for cell-cycle control,^[14] at low concentration (0.2–5 μM) with higher reactivity compared with other categories of DNA. The alkylation sites are highly dependent on G-4-forming sequences. Adenine was alkylated in H22, whereas guanine (G) and adenine (A) were alkylated in c-MYC. Furthermore, PDS-Chl selectively impairs growth of cells with deficiencies in nucleotide excision repair (NER) but shows no sensitivity for the repair gene *BRCA2*, suggesting it preferentially generates intrastrand cross-links through selective G-4 recognition. This result provides new insight into potential therapeutic treatments for genetically impaired NER cancers.

Nonetheless, the characteristics and specificities of G-4-directing alkylating agents are determined by various factors such as the nature of the alkylating warheads, the choice of G-4-directing ligands, and the spacers. Alkylating reactivity, specificity, and topology-dependent site-selectivity could be greatly modulated by adjusting how these three crucial structural parameters are integrated in the alkylating probe design. One of the many reported telomere stabilizers, 3,6-bis(1-methyl-4-vinylpyridinium iodide)carbazole (BMVC), binds the human telomeric quadruplex with a nanomolar dissociation constant^[15] and selectively enters cancerous nuclei.^[16] In addition, bifunctional alkylating warheads could improve alkylation efficiency and may provide a therapeutic benefit. In this study, we used BMVC as the G-4-directing ligand for integration with aniline mustard to search for optimized G-4 alkylating agents. To preserve the binding ability of BMVC, an alkylating warhead was incorporated at the N9-position of BMVC.^[17] The BMVC-C3M (also abbreviated as C3M in some figures) were designed and synthesized by integrating aniline mustard with BMVC through flexible linkers (CH_2)₃ to provide a suitable reaction trajectory toward polymorphic G-4 DNAs (Scheme 1). The elaborated in-



Scheme 1. Reagents and conditions: i) 4-iodoaniline, K_2CO_3 , acetone, reflux, 12 h, 38%; ii) 1,3-bis(bromomethyl)propane, K_2CO_3 , acetone, reflux, 12 h, 62%; iii) SnCl_2 , ethanol, reflux, 9 h, 52%; iv) 10% aq. NaOH, 2-chloroethanol, 90 °C, 12 h, 69%; v) TBSCl, imidazole, CH_2Cl_2 , RT, 8 h, 74%; vi) 4-vinylpyridine, $\text{Pd}(\text{OAc})_2$, $\text{P}(\text{o-tol})_3$, K_2CO_3 , DMF, 80 °C, 9 h, 58%; vii) TBAF, THF, RT, 6 h, 85%; viii) PPh_3 resin, CCl_4 , CH_2Cl_2 , 0 °C to RT 2 h, then reflux 2 h, 66%; ix) CH_3I , acetone, RT, 12 h, 98%.

vestigation of the BMVC-C3M alkylation reaction profiles, and its mechanism of action on several different G-4-forming sequences and DNA structures in vitro, is anticipated to inform the effective design of G-4-directing alkylation agents.

Results and Discussion

Synthetic strategies and synthesis

Two strategies may be used to synthesize BMVC-C3M. One is a concise modular process of synthesizing BMVC, linkers, and aniline mustard derivatives, followed by final assembly. However, target synthesis by using this method was hampered by the rather inert reactivity of the phenolic group of aniline mustard toward nucleophilic substitution as well as in the Mitsunobu reaction that is used to incorporate the linker. Therefore, the consecutive synthetic strategy shown in Scheme 1 was adopted. First, carbazole was diiodinated at the 3,6-position as described previously^[18] to provide 1 for implementation of a vinylpyridine group through Heck coupling at a later stage. Commercially available 1,3-dibromopropane reacted with *p*-nitrophenol to give monosubstituted *p*-nitrophenyl ether, leaving another bromo group free to incorporate 1 through nucleophilic substitution to afford 3. Reduction of the nitro group on 3 followed by alkylation with chloroethanol resulted in the formation of diethanolaniline derivative 5. The hydroxyl groups were protected as the *tert*-butyldimethylsilyl ether before the compound was submitted to Heck coupling conditions with 4-vinylpyridine to afford 7. Without protection, the Heck coupling reaction was sluggish. Subsequent deprotection with

tetra-*n*-butylammonium fluoride (TBAF), followed by chlorination of the resultant hydroxyl groups with polystyrene-bound triphenylphosphine, and methylation of the pyridine moiety with methyl iodide, afforded the target compound BMVC-C3M in 55% yield in three steps. Detailed synthetic procedures and structural characterization are provided in the Supporting Information.

G-4 DNAs used to evaluate the alkylation profile of BMVC-C3M

A diverse array of structures can be formed by G-4 DNAs,^[19] however, no rules have been developed to predict the folding propensity or topology of G-4 structures based only on their sequence.^[20] Given that each G-4 topology adopts a unique three-dimensional spatial arrangement, small molecules may selectively recognize one form over others or even convert one form into another. For instance, telomestatin induces the formation of basket-type G-4 structures from hybrid-type G-4 in the telomeric region.^[21] The human telomeric sequence H22 is commonly used to evaluate the behavior of alkylation probes. This sequence folds into an antiparallel basket-type G-4 structure in solution in the presence of Na⁺, as revealed by NMR spectroscopic analysis,^[22] and forms a parallel propeller structure in crystals in the presence of K⁺.^[23] The co-existence of multiple topological isomers complicates the structural analysis of H22 in K⁺-containing buffer.^[24] The [T₂A(G₃T₂A)₃G₃T₂](H26) sequence adopts a definite hybrid-2 G-4 structure in the presence of K⁺ in 75% of the population, as revealed by high-field solution NMR spectroscopy;^[25] the remaining 25% consists of other undetermined topological forms.

We chose H26 in K⁺-containing buffer to investigate the alkylation profiles of BMVC-C3M. To test the G-4 structure-discriminating ability of BMVC-C3M, [T₂AG₃(T₂AG₃)₃](H24) in Na⁺-containing buffer sharing the same sequence with H22 with two extra thymine bases at the 5'-end, was used for validation as well as for comparison. Besides telomeric G-4-forming sequences, the c-MYC silencer element [TGAG₃TG₃TAG₃TG₃TA₂], adopting a parallel G-4 topology in K⁺-containing buffer,^[26] was also used to profile the alkylation pattern of BMVC-C3M. By investigating the alkylation profiles of all three distinct G-4 topological structures, we hope to facilitate the design of alkylating agents that effectively stabilize telomeric sequences, preventing over-proliferation and advancing anticancer drug development to rectify genetic malfunctions. Hence, H26 and c-MYC were folded into a hybrid-2-type and parallel G-4 structure in K⁺-containing buffer, respectively, whereas H24 was folded into an anti-parallel G-4 structure in Na⁺-containing buffer.

Probing structural changes and stability of G-4s in the presence of BMVC-C3M based on circular dichroism and melting point measurements

Circular dichroism (CD) spectra and melting point determinations are commonly used to probe structural changes and to

evaluate the stability of G-4 upon ligand treatment. Thus, CD and the melting temperature were used to evaluate binding of BMVC-C3M and G-4. Two representative G-4-forming sequences were chosen, including the human telomeric sequence H26 and the gene promoter c-MYC. H26 produced a characteristic hybrid-2 structure with positive signals at 265 and 290 nm in the CD spectra (see the Supporting Information, Figure S1 a),^[27] whereas c-MYC produced a positive signal at 265 nm and a negative signal at 245 nm, corresponding to a parallel structure (Figure S1 b). BMVC and BMVC-C3M were then incubated with DNA at 37 °C for 20 h, followed by CD measurement. In the presence of H26, BMVC-C3M exhibited a similar CD response, with a significant increase at 265 nm and a decrease at both 245 and 295 nm (Figure S1 a). The observed CD profile is similar to that induced by BMVC, but with a smaller amplitude, suggesting that G-4 undergoes a conformational adjustment in the presence of BMVC-C3M, similar to that of BMVC.^[28] In contrast to the remarkable signal transition observed in H26, minor spectral variations were detected when BMVC or BMVC-C3M was incubated with c-MYC (Figure S1 b). As expected, the structural modification at the N9-position of BMVC seems to create a minor perturbation of the interactions with H26 and c-MYC.

The quadruplex melting curves monitored at 265 nm showed an approximate 20 °C increase in the melting temperature of the folded H26, suggesting that BMVC-C3M is able to stabilize the G-4 structure but to a lesser extent than BMVC, representing a nearly 30 °C increase (Figure S1 c). This result indicates that the structural modification of BMVC influences the stabilization ability towards H26. Given that c-MYC quadruplex is a relatively stable parallel structure, with a melting temperature of 90.8 °C in 150 mM KCl solution, more than 95 °C melting temperature was observed in the presence of either BMVC or BMVC-C3M (Figure S1 d).

Photophysical behavior and the intracellular localization of BMVC-C3M

Photophysical spectroscopy is a useful tool to delineate the modes of binding based on spectral changes. Thus, the photophysical features of BMVC-C3M were examined to investigate how the structural modification influences absorption, fluorescence, and binding behavior upon incubation with H26, H24, and c-MYC (Figure S2). The absorption spectra showed a bathochromic shift and hypochromic effect upon addition of either H26, H24, or c-MYC to BMVC-C3M at a 1:1 ratio. The absorption profiles are very similar to that of BMVC, indicating that BMVC-C3M acts in a similar fashion, possibly by external stacking with the G-4 structure.^[29]

BMVC is a turn-on G-4 fluorescent probe upon binding.^[29–30] BMVC-C3M exhibited a similar fluorescence profile but with smaller amplitude in the presence of H26, H24, or c-MYC; the fluorescence color of the solution changes from orange-red to yellow with a green tint, accordingly. No particular specificity towards three different G-4-forming sequences was observed. The fluorescence enhancement can be attributed to restriction of the linker segmental mobility and intercalation of the BMVC

fluorophore.^[31] BMVC-C3M exhibited the fluorescence with 10-fold enhancement for H26, 14-fold for H24, and 12-fold for c-MYC. The smaller increase in BMVC-C3M versus BMVC fluorescence may be due to the change in G-4 stacking caused by the reactive groups. Attempts to use UV/Vis or fluorescence changes to determine the binding constant for BMVC-C3M failed because alkylation reactions occur during the course of the titration, thus complicating the absorption and fluorescence spectra.

Another remarkable trait of BMVC is its nuclear permeability and strong fluorescence.^[32] To investigate whether BMVC-C3M retains its nuclear-penetrating ability, confocal fluorescence microscopy was used to observe MCF7 cells incubated with BMVC and BMVC-C3M and co-stained with nuclear-specific DAPI. The fluorescence images showed much higher fluorescence in the nuclei than in the cytoplasm (Figure S3). The results suggest that appreciable amounts of BMVC-C3M enter the nucleus and may react with G-4 DNAs, although they may also react with other cellular nucleophilic components such as proteins and low-molecular-weight thiols.

Alkylating reactivity, selectivity, and specificity toward different types of DNA

After confirming that the directing and stabilizing abilities of the BMVC-C3M agents toward telomeric G-4 DNA (H26) and c-MYC were preserved based on CD analysis and on quadruplex-melting estimation, we evaluated the ability of BMVC-C3M to alkylate G-4. To this end, the alkylating reactivity of the recognition unit of BMVC and the alkylating warhead Chl toward H26 were assessed and compared with the alkylating ability of BMVC-C3M (Figure S4). The covalently hybridized adducts were resolved by high-resolution denaturing polyacrylamide gel electrophoresis (PAGE). The slower migrating bands corresponding to the alkylated adducts (ALK) were separated from non-reacted DNA (NR). The results showed that the formation of retarded bands was observed in the reaction of BMVC-C3M (Lanes 2). This preliminary result suggests that BMVC-C3M would indeed alkylate G-4 DNAs. In contrast, no alkylated adducts were detected in the presence of 2 μM BMVC or Chl (Lanes 3 and 4), even when the concentration of Chl was increased to 50 μM (Lane 5). The results suggest either that Chl could not effectively alkylate the G-4 sequence or that the lack of a positive charge on Chl may prohibit the resolution of Chl-alkylated adducts from non-reacted DNA by PAGE. If the latter is the case, hydrolysis of the Chl-DNA mixture would facilitate the identification of Chl-DNA

adducts. Accordingly, hydrolysis of the mixture was attempted and the results are described in the footprinting analysis section. Furthermore, dose-dependence experiments indicated that a 2:1 ratio of BMVC-C3M and H26 resulted in saturated conversion (Figure S5). Increasing amounts of BMVC-C3M beyond two equivalents did not produce a higher alkylation yield, perhaps because the G-4 structure had reached its capacity and could not accommodate more BMVC-C3M. Therefore, initial assessment of the ability of BMVC-C3M to alkylate fluorescein (FAM)-labeled DNA was achieved by variously timed reactions at a fixed 2:1 ratio.

The alkylating reactivity and selectivity of BMVC-C3M were then examined against different DNA structures adopted by scrambled single-stranded R26 (ss), and double-stranded H26 sequences (ds), in addition to the folded H26 (G-4). As revealed in Figure 1 a, the use of BMVC-C3M led to highly selective DNA alkylation. In the case of G-4 DNA, the alkylated product could be detected after 20 min in 7% alkylation yield after the addition of BMVC-C3M (Lane 1). A single alkylated band was observed even after longer incubation times. In contrast, obscure alkylated bands in less than 6% alkylation yield were obtained

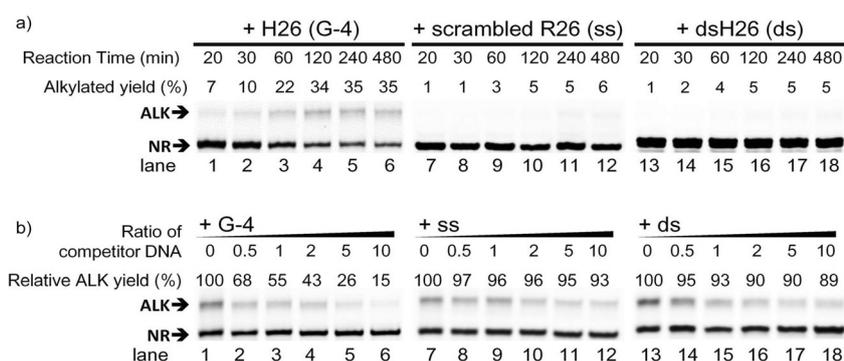


Figure 1. Evaluation of BMVC-C3M-mediated DNA alkylation. a) BMVC-C3M (2 μM) was incubated with 5'-FAM H26 (1 μM), scrambled ssR26 (1 μM), or dsH26 (1 μM) DNA for varying time intervals. DNA alkylation yield was calculated by integrating the fluorescence intensities of the corresponding bands. b) BMVC-C3M (2 μM) was incubated with 5'-FAM H26 (G-4*; 1 μM) in the presence of different molar ratios (0, 0.5, 1, 2, 5, and 10) of unlabeled H26 (G-4), scrambled ssR26 (ss), or dsH26 (ds) for 20 h at 37 °C. Alkylated DNA (ALK) and nonreacted DNA (NR) were resolved by using 20% denaturing PAGE. Three runs were performed for each alkylation experiment. Relative yields were calculated; adduct formation in the absence of competitors was considered 100% in each experiment.

in the case of ss and ds DNA. BMVC-C3M displayed higher reactivity (34% in 4 h) toward G-4 DNA than the other types of DNA (< 6%). These results indicate that the reactivity of BMVC-C3M is regulated by the structure of the target DNA. H24 and c-MYC were also employed to react with BMVC-C3M. In the case of H24, BMVC-C3M showed a high alkylation yield of 45% (Figure S6), whereas BMVC-C3M/c-MYC adducts were unstable under the denaturing PAGE experimental conditions, resulting in the fragmentations and rendering the identification of the alkylation sites infeasible.^[33]

To further investigate the alkylating specificities of BMVC-C3M in the presence of various types of DNA, competition experiments were conducted. A constant amount of BMVC-C3M

(2 μM) was incubated with a fixed amount (1 μM) of 5'-FAM H26 in the presence of various concentrations of unlabeled G-4, scrambled ssDNA, and dsDNA for 20 h at 37 °C. The experimental results obtained for BMVC-C3M are shown in Figure 1b. Only unlabeled G-4 DNA could effectively compete with labeled G-4 DNA for adduct formation with BMVC-C3M. The alkylated band faded in the presence of higher concentrations of cold G-4 DNA (Lanes 1–6). In the presence of a 10-fold excess of the G-4 competitor, the alkylated band was barely detectable (15% relative alkylation yield; Lane 6), but the formation of alkylated products was only slightly affected upon addition of ssR26 (Lanes 7–12) and was moderately affected upon addition of dsH26 (Lanes 13–18), indicating that BMVC-C3M exhibited definitive directing ability with structural specificity. Notably, at a 1:1 molar ratio of G-4 competitor to labeled G-4 DNA (Lane 3), the relative alkylation yield was approximately 50%, with no preference for unlabeled G-4 or 5'-FAM-labeled G-4 DNA, indicating that the FAM label does not interfere with BMVC-C3M alkylation.

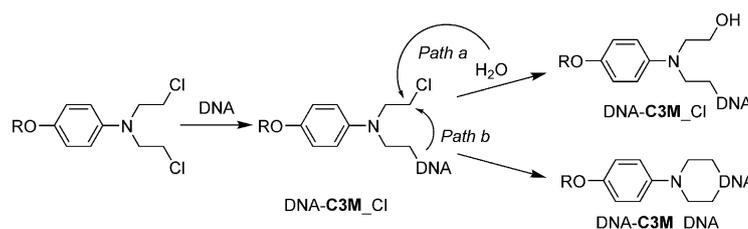
ESI-MS analysis of alkylated adducts

Chl is a bifunctional alkylating agent that is known to form covalent adducts with double-helical DNA, especially with the N7 of guanines, through nucleophilic reaction either within the same strand (intrastrand) or between the two opposite DNA strands (interstrand).^[34] However, the Balasubramanian group recently demonstrated that tethering Chl with a G4-targeting ligand exhibited a new mechanism of action that preferentially formed intrastrand cross-links with c-MYC in vitro as well as in vivo through G4 structural recognition.^[13] Their finding paves the way for new interventions in diseases associated with genetically impaired NER. It is thus important to exploit the nature of G-4-alkylated adducts formed by BMVC-C3M. To address this aim, ESI in combination with matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) MS were conducted to elucidate the identity of the resulting alkylated DNA adducts and possibly unveil the mechanism of alkylation at the molecular level. Selected sections of the ESI-MS spectra of BMVC-C3M reacting with either folded H26, H24, or c-MYC are shown in Figure 2. The peaks of several alkylated DNA adducts and unreacted DNAs carrying different numbers of negative charges were observed. Notably, the intensity of the peaks corresponding to unreacted DNA was relatively weak in the case of c-MYC (Figure 2c), suggesting that BMVC-C3M possesses extremely good alkylating ability toward c-MYC.

The most abundant peaks corresponding to $[\text{H26-C3M}]^{7-}$, $[\text{H24-C3M}]^{7-}$, and $[\text{c-MYC-C3M}]^{8-}$ were further analyzed to identify possible adducts. Three sets corresponding to three plausible BMVC-C3M-DNA adducts were found, and the calculated m/z values of these peaks shown in the insets were obtained according to the formula: $[\text{mass of DNA adducts} - (\text{number of charge} + 2)] / \text{number of charge}$ (Figure 2 inset).^[35] The peak centered at m/z 1256.80 (H26), 1169.79 (H24), or 950.56 (c-MYC) is attributed to the

adduct with one chloro-group of BMVC-C3M substituted by DNA and another hydrolyzed (abbreviated as DNA-C3M_OH), whereas the peak centered at m/z 1259.37 (H26), 1172.36 (H24), or 952.81 (c-MYC) is assigned to the adduct with one DNA attached and one chloro-group preserved (abbreviated as DNA-C3M_Cl). The rather small peaks at m/z 1254.37 (H26), 1167.36 (H24), or 948.44 (c-MYC) resulted from the intrastrand DNA cross-linked by BMVC-C3M (abbreviated as DNA-C3M_DNA). No higher molecular weight corresponding to the interstrand cross-linked DNA adduct was observed. The results of the intrastrand DNA cross-linking experiment are consistent with the analysis of the enzymatic digestion product of PDS-Chl against c-MYC by LC-MS reported by the Balasubramanian group.^[13] To confirm that the observed distribution of adducts is universal for differently charged species, analysis of signals for $[\text{H24-C3M}]^{9-}$ was conducted and a similar distribution of these three alkylated adducts was observed (Figure S7), indicating that the number of the charge does not affect the adduct distribution. To our knowledge, this is the first observation of preserving chloro-group species of aniline mustard as intact G-4 adducts. We reasoned that the direct analysis of the intact DNAs after alkylation led to these findings, which are critical to the interpretation of how BMVC-C3M alkylates DNAs.

To confirm that no interstrand cross-linked DNA was formed, higher concentrations of BMVC-C3M (5 μM) were added to H24 (1 μM) and subjected to MALDI-TOF mass measurement. The results revealed that only over-alkylated DNA adducts were detected and that no interstrand cross-link product was present (Figure S8). In addition, no products with more than three alkylations were found, and the relative intensity declined with increasing number of modifications. Given that we used a 1:1 ratio of DNA/BMVC-C3M for ESI-MS spectroscopic studies, it is reasonable that we only observed 1:1 adduct formation. On the basis of mass data, we believe the intrastrand cross-link alkylation is the result of a two-step process (Scheme 2). BMVC-C3M was first directed toward G-4 DNA through the G-4 directing group and was alkylated by the proximal nucleobase (DNA-C3M_Cl). Another chloro-group was subsequently either hydrolyzed by water (DNA-C3M_OH) or substituted by the other nucleobase on the same G-4 preferentially to yield the cross-linking of an intramolecular G-4 structure (DNA-C3M_DNA).



Scheme 2. Proposed mechanism for stepwise DNA dialkylation with BMVC-C3M.

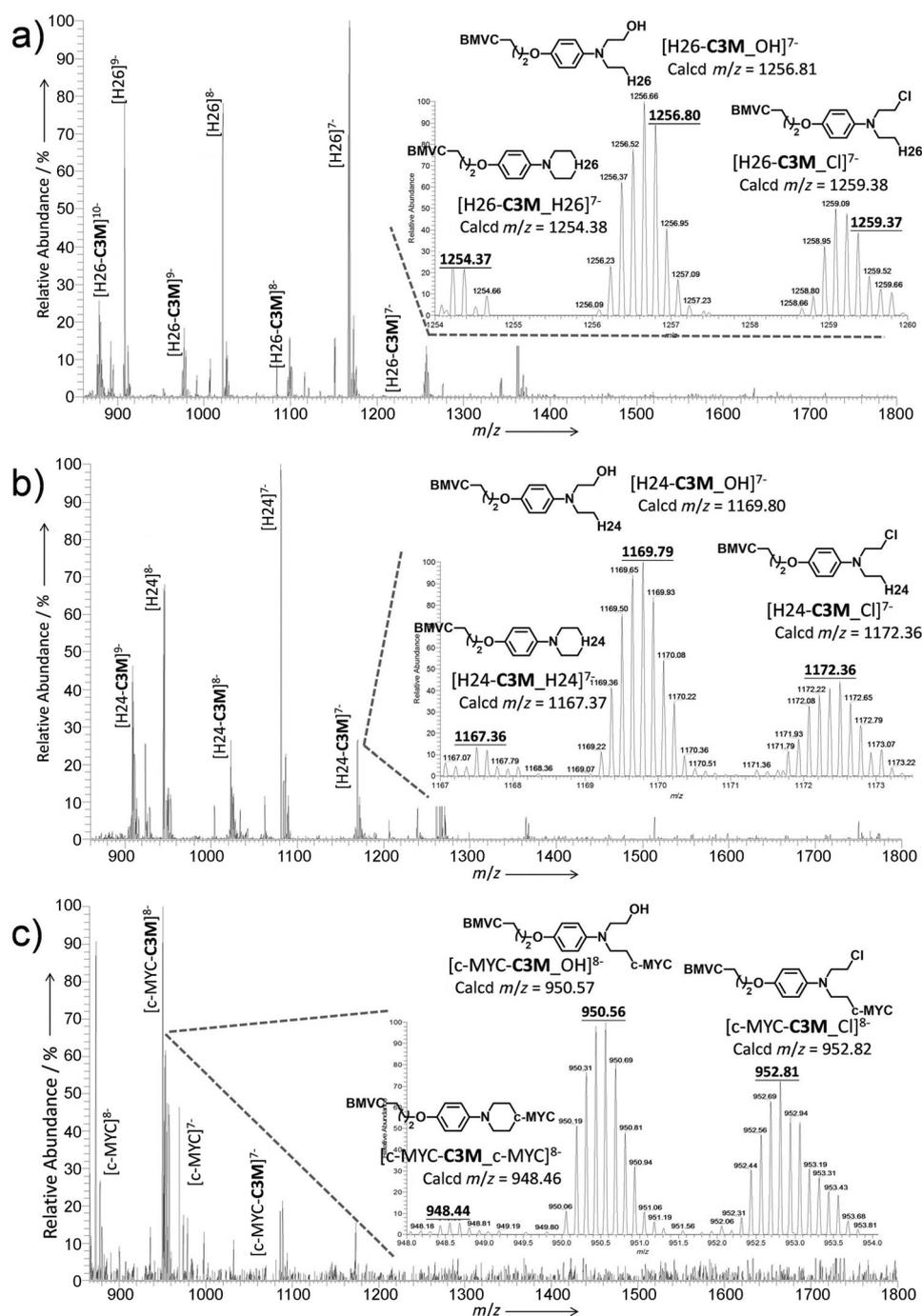


Figure 2. ESI-MS spectrum of samples obtained by reacting BMVC-C3M (1 μM) with folded DNA a) H26, b) H24, or c) c-MYC (1 μM) for 20 h. The insets show the expanded signals for $[\text{H26-C3M}]^{7-}$, $[\text{H24-C3M}]^{7-}$, or $[\text{c-MYC-C3M}]^{8-}$, which are indicated by the dashed line to the corresponding peaks, respectively. Detected m/z signals related to calculated m/z values are underlined.

Identification of nucleotides involved in the alkylation by LC-MS analysis

The nucleotides involved in the alkylation remain to be elucidated. To this end, BMVC-C3M (1 μM) was reacted with H26, H24, and c-MYC (1 μM), respectively, and the reacted mixtures (50 μL , 20 $\text{ng } \mu\text{L}^{-1}$) were digested for 12 h according to a previously reported protocol^[36] to release the monoalkylated and in-

trastrand cross-linking adducts as nucleoside and dinucleoside remnants, and subjected to LC-MS analyses. The results based on the selected ion chromatograms (SICs) of monoalkylated adducts are summarized in Table S1. Regardless of the types of topologies used, only A- or G-alkylated adducts were evident; no T-alkylated adduct was observed. The intrinsic low nucleophilicity of thymine may account for these observations. Further analyses of the adducts showed that BMVC-C3M reacted preferentially with the A of human telomeric sequences and with the G of c-MYC. The results indicated that the topology influences the site of alkylation.

The most abundant dinucleoside adducts of BMVC-C3M were then examined after screening all possible dialkylated products according to previous reports.^[37] The SICs of BMVC-C3M, signaling the presence of the dinucleoside adducts, are shown in Figure 3. Two types of alkylation modes were possible. BMVC-C3M bridged two adenines (dA-C3M-A, one 2'-deoxyribose of 2'-deoxyadenosine was eliminated) (Figure 3 a) in the case of reacting with H26, and two guanines (G-C3M-G, two 2'-deoxyribose of alkylated 2'-deoxyguanosine was eliminated) in reacting with either H24 or c-MYC (Figure 3 b,c). The four individual peaks in the given m/z assigned to the G-nucleosides were eluted at different times (4.73, 5.54, 6.00, and 6.10 min, respectively), indicating that they originated from four different G-alkylation sites. Moreover, the relatively large amount of dialkylated adducts of c-MYC suggested that the formation of intrastrand cross-linking adducts is rather easy compared with other topologies of G-4 (Figure 3 c). In addition, bis-adducts with both A and G were detected in the digestion of the alkylated mixture of BMVC-C3M with c-MYC (Figure S9).

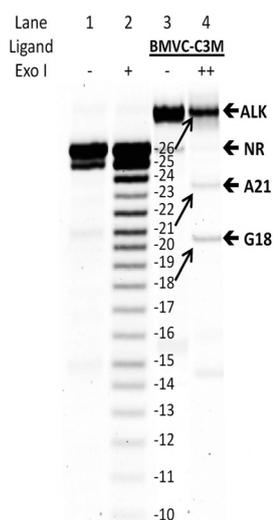


Figure 5. Alkylation site analysis of 5'-FAM H26 (1 μM) in K^+ -containing buffer with BMVC-C3M (1 μM). 3'-Exonuclease digestion of nonalkylated (Lanes 1 and 2) and alkylated H26 (Lanes 3 and 4) were analyzed by 20% denaturing PAGE. Enzyme units (μL): – for 0; + for 0.001; and ++ for 0.08. Arrows indicate the upward shift of bands in the alkylated sample, which provided a direct indication of the alkylated position. The hydrolysis band below 9 was too faint to be detected and no alkylation site was found in this region.

Considering that the product of BMVC-C3M-treated DNA undergoes hydrolysis during purification, the amount of purified alkylated product was reduced. As a result, the alkylation site at A15 was undetectable and more stable alkylation sites at G18 and A21 were detected instead (Lane 4). Notably, the alkylation sites of BMVC-C3M were mainly located at the adenine nucleobases, and not guanine, which possesses the highest intrinsic nucleophilicity among the four deoxyribonucleosides.^[39] These results suggest that adenine nucleobases might be more accessible to the incoming electrophiles.

Footprinting analysis of 5'-FAM H24 (1 μM) treated with BMVC-C3M is shown in Figure 4b. Compared with the alkylated adducts of H26, the alkylated H24 adducts were mostly hydrolyzed in a hot alkaline solution. The abasic sites of BMVC-C3M–H24 adducts were located at the Gs. G12 was the main alkylation site (Lane 3) at the concentration. Upon increasing concentration, additional G4- and G6-alkylated adducts were observed (Lane 5). Compared with H26, H24 in Na^+ -containing buffer tended to adopt an antiparallel G-4 structure with a looser quartet network and relatively naked Gs, rendering these Gs susceptible to attack by the electrophilic warhead. Notably, the alkylation profile obtained from our chemical footprinting analysis of alkylated H24 is similar to that revealed by the enzymatic footprinting analysis in the reaction of H22 with Pt-quinacridine (Pt-MPQ), which displayed three main alkylation sites on G2, G10, and G22 in H22, corresponding to G4, G12, and G24 in H24.^[8a] G4 and G12 in H24 were detected in the reaction with BMVC-C3M, whereas G24 was not found in our study, implying that stacking of the BMVC moiety on the antiparallel structure prevents the alkylation of G24.

As previously reported,^[40] the sequences of G-4 are not only expected to influence G-4 topologies but also the alkaline cat-

ions present in the solution. To evaluate whether cations influence the alkylation profiles of BMVC-C3M, footprinting analyses of alkylated H26- and H24-adducts of reacting BMVC-C3M were conducted by swapping the cations present in the two sequences. The alkylation profiles of H26 in Na^+ -containing buffer were similar to those of H24 in the same cation-containing buffer (G12 for BMVC-C3M, Figure S11 a). Notably, the reference marker from G17 to G24 (Lanes 1 and 2) became blurred and the alkylated products were not fully hydrolyzed, indicating that other minor G-4 topologies complicate the appearance of the bands. These results suggested that H26, with two additional nucleotides, assumed one major topology similar to that of H24, but was still capable of assuming other topologies. In the footprinting analysis of H24 in K^+ -containing buffer, the reference marker became very difficult to identify, rendering the determination of the alkylation sites infeasible (Figure S11 b). Presumably, H24 adopted dynamic polymorphic structures in the presence of K^+ . These results confirmed that the topologies and alkylation profiles are highly cation-dependent.

The DNA footprinting analysis of c-MYC (1 μM) treated with BMVC-C3M (0.5, 1, and 2 μM), shown in Figure 4c, revealed that G2 near the 5'-end of the c-MYC sequence is the primary alkylation site at the various concentrations tested. Presumably, the relatively exposed nucleophilic functionality of G2, which does not participate in the hydrogen-bonding network, is susceptible to attack by the reactive electrophile.

Delineation of the specific alkylation sites in combination with footprinting and LC-MS data

Based on the results of chemical and enzymatic footprinting, and on LC-MS data, the recognition and alkylation profile of BMVC-C3M toward G-4 topological structures can be delineated with precision. The major alkylated adducts detected by LC-MS and the authentic alkylation sites determined by footprinting analyses are summarized in Table 1 for convenient comparison. In some cases, when the experimental data were not sufficient to firmly underpin the intrastrand cross-linking sites, *in silico* modeling was used to rationally extrapolate the possible cross-linking site. The model assumed that when one reactive group of the aniline mustard was alkylated by the first nucleobase, the mobility of another reactive warhead was restricted. The nucleophilicity and accessibility of the nucleobases are not the only factors that affect the reacted nucleobases in the disubstituted reaction; the geometric orientation and spatial arrangement are also critical elements that determine whether the reacting partners form intrastrand cross-linked adducts. To identify the dialkylation sites, the distance between candidate nucleobases in the dialkylating reaction was measured according to the folded G-4 structures stored at Protein Data Bank [code 2JPZ (H26), 143D (H24), and 1XAV (c-MYC)] by using Discovery Studio 4.0 Client (Figure S12). We hypothesized that the major alkylation site would be the first alkylation in the dialkylation reaction. To minimize structural perturbation introduced in the G-4 structure, the ideal distance between the two react-

Table 1. Summary of major monoalkylated and dialkylated adducts detected by footprinting and SICs of LC-MS analysis.

	H26 (Hybrid-2)	H24 (Antiparallel)	c-MYC (Parallel)
Alkylation efficiency [%]	35	45	nd ^[a]
Monoalkylation	A(1) G(2)	A(1) G(2) ^[c]	G
LC-MS results ^[b]	Intrastrand cross-link	G–G(1) A–A(2)	G–G(1) A–G(2)
Footprinting results ^[d]	Alkylation sites	A15/G18/A21 ^[e]	G2
Comprehensive results	Intrastrand cross-linking sites	G4–G12(1) A3–A15(2)	G2–G4 G2–G17(1) G2–A3(2)

[a] No yield is given for c-MYC because of the instability of the alkylation adduct (nd: not determined). [b] The difference in relative abundance by LC-MS was less than 10-fold; thus, the major and minor adducts are indicated as 1 or 2 in parentheses, respectively. [c] A-alkylated (area: 6.9×10^5) and G-alkylated adducts (area: 4.0×10^5) were observed with the same order of magnitude. [d] Bold letters correspond to major alkylation sites. The other minor alkylation sites are indicated in regular font. [e] Chemical and enzymatic footprinting results are listed. A15 was determined by chemical footprinting; G18 and A21 were determined by enzymatic footprinting.

ed nucleobases should be within 7.2 Å to match the maximum stretch of the two reactive chloroethyl groups of ChI.^[41]

In the alkylation reaction of BMVC-C3M with H26, the LC-MS data revealed that A and G are responsible for the alkylation, and the footprinting results confirmed that the plausible alkylation sites were located at A15, A21, and G18. These may be the product of the monoalkylated or cross-linked dialkylated adduct. The SICs of the LC-MS data provided a further clue that BMVC-C3M bridged two adenines (dA–C3M–A), clearly ruling out the possibility of G18 being involved in the cross-linking reaction. Examination of hybrid-2-type topology adopt-

ed by H26 revealed that three adenines (A3, A15, and A21) near the 5'-end are within a reasonable distance to undergo the intramolecular cross-linking reactions (Figure 6a). Given that A15 was the major band that emerged in the footprinting analysis, it is reasonable to assume that A15 is the major monoalkylation site as well as the first alkylation site of the intramolecular cross-link. The second alkylation may occur either at A3 or A21, which are within a suitable distance (3.6–6.8 Å) to undergo the cross-linking reactions (Figure S12a). However, only A21 and not A3 was observed in the footprinting analysis. Thus, A21 appears to be the plausible nucleobase to proceed toward the second substituted reaction, yielding the A15–C3M–A21 cross-link. In contrast, the reactivity of styrene oxide towards all four deoxyribonucleosides decreased in the order: dG > dC > dA ≫ dT.^[39] The nucleophilicity of dA ranked 3rd among the four nucleosides, and it exhibited greater susceptibility to nucleophilic attack in our experiments. Presumably, the accessibility of the intrinsically more reactive Gs is blocked by the hydrogen-bonding network of G-4

DNA in the hybrid-type topology, and thus alkylation occurring at the loop adenine nucleobases is kinetically favored.^[42] Similar results were also observed for NDI-oxirane and PDS-ChI when a similar hybrid-type structure of G-4 but a different DNA sequence was used.^[9d,13] With the alkylation sites determined precisely, the binding of BMVC-C3M to H26 can be depicted. BMVC-C3M seems to stack preferentially on the 5'-side of the external quartets, allowing the alkylation reactions with the surrounding nucleobases to occur (Figure 6a). Given that A3 is located above the proposed binding position and gained additional stacking stabilization on BMVC-C3M, its spatial ar-

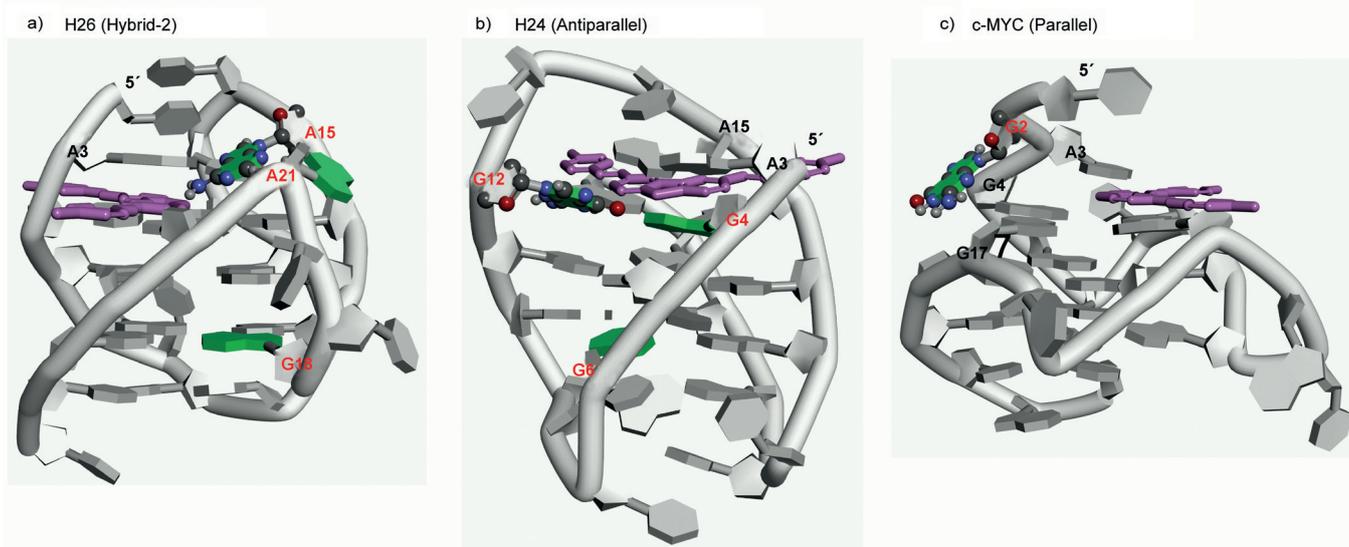


Figure 6. Major G-4 topological structures of human telomere a) H26 in K^+ , b) H24 in Na^+ , and oncogene promoter c) c-MYC in K^+ -containing buffers. Alkylation sites determined by footprinting analysis are highlighted in green and labeled in red. Primary alkylated nucleobases are indicated by ball-and-stick atoms (dark: carbon, blue: nitrogen, red: oxygen, and gray: hydrogen). Other plausible alkylated nucleobases are labeled in black. Backbone and the other nucleobases are shown in gray. Plausible binding position of the BMVC moiety on BMVC-C3M (purple) is shown as stick atoms.

rangement might prohibit the alkylation. This experimental observation is in accordance with the end-binding model of BMVC towards telomeric G-4 DNA, which was suggested by the results of a previous molecular simulation.^[15]

In the alkylation reaction of BMVC-C3M with H24, which adopted an antiparallel topology under the experimental conditions, only G-alkylation sites at G4 (minor), G6 (minor), and G12 (major) were observed by chemical footprinting analyses, whereas the A and G sites were the primary and secondary alkylation sites, respectively, according to the LC-MS data. The apparent discrepancy might be ascribed to the greater stability of the A-alkylated adducts compared with the G-alkylated adducts under chemical hydrolysis, causing only G-alkylated adducts to be detected in the footprinting experiment. Two intra-strand cross-linking adducts of BMVC-C3M, G-C3M-G (major) and A-C3M-A (minor), were detected in SIC monitoring. Given that G12 is the primary alkylation site, only Gs within 7.2 Å of G12 can react with G12 to form a cross-link adduct. Examination of the topology of H24 revealed that G12 is flanked by G4 and G24, located on the same quartet plane. G12-G4 and G12-G24 were thus candidate cross-link adducts; however, only G4 was observed in the footprinting analysis, ruling out the formation of G12-G24. In addition, G4 with a suitable spatial arrangement (2.8–5.6 Å) readily proceeded to the second substituted reaction to give the G4-C3M-G12 cross-link after alkylation at G12 (Figure S12b). Although no abasic sites on A were detected in the footprinting analysis, A3 and A15 (located near the upper G-quartet plane) were within reacting distance as estimated by *in silico* modeling, and these were speculated to form the cross-linked adduct A3-C3M-A15 (Figure S12c). The alkylation profile suggested that BMVC-C3M tends to stack over two adjacent guanines (G16 and G24) on the upper quartet near the diagonal loop of the 5'-end (Figure 6b), which might mask the alkylating ability of G16 and G24, and thus allow G4 and G12 to be alkylated effectively.

In the reaction of BMVC-C3M with *c*-MYC, the LC-MS data showed that the major monoalkylated adduct is at the G site; G2 was the only G site observed upon footprinting. Given this information, the major alkylation site was assigned to G2 of the 5'-flanking region of the parallel *c*-MYC (Figure 6c). Furthermore, dA-C3M-G and G-C3M-G dinucleoside remnants were detected after analyzing the nuclease-digested products, and G-C3M-G was present at a higher relative abundance. The recent NMR structure of a 2:1 complex between monosubstituted quindoline and the parallel *c*-MYC quadruplex indicated that both 5'- and 3'-sides of the external quartets were accessible and able to accommodate the bound molecule.^[43] According to the end-binding model described above, the BMVC moiety of BMVC-C3M stacked over two adjacent guanines (G8 and G13) on the 5'-side of the external quartets, rendering the alkylating warhead prone to covalent trapping by the proximal G2. Another reactive warhead on G2-substituted BMVC-C3M was allowed to approach the four possible Gs, including G4, G8, G13, and G17, on the 5'-side of the external quartets to achieve a G-G cross-link. *In silico* modeling showed that only G4 and G17 were within 7.2 Å of G2 and are not involved in stacking of the BMVC moiety. Thus, the major intrastrand

cross-link might be G2-C3M-G4 and G2-C3M-G17 for the G-G cross-link, yielding 4.3–7.1 and 4.6–6.7 Å linkages (Figure S12d). We also examined candidate As (A3, A12, A21, and A22) on *c*-MYC to identify which is involved in forming the minor dinucleoside remnant A-C3M-G (Figure S12e). However, the bridge lengths between all four adenines and G2 were longer than 7.2 Å. Of these, the shortest bridge was between A3 and G2 (7.9 Å), suggesting that A3 might be the partner to undergo the second substitution, yielding the A-C3M-G cross-link. Given that A3 is located on the highly flexible 5'-flanking region, the longer bridge length (> 7.2 Å) between G2 and A3 indicates that adjustment of the 5'-flanking region of the parallel *c*-MYC was required to achieve the A3-C3M-G2 cross-link. Notably, both dA-C3M-G (minor) and G-C3M-G (major) dinucleoside remnants were detected in the digested products of BMVC-C3M treated with *c*-MYC, which displayed different alkylation site selectivity from the PDS-Chl treated with *c*-MYC. We reason that different G-4-directing groups might reside at different locations at *c*-MYC, and thus disposed the alkylation warhead towards different reaction zones.

Conclusion

We have designed and synthesized a new G-4 alkylator, namely BMVC-C3M, featuring a bifunctional alkylating warhead and fluorescent G-4-specific binder with a suitable linker length. Our results indicate that this probe can alkylate G-4 DNAs with good reactivity and high specificity. Analysis of the intact alkylation adducts by ESI-MS revealed the stepwise DNA alkylation mechanism of aniline mustard for the first time. Based on the results of footprinting analysis in combination with mass spectroscopic techniques and *in silico* modeling, the monoalkylation sites and intrastrand cross-linked sites were determined and found to be dependent on G-4 topology. The alkylation profile observed for each G-4 topological structure is the overall results of the specific recognition of BMVC-C3M towards each G-4 structure as well as the nucleophilicity, accessibility, geometric orientation, and the spatial arrangement of the nucleobases in the vicinity of the warhead after the recognition. Collectively speaking, our probe could provide a credible tool for the structural and functional characterization of the prevalence of G-4 DNAs in biological systems. Furthermore, this versatile G-4-selective alkylator with fluorescent readout will be used to explore putative quadruplex sequences by genomic DNA profiling.

Experimental Section

DNA preparation

DNA oligonucleotides used in the experiments were H24 (5'-TTA GGG TTA GGG TTA GGG-3'), H26 (5'-TTA GGG TTA GGG TTA GGG TTA GGG TT-3'), *c*-MYC (5'-TGA GGG TGG GTA GGG TGG GTA A-3'), scrambled R26 (5'-TGG GAT TGT GTG AAG TTG GTG AGT TG-3'), and dsH26 (5'-TTA GGG TTA GGG TTA GGG TTA GGG TT and the complementary sequence AA CCC TAA CCC TAA CCC TAA CCC TAA-3'). Oligonucleotides and derivatives labeled at the 5'-end with fluorescent 6-carboxyfluorescein (FAM) were all purchased from Bio

Basic Canada Inc. and used without further purification. They were dissolved in 10 mM Tris-HCl (pH 7.4) buffer in the presence of 150 mM NaCl or KCl. All DNAs were annealed by heating at 95 °C for 5 min, gradually cooled to RT, and then incubated at 4 °C overnight before use. Double-stranded DNAs (dsDNA) were annealed by using the prepared DNA solution with an equal amount of the complementary oligonucleotide by the same heating-cooling temperature program used for the preparation of DNA mentioned above.

CD and melting point measurements

CD spectra of the oligonucleotides were collected with a Jasco J-715 spectropolarimeter (Jasco, Japan). A quartz cuvette with 1 cm optical path length was used to obtain spectra, which were recorded from 210 to 350 nm at 2 nm bandwidth and an instrument scanning speed of 50 nm min⁻¹ with 8 s response time. The measurements were the averages of two repetitions recorded at RT. Spectra were baseline-corrected and the signal contributions of the buffer were subtracted. Melting points (T_m) of H26 and c-MYC were measured by monitoring the CD signal at 265 nm, which is a characteristic peak of the quadruplex, in the absence and presence of compounds at the indicated ratios. The temperature was elevated gradually from 10 to 95 °C at a rate of 1.0 °C min⁻¹.

DNA alkylation, denaturing polyacrylamide gel electrophoresis, and gel visualization

DNA alkylation experiments were generally conducted at 37 °C for the given time, and were terminated by the addition of 2-mercaptoethanol (10 μL), and left for an additional 3 min. To the reaction mixture was then added Herring sperm DNA (1 μL) (Invitrogen), and precipitation was induced by the addition of cold ethanol. Precipitated DNA was then dissolved in gel loading buffer II (Ambion), denatured at 90 °C for 5 min, and immediately resolved in 20% polyacrylamide 7 M urea sequencing gel run at 1500 V constant Voltage for 3 h. The resulting gels were excited at 488 nm and the fluorescence of FAM was detected with a Typhoon 9400 (GE Healthcare, America) imager. The fluorescence intensity signals were further processed by using ImageQuant 5.2 software. The alkylation yield was calculated from the integral area of the alkylated bands divided by the total integral area of the alkylated bands and the nonalkylated band. The reported values, which are given as percentages, are the mean of at least three runs.

Competition experiments using different molar ratios of cold competitors

5'-FAM labeled H26 (1 μM) and BMVC-C3M (2 μM) were mixed with various concentrations of respective cold DNA competitors including H26, scrambled ssR26, and dsH26 to give the molar ratios of competitor/labeled H26 as 0.5, 1, 2, 5, and 10. After incubation at 37 °C for 20 h, standard workup, gel running conditions and visualization described above in the general method were used to obtain the gel images. The relative yield assuming the adduct formation in the absence of competitor as 100% in each experiment, is reported. The reported values, which are given as percentages, are the mean of at least three runs.

Mass spectroscopic analysis of BMVC-C3M/DNA adduct

Either 1 μM H26, H24, or c-MYC was reacted with 1 μM BMVC-C3M at 37 °C for 20 h. The respective solution was exchanged by ultrafiltration to 150 mM ammonium acetate (pH 7.0) and concentrated

to 30 μM followed by the addition of 10% volume of 2-propanol. A linear ion trap-orbitrap mass spectrometer (Orbitrap Elite, Thermo Fisher Scientific, Bremen, Germany) equipped with a home-built nanoflow source was used to provide the mass spectra of alkylation adducts. Infusion with the injected gradient of 80% acetonitrile and 20% 150 mM ammonium acetate (pH 7.0) was loaded into quartz emitters pulled in house, which were held at an ionization voltage of ca. 1.5 kV and a capillary temperature of 150 °C.

Nuclease digestion and LC-MS monitoring of alkylated adducts

The reacted mixtures (50 μL, 20 ng μL⁻¹) were digested for 12 h according to a reported protocol.^[36] The enzymes were removed by chloroform extraction at the end of enzymatic digestion. The resulting aqueous layers were dried, reconstituted in doubly distilled water, and analyzed by LC-MS. A linear ion trap-orbitrap mass spectrometer (Orbitrap Elite, Thermo Fisher Scientific, Bremen, Germany) coupled online with a UHPLC system (ACQUITY UPLC, Waters, Millford, MA) was used. The hydrolyzed residuals were separated by using a BEH-C18 column (Waters). The column temperature was maintained at 45 °C. Eluting buffers were Buffer A (2% acetonitrile and 0.1% formic acid in ddH₂O) and Buffer B (0.1% formic acid in acetonitrile) using gradients of 0 min (0.5% B), 1 min (1% B), 4 min (4% B), 10 min (95% B). The mass spectrometer was operated in the positive ion mode and set to one full FT-MS scan (m/z 200–1200) with 60,000 resolution.

Determination of the alkylation sites by treatment with piperidine

BMVC-C3M (0.5, 1, and 2 μM) was incubated with 5'-FAM labeled DNA (1 μM) in K⁺- or Na⁺-containing buffers for 20 h at 37 °C, respectively. Alkylation was terminated by the addition of 2-mercaptoethanol (10 μL) and the mixture was left for an additional 3 min. Maxam-Gilbert dG and dG+dA sequencing conditions were applied to provide the reference for the fragment migration. Dimethyl sulfate (DMS) (dG lane reference): the nonalkylated DNA was dissolved in 4 μL of DMS for 10 s at 37 °C, and terminated by the addition of 2-mercaptoethanol (10 μL). Formic acid (dG+dA lane reference): the nonalkylated DNA was dissolved in 50 μL formic acid for 3 min at 37 °C. Herring sperm DNA (1 μL) (Invitrogen) was then added. After ethanol precipitation, the resulting residues were directly incubated with 50 μL of 0.7 M piperidine at 90 °C for 25 min, which induced cleavage at the depurinated sites. After evaporation of piperidine, the gel running conditions and visualization described above in the general method were used to obtain the gel images.

Determination of the alkylation sites by 3'-exonuclease digestion

Nonalkylated H26 and alkylated adducts were first resolved by 20% polyacrylamide 7 M urea gel, then the band of interest was visualized by UV shadowing, excised, and extracted by using the crush-and-soak method. The eluted solutions were then concentrated by using an Amicon Ultra 0.5 mL 3 kDa centrifugal filter. The resulting alkylated H26 were incubated in 10 mM Tris-HCl buffer (pH 8.0) containing 2 mM MgCl₂ and 0.5 mg mL⁻¹ transfer RNA with the 3'-exonuclease phosphodiesterase I from *Crotalus adamanteus* venom (sigma) at 0.08 U μL⁻¹ for 30 min at 37 °C. The partial digestion of nonalkylated H26 was incubated with 0.001 U μL⁻¹. The digested fragments were subjected to the gel running conditions

and visualization described above in the general method to obtain the gel images.

Acknowledgements

We thank Professor Sheh-Yi Sheu at National Yang-Ming University for helpful discussions on molecular binding models and Professor Ta-Chau Chang's group at Academia Sinica for assistance in acquiring CD spectra and quadruplex-melting points at the beginning of the study. The mass spectrometry analysis was supported by the Metabolomics Facilities of Scientific Instrument Center at Academia Sinica. This work was supported by the Ministry of Science and Technology of the Republic of China, Taiwan [MOST 103–2113-M-002-016-MY3], and by a joint LAKER grant from the National Taiwan University and Academia Sinica.

Keywords: alkylation · DNA recognition · DNA structures · G-quadruplexes · molecular recognition

- [1] D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364–366.
- [2] H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422.
- [3] N. Maizels, L. T. Gray, *PLoS Genet.* **2013**, *9*, e1003468.
- [4] L. A. Cahoon, H. S. Seifert, *Science* **2009**, *325*, 764–767.
- [5] Y. Hai, W. Cao, G. Liu, S.-P. Hong, S. A. Elela, R. Klinck, J. Chu, J. Xie, *Nucleic Acids Res.* **2008**, *36*, 3320–3331.
- [6] a) D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, J. K. C. Lim, D. Von Hoff, K. Anderes, W. G. Rice, *Cancer Res.* **2009**, *69*, 7653–7661; b) S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.
- [7] a) H. Han, D. R. Langley, A. Rangan, L. H. Hurley, *J. Am. Chem. Soc.* **2001**, *123*, 8902–8913; b) P. Wang, L. Ren, H. He, F. Liang, X. Zhou, Z. Tan, *ChemBioChem* **2006**, *7*, 1155–1159; c) R. Rodriguez, S. Müller, J. A. Yeoman, C. Trentesaux, J.-F. Riou, S. Balasubramanian, *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759; d) G. W. Collie, R. Promontorio, S. M. Hampel, M. Micco, S. Neidle, G. N. Parkinson, *J. Am. Chem. Soc.* **2012**, *134*, 2723–2731; e) M. Nikan, M. Di Antonio, K. Abecassis, K. McLuckie, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2013**, *52*, 1428–1431; *Angew. Chem.* **2013**, *125*, 1468–1471; f) M. Micco, G. W. Collie, A. G. Dale, S. A. Ohnmacht, I. Pazitna, M. Gunaratnam, A. P. Reszka, S. Neidle, *J. Med. Chem.* **2013**, *56*, 2959–2974; g) Y. Xu, *Chem. Soc. Rev.* **2011**, *40*, 2719–2740; h) S. A. Ohnmacht, S. Neidle, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2602–2612.
- [8] a) H. Bertrand, S. Bombard, D. Monchaud, M.-P. Teulade-Fichou, *J. Biol. Inorg. Chem.* **2007**, *12*, 1003–1014; b) H. Bertrand, S. Bombard, D. Monchaud, E. Talbot, A. Guedin, J.-L. Mergny, R. Grunert, P. J. Bednarski, M.-P. Teulade-Fichou, *Org. Biomol. Chem.* **2009**, *7*, 2864–2871.
- [9] a) M. Di Antonio, F. Doria, S. N. Richter, C. Bertipaglia, M. Mella, C. Sissi, M. Palumbo, M. Freccero, *J. Am. Chem. Soc.* **2009**, *131*, 13132–13141; b) M. Nadai, F. Doria, M. Di Antonio, G. Sattin, L. Germani, C. Percivalle, M. Palumbo, S. N. Richter, M. Freccero, *Biochimie* **2011**, *93*, 1328–1340; c) F. Doria, M. Nadai, M. Folini, M. Di Antonio, L. Germani, C. Percivalle, C. Sissi, N. Zaffaroni, S. Alcaro, A. Artese, S. N. Richter, M. Freccero, *Org. Biomol. Chem.* **2012**, *10*, 2798–2806; d) F. Doria, M. Nadai, M. Folini, M. Scalabrin, L. Germani, G. Sattin, M. Mella, M. Palumbo, N. Zaffaroni, D. Fabris, M. Freccero, S. N. Richter, *Chem. Eur. J.* **2013**, *19*, 78–81; e) L. Yuan, T. Tian, Y. Chen, S. Yan, X. Xing, Z. Zhang, Q. Zhai, L. Xu, S. Wang, X. Weng, B. Yuan, Y. Feng, X. Zhou, *Sci. Rep.* **2013**, *3*, 1811.
- [10] D. Verga, F. Hamon, F. Poyer, S. Bombard, M.-P. Teulade-Fichou, *Angew. Chem. Int. Ed.* **2014**, *53*, 994–998; *Angew. Chem.* **2014**, *126*, 1012–1016.
- [11] J. Cadet, E. Sage, T. Douki, *Mutat. Res.* **2005**, *571*, 3–17.
- [12] a) S. Rickling, L. Ghisdavu, F. Pierard, P. Gerbaux, M. Surin, P. Murat, E. Defrancq, C. Moucheron, A. Kirsch-De Mesmaeker, *Chem. Eur. J.* **2010**, *16*, 3951–3961; b) E. Wachter, B. S. Howerton, E. C. Hall, S. Parkin, E. C. Glazer, *Chem. Commun.* **2014**, *50*, 311–313; c) M. Nadai, F. Doria, L. Germani, S. N. Richter, M. Freccero, *Chem. Eur. J.* **2015**, *21*, 2330–2334.
- [13] M. Di Antonio, K. I. E. McLuckie, S. Balasubramanian, *J. Am. Chem. Soc.* **2014**, *136*, 5860–5863.
- [14] T. J. Phesse, K. B. Myant, A. M. Cole, R. A. Ridgway, H. Pearson, V. Muncan, G. R. van den Brink, K. H. Voudsen, R. Sears, L. T. Vassilev, A. R. Clarke, O. J. Sansom, *Cell Death Differ.* **2014**, *21*, 956–966.
- [15] C.-C. Chang, C.-W. Chien, Y.-H. Lin, C.-C. Kang, T.-C. Chang, *Nucleic Acids Res.* **2007**, *35*, 2846–2860.
- [16] C.-C. Chang, I. C. Kuo, J.-J. Lin, Y.-C. Lu, C.-T. Chen, H.-T. Back, P.-J. Lou, T.-C. Chang, *Chem. Biodiversity* **2004**, *1*, 1377–1384.
- [17] D.-Y. Yang, S.-Y. Sheu, *J. Phys. Chem. A* **2007**, *111*, 9224–9232.
- [18] S. M. Bonesi, R. Erra-Balsells, *J. Heterocycl. Chem.* **2001**, *38*, 77–87.
- [19] S. M. Kerwin, *Curr. Pharm. Des.* **2000**, *6*, 441–471.
- [20] A. Risitano, K. R. Fox, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2047–2050.
- [21] E. M. Rezler, J. Seenisamy, S. Bashyam, M.-Y. Kim, E. White, W. D. Wilson, L. H. Hurley, *J. Am. Chem. Soc.* **2005**, *127*, 9439–9447.
- [22] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.
- [23] G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876–880.
- [24] a) K. N. Luu, A. T. Phan, V. Kuryavyy, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970; b) J. Li, J. J. Correia, L. Wang, J. O. Trent, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, 4649–4659.
- [25] J. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
- [26] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
- [27] A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
- [28] Z.-F. Wang, T.-C. Chang, *Nucleic Acids Res.* **2012**, *40*, 8711–8720.
- [29] C.-C. Chang, J.-Y. Wu, C.-W. Chien, W.-S. Wu, H. Liu, C.-C. Kang, L.-J. Yu, T.-C. Chang, *Anal. Chem.* **2003**, *75*, 6177–6183.
- [30] E. Largy, A. Granzhan, F. Hamon, D. Verga, M.-P. Teulade-Fichou, *Top. Curr. Chem.* **2012**, *330*, 111–177.
- [31] Y.-L. Tsai, C.-C. Chang, C.-C. Kang, T.-C. Chang, *J. Lumin.* **2007**, *127*, 41–47.
- [32] C.-C. Kang, W.-C. Huang, C.-W. Kouh, Z.-F. Wang, C.-C. Cho, C.-C. Chang, C.-L. Wang, T.-C. Chang, J. Seemann, L. J.-S. Huang, *Integr. Biol.* **2013**, *5*, 1217–1228.
- [33] M.-H. Chung, H. Kiyosawa, E. Ohtsuka, S. Nishimura, H. Kasai, *Biochem. Biophys. Res. Commun.* **1992**, *188*, 1–7.
- [34] D. M. Noll, T. M. Mason, P. S. Miller, *Chem. Rev.* **2006**, *106*, 277–301.
- [35] C. M. Castleberry, L. P. Rodicio, P. A. Limbach in *Curr. Protoc. Nucleic Acid Chem., Unit 10.2: Electrospray Ionization Mass Spectrometry of Oligonucleotides*, Wiley, New York, **2008**.
- [36] E. P. Quinlivan, J. F. Gregory III, *Anal. Biochem.* **2008**, *373*, 383–385.
- [37] a) E. Haapala, K. Hakala, E. Jokipelto, J. Vilpo, J. Hovinen, *Chem. Res. Toxicol.* **2001**, *14*, 988–995; b) D. Florea-Wang, E. Haapala, J. Mattinen, K. Hakala, J. Vilpo, J. Hovinen, *Chem. Res. Toxicol.* **2003**, *16*, 403–408.
- [38] C. J. Burrows, J. G. Muller, *Chem. Rev.* **1998**, *98*, 1109–1152.
- [39] K. Savela, A. Hesso, K. Hemminki, *Chem.-Biol. Interact.* **1986**, *60*, 235–246.
- [40] Z.-F. Wang, M.-H. Li, S.-T. D. Hsu, T.-C. Chang, *Nucleic Acids Res.* **2014**, *42*, 4723–4733.
- [41] S. Mukherjee, A. Guainazzi, O. D. Schärer, *Nucleic Acids Res.* **2014**, *42*, 7429–7435.
- [42] L. Rao, U. Bierbach, *J. Am. Chem. Soc.* **2007**, *129*, 15764–15765.
- [43] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.

Received: July 3, 2015

Published online on October 13, 2015