0026-895X/07/7104-1165–1178$20.00
MOLECULAR PHARMACOLOGY
Vol. 71, No. 4
30072/3194326
Mol Pharmacol 71:1165–1178, 2007
Printed in U.S.A.

Structure of the d(CGCGAATTCGCG)$_2$ Complex of the Minor Groove Binding Alkylating Agent Alkamin Studied by Mass Spectrometry[S]

Amin M. S. Abdul Majid, George Smythe, William A. Denny, and Laurence P. G. Wakelin

School of Pharmaceutical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia (A.M.S.A.M.); School of Medical Sciences (A.M.S.A.M., L.P.G.W.) and Bioanalytical Mass Spectrometry Facility (G.S.), University of New South Wales, New South Wales, Australia; Auckland Cancer Society Research Centre, Faculty of Medicine and Health Science, University of Auckland, Auckland, New Zealand (W.A.D.)

Received August 22, 2006; accepted January 18, 2007

ABSTRACT

Nitrogen mustard alkylating agents are important cancer drugs. Much interest has been focused on redirecting their covalent adducts from the N7 atoms of guanine in the major groove of DNA to the N3 atoms of adenine in the minor groove by attaching mustard groups to AT-selective minor groove binding ligands. Here we describe the use of electrospray ionization and matrix-assisted laser desorption ionization/time-of-flight mass spectrometry to study the structure of the DNA complexes of two minor groove binding polybenzamide mustards, alkamin and alkamini; the former is a bis-half-mustard in which reactive groups are disposed at each end of the ligand, and the latter is its monofunctional analog. Alkamin is potently cytotoxic and active in experimental mouse tumor models, whereas alkamini is not. We have studied their interaction with the DNA dodecamer d(CGCGAATTCGCG)$_2$, designated A2T2, and we provide a detailed analysis of the observed DNA-ligand adduct ions and their fragmentation products. We find that alkamini alkylates A2T2 at guanine G4 and adenines A5 and A6 in a manner consistent with covalent attack on purine N3 atoms from the minor groove of the AT tract. Alkamin also forms monofunctional adducts at G4 and both adenines in which the second mustard arm is hydrolyzed but, in addition, forms a variety of interstrand cross-links between adenines A5/A6 and A5′/A6′, an interstrand cross-link between G4 and A6′, and an intrastrand cross-link between G4 and A6. We conclude that the marked cytotoxicity of alkamin and its experimental antitumor activity could be the consequence of its ability to cross-link cellular DNA at AT tract sequences.

Nitrogen mustard alkylating agents are important drugs in the treatment of cancer, but their usefulness is limited by their lack of intrinsic specificity for DNA. In addition to forming cytotoxic DNA interstrand cross-links, they are susceptible to hydrolysis, form complexes with proteins, can be deactivated by cellular thiols, and yield mutagenic and carcinogenic monofunctional adducts (Gamcsik et al., 1999; Panasci et al., 2001; Drablos et al., 2004). These deficiencies led to attempts to direct mustard groups to DNA by attaching them to minor groove binding (MGB) ligands and to intercalating agents, the major objectives being to increase dose potency and to enhance interstrand cross-link formation at the expense of monofunctional reaction (Gravatt et al., 1991; Denny, 2001). In general, the former has been achieved, but the latter remains an objective. Attaching mustards to MGBs redirects alkylation from the N7 of guanine in the major groove to the N3 of adenine, thereby circumventing efficient guanine adduct repair systems (Denny, 2001; Drablos et al., 2004). Several classes of MGB-directed aniline mustards have been described involving oligopyrrole peptide, bisbenzimidazole, bis-quaternary ammonium heterocycle, and polybenzamide carriers (see Denny, 2001; Fig. 1). Most notable among these agents is the clinical trial candidate tallimustine, which binds to AT tracts alkylating the N3 of adenine in

This work was supported financially by University Science Malaysia (A.M.S.A.M.), the Australian National Health and Medical Research Council (L.P.G.W.), and the Auckland Division of the Cancer Society of New Zealand (W.A.D.). The mass spectrometric analysis for this work was carried out at the Bioanalytical Mass Spectrometry Facility, University of New South Wales, New South Wales, Australia, and was supported in part by grants from the Australian Government Systemic Infrastructure Initiative and Major National Research Facilities Program (University of New South Wales node of the Australian Proteome Analysis Facility).

Article, publication date, and citation information can be found at http://molpharm.aspetjournals.org.

doi:10.1124/mol.106.030072.

[S] The online version of this article (available at http://molpharm.aspetjournals.org) contains supplemental material.

ABBREVIATIONS: MGB, minor groove binding; MALDI, matrix-assisted laser desorption ionization; ESI, electrospray ionization; MS/MS, tandem mass spectrometry; TOF, time of flight; RHF, right hand fragment; LHF, left hand fragment; SS, single-stranded; deA, deadenylated; A2T2, d(CGCGAATTCGCG)$_2$. See Supplementary Figs. S1–S3 and Table S1 for abbreviations describing molecular ion assignments.

the sequences 5′-TTTTGA and 5′-TTTTAA (Broggini et al., 1995; Rossi et al., 1996; Turner et al., 1998) without reaction at guanine (Tagliabue et al., 1997). Tallimustine, despite being a full mustard, does not form interstrand cross-links (Bellorini et al., 1995).

In an attempt to maximize cross-link frequencies at sequences such as AATT and AAATTT, Denny and colleagues (Prakash et al., 1991; Atwell et al., 1995) synthesized ligands that incorporate single-armed mustard groups disposed at either end of a terephthalic acid framework (Fig. 1), the intent being to alkylate adenines four residues apart in complementary strands. Dimethylaminomethyl groups provide positive charge to promote reversible binding before alkylation. The annular shape of the ligand is intended to position the functional groups so as to promote hydrogen bonding in the center of the AT tract and alkylation of the distal adenine N3 nitrogens. DNA sequencing and molecular modeling experiments show that alkamin, the parent bifunctional polybenzamide mustard (Fig. 1), binds to AT tracts in a manner consistent with these design principles (Prakash et al., 1991; Turner et al., 1999) and that its monofunctional analog, alkamini, shares this binding mode and sequence selectivity (Turner et al., 1999). Alkamin is potently cytotoxic, with an IC_{50} value of 7 nM against P388 mouse leukemia cells, has in vivo activity in a stringent single dose assay, and shows a positive hypersensitivity ratio of 15 for activity against a pair of Chinese hamster ovary cell lines designed to reveal whether the mode of drug action involves DNA interstrand cross-links (Prakash et al., 1991; Atwell et al., 1995). In these assays, alkamini is 35-fold less cytotoxic and lacks in vivo activity, pointing to the importance of alkamin cross-links as a major determinant of its biological activity (Atwell et al., 1995). As a reference, the clinically used alkylating agent chlorambucil, which reacts virtually exclusively at guanine N7 sites on DNA (Mattes et al., 1986a), has an IC_{50} of 6.75 μM against P388 leukemia, and the same in vivo efficacy as alkamin but at 25-fold higher doses (Atwell et al., 1995).

To date, the most informative approaches to studying alkylator-DNA complexes have used ^{32}P end-labeled DNA fragments and Maxam-Gilbert sequencing techniques (Kohn et al., 1987). These methods rely on the fact that the agents quaternize purine ring nitrogens, which labilizes the glycosidic bond to hydrolysis and leads to hydrolytic cleavage of the sugar-phosphate backbone at the apurinic sugar (Mattes et al., 1986a,b). In seeking new approaches that might reveal more detail about the nature of the DNA-ligand complex, we have explored the use of oligonucleotides of defined sequence and MALDI and ESI mass spectrometry to characterize DNA complexes of alkamin and alkamini. At the outset of this work, little was known about the suitability of electrospray mass spectrometry to study covalent DNA-ligand interactions of this kind, and still there are no reports of the use of MALDI spectroscopy in studies of such complexes. Sheil and colleagues (Wickham et al., 1995; Iannitti and Sheil, 1997; Iannitti-Tito et al., 2000) used electrospray methods to investigate covalent oligonucleotide complexes of the N7 guanine-alkylating agents hedamycin and a phenanthridinium alkylbromide and report a defined pathway for gas-phase fragmentation of the complex molecular ion that reveals the position of the quaternized base. We have found that both MALDI and ESI methods are well suited to analyze oligonucleotide-MGB alkylator adducts, and that they are capable of providing detailed insights into the molecular nature of the complexes, particularly with regard to the formation of intra- and intermolecular cross-links, issues that are difficult to resolve at base resolution by other means. We conclude that the interstrand cross-link frequency for alkamin binding at sites with appropriately disposed adenines is much higher than previously realized, and we speculate that the formation of these cross-linked species lies at the heart of its biological activity.

Materials and Methods

Materials. Ammonium acetate was purchased from APS Finechem (Seven Hills, NSW, Australia), and a 300 mM stock solution was adjusted to pH 7.0 with glacial acetic acid (Scharlau, Barcelona, Spain). The stock solution was diluted with water to produce buffer solutions of concentration 20, 50, and 100 mM in ammonium acetate, which were stored in plastic bottles at 4°C. Analytical grade sterile Baxter water (Baxter, Toongabbie, NSW, Australia) was used throughout. All volumetric measurements were carried out using Gilson pipettes (Interpath Services, Caringbah, NSW, Australia), and extreme measures were taken to avoid the use of glass containers to minimize salt contamination. Nevertheless, both MALDI and ESI spectra are found to include adventitious alkali metal ion adducts. The aniline mustards alkamin, molecular weight 714.6, and alkamini, molecular weight 609.0, were synthesized as the dihydrochloride salts as described previously (Prakash et al., 1991: Atwell et al., 1995). All ligands were dissolved to a concentration of 2.0 mM in a mixture of isopropanol and acetonitrile (APS Finechem), to prevent

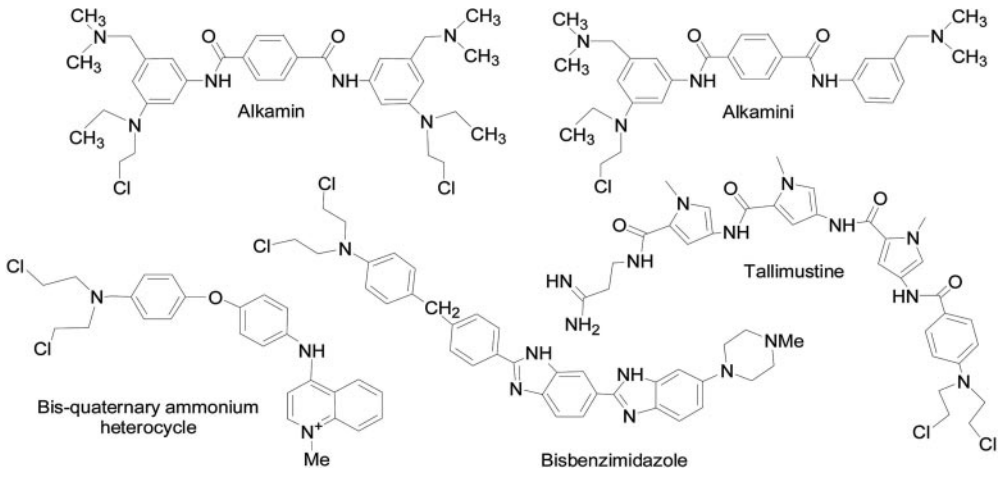


Fig. 1. Structures of MGB alkylating mustards.

hydrolysis and loss of functional groups and stored at −20°C. The ammonium salt of the dodecanucleotide d(CGCGAATTCGCG)$_2$, designated A2T2, single-strand molecular weight 3833.4, was purchased from Oswel DNA Services (Southampton, UK), and dissolved to a concentration of 4 mM in duplex in deionized water (3.9 mg of A2T2 in 127 μl). The stock solution was diluted to 2 mM concentration in 100 mM ammonium acetate buffer, annealed by heating to 90°C, slowly cooled to room temperature, and further diluted to 100 μM in 20 mM ammonium acetate. All DNA samples were stored at −20°C. Bases are numbered from the 5′ end on the Watson strand (e.g., C1G2C3G4A5) and are denoted as their prime on the Crick strand (e.g., C1′G2′C3′G4′A5′). Thermal melting measurements (Cary Bio5000 spectrophotometer) at a DNA concentration of 2 to 4 μM in duplex in 20 mM sodium acetate buffer pH 7.0 gave biphasic melting curves with a melting temperature of approximately 45°C.

A QSTAR Pulsar *i* electrospray mass spectrometer (Applied Biosystems, Foster City, CA) fitted with a NanoSpray ion source was employed to obtain spectra using the electrospray ionisation method, and a Voyager Elite XL Biospectrometry Workstation (Applied Biosystems) used for the acquisition of mass spectra via the matrix-assisted laser desorption/ionization method. The Voyager instrument was equipped with a pulsed nitrogen laser (337 nm) as the desorption/ionization source. Both instruments are located in the Bioanalytical Mass Spectrometry Facility, University of New South Wales, NSW, Australia. Electrospray analyses were carried out using freshly prepared solutions containing 8 μl of a 50% water/acetonitrile mixture and 2 μl of the aqueous DNA sample (i.e., an ESI solvent-to-sample ratio of 5:1), and MALDI measurements employed a freshly prepared 3-hydroxypicolinic acid matrix (3-hydroxypyridine-2-carboxylic acid; Lancaster Synthesis, Lancaster, England) as outlined in the *Voyager Biospectrometry Workstation User's Guide* Version 5.1. The matrix was prepared by dissolving 50 mg of 3-hydroxypicolinic acid in 1 ml of a 50% aqueous solution of acetonitrile, and this was mixed with 1 ml of a 50 mg/ml aqueous solution of ammonium citrate (dibasic salt; Research Organics Inc., Cleveland, OH). The final mixture was vigorously vortexed to effect complete dissolution.

Preparation of DNA-Ligand Complexes. All DNA-ligand complexes were prepared in 20 mM ammonium acetate (NH_4OAc) buffer, pH 7.0, at a stoichiometry of one drug molecule added per duplex by mixing 2 μl of 2 mM ligand in isopropanol with 38 μl of 100 μM DNA in 20 mM NH_4OAc solution. Thus, the final conditions are 100 μM ligand, 95 μM DNA, and 2.5% isopropanol in 20 mM NH_4OAc, pH 7.0. Complexes were prepared in 1.5-ml plastic Eppendorf tubes and were incubated in a heating block (Thermoline, Smithfield, NSW, Australia) at 37°C for 24 h. Each tube was individually wrapped with aluminum foil to facilitate even heat distribution during incubation, during which they were placed in an inverted position to minimize sample evaporation. Control DNA samples, in which the A2T2 DNA was incubated without drug, were subjected to the same procedures. Freshly prepared complexes were analyzed immediately using both MALDI and ESI mass spectrometers after diluting the DNA concentration to 20 μM with matrix or organic solvent (see below). Where appropriate, MS/MS analysis of the complexes was performed using the ESI instrument to confirm ion peak identity.

MALDI-TOF Measurements. All DNA samples were mixed with freshly prepared 3-hydroxypicolinic acid matrix solution at a 1:5 (v/v) ratio before MALDI-TOF analysis (typically 2 μl of DNA sample plus 8 μl of matrix solution). Two microliters of this mixture was then spotted onto a standard 100-well MALDI stainless steel sample plate and allowed to dry under vacuum. Once dried, a further 2 μl of the sample was applied to the same spot, followed by more vacuum drying. All analyses were carried out immediately after sample spotting. MALDI spectra were recorded in positive ion mode, because this mode yielded the same number of DNA-related peaks as negative ion mode, but also included peaks for the base-ligand fragments, which are absent in the latter. All spectra were obtained using the reflector mode, because this gives the highest resolution, and the system was not limited by sensitivity considerations. In confirmation of this, no missing peaks were detected when comparing results obtained from reflector and linear modes of operation. The mass spectrometer was set to delayed extraction mode, with a delay time of 80 ns, because this gives superior resolution compared with the continuum mode. The accelerating voltage was set to 20,000 V with the grid voltage at 70%. The mirror voltage ratio was set to 1:12, and the guide wire percentage was fixed at 0.04%. Spectra were acquired in four mass ranges: 500 to 8500 Da, 500 to 1000 Da, 1000 to 4000 Da, and 4000 to 8500 Da. The number of laser shots per spectrum was set to 240, with a repeat rate of 5.2 Hz, because this produced a good signal-to-noise ratio; the laser intensity was adjusted to maximize said ratio. It was found that setting the laser intensity to 2746, on a scale up to 4000, generally gave the best results, although a value of 2670 gave cleaner spectra in the lowest mass range. To further minimize low mass peak saturation, the low mass gate was set at 499 Da. The timed-ion selector was switched off, and the analog-to-digital converter bin size was set to 2 ns to enable recording of the maximum number of data points per run, thus improving resolution. An A-to-D converter input bandwidth of 20 MHz provided appropriate noise limitation, and the amplifier vertical scale was set to 50 mV with a vertical offset fixed at −5%. All spectra were recorded using manual calibration; the scale was calibrated to the mass of the individual unreacted single stranded DNA peaks.

Electrospray Measurements. All electrospray experiments were performed using a NanoSpray source involving a small Pyrenea-coated glass needle (Proxeon, Odense, Denmark), which allows a lower sample throughput compared with normal electrospray methods. The sample-to-buffer ratio (v/v) was kept at 1:5 (2 μl of sample and 8 μl of solvent), to keep the ammonium salt concentration at a low level, and to minimize clogging of the ion source needle. These conditions produce spectra with good signal intensity and high resolution. The needle tip was positioned ~10 mm from the orifice, and nitrogen was used as the curtain gas. Two-microliter samples were injected over a period of 5 min with a flow rate of 20 nl per min. The QSTAR instrument was calibrated against a mixture of ALILTLVS peptide and CsI. Spectra were recorded from *m/z* 100 to 2000. The curtain gas was limited to a flow of 20 with the ion spray voltage set at approximately −700 V in the negative ion mode and +750 V in the positive ion mode. For all measurements, the declustering potential was set at −50 V and the declustering potential 2 at −10 V. The collision gas concentration, which aids in molecular ion focusing, was set to a low value of 5 units, and the focusing potential value fixed at −220 V. The ion-release delay was limited at 6 ns and the ion-release pulse width set at 5 ns. In the resolution settings, the Q1 (quadrupole 1), was set to run at unit resolution and the ion energy 1 setting was fixed at −1. The grid voltage was set at −3 with the Enzil lens focus set at −10 and the Enzil lens steering at 0.5. The multichannel plate detector (MCP) was set at 2000 V. ESI spectra were recorded in both negative and positive ion mode.

Data were analyzed by using *Analyst QS*, the primary mass-to-charge ratio (*m/z*) spectra containing singly and multiply charged species being reconstituted into a simple mass spectrum using the *Bayesian Peptide Reconstruction* (*BPR*) function. The *BPR* was set to scan from 100 to 9000 Da, with mass tolerance and signal-to-noise threshold values set to 0.2 Da and 10, respectively. Peaks with areas less than 1% of the largest peak in the primary spectra were eliminated from further analysis, and all peaks in the reconstituted spectra were checked for authenticity by visually inspecting their multiply charged components in the primary spectra. Background peaks that have their origins in the buffer solution were not taken into account. It was a common finding that singly charged species in the primary spectra failed to be automatically incorporated into the reconstituted spectra by the *BPR* function.

ESI Tandem MS Measurements. Where possible, negative ion MS/MS measurements were made on ligand-DNA complexes using the MS/MS mode of the QSTAR spectrometer. Two microliters of the 100 μM sample solution was dissolved in 10 μl of a 50:50 water/

acetonitrile aqueous buffer (5 mM NH_4OAC) and loaded into a Nano-Spray needle with the tip positioned ~10 mm from the orifice. Instrument settings were as for the nontandem ESI measurements. Nitrogen was used as curtain gas and a potential of −800 V was applied to the needle. A TOF-MS scan was acquired (*m/z* 400-2500, 1 s) and accumulated for ~2 min into a single file. Precursor masses determined from the TOF MS scan were selected by Q1 for MS-MS analysis. Nitrogen was used as collision gas, and optimal collision energy was chosen for each precursor ion.

Molecular Mass Calculations. Ligand molecular masses were calculated manually and by using *ChemDraw Ultra 7.0*, as their average molecular weights. Average DNA molecular masses were calculated using the *Nucleic Acid Editor* function in the Micromass Ltd *Masslynx Biolynx* software package. The masses of DNA hydrolytic fragments, and DNA-ligand complexes, were calculated manually with the assistance of the *Nucleic Acid Editor* program.

Results

MALDI and ESI spectra of alkamin- and alkamini-A2T2 DNA complexes were acquired as described under *Materials and Methods*. In all, six independent A2T2 complexes of alkamin and alkamini were prepared, and dozens of spectra were measured for each complex. ESI and MALDI measurements were also made with the drug-free DNA. The displayed figures show a particular, representative spectrum, whereas the tables record the mean data obtained over all the complexes studied. Raising the complex input stoichiometry to 2 ligand molecules to one DNA duplex did not introduce new species into the spectra but did increase peak ion intensities. The mass spectra indicate that alkamin forms only a single complex at an input ratio of 1:1, whereas alkamini forms both 1:1 and 2:1 complexes at this stoichiometry. In the Supplementary Material, we describe the types of chemical species that accompany alkylation to assist the reader in understanding the complexity of the observed mass spectra, and to help facilitate their molecular interpretation. We also provide a nomenclature for describing the observed species in Table S1. The main chemical processes leading to ligand-induced fragmentation of A2T2 center on hydrolysis and dehydration reactions. The sequence of events starts with quaternization at the N3 atoms of adenine or guanine, which initiates hydrolytic cleavage of the glycosidic bond to give a wide range of fragments as shown in Figs. S1 and S2. The Supplementary Material also includes a description of the MS/MS gas-phase degradation pathways (Fig. S3; note that fragment nomenclature is given in italics), expanded MALDI spectra (Figs S4–S6), and a discussion of the problems associated with quantifying the relative amounts of the observed species.

MALDI Mass Spectrum of the Alkamini-A2T2 Complex. Figure 2a shows the positive ion MALDI spectrum of the alkamini-A2T2 DNA complex. The masses observed are listed in Table 1 along with their molecular assignments. In the MALDI spectra, DNA and its fragment ions generally appear as the singly positively charged free acid; i.e., each phosphate group bears a hydrogen atom. Figure 2a is the full-range spectrum, and Fig. S4, a–e, displays expanded portions. The observed peaks can be conveniently separated into five structurally relevant mass range groups (i.e., *m/z* 500-1000, *m/z* 1000–3000, *m/z* 3000–3600, and *m/z* 3600–4500, and *m/z* 7000–8000). The six peaks between *m/z* 7000 and 8000 represent duplex forms of A2T2 DNA and related species. These include the unreacted A2T2 duplex itself at *m/z* 7291.8, its ion-pair with a protonated ligand-adenine adduct derived from a depurination event (*m/z* 7925.1), the intact duplex-alkamini adduct (*m/z* 7791.0), the singly deadenylated form of the 2:1 adduct (*m/z* 7674.1), and singly and doubly deadenylated duplex ions (*m/z* 7174.8 and 7057.8). The group of peaks between *m/z* 3600 and 4500 include ion pairs of unreacted single-stranded DNA and single-stranded DNA-alkamini adducts. Here we observe an intense peak representing the single stranded DNA bound to an alkamini molecule at *m/z* 4146.2, its K^+ salt at *m/z* 4184.4, and a deadenylated derivative of the ligand-DNA adduct at *m/z* 4029.2. The identity of the SS-L peak is confirmed by MS/MS ESI measurements on the parent ion (*m/z* 1035.3, z = −3, see Fig. 4a) that yield gas phase peaks for deadenylated single-stranded DNA, *SS deA*, right-hand fragments (*RHFs*), and left hand fragments (*LHFs*) for reaction with adenines A5 and A6, and the adenine-alkamini adduct. The remaining peaks found within the *m/z* 3600 and 4500 mass range are assigned as various ion pairs of intact single-stranded DNA, the pairing agents being protonated ligand-adenine adducts and Na^+ and K^+. The assemblage of peaks within the range *m/z* 3000 to 3600 represent deadenylated and deguanylated single-stranded DNAs resulting from depurination of the quaternized bases. The most structurally significant of these are the peaks at *m/z* 3530.4, 3592.4 and 3577.4, representing deadenylated DNA species and the Na^+ and K^+ salts of their dehydration product [see Fig. S1, species (**2**) and (**9**)]; the peak at *m/z* 3514.4 assigned to deguanylated DNA; and the peak at *m/z* 3397.2 in which the DNA has lost both a guanine and an adenine to hydrolysis. The assignments of the SS deA and SS deA β-elim species are verified by MS/MS ESI measurements on the parent ions (SS deA, *m/z* 881.0, z = −4, see Fig. 4b; SS deA β-elim, *m/z* 701.2, z = −5, see Fig. 4c) that again yield peaks for *RHFs* and *LHFs* for reaction with adenines A5 and A6.

The array of peaks between *m/z* 1000 and 3000 represent the fragmented DNA products resulting from phosphate hydrolysis at the alkylated and depurinated sites. Here, we see a set of RHFs (see Figs. S1 and S2) resulting from reaction with guanine G4, *m/z* 2490.6; adenine A5, *m/z* 2177.4; and adenine A6, *m/z* 1864.4. Also apparent is the product where phosphate hydrolysis at A6 cleaved in the alternate manner to yield a DNA fragment with a 5′-OH terminus, *m/z* 1785.6 (see Fig. S1). In contrast to the abundance of RHFs, there is a paucity of matching LHFs, so the only LHF species observable are derivatives of reaction at adenine A6: its dehydration product at *m/z* 1667.5 and its 3′-phosphorylated derivative with *m/z* 1766.2. This region of the spectrum also contains peaks representing doubly charged single-stranded DNA, *m/z* 1824.8, and the doubly charged DNA-ligand adduct at *m/z* 2074.2. Finally, in the low molecular mass region *m/z* 500 to 1000, we find peaks representing ligand-base adducts resulting from depurination, hydrolyzed unreacted ligand, and an assortment of matrix-derived materials. The most abundant complex-derived species is the alkamini-adenine adduct at *m/z* 635.5, followed by the hydrolyzed free ligand, *m/z* 518.2, and the guanine-ligand adduct, *m/z* 651.8.

ESI Mass Spectrum of the Alkamini-A2T2 Complex. Fig. 2b shows the primary ESI mass spectrum for the alkamini-A2T2 complex acquired in negative ion mode, the observed peaks, and their molecular assignments being

recorded in Table 1. In the primary spectrum, DNA and its fragment ions appear as multiply charged negative ions because of the loss of protons from the acid phosphate groups. Fig. 2b is the full-range spectrum, and its singly charged Bayesian reconstituted form is shown in Fig. 2c. In the ESI spectrum, the peaks fall into four distinct groups, with *m/z* values in the ranges 500 to 1000, 1000 to 3000, 3000 to 3600, and 3600 to 4500; the fifth group of peaks representing the duplex species that appear in the MALDI spectrum around 7000 to 8000 Da are notably absent. As in the MALDI spectrum, peaks within the *m/z* range 3600 to 4500, Fig. 2c, represent single-stranded DNA-ligand complexes and their K^+ salts and free single-stranded DNA and its Na^+, K^+, and protonated ligand-base adduct ion pairs. The most significant peak in this region, from a structural standpoint, is that of the single-

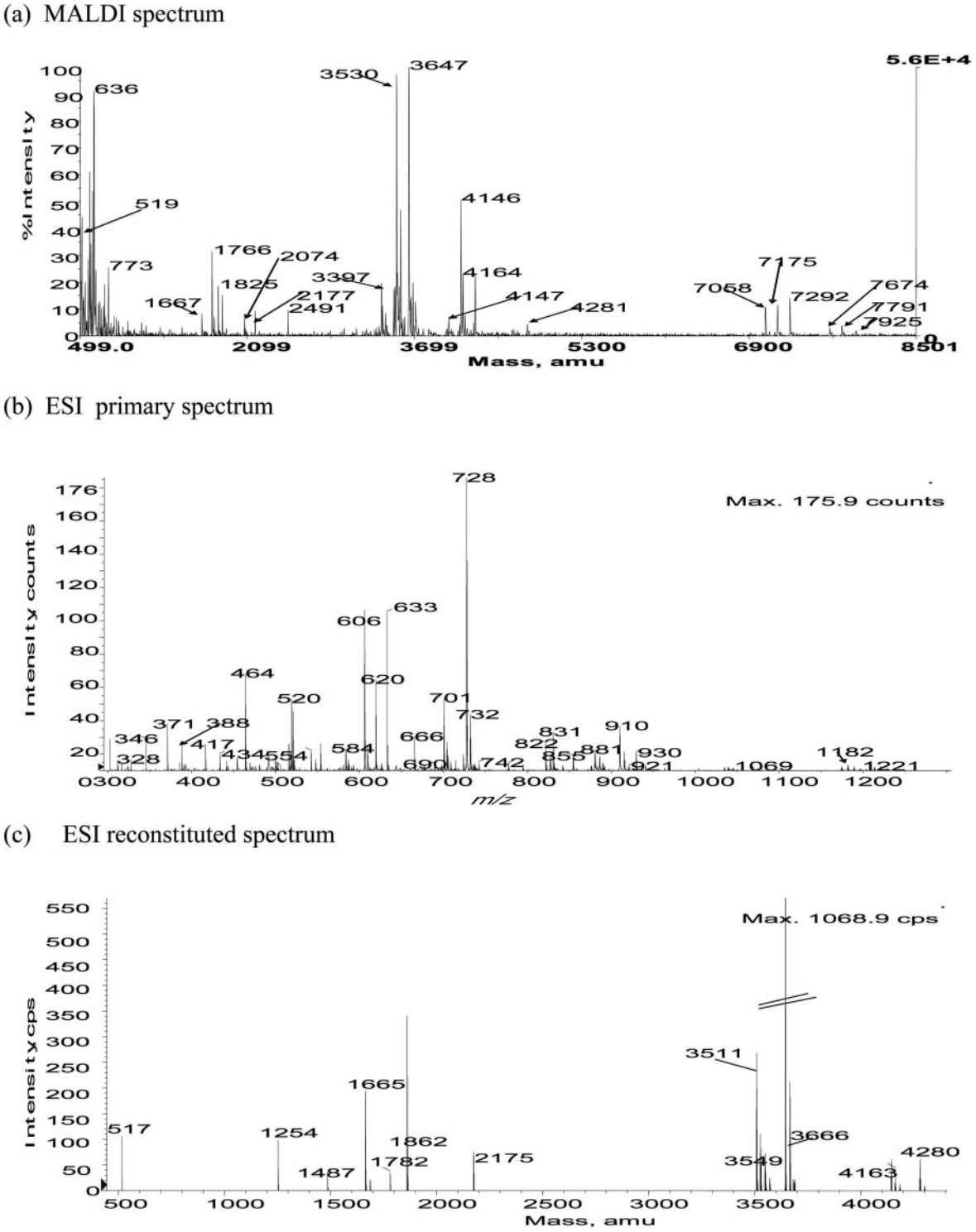


Fig. 2. Mass spectra for the 1:1 alkamini-A2T2 complex. Positive ion MALDI spectrum, *m/z* 500 to 8500 (a), negative ion ESI primary spectrum, *m/z* 300 to 1300 (b), and Bayesian reconstituted negative ion ESI spectrum, *m/z* 500 to 4400 (c).

stranded DNA-ligand adduct. Its observed negative ion mass of 4145.2 is 2 Da lower than the calculated mass of the complex, reflecting the necessity of deprotonating the quaternized adduct to yield a negative molecular ion. Fig. 2c also shows the peaks in the range *m/z* 3500 to 3600, which includes the depurinated single stranded DNAs, in this case the deadenylated species, *m/z* 3528.8, their Na^+ salts, and their dehydrated β-elimination products (see Fig. S1), *m/z* 3510.7. The DNA fragments resulting from hydrolysis at these apurinic sites are shown in Fig. 2c over the *m/z* range of 1000 to 3000. Here, we find strong peaks representing the RHFs for alkylation at adenines A5, *m/z* 2175.8, and A6, *m/z* 1862.7, together with a derivative of the latter in which the phosphate group cleaved in the alternate manner (see Fig. S1), *m/z* 1782.7. The LHF for reaction at adenine A6 appears as its β-elimination product, *m/z* 1665.7, together with its Na^+ adduct, *m/z* 1687.7, and a further degradation product in which both the apurinic sugar and its 5′-phosphate have been lost to yield CGCGA, *m/z* 1487.5. Reaction at A5 gives a peak representing its LHF, which has lost the sugar ring, giving CGCG, *m/z* 1254.4.

In the low molecular mass region of the spectrum (Fig. 2c), the only ligand-related species observable is the hydrolyzed free ligand, L-OH, at *m/z* 517.3. To detect the base-ligand adducts in ESI spectra, it was necessary to collect data in the positive ion mode, as shown in the Bayesian transformation in Fig. S6a. Here we observe peaks representing the ligand-adenine adduct, *m/z* 634.0, the ligand-guanine adduct, *m/z* 650.6, and the hydrolyzed ligand, *m/z* 517.3. In contrast, no

TABLE 1

Molecular assignments of the ions observed in the ESI and MALDI spectra of the 1:1 alkamini-A2T2 DNA complex

The calculated masses are for the neutral species, the MALDI data were acquired in the positive ion mode, and the ESI data represent the mass of singly-charged negative ions derived from the Bayesian transformation of the primary electrospray spectrum. The charged states involved in the transformations are indicated. The nomenclature for the species assignments is given in Table S1 and Figs. S1 and S2.

Species Assignments	Mass (M^0) Calculated	Mass MALDI (M^+)		Mass ESI (M^-)		
		Observed	Area	Observed	Area	Charges
			%		%	
DS + $[HL\text{-}A]^+$	7927.5	7925.1	0.1			
DS-L	7793.5	7791.0	0.4			
DS-L deA	7676.5	7674.1	0.4			
DS DNA	7292.8	7291.8	1.5			
DS deA	7175.8	7174.8	1.2			
DS deA deA	7058.8	7057.8	1.2			
SS DNA + $[HL\text{-}A]^+ + [Na^+]$	4303.1			4301.7	0.2	4
SS DNA + $[HL\text{-}A]^+$	4281.1	4281.5	2.1	4280.1	1.8	5,4
SS-L + $[K^+]$	4185.1	4184.4	0.7	4185.2	0.3	4
SS DNA + $[HL\text{-}OH]^+$	4164.1	4164.2	2.7	4163.3	0.9	5,4
SS-L	4147.1	4146.2	5.6	4145.2	2.4	6,5,4
SS-L deA	4030.1	4029.2	2.7			
UA		3717.2	0.4			
SS DNA + $3[Na^+]$	3712.4	3710.5	1.5			
SS DNA + $2[Na^+]$	3690.4			3689.6	1	5,4,3
SS DNA + $[K^+]$	3684.4	3685.7	1.6	3683.4	0.4	5,4
SS DNA + $[Na^+]$	3668.4	3669.4	0.1	3667.6	9.5	7,6,5,4,3
UA		3663.2	0.4			
UA		3656.1	1.6			
SS DNA	3646.4	3647.4	13	3645.4	39	8,7,6,5,4,3
SS deA + $[Na^+]$,$[K^+]$	3605.4	3606.8	0.5			
SS deA β-elim + $[K^+]$, $2[Na^+]$	3592.4	3592.4	0.7			
SS deA β-elim + $3[Na^+]$	3576.4	3577.4	0.5			
SS deA + $2[Na^+]$	3573.4			3572.3	0.8	4,3
SS deA + $[K^+]$	3567.4	3568.5	6.3			
SS deA + $[Na^+]$	3551.4	3552.6	4.1	3550.6	2.8	5,4,3
SS deA	3529.4	3530.4	13	3528.8	6.4	7,6,5,4,3
SS deG	3513.4	3514.4	2.4			
SS deA β-elim	3511.4			3510.7	8.5	7,6,5,4
UA		3423.4	1.1			
SS deA deG	3396.3	3397.2	1.6			
UA		3385.3	0.9			
RHF G4	2489.6	2490.6	0.8			
RHF A5	2176.4	2177.4	0.9	2175.8	3.1	5,4,3
$[SS\text{-}L]^{2+}$	2073.6	2074.2	0.7			
RHF A6	1863.2	1864.4	1.1	1862.7	12	5,4,3,2
$[SS\ DNA]^{2+}$	1824.2	1824.8	1.3			
RHF A6–PO_3	1784.2	1785.6	0.1	1782.7	0.5	3,2
LHF A6 + PO_3	1764.2	1766.2	2.4			
LHF A6 β-elim + $[Na^+]$	1688.7			1687.7	0.3	2
LHF A6 β-elim	1666.2	1667.5	0.6	1665.7	3.6	3,2
LHF A6–sugar–PO_3	1488.0			1487.5	0.4	2
LHF A5–sugar	1254.8			1254.4	1.2	3,2
L-OH	517.7			517.4	0.8	1
L-G	650.7	651.8	1.9	650.6	4	+2
L-A	634.7	635.5	9.8	634.0	59	+2
L-OH	517.7	518.2	5.2	517.3	37	+1

DS, double-stranded; UA, unassigned peaks; deG, deguanylated; β-elim, β-elimination reaction.

DNA-related peaks were observable in the positive ion ESI spectrum for this DNA-ligand pair, or for the alkamin complex, at the usual scale expansions.

MALDI Mass Spectrum of the Alkamin-A2T2 Complex. The positive ion MALDI mass spectrum of the alkamin-A2T2 complex is presented in Fig. 3a; the observed peaks and their molecular assignments are listed in Table 2. Fig. 3a is the full spectrum from *m/z* 600 to 8500 Da; Fig. S5, a to e, is the expanded portion. As with the MALDI analysis of the alkamini-DNA complex, we find the peaks fall into five major groups arranged by molecular mass. The peaks observed at *m/z* 7291.7, *m/z* 7174.6, and *m/z* 7057.5 in Fig. S5a represent duplex DNA and its deadenylated and doubly deadenylated derivatives, the latter arising from depurination events at alkylated adenines. However, in distinction to the alkamini complex, we do not see peaks representing duplex-ligand adducts. In the *m/z* range 3600 to 4500 Da, Fig. S5, b and c, we find peaks representing single-stranded DNA and its Na^+ and K^+ ion pairs, as well as ion pairs formed with charged hydrolyzed ligand ($[HO\text{-}HL\text{-}OH]^+$, *m/z* 4250.8) and with a protonated ligand-adenine base adduct ($[A\text{-}HL\text{-}A]^+$, *m/z* 4485.2). Most significantly, in this mass range we observe

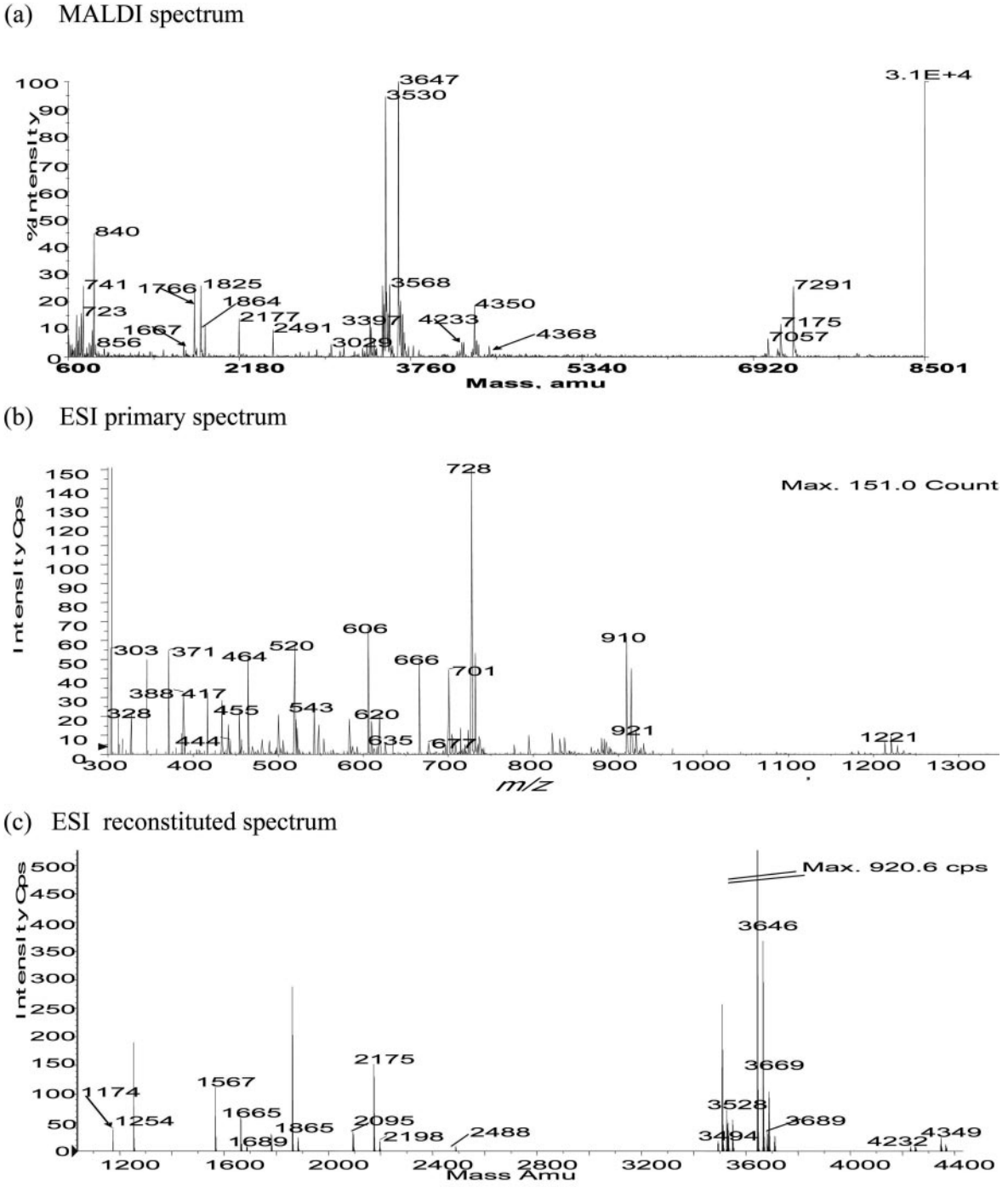
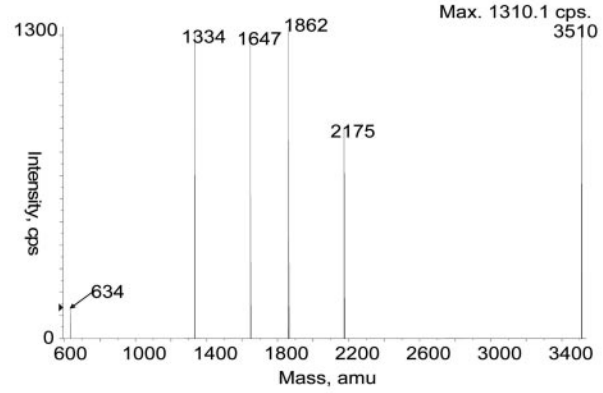


Fig. 3. Mass spectra for the 1:1 alkamin-A2T2 complex. Positive ion MALDI spectrum, *m/z* 600 to 8500 (a), negative ion ESI primary spectrum, *m/z* 300 to 1300 (b), and Bayesian reconstituted negative ion ESI spectrum, *m/z* 1100 to 4400 (c).

peaks representing single stranded DNA-ligand adducts that fall into two classes: the peak at *m/z* 4233.0 betokens monofunctional alkylation in which the second mustard arm of the ligand is hydrolyzed (SS-L-OH), and the peaks at *m/z* 4349.9 and 4368.2 represent bifunctional reaction where the second mustard group is attached to adenine (SS-L-A) and guanine (SS-L-G) respectively. These latter species necessarily derive from a depurination event at one arm of interstrand cross-linked complexes. The spectrum also contains an alkali metal salt of one of these species, either the Na^+ salt of SS-L-G or the K^+ salt of SS-L-A; the assignment is ambiguous because of the identity of the calculated molecular masses. The assignments of SS-L-A and SS-L-G are confirmed by MS/MS ESI measurements on the parent ions as shown in Fig. 5a (SS-L-A, *m/z* 868.9, z = −5), and Fig. 5b (SS-L-G, *m/z* 872.5, z = −5). SS-L-A yields product ions corresponding to *SS deA* and *SS deG*, *RHFs* for A5 and A6, *LHFs* for G4 and A5, and the base-ligand adduct A-L-A, which, taken together, indicate the formation of interstrand cross-links between adenine A5′ or A6′ and G4, A5, and A6. Conversely, the tandem measurements with SS-L-G give rise to daughter ions for the *RHFs* of A5 and A6, revealing the G4 interstrand cross-link to involve reaction with both adenines A5′ and A6′.

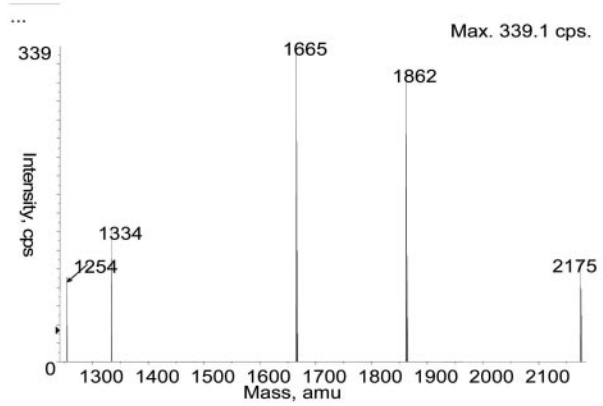
In the depurinated DNA mass range, *m/z* 3000 to 3600 (Fig. S5c), we find peaks representing deadenylated single-stranded DNA, *m/z* 3530.2, and its Na^+ and K^+ salts, degua-

(a) MS/MS of SS-L



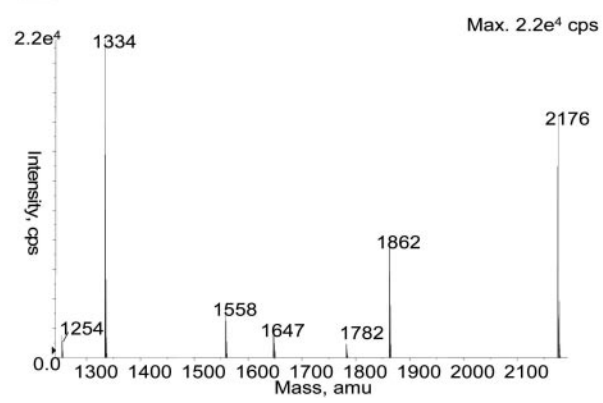
Assignment	Mass (M^-)	Area %
SS deA	3510.6	29.8
RHF A5(a)	2176.4	14.6
RHF A6(a)	1862.8	21.5
LHF A6(a)	1647.8	17.4
LHF A5(a)	1334.6	15.4
L-A	634.6	1.31

(b) MS/MS of SS deA



Assignments	Mass (M^-)	Area %
RHF A5(a)	2175.9	10.3
RHF A6(a)	1863.0	31.8
LHF A6(b)	1665.9	32.3
LHF A5(b)	1334.5	10.2
LHF A5-sugar	1254.4	5.31
SS deA	881.0	10.2

(c) MS/MS of SS deA β-elim



Assignments	Mass (M^-)	Area %
RHF A5(a)	2176.4	35.0
RHF A6(a)	1863.3	17.0
RHF A6(c)	1783.3	1.8
LHF A6(a)	1648.3	3.7
UA	1559.1	7.0
LHF A5(b)	1335.0	32.0
LHF A5 -sugar	1254.9	2.3
Parent ion	701.1	2.3

Fig. 4. Negative ion ESI MS/MS spectra for the 1:1 alkamini-A2T2 complex. Parent ion is the SS-L species, *m/z* 1035.3, z = −3 (a); parent ion is SS deA, *m/z* 881.0, z = −4 (b); and parent ion is SS deA β-elim, *m/z* 701.2, z = −5 (c). For species assignments, see Fig. S3.

nylated DNA, *m/z* 3514.6, and doubly depurinated DNA that has lost both a guanine and an adenine base to hydrolysis, *m/z* 3397.2. In the DNA fragmentation portion of the spectrum, Fig. S5d, we observe clear RHF signals for covalent reaction at guanine G4, *m/z* 2490.7, adenine A5, *m/z* 2177.0, and adenine A6, *m/z* 1864.2. In addition, we see the cognate LHFs for reaction at the adenines (A5 *m/z* 1570.0, A6 *m/z* 1685.4), their dehydrated β-elimination products (A5 β-elim *m/z* 1352.0, A6 β-elim *m/z* 1667.3), the LHF of A5 lacking its apurinic sugar (*m/z* 1254.6) and the LHF of A6 bearing a 5′ phosphate (*m/z* 1766.2). In contrast, the LHF of G4 and its derivatives are absent from the spectrum. Figure S5e shows the low molecular mass region where a variety of ligand-base adducts resulting from depurination events can be seen. Once again, these species fall into two groups: those that derive from monofunctional alkylation at guanine and adenine bases (A-L-OH, *m/z* 722.6; G-L-OH, *m/z* 739.9), and those that represent bifunctional reaction involving either two adenines, A-L-A, *m/z* 839.8, or an adenine and a guanine, A-L-G, *m/z* 855.8. In addition, there is also evidence for hydrolyzed unreacted alkamin, HO-L-OH, *m/z* 605.2.

ESI Mass Spectrum of the Alkamin-A2T2 Complex. The primary negative ion ESI mass spectrum of the alkamin-A2T2 complex is shown in Fig. 4b, and the masses of the

TABLE 2

Molecular assignments of the ions observed in the ESI and MALDI spectra of the 1:1 alkamin-A2T2 DNA complex

The calculated masses are for the neutral species, the MALDI data were acquired in the positive ion mode, and the ESI data represent the mass of singly-charged negative ions derived from the Bayesian transformation of the primary electrospray spectrum. The charged states involved in the transformations are indicated. The nomenclature for the species assignments is given in Table S1 and Figs. S1 and S2.

Species Assignments	Mass (M^0) Calculated	MALDI (M^+)		ESI (M^-)		
		Observed	Area	Observed	Area	Charges
			%		%	
DS DNA	7292.8	7292.1	3.8			
DS deA	7175.8	7175.2	1.5			
DS deA deA	7058.8	7058.2	0.9			
SS DNA + [A-HL-A]$^+$	4485.1	4485.2	0.6			
SS-L-G + [Na^+] *	4389.3	4387.1	0.6			
SS-L-G	4367.3	4368.2	1.0	4367.6	0.1	6
SS-L-A	4351.3	4349.9	2.3	4349.2	0.5	5,4
SS DNA + [HO-HL-OH]$^+$	4250.6	4250.8	0.6	4251.1	0.3	5
SS-L-OH	4234.2	4233.0	0.8	4232.1	0.1	5
SS DNA + 3[Na^+]	3711.7			3711.4	1.1	5,4
UA		3701.7	1.0			
SS DNA + 2[Na^+]	3690.4			3689.5	3.8	7,6,5,4,3
SS DNA + [K^+]	3684.4	3685.4	1.6	3683.5	1.5	7,6,5,4,3
SS DNA + [Na^+]	3668.5	3668.8	3.1	3667.5	15.1	8,7,6,5,4,3
SS DNA	3646.4	3647.4	18.0	3645.4	33.8	8,7,6,5,4,3
SS deA + [K^+]	3567.4	3568.4	3.1			
SS deA + [Na^+]	3551.4	3552.4	1.7	3550.5	1.8	6,5,4,3
SS deA β-elim + [Na^+]	3533.5			3533.1	2.1	7,6,5,4
SS deA	3529.4	3530.2	16.0	3528.4	2.5	7,6,5,4
SS deG	3513.4	3514.6	2.8			
SS deA β-elim	3511.3			3510.7	8.5	8,7,6,5,4
SS deG β-elim	3495.3			3494.3	0.3	7
UA		3422.7	0.7			
SS deA deG	3396.3	3397.2	1.3			
UA		3384.8	1.7			
RHF G4	2489.6	2490.7	1.0	2489.0	0.2	4
RHF A5 + [Na^+]	2198.4			2198.0	0.3	5
RHF A5	2176.4	2177.0	1.5	2175.9	5.0	5,4,3,2
RHF A5–PO_3	2096.4			2096.1	0.5	4
RHF A6 + [Na^+]	1885.2			1885.1	0.4	2
RHF A6	1863.2	1864.2	1.1	1862.8	8.5	5,4,3,2
[SS DNA]2+	1824.2	1824.7	2.8			
RHF A6–PO_3	1783.2			1782.8	0.7	3,2
LHF A6 + PO_3	1764.2	1766.2	3.0			
LHF A6 β-elim + [Na+]	1688.1			1688.0	0.2	3,2
LHF A6	1684.1	1685.4	0.2			
LHF A6 β-elim	1666.1	1667.3	0.5	1665.8	1.2	3
LHF A6-sugar	1568.1			1567.6	1.0	3
UA		1480.1	0.3	1478.7	0.5	2
LHF A5	1371.0	1371.7	0.1			
LHF A5 β-elim	1353.0	1354.1	0.1			
LHF A5–sugar	1254.0	1254.6	0.2	1254.5	2.4	4,3
LHF A5–sugar-PO_3	1174.8			1174.7	0.4	3
A-L-G	855.0	855.8	0.5			
A-L-A	838.7	839.8	9.1	838.7	11.8	+2
UA				754.4	12.3	+2,+1
UA				732.3	18.6	+2,+1
G-L-OH	737.9	739.9	5.4	738.4	31.1	+2,+1
A-L-OH	721.9	722.6	2.8	721.7	26.1	+3,+2
UA		678.1	3.0			
HO-L-OH	604.2	605.3	0.7			

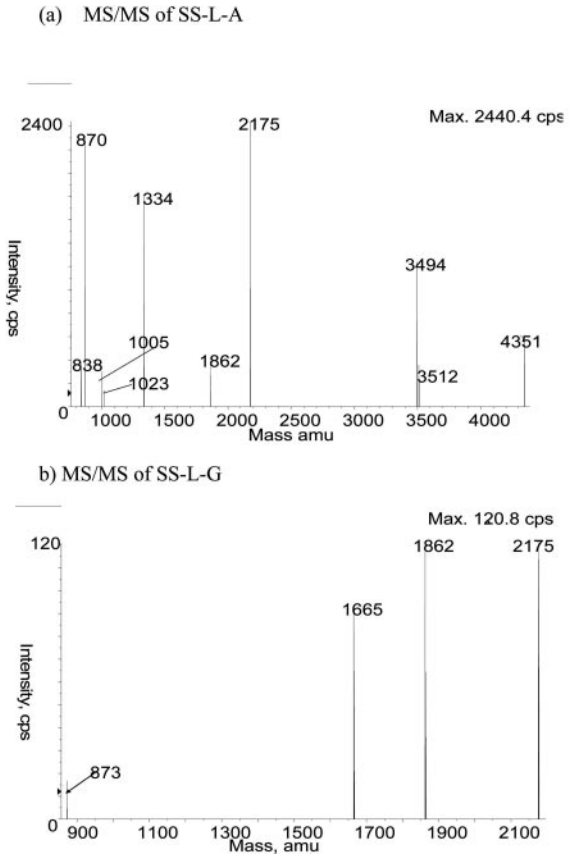
DS, double-stranded; UA, unassigned peaks; deG, deguanylated; β-elim, β-elimination reaction.

observed peaks, their assignments, and charged states are listed in Table 2. Figure 4c is the complete reconstituted spectrum over the molecular mass range 700 to 4400 Da. Once again, we find that the observed peaks fall into four main classes based on the nature of their molecular identity. As found with the alkamini complex, peaks representing duplex species are missing from the ESI spectrum, and the most notable highest mass ions derive from single stranded DNA ligand complexes (Fig. 4c). In accord with the MALDI data, there is clear evidence for monofunctional reaction to yield SS-L-OH adducts at *m/z* 4232.1 and for bifunctional reaction giving the species SS-L-A at *m/z* 4349.2 and SS-L-G at *m/z* 4369.7. Again, these peaks generally appear with masses 2 units below the expected value as a result of the requirement to lose two protons to yield a negative ion from the quaternized complexes (Table 2). This is not the case for the SS-L-G adducts, which fact we attribute to an artifact resulting from the lack of inclusion of low-intensity isotope peaks that fall below the experimental cut-off. Within the high molecular mass grouping, we again see single-stranded DNA and its Na^+, K^+, and protonated hydrolyzed ligand adducts (Fig. 4c), together with a variety of depurinated species. Most abundant among the latter are deadenylated DNAs, *m/z* 3528.4, their β-elimination products, *m/z* 3517.0, and their sodium adducts.

Reaction at guanine is revealed by the presence of the dehydration product of deguanylated DNA, *m/z* 3494.3, but unlike the MALDI data, there is no direct evidence for a doubly depurinated species. In the middle mass range (Fig. 4c), we observe RHFs for reaction with guanine G4, *m/z* 2489.0, adenine A5, *m/z* 2175.9, and adenine A6, *m/z* 1862.8, and, in distinction to the MALDI data, we note that hydrolysis at the deadenylated sites also results in the alternate cleavage of the phosphate group (RHF A5-PO_3, *m/z* 2096.1; RHF A6-PO_3, *m/z* 1782.8). With respect to the cognate LHFs (Fig. 4c), we find evidence for the β-elimination product for alkylation of adenine A6, *m/z* 1665.8, its Na^+ salt, *m/z* 1688.0, and its fragmentation product lacking the apurinic sugar, *m/z* 1567.6, but no ions relating to reaction at G4. Alkylation at adenine A5 yields an LHF lacking the apurinic sugar, *m/z* 1254.5, and one that has lost both sugar and the 3′ phosphate group, *m/z* 1174.7, to give CGCG. Although ligand-base adducts are absent from the negative ion spectrum, they are visible in the positive ion spectrum, shown in the Bayesian reconstituted form in Fig. S6b, although their intensity is substantially lower than that found for alkamini adducts in this circumstance. In contrast to the MALDI data, there are no peaks associated with free hydrolyzed ligand or with the cross-linked adenine-ligand-guanine adduct, but the peaks representing monofunctional reaction with guanine, G-L-OH, *m/z* 738.4, and adenine, A-L-OH *m/z* 721.7, are in abundance, as is the peak describing the cross-linked adenine-ligand-adenine adduct, A-L-A *m/z* 838.7. In addition, there are two major unidentified peaks at *m/z* 754.4 and 732.3 that seem to be associated with alkamin-base fragments.

Discussion

We have used ESI and MALDI mass spectrometry to characterize oligonucleotide complexes of alkamin and alkamini to help define the relationships between DNA binding properties and antitumor activity for polybenzamide mustards. Because the use of these methods is novel in studies of



Assignments	Mass (M^-)	Area %
SS-L-A	4349.2	10.4
SS deA	3511.3	4.8
SS deG	3494.5	9.9
RHF A5(a)	2176.0	32.0
RHF A6(a)	1862.7	4.0
LHF A5(a)	1334.6	15
LHF G4(b)	1023.4	1.0
LHF G4(a)	1005.4	2.1
Parent ion	870.7	17.0
A-L-A	838.9	3.7

Assignments	Mass (M^-)	Area %
RHF A5(a)	2175.4	41.7
RHF A6(a)	1862.4	30.6
LHF A6(b)	1665.3	24.8
Parent ion	872.7	2.95

Fig. 5. Negative ion ESI MS/MS spectra for the 1:1 alkamin-A2T2 complex. Parent ion is the SS-L-A species, *m/z* 868.9 z = −5 (a); parent ion is SS-L-G, *m/z* 872.5 z = −5 (b). For species assignments, see Fig. S3.

covalent oligonucleotide-drug complexes, our work also provides an assessment of the suitability of each approach. We find that positive ion MALDI and Bayesian-reconstituted negative ion ESI spectra have the same general form and contain sufficient information to construct a detailed model of the intact DNA-ligand complex. Both spectra contain ions that derive from a binding scheme in which the ligand alkylates an adenine or guanine in the GAATTC segment, thereby quaternizing it. The DNA then undergoes a succession of hydrolytic events starting with depurination followed by phosphate cleavage, the resulting DNA fragments revealing the site of alkylation (see Figs. S1 and S2 and Table S1). We find that interstrand covalently cross-linked duplexes resulting from reaction with alkamin are unstable in both MALDI and ESI. In both cases, one arm of the cross-link undergoes a depurination event, yielding a single-stranded DNA bearing a ligand attached to a base (Table 2, Figs. 3, 5, and S5): an ion that becomes a powerful hallmark for interstrand cross-link formation that can lead to identification of the sites of attachment to each strand. Our findings enable a detailed description of the nature of DNA-polybenzamide mustard adducts and have provided insights into the molecular determinants of the antitumor activity of alkamin.

Structure of the Alkamini-A2T2 Complex. We conclude from the observed peaks representing the ligand-base adducts, the collection of phosphate hydrolysis fragments, the depurinated DNAs, and the intact single and double-stranded DNA-ligand adducts (Table 1, Figs. 2 and S4) that alkamini alkylates G4, A5, and A6. Reaction with the adenines is directly confirmed by the MS/MS measurements (Fig. 4). This conclusion accords with the notion that the ligand binds reversibly to the minor groove in the AATT segment, and that it shuffles up and down, alkylating the N3 atoms of the GAA section. Thus, at the completion of reaction, alkamini is distributed between DNA molecules to give a population of three distinct 1:1 complexes, as shown in Fig. 6a, molecular models indicating that the binding site extends over four base pairs. Taking the relative normalized peak areas of the ligand-base adducts as an indication of the gross reactivity at adenines or guanines, alkamini seems to have a 5- to 10-fold preference for reaction at adenines A5 plus A6 compared with guanine G4. The MALDI analysis provides evidence for additional alkamini complexes involving two bound ligand molecules per duplex: deguanylated and deadenylated single-stranded DNA (Table 1, Fig S4c) and a deadenylated single-stranded DNA-ligand complex (Table 1, Fig. S4b). Both species are consistent with two alkamini molecules reacting with the same DNA strand to alkylate an adenine and a guanine, presumably G4 and A5 or A6. Similar reasoning suggests that a doubly deadenylated duplex DNA ion (Table 1, Fig. S4a) indicates the presence of another 2:1 complex in which an alkamini molecule is bound to an adenine in each strand, presumably A5 or A6 and A5′ or A6′. The existence of two classes of 2:1 complexes is further supported by a deadenylated duplex-ligand adduct (Table 1, Fig. S4a), which could be derived from either form. Figure 6b illustrates some possibilities for the structure of these 2:1 complexes, in which ligand molecules bind in a head-to-head fashion without intermolecular overlap. We have rejected models involving head-to-tail arrangements because they necessarily involve ligand-ligand interactions. We suppose that the first molecule to bind initially occupies the AT tract, from where it alkylates a purine, and that thermal motion displaces its polybenzamide moiety from the AT segment, allowing reaction of the second ligand in the vacated narrow minor groove. For the intrastrand G/A complex, at equilibrium, one ligand binds to G4 extending into the GC-region, and the other binds at A5 or A6, thus filling the AT tract. For the interstrand A/A complex, the most likely arrangement is reaction with A5 and A5′, but other dispositions are possible. The occurrence of 2:1 alkamini complexes in a reaction mixture of 1:1 stoichiometry implies the existence of some degree of cooperativity in the binding mechanism.

Structure of the Alkamin-A2T2 Complex. There is no evidence for 2:1 alkamin/A2T2 complexes in the mass spectra, whereas it is clear that there is an assortment of monofunctional and bifunctional adducts involving both adenines and guanines. There is abundant confirmation that alkamin forms monofunctional adducts with G4, A5, and A6, as confirmed by the presence of guanine and adenine base-ligand adducts in which the second mustard is hydrolyzed, by a variety of RHFs and LHFs for reaction with G4, A5, and A6, by the presence of deadenylated and deguanylated single-stranded DNAs and their dehydration products, by the finding of intact single-stranded DNA and its adduct bearing a hydrolyzed ligand, and by the observation of deadenylated

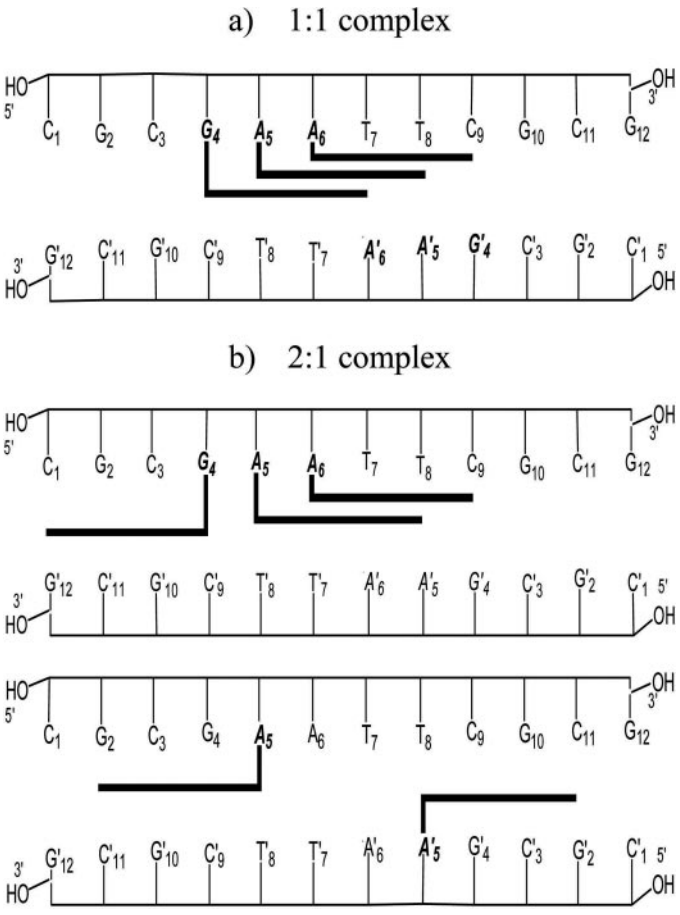


Fig. 6. Schematic diagram of modes of alkamini binding to A2T2 DNA. a, 1:1 covalent complex. Here alkamini binds as a 1:1 complex, either to G4, A5, or A6. The data indicate that all sites are populated. b, 2:1 covalent complexes. In the intrastrand complex, alkamini binds to G4 and either A5 or A6; the data cannot distinguish which adenine is involved. In the interstrand complex, the ligand is bound to an adenine in each strand, most likely to the A5/A5′ pair.

duplex DNA (Table 2, Figs. 3, 5, and S5). Taking these data together, we conclude that alkamin forms monofunctional adducts with A2T2 DNA in the same manner as alkamini, in which alkylation occurs at G4, A5, and A6 in the minor groove as shown in Fig. 7a. Unfortunately, it was not possible to confirm the assignment of monofunctional reaction at these three purines by tandem measurements because the intensity of the parent SS-L-OH ion was too weak.

Alkamin interstrand cross-links are made evident by several criteria, including the appearance of the ligand-base adducts A-L-A and A-L-G derived from bifunctional alkylation of two bases; by the observation of the intact single-stranded DNA-ligand adducts SS-L-A and SS-L-G, in which the second reactive arm is bound to an adenine or a guanine base; and the finding of doubly deadenylated duplex DNA, indicating reaction with an adenine in each DNA strand (Table 2, Figs. 3, 5, and S5). Given that both ESI and MALDI spectra indicate that alkylation is confined to reaction with G4, A5, and A6 (Table 2), we conclude that the alkamin interstrand cross-links must necessarily involve these same bases. Therefore, for the guanine-adenine interstrand cross-link, we propose that the ligand is attached to G4 and A6′ and A5′ in the complementary strand as shown in Fig. 7(b). This assertion is substantiated by the MS/MS measurements on the SS-L-G adduct, which shows attachment to the single stranded DNA at both adenines, and by similar measurements on the SS-L-A fragment confirming reaction with G4 (see Fig. 5). These findings place the polybenzamide moiety toward the 5′ end of the AT tract, providing cross-links between bases 4 (designated 1→4) and 5 (designated 1→5) nucleotides apart. Molecular models suggest that the latter is the maximum distance by which the carbon atoms carrying the mustard chlorides can be separated, when the molecule is fixed in a banana-shaped planar conformation suitable for optimal occupation of the minor groove. We note that the guanine-adenine interstrand cross-link forms two well defined complexes. By contrast, identifying precisely which bases are involved in the adenine-adenine interstrand cross-link is more circumspect. Molecular models suggest that by rotating about the two mustard nitrogen-phenyl carbon bonds, alkamin can form 1→5, 1→4, 1→3, and possibly 1→2 (equivalent to 2→3 within the binding site) interstrand cross-links when confined to the optimum minor groove binding conformation, because there are three potential ways of aligning the mustard groups, as shown in Fig. 7, c to e. Thus, we conclude that the data are consistent with formation of a

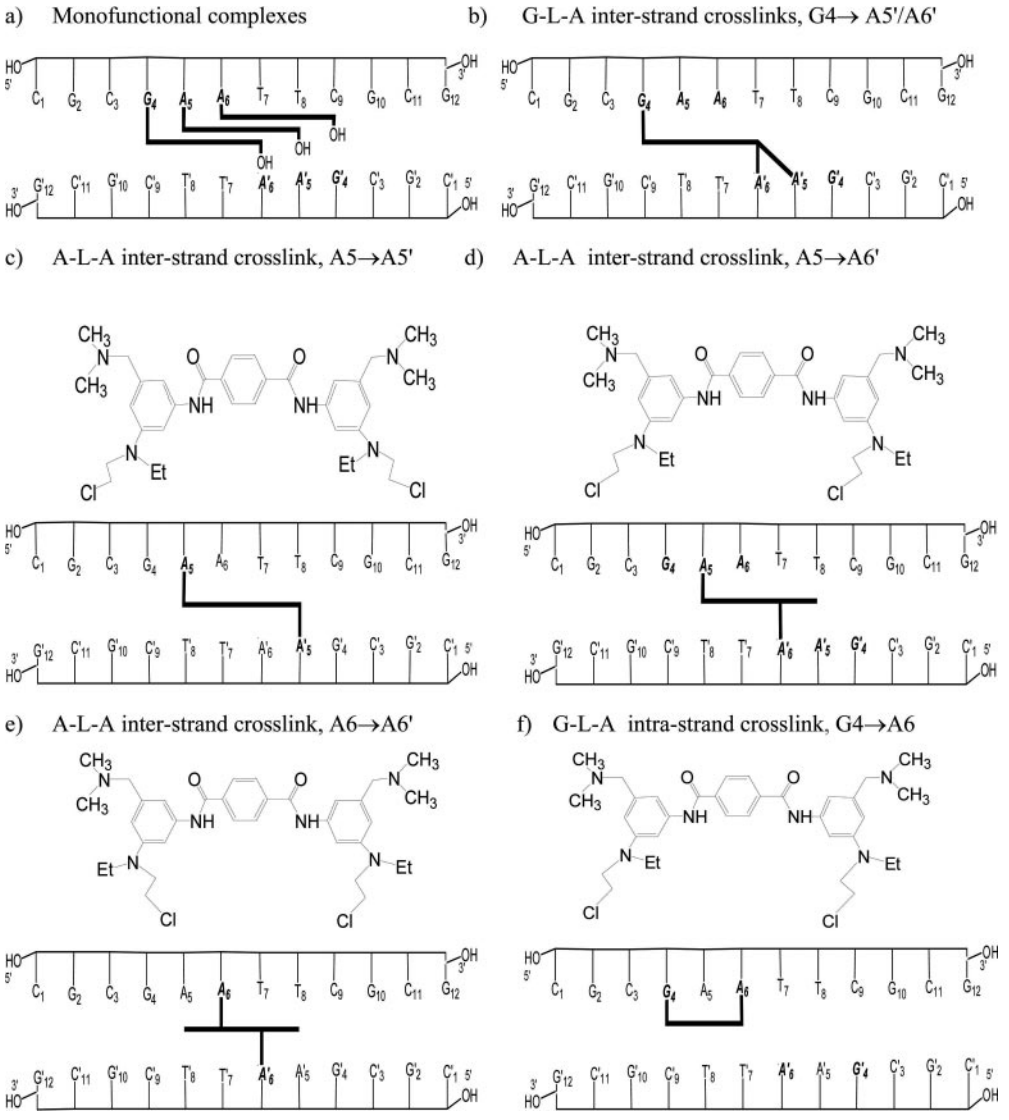


Fig. 7. Schematic diagram of modes of alkamin binding to A2T2 DNA. a, Monofunctional complexes. Here alkamin binds monofunctionally to either G4, A5, or A6. The data indicate that all sites are populated and that the second mustard is hydrolyzed. b, interstrand cross-link between G4 and A5 or A6. MS/MS data indicate that both A5 and A6 are involved in the cross-link. c to e, interstrand cross-links involving adenines. The data indicate that both A5 (A5′) and A6 (A6′) are involved in the adenine-ligand-adenine cross-links. Molecular modeling suggests that all three distinguishable A-ligand-A attachments are sterically possible because of conformational flexibility about the mustard nitrogen-phenyl bond. f, intrastrand guanine-ligand-adenine cross-link. The data are consistent with an intrastrand cross-link between G4 and A6.

1→4 cross-link between A5 and A5′, a 1→3 cross-link between A5 and A6′, and a 1→2 (equivalent to 2→3) cross-link between A6 and A6′. The MS/MS measurements on SS-L-A support these assertions and show that both A5 and A6 are involved in the adenine-ligand-adenine cross-links (Fig. 5).

The appearance of a doubly depurinated single-stranded DNA that has lost both an adenine and a guanine (Table 2, Figs. 3a and S5c), we take as evidence that the ligand can also form intrastrand cross-links involving these bases. Given the above molecular interpretation, we propose that in this complex, alkamin forms a 1→3 cross-link between G4 and A6, as shown in Fig. 7f. Direct evidence for such an intrastrand cross-link is provided by a MALDI peak (data not shown) at *m/z* 4216.2 identified as single-stranded DNA bearing a ligand fragment lacking both chlorines, in complexes in which the input stoichiometry is raised to 2:1, conditions at which peak areas are enhanced without forcing additional covalent complexes.

Distribution between Bound Species for Alkamin. Taking the relative normalized peak areas of the ligand-base adducts as an indication of the gross reactivity at adenines and guanines, the MALDI data for the monofunctional alkamin adducts suggests that there is approximately twice as much reaction at G4 as there is at A5 and A6 combined, because G-L-OH represents approximately 25% of the observed total Pu-L-X species and A-L-OH is approximately 15% (proportions derived from Table 2). This is at first sight surprising, until it is realized that most (50%) of the ligand is found in A-L-A interstrand cross-links, with a minor component (10%) in the form of A-L-G cross-links, some of which is interstrand and some of which is intrastrand (proportions derived from Table 2). Thus a little over half the bound alkamin is found in cross-linked complexes, the vast majority being interstrand. All in all, approximately one third of the bound alkamin is located on guanine G4.

Comparison with Previous Studies. The sequencing studies of Prakash et al. (1991) and Turner et al. (1999, 2000), did not identify AATT among the most occupied alkamin sites. Indeed, Turner et al. (1999) show no such sites, and Prakash et al. (1991) provide only one, at a CAATTT sequence. However, both studies show multiple examples of reaction with A_mT_n tracts, many having the capacity for the formation of adenine-adenine interstrand cross-links. One possible reason for the lack of appearance of AATT in these studies is incomplete reaction for kinetic reasons, because Atwell et al. (1995) show that alkamin alkylates calf thymus DNA with a half-life of 3.5 h at 37°C, a value some 3.5- to 7-fold longer than the incubation period used in the sequencing experiments. Although previous studies provide evidence using linearized pBR 322 DNA that alkamin causes interstrand cross-links, none was able to characterize its nature (Prakash et al., 1991; Atwell et al., 1995; Turner et al., 2000). Turner et al. (2000) attempted to determine whether cross-linking is limited to binding at AT tracts by challenging alkamin binding with Hoechst 33285. In general, they conclude that it is not, but whatever may be the correct interpretation of their findings, the mass spectrometry data unequivocally demonstrate that alkamin forms interstrand cross-links at AATT sequences. Prakash et al. (1991) estimated the cross-link frequency as 1 per 10 alkamin molecules bound, which compares with an interstrand cross-link frequency of one per two bound ligand molecules for reaction with A2T2 DNA (this work). We conclude from this that the cross-link frequency for alkamin binding at sites with appropriately disposed adenines is much higher than previously realized. Because alkamin, but not alkamini, has experimental antitumor activity (Prakash et al., 1991; Atwell et al., 1995) and is much more cytotoxic (Atwell et al., 1995), we speculate that the formation of such cross-linked species lies at the heart of its biological activity. If so, the mass spectrometry studies indicate the intricacy of the potential interactions of alkamin with AT tract sequences and highlight the difficulty of identifying which particular cross-links form the cytotoxic lesion.

Last, we comment on the possible extension of these measurements to probe minor groove binding alkylator adducts in intact living cells. In principle, there is no bar to this except for the issue of sensitivity. Successful in vivo investigations would require application of the most sensitive mass spectrometers, perhaps Fourier transform instruments (Marshall et al., 1998), and the targetting of an appropriate isolated AT tract bounded by two restriction sites spaced 12 to perhaps 30 base pairs apart. To ensure sufficient material for analysis, it may also be necessary for the target to be available in multiple copies such as might be found in repetitive DNA sequences like tandem and interspersed repeats (Dewannieux and Heidmann, 2005). By such means, it should be possible to identify alkamin binding sites in vivo and to follow their evolution and repair.

Acknowledgments

We thank Dr. Mark Raftery for technical discussions concerning the mass spectrometers used, and Professor Margaret Sheils for initial discussions concerning electrospray mass spectrometry of DNA-ligand complexes.

References

Atwell GJ, Yaghi BM, Turner PR, Boyd M, O'Connor CJ, Ferguson LR, Baguley BC, and Denny WA (1995) Synthesis, DNA interactions and biological activity of DNA minor groove targeted polybenzamide-linked nitrogen mustards. *Bioorg Med Chem* **3:**679–691.

Bellorini M, Moncollin V, D'Incalci M, Mongelli N, and Mantovani R (1995) Distamycin A and tallimustine inhibit TBP binding and basal *in vitro* transcription. *Nucleic Acids Res* **23:**1657–1663.

Broggini M, Coley HM, Mongelli N, Pesenti E, Wyatt MD, Hartley JA, and D'Incalci M (1995) DNA sequence-specific adenine alkylation by the novel antitumor drug tallimustine (FCE 24517), a benzoyl nitrogen mustard derivative of distamycin. *Nucleic Acids Res* **23:**81–87.

Denny WA (2001) DNA minor groove alkylating agents. *Curr Med Chem* **8:**533–544.

Dewannieux M and Heidmann T (2005) LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. *Cytogenet Genome Res* **110:**35–48.

Drablos F, Feyzi E, Aas PA, Vaagbo CB, Kavli B, Bratlie MS, Pena-Diaz J, Otterlei M, Slupphaug G, and Krokan HE (2004) Alkylation damage in DNA and RNA—repair mechanisms and medical significance. *DNA Repair (Amst)* **3:**1389–1407.

Gamcsik MP, Dolan ME, Andersson BS, and Murray D (1999) Mechanisms of resistance to the toxicity of cyclophosphamide. *Curr Pharm Des* **5:**587–605.

Gravatt GL, Baguley BC, Wilson WR, and Denny WA (1991) DNA-directed alkylating agents. 4. 4-anilinoquinoline-based minor groove directed aniline mustards. *J Med Chem* **34:**1552–1560.

Iannitti P and Sheil MMWG (1997) High sensitivity and fragmentation specificity in the analysis of drug-DNA adducts by electrospray tandem mass spectrometry. *J Am Soc Mass Spectrom* **119:**1490–1491.

Iannitti-Tito P, Weimann A, Wickham G, and Sheil MM (2000) Structural analysis of drug-DNA adducts by tandem mass spectrometry. *Analyst* **125:**627–633.

Kohn KW, Hartley JA, and Mattes WB (1987) Mechanisms of DNA sequence selective alkylation of guanine-N7 positions by nitrogen mustards. *Nucleic Acids Res* **15:**10531–10549.

Marshall AG, Hendrickson CL, and Jackson GS (1998) Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* **17:**1–35.

Mattes WB, Hartley JA, and Kohn KW (1986a) DNA sequence selectivity of guanine-N7 alkylation by nitrogen mustards. *Nucleic Acids Res* **14:**2971–2987.

Mattes WB, Hartley JA, and Kohn KW (1986b) Mechanism of DNA strand breakage by piperidine at sites of N7-alkylguanines. *Biochim Biophys Acta* **868:**71–76.

Panasci L, Paiement JP, Christodoulopoulos G, Belenkov A, Malapetsa A, and Aloyz R (2001) Chlorambucil drug resistance in chronic lymphocytic leukemia: the emerging role of DNA repair. *Clin Cancer Res* **7:**454–461.

Prakash AS, Valu KK, Wakelin LP, Woodgate PD, and Denny WA (1991) Synthesis

and anti-tumour activity of the spatially-separated mustard bis-N,N′-[3-(N-(2-chloroethyl)-N-ethyl)amino-5-[N,N-dimethylamino)methyl)-aminophenyl]-1,4-benzenedicarbox amide, which alkylates DNA exclusively at adenines in the minor groove. *Anticancer Drug Des* **6:**195–206.

Rossi R, Montecucco A, Capolongo L, Mezzina M, Chevallier-Lagente O, Sarasin A, and Ciarrocchi G (1996) The alkylating antitumor drug tallimustine does not induce DNA repair. *Anticancer Res* **16:**3779–3783.

Tagliabue G, Filippini C, Ubezio P, and D'Incalci M (1997) Combination of the new minor groove alkylator tallimustine and melphalan. *Eur J Cancer* **33:**284–287.

Turner PR, Ferguson LR, and Denny WA (1998) Binding of polybenzamides to DNA: studies by DNase I and chlorambucil interference footprinting and comparison with Hoechst 33258. *Anticancer Drug Des* **13:**941–954.

Turner PR, Denny WA, and Ferguson LR (2000) Role of DNA minor groove alkylation and DNA cross-linking in the cytotoxicity of polybenzamide mustards. *Anticancer Drug Des* **15:**245–253.

Turner PR, Ferguson LR, and Denny WA (1999) Polybenzamide mustards: structure-activity relationships for DNA sequence-specific alkylation. *Anticancer Drug Des* **14:**61–70.

Wickham G, Iannitti P, Boschenok J, and Sheil MM (1995) The observation of a hedamycin-d(CACGTG)2 covalent adduct by electrospray mass spectrometry. *FEBS Lett* **360:**231–234.

Address correspondence to: Dr. Amin Malik Shah Abdul Majid, School of Pharmacy, University of Science Malaysia, Minden, 11800 Penang, Malaysia. E-mail: aminmalikshah@usm.my