

The DNA sequence specificity of bleomycin cleavage in a systematically altered DNA sequence

Shweta D. Gautam¹ · Jon K. Chen¹ · Vincent Murray¹

Received: 7 February 2017 / Accepted: 11 May 2017 / Published online: 16 May 2017
© SBIC 2017

Abstract Bleomycin is an anti-tumour agent that is clinically used to treat several types of cancers. Bleomycin cleaves DNA at specific DNA sequences and recent genome-wide DNA sequencing specificity data indicated that the sequence 5'-RTGT*AY (where T* is the site of bleomycin cleavage, R is G/A and Y is T/C) is preferentially cleaved by bleomycin in human cells. Based on this DNA sequence, we constructed a plasmid clone to explore this bleomycin cleavage preference. By systematic variation of single nucleotides in the 5'-RTGT*AY sequence, we were able to investigate the effect of nucleotide changes on bleomycin cleavage efficiency. We observed that the preferred consensus DNA sequence for bleomycin cleavage in the plasmid clone was 5'-YYGT*AW (where W is A/T). The most highly cleaved sequence was 5'-TCGT*AT and, in fact, the seven most highly cleaved sequences conformed to the consensus sequence 5'-YYGT*AW. A comparison with genome-wide results was also performed and while the core sequence was similar in both environments, the surrounding nucleotides were different.

Keywords Bleomycin · Cleavage sites · DNA damage · DNA sequence specificity · Consensus sequence · Plasmid vs genome-wide

Abbreviations

CE-LIF Capillary electrophoresis with laser-induced fluorescence

dNTP Deoxynucleotide triphosphates
Exo-SAP Exonuclease I and shrimp alkaline phosphatase
FAM-dUTP Aminoallyl-dUTP-6-FAM

Introduction

Bleomycin was originally isolated from *Streptomyces verticillus* by Umezawa et al. [1] and belongs to a group of water-soluble glycopeptides. Bleomycin has potent tumour destroying properties and is an important drug in cancer chemotherapy [2, 3]. Bleomycin is isolated as a copper complex but is administered to patients in the metal-free form, which reduces irritation at the injection site [4–6]. The drug binds to Cu(II) in the blood plasma and is transported into the cell. The bleomycin–Cu(II) complex is reduced to bleomycin–Cu(I) and is then displaced by Fe(II) [7]. Bleomycin is used effectively in combination chemotherapy to treat Hodgkin's lymphoma, squamous cell carcinoma and testicular cancer [7–9]. Clinically, bleomycin is 90% curative for testicular cancer when used in combination with etoposide and cisplatin [7, 10].

The bleomycin molecule is made up of four domains: a metal-binding domain, a linker region, a carbohydrate moiety and a bithiazole tail [11–14]. The metal-binding domain contains ligands that form complex bonds with the transition metals. The collaborative interaction of the pyrimidine moiety of bleomycin and the bithiazole tail is responsible for DNA binding and sequence specificity [15]. Drug–DNA binding can occur due to intercalation or minor groove interaction [15, 16] although more recent research indicates that intercalative binding is the main form of DNA binding [11, 17]. The linker region improves the DNA cleavage efficiency. The role of the

✉ Vincent Murray
v.murray@unsw.edu.au

¹ School of Biotechnology and Biomolecular Sciences,
University of New South Wales, Sydney, NSW 2052,
Australia

carbohydrate remains unclear [11, 18]; however, there is evidence that it is involved in tumour cell uptake [19, 20].

Studies have suggested that the cytotoxic properties of bleomycin are due to its effect on the physical integrity of the DNA by causing strand scission. The incubation of DNA with bleomycin under appropriate conditions results in both single- and double-strand breaks [2, 5, 15, 21]. In vitro studies show that a single bleomycin molecule can cause both single- and double-strand DNA breaks [11]. The ratio of single-strand breaks to double-strand breaks is approximately 6–10 [22–25].

The anti-tumour properties of bleomycin are the result of the ability of the drug to cause oxidative degradation of DNA [26]. Previous studies have established that DNA strand scission requires oxygen, a reductant and a transition metal ion (with Fe^{2+} as the most extensively studied metal ion) [11, 27–30]. Bleomycin and a ferrous ion form a coordinated complex that reduces molecular oxygen and results in a ferric–iron–bleomycin complex (bleomycin–Fe(III)–OOH) [30–32]. This complex produces a free radical at the C-4' deoxyribose and in the presence of oxygen, strand breaks occur with the production of 3'-phosphoglycolate and 5'-phosphate ends at the DNA cleavage site [11, 33]. In the absence of oxygen, a 4'-oxidised abasic site is formed [34, 35].

Bleomycin preferentially cleaves DNA at pyrimidine nucleotides. The primary targeted sites are 5'-GY (5'-GT and 5'-GC) [2, 15, 36–49]. The cellular bleomycin DNA sequence cleavage specificity also occurs at 5'-GT and 5'-GC dinucleotides [25, 50–56]. Using next-generation Illumina sequencing, we recently determined the genome-wide DNA sequence specificity of bleomycin cleavage in human cells [57–59]. Our analysis of over 200 million bleomycin double-strand break sites revealed a longer consensus sequence than previously determined. Bleomycin preferentially cleaved at the sequence 5'-RTGT*AY (where R is G/A and Y is T/C) in human cells [59]. In parallel, we also investigated the bleomycin genome-wide DNA sequence specificity in purified human DNA and found a slightly different sequence preference of 5'-TGT*AT [59]. In both cellular and purified human DNA, a core trinucleotide 5'-GT*A and tetranucleotide 5'-TGT*A were found.

There were two main aims in this paper. First, we constructed a plasmid clone (Fig. 1) that contained core 5'-GT, 5'-GTA and 5'-TGTA sequences with systematic flanking single nucleotide variations that enabled us to elucidate the crucial sequences involved in the bleomycin sequence cleavage recognition. Second, the plasmid clone contained the DNA sequence, 5'-RTGT*AY with nucleotide variations that enabled our genome-wide bleomycin sequence specificity studies to be compared with the same sequences in purified plasmid DNA.

We utilised 5'- and 3'-end fluorescent labelling and analysis on both strands to precisely determine the bleomycin sequence specificity in the plasmid clone. The use of capillary electrophoresis with laser-induced fluorescence detection (CE-LIF) [60, 61] enabled the exact determination of the sites of bleomycin cleavage and also permitted accurate quantification of the cleavage intensity [62, 63]. If a labelled DNA molecule is cleaved more than once, then only the cleavage site closest to the labelled end is observed [56, 62, 64–67]. The use of both 5'- and 3'-end fluorescent labelling enables the removal of this end-label bias problem [68].

Materials and methods

The bleomycin preparation, Blenoxane[®] (Bristol Laboratories), was used in this study and is a mixture of bleomycin A2 ~ 60%, bleomycin B2 ~ 30% and other minor components. Blenoxane[®] was utilised since this bleomycin preparation is used in the clinic as a cancer chemotherapeutic agent. The RTGTAY plasmid was synthesised by GenScript [35]. The oligonucleotides were obtained from Invitrogen and the deoxynucleotide triphosphates (dNTPs) were from Thermo Fischer Scientific. The Klenow fragment (DNA polymerase I) and exonuclease I were purchased from New England Bio-labs. Shrimp alkaline phosphatase (SAP) was sourced from Promega. Aminoallyl-dUTP-6-FAM (FAM-dUTP) was from Jena Bioscience (Germany). *Taq* DNA polymerase was obtained from Life Technologies.

3'-End-labelling procedure

The 3'-end-labelling procedure [61, 63, 69] produces a PCR product labelled at only one 3'-end. The primers used for the top strand were SEQ3 (5'-TGTGCTGCAA GGCGA-3') and REV2 (5'-ATTGTGAGCGGATAAC-3'); and for the bottom strand REV3 (5'-TTGTGAGCGCGGA TAAC-3') and SEQ2 (5'-ATGTGCTGCAAGGCGA-3'). A 20- μl reaction volume was prepared which consisted of approximately 3 ng plasmid DNA, 10 pmol of each primer, 0.3 mM each dNTP, 16.6 mM $(\text{NH}_4)_2\text{SO}_4$, 67 mM Tris–HCl, pH 8.8, 6.7 mM MgCl_2 and 0.1 U *Taq* DNA polymerase. The Bio-Rad DNA Engine Dyad Peltier thermal cycler was utilised and the thermal cycling conditions were 95 °C for 5 min; 25 cycles of 95 °C for 45 s, 55 °C for 1 min, 72 °C for 2 min; and a single cycle of 72 °C for 10 min.

The PCR products were treated with Exo-SAP (exonuclease I and shrimp alkaline phosphatase) which removes excess primers and dNTPs [61]. The reaction mixture (12 μl) which consisted of 10 μl PCR product, 0.1 μl of 20 U/ μl exonuclease I, 0.1 μl of 1 U/ μl of shrimp alkaline phosphatase, 0.2 μl of 10 \times exonuclease I buffer and 1.6 μl Milli-Q H_2O was then incubated at 37 °C for

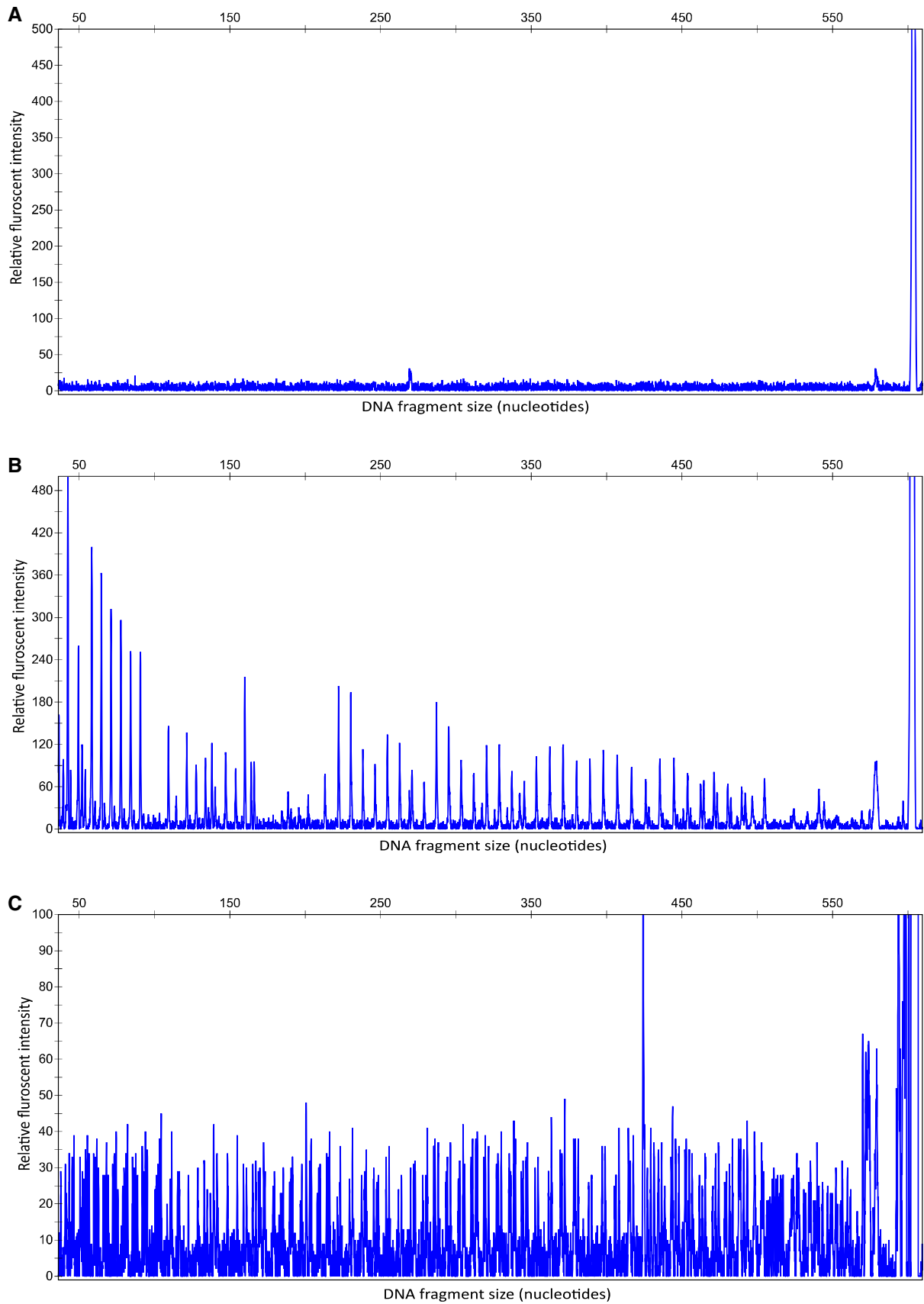


Fig. 2 Electropherogram showing the bleomycin damage on the 3'-end labelled PCR product (*top strand*) derived from the RTGTAY plasmid. Relative fragment sizes (nucleotides) are shown on the *x-axis* whereas, the relative fluorescent intensity is along the *y-axis*. **a** A no drug control. **b** The 3'-end labelled PCR product treated with 0.05 mM bleomycin. **c** The Maxam–Gilbert G+A size marker for the 3'-end labelled PCR product. The peak at the far-right end is the full-length product at 590 bp

mixture consisted of approximately 12 ng of the end-labelled DNA, 720 ng of purified chicken DNA as carrier DNA [61–63, 70] and an equal concentration (0.05 mM) of FeSO₄ and bleomycin. Carrier DNA was included to enable experimental consistency, but required a higher bleomycin concentration for effective cleavage. The reaction mixture was incubated at 37 °C for 30 min followed by ethanol precipitation and dissolved in 10–15 μl 10 mM Tris–HCl, pH 8; 0.1 mM EDTA.

DNA sequencing reactions as molecular weight size markers

Maxam–Gilbert G+A cleavage and dideoxy sequencing were used as size markers for 3'- and 5'-end labelling, respectively, and were produced as previously described [61, 63, 71, 72]. The bleomycin cleavage produces 3'-phosphoglycolate ends that show unusual mobility in capillary electrophoresis [63, 71]. The 5'-end labelled samples were treated with endonuclease IV as previously described [63] before comparison with the dideoxy sequencing reactions. Endonuclease IV treatment converts the 3'-phosphoglycolate to 3'-hydroxyl termini.

DNA cleavage analysis

The bleomycin-treated DNA samples (2 μl) were sent for CE-LIF fragment analysis at the Ramaciotti Centre for Genomics, University of NSW. The samples were processed on an ABI3730 Sequencer and the data were analysed with GeneMapper software version 3.7 (Applied Biosystems) [60–62]. The relative fluorescence intensity for each damaged peak was quantified using the area under the peak using GeneMapper. The damage intensity values were calculated after subtracting the values of background peak area from the equivalent damaged peak area. The percentage damage was calculated for each damaged site and the end-label bias was corrected using an algorithm [60–63]. The damage percentage was then calculated for each peak by dividing it by the sum of all peaks including the full-length peak [61]. Three independent bleomycin experiments were performed for each strand and at both the 5'- and 3'-ends.

Results

Bleomycin cleavage with end-labelled PCR products

Bleomycin cleavage was examined on both strands of the RTGTAY plasmid insert DNA sequence (Fig. 1). Using the RTGTAY plasmid as template, PCR products were generated that were fluorescently labelled at either the 3'- or 5'-ends. These PCR products were gel purified and subjected to bleomycin damage. These bleomycin-treated samples were then analysed by capillary electrophoresis (ABI 3730) and the cleavage sites were quantified using Gene Mapper software. Maxam–Gilbert G+A chemical sequencing (for 3'-end labelled fragments) and dideoxy sequencing products (for 5'-end labelled fragments) were used as size markers to determine the precise sites of bleomycin cleavage in the analysed sequence.

The experiments were performed in triplicate. Both strands of the RTGTAY plasmid were analysed and each strand was labelled at either the 5'- or 3'-ends. Thus, four sets of triplicate experiments were performed. After quantification and use of an algorithm to remove end-label bias [56, 62], the 5'- and 3'-labelled experiments for the top strand were combined to give an accurate quantified profile of the bleomycin cleavage sites on the top strand. Similarly, the bottom strand 5'- and 3'-labelled experiments were combined to provide the intensity of the bleomycin cleavage sites on the bottom strand.

Quantitative analysis

The fluorescently labelled PCR products were treated with 0.05 mM bleomycin and electrophoresed on an ABI 3730 capillary sequencer. The electropherogram in Fig. 2 shows the bleomycin cleavage profile for the top strand of the RTGTAY plasmid that was fluorescently labelled at the 3'-end. The electropherogram shows a no drug control, a 0.05 mM bleomycin-treated sample along with a G+A size marker. The cleavage pattern in the damaged sample was well defined since the no drug control had a very low background. The far-right peak at the end of no drug control and damaged sample is the full-length PCR product (590 bp). Higher levels of cleavage were seen with increasing concentration of bleomycin; however, the cleavage sites were highly biased towards the labelled end and presented as shorter fragments.

The bleomycin cleavage sites were quantified and sites having a damage percentage above 0.02% were used for further analysis. The quantified damage sites are depicted as bar charts and Fig. 3 represents the percentage cleavage of bleomycin for the insert sequence in the top strand

of the RTGTAY plasmid. The quantified damage percentage for each cleavage site with 5'-end labelling is shown above x-axis and the data for 3'-end labelling are shown below x-axis. The standard error of mean is depicted as error bars.

The bleomycin DNA damage intensities from the 5'- and 3'-end labelling experiments were combined to give an average intensity at each site. These combined intensities can be seen in Fig. 4 for the top strand of the RTGTAY plasmid. Figure 4 shows the analysed sequence in the insert region (97–467 bp) of the RTGTAY plasmid excluding the telomeric and polyguanine repeats. The DNA damage percentage for the top strand varied from 0.03 to 2.70%.

The strongest cleavage sites for bleomycin were found to be at 5'-GT and 5'-GC dinucleotides. The top strand of RTGTAY plasmid shows cleavage at all 5'-GT and all 5'-GC dinucleotides sites, but none of the 5'-GA sites were cleaved by bleomycin. For the bottom strand, the most intense cleavage sites were at 5'-GT and 5'-GC dinucleotides and 5'-GTA trinucleotides.

The 30 most intense bleomycin cleavage sites from the insert region of the top strand of the RTGTAY plasmid are summarised in Table 1. The cleavage site (read 5'–3') is depicted as six nucleotides with the dinucleotide cleavage site represented in capital letters and the * indicating the cleavage site. Of the 30 most intense bleomycin sites in Table 1, 29 contain the dinucleotide 5'-GT* with one 5'-GC*

as the 27th most intense site. The top 26 most intense sites all have 5'-GT*A at the bleomycin cleavage site.

Systematic variation of neighbouring nucleotides around the bleomycin cleavage site

The sequence on the top strand of the RTGTAY plasmid was specifically designed to have systematic variation of neighbouring nucleotides around 5'-GT, 5'-GTA and 5'-TGTA sequences. This detail from the top strand was used to determine the effect of neighbouring nucleotides on the damage intensity. This region is shown in Fig. 4 and is the insert region from 97 to 467 bp of the top strand of RTGTAY plasmid. Figure 4 also shows the bleomycin cleavage intensities at the various sites.

Table 2 reveals the influence of single nucleotide variations on bleomycin cleavage intensity for bleomycin cleavage sites on the top strand of the RTGTAY plasmid. The data are presented as nucleotides positioned at –3, –2, –1, 0, 1 and 2 with respect to the bleomycin cleaved nucleotide at position 0. X is the variant nucleotide and the values for G, A, T and C are indicated.

In Table 2a, the effect of nucleotide variation at position –3 is shown for the sequence 5'-XTGT*A that is located in the region 100–240 bp. The C nucleotide has the highest percentage damage, followed by T, A and G. Hence the preferred nucleotide at position –3 is a pyrimidine, C or T.

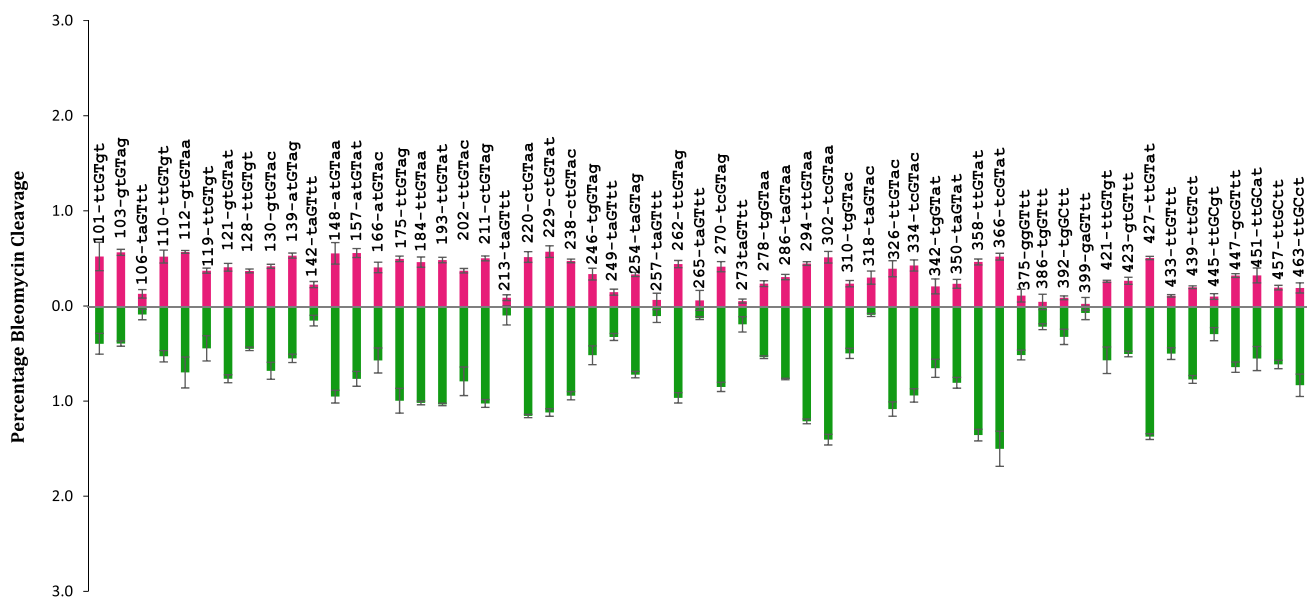


Fig. 3 The percentage bleomycin cleavage for the *top* strand of the RTGTAY plasmid. The sites are ordered in 5'–3' direction (*left to right*). The percentage bleomycin cleavage is shown on the *y*-axis. Each bar represents a bleomycin cleavage site. The *bars* (in *pink*) above the *x*-axis represent the cleavage sites from 5'-end labelling and the *bars* (in *green*) below the *x*-axis from 3'-end labelling. Each

bar is labelled with the base number and the hexanucleotide sequence for the cleaved site. For example, the bar labelled 58-ttGTAA indicates that the nucleotide 'T', at base pair 58, is cleaved. These values are derived from three independent experiments and the *error bars* represent the standard error of the mean

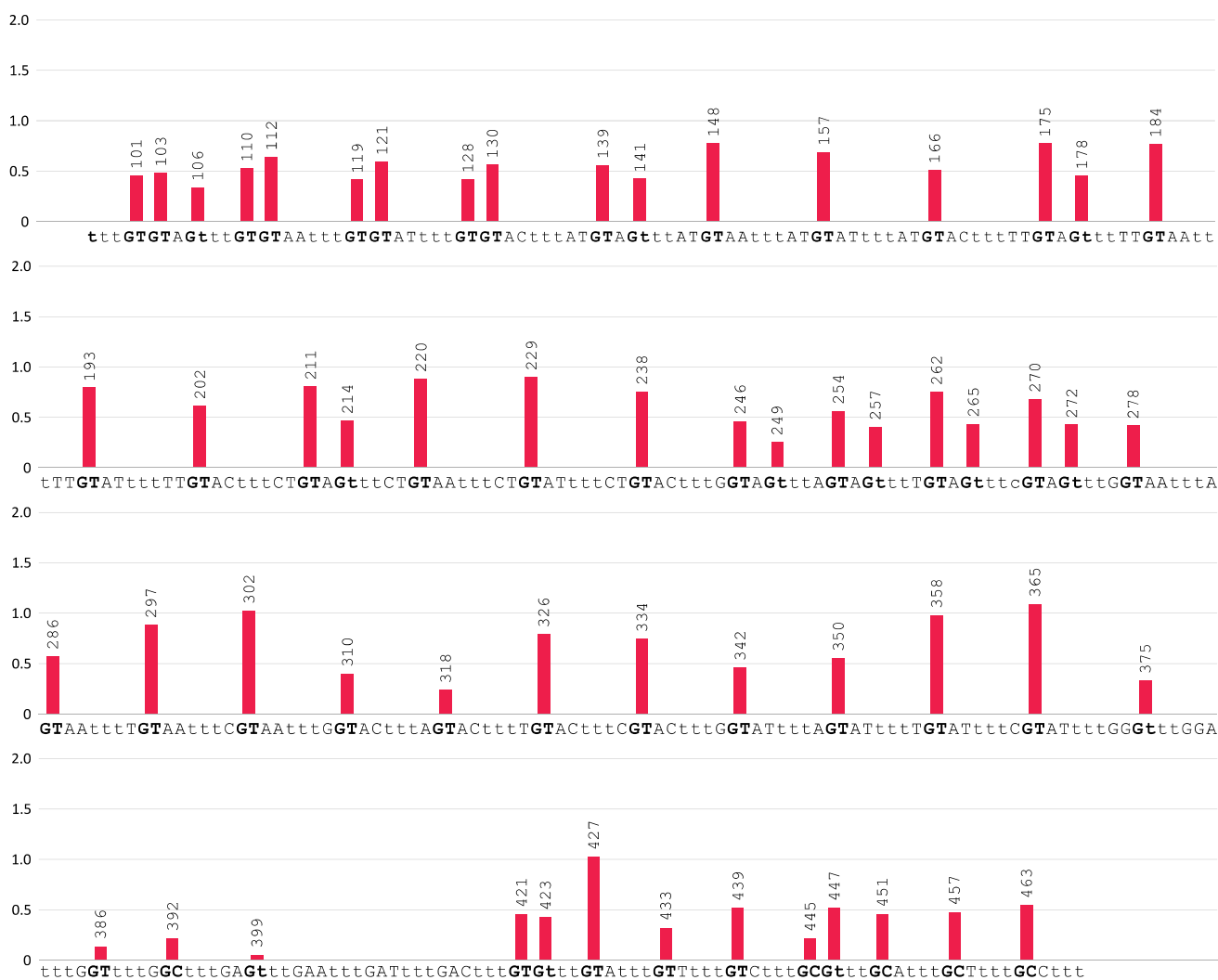


Fig. 4 Percentage bleomycin damage in the insert region of the top strand of the RTGTAY plasmid. The insert sequence is 371 base pairs long and corresponds to bp 97–467. The sites are ordered in 5′–3′ direction (left to right). The region has core sequences of 5′-GT, 5′-GTA** and 5′-TGTA with comprehensive variations in the neighbouring nucleotide sequences. The percentage damage is shown on

the y-axis and is the average of data from 5′- and 3′-end labelling. Each bar (in pink) above the horizontal axis represents a bleomycin cleavage site. The cleavage sites are marked in bold and is labelled with the corresponding base pair. The figure shows the effect of the neighbouring nucleotide on the bleomycin cleavage intensity

For Table 2b, nucleotide variation at position –2 in the sequence 5′-XGT*A is depicted. The order of nucleotides with the highest percentage damage is C = T > A > G. Thus the preferred nucleotide at position –2 is C or T.

At the –1 position, G is always present. In Table 2c, the variation at position 0 is shown for the sequence 5′-GX*. In Table 2c (i), the cleavage sites in the region 372–464 are presented. The T nucleotide has the highest percentage damage at position 0, followed by C. In Table 2c (ii), a larger number of sites are shown covering the whole insert region and T again had the highest percentage damage at position 0, followed by C. For G and A, none of the sites

had appreciable cleavage. Thus T is the preferred nucleotide at position 0.

For Table 2d, nucleotide variation at position +1 is depicted. In Table 2d (i), the cleavage sites in the region 372–464 are shown for the sequence 5′-GN*X. The 5′-GT*A sequence has the highest percentage damage with a value of 1.03% that is larger than the other values by an appreciable margin. For 5′-GT*X, the intensity order was A ≫ C > G > T, while for 5′-GC*X it was C > T > A > G. Hence for the +1 position, the A nucleotide was only preferred when present as 5′-GT*A and not as 5′-GC*A. The 5′-GG*X and 5′-GA*X sites had no appreciable cleavage.

Table 1 The 30 most intense bleomycin cleavage sites in the insert sequence of the top strand of RTGTAY plasmid

Base pair number	Damage %	Sequence
366	1.09	tcGT*at
427	1.03	ttGT*at
302	1.02	tcGT*aa
358	0.98	ttGT*at
229	0.90	ctGT*at
294	0.89	ttGT*aa
220	0.88	ctGT*aa
211	0.80	ctGT*ag
193	0.79	ttGT*at
326	0.79	ttGT*ac
148	0.78	atGT*aa
175	0.78	ttGT*ag
184	0.77	ttGT*aa
238	0.76	ctGT*ac
262	0.75	ttGT*ag
334	0.74	tcGT*ac
157	0.69	atGT*at
270	0.68	tcGT*ag
112	0.64	gtGT*aa
202	0.61	ttGT*ac
121	0.60	gtGT*at
286	0.57	taGT*aa
130	0.56	gtGT*ac
254	0.56	taGT*ag
139	0.56	atGT*ag
350	0.55	taGT*at
463	0.54	ttGC*ct
110	0.53	ttGT*gt
447	0.52	gcGT*tt
439	0.52	ttGT*ct

The 30 most intense bleomycin cleavage sites from the insert region of the top strand of the RTGTAY plasmid from 97 to 467 bp are shown. The position of the bleomycin cleavage is shown in bp. The sites are sorted by damage percentage in decreasing order. The cleavage site (read 5′–3′) is depicted as six nucleotides with the dinucleotide cleavage site represented in capital letters and the * indicating the cleavage site. For example, for the sequence 5′-tcGT*at, the T* nucleotide is destroyed during the bleomycin cleavage reaction

A larger number of sites was examined in Table 2d (ii) for the 5′-GT*X sequence at position +1. The percentage damage order was A > C > G > T. Hence at the +1 position, the preferred nucleotide was A.

Table 2e reflects the effect of nucleotide variation at position +2. In Table 2e (i), the region 100–240 bp was examined for the sequence 5′-TGT*AX. The nucleotide preference was A > T > G > C. In Table 2e (ii), the region 244–368 bp was investigated for the sequence 5′-GT*AX. The nucleotide preference was T > A > G > C. In Table 2e (iii), the values

in parts (i) and (ii) are averaged for the eight sites and the nucleotide preference was T = A > G > C. Hence at the +2 position, the preferred nucleotide was T or A.

In summary, from the above results, we can infer the bleomycin cleavage preferences for positions from –3 to +2. For position –3, it was C > T > A > G; for position –2, C = T > A > G; for position –1, G; for position 0, T > C; for position +1, A > C > G > T; for position +2, T = A > G > C. A briefer nucleotide preference that only shows the most highly preferred nucleotides is shown in Table 3. From the preferences in Table 3, a consensus sequence for bleomycin cleavage can be derived as 5′-YYGT*AW (where Y = C/T and W = A/T). Also shown in Table 3 is human genome-wide data from our previous paper [59] that reveals a consensus sequence of 5′-WTGT*AW.

From Table 1, the most highly cleaved sequence was 5′-TCGT*AT and the seven most highly cleaved sequences conformed to the consensus sequence 5′-YYGT*AW. Eight out of the ten most cleaved sequences also conformed to this consensus sequence.

Discussion

In this study, the DNA sequence specificity of the cancer chemotherapeutic agent, bleomycin, was investigated. In this paper, we wished to comprehensively define the DNA sequence specificity of bleomycin cleavage. This paper represents the first occasion that a systematically altered DNA sequence has been employed to examine the DNA sequence specificity of bleomycin. In previous experiments, random sequences were used that did not include every possible DNA sequence combination. When random DNA sequences are employed, there is the possibility that a particular sequence is missing from the analysed sequence. In this paper, every possible combination of DNA sequences around the bleomycin cleavage site was examined.

Bleomycin utilises the oxidation of ferrous ions to ferric ions as the driving force for the DNA damage reaction and results in the cleavage of specific DNA sequences. The bleomycin cleavage data from the insert region of the RTGTAY plasmid was used to study the sequence specificity of bleomycin. This insert region contained systematic nucleotide variations in the flanking DNA sequences around the bleomycin cleavage site and allowed a precise evaluation of the crucial sequences involved in the bleomycin DNA cleavage reaction. A consensus sequence for the most cleaved bleomycin sites was found to be 5′-YYGT*AW (Table 3). This plasmid DNA sequence specificity data also permitted a comparison with recent

Table 2 The influence of single nucleotide alterations on bleomycin cleavage intensity

Position (bp)	Sequence	X = G	X = A	X = T	X = C
(a) Effect of nucleotide variation at position -3					
100–132	XTGT*A	0.48	0.56	0.78	0.80
136–168	XTGT*A	0.64	0.78	0.77	0.88
172–204	XTGT*A	0.60	0.69	0.79	0.90
208–240	XTGT*A	0.56	0.51	0.61	0.75
	Average	0.57 (4)	0.63 (4)	0.74 (4)	0.83 (4)
(b) Effect of nucleotide variation at position -2					
244–272	XGT*A	0.46	0.56	0.75	0.68
276–304	XGT*A	0.41	0.57	0.88	1.02
308–336	XGT*A	0.40	0.24	0.79	0.74
340–368	XGT*A	0.46	0.55	0.98	1.09
	Average	0.43 (4)	0.48 (4)	0.85 (4)	0.88 (4)
(c) (i) Effect of nucleotide variation at position 0					
372–392	GX*	0	0	0.46	0.22
396–416	GX*	0	0	1.02	0.46
420–440	GX*	0	0	0.32	0.48
444–464	GX*	0	0	0.52	0.55
	Average	0 (4)	0 (4)	0.58 (4)	0.43 (4)
(c) (ii) Effect of nucleotide variation at position 0					
–	GX*	0 (8)	0 (4)	0.58 (53)	0.43 (5)
(d) (i) Effect of nucleotide variation at position +1					
372–392	GG*X	0 (1)	0 (1)	0 (1)	0 (1)
396–416	GA*X	0 (1)	0 (1)	0 (1)	0 (1)
420–440	GT*X	0.46 (1)	1.03 (1)	0.32 (1)	0.52 (1)
444–464	GC*X	0.22 (1)	0.46 (1)	0.48 (1)	0.55 (1)
(d) (ii) Effect of nucleotide variation at position +1					
–	GT*X	0.46 (5)	0.69 (33)	0.22 (14)	0.52 (1)
(e) (i) Effect of nucleotide variation at position +2					
100–132	TGT*AX	0.48	0.64	0.60	0.56
136–168	TGT*AX	0.56	0.78	0.69	0.51
172–204	TGT*AX	0.78	0.77	0.79	0.61
208–240	TGT*AX	0.80	0.88	0.90	0.75
		0.66 (4)	0.77 (4)	0.75 (4)	0.61 (4)
(e) (ii) Effect of nucleotide variation at position +2					
244–272	GT*AX	0.46	0.41	0.46	0.40
276–304	GT*AX	0.56	0.57	0.55	0.24
308–336	GT*AX	0.75	0.88	0.98	0.79
340–368	GT*AX	0.68	1.02	1.09	0.74
		0.61 (4)	0.72 (4)	0.77 (4)	0.54 (4)
(e) (iii) Effect of nucleotide variation at position +2					
100–368	Average	0.63 (8)	0.75 (8)	0.76 (8)	0.58 (8)

The influence of single nucleotide variation on the bleomycin cleavage intensity is shown for cleavage sites on the top strand of the RTGTAY plasmid from 97 to 467 bp. The data are presented as nucleotides positioned at -3, -2, -1, 0, 1 and 2, respectively. The position 0 (with an *) defines the bleomycin-cleaved nucleotide and X is the variant nucleotide. The values for the variant X nucleotides, G, A, T** and C, are the average percentage bleomycin damage from three independent experiments. The combined average is written at the bottom of each part and the number of sites is written in brackets. Section a shows variants at position -3; section b at position -2; section c at position 0; section d at position +1; section e at position +2

genome-wide data in purified human DNA and in human cells [59].

The effect of neighbouring nucleotides on the intensity of bleomycin cleavage

The RTGTAY plasmid contained 5'-GT, 5'-GT*A and 5'-TGT*A as core sequences with comprehensive variations of the flanking DNA sequences. The study involved bleomycin cleavage of both 5'- and 3'-end labelled DNA that enabled the elimination of end-label bias [68]. The use of fluorescently labelled DNA and detection by CE-LIF permitted a precise and accurate quantitative determination of the bleomycin DNA sequence specificity.

The bleomycin sequence specificity was investigated by quantifying the intensity of the cleavage sites in the insert sequence of the top strand of RTGTAY plasmid (Tables 1, 2, 3). The bleomycin cleavage intensity varied with alteration of the surrounding nucleotides. The top cleaved trinucleotide for bleomycin was found to be 5'-GT*A that had been previously demonstrated [61, 68, 73]. For tetranucleotides, it was 5'-YGT*A. The most intense bleomycin DNA cleavage sites conformed to the sequence 5'-YYGT*AW (Table 3). Eight of the ten most cleaved sequences conformed to this consensus sequence (Table 1).

There are previous studies that sought to determine a larger bleomycin DNA cleavage consensus sequence than a simple dinucleotide. Murray and Martin [39] found that T at the -2 position produced more intense bleomycin cleavage sites. Chen et al. [73] observed a consensus sequence of 5'-TTGT*AW that was consistent with the consensus sequence 5'-YYGT*AW found in this paper.

Earlier studies have reported that alternating purine-pyrimidine nucleotide sequences are strong damage sites for bleomycin [25, 39, 40]. This was proposed to be due to a specific conformation of DNA at this alternating purine-pyrimidine sequence [74]. The central part of the 5'-YYGT*AW consensus sequence, 5'-YGT*A, corresponds to an alternating purine-pyrimidine nucleotide sequence and could be extended to the +2 position when T is present at the W position. However, Y at the -3 position is not consistent with an alternating purine-pyrimidine sequence.

One drawback with the current study was the non-random nature of the RTGTAY plasmid. In order to have the presence of all possible DNA sequence combinations, the analysed sequence was necessarily non-random. Chung and Murray [68] utilised two random sequences for the determination of the DNA sequence specificity of bleomycin and found a consensus sequence of 5'-TGT*A. This was very similar to the consensus sequence found in this paper but had a minor difference at the -2 position where there was a T instead of Y (T/C).

Table 3 The preferred neighbouring nucleotides at the bleomycin cleavage site

Position	RTGTAY plasmid insert preferred nucleotide	Genome-wide preferred nucleotide (50 k purified DNA)
−3	C > T	A > T
−2	C = T	T
−1	G	G
0	T	T
+1	A	A
+2	T = A	T > A

The preferred nucleotides at the bleomycin cleavage site are shown from the −3 to the +2 positions. The data for the RTGTAY plasmid insert are from Table 2. Only the most highly preferred nucleotides are shown. The preferred nucleotides from genome-wide data from purified human DNA are depicted for the top 50,000 cleavage sites (data from [59])

Comparison of bleomycin sequence specificity in plasmid DNA with the entire human genome

The second aim of this paper was to compare the bleomycin genome-wide DNA sequence specificity with the same sequences in purified plasmid DNA. A comparison of the bleomycin DNA sequence specificity in plasmid DNA with the human genome-wide data is shown in Table 3. For the most highly cleaved sequences, it can be observed that a consensus sequence of 5'-YYGT*AW can be derived from the plasmid and 5'-WTGT*AW for the human genome-wide data [59]. The human genome-wide data in Table 3 was obtained from the 50,000 most highly cleaved sites with purified human DNA (not cellular DNA). A comparison with purified human DNA, and not human cellular DNA, was considered to be more appropriate since cellular DNA is complexed with chromosomal proteins that could affect the bleomycin cleavage reaction (see below).

The core 5'-GT*A sequence was the same for both environments. The difference at position −3 was C > T for the plasmid and A > T for the genome-wide environment. For the genome-wide data, the −3 position was not found to be significantly different to the surrounding sequence as assessed by the Mahalanobis distance [59]. Hence, a comparison at this position is not likely to be valid.

For the −2 position, C = T was found for the plasmid sequence and T for the genome-wide data; however, with genome-wide a C was the lowest ranked nucleotide at the −2 position. At the +2 position, T = A is observed for the plasmid and T > A for the genome-wide data.

Hence, there are minor differences at the −2, +2 and possibly the −3 positions between the two environments. There are several possible reasons for these differences [68]. First, the plasmid data detect single-strand breaks

whereas the genome-wide data detect double-strand breaks. As mentioned previously, there are thought to be 6–10 single-strand breaks for every double-strand break [22–25]. A double-strand break is thought to be derived from a similar process to a single-strand break but is a more extreme event [11, 75].

Second, during the Illumina genome-wide procedure oligonucleotide linkers are ligated to the double-strand breaks before being placed on the flowcell. These processes may not be DNA sequence independent and may introduce differences compared with the simpler CE-LIF process for the plasmid data.

Third, even though the plasmid has been designed to contain all of the possible DNA sequence combinations, there were fewer DNA sequences analysed with the plasmid data. In the genome-wide data, the most frequent 50,000 sites were examined; hence a greater number of sites were investigated that leads to a higher level of confidence in the genome-wide results.

Fourth, for the plasmid sequences the neighbouring sequences to the hexanucleotide cleavage site were three or four consecutive T nucleotides because these oligo T sequences are never cleaved by bleomycin. In contrast for the genome-wide data, random sequences were present next to the bleomycin cleavage sites.

With reference to the cellular genome-wide data, the consensus DNA sequence was 5'-RTGT*AY [59]. Again the 5'-GT*A core was the same in the two environments, and the flanking sequences at positions −2 and +2 were slightly different. The main difference was at the −3 position and that position was also different for the purified genome-wide DNA. The cellular environment is very different to the purified DNAs because of the presence of chromosomal proteins, different chemical constituents, and the dynamic nature of the DNA inside the cells with ongoing replication and transcription processes [59]. Bleomycin preferentially cleaves at the linker regions of nucleosomes in the cellular DNA. The DNA encased around the nucleosome core is supercoiled and distorted which alters the bleomycin cleavage intensity [50, 51, 57, 76, 77]. The different cellular environment, especially proteins bound to DNA, will affect the bleomycin cleavage reaction and could be responsible for these differences.

In conclusion, this paper has investigated a longer consensus DNA sequence for the bleomycin cleavage site. By use of systematic alteration of flanking sequences, the effect of neighbouring nucleotides on the intensity of bleomycin was evaluated. We attempted in this paper to fully document the DNA sequences that were preferentially cleaved by bleomycin. Knowledge of the sequence specificity will help us identify the crucial gene sequences or gene features that are cleaved by bleomycin and are important for the anti-tumour activity of bleomycin. This

deeper understanding of the bleomycin DNA cleavage process may enable the development of more effective cancer chemotherapeutic agents based on bleomycin.

Acknowledgements Support of this work by the University of New South Wales, Science Faculty Research Grant Scheme is gratefully acknowledged. SDG was funded by an Australian Postgraduate Award.

References

1. Umezawa H, Maeda K, Takeuchi T, Okami Y (1966) *J Antibiot* 19:200–209
2. Kross J, Henner WD, Hecht SM, Haseltine WA (1982) *Biochemistry* 21:4310–4318
3. Suzuki H, Nagai K, Akutsu E, Yamaki H, Tanaka N, Umezawa H (1970) *J Antibiot* 23:473–480
4. Burger RM (1998) *Chem Rev* 98:1153–1170
5. Horwitz SB, Sausville EA, Peisach J (1979) In: Hecht SM (ed) *Bleomycin: chemical, biochemical, and biological aspects*. In: *Proceedings of US–Japan Symposium, East-West Center, Honolulu, July 18–22, 1978*. Springer New York, New York, NY, pp 170–183
6. Claussen CA, Long EC (1999) *Chem Rev* 99:2797–2816
7. Einhorn LH (2002) *Proc Natl Acad Sci USA* 99:4592–4595
8. Stoter G, Kaye SB, de Mulder PH, Levi J, Raghavan D (1994) *J Clin Oncol* 12:644–645
9. Neese F, Zaleski JM, Loeb Zaleski K, Solomon EI (2000) *J Am Chem Soc* 122:11703–11724
10. Burger RM, Projan SJ, Horwitz SB, Peisach J (1986) *J Biol Chem* 261:5955–5959
11. Chen J, Stubbe J (2005) *Nat Rev Cancer* 5:102–112
12. Stubbe J, Kozarich JW (1987) *Chem Rev* 87:1107–1136
13. Stubbe J, Kozarich JW, Wu W, Vanderwall DE (1996) *Acc Chem Res* 29:322–330
14. Hecht SM (2000) *J Nat Product* 63:158–168
15. D'Andrea AD, Haseltine WA (1978) *Proc Natl Acad Sci USA* 75:3608–3612
16. Povirk LF, Hogan M, Dattagupta N (1979) *Biochemistry* 18:96–101
17. Goodwin KD, Lewis MA, Long EC, Georgiadis MM (2008) *Proc Natl Acad Sci USA* 105:5052–5056
18. Wu W, Vanderwall DE, Turner CJ, Kozarich JW, Stubbe J (1996) *J Am Chem Soc* 118:1281–1294
19. Yu Z, Schmaltz RM, Bozeman TC, Paul R, Rishel MJ, Tsosie KS, Hecht SM (2013) *J Am Chem Soc* 135:2883–2886
20. Schroeder BR, Ghare MI, Bhattacharya C, Paul R, Yu Z, Zaleski PA, Bozeman TC, Rishel MJ, Hecht SM (2014) *J Am Chem Soc* 136:13641–13656
21. Steighner RJ, Povirk LF (1990) *Proc Natl Acad Sci USA* 87:8350–8354
22. Povirk LF, Wübker W, Köhnlein W, Hutchinson F (1977) *Nucleic Acids Res* 4:3573–3580
23. Cullinan EB, Gawron LS, Rustum YM, Beerman TA (1991) *Biochemistry* 30:3055–3061
24. Tounekti O, Kenani A, Foray N, Orłowski S, Mir LM (2001) *Br J Cancer* 84:1272–1279
25. Grimwade JE, Cason EB, Beerman TA (1987) *Nucleic Acids Res* 15:6315–6329
26. Hashimoto S, Wang B, Hecht SM (2001) *J Am Chem Soc* 123:7437–7438
27. Absalon MJ, Kozarich JW, Stubbe J (1995) *Biochemistry* 34:2065–2075
28. Absalon MJ, Wu W, Kozarich JW, Stubbe J (1995) *Biochemistry* 34:2076–2086
29. Sam JW, Takahashi S, Lippai I, Peisach J, Rousseau DL (1998) *J Biol Chem* 273:16090–16097
30. Sausville EA, Stein RW, Peisach J, Horwitz SB (1978) *Biochemistry* 17:2746–2754
31. Oberley LW, Buettner GR (1979) *FEBS Lett* 97:47–49
32. Caspary WJ, Lanzo DA, Niziak C (1982) *Biochemistry* 21:334–338
33. Dorr RT (1992) *Semin Oncol* 19:3–8
34. Rabow LE, Stubbe J, Kozarich JW (1990) *J Am Chem Soc* 112:3196–3203
35. Chen JK, Murray V (2016) *J Biol Inorg Chem* 21:395–406
36. Takeshita M, Grollman AP, Ohtsubo E, Ohtsubo H (1978) *Proc Natl Acad Sci USA* 75:5983–5987
37. Takeshita M, Kappen LS, Grollman AP, Eisenberg M, Goldberg IH (1981) *Biochemistry* 20:7599–7606
38. Mirabelli CK, Ting A, Huang CH, Mong S, Crooke ST (1982) *Cancer Res* 42:2779–2785
39. Murray V, Martin RF (1985) *Nucleic Acids Res* 13:1467–1481
40. Murray V, Tan L, Matthews J, Martin RF (1988) *J Biol Chem* 263:12854–12859
41. Nightingale KP, Fox KR (1993) *Nucleic Acids Res* 21:2549–2555
42. Murray V (2000) *Prog Nucl Acid Res Mol Biol* 63:367–415
43. Lewis MA, Long EC (2006) *Bioorg Med Chem* 14:3481–3490
44. Ma Q, Akiyama Y, Xu Z, Konishi K, Hecht SM (2009) *J Am Chem Soc* 131:2013–2022
45. Bozeman TC, Nanjunda R, Tang C, Liu Y, Segerman ZJ, Zaleski PA, Wilson WD, Hecht SM (2012) *J Am Chem Soc* 134:17842–17845
46. Segerman ZJ, Roy B, Hecht SM (2013) *Biochemistry* 52:5315–5327
47. Roy B, Hecht SM (2014) *J Am Chem Soc* 136:4382–4393
48. Roy B, Tang C, Alam MP, Hecht SM (2014) *Biochemistry* 53:6103–6112
49. Tang C, Paul A, Alam MP, Roy B, Wilson WD, Hecht SM (2014) *J Am Chem Soc* 136:13715–13726
50. Murray V, Martin RF (1985) *J Biol Chem* 260:10389–10391
51. Cairns MJ, Murray V (1996) *Biochemistry* 35:8753–8760
52. Kim A, Murray V (2000) *Int J Biochem Cell Biol* 32:695–702
53. Kim A, Murray V (2001) *Int J Biochem Cell Biol* 33:1183–1192
54. Temple MD, Freebody J, Murray V (2004) *Biochim Biophys Acta* 1678:126–134
55. Temple MD, Murray V (2005) *Int J Biochem Cell Biol* 37:665–678
56. Nguyen HTQ, Murray V (2012) *J Biol Inorganic Chem* 17:1209–1215
57. Murray V, Chen JK, Galea AM (2014) *Cell Mol Life Sci* 71:1505–1512
58. Murray V, Chen JK, Galea AM (2014) *Mutat Res Fund Mol Mech Mutagen* 769:93–99
59. Murray V, Chen JK, Tanaka MM (2016) *Mol Biol Report* 43:639–651
60. Paul M, Murray V (2012) *Biomed Chromatogr* 26:350–354
61. Nguyen TV, Murray V (2012) *J Biol Inorg Chem* 17:1–9
62. Murray V, Nguyen TV, Chen JK (2012) *Chem Biol Drug Des* 80:1–8
63. Nguyen TV, Chen JK, Murray V (2013) *J Chromatogr B* 913:113–122
64. White RJ, Phillips DR (1989) *Biochemistry* 28:4277–4283
65. Gordon LK, Haseltine WA (1980) *J Biol Chem* 255:12047–12050
66. Lobachevsky PN, Martin RF (2000) *Radiat Res* 153:271–278
67. Lobachevsky PN, Martin RF (1996) *Acta Oncol* 35:809–815
68. Chung LH, Murray V (2016) *J Chromatogr B* 1008:87–97

69. Inazuka M, Tahira T, Hayashi K (1996) *Genome Res* 6:551–557
70. Murray V, Martin RF (1994) *Nucleic Acids Res* 22:506–513
71. Nguyen TV, Murray V (2013) *Biomed Chromatogr* 27:390–395
72. Belikov S, Wieslander L (1995) *Nucleic Acids Res* 23:310
73. Chen JK, Yang D, Shen B, Neilan BA, Murray V (2016) *Bioorg Med Chem* 24:6094–6101
74. Calladine CR (1982) *J Mol Biol* 161:343–352
75. Povirk LF, Han YH, Steighner RJ (1989) *Biochemistry* 28:5808–5814
76. Kuo MT, Hsu TC (1978) *Nature* 271:83–84
77. Galea AM, Murray V (2010) *Chem Biol Drug Des* 75:578–589