The Genome Factor in Region-Specific DNA Damage: The DNA-Reactive Drug U-78779 Prefers Mixed A/T-G/C Sequences at the Nucleotide Level but Is Region-Specific for Long Pure AT Islands at the Genomic Level†

Maryanne C. S. Herzig,‡ Karl A. Rodriguez,‡ Alex V. Trevino,‡ Jaroslaw Dziegielewski,‡,§ Brenda Arnett,‡ Laurence Hurley,|| and Jan M. Woynarowski*,‡

Cancer Therapy and Research Center, 14960 Omicron Drive, San Antonio, Texas 78245, and University of Arizona Cancer Center, 1515 North Campbell Avenue, P.O. Box 245024, Tucson, Arizona 85724-5024

Received October 9, 2001; Revised Manuscript Received November 21, 2001

ABSTRACT: Bizelesin is the first anticancer drug capable of damaging specific regions of the genome with clusters of its binding sites T(A/T)$_4$A. This study characterized the sequence- and region-specificity of a bizelesin analogue, U-78779, designed to interact with mixed A/T-G/C motifs. At the nucleotide level, U-78779 was found to prefer runs of A/Ts interspersed with 1 or 2 G/C pairs, although 25% of the identified sites corresponded to pure AT motifs similar to bizelesin sites. The *in silico* computational analysis showed that the preferred mixed A/T-G/C motifs distribute uniformly at the genomic level. In contrast, the secondary, pure AT motifs (A/T)$_6$A were found densely clustered in the same long islands of AT-rich DNA that bizelesin targets. Mapping the sites and quantitating the frequencies of U-78779 adducts in model AT island and non-AT island naked DNAs demonstrated that clusters of pure AT motifs outcompete isolated mixed A/T-G/C sites in attracting drug binding. Regional preference of U-78779 for AT island domains was verified also in DNA from drug-treated cells. Thus, while the primary sequence preference gives rise to non-region-specific scattered lesions, the clustering of the minor pure AT binding motifs seems to determine region-specificity of U-78779 in the human genome. The closely correlated cytotoxic activities of U-78779 and bizelesin in several cell lines further imply that both drugs may share common cellular targets. This study underscores the significance of the genome factor in a drug's potential for region-specific DNA damage, by showing that it can take precedence over drug binding preferences at the nucleotide level.

A nonrandom distribution of DNA lesions in various domains of the genome may be critically important for the selectivity and pharmacological properties of DNA-reactive anticancer drugs. The life or death fate of drug-treated tumor cells may well be determined by a few lesions in defined DNA loci that are crucial for cell growth. Less lethal lesions elsewhere may be superfluous for cancer cell killing, yet they are likely to contribute to toxic and mutagenic effects in normal tissues.

Our recent studies identified bizelesin, a cyclopropylpyrroloindole (CPI)[1] DNA minor groove binding drug (Figure 1), as the first small molecule capable of preferential damage to defined genomic regions (*1*, *2*). These regions are long (200−1000 bp) minisatellite repeats of AT-rich DNA, referred to as "AT islands". Bizelesin's preference for AT islands is consistent with the drug's specificity at the nucleotide level. Bizelesin and several other CPI drugs strongly prefer pure AT motifs, alkylating adenine residues at their 3′ ends (*3*−*10*). The high preference of bizelesin for 5′-T(A/T)$_4$A-3′ motifs stems from the formation of interstrand cross-links mediated by the two CPI moieties (Figure 1) (*3*, *9*−*12*). Bizelesin's regional specificity reflects the nonrandom distribution of T(A/T)$_4$A motifs; while these motifs cluster in the targeted AT island, they are infrequent elsewhere in the genome (*1*, *2*).

Whereas bizelesin prefers purely A/T motifs, analogues such as U-78779 (Figure 1) were designed to accommodate some G/C pairs within their binding sites by insertion of an additional ring system (pyrrole moiety) between the two alkylating CPI moieties (*13*). Consistent with the design, the preferred binding sites for U-78779, examined in short model oligonucleotides, were found to contain 1−2 G/C bp (*13*).

† This study was supported in part by a grant from the National Cancer Institute (CA71969) and a grant from Children Cancer Center, University of Texas Health Science Center at San Antonio, San Antonio, TX. A preliminary account of this study has been presented in part at the 91st Annual Meeting of the American Association for Cancer Research, April 1−5, 2000, San Francisco, CA (*40*), and at the 92nd Annual Meeting of the American Association for Cancer Research, March 24−28, 2001, New Orleans, LA (*41*).

* To whom correspondence should be addressed: Head, Molecular Pharmacology, Cancer Therapy and Research Center, Institute for Drug Development, 14960 Omicron Dr., San Antonio, TX 78245-3217. Phone: 210-677-3832, Fax: 210-677-0058, E-mail: jmw1@saci.org.

‡ Cancer Therapy and Research Center.

§ Present address: Roswell Park Cancer Institute, Elm and Carlton St., Buffalo, NY 14263.

|| University of Arizona Cancer Center.

[1] Abbreviations: bp, base pair(s); CPI, cyclopropylpyrroloindole; dsb, double-strand break(s); MAR, matrix attachment region; nt, nucleotide(s); PCR, polymerase chain reaction; QPCR, quantitative polymerase chain reaction; RPE, repetitive primer extension; SV40, simian virus 40; TE, 10 mM Tris-HCl, 1 mM EDTA, pH 7.4; GI_{50}, drug concentration inhibiting relative cell growth by 50%.

10.1021/bi011907s CCC: $22.00
Published on Web 01/11/2002

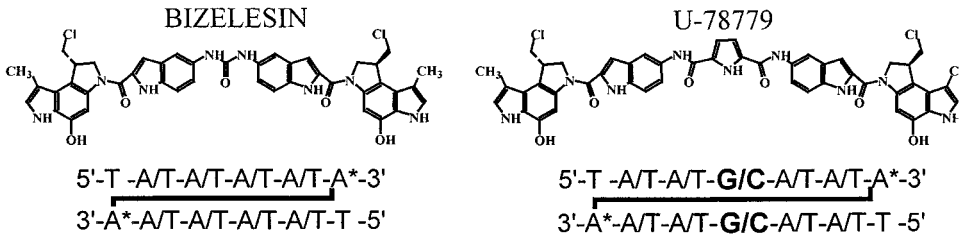

FIGURE 1: Structures of the CPI drugs bizelesin and U-78779 and examples of preferred drug binding motifs. Asterisks in these motifs indicate adducted adenine residues.

U-78779 can also form interstrand cross-links by reacting with the adenines at the 3′ ends similar to bizelesin, but the larger U-78779 molecule spans 7 bp, not 6 bp (*13*). It remained unknown, however, whether U-78779 retains the preference for mixed A/T-G/C sites in cellular systems and whether such a preference offers an advantage over drugs that bind pure AT motifs. In particular, a shift in the sequence specificity, compared to bizelesin, was likely to affect which specific regions in genome could be targeted. Indications of a high cytotoxic potency of U-78779 (*13*) further warranted a systematic examination of this drug.

Our previous studies with several drugs demonstrated that a combination of molecular pharmacology and bioinformatics tools reliably predicts the relative vulnerability of specific regions in cellular DNA (*1*, *2*, *14*, *15*). In this report, an analogous experimental and computational approach was applied to examine the DNA reactivity, including sequence- and region-specificity, of U-78779. The results confirmed that U-78779 prefers mixed A/T-G/C motifs in its binding to both naked and cellular DNA. However, these preferred binding motifs are randomly scattered in the genome and are unlikely to confer any region specificity. Nevertheless, U-78779 can damage specific regions in the human genome, but due to its minor, pure AT motifs, which cluster in the same AT islands that are preferentially damaged by bizelesin. These results underscore the significance of the genome factor in region-specific DNA damage, in addition to drug specificity at the nucleotide level.

METHODS

Drugs and Treatment Conditions. Bizelesin and its analogue U-78779 were generously provided by Dr. Robert Kelly at Pharmacia Upjohn Co. (Kalamazoo, MI). Drugs were dissolved in dimethylacetamide and stored at −20 °C protected from light. Unless noted otherwise, naked DNA and intact cells or nuclei were incubated with the indicated drug concentrations for 4 h at 37 °C. Upon heating, the N3 adenine adducts, such as those formed by CPI drugs, are cleaved between the deoxyribose at the drug-adducted adenine and the phosphate on the 3′ side (*5*). Several assays indicated below employed thermal conversion of drug adducts to strand breaks, which refers to drug-treated samples being additionally heated for 15 min at 95 °C.

Drug Reactivity with Purified SV40 DNA. The reactivity of U-78779 with naked circular SV40 DNA was assessed based on thermal conversion of drug adducts to strand breaks by following the changes in DNA topological forms (*15*, *16*). Briefly, following drug treatment, and solvent extraction to remove unreacted drug, samples were heated to convert DNA adducts to strand breaks, and topological forms of SV40 DNA were separated by agarose gel electrophoresis. Gels were stained with Sybr Green I (Molecular Probes, Inc.) and photographed with a UV transilluminator. The negatives were scanned in a Molecular Dynamics laser densitometer. Bands corresponding to Form I (supercoiled DNA), Form II, (relaxed circular), and Form III (linear) were quantitated using Image Quant software (Molecular Dynamics). For each lane, Form I DNA was normalized first to total DNA in that lane and next to the fraction of Form I in mock-treated samples to yield the fraction of remaining form I (FI_R). The frequencies of drug lesions were estimated based on Poisson distribution using the equation:

$$f = [-\ln(FI_R)]/L \quad (1)$$

where f is the number of breaks per base pair and L is the length of SV40 DNA (5243 bp).

Total Adducts in Bulk Cellular DNA of Intact Cells. Covalent adducts of U-78779 in genomic DNA of intact human leukemia CEM cells (from Dr. William T. Beck, University of Illinois, Chicago) were quantitated following thermal conversion to strand breaks by an alkaline sucrose gradient sedimentation procedure as described elsewhere (*1*, *15*). Briefly, CEM cells were prelabeled with [^{14}C]thymidine and then incubated with drugs as indicated. After drug treatment, cells were harvested and lysed in 1% Sarkosyl, 1 M NaCl, 10 mM EDTA, pH 8.6. Following heating to convert DNA adducts to strand breaks, aliquots of cell lysates were loaded onto alkaline sucrose gradients consisting of a 0.5 mL of 1% Sarkosyl/2.5% sucrose lysing layer, a 10 mL 5−30% sucrose layer, and a 0.5 mL 60% sucrose cushion (all in 2 mM EDTA, 300 mM NaOH, 700 mM NaCl). Following an additional lysis time (1 h at 20 °C), gradients were centrifuged at 36500*g* for 19 h in an SW40Ti rotor (Beckman Instruments) followed by fractionation and processing of gradient fractions as described (*16*).

The frequency of drug-induced heat-labile sites (*f*, the number of breaks per base pair, which is proportional to drug adducts) was estimated from the shifts in the sedimentation profiles toward the top of the gradients as described (*16*, *17*) with the modification that external ^{14}C-radiolabeled controls were used (*1*, *15*) as indicated by eq 2:

$$f = \{[(MW_n)_A/(MW_n)_B] - 1\}/[(MW_n)_A/640] \quad (2)$$

where $(MW_n)_A$ and $(MW_n)_B$ are the number average molecular weights of control and drug-treated ^{14}C-radiolabeled DNA peaks, respectively. The value 640 represents the average molecular weight of one DNA base pair.

Interstrand Cross-Links in Nuclei from CEM Cells. Interstrand cross-link determinations were carried out as described previously using sedimentation analysis in alkaline sucrose gradients (*10*, *18*). Briefly, nuclei were isolated from cells

prelabeled with [^{14}C]-thymidine under isotonic conditions exactly as described previously (*18*) and incubated with drugs for the indicated times. Aliquots of drug-treated nuclei were loaded onto alkaline sucrose gradients composed as stated above for total adduct determinations, except that a 5−20% sucrose gradient was used. Following a 10 h lysis, sedimentation was carried out in an SW41 rotor (Beckman Instruments) at 20 °C for 11 h at 36500*g*. The frequency of drug-induced interstrand cross-links was estimated from the shifts in the sedimentation profiles toward the bottom of the gradients using the mathematical treatment of Roberts and Friedlos (*19*) (eq 3):

$$f = [-\ln(1 - \mathrm{FD})]/[\mathrm{MW}_n/640] \quad (3)$$

where MW_n is the number average molecular weight for control DNA, 640 is the average molecular weight of 1 bp, and FD, Fractional Difference, is the sum of the differences between the control and treated DNA profiles computed as described elsewhere (*10*).

Sites of Drug Adducts: Repetitive Primer Extension (RPE). The sites of drug adducts in intracellularly treated SV40 were determined as described previously (*10*, *14*, *15*). Briefly, SV40 virus-infected green monkey BSC-1 cells were treated with U-78779, and SV40 DNA was isolated. Localization of U-78779 adducts was determined based on prematurely terminated primer extensions on drug-modified DNA template in RPE reactions (*10*, *14*, *15*).

Primers used in RPE, referred to as MAR and NO, covered the regions that are relatively enriched and devoid of AT runs, respectively. The sequences of MAR and NO primers, their preparation and 5′-end-labeling, RPE cycling conditions, and product analysis were described in detail elsewhere (*1*, *14*, *15*). The positions of drug-induced stop sites, determined based on dideoxy sequencing ladders run in the same gel, are assumed to be within ±1−2 bp from the positions of adducted bases (*10*).

In Silico Motif Distribution Analysis. The distribution of potential drug binding sites at the genomic level was analyzed *in silico* as described previously (*2*) by tabulating the positions of the exact matches to specified binding motifs in each sequence to generate hit distribution histograms (such as in Figure 6). The total length of DNA evaluated for U-78779 binding motifs was approximately 2×10^7 bp, which well exceeds the minimal number of hits ($\sim 10^5$) needed for stable distribution parameters (*2*).

Levels and Sites of Drug Adducts in Model Linear DNA. The sites of drug adducts were mapped in a model AT island as described elsewhere for bizelesin adducts (*2*). The PCR-generated AT island DNA used in these experiments consisted mainly of the 859 bp GenBank AF385609 sequence, with ∼15% of the longer variant corresponding to a 1025 bp AT island in GenBank Z79699. Since the flanks of both variants show complete homology over several hundred base pairs, drug adducts in either product generate identical end-labeled subfragments in those areas of the gel where precise positions of drug adducts can be assessed (*2*).

Drug-induced lesions in model linear DNAs were quantitated by agarose gel electrophoresis following thermal conversion of drug adducts to strand breaks exactly as described previously (*2*). Under these assay conditions, drug-induced interstrand cross-links and/or closely spaced mono-adducts give rise to double strand breaks in DNA. DNA fragments for these experiments (the 859 bp GenBank: AF385609, and 536 bp β-globin fragment) were either ^{32}P end-labeled or ^{32}P uniformly labeled as described (*2*). After phoshorimaging of the dried agarose gels, the disappearance of full-length fragments was quantitated using ImageQuant software (Molecular Dynamics). Adduct frequencies were estimated based on Poisson distribution, analogous to eq 1, except that FI_R was substituted by the remaining fraction of full-length DNA.

Region-Specific Lesions in Cellular DNA. QPCR stop assay monitored drug adducts based on the progressive elimination of the amplified DNA signal (*1*, *2*, *14*, *15*). The QPCR determinations for (i) the above-mentioned AT island variants in GenBank Z79699/AF385609 and (ii) a non-AT-rich region in the β-globin gene were run as previously described (*1*, *2*, *15*) except that both pairs of primers were used together (at 0.1 and 0.025 μM, respectively) in a multiplex format. PCR was carried out with initial denaturation at 95 °C for 60 s followed by 27 cycles of 94 °C for 30 s, 55 °C for 15 s, and 60 °C for 45 s and a terminal extension at 60 °C for 3 min. For Z79699/AF385609, both 859 and 1025 bp products were quantitated together.

Cytotoxic Activity in Human Tumor Cell Lines. Drug cytotoxic activities were assayed in CEM and several other human cell lines: Colo320DM colon carcinoma cells, MCF-7 and MDA-MB-231 breast cancer, and A2780 ovarian carcinoma purchased from American Type Culture Collection (Manassas, VA), and LnCAP-Pro5 prostate cancer cells (generously provided to Dr. B. Woynarowska by Dr. C. A. Pettaway at M. D. Anderson Cancer Center). All cells lines were cultured either as described previously [CEM, LnCaP-Pro5 (*20*, *21*)] or as recommended by the provider.

Drug effects on cell proliferation were measured as described previously (*14*) by a standard colorimetric assay that follows the accumulation of a formazan dye generated by the conversion of 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) by mitochondrial dehydrogenases in living cells (*22*). Exponentially growing cells (∼3000 cells/well in a 96-well plate) were incubated with a series of drug concentrations for 72 h. MTT was then added, and the cells were incubated an additional 4 h. Absorbance at 595 nm was measured on a BioRad model 3550 microplate reader. The results were expressed as GI_{50} values (drug concentrations inhibiting relative cell growth, RG, by 50%), using the equation: $\mathrm{RG} = (\mathrm{Tr} - T_0)/(\mathrm{Con} - T_0)$, where Tr and Con are the MTT absorbance signals for drug-treated and control wells, respectively, and T_0 is the initial (time zero) absorbance.

RESULTS

This report consists of three parts. Initial studies examined the overall ability of U-78779 to react with naked and cellular DNA, and identified its preferred binding motifs. Next, a long-range *in silico* analysis was used to assess potential genomic targets for various identified binding motifs. Finally, the *in silico* predictions were experimentally verified by determining drug-induced lesions in model regions, both in drug-treated cells and in naked DNA.

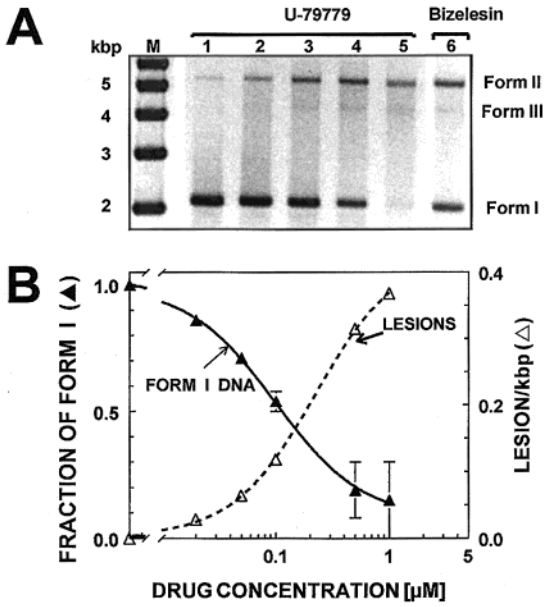

FIGURE 2: Induction of thermally labile adducts by U-78779 in naked SV-40. Panel A: Representative nondenaturing agarose electrophoresis of SV40 DNA incubated for 4 h at 37 °C with U-78779 at 0 μM (lane 1), 0.02 μM (lane 2), 0.05 μM (lane 3), 0.1 μM (lane 4), and 0.5 μM (lane 5). Lane 6 contained DNA treated with 0.05 μM bizelesin. All the SV40 samples were heated 15 min at 95 °C after drug treatment to convert drug adducts to strand breaks. The proportions of Forms I and II in unheated drug-treated samples did not differ from those in control samples (not shown). Lane denoted "M" shows size markers. Panel B: Quantitation of thermally labile adducts based on topological forms conversion after 4 h incubation of SV40 DNA with drugs. The topological conversion of Form I, normalized as the remaining fraction of Form I initially present (FI_R, solid lines, ▲), is shown. The results are averaged values (±SE) from 3 independent experiments, except for 1 μM drug (2 experiments). The corresponding frequency of U-78779 lesions is also shown (dashed line, △).

(A) Characterization of DNA Reactivity, Types of Lesions, and Preferred Motifs of U-78779

U-78779 Is Nearly as Reactive as Bizelesin with Naked SV40 DNA. DNA adducts formed by CPI drugs can be quantitatively converted to strand breaks upon heating. Thus, by monitoring these heat-induced strand breaks, one can quantitate the original DNA lesions (*15*, *16*). We used this strategy first to compare the overall reactivity of U-78779 and bizelesin with naked DNA. SV40 DNA, used as a model, offers a balanced AT/GC composition (59.2 and 40.8%, respectively) that is close to that of the entire human genome. With this supercoiled DNA, a single strand break effects the conversion of supercoiled Form I to relaxed circular Form II, whereas double strand breaks would result in conversion to linear Form III.

Treatment with U-78779 drastically reduced the percentage of Form I with concurrent increase in Form II (Figure 2A). The effect could be detected at U-78779 concentrations as low as 0.02 μM with nearly all Form I DNA disappearing at 0.5 μM drug. The disappearance of Form I DNA is proportional to an increasing number of drug adducts (Figure 2B). U-78779 produces an estimated 1.35 ± 0.07 lesions kbp^{-1} μM^{-1} (mean ± SE), based on the initial slope of the lesions frequency plot (for 0.05−0.1 μM drug). Under analogous conditions, estimated bizelesin lesions in SV40 DNA amount to 2.05 ± 0.2 lesions kbp^{-1} μM^{-1} [Figure 2A and data not shown, also data from (*16*)]. Thus, U-78779 is somewhat less reactive with SV40 DNA than bizelesin.

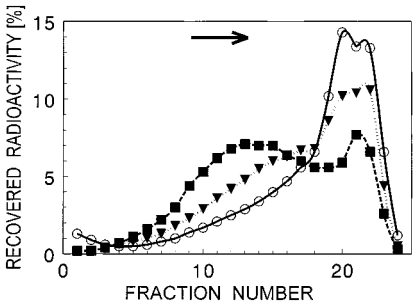

FIGURE 3: Sedimentation analysis of thermally labile DNA adducts in U-78779-treated CEM cells. Cells were treated with 0 (○), 0.2 (▼), or 0.4 μM U-78779 (■) for 4 h at 37 °C. Following cell lysis and thermal conversion of drug adducts to strand breaks, DNA was analyzed by sedimentation in alkaline sucrose gradients. Sedimentation profiles shown are from a representative experiment from the 4 independent experiments performed. The arrow indicates the direction of sedimentation.

U-78779 Binds to Genomic DNA of CEM Cells and Forms Interstrand Cross-Links in Isolated Nuclei. The ability of U-78779 to damage genomic DNA of CEM cells was monitored by sedimentation analysis on alkaline sucrose gradients following heat-induced conversion of DNA adducts to strand breaks (*15*, *16*). The observed slower sedimentation of DNA from U-78779-treated cells is indicative of thermally induced strand breakage and proportional to the frequency of drug−DNA adducts (Figure 3). U-78779 forms an estimated 0.22 ± 0.09 lesions kbp^{-1} μM^{-1} (mean ± SE), compared to the previously determined value of 0.87 ± 0.04 lesions kbp^{-1} μM^{-1} for bizelesin (*2*).

The two CPI moieties of U-78779 should allow for interstrand cross-links between suitably spaced 3′-adenine residues (*13*). The U-78779 potential to cross-link genomic DNA was investigated in the nuclei of CEM cells treated with the drug and then directly subjected to sedimentation on alkaline sucrose gradients (no conversion of adducts to breaks, Figure 4). Under these conditions, monoadducts by CPI drugs do not affect DNA sedimentation (*17*). By contrast, interstrand cross-linking prevents alkali-induced strand separation, resulting in the appearance of a rapidly sedimenting DNA (*10*). Like bizelesin, U-78779 also induces such shifts in the sedimentation profiles of DNA from drug-treated nuclei (Figure 4). Thus, U-78779 induces DNA interstrand cross-links in nuclear DNA. Estimates based on the sedimentation profiles suggest that U-78779 forms ∼1.7 cross-links kbp^{-1} μM^{-1} compared to ∼1.9 cross-links kbp^{-1} μM^{-1} for bizelesin.

U-78779 Adducts Are Preferentially Formed at Mixed A/T-G/C Sites in Intracellular SV40 DNA. Having established the overall DNA reactivity of U-78779, we next mapped the sites of U-78779 adducts using the repetitive primer extension (RPE) assay and SV40 DNA from drug-treated, SV40-virus-infected cells. The utility and advantages of this system have been shown previously (*10*, *15*, *23*, *24*). In the RPE assay, drug adducts in the template strand prematurely terminate *Taq* polymerase-catalyzed primer extension, resulting in nascent strands whose length indicates the position of drug adducts with a resolution of 1−2 bp (*10*, *14*). RPE analysis for U-78779 adducts covered a total of nearly 450 bp of

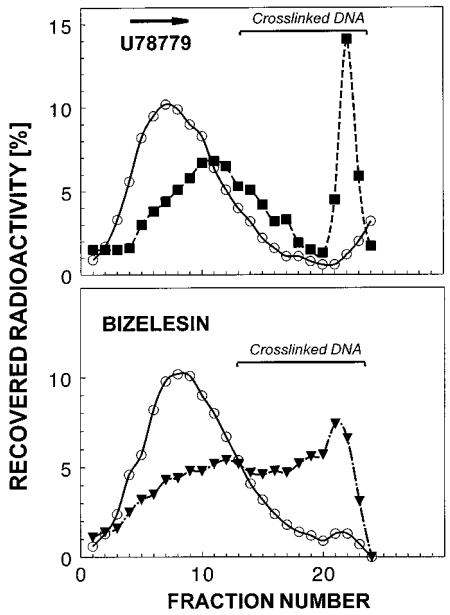

FIGURE 4: Interstrand cross-links in nuclei from CEM cells treated with U-78779 and bizelesin. Nuclei isolated from CEM cells were treated for 4 h at 37 °C with no drug (○), 0.02 μM U-78779 (■), or 0.02 μM bizelesin (▼). Nuclei were then subjected to sedimentation analysis in alkaline sucrose gradients. An increase above control in the fastest sedimenting material (horizontal bar) is indicative of cross-linked DNA. Sedimentation profiles shown are from a representative experiment from the 2 independent experiments performed. The arrow indicates the direction of sedimentation.

SV40 DNA in a relatively AT-rich MAR region and a non-AT-rich area (NO primer).

The results of the RPE assay clearly demonstrate that lesions formed by U-78779 in intracellular SV40 DNA are highly sequence-specific. The RPE reactions with SV40 DNA from untreated control cells (Figure 5A, lane 1) resulted mainly in long unimpeded extension products that migrated near the top of the gel (beyond the region shown in Figure 5A). In contrast, DNA samples from drug-treated cells produced several distinct bands corresponding to shorter, prematurely terminated, nascent strands. The examples of such bands for NO primer in Figure 5A (lanes 2 and 3) identify drug adducts formed at several mixed G/C-A/T sequences. The strongest site identified in this region (at position 4165) features a centrally located GC pair. Thus, U-78779 can bind mixed A/T-G/C motifs in intracellular DNA.

An increased tolerance of U-78779 for G/C base pairs is evident given that approximately three-fourths of all positively identified U-78779 adduct sites were found within various mixed A/T-G/C motifs containing 1 or 2 G/C pairs (Figure 5B). An alternative adduct detection, based on thermally induced strand breaks of a drug-treated naked PCR product corresponding to the SV40 NO region, confirmed the definite preference of U-78779 for mixed A/T-G/C sites (fragment length analysis by capillary electrophoresis using an automated sequencing system, data not shown). However, 25% of all sites mapped by RPE were found in pure A/T runs (Figure 5B). This ratio closely matches the previously observed pattern of sites with model naked oligonucleotides (*13*). Thus, U-78779 retains some ability to react within pure AT tracts.

Several of the identified submotifs for U-78779 adducts (Figure 5B) are consistent with the formation of interstrand

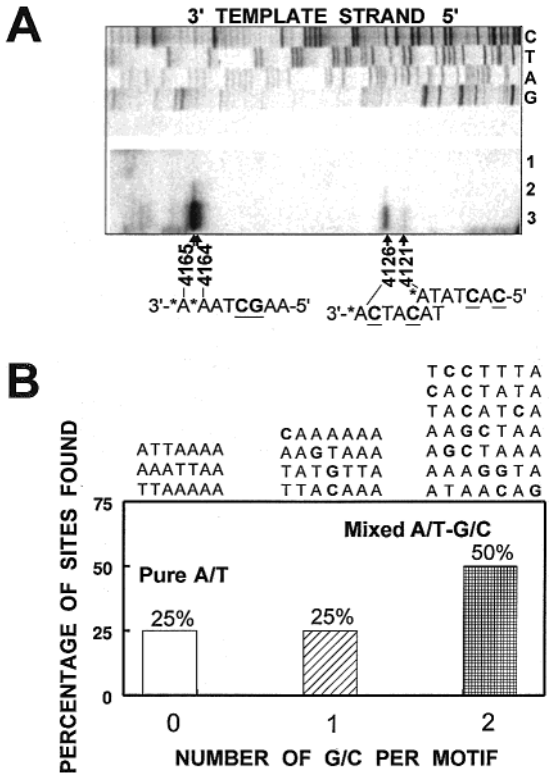

FIGURE 5: Sequence specificity of U-78779 in intracellular SV40 DNA. BSC-1 cells infected with SV40 virus were treated with the drug for 4 h, followed by the isolation of SV40 DNA and mapping of the sites of drug adducts using repetitive primer extension (see Methods for details). Panel A: Representative example of RPE (a fragment of a sequencing polyacrylamide gel) using a primer for the MAR region derived from the lower strand of SV40 sequence. Template SV40 DNA was from cells treated with U-78779 at 0 μM (lane 1), 0.5 μM (lane 2), and 5 μM (lane 3). Dideoxy sequencing reactions generated with the same primer and untreated SV40 DNA as a template are shown in lanes designated C, T, A, and G (denoting bases in the template, i.e., top strand of the SV40 sequence). Specific stop sites are marked with both position and sequence. Asterisks indicate putatively adducted bases. Panel B: List of sites identified by RPE containing 0, 1, and 2 G/C bases (top) and their relative abundance (bottom). All the listed sites reproducibly appeared in at least two independent experiments.

cross-links between adenines at the 3′ ends of drug binding sites [5′-T(N)$_5$A-3′]. Two additional sites might reflect cross-links involving a G adduction at one arm, a possibility that has been previously postulated (*13*). Collectively, the RPE data indicate that binding specificities of U-78779 in intracellular DNA are distinct but complex; the drug is likely to bind to both cross-linking and monoadduct sites, preferring mixed A/T-G/C motifs, but pure AT motifs (both cross-linking and monoadduct) need to be considered as minor binding sites.

(B) In Silico Long-Range Sequence Analysis for Drug Binding Motifs

Having established drug binding preferences with intracellular DNA, we analyzed *in silico* the distribution of U-78779 binding motifs in DNA sequences of the human genome. For various drugs of diverse binding preferences at the nucleotide level, this approach has been proven to provide reliable predictions of which regions of the genome are likely and which are unlikely to be preferentially targeted (*1, 2, 14, 15*).

Table 1: Long-Range *in Silico* Sequence Analysis for the Distribution of Possible Binding Sites for U-78779[a]

drug	motif searched	predicted[b]	average hits/0.25 kbp (±SD)	"Hot" loci[c] hits/0.25 kbp	no. of "Hot" loci/Mbp	ratio of "Hot" loci to average	expected region specificity
U-78779	generalized mixed A/T-G/C (T/A)(T/A$_x$G/C$_y$)A[d]	26.8	22.1 ± 11.4	90−108	0.3	4−5	no
U-78779	pure AT (cross-links + monoadducts) (T/A)$_6$A	5.3	6.8 ± 7.2	112−232	0.8	16−34	yes
bizelesin[e]	T(A/T)$_4$A (cross-links)	2.9	2.8 ± 3.6	40−99	1.0	14−34	yes

[a] Human DNA sequences covering ~20 × 10^6 bp were analyzed for the distribution of the indicated motifs. The "hits" recorded are exact matches to these motifs on both DNA strands and are given as average values per 0.25 kbp sequence sections ("bins", cf. Figure 2). [b] Predicted frequency of specific motifs was calculated based on the probability of a random occurrence of such motifs (*39*) assuming the overall DNA composition of 60% AT and 40% GC. [c] Defined as loci in which peak hits/0.25 kbp exceeded the average peak hits/0.25 kbp in all the sequences analyzed by more than 2.5 × SD (rounded). The average maximum for the generalized mixed A/T-G/C motifs was 57.6 ± 12.5 and was 44 ±27.4 for the pure AT motif. [d] All the permutations with $y = 1$ or 2 vicinal G/C pairs in the central 5 bp section ($x + y = 5$) of the 7 bp binding site. [e] Data from reference (*2*).

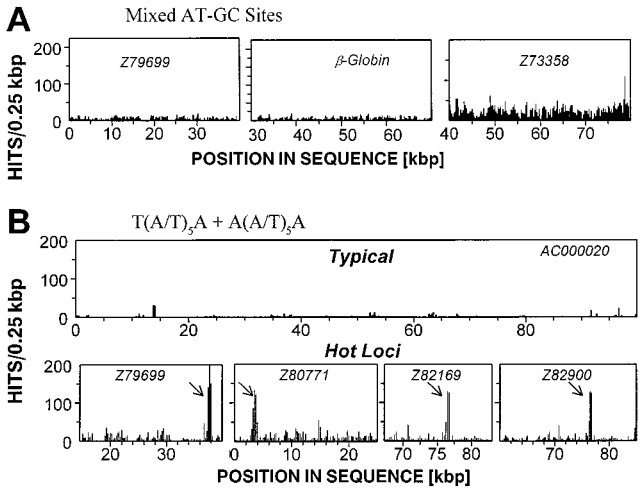

FIGURE 6: Examples of the *in silico* analysis for long-range distribution of potential binding sites for the U-78779 predominant mixed G/C-A/T motif (A) and minor pure AT motifs (B). The histograms depict the number of occurrences of drug binding motifs in bins of 250 bp along the indicated GenBank sequences. Arrows point to the peaks of local density of pure AT binding motifs that localize to long (200−1000 kbp) AT islands.

Mixed A/T-G/C Motifs Do Not Confer Region Specificity. We analyzed first the prevailing, mixed A/T-G/C motifs of U-78779. In addition to the explicit limited submotifs listed in Figure 5B, we used a broader, generalized consensus that covered all permutations with 1 or 2 vicinal GC within the 7 bp span and included both cross-linking as well as monoadduct sites. The results of this search, summarized in Table 1, show that the collective abundance of all the mixed A/T-G/C permutations is modestly high (on average 22.1 hits/250 kbp). This value is close to the frequency anticipated for a random occurrence of the examined motifs in the genome (26.8 hits/250 bp). Moreover, all the analyzed U-78779-preferred mixed A/T-G/C motifs (~1.7 × 10^6 hits in ~20 × 10^6 bp) were fairly randomly distributed. Occasional loci with modestly elevated numbers of hits are encountered approximately once per 3 Mbp, but the maximal hit density (90−108 hits/0.25 kbp) exceeded only 4−5-fold the genome-wide average (Table 1). Examples of such rather unremarkable motif distribution for a few GenBank loci are shown in Figure 6A. It needs to be noted that an analogous pattern of nearly uniformly distributed sites was found when selected mixed A/T-G/C submotifs were analyzed individually, except that, as anticipated, the occurrences of individual submotifs were proportionately less frequent (data not shown). Given the known overall preference of CPI drugs for adenine alkylation, the motifs that assumed a G-A cross-link, like some of the mentioned actual sites (viz., Figure 5B), were not included in the generalized consensus motif. When such motifs were separately analyzed (data not shown), they were also only marginally elevated at best in some genomic loci. In light of this unremarkable distribution of the mixed A/T-G/C motifs, it can be concluded that binding of U-78779 to its preferred sites is devoid of any region specificity.

Pure AT Motifs Clustered in Long AT Islands May Confer Region Specificity. Another round of *in silico* search analyzed the genomic distribution of the pure AT minor binding motifs of U-78779, with submotifs covering both cross-links and monoadducts. In profound contrast to the results for mixed A/T-G/C sites, the less preferred pure A/T motifs were found to be nonrandomly distributed (Table 1). The genome-wide average frequency (±SD) of these motifs (6.8 ± 7.2 hits/250 bp) was close to the value anticipated for random occurrence (5.3 hits/250 bp). However, a number of loci were identified in which these motifs clustered with a hit density of 112−232 hits/250 bp, which is 16−34-fold greater than the genome-wide average. These clusters of binding motifs were found approximately every 0.8 × 10^6 bp. Consequently, the distribution parameters for the minor pure AT sites of U-78779 strikingly resemble those for bizelesin sites (Table 1). Thus, like bizelesin, the minor U-78779 sites cluster in discrete genomic loci. Moreover, the domains with the clusters of U-78779 pure AT sites (see examples in Figure 6B) exactly coincide with the AT islands identified as a preferential target for bizelesin (*2*). If pure AT cross-linking and monoadduct motifs are analyzed separately, both distributions practically overlap at the genomic level (data not shown). Hence, the *in silico* results taken together suggested that U-78779 binding to the minor, pure AT motifs could confer region specificity at the genomic level for AT islands as a preferred target.

(C) Experimental Verification of the in Silico Predictions

The outcome of *in silico* analysis was verified using a model AT island (the 859 bp locus GenBank: AF385609), and a model non-AT-rich domain (a 536 bp fragment of the β-globin gene) (*2*). The 859 bp AT island DNA has 384

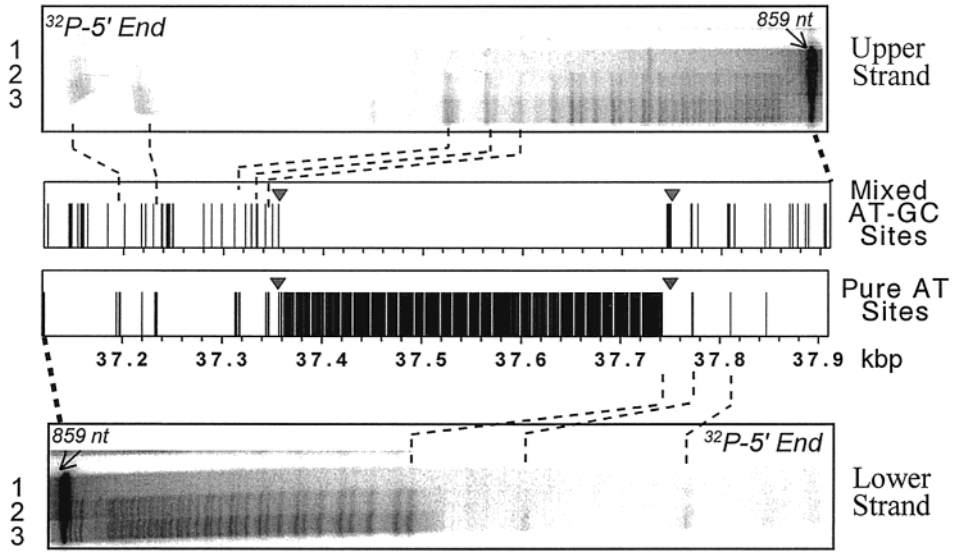
32P-5′ End
859 nt
1
2
3
Upper Strand
Mixed AT-GC Sites
Pure AT Sites
37.2 37.3 37.4 37.5 37.6 37.7 37.8 37.9 kbp
859 nt
32P-5′ End
1
2
3
Lower Strand

FIGURE 7: Predicted and actual sites of U-78779 adducts in the naked DNA of a model AT island. The distribution of the predicted mixed G/C-A/T and pure AT U-78779 binding sites is shown in the AF385609 AT island as spikes corresponding to every occurrence of the indicated motifs. Triangles point to the positions of the last G/C base pairs flanking the pure AT core. The actual drug sites were determined in AT island DNA (singly ^{32}P-end-labeled at either 5′ end) following treatment for 4 h at 37 °C with U-78779 at 0 μM (lanes 1), 0.005 μM (lanes 2), and 0.02 μM (lanes 3), thermal conversion of drug adducts to strand breaks, and electrophoresis on 8% polyacrylamide/urea gels. The autoradiographic images of gel lanes (examples for 4 independent experiments) are oriented to align the labeled 5′ ends of the fragments with regard to the DNA sequence. The bands corresponding to full-length 859 nucleotide fragments reflect sequence positions of 37101 and 37959 for the labeled lower and upper strand, respectively. The numbering scheme follows that of the longer variant of that AT island in GenBank Z79699, assuming that position 1 of AF385609 corresponds to position 37101 in GenBank Z79699.

pure AT sites, mostly clustered in the 100% AT central core, and 111 of the mixed A/T-G/C type located at the flanks of that sequence (Figure 7, central panel). The sequence of the β-globin fragment has only 28 potential U-78779 binding sites available, all of them being of the mixed A/T-G/C type. U-78779 ability to damage these sequences was examined first using naked model DNA fragments and was next corroborated with genomic DNA from drug-treated cells.

Agreement between the Predicted and Actual Sites of U-78779 Adducts in a Model AT Island DNA. U-78779 binding sites were mapped in the naked DNA of the model AT island sequence. Following drug treatment and the conversion of adduct to strand breaks, sequencing polyacrylamide gels were used to characterize the distribution of actual drug binding sites on both strands.

The results demonstrate an excellent agreement between the predicted and actual sites of drug adducts (Figure 7). The flanks of this AT island sequence contain fairly numerous mixed A/T-G/C sites, and some of the identified adducts indeed correspond to such sites. However, the bulk of U-78779 adducts are found within the 100% A/T core, as indicated by very strong, multiple bands seen for both top and bottom strands. The observed massive damage to the 100% A/T core follows closely the distribution of the predicted pure AT binding sites (Figure 7, central panels). Accordingly, the majority of sites whose exact positions could be identified were found to co-locate with pure AT motifs.

U-78779 Preferentially Damages a Model AT Island Naked DNA. To assess the overall U-78779 binding to model naked DNAs, drug adducts formed in respective fragments were thermally converted to strand breaks. Under this heat treatment, drug-induced cross-links, as well as closely spaced monoadducts, are expected to give rise to double stranded breaks (dsb) in DNA. The resulting fragments were thus analyzed by nondenaturing agarose electrophoresis, and the disappearance of full-length fragments was used to assess the reactivity of U-78779 (Figure 8A,B). Whereas this assay does not differentiate between interstrand cross-links (a main lesion type for bizelesin and presumably also for U-78779) and closely spaced monoadducts, it provides a convenient measure of the overall reactivity of the drug with specific target sequences.

In the AT island DNA, the first drug-induced thermally inducible dsb are detected at 0.005 μM U-78779, with a corresponding frequency of ~0.6 lesion/kbp. At 0.1 μM drug, there is almost no full-length DNA remaining, reflecting >2.3 lesions/kbp. Drug effects in non-AT island DNA, which has only mixed A/T-G/C sites, are much less dramatic. Only a marginal level of lesions (~0.2 dsb/kbp) was observed in the β-globin sequence at 0.1 μM drug, and even at drug concentrations as high as 0.5−1 μM, lesion frequency remained below 1 dsb/kbp. Normalization to the number of drug lesions per unit of drug concentration (initial slope of concentration dependence) reveals that with naked DNA, U-78779 is ~65-fold more reactive toward AT island sequences than toward non-AT island fragment (116 ± 13 vs 1.8 ± 0.3 lesions kbp^{-1} (μM drug)$^{-1}$, respectively, mean ± SE).

The preferential binding of U-78779 to AT island over non-AT island DNA is further corroborated by competition experiments in which an excess of either sequence in unlabeled form was used as a competitor for drug binding to labeled DNA (Figure 8C). A 5-fold excess of unlabeled competitor AT island DNA resulted in nearly complete abrogation of drug adduct formation in labeled material. Nearly 100% of labeled AT island DNA remained as the full-length material in the presence of unlabeled competitor, compared to only ~16% remaining in its absence. Unlabeled non-AT island DNA (a β-globin fragment) also afforded

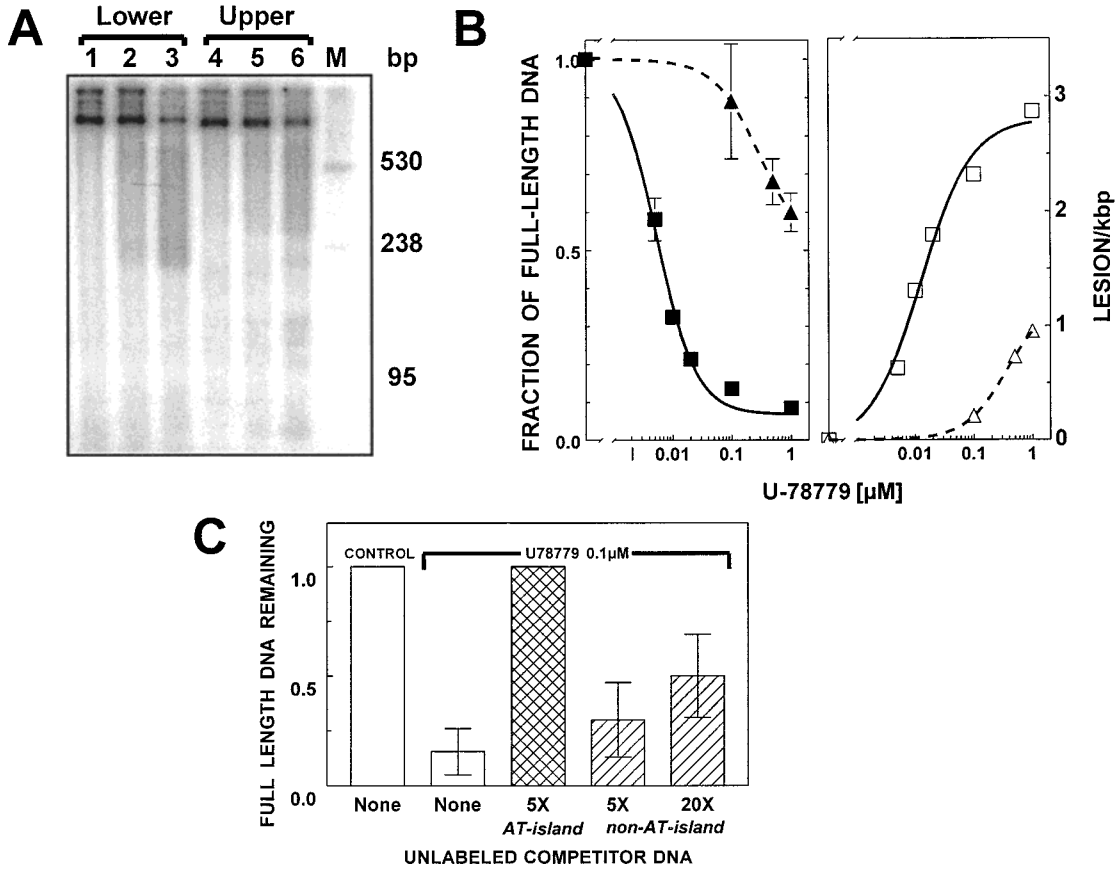

FIGURE 8: Preferential formation of U-78779 adducts in naked DNA of AF385609 AT island versus a non-AT island β-globin sequence. Singly ^{32}P-end-labeled DNA was generated with either upper or lower labeled primer and drug-treated for 4 h at 37 °C with U-78779. Drug adducts were thermally converted to strand breaks, and the resulting fragments were analyzed in nondenaturing agarose gels. Panel A: Representative agarose gel of AF385609 DNA treated with U-78779 at 0 μM (lanes 1, 4), 0.005 μM (lanes 2, 5), and 0.02 μM (lanes 3, 6). Lane M denotes size markers. Panel B: Quantitation of drug-induced disappearance of full-length DNA (left, solid symbols, mean ± SE from 2−6 determinations carried out either with ^{32}P-end-labeled or with uniformly labeled DNAs) and the corresponding frequency of double strand breaks/kbp (right, open symbols) in AF385609 AT island DNA (■, □) and a non-AT island β-globin sequence (▲, △). Panel C: Competition by unlabeled AT island and non-AT island DNA. The ordinate shows the fraction of remaining full-length labeled AT island DNA (normalized to untreated control, open bar) after treatment with the drug. Drug treatment was carried out as in panels A and B in the absence (closed bars) or presence of excess unlabeled AT island DNA (cross-hatched bars, 5× excess) or unlabeled non-AT island β-globin fragment (right-hatched bars at 5× and 20× as indicated).

some protection of the labeled DNA from adduct formation but was obviously less efficient even at a 20-fold excess. In terms of the number of competing mixed A/T-G/C binding sites, the 5-fold excess of AT island DNA is equivalent to the 20-fold excess of unlabeled β-globin fragment (111 and 28 sites/molecule, respectively). However, the AT island DNA contains in addition numerous competing pure AT sites that are absent from the β-globin fragment. Thus, the disproportionate ability of the AT island DNA to compete for drug binding corroborates the significant contribution of pure AT sites in the binding of U-78779 to DNA.

U-78779 Preferentially Damages a Model AT Island in Drug-Treated Human Cancer Cells. The regional preferences of U-78779 in drug-treated cancer cells were followed by determining drug-induced lesions in the model AT island and non-AT island regions of the cancer cell genome using a multiplex QPCR stop assay. The results show that U-78779 treatment of CEM cells progressively impedes the amplification of both domains, which is indicative of adduct formation in both genomic regions (Figure 9). It is clear, however, that the AT island region is more severely affected, with a drastic reduction in PCR signal at drug concentrations between 0.05 and 0.1 μM (Figure 9A). The corresponding initial slope in the lesion frequency curve amounts to an estimated 5.7 lesions kbp^{-1} μM^{-1}. In contrast, the amplification of the model non-AT island (in the β-globin locus) is less affected (Figure 9B), resulting in an estimated initial slope of only 0.8 lesion kbp^{-1} μM^{-1}. Thus, the actual frequency of U-78779 lesions in the AT island is ~7-fold greater than the respective frequency in the non-AT island region. This comparison demonstrates that clusters of pure AT sites, such as in the model AT island, can effectively compete with scattered mixed AT-G/C sites for the binding of U-78779 in cellular settings.

The Pattern of Cytotoxicity of U-78779 Correlates with Bizelesin Cytotoxicity. The *in silico* analysis and the determinations of region-specific lesions suggested that the pure AT fraction of U-78779 adducts could be formed in the same domains that are presumably responsible for hyper-lethality of bizelesin lesions. If U-78779 indeed targets the same critical domains in the genome as bizelesin, cytotoxic responses to both drugs should correlate. The determinations of drug cytotoxic activities in several cell lines demonstrated that U-78779 retains the extreme cytotoxicity seen with bizelesin, with GI_{50} values in the picomolar range (Figure 10). In addition, cell lines that were most sensitive to

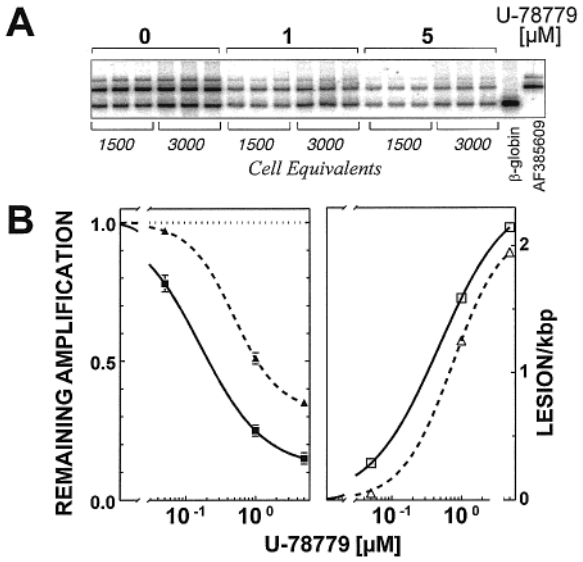

FIGURE 9: U-78779-induced lesions in specific regions of cellular DNA. CEM cells were incubated with U-78779 for 4 h followed by DNA isolation and determination of region-specific damage by QPCR stop assay. Panel A: Multiplex PCR: examples of agarose electrophoresis of amplified DNA from control cells and cells treated with 1 and 5 μM U-78779 using primers for the AF385609 system and a 536 bp fragment in the β-globin gene. The determinations were carried out in triplicate using two different DNA amounts (expressed as cell equivalents). PCR reactions of control DNA with the primer systems individually are to the right. Panel B: Quantitation of region-specific DNA damage in the model AT island AF385609 (■, □) and in the model non-AT island in the β-globin gene (▲, △). Amplification of DNA from U-78779-treated cells is normalized relative to the amplification of DNA from control cells (relative amplification, closed symbols, left panel). Lesions/kbp calculated from a Poisson distribution of the data are plotted in the right panel (open symbols). The data represent average values (±SE) from 3 independent experiments performed in multiplex QPCR as in the examples shown in panel A.

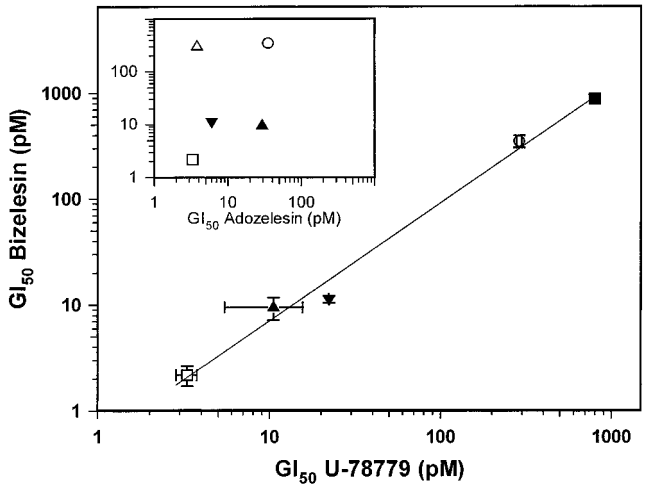

FIGURE 10: Correlation between the cytotoxic activities of U-78779 and bizelesin against cancer cell lines. Drug cytotoxicity was assessed by MTT assay after 72 h treatment of the following cell lines: CEM (□), LNCaP-Pro5 (▲), A2780 (▼), Colo320DM (○), and MDA-MB-231 (■) or MCF-7 (△). The plot shows average GI_{50} values (±SEM) from 2−8 and 1−5 experiments for U-78779 (ordinate) and bizelesin (abscissa), respectively. The regression line reflects the correlation coefficient $r = 0.99$. The inset compares GI_{50} values (±SEM contained within symbols) from 1−4 experiments for adozelesin (ordinate) and bizelesin (abscissa), which show no correlation ($r < 0.01$).

bizelesin were also most sensitive to U-78779. The correlation between cytotoxic responses to both drugs is highly significant ($r = 0.99$, $p < 0.0005$). By contrast, no such correlation ($r = 0.01$, Figure 10 inset) was seen between the cytotoxic activities of bizelesin and adozelesin, a closely related drug that is less sequence specific and binds promiscuously in various genomic regions with only a marginal preference for AT islands (*1*).

DISCUSSION

Understanding how to rationally target specific DNA domains in the genome is of broad significance for the development of more selective anticancer drugs. Bizelesin, an anticancer drug that preferentially targets long (200−1000 bp) islands of AT-rich DNA, provides a paradigm that region-specific damage to cellular DNA by small molecules is possible. This paradigm is further corroborated and expanded by findings with a bizelesin analogue, U-78779. U-78779 is highly cytotoxic and sequence-specific, preferring mixed A/T-G/C motifs in naked and intracellular DNA. Like bizelesin, U-78779 also seems capable of region-specific DNA damage. However, the region specificity of U-78779 is not due to, but rather *despite*, its main reactivity preferences at the nucleotide level. This seemingly paradoxical finding can be explained by the genome factor: in contrast to the uniform distribution of U-78779 preferred sites in the human genome, the minor pure AT motifs are highly clustered in the same long AT islands as bizelesin sites.

U-78779 shares several but not all DNA binding properties with bizelesin. U-78779 reactivity with naked SV40 DNA (Figure 2) and cellular DNA (Figure 3) is somewhat lower than that of bizelesin, but the differences are not dramatic. Given this reduced overall DNA reactivity, the similar level of cross-links found in isolated nuclei (Figure 4) implies that this lesion type is at least as prevailing for U-78779 as it is for bizelesin. Interstrand cross-links by U-78779 were also observed in defined model oligonucleotides (*13*). Reflecting a larger molecule, interstrand cross-links formed by U-78779 were generally consistent with drug molecules occupying 7 bp and forming covalent bonds to DNA via the adenine residues at the 3′ ends of their binding sites (1,7 A,A) (*13*), as opposed to 6 bp-spanning bizelesin cross-links (1,6 A,A, Figure 1). For both drugs, occasional non-adenine adducts have been noted (*3*, *6*, *10*, *13*).

Whereas both drugs can be regarded as "AT-specific", they differ profoundly in their binding preferences at the nucleotide level. Experimental determinations in a model system of intracellular SV40 confirmed that U-78779 prefers mixed A/T-G/C motifs (Figure 5). Binding to pure AT sequences, the hallmark of bizelesin adducts, was also noted for U-78779, but such sites comprised only 25% of all binding sites observed in the several hundred base pairs analyzed. These results with intracellularly treated SV40 DNA are in full agreement with the previous observations using model oligonucleotides (*13*), that found a similar proportion of pure AT sites to mixed A/T-G/C sites. Thus, in its DNA reactivity, U-78779 resembles bizelesin, except that it prefers the mixed A/T-G/C motifs over pure AT motifs at the nucleotide level.

At the genomic level, however, the mixed G/C-A/T motifs of U-78779 are scattered more or less uniformly (Table 1 and Figure 6). The unremarkable distribution of these motifs resembles the omnipresent sites for non-sequence-specific

drugs such as cisplatin (*15*). This is in direct contrast to the regional clustering of the pure A/T motifs of U-78779 in the same long AT islands that are targeted by bizelesin. Thus, the *in silico* analysis reveals a distinct potential for region-specificity of the *minor* but not major U-78779 binding motif.

The *in silico* predictions are fully confirmed by the experimental data with model naked DNA sequences and with genomic DNA from drug-treated cells. Mapping of U-78779 adducts in a model AT island sequence demonstrated that, while mixed G/C-A/T sites present at the flanks of that sequence can be adducted, the majority of drug adducts are formed in the core of pure AT sequence (Figure 7). Quantitation of drug-induced lesions in this model AT island and in a model non-AT island, including direct cross-competition experiments, showed a profound preference of U-78779 for the former region (Figure 8). Finally, determinations of region-specific lesions in DNA from drug-treated cells confirmed in genomic DNA that the test AT island was more damaged by U-78779 than the non-AT island locus (Figure 9).

The determinations of drug-induced lesions in these model regions further highlight both similarities and differences between U-78779 and bizelesin. With either naked or cellular DNA, both drugs definitely prefer the model AT island over the non-AT island locus. However, bizelesin, which essentially binds only pure AT motifs that are absent in the model non-AT island employed, only marginally reacted with that DNA locus (*2*). Reactivity of U-78779 with this locus is lower than with the AT island, but clearly detectable, and is consistent with U-78779 binding to mixed A/T-G/C sites present in this non-AT island.

With its dual binding preferences, U-78779 may thus reveal a "Jekyll and Hyde" nature at the level of the human genome. The consequences of its uniformly distributed lesions at mixed A/T-G/C sites are likely to be relatively benign (Dr. Jekyll). The precedence for this is provided by tallimustine, a drug that, like U-78779, prefers mixed A/T-G/C motifs (TTTTGPu) (*15*, *25*, *26*), which are also uniformly distributed (*15*). These sequence-specific but not region-specific lesions are nearly 2 orders of magnitude less lethal than region-specific lesions formed in AT islands by bizelesin (*15*). Analogously, the less preferred pure AT sites that co-locate with the presumably lethal bizelesin lesions in the long AT islands are likely to be the main "killers" (Mr. Hyde) for U-78779. The extrapolation of U-78779 DNA binding data with cytotoxicity in CEM cells suggests that <10 lesions/cell are sufficient for cell growth inhibition, similar to the previous estimates for bizelesin (*15*). Some of these infrequent lethal lesions are likely to consistently form within the limited number ($\sim 3 \times 10^3$/human genome) of distinct AT islands. The rest, perhaps even the majority, of these few adducts associated with cell kill may form in mixed AT-G/C sites. However, these non-region-specific lesions could be in any of the omnipresent mixed AT-G/C sites ($\sim 2.6 \times 10^8$/genome) and would be unlikely to consistently affect any specific kind of genomic loci. The close correlation between the cytotoxic activities of U-78779 and bizelesin (Figure 10) is consistent with the idea that the lethality of U-78779 may also be determined by targeting AT islands.

The high lethality of lesions in AT islands, explicitly demonstrated for bizelesin and strongly implied also for U-78779, suggests that these regions are critical for the growth of cancer cells. The identified AT islands have a potential to function as strong matrix attachment regions (MARs) (*2*), which are essential components of nuclear organization (*27*−*30*). The critical function of MAR domains in replication is consistent with the potent lethality of drug-induced lesions in AT islands. While no data are available on the effects of U-78779 on DNA replication, bizelesin is known to impede replication at the level of the new replicon initiation (*24*, *31*). A single bizelesin adduct seems to inhibit the initiation of several replicons (*24*).

It may be cell-type-dependent whether lesions in a specific AT island are critical. The functional organization of various specific MAR-capable domains (association with the nuclear matrix versus presence in loop DNA) is known to differ between tumor and normal cells (*27*). Furthermore, due to the inherent genetic instability of minisatellites (*32*−*34*), including the known instability of several long AT islands, their abundance is likely to vary from cell type to cell type. Such differences in function and abundance might explain why certain cell lines appear to be hypersensitive to AT island targeting drugs (Figure 10). Our recent studies (Woynarowski et al., unpublished data) showed that bizelesin forms similar levels of bulk and region-specific DNA lesions in cancer CEM and normal WI-38 cells, but the drug is ~80-fold less cytotoxic in the normal cell line. In parallel, we found that certain AT island sequences are significantly more abundant and function as in vivo MARs in CEM cells but not in WI-38 cells. Clearly, further investigations are warranted to explore which AT island subtypes and which tumor types can maximize the benefits of AT island targeting strategies.

The existence of a target with clusters of potential drug binding sites in the genome could be essential for region specificity of small molecule drugs (*1*, *2*, *15*). This prerequisite is not met by the scattered mixed A/T-G/C sites of U-78779 and also by binding motifs of various unrelated sequence-specific drugs. For instance, DSB-120 is a drug that forms cross-links between guanine residues in mixed G/C-A/T motifs with 1−2 central AT base pairs (*35*, *36*) that resemble the inverted (A/T to G/C) main binding preferences of U-78779. The distribution of DSB-120 binding sites, examined in several genes totaling 80 kbp (*37*) as well as in long-range analysis of ~20 Mbp of human sequences (Woynarowski, J. M., unpublished data), was unremarkable, like that of the U-78779 mixed A/T-G/C motifs. Like a number of other drugs that were initially found promising because of their high sequence-specificity at the nucleotide level, DSB-120 showed mediocre antitumor properties (*38*). It is tempting to speculate that the lack of region-specificity at the genomic level might have contributed to the ultimate disappointment with such agents.

Collectively, this study underscores the significance of the genome factor by showing that it can take precedence over drug binding preferences at the nucleotide level. Lesions in mixed A/T-G/C motifs preferred by U-78779 might be relatively inconsequential for drug effects in the cell, if not detrimental by competing with more critical targets for drug molecules. The presence of compatible targets for specific drug binding motifs may profoundly vary among different genomes. For instance, amplification of a specific locus in a cancer cell genome can make it a suitable target for drug

design. Targeting various organisms will also require matching drug binding preferences to the genome of a specific organism. In addition to such fairly obvious aspects, this report demonstrates that the genome can affect drug action in a far less intuitive way. These findings emphasize the need to routinely and rigorously assess the long-range distribution of the drug binding motif in the evaluation of sequence-specific agents.

ACKNOWLEDGMENT

We thank Dr. Barbara Woynarowska for frequent and inspiring discussions and for critical reading of the manuscript.

REFERENCES

1. Woynarowski, J. M., Napier, C., Trevino, A. V., and Arnett, B. (2000) *Biochemistry 39*, 9917−9927.
2. Woynarowski, J. M., Trevino, A. V., Rodriguez, K. A., Hardies, S. C., and Benham, C. J. (2001) *J. Biol. Chem. 276*, 40555−40566.
3. Ding, Z. M., and Hurley, L. H. (1991) *Anticancer Drug Des. 6*, 427−452.
4. Hurley, L. H., Reynolds, V. L., Swenson, D. H., Petzold, G. L., and Scahill, T. A. (1984) *Science 226*, 843−844.
5. Reynolds, V. L., Molineux, I. J., Kaplan, D. J., Swenson, D. H., and Hurley, L. H. (1985) *Biochemistry 24*, 6228−6237.
6. Lee, C. S., and Gibson, N. W. (1993) *Biochemistry 32*, 2592−2600.
7. Weiland, K. L., and Dooley, T. P. (1991) *Biochemistry 30*, 7559−7565.
8. Sun, D., and Hurley, L. H. (1993) *J. Am. Chem. Soc. 115*, 5925−5933.
9. Lee, C. S., Pfeifer, G. P., and Gibson, N. W. (1994) *Biochemistry 33*, 6024−6030.
10. Woynarowski, J. M., Chapman, W. G., Napier, C., and Herzig, M. C. S. (1999) *Biochim. Biophys. Acta 1444*, 201−217.
11. Thompson, A. S., and Hurley, L. H. (1995) *J. Mol. Biol. 252*, 86−101.
12. Lee, C. S., and Gibson, N. W. (1993) *Biochemistry 32*, 9108−9114.
13. Park, H. J., Kelly, R. C., and Hurley, L. H. (1996) *J. Am. Chem. Soc. 118*, 10041−10051.
14. Woynarowski, J. M., Chapman, W. G., Napier, C., Herzig, M. C. S., and Juniewicz, P. (1998) *Mol. Pharmacol. 54*, 770−777.
15. Herzig, M. C. S., Trevino, A. V., Arnett, B., and Woynarowski, J. M. (1999) *Biochemistry 38*, 14045−14055.
16. Woynarowski, J. M., McHugh, M. M., Gawron, L. S., and Beerman, T. A. (1995) *Biochemistry 34*, 13042−13050.
17. Zsido, T. J., Woynarowski, J. M., Baker, R. M., Gawron, L. S., and Beerman, T. A. (1991) *Biochemistry 30*, 3733−3738.
18. Woynarowski, J. M., Faivre, S., Herzig, M. C. S., Arnett, B., Chapman, W. G., Trevino, A. V., Raymond, E., Chaney, S. G., Vaisman, A., Varchenko, M., and Juniewicz, P. E. (2000) *Mol. Pharmacol. 58*, 920−927.
19. Roberts, J. J., and Friedlos, F. (1981) *Biochim. Biophys. Acta 655*, 146−151.
20. Woynarowski, J. M., Napier, C., Koester, S. K., Chen, S.-F., Troyer, D., Chapman, W., and MacDonald, J. R. (1997) *Biochem. Pharmacol. 54*, 1181−1193.
21. Woynarowska, B. A., Woynarowski, J. M., Herzig, M. C. S., Roberts, K., Higdon, A. L., and MacDonald, J. R. (2000) *Biochem. Pharmacol. 59*, 1217−1226.
22. Arnould, R., Dubois, J., Abikhalil, F., Libert, A., Ghanem, G., Atassi, G., Hanocq, M., and Lejeune, F. J. (1990) *Anticancer Res. 10*, 145−154.
23. Woynarowski, J. M., McHugh, M., Gawron L. S., and Beerman T. A. (1995) *Biochemistry 34*, 13042−13050.
24. Woynarowski, J. M., and Beerman, T. A. (1997) *Biochim. Biophys. Acta 1353*, 50−60.
25. Broggini, M., Ponti, M., Ottolenghi, S., D'Incalci, M., Mongelli, N., and Mantovani, R. (1989) *Nucleic Acids Res. 17*, 1051−1059.
26. Broggini, M., Erba, E., Ponti, M., Ballinari, D., Geroni, C., Spreafico, F., and D'Incalci, M. (1991) *Cancer Res. 51*, 199−204.
27. Berezney, R., Dubey, D. D., and Huberman, J. A. (2000) *Chromosoma 108*, 471−484.
28. Nickerson, J. (2001) *J. Cell Sci. 114*, 463−474.
29. DePamphilis, M. L. (2000) *J. Struct. Biol. 129*, 186−197.
30. Davie, J. R., Samuel, S. K., Spencer, V. A., Holth, L. T., Chadee, D. N., Peltier, C. P., Sun, J. M., Chen, H. Y., and Wright, J. A. (1999) *Biochem. Cell. Biol. 77*, 265−275.
31. McHugh, M. M., Kuo, S. R., Walsh-O'Beirne, M. H., Liu, J. S., Melendy, T., and Beerman, T. A. (1999) *Biochemistry 38*, 11508−11515.
32. Richards, R. I. (2001) *Trends Genet. 17*, 339−345.
33. Mangelsdorf, M., Ried, K., Woollatt, E., Dayan, S., Eyre, H., Finnis, M., Hobson, L., Nancarrow, J., Venter, D., Baker, E., and Richards, R. I. (2000) *Cancer Res. 60*, 1683−1689.
34. Handt, O., Sutherland, G. R., and Richards, R. I. (2000) *Mol. Genet. Metab. 70*, 99−105.
35. Mountzouris, J. A., Wang, J. J., Thurston, D., and Hurley, L. H. (1994) *J. Med. Chem. 37*, 3132−3140.
36. Bose, D. S., Thompson, A. S., Ching, J., Hartley, J. A., Berardini, M., Jenkins, T. C., Neidle, S., Hurley, L. H., and Thurston, D. E. (1992) *J. Am. Chem. Soc. 114*, 4939−4941.
37. Neidle, S., Puvvada, M. S., and Thurston, D. E. (1994) *Eur. J. Cancer 4*, 567−568.
38. Walton, M. I., Goddard, P., Kelland, L. R., Thurston, D. E., and Harrap, K. R. (1996) *Cancer Chemother. Pharmacol. 38*, 431−438.
39. Kramer, J. A., Singh, G. B., and Krawetz, S. A. (1996) *Genomics 33*, 305−308.

BI011907S