THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 44, Issue of November 2, pp. 40555–40566, 2001

Printed in U.S.A.

AT-rich Islands in Genomic DNA as a Novel Target for AT-specific DNA-reactive Antitumor Drugs*[S]

Received for publication, April 17, 2001, and in revised form, June 13, 2001
Published, JBC Papers in Press, August 3, 2001, DOI 10.1074/jbc.M103390200

Jan M. Woynarowski‡§, Alex V. Trevino‡, Karl A. Rodriguez‡, Stephen C. Hardies¶, and Craig J. Benham‖

From the ‡Cancer Therapy and Research Center, San Antonio, Texas 78245, the ¶University of Texas Health Science Center, San Antonio, Texas 78284, and ‖UC Davis Genome Center, University of California, Davis, California 95616

Interstrand cross-links at T(A/T)$_4$A sites in cellular DNA are associated with hypercytotoxicity of an anticancer drug, bizelesin. Here we evaluated whether these lethal effects reflect targeting critical genomic regions. An *in silico* analysis of human sequences showed that T(A/T)$_4$A motifs are on average scarce and scattered. However, significantly higher local motif densities were identified in distinct minisatellite regions (200–1000 base pairs of ~85–100% AT), herein referred to as "AT islands." Experimentally detected bizelesin lesions agree with these *in silico* predictions. Actual bizelesin adducts clustered within the model AT island naked DNA, whereas motif-poor sequences were only sparsely adducted. In cancer cells, bizelesin produced high levels of lesions (~4.7–7.1 lesions/kilobase pair/μM drug) in several prominent AT islands, compared with markedly lower lesion levels in several motif-poor loci and in bulk cellular DNA (~0.8–1.3 and ~0.9 lesions/kilobase pair/μM drug, respectively). The identified AT islands exhibit sequence attributes of matrix attachment regions (MARs), domains that organize DNA loops on the nuclear matrix. The computed "MAR potential" and propensity for supercoiling-induced duplex destabilization (both predictive of strong MARs) correlate with the total number of bizelesin binding sites. Hence, MAR-like AT-rich non-coding domains can be regarded as a novel class of critical targets for anticancer drugs.

Cellular DNA is not a uniform target for DNA-reactive drugs. At the nucleotide level, drugs recognize and bind short motif(s) of a few base pairs. The location of drug adducts at the genomic level, however, depends on how these short motifs are distributed in larger domains (hundreds of base pairs) that may have distinct structural and functional properties. This aspect, referred to as region specificity, may be critical for the biological outcome of drug action (1, 2).

The classical antitumor DNA-reactive drugs currently in the clinic display low sequence specificity, binding virtually indiscriminately to cellular DNA. Not only are such drugs non-region-specific (3, 4), but they also form most of their adducts with macromolecules other than DNA (5). In contrast, certain minor groove-binding agents (MGBs)[1] combine a high sequence specificity for AT motifs with a lack of reactivity with non-DNA targets. Bizelesin and adozelesin of the cyclopropylpyrroleindole family are two novel anticancer drugs with these properties (4, 6–10). These sequence-specific cyclopropylpyrroleindole drugs, as well as several other AT-specific MGBs, are currently in phase I clinical trials (11–17).

Despite the growing interest in novel sequence-specific small molecules, their potential for producing region-specific DNA damage remains largely unexplored. Our recent studies of AT-specific MGBs demonstrated that some, but not all, AT-specific MGBs are able to preferentially damage specific regions of genomic DNA (4, 8–10). Among the small molecules tested, the cyclopropylpyrroleindole drug bizelesin was found to have the greatest potential for producing region-specific damage. In model sequences analyzed, bizelesin adducts were non-randomly distributed and reflected mainly interstrand cross-links at T(A/T)$_4$A sites, although some monoadducts at A(A/T)$_4$A sites are also observed (Fig. 1) (8–10).

Bizelesin is one of the most cytotoxic compounds ever identified, requiring only a few ($<10^1$) drug adducts per cell for cell growth inhibition (4, 10). Since typical DNA-reactive drugs must form several thousand lesions/cell to achieve similar effects (4), one can infer that bizelesin must damage DNA specifically within regions that are crucial for continued cell growth. Based on various observations, including a potent and selective inhibition of DNA replication, we previously hypothesized (10) that the basis of bizelesin's exceptional potency may be its targeting of AT-rich matrix-associated regions (MARs), domains of critical importance for replication.

To investigate the possibility that bizelesin could kill cancer cells by targeting such specific, potentially critical regions, we have developed a new approach that combines bioinformatics and pharmacogenomics with molecular pharmacology. The results demonstrate that bizelesin preferentially targets AT-rich minisatellite loci, in which bizelesin binding sites are highly

* This work was supported in part by Grant CA71969 from the NCI, National Institutes of Health and by a grant from Children's Cancer Center, University of Texas Health Science Center, San Antonio, TX (to J. M. W.) and by Grant DBI 99-04549 from the National Science Foundation and Grant RO1-HG01973 from the National Institutes of Health (to C. J. B.). A preliminary account of this work was presented in part at the 10th NCI-EORTC Symposium on New Drugs in Cancer Therapy, June 16–19, 1998, Amsterdam (Abstr. 528). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains an extended list of ~40 prominent regions identified in this study.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF385609.

§ To whom correspondence should be addressed: Cancer Therapy and Research Center, Inst. for Drug Development, 14960 Omicron Dr., San Antonio, TX 78245-3217. Tel.: 210-677-3832; Fax: 210-677-0058; E-mail: jmw1@saci.org.

[1] The abbreviations used are: MGB, minor groove binding drug; MAR, matrix-associated region; SIDD, stress-induced duplex destabilization; QPCR, quantitative polymerase chain reaction; PCR, polymerase chain reaction; bp, base pair(s); kbp, kilobase pair(s).

clustered. The identified domains have a variety of properties known to be associated with MAR function.

EXPERIMENTAL PROCEDURES

In Silico Motif Distribution Analysis—The long range distribution of drug binding motifs was analyzed using the custom Msearch program (developed by S. C. Hardies) written in Fortran to run under MS DOS, which enables batch processing of multiple GenBank™ entries to tabulate the positions of the exact matches to specified binding motifs in each sequence. The output files were further processed using custom scripts for Excel (Microsoft, Redmond, WA) to generate distribution histograms showing the number of hits in each 250-bp region along the sequence, and to catalog the analyzed sequences and the results of analysis.

The majority of the analyzed sequences were selected at random. To ensure that both coding and non-coding regions were adequately represented, contiguous sequences of 30–300 kbp were favored. Entries shorter than 10 kbp were analyzed only when warranted by specific information, such as the reported presence of an AT-rich region. Depending on the motif examined, the total length of human DNA sequences evaluated ranged from 3.1×10^7 bp to 4.3×10^7 bp as indicated in Table I. The distribution parameters (Table I) remained virtually unchanged after the number of analyzed hits exceeded $\sim 10^5$. For the bizelesin cross-linking motif $T(A/T)_4A$, this number of hits was reached after analyzing $\sim 10^7$ bp of DNA sequences.

For some sequences, the distribution of drug binding sites was compared with the distribution of known AT elements such as TATA boxes and polyadenylation signals. In such cases, the sites of potential TATA boxes and polyadenylation signals were determined using an Internet tool (www.itba.mi.cnr.it/webgene/).

Sites of Drug Adducts in Naked Model AT Island DNA—Sites of drug adducts in naked DNA were determined as described previously using PCR-generated, uniquely end-labeled DNA (9), except that DNA from untreated CEM cells was used as template (at 6000 cell eq/reaction) and primers designed for the Z79699 AT island (4). PCR conditions were as described (9), except that 28 cycles were used and one of the primers was 5′-end-labeled with [γ-^{32}P]dATP (9). Following the removal of unincorporated primers using Spin Column 200 (Sigma), the labeled DNA ($\sim$0.1 μg/50 μl in 10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl) was incubated for 4 h with the indicated bizelesin concentrations. After ethanol precipitation to remove unreacted drug, DNA samples were heated for 15 min at 95 °C to convert adducts to breaks. Drug-treated DNA samples were analyzed by sequencing polyacrylamide gel electrophoresis followed by phosphorimaging on a Storm system (Molecular Dynamics, San Jose, CA) . To determine precise adduct positions in the sequence, sequencing reactions using the same primers as the primers for the generation of 5′-end-labeled PCR products were run in parallel on the same gels (9).

The PCR system for the Z79699 AT island used in these experiments generates two related products. The main product ($\sim$85%) is an 859-bp band. A minor product is a 1025-bp band, consistent with the sequence of Z79699 in GenBank™. Cloning and sequencing of the 859-bp product (GenBank™ accession no. AF385609) confirmed that it differs from the 1025-bp Z79699 GenBank™ sequence in the number of repeats in the central segment, with the complete overlaps of extensive flanking areas from both 5′ and 3′ ends (see Fig. 3). Thus, drug adducts in either product would generate identical end-labeled subfragments in those areas of the gel where precise positions of drug adducts can be assessed.

Levels of Drug Adducts in Naked Model DNA by Agarose Electrophoresis—Agarose electrophoresis was used to quantitate DNA adducts (following thermal conversion to strand breaks) as described elsewhere (9). Labeled DNA for these experiments was generated by PCR and included (i) model AT island DNA made as described above for the sites of drug adducts and (ii) a model non-AT island DNA (a 536-bp β-globin fragment generated using the previously described primer system and PCR conditions (3)). These DNAs were either ^{32}P-end-labeled exactly as described for the sites of drug adducts or uniformly labeled with [^{32}P]dGTP during the PCR reaction. Following drug treatment and processing as for the sites of drug adducts, samples were analyzed by agarose electrophoresis. After phosphorimaging, the disappearance of full-length materials was quantitated using ImageQuant software (Molecular Dynamics). Based on these data, adduct frequencies were estimated as described (9). The results from both labeling protocols were pooled for lesion quantitation. In some experiments, unlabeled competitor DNA was added as indicated in Fig. 4*C*.

Drug Adducts in Specific Regions by QPCR Stop Assay—Human leukemia CEM cells were cultured as described elsewhere (4). Human colon carcinoma COLO320DM cells (purchased from the ATTC, Manassas, VA) were maintained in RPMI 1640 medium supplemented with 10% fetal calf serum (Life Technologies, Inc.). Prior to drug treatment, cells were prelabeled with [^{14}C]thymidine and the incorporated radioactivity was used to determine cpm/cell ratio (3, 4, 10). Following drug treatment for 4 h, cellular DNA was purified using either PureGene or Quiagen kits (3, 4, 10).

The details of the QPCR assay for region-specific DNA damage have been given elsewhere (3, 4, 10). The assay was run under conditions (cycle number, template amounts) that ensure the linearity of the signal as a function of the amount of undamaged template. The amounts of template DNA used in these experiments are expressed as cell eq, based on the ^{14}C cpm/cell ratio. Specific PCR primers for various regions and cycling conditions used were as described (4, 10), except for the AT island in Z80771. The Z80771 system used 5′-TTCCATTTTATAGTAGAACATGCGTAGA as the upper primer and 5′-AAATGCTGTTGGTATTGTGTTGATAC as the lower primer to generate a product consisting of 928 bp. PCR cycling conditions for the Z80771 system were identical to those described previously for the Z79699/AF385609 system (4), except that typically 500 and 1000 cell eq of template DNA and 22 cycles were used.

Following agarose electrophoresis, autoradiography, and/or phosphorimaging, the signals of the amplified PCR products were quantitated, normalized to signals for DNA from untreated control cells, and converted to lesion frequencies as described previously (3, 4, 10). For the Z79699 system, which generates two products (see above), the quantitation was based on the data for the shorter, 859-bp variant (main product). The results reported here are from two to four independent experiments, each carried out in triplicate, typically at two different amounts of DNA template.

The frequency of lesions was estimated based on the Poisson distribution and normalized per unit length of DNA (3, 4, 10). To compare drug effects on diverse regions, lesion frequencies were further normalized by drug concentration to yield lesions/kbp/μM (Fig. 6). Only those drug concentrations producing between 15 and 85% inhibition were included in the second normalization.

Repeat Consensus Analysis—The consensus repeats for the AT islands with the clusters of bizelesin binding sites were determined using a Tandem Repeats Finder version 2.02 program (18) made available by its author, Dr. Benson.

Duplex Stability—The thermodynamic stability of the DNA duplex is calculated at each nucleotide position in a given sequence as the melting temperature of the 25-bp oligonucleotide contained within a window of that length that begins at a given position. This calculation is based on a nearest-neighbor algorithm (19), as implemented in the Oligo program (Molecular Biology Insights, Inc., Cascade, CO), and assumes 120 mM NaCl and 10 mM $MgCl_2$ concentrations.

Stress-induced Duplex Destabilization (SIDD) Analysis—Computational techniques to assess the destabilization properties of a DNA sequence that is subjected to a superhelical stress have been described elsewhere (20, 21). For any user-specified level of superhelicity, the algorithms calculate: (i) the equilibrium probability of denaturation of each base pair along the DNA sequence, and (ii) the incremental free energy $G(x)$ needed to force the base pair at position x to always be separated (22, 23). The results of such calculations agree precisely with the experimental determinations of the sites that denature and the magnitude of transition at each site at varying levels of superhelicity (20, 24). Calculations in this study assumed superhelical densities varying from -0.04 up to -0.065, which correspond to moderately low physiological values. Most of the results reported here are for a superhelical density of -0.04 to center on regions that are first to become destabilized. In one case of less destabilized sequence, the calculation used a density of -0.045.

To compare various destabilized regions, the integrated SIDD potential was calculated for entire regions within sequence positions stated in Table III by integrating the area under the $G(x)$ *versus* x curve. The integration positions were defined by the cut-off level of $G(x) < 7$. Based on the previous studies (24, 25), distinct MAR domains are expected to exhibit integrated SIDD potential of at least 1500.

DNA Flexibility—The flexibility of DNA in the identified AT islands along with the adjacent sequences was analyzed using the method of Sarai *et al.* (26), as implemented in the FlexStab program (available at leonardo.ls.huji.ac.il/departments/genesite/faculty/bkerem.htm). This program calculates the flexibility of a given DNA sequence by considering a series of overlapping windows. Within each window, the flexibilities for each dinucleotide step are analyzed to give a window-averaged flexibility. The full GenBank™ entries that contain the

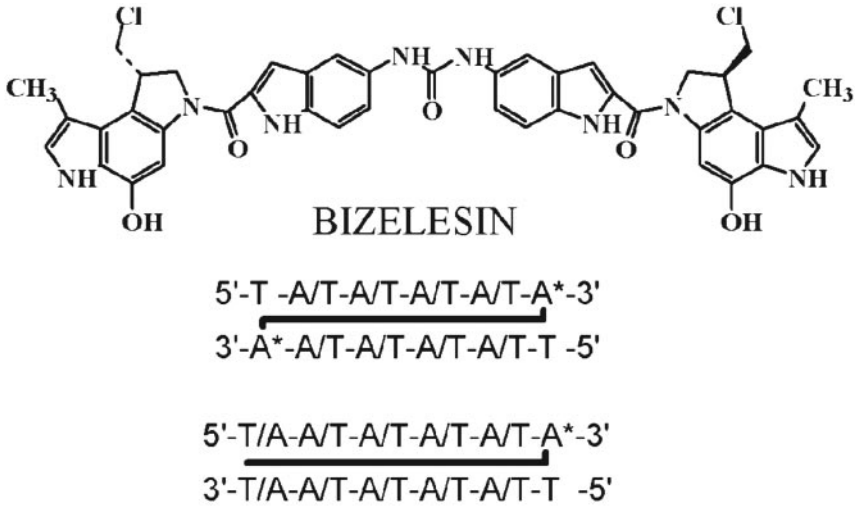

FIG. 1. **Structure of bizelesin, and its preferred DNA binding motifs for interstrand cross-links (*top*) and monoadducts (*bottom*).** *Asterisks* in these motifs indicate adducted adenine residues.

indicated AT islands were analyzed using a window length of 250 bp and a shift of 25 bp.

MAR Potential—To assess the potential of bizelesin-targeted AT islands to function as MARs, an Internet tool (MAR Finder; www.ncgr.org/MarFinder/) developed by Krawetz and co-workers (27, 28) was used. MAR Finder assesses several known properties of MAR sites and then displays their cumulative statistically weighted distribution. Unless noted otherwise, the analysis was run at the standard, default settings (28). MAR potential *versus* sequence position was plotted after the highest peak in the analyzed sequence is normalized to 1 (*i.e.* Fig. 7*E*). MAR Finder tool reports also the peak value used for normalization and the integrated (normalized) MAR potential. To compare MAR potentials of AT islands in various sequences (Table III), the integrated MAR potentials were un-normalized by multiplying these values by the peak value. These data were obtained by analyzing 25-kbp segments of the indicated sequences containing a specific AT island. If the entire sequence was shorter than 25 kbp, then the complete sequence was analyzed.

RESULTS

The study reported here consists of three components. First, regions that could potentially be targeted by bizelesin were identified at the genome level by a long range *in silico* analysis. Next, the *in silico* predictions were verified by experimentally determining drug-induced lesions, both in naked DNA and in drug-treated cells. Finally, the identified targeted regions were evaluated for attributes that could provide insight into their possible functional roles in the genome.

Long Range Distribution of Bizelesin Binding Motifs: In Silico Analysis

To identify domains likely to be targeted by bizelesin at the genomic level, we examined the distribution of drug binding motifs within a variety of sequences that total between 1 and 1.5% of the human genome. Specifically, we searched for matches to the previously established bizelesin drug binding motifs: (i) interstrand cross-linking sites $T(A/T)_4A$ (the main type of bizelesin adducts, Fig. 1) and (ii) monoadduct sites $A(A/T)_4A$ (minor adducts).

The "density" of $T(A/T)_4A$, bizelesin's cross-linking motif, was low in the bulk of analyzed sequences, with an average of 2.8 hits/0.25 kbp (Table I). A typical distribution histogram (Fig. 2*A*, *upper panel*) displays the low number of $T(A/T)_4A$ potential drug binding sites that are scattered more or less uniformly throughout the 100-kbp length of the sequence analyzed in this example. Bizelesin's minor motif $A(A/T)_4A$ is somewhat more common than the cross-linking motif (Table I), but follows a similar pattern of approximately uniform distribution in typical genomic regions (data not shown). The observed average frequencies of both bizelesin motifs are similar to the frequencies predicted in a random sequence.

The overall distribution of bizelesin motifs, however, is nonrandom. A number of domains contain large and dense clusters of potential bizelesin binding sites, which strongly contrast to the situation in typical stretches of DNA. Motif distribution profiles for several of these hot spot loci are given in Fig. 2, *A* (*lower panels*) and *B*. Hot spots for the cross-linking motif exceeded the average motif density by 14–35-fold (with peak bizelesin motif density of 40–99 hits/0.25 kbp), far beyond the levels achievable as outliers of a random distribution. The clusters of the minor, monoadduct motif $A(A/T)_4A$ tend to occur at the same loci as the cross-linking motif hot spots, although certain loci (*i.e.* the hot spot in Z82900) are markedly richer in the $A(A/T)_4A$ motif (Fig. 2*B*) than in the $T(A/T)_4A$ motif (Fig. 2*A*, *lower panel*). Together, these results indicate that the human genome contains extended regions of highly atypical sequence composition that are likely to be preferentially targeted by bizelesin.

Bizelesin Targets Islands of AT-rich DNA—Table II lists several representative examples of the identified bizelesin hot spot domains, arranged in descending order by the total number of perfect matches to the bizelesin cross-linking motif. An extended list of ~40 prominent regions identified in this study is provided in the supplement (available in the on-line version of this article).

The regions with high densities of bizelesin motifs are highly AT-rich (between 80 and 100% A/T) and span between 200 and 1000 bp (Table II). The total number of bizelesin sites per locus varies with the length of the island involved. More than 200 occurrences of the cross-linking motif were found in one AT island (in Z79699), and more than 100 hits were present in several other regions (Table II and on-line supplement). Prominent bizelesin-targeted AT islands are typically isolated from one another by many kbp, with the average occurrence of ~1 AT island/megabase pair. Two prominent AT islands in the same sequence were found only twice (in rather long GenBank™ sequences of 215 and 78 kbp; see on-line supplement).

The available information on the identified AT islands is too limited to classify all these domains in the context of known AT-rich elements. However, the AT islands in GenBank™ sequences X04682 and X00364 correspond precisely to the known AT-rich MAR domains in the *apoB* and c-*myc* genes, respectively (29, 30). The potential bizelesin targets also include AT islands in the AT-rich fragile sites (31), FRA16D (in Gen-

TABLE I

Long range in silico sequence analysis for the distribution of possible bizelesin binding sites

Human DNA sequences covering 43.0 and 31.4 × 10^6 bp were analyzed for T(A/T)$_4$A and A(A/T)$_4$A motifs, respectively. 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.* Fig. 2).

Drug	Motif searched	Predicted[a]	Average	"Hottest" loci[b]	No. of hottest loci	Ratio hottest loci to average	Expected region specificity
		hits/0.25 kbp	*hits/0.25 kbp*	*hits/0.25 kbp*	*loci/Mbp*		
Bizelesin	T(A/T)$_4$A (cross-links)	2.92	2.8 ± 3.6	40–99	1.0	14–34	Yes
	A(A/T)$_4$A (mono-adducts)	5.83	8.6 ± 2.6	60–129	0.6	10–19	Yes

[a] Predicted frequency of specific motifs was calculated based on probability of random occurrence of such motifs (27) assuming the overall DNA composition of 60% AT and 40% GC.

[b] 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 × S.D. (rounded).

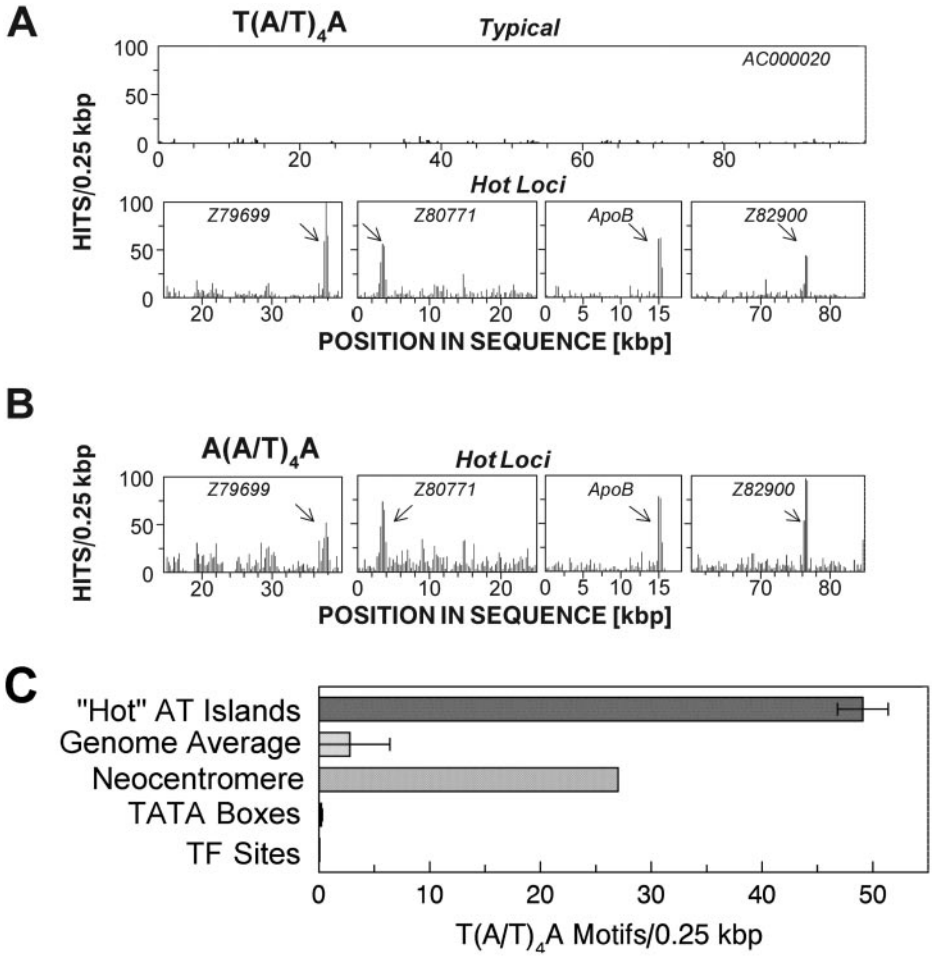

FIG. 2. **The *in silico* analysis for long range distribution of potential bizelesin binding sites.** *A* and *B*, examples of distribution histograms for cross-linking motif, T(A/T)$_4$A and monoadduct motifs A(A/T)$_4$A, respectively. The *histograms* depict the number of occurrences of drug binding motifs in bins of 250 bp along the indicated sequences. GenBank™ entry AC000020 in *A* (*top*) represents a typical locus whereas sequences Z79699, Z80771, ApoB, and Z82900 in *B* are loci that contain hot spots with clusters of respective drug binding motifs (*arrows*). Sequence positions for the ApoB gene are given using the numbering of Levy-Wilson and Fortier (30). The AT island in the ApoB gene (at positions 14817..15387) is contained within GenBank™ X04682, positions 189..759. *C*, densities of the cross-linking motif T(A/T)$_4$A/0.25 kbp (± S.E. where applicable) in various types of AT-rich domains/elements: hot spot AT islands (defined as indicated in Table I; for comprehensive list see the on-line supplement), all the sequences analyzed, a neocentromeric sequence (GenBank™ AF042484), TATA boxes (TATAAA consensus), and transcription factor (*TF*) binding sites. Although binding sites for nine vertebrate transcription factors were found to be compatible with the T(A/T)$_4$A motif, 99% of the identified occurrences corresponded to the NNNNNTAATTAN motif of the homeobox protein Prx2 (S8) (32).

Bank™ AF217490, with a total of 60 bizelesin cross-linking sites, Table II) and FRA16B (see on-line supplement).

Various other AT-rich elements are much less significant or insignificant as potential targets for bizelesin (Fig. 2*C*). For instance, the hottest centromeric sequence identified features only modest local densities and total number of hits for bizelesin motifs. In addition, although potential TATA boxes conform to bizelesin binding motifs), such sites are infrequent and scattered. Moreover, many sites corresponding to TATA box consensus sequence are found within AT islands. Polyadenylation signals are similarly infrequent and essentially incompatible with the cross-linking motif, although they may harbor the monoadduct motif (data not shown). Thus, TATA boxes and polyadenylation signals outside of AT islands are unlikely to be among the preferred targets. Likewise, bizelesin cross-linking motifs are very rarely found in binding sites for transcription factors (Fig. 2*C*). The secondary monoadduct motif is harbored in a larger variety of elements, but these sites still represent only a minuscule fraction of all the monoadduct sites. Hence, the isolated occurrences of either cross-linking or monoadduct motifs in transcription-related elements are highly unlikely to be effective in competing with the clusters of these sites, such as found in the prominent AT islands.

Determination of Drug-induced Lesion Sites in Naked DNA and in Intact Cells

Preferential Bizelesin Damage to a Model AT Island in Naked DNA—Further studies experimentally determined bizelesin binding to the model AT island sequence in naked DNA. This model AT island corresponded to a previously uncharacterized shorter variant of the AT island in Z79699 sequence that prevails in CEM cells (Table II). Both variants have extensive sequence overlap on both 5′ and 3′ ends and differ mainly in the numbers of tandem repeats in their pure AT sections (Fig. 3*A*). The PCR-generated material used in these

TABLE II

AT islands with clusters of T(A/T)$_4$A bizelesin binding motif, identified in the long range in silico analysis of ~43 Mbp of human sequences

In comparison to some of the hottest loci (items 1–8 selected from ~40 loci identified), c-Myc MAR is listed as an example of regions with a modest number of hits.

GenBank™ entry[a] (sequence positions)	T(A/T)$_4$A motif		AT island length[c]	AT content[c]	Location/function
	Peak density[b]	Hits in the entire AT island[c]			
	hits/250 bp		*bp*	*%*	
Z79699 (37120..38119)	99	229	1000	85.7	Chromosome X
Shorter variant (AF385609)[d]	65	162	834	82.7	*In vitro* MAR
Z80771 (3047..3882)	56	163	836	90.7	Chromosome X
X04682[e] (189..759)	73	150	571	95.7	MAR at the 3′ end of *apoB* gene (30), known hypervariable minisatellite (37,53)
AC005195 (63229..63874)	53	136	646	87.3	Chromosome 5p1
Z72519 (33675..34217)	65	114	543	86.4	Chromosome X
Z82900 (76152..76715)	44	97	564	96.1	Chromosome X
AF217490 (147241..147431)	60	61	191	95.3	Fragile site FRA16D (chromosome 16q23)(34)
X00364 (7378..7661)	11	21	284	80.3	c-*myc* proto-oncogene MAR (29)

[a] For extended list of sites, see on-line supplement. The positions listed correspond to the peaks of the "density" of bizelesin binding motif.
[b] Peak of local density of exact matches to bizelesin cross-linking motif (per bins of 0.25 kbp) in the motif distribution histogram (see Fig. 2 for details).
[c] In the position range specified for each domain.
[d] AF385609 is a shorter variant of the AT island in Z79699 identified in DNA from CEM cells (see Fig. 3 and "Experimental Procedures" for details).
[e] Sequence positions for X04682 correspond to positions 14817..15387 in the entire ApoB locus using the numbering of Levy-Wilson and Fortier (30).

experiments consisted mainly of the shorter variant 859-bp product. Following drug treatment, the adduct sites within this model AT island DNA were thermally converted to strand breaks. Sequencing gels used to characterize the distribution of actual drug binding sites on both strands are shown in Fig. 3*B*, along with the plot of the predicted cross-linking and monoadduct sites.

These results show excellent agreement between the predicted and actual locations of drug adducts. The majority of sites whose exact positions could be identified were found to collocate with cross-linking motifs, although some sites consistent with monoadducts were also found. Very strong, multiple bands are seen for both top and bottom strands in the areas corresponding to the pure A/T core. The same area contains high number of predicted sites (both cross-linking and monoadduct motifs). By contrast, infrequent weak sites were found in the flanking sequences, as was predicted *in silico*. The pattern of multiple strong sites within the pure A/T core unequivocally demonstrates the preferential recognition of this domain.

Drug-treated naked DNA was also analyzed by non-denaturing agarose gels. Under these conditions, drug-induced interstrand cross-links and/or closely spaced monoadducts give rise to double strand breaks in DNA. Fig. 4*A* demonstrates a drug concentration-dependent disappearance of the 859-bp band (the initial length of AT island DNA). The sizes of generated sub-bands are consistent with the claim that drug adducts are formed mainly within the 100% A/T core of the AT island sequence.

The disappearance of full-length DNA was next used to assess the frequency of bizelesin lesions in the AT island DNA and in a model non-AT island sequence (a fragment of the β-globin gene) that is poor in bizelesin motifs (Fig. 4*B*). In the AT island DNA, the first drug adducts are detected at 0.005 μM bizelesin, with a frequency of ~0.8 adducts/kbp. At 0.1 μM drug, there is almost no full-length DNA remaining and the adduct density approaches 2.5 adducts/kbp. Drug effects in non-AT island DNA are much less dramatic. Only a marginal level of lesions (~0.5–0.8 adducts/kbp) was observed in the β-globin sequence at drug levels as high as 0.5–1 μM. Normalization by the number of drug lesions per unit of drug concentration reveals that bizelesin is 2 orders of magnitude more reactive toward AT island than toward non-AT island naked DNA (95.2 ± 14.2 *versus* 1.1 ± 0.2 lesions/kbp/μM drug).

The high preference of bizelesin for AT island over non-AT island DNA is further documented by competition experiments in which an excess of either sequence was used as unlabeled DNA (Fig. 4*C*). A 5-fold excess of unlabeled competitor AT island DNA resulted in markedly higher level of the full-length labeled fragment remaining (38% *versus* ~5% for no competitor). These values correspond to an estimated 5.4-fold reduction in drug adducts due to the competition, which is close to the theoretically expected 5-fold reduction. In contrast, even a 20-fold excess of unlabeled β-globin fragment only marginally diminished drug binding to labeled AT island.

Preferential Bizelesin Damage to AT Islands in Drug-treated Human Cancer Cells—Further experiments assessed whether bizelesin preferentially targets AT islands in drug-treated cancer cells. QPCR stop assay, used in these studies, follows drug-induced lesions in specific regions based on the progressive elimination of the amplified product (Fig. 5*A*). The reduced PCR signal allows one to estimate the frequency of lesions in the region defined by the PCR primers (Fig. 5, *B* and *C*). Data for the AT islands within the Z79699 and Z80771 sequences illustrate a definite reduction in PCR signal at drug concentrations between 0.05 and 0.1 μM, with a nearly complete elimi-

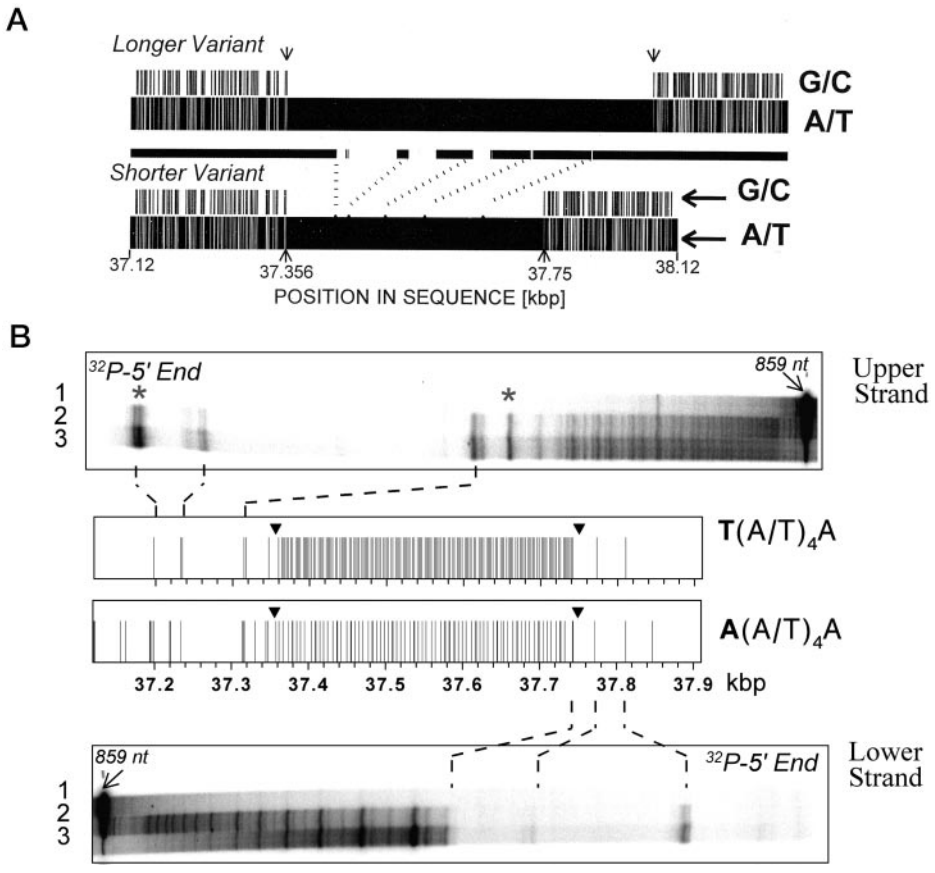

FIG. 3. **Model AT island in Z79699 GenBank™ sequence and its shorter variant (GenBank™ AF385609™) identified in CEM cells (*A*) and predicted and actual sites of bizelesin adducts in the naked DNA of model AT island (*B*).** *A*, DNA sequences for both AT island variants in *A* are plotted as the distribution of G/C and A/T base pairs. The alignment of these variant sequences and the gaps in the shorter variant are depicted by the *bar* in the *middle*. Positions in sequence, including the positions of the last G/C base pairs flanking the pure AT core (*arrows*), are indicated using the numbering as in GenBank™ Z79699. *B*, the distribution of the predicted $T(A/T)_4A$ and $A(A/T)_4A$ bizelesin binding sites (cross-links and monoadducts, respectively) in the shorter (AF385609) AT island is shown 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 (*cf. A*). The actual drug sites were determined in the AT island DNA (singly ^{32}P-end-labeled at either 5′ end) following the treatment for 4 h at 37 °C with bizelesin at 0 μM (*lanes 1*), 0.005 μM (*lanes 2*), and 0.02 μM (*lanes 3*), and thermal conversion of drug adducts to strand breaks. The autoradiographic images of sequencing polyacrylamide gel electrophoresis (examples for four independent experiments) are reoriented to align the labeled 5′ ends of the fragments in regard to DNA sequence. The bands corresponding to full-length 859-nucleotide fragments reflect sequence positions of 37101 and 37959 for the lower and upper strand labeled, respectively. *Asterisks* indicate the positions of some bands that appear to correspond to monoadducts at $A(A/T)_4A$ sites.

nation of the signal at 5 μM bizelesin. At 0.1 μM drug, the PCR signals correspond to estimated 0.89 and 0.78 lesions/kbp in the Z79699 and Z80771 AT islands, respectively.

The experimentally determined frequencies of bizelesin lesions (normalized by drug concentration) are compared for several AT islands, non-AT island regions, and bulk DNA (Fig. 6*A*). Also shown are the binding motif densities predicted by *in silico* analysis (Fig. 6*B*). As noted previously (10), lesions in the Apo B MAR (AT island in X04682) exceed by severalfold the average damage seen in bulk DNA (7.06 ± 1.35 *versus* 0.87 ± 0.04 lesions/kbp/μM, respectively). Regions identified only through the *in silico* analysis as potentially vulnerable (the AT islands in Z79699, Z80771) proved to be among the experimentally most sensitive targets with 5.54 ± 0.97 and 4.70 ± 0.72 actual lesions/kbp/μM, respectively (Fig. 6*A*). Lower levels of bizelesin lesions in the c-Myc MAR, AT island in X00364, (2.13 ± 0.52 lesions/kbp/μM) are consistent with its shorter length and a relatively modest motif count in the *in silico* analysis. Several loci of low average densities of *in silico* predicted potential binding sites showed low levels of actual bizelesin lesions similar to average damage in bulk DNA (*e.g.* <0.4 lesions/kbp/μM in the β-globin locus). Overall, close agreement is seen between the actual lesion frequencies and the predictions based on motif distribution (r^2 = 0.86, Fig. 6*B*, *inset*). These findings demonstrate that treatment of CEM cells with bizelesin produces markedly more lesions in AT island regions than in either non-AT island loci or bulk DNA.

Additional confirmation of the regional specificity of bizelesin lesions was obtained with human colon carcinoma COLO320DM cells (QPCR concentration profiles not shown). In these cells, bizelesin induced 3.77 ± 0.38 lesions/kbp/μM drug in the Z79699/AF385609 AT island, compared with 0.97 ± 0.44 lesions/kbp/μM in bulk DNA (values normalized by drug concentration). Only 1.46 ± 0.07 lesions/kbp/μM were found in the motif-poor c-Myc ORI domain. It needs to be emphasized that COLO320 DM cells have massive genetic rearrangements, which include ~30–50-fold amplification of c-Myc sequences (33). The regional density of bizelesin adducts in c-Myc MAR domains is similar in COLO320DM and CEM cells (2.73 ± 0.12 and 2.13 ± 0.52 lesions/kbp/μM, respectively). These sequences in COLO320 DM cells, however, are likely to bind more drug, proportionately to their greater abundance. Such increased binding to amplified c-Myc loci may explain the reduced drug binding to loci of higher local motif density (such as that within Z79699), compared with cells with no c-Myc amplification.

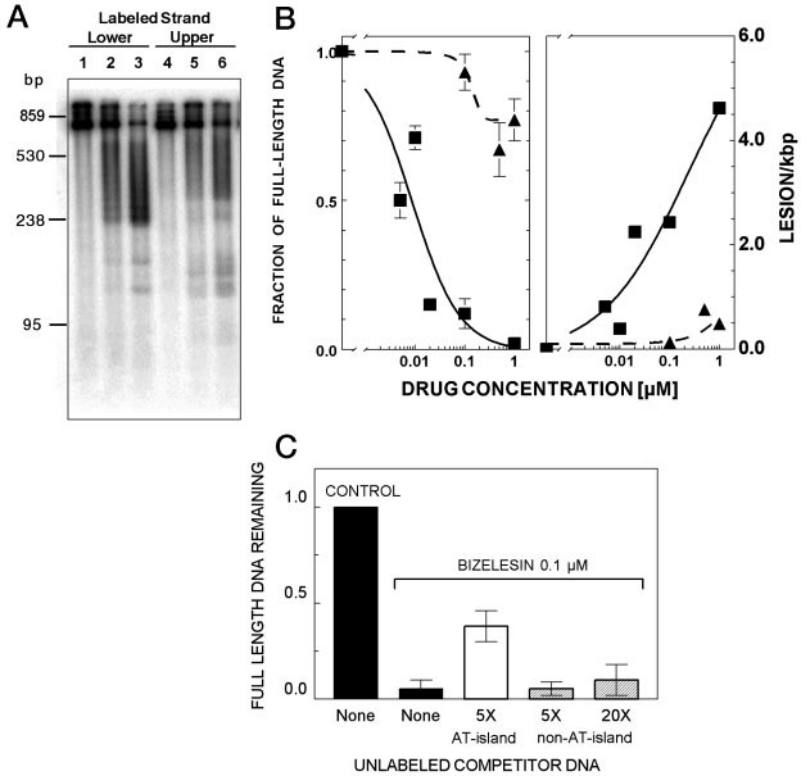

FIG. 4. **Preferential formation of bizelesin adducts in naked DNA of a model AT island *versus* a non-AT island β-globin sequence.** Singly ^{32}P-end-labeled DNA was generated and drug treated as described in the legend to Fig. 3. Drug adducts were thermally converted to strand breaks, and the resulting fragments were analyzed in non-denaturing agarose gels. *A*, representative agarose gel of Z79699 DNA treated with bizelesin at 0 μM (*lanes 1* and *4*), 0.005 μM (*lanes 2* and *5*), and 0.02 μM (*lanes 3* and *6*). *B*, quantitation of drug-induced disappearance of full-length DNA (*left*, mean ± S.E. from two to six determinations) and the corresponding frequency of double-strand breaks/kbp (*right*) in Z79699 AT island DNA (■) and a non-AT-island β-globin sequence (▲). *C*, competition of 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) after treatment with the drug. Drug treatment was carried out as in *A* and *B* except that the indicated excess amounts of unlabeled competitor DNA (either AT island or non-AT island: Z79699 fragment and a β-globin fragment, respectively) were present.

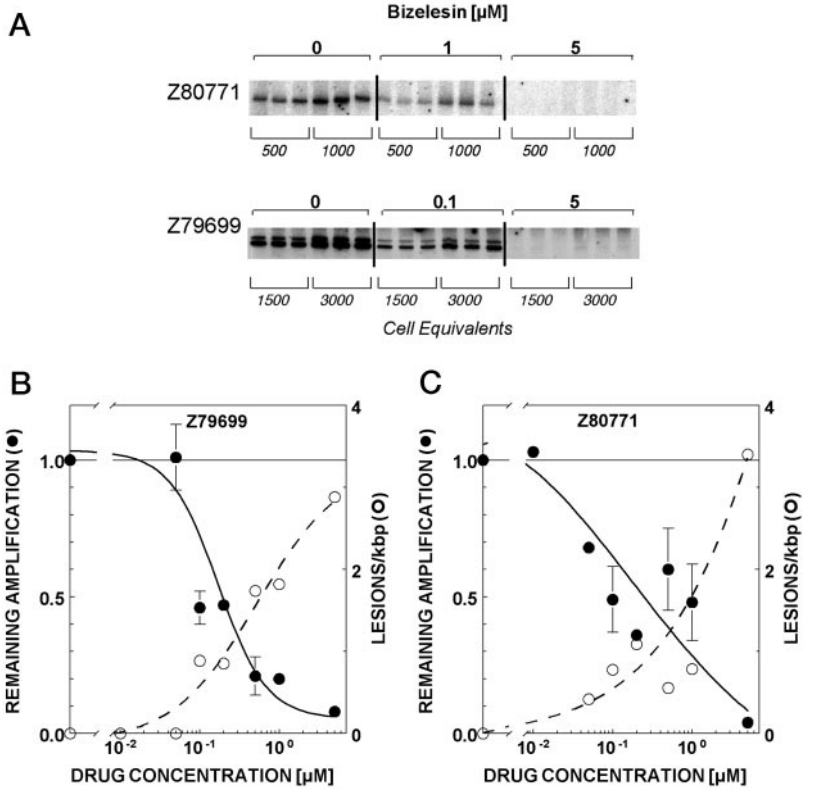

FIG. 5. **Region-specific damage to AT islands in drug-treated human cancer cells.** CEM cells were incubated with bizelesin for 4 h followed by DNA isolation and determination of region-specific damage by QPCR stop assay. Representative examples of agarose gel electrophoresis for QPCR stop assay products are shown for AT islands in GenBank™ sequences Z79699/AF385609 and Z80771 (*top panels*). Different amounts of template DNA are indicated as cell eq. *Bottom panels* illustrate the decrease in the amplified signal as a function of drug concentrations and the corresponding frequency of drug-induced lesions. Data shown represent means (± S.E.) from two to six independent experiments carried out in triplicate using two levels of template DNA. For the PCR system for Z79699/AF385609, the quantitation is based on the shorter variant (AF385609) prevailing in CEM cells.

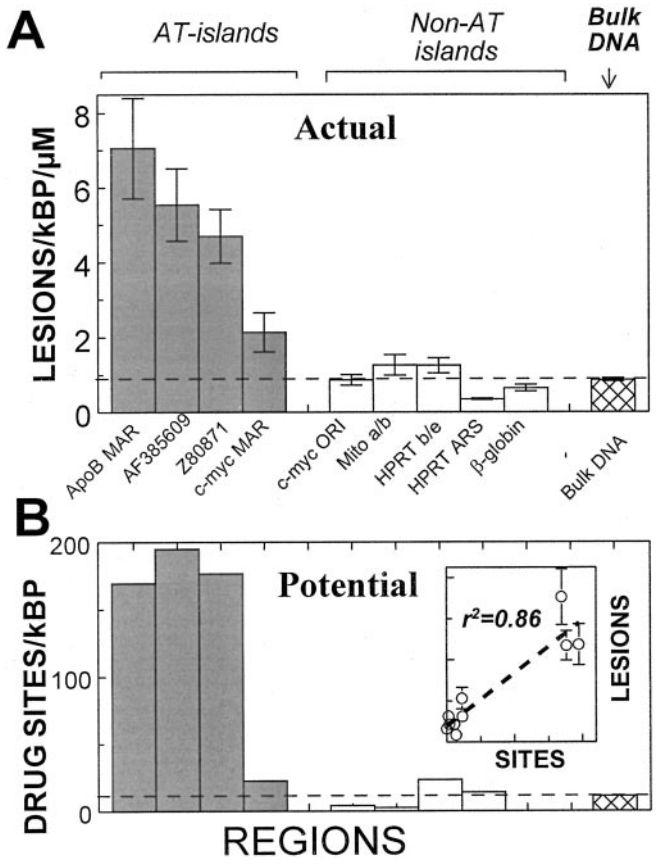

FIG. 6. **Preferential targeting of AT islands detected in bizelesin-treated CEM cancer cells: actual lesions (*A*) *versus* predicted motif density (potential sites, *B*).** *A*, actual lesion frequencies in specific regions of DNA from cells incubated with bizelesin for 4 h are based on QPCR stop assay systems (such as in Fig. 5). The decrease in signal amplification for each region (compared with DNA from untreated cells) has been typically quantitated for three to four drug levels in two to six independent experiments carried out in triplicate using two levels of template DNA. The resulting estimates of lesion frequencies were normalized to 1 μM drug (means ± S.E.). Data for ApoB MAR, c-myc MAR, and non-AT island domains have been recalculated based on the previously published full concentration dependence data (10). AF385609 denotes a shorter variant of AT island in GenBank™ Z79699. Lesions in bulk DNA (*cross-hatched bar* and *broken line*) have been determined previously by sedimentation analysis; Ref. 10). *B*, the predicted densities of bizelesin binding motif $T(A/T)_4A$ in individual regions are averaged over the entire lengths of respective QPCR products and therefore are somewhat lower than the peak motif densities shown for some of these regions in Table II. Motif density in bulk DNA (*cross-hatched bar* and *broken line*) represents the overall average frequency of bizelesin sites identified in the long range *in silico* analysis (Table I). The *inset* depicts the correlation between the levels of actual lesions in DNA from CEM cells (data from *panel A*) and the predictions based on motif density.

Properties of Bizelesin-targeted AT Islands

To better understand the nature of bizelesin-targeted domains, bioinformatics tools were used to evaluate (i) the minisatellite repeat consensus, (ii) the AT richness, (iii) potential for superhelically induced duplex instability, (iv) increased DNA flexibility, and (v) the putative MAR-like character of these regions. The attributes derived from these computational methods for several AT islands are listed in Table III. Fig. 7 depicts the properties of two selected AT islands (Z79699 and Z80771) in the context of their adjacent genomic DNA sequences.

No Common Consensus Other Than AT Richness—No single repeat consensus sequence is common to all these AT-rich islands, although specific types of $(A/T)_n$ repeats, occasionally interspersed with A/T-G/C motifs, can be defined for individual islands (Table III). The identified AT islands have cores containing from 200 to 800 bp of nearly pure or completely pure AT sequence (average A/T content of nearly 90%, Table III). For example, the AT island in Z79699 corresponds to either 578 and 324 bp of pure AT sequence, depending on the variant (*cf.* Fig. 3*A*). This magnitude of AT richness is unmatched at any other sites that are not preferential bizelesin targets throughout the analyzed genomic regions (Fig. 7*B*).

The AT Islands Targeted by Bizelesin Have Lowered Thermodynamic Stability and Are Easily Destabilized by Superhelicity—Consistent with the high AT content of the bizelesin targeted loci, these sites have the lowest stability of any positions within their genomic neighborhoods. This decreased duplex stability extends over several hundred base pairs (Fig. 7*C*) and corresponds precisely to the positions of the clusters of bizelesin binding sites (*cf.* Fig. 7*A*).

Local DNA destabilization *in vivo* is driven by superhelical stresses imposed on the molecule. Because local denaturation alters the level of stress throughout the domain, the imposed stresses couple together the destabilization behavior of all base pairs experiencing them. Accurate theoretical methods are available that analyze this global coupling and predict SIDD (20, 21). In these calculations, the extent of DNA destabilization is inversely related to $G(x)$, which is the free energy required to force the base pair at position x to always be in the denatured state. The negative values of $G(x)$ indicate sites with the high probability of being denatured.

As seen for long stretches of DNA sequences in Fig. 7*D*, most of the base pairs in a superhelical domain are not significantly destabilized, with high values of $G(x)$. The AT island regions of these sequences, however, show a dramatic reduction in the $G(x)$ values. Thus, the SIDD profiles for bizelesin-targeted AT islands shown in Fig. 7*D* predict that these two AT islands are strongly destabilized over several hundred bp. Analogous pattern has been observed with all the other AT islands listed in Table III ($G(x)$ profiles not shown). With one exception (the AT island in AC005195), each bizelesin-targeted AT island is the first area to become destabilized as the level of negative superhelical stress is increased. Furthermore, the bizelesin-targeted AT islands all exhibit distinctive destabilization properties observed in MAR domains (24, 25): at least 30 contiguous base pairs that are extremely SIDD-prone ($G(x) < 1$), surrounded by additional destabilized base pairs with $G(x) < 7$. As a numerical measure of both the length of destabilized sequences and the magnitude of destabilization, we used integrated SIDD potential assessed by integrating area under the curve for $G(x)$ minima (within the $G(x) < 7$ constraints). The values of integrated SIDD potential for several bizelesin-targeted AT islands exceed 5000 (Table III), much above the minimal value of 1500 expected for a distinct MAR domain (24, 25). Moreover, there is only one non-AT island site in the analyzed sequences that satisfies these criteria for a MAR with integrated SIDD measure of ~1900 (positions 44196 and 44433 in the AC005195). Thus, bizelesin-targeted AT islands tend to be primary sites of destabilization in their sequence context, which is consistent with a role as strong MAR domains. The integrated SIDD potential correlates well with the total number of bizelesin sites/AT island (Fig. 8*A*, $r^2 = 0.63$, $n = 8$).

AT Islands Targeted by Bizelesin Are Loci of Unusually High DNA Flexibility—Elevated local DNA flexibility, computed as the possible average propeller twist angles (26), has been suggested to accompany AT-rich elements of genetic instability (fragile sites) (34). Analogous calculations for DNA sequences containing the identified AT islands also show enhanced DNA flexibility. All the analyzed prominent bizelesin-targeted AT islands exhibit broad peaks of flexibility, exceeding 16° in sev-

TABLE III
Properties of AT islands targeted by bizelesin
Details on the determination of specific attributes listed in the table are given under "Experimental Procedures."

GenBank™ entry	Consensus repeat sequence and perfect repeat length[a]	AT-rich core sequence	Duplex stability T_m (minimum positions)	Integrated SIDD potential (site positions)	DNA flexibility twist angle (peak positions)	Integrated MAR potential (peak positions)
		% match	*°C*		*degrees*	*× 10^{-3}*
Z79699	ATATATATTTATATATA TATTTATATTT	$(A/T)_{578}$ (98)	42.4	6819.7	16.4	195
	28-mer (n = 19.5, 97)		(37360..7880)	(37330..37931)	(37126..37826)	(37300.38000)
Z80771	TTATATATAAGTATATA TTTATATAATT	$(A/T)_{804}$ (85) or $[(A/T)_{23}G/C(A/T)_3]_{27}$ (86)	45.0	6591.1	15.4	75
	28-mer (n = 21.6, 91)		(3100.3694)	(3060..3744)	(2876..4050)	(3100.3700)
X04682	TTTTATAATTAAAATAT TTATAATTAAATA	$(A/T)_{556}$ (92)	49.6	5002.1	13.5	73
	30-mer (n = 18.2, 93)		(14817..15336)	(14798..15372)	14676..15226)	(14700..15100)
AC005195	ATATATATATATTCC	$(A/T)_{646}$ (87) or $[(A/T)_{13}(G/C)_2]_{43}$ (92)	47.8	6622.6	16.6	128
	15-mer (n = 417.7, 92)		(63264..63850)	(63205..63888)	(63051..64050)	(63200..63900)
Z72519	TA 2-mer (n = 161.5, 78) $(A/T)_{14}(G/C)(A/T)_3$ 18-mer (n = 28.6, 81)	$(A/T)_{318}$ (88)	42.8 (33680..33975)	2709.5 (33670..34007)	16.8 (33501..34201)	51 (33500..34200)
Z82900	ATATACTAATATATTAA TATAAATAATATATTAAT	$(A/T)_{560}$ (92)	43.6	5176.6	15.9	12
	35-mer (n = 11.6, 84)		(76163–76678)	(76150..76739)	(75476..76925)	(76100..76800)
AF217490	TA 2-mer (78)	$(A/T)_{191}$ (91)	38.8 (147255..147390)	1966.3 (147225..147418)	13.7 (38951..39026)	62 (146900..14780)
X00364	No consensus identified	$[(A/T)_8(G/C)(A/T)_2(G/C)]_{24}$ (73)	55.8	2883.3	11.5	7
			(7418..7572)	(7362..7709)	(at 7451)	(7200..7500)

[a] Copy no. (n) and % match are given in parentheses.

eral cases (Fig. 7*E*, Table III). These peak values are well above the level typical of the surrounding local genomic sequences (<10°).

AT Islands Targeted by Bizelesin Co-localize with Loci of High MAR Potential—MAR-binding regions within genomic DNA do not have any single, common consensus sequence. However, an algorithm has been developed that assesses the potential for MAR function by constructing weighted averages of several properties known to occur at some MARs, but not at all MARs (27).

This approach was used to assess the MAR potential of the identified AT islands. The analysis indicated that most bizelesin binding hot spots collocate with domains that are computationally identified as likely MAR candidates. The peaks of MAR potential found in AT islands dominated other peaks within tens, and in some cases hundreds, of kbp (Fig. 7*F* and data not shown). Whereas the MAR potential prediction algorithm takes 75% of the maximal peak as the threshold for significance of adjacent potential MARs (27), sequences near the prominent AT islands commonly displayed negligible background values (often <10–15% of the peaks). Analysis of selected loci using individual criteria of the algorithm rather than the collective set indicated that the MAR potential of AT islands reflects their AT richness, AT motifs implicated in the origins of replication, and topoisomerase II sites. On the other hand, GC richness, bending, and kinking were found to be irrelevant (data not shown).

Analysis of a data base of MAR sequences (transfac.gbf.de/SMARtDB/index.html) for bizelesin binding sites indicates that not all known MARs are among bizelesin targets (data not shown). However, all the identified regions with dense and extended clusters of bizelesin binding sites exhibit high values of the integrated MAR potential, exceeding manyfold the value of 7×10^3 for the known MAR in the c-Myc and comparable or greater than the value of 73×10^3 for the known MAR in ApoB (X00364, and X04682, respectively, Table III). Collectively, the integrated MAR potential assessed for these AT islands and other bizelesin-targeted loci (see on-line supplement) correlates with the number of drug binding sites (Fig. 8*B*, r^2 = 0.46, n = 36).

DISCUSSION

Lesions in domains of cellular DNA that are crucial for tumor cell functioning are likely to have particular significance for selectively and predictably inducing cancer cell death. By contrast, damage to unessential targets may contribute little to the killing of cancer cells, yet produce adverse (*e.g.* mutagenic) effects in normal cells (5, 35). Thus, drug region specificity could both focus damage at critical regions and ameliorate undesirable "collateral" damage. Bizelesin, an AT-specific drug, is the first small molecule drug with *in vivo* anticancer activity found to have a potential for region-specific DNA damage (10). This report comprehensively verifies the region specificity of bizelesin at the genomic level and characterizes its preferential targets, domains of AT-rich DNA.

The computational (*in silico*) analysis of genomic sequences proved to be useful for identifying drug-targeted regions in the genome. The distributions of the previously determined bizelesin binding motifs, $T(A/T)_4A$ for interstrand cross-links (major sites) and $A(A/T)_4A$ for monoadducts (minor sites) (8), were analyzed in a large selection of human sequences representing ~1–1.5% of the human genome. In the majority of sequences examined, both motifs are infrequent and scattered. However, hot spots are found (~1/10^6 bp) that have dense clusters of potential binding sites with local motif densities up to ~35 times the average motif density (Fig. 2, Table I). Consistent

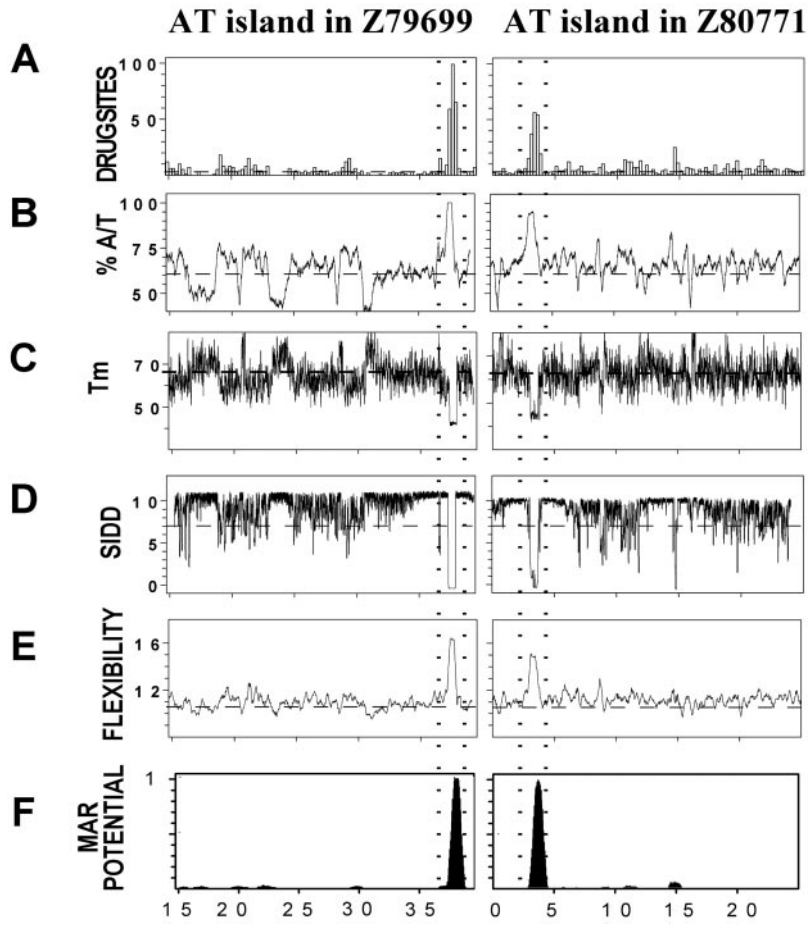

FIG. 7. **Characteristics of the AT islands identified in Z79699 and Z80771 loci in their respective sequence contexts (computational analysis).** *A*, distribution of bizelesin binding sites (cross-linking motif, re-plotted from Fig. 2*A*; for details see legend to Fig. 2). *Vertical dotted lines* are drawn down from *panel A* to indicate the location of bizelesin-targeted regions in the other panels. *B*, percentage of A/T bases (250-bp moving window). *C*, thermodynamic duplex stability (oligonucleotide melting temperatures). *D*, SIDD. The minimum in duplex incremental free energy $G(x)$ (*ordinate*) corresponds to the maximum probability for duplex opening at that region. *E*, DNA flexibility. The possible variations in propeller twist angle of DNA bases. *F*, MAR potential. Peaks in MAR potential are indicative of sequences having properties that are associated with known MAR domains. *Horizontal broken lines* in *A–C* and *E* correspond to an average value of a specific parameter across several long (>35 kbp) GenBank™ sequences listed in Table III. *Horizontal broken lines* in *D* indicate the level of $G(x) = 7$, a cut-off for destabilized sites. The data cover 25-kbp sections of each GenBank™ sequence including a respective AT island, except for *apoB* gene, for which a complete sequence (~16 kbp) is shown. For details on the computation of specific parameters, see "Experimental Procedures."

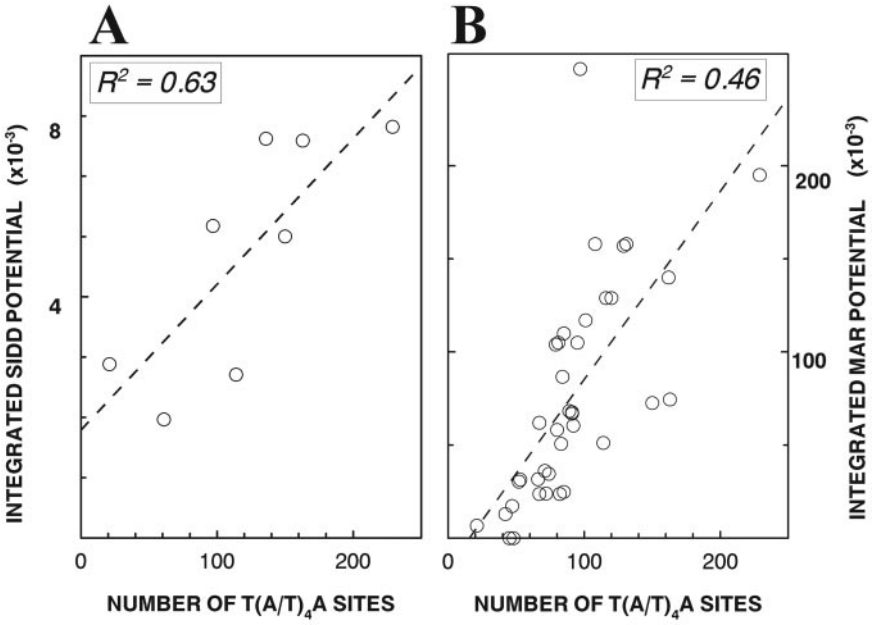

FIG. 8. **Correlation between the total number of potential bizelesin binding sites per AT island and integrated SIDD effect (*A*) and integrated (un-normalized) MAR potential (*B*).** Data shown are for the AT islands listed in Table III (*both panels*) and several additional other AT islands (*B*) listed in the on-line supplement.

with the AT-specific binding motifs of bizelesin, its targeted hot spots are islands of extremely AT-rich DNA.

The profound difference between the high local motif density in the identified AT islands and the low average motif density elsewhere suggests the potential of bizelesin for producing highly region-specific DNA damage on a genome-wide scale. Both bizelesin motifs confer region specificity by recognizing the same hot spots. In contrast, a less specific binding motif $(A/T)_3A$ for a related drug, adozelesin, is not sufficient for region specificity due to its high abundance outside of the hot spot areas (data not shown). Accordingly, adozelesin forms markedly more lesions in various regions than bizelesin, although lesions by bizelesin are more cytotoxic (10).

Corroborating these *in silico* predictions, bizelesin adducts in both naked DNA and in drug-treated cancer cells show remarkable region specificity. Drug adducts to naked DNA co-localize with the AT-richest core of a model AT island (Fig. 3). Compared with this AT island DNA, a model non-AT island region

shows two orders of magnitude lower frequency of bizelesin adducts (Fig. 4). Bizelesin lesions in DNA from drug-treated cancer cells follow a similar pattern. A much higher frequency of drug adducts was found in the prominent AT islands, which contain clusters of potential binding sites, than in both several other specific regions and in bulk DNA (Figs. 5 and 6). An excellent correlation exists between the predicted site densities and the actual region-specific lesion frequency determined in cellular DNA (r^2 = 0.86 in CEM cells). Along with our earlier investigations (4, 10), these results prove the power of *in silico* predictions to correctly identify drug-targeted regions in cancer cell genomes.[2]

Bizelesin-targeted AT islands are distinguished from other AT tracts by their size and the magnitude of their AT richness (from ~80 to 100% A/T, extending ~200–1000 bp). These long AT islands are repetitive minisatellites, although no shared sequence consensus other than the generic $(A/T)_n$ can be identified. At least some of these AT islands exist in multiple variants, as has been seen for the AT islands in Z79699/AF385609 (Fig. 3) and the ApoB MAR (10, 36, 37). The AT island in Z80771, however, exists in a single form, at least in CEM cells (*cf.* Fig. 5*A*). Being minisatellites, these identified domains are likely to be hypervariable and genetically unstable due to polymerase slippage and/or recombinational events (31, 38, 39). Certain AT-rich sites of known genomic instability, including the fragile sites FRA16D and FRA16B, have been implicated in leukemias and lymphomas (31, 34, 40–43). Both FRA16D and FRA16B are among the AT islands that are targeted by bizelesin. The expansion and/or amplification of AT islands, characteristic of both FRA16D and FRA16B, would profoundly increases their size making the expanded AT islands bigger as drug targets. As proposed for fragile sites (31), resulting size changes (dynamic mutations) may also contribute to the carcinogenic phenotypes.

The extreme AT richness of the identified AT islands results in their highly unusual properties that are unmatched in the long stretches of DNA spanning from tens to hundreds of kbp (Tables II and III; Fig. 7). The thermodynamic instability of AT island domains is reflected by a decrease in the local duplex melting temperature at their positions. They also show an unusually high DNA flexibility, allowing for a wide range of possible propeller twist angles. Moreover, AT islands entirely dominate adjacent sequences in their propensity for partial base unpairing induced by superhelical stress (SIDD). The SIDD results strongly suggest that the identified AT islands can become unpaired even under modest levels of supercoiling that are commonly encountered in eukaryotic cells.

Their SIDD properties and other sequence characteristics suggest that the bizelesin hot spot AT islands are likely to coincide with prominent MARs, which are specialized loci where chromosomal DNA is anchored to the nuclear matrix (44, 45). Computational analysis invariably indicated that prominent bizelesin hot spots were the only regions identified as strong MAR candidates (high MAR potential) in their entire respective sequence environments (Fig. 7). Both SIDD potential and MAR potential correlate well with the number of drug binding sites in AT islands, implying that strong MARs are the most likely targets for bizelesin (Fig. 8). Some of the identified AT islands have been experimentally shown to function as MARs. Examples include the c-Myc MAR (29) and the ApoB MAR (30). Our preliminary results confirm also that the Z79699/AF385609 AT island, the "hottest" identified region for bizelesin binding motifs, binds specifically to the isolated nuclear matrix, and hence functions as an *in vitro* MAR. The prominent MAR-like AT islands identified in this report as well as the known most prominent strong MAR domains (46–49) are both found with typical frequencies of $1/10^6$ bp.

It seems likely that bizelesin binding to MAR-like AT island domains affects their MAR function. Whereas the unusual structural properties of AT islands do not interfere with bizelesin binding, any bizelesin adduct, especially interstrand cross-links, will markedly stabilize these inherently destabilized duplexes. Bizelesin binding in the minor groove is also expected to stiffen DNA by reducing the flexibility of propeller twisting (50, 51). Elimination of the native partially unwound and flexible conformation of AT islands may thus cause structurally similar outcomes as the A/T to G/C mutations that are known to abrogate MAR function (25). This possibility is consistent with the potent and selective inhibition of DNA replication by bizelesin. Our previous studies in BSC-1 cells (52) and analogous recent results in CEM cells[2] found that the frequency of replication-inhibiting bizelesin adducts ($2–3/10^6$ bp) is comparable to the average size of replication factories ($\sim 10^6$ bp), in which strong MARs are thought to control the nearly synchronous initiation of multi-replicon clusters (48). Interference with critical elements of the replication apparatus provides a plausible explanation for why targeting AT islands is several orders of magnitude more lethal than is non-region-specific DNA damage (4, 8–10, 52).

Given the promising antitumor properties of bizelesin, a potential utility of AT island targeting warrants further exploration. Ongoing studies do suggest that lesions in MAR-like AT islands can be more lethal to cancer CEM cells than to normal WI-38 cells. It remains to be established what is the precise mechanism by which damage to AT islands impedes cell function and which tumors would most profoundly differ from normal cells in their responses. Further studies are also needed to establish whether certain subclasses of AT islands may be more critical than others. The exact genomic locations of the very infrequent bizelesin lesions that are sufficient for cell growth inhibition ($<10^1$ lesions/cell) (4, 10) remain unknown. Nonetheless, it seems highly likely that at least some of these few lesions responsible for lethal effects occur in distinct AT islands, in which drug motif clusters. Whereas sporadic lesions might occur elsewhere at numerous scattered sites, they can be expected to contribute less to bizelesin's biological effects.

Bizelesin-targeted MAR-like AT islands provide a proof of principle that targeting repetitive sequences crucial for the survival of cancer cells is feasible. Targeting repetitive sequences seems, in general, an alternative to targeting unique regions within specific genes. In contrast to unique regions, clusters of drug binding sites in repetitive sequences constitute abundant targets that are far easier to selectively hit with small molecules. Bioinformatics/pharmacogenomics tools can identify binding sites and targeted regions for any DNA-reactive agent, ultimately throughout the entire human genome. The use of these tools should accelerate and reduce the cost of development of new region-specific molecules. Hence, further studies are warranted not only to better understand the utility of AT islands as a target but also to elucidate the general rules of regional DNA targeting and their applicability to rational design of more specific anticancer drugs.

Acknowledgments—We thank Brenda Arnett for excellent technical assistance, Dr. Ryszard Konopinski for cloning the Z79699 fragment, and Drs. Laurence Hurley, Terence Herman, and Barbara Woynarowska for numerous stimulating and inspiring discussions.

[2] M. C. S. Herzig and J. M. Woynarowski, unpublished data.
[3] J. M. Woynarowski, A. V. Trevino, K. A. Rodriguez, S. C. Hardies, and C. J. Benham, unpublished data.

REFERENCES

1. Kohn, K. W., Hartley, J. A., and Mattes, W. B. (1988) *Biochem. Pharmacol.* **37,** 1799–1800
2. Hartley, J. A., Lown, J. W., Mattes, W. B., and Kohn, K. W. (1988) *Acta Oncol.* **27,** 503–510
3. Woynarowski, J. M., Chapman, W. G., Napier, C., Herzig, M., and Juniewicz, P. (1998) *Mol.. Pharmacol.* **54,** 770–777
4. Herzig, M. C., Trevino, A. V., Arnett, B., and Woynarowski, J. M. (1999) *Biochemistry* **38,** 14045–14055
5. Lawley, P. D., and Phillips, D. H. (1996) *Mutat. Res.* **355,** 13–40
6. Lee, C. S., Pfeifer, G. P., and Gibson, N. W. (1994) *Biochemistry* **33,** 6024–6030
7. Weiland, K. L., and Dooley, T. P. (1991) *Biochemistry* **30,** 7559–7565
8. Woynarowski, J. M., Chapman, W. G., Napier, C., and Herzig, M. C. S. (1999) *Biochim. Biophys. Acta* **1444,** 201–217
9. Woynarowski, J. M., McHugh, M., Gawron, L. S., and Beerman, T. A. (1995) *Biochemistry* **34,** 13042–13050
10. Woynarowski, J. M., Napier, C., Trevino, A. V., and Arnett, B. (2000) *Biochemistry* **39,** 9917–9927
11. Punt, C. J., Humblet, Y., Roca, E., Dirix, L. Y., Wainstein, R., Polli, A., and Corradino, I. (1996) *Br. J. Cancer* **73,** 803–804
12. Weiss, G. R., Poggesi, I., Rocchetti, M., DeMaria, D., Mooneyham, T., Reilly, D., Vitek, L. V., Whaley, F., Patricia, E., Von Hoff, D. D., and O'Dwyer, P. (1998) *Clin. Cancer Res.* **4,** 53–59
13. Foster, B. J., LoRusso, P. M., Poplin, E., Zalupski, M., Valdivieso, M., Wozniak, A., Flaherty, L., Kasunic, D. A., Earhart, R. H., and Baker, L. H. (1995) *Invest. New Drugs* **13,** 321–326
14. Fleming, G. F., Ratain, M. J., O'Brian, S. M., Schilsky, R. L., Hoffman, P. C., Richards, J. M., Vogelzang, N. J., Kasunic, D. A., and Earhart, R. H. (1994) *J. Natl. Cancer Inst.* **86,** 368–372
15. Shamdas, G. J., Alberts, D. S., Modiano, M., Wiggins, C., Power, J., Kasunic, D. A., Elfring, G. L., and Earhart, R. H. (1994) *Anticancer Drugs* **5,** 10–14
16. Schwartz, G. H., Aylesworth, C., Stephenson, J., Johnson, T., Campbell, E., Hammond, L., Von Hoff, D. D., and and Rowinsky, E. K. (2000) *Proc. Amer. Soc. Clin. Oncol.* **19,** 235a
17. Pavlidis, N., Aamdal, S., Awada, A., Calvert, H., Fumoleau, P., Sorio, R., Punt, C., Verweij, J., Van Oosterom, A., Morant, R., Wanders, J., and Hanauske, A. R. (2000) *Cancer Chemother. Pharmacol.* **46,** 167–171
18. Benson, G. (1999) *Nucleic Acids Res.* **27,** 573–580
19. Rychlik, W., and Rhoads, R. E. (1989) *Nucleic Acids Res.* **17,** 8543–8551
20. Benham, C. J. (1992) *J. Mol. Biol.* **225,** 835–847
21. Fye, R. M., and Benham, C. J. (1999) *Phys. Rev. E* **59,** 3408–3426
22. Benham, C. J. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2999–3003
23. Benham, C. J. (1996) *J. Mol. Biol.* **255,** 425–434
24. Benham, C., Kohwi-Shigematsu, T., and Bode, J. (1997) *J. Mol. Biol.* **274,** 181–196
25. Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., and Kohwi-Shigematsu, T. (1992) *Science* **255,** 195–197
26. Sarai, A., Mazur, J., Nussinov, R., and Jernigan, R. L. (1989) *Biochemistry* **28,** 7842–7849
27. Kramer, J. A., Singh, G. B., and Krawetz, S. A. (1996) *Genomics* **33,** 305–308
28. Singh, G. B., Kramer, J. A., and Krawetz, S. A. (1997) *Nucleic Acids Res.* **25,** 1419–1425
29. Chou, R. H., Churchill, J. R., Flubacher, M. M., Mapstone, D. E., and Jones, J. (1990) *Cancer Res.* **50,** 3199–3206
30. Levy-Wilson, B., and Fortier, C. (1989) *J. Biol. Chem.* **264,** 21196–21204
31. Sutherland, G. R., Baker, E., and Richards, R. I. (1998) *Trends Genet.* **14,** 501–506
32. Norris, R. A., and Kern, M. J. (2001) *DNA Cell Biol.* **20,** 89–99
33. Bolzan, A. D., Paez, G. L., Bianchi, M. S., and Bianchi, N. O. (2000) *Cancer Genet. Cytogenet.* **120,** 166–170
34. Ried, K., Finnis, M., Hobson, L., Mangelsdorf, M., Dayan, S., Nancarrow, J. K., Woollatt, E., Kremmidiotis, G., Gardner, A., Venter, D., Baker, E., and Richards, R. I. (2000) *Hum. Mol. Genet.* **9,** 1651–1663
35. Sanderson, B. J., and Shield, A. J. (1996) *Mutat. Res.* **355,** 41–57
36. Wang, D. M., Taylor, S., and Levy-Wilson, B. (1996) *J. Lipid Res.* **37,** 2117–2124
37. Desmarais, E., Vigneron, S., Buresi, C., Cambien, F., Cambou, J. P., and Roizes, G. (1993) *Nucleic Acids Res.* **21,** 2179–2184
38. Glover, T. W. (1998) *Recent Res. Cancer Res.* **154,** 185–199
39. Handt, O., Sutherland, G. R., and Richards, R. I. (2000) *Mol. Genet. Metab.* **70,** 99–105
40. Takahashi, E., Kaneko, Y., Ishihara, T., Minamihisamatsu, M., Murata, M., and Hori, T. (1988) *Hum. Genet.* **80,** 124–126
41. 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
42. Murata, M., Takahashi, E., Ishihara, T., Minamihisamatsu, M., Takagi, T., Kaneko, Y., and Hori, T. (1987) *Cancer Genet. Cytogenet.* **25,** 81–86
43. Le Beau, M. M., and Rowley, J. D. (1986) *Adv. Hum. Genet.* **15,** 1–54
44. Boulikas, T. (1995) *Int. Rev. Cytol.* **162A,** 279–388
45. 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
46. Khodarev, N. N., Sokolova, I. A., and Vaughan, A. T. M. (1998) *J. Cell. Biochem.* **70,** 604–615
47. Vassetzky, Y. S., Bogdanova, A. N., and Razin, S. V. (2000) *J. Cell. Biochem.* **79,** 1–14
48. Berezney, R., Dubey, D. D., and Huberman, J. A. (2000) *Chromosoma* **108,** 471–484
49. Cockerill, P. N., and Garrard, W. T. (1986) *Cell* **44,** 273–282
50. Ha Duong, T., and Zakrzewska, K. (1997) *J. Biomol. Struct. Dyn.* **14,** 691–701
51. Thompson, A. S., and Hurley, L. H. (1995) *J. Mol. Biol.* **252,** 86–101
52. Woynarowski, J. M., and Beerman, T. A. (1997) *Biochim. Biophys. Acta* **1353,** 50–60
53. Boerwinkle, E., Xiong, W. J., Fourest, E., and Chan, L. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 212–216