Combinatorial Identification of a Novel Consensus Sequence for the Covalent DNA-Binding Polyamide Tallimustine†

Gulshan Sunavala-Dossabhoy‡ and Michael W. Van Dyke*

Department of Molecular and Cellular Oncology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030

Received September 30, 2004; Revised Manuscript Received November 22, 2004

ABSTRACT: Many agents successfully used in cancer chemotherapy either directly or indirectly covalently modify DNA. Examples include cisplatin, which forms a covalent adduct with guanines, and doxorubicin, which traps a cleavage intermediate between topoisomerase II and torsionally strained DNA. In most cases, the efficacy of these drugs depends on the efficiency and specificity of their DNA binding, as well as the discrimination between normal and neoplastic cells in their handling of the drug−DNA adducts. While much is known about the chemistry of drug−DNA adducts, little is known regarding the overall specificity of their formation, especially in the context of a whole human genome, where potentially billions of binding sites are possible. We used the combinatorial selection method restriction endonuclease protection, selection, and amplification (REPSA) to determine the DNA-binding specificity of the semisynthetic covalent DNA-binding polyamide tallimustine, which contains a benzoic acid nitrogen mustard appended to the minor groove DNA-binding natural product distamycin A. After investigating over 134 million possible sequences, we found that the highest affinity tallimustine binding sites contained one of two consensus sequences, either the expected distamycin hexamer binding sites followed by a CG base pair (e.g., 5′-TTTTTTC-3′ and 5′-AAATTTC-3′) or the unexpected sequence 5′-TAGAAC-3′. Curiously, we found that tallimustine preferentially alkylated the N7 position of guanines located on the periphery of these consensus sequences. These findings suggested a cooperative binding model for tallimustine in which one molecule noncovalently resides in the DNA minor groove and locally perturbs the DNA structure, thereby facilitating alkylation by a second tallimustine of an exposed guanine on another side of the DNA.

Even with the development of novel approaches for the treatment of cancer, for example, antibodies directed against growth factor receptors and antiangiogenic peptides, there is increasing recognition that the effective treatment of disseminated disease will also require the use of more conventional cytotoxic agents. Many drugs most commonly used in the treatment of solid and hematological malignancies incorporate the formation of covalent adducts with DNA as part of their presumed mechanism of action (reviewed in ref *1*). Some, such as the alkylators cisplatin and cyclophosphamide, form adducts by reacting directly with DNA. Others, such as the intercalators doxorubicin and topotecan, bind DNA noncovalently but stabilize covalent intermediates between certain DNA-modifying enzymes and their substrates. In most cases, the effectiveness and toxicity of these small molecule (<1 kDa) drugs rely in large part on their specific interactions with DNA and the cellular responses to these unique (protein) drug−DNA complexes. Thus, improved understanding of the recognition of DNA by these small molecule drugs and the repair of their resulting adducted DNAs should be instrumental in developing more efficacious and less injurious cytotoxic antineoplastic agents.

Most small molecules that form adducts directly with DNA do so with limited sequence specificity (*2*). This in part reflects the specificity of the chemical reaction that forms the covalent bond and the limited interface between the small molecule and DNA. However, some small molecules that can form adducts also interact noncovalently with a relatively large surface on duplex DNA. Examples include adozelesin, carzelesin, and bizelesin, structural derivatives of the antibiotic CC-1065, and tallimustine, a derivative of the antibiotic distamycin A (*3*, *4*). These small molecule ligands primarily interact with duplex DNA through its minor groove and typically reside in AT-rich sequences. Since only a subset of all AT-rich sequences are preferred noncovalent binding sites for these minor groove binders, their intrinsic noncovalent binding specificity can ultimately influence where covalent adduct formation occurs.

Alkylating agents that bind DNA with high sequence specificity have been of great interest for their potential use as antitumor agents. Tallimustine (FCE 24517, PNU-152241) (Figure 1A) is one of the distamycin derivatives that has reached clinical trials because of its potent antitumor activity,

† This research was supported by grants from the American Cancer Society (RPG-97-028-03-LBC) and the Robert A. Welch Foundation (G-1199).

* Corresponding author address: Department of Molecular and Cellular Oncology, Unit 079, U. T. M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030-4009. Phone: +1-713-792-8954. Fax: +1-713-794-0209. E-mail: mvandyke@mdanderson.org.

‡ Present address: Department of Biochemistry and Molecular Biology, Louisiana State University Health Science Center−Shreveport, 1501 Kings Highway, Shreveport, LA 71130.

10.1021/bi047877l CCC: $30.25
Published on Web 01/28/2005

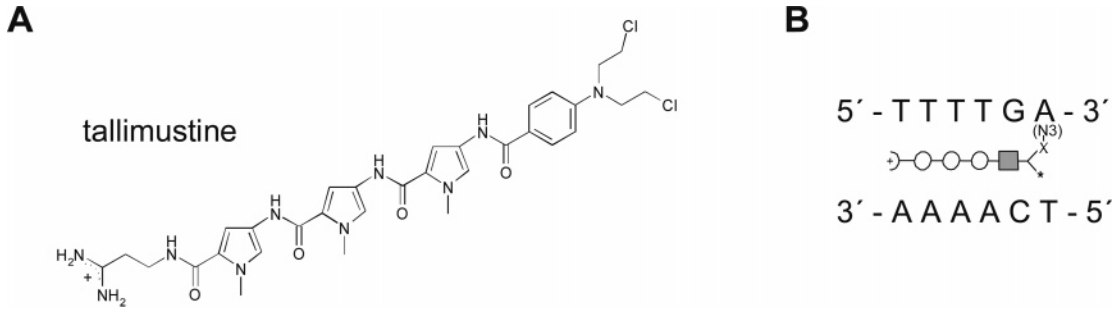

FIGURE 1: (A) Chemical structure of tallimustine (FCE24517, PNU152241), a benzoyl nitrogen mustard derivative of distamycin A. (B) Schematic representation of tallimustine binding in the DNA minor groove to its literature consensus sequence. Polyamide subunits include 3-aminopropionamidine (⊃), 4-amino-1-methyl-1*H*-pyrrole-2-carboxylic acid (○), and 4-[bis(2-chloroethyl)amino]benzoic acid (■<). Reacted (X) and unreacted (*) 2-chloroethylamines, as well as the site of covalent attachment (e.g., N3 position on adenine), are indicated.

though its utility has been tempered by its severe myelotoxicity (*5*−*9*). Tallimustine was synthesized by tethering the chemically reactive benzoyl nitrogen mustard moiety to the DNA-binding frame of distamycin A (*10*). The *N*-methylpyrrole-carboxamide framework of distamycin binds noncovalently in the DNA minor groove with an AT-rich sequence preference, while the nitrogen mustard moiety alkylates DNA bases to form covalent adducts. Studies on its mechanism of action suggest that, similar to CC-1065, tallimustine binds in the minor groove of the DNA, where it alkylates the N3 position of adenine within the minor groove (*11*, *12*). Thus, unlike the classical nitrogen mustards, tallimustine apparently does not alkylate guanine N7 in the major groove, or exhibit a strong potential for interstrand DNA cross-linking.

Repetitive primer extension studies and footprinting data demonstrated the binding of tallimustine to AT-rich regions of the DNA (*11*, *13*). Tallimustine recognized discrete sequence elements and was found to be more specific in sequence recognition than distamycin. Tallimustine displayed a high sequence specificity for 5′-TTTTGA-3′ and 5′-TTTTAA-3′ motifs and preferentially alkylated the N3 position of the 3′ terminal adenine in these sequences (Figure 1B) (*12*, *14*). A recent study assessing tallimustine sites on cellular DNA showed that, despite its high sequence specificity, tallimustine also demonstrated an affinity for many atypical motifs not corresponding to the previously reported consensus sequences 5′-TTTTGPu-3′ and 5′-TTTTAA-3′ (*15*). The aim of our present study was to examine the full DNA-binding sequence selectivity of tallimustine and determine those molecular features that direct tallimustine−DNA adducts. Using a modification of our combinatorial approach restriction endonuclease protection, selection, and amplification (REPSA),[1] we identified two new consensus tallimustine−DNA binding sequences and obtained evidence that the preferred tallimustine alkylation sites may not be the N3 position of 3′ terminal adenines but instead the N7 position of adjacent guanines.

MATERIALS AND METHODS

Chemicals. Tallimustine was a generous gift from Pharmacia & Upjohn (Milan, Italy). It was prepared as a 10 mM stock solution in dimethyl sulfoxide and stored in amber tubes at −20 °C. Further dilutions were made in sterile distilled water immediately before use.

Oligonucleotides. Phosphodiester oligonucleotides were prepared on a Millipore (Bedford, MA) Cyclone Plus DNA synthesizer using standard phosphoramidite chemistry, essentially as previously described (*16*). The nucleotide sequences of oligonucleotides used in this study were (5′ → 3′): 63AL, CTAGGAATTCGTGCAGAGGTGA-3′; 63AR, GTCCAAGCTTCTGGAGGGATGGTAA; 63R14, CTAGGAATTCGTGCAGAGGTGA-N_{14}-TTACCATCCCTCCAGAAGCTTGGAC. For oligonucleotide 63R14, the sites containing mixed bases (N) were synthesized using an equimolar mixture of each phosphoramidite.

Selection template ST-2 (Figure 2A) was synthesized by six rounds of a standard polymerase chain reaction (PCR), using the oligonucleotide 63R14 as a template and oligonucleotides 63AL and 63AR as primers. The distribution of nucleotides incorporated into the central random cassette was 27% A, 18% C, 16% G, and 39% T, as determined by sequencing of eight individual clones derived from the starting material (*16*).

REPSA. A flowchart of the steps used in REPSA for covalent DNA-binding ligands is shown in Figure 2B. To achieve site-specific tallimustine binding, 4 ng (96 fmol) of ST-2 template was incubated for 5 h at 37 °C in a 20-μL volume containing buffer A (20 mM Tris-acetate, pH 7.9, 50 mM K-acetate, 10 mM Mg-acetate, and 0.05% Nonidet P-40) and 100 μM tallimustine where indicated. Afterward, the reaction mixtures were diluted 5-fold with water containing 400 ng of poly(dG-dC) and then extracted thrice with 1 mL of water-saturated 1-butanol to remove unbound tallimustine before precipitation with 1 mL 1-butanol. The carrier DNA poly(dG-dC) was chosen since it has a very low affinity for distamycin, lacks any type IIS restriction endonuclease (IISRE) binding sites, and is not complementary to either primer used in PCR. To probe for tallimustine-bound sequences, the dried DNA pellets were resuspended in a 20-μL volume containing the appropriate IISRE buffer and digested at 37 °C with 2 units of either *Bpm*I or *Fok*I (New England Biolabs, Beverly, MA) for either 30 or 15 min, respectively. Different IISREs were used in alternate rounds of REPSA selection to minimize the selection of individual IISRE binding sites within the random sequence region (*17*). Finally the products of the restriction digestion were directly amplified in a PCR reaction containing 100 ng of 5′ ^{32}P end-labeled 63AL and 100 ng of unlabeled 63AR primers, 0.1

[1] Abbreviations: bp, base pair; IISRE, type IIS restriction endonuclease; N, random nucleotide; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; REPA, restriction endonuclease protection assay; REPSA, restriction endonuclease protection selection and amplification; SSC, saline sodium citrate buffer; ST-2, selection template 2; Tal, tallimustine.

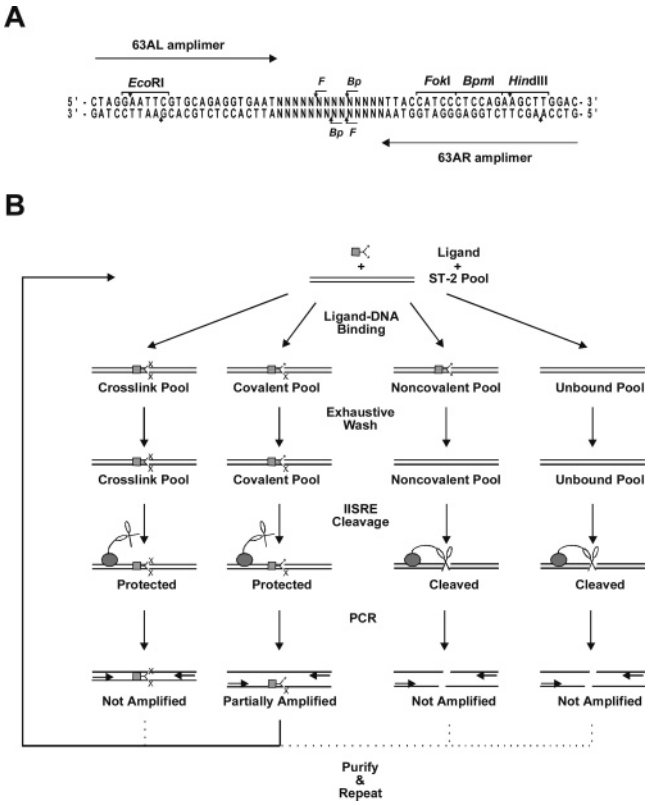

FIGURE 2: (A) Schematic representation of the REPSA selection template ST-2. Restriction endonuclease binding sites are indicated by brackets, and their cleavage sites are indicated by arrows. *Fok*I (*F*), *Bpm*I (*Bp*). (B) Flowchart for the combinatorial method REPSA as applied to ligands that covalently bind duplex DNA. A bis-reactive ligand is represented by the symbol ■<, and a type IIS restriction endonuclease is represented by scissors (DNA-cleaving domain) tethered to a gray oval (DNA-binding domain). Dotted lines indicate reduced contributions to pool.

mM concentrations of each dNTP, and 10 units of *Taq* DNA polymerase in buffer B (10 mM Tris-Cl, pH 9.0, 50 mM KCl, 2.5 mM $MgCl_2$, and 0.1% Triton X-100). The amplification profile was 94 °C for 1 min, 50 °C for 3 min, and 72 °C for 1 min, for a total of nine cycles. The amplified DNA was purified by phenol−chloroform extraction followed by spin filtration through Ultrafree-MC 5000 NMWL spin filters (Millipore, Billerica, MA) at 15000*g* for 30 min. The DNA retained on the filter was resuspended in 20 μL of TE buffer (10 mM Tris-Cl, pH 7.5, 1 mM ethylenediaminetetraacetic acid [EDTA]) and quantitated by an ethidium dot fluorescence assay (*18*). Two nanograms of this DNA served as the starting template for the next round of REPSA, while an aliquot of the remainder was analyzed by nondenaturing 8% polyacrylamide gel electrophoresis and autoradiography.

After seven rounds of REPSA, the selected DNAs were digested with *Eco*RI and *Hin*dIII and ligated into similarly cut plasmid pUC12. The plasmids were transformed into *Escherichia coli* strain XL1-blue cells, and the cells were cultured on LB-agar plates containing 100 μg/mL ampicillin. The clones were screened for inserts by blue-white color selection, and plasmid minipreps were made from the insert containing clones. The DNA inserts were sequenced using a Sequenase v2.0 kit (USB Corp., Cleveland, OH) and universal primers following manufacturer's protocols.

REPA. Radiolabeled fragments containing the DNA insert from individual clones were generated by PCR amplification using ^{32}P end-labeled 63AL and unlabeled 63AR primers and purified as described above. Radiolabeled DNA (96 fmol, 10 kcpm) was incubated for 15 min at 37 °C in a 20-μL volume containing buffer A and 50 μM tallimustine where indicated. Afterward, samples were processed with 1-butanol and digested with 1 unit of *Fok*I as described above. The resulting DNA products were resolved on an 8% nondenaturing polyacrylamide gel and visualized by autoradiography.

DNase I Footprinting. Radiolabeled DNA probes for DNAse I footprinting were generated by 3′-end filling *Hin*dIII-cut plasmid DNA with α[^{32}P]dATP and unlabeled deoxyribonucleotides, then cleaving with *Eco*RI and isolating the 53-bp fragments essentially as previously described (*19*). Ten thousand counts per minute of the labeled DNA probe was incubated for 15 min at 37 °C in 20-μL volume containing buffer A and 50 μM tallimustine, where indicated. After extraction and precipitation with 1-butanol, samples were resuspended in 20 μL of room temperature buffer C (10 mM Tris-HCl, pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, and 0.05% Nonidet P-40). DNase I (40 ng) was then added and digestion allowed to occur for 30 s before stopping and purification as previously described (*19*). DNA cleavage products were resolved by high-resolution denaturing PAGE and visualized by autoradiography. Products from purine-specific chemical cleavage reactions were used as markers.

Piperidine and Thermal Cleavage Assays. Identification of covalent tallimustine−DNA sites was carried out using radiolabeled DNA probes generated by PCR amplification of plasmid DNA clones with 63AL and 63AR primers. Depending on the DNA strand being investigated, either 63AL or 63AR was singly labeled on its 5′ end. Ten thousand counts per minute of the labeled DNA probe was incubated in a 20-μL volume under a variety of conditions (e.g., tallimustine concentration, buffer composition, pH, incubation time, and temperature). Following incubation, 1 μg of carrier plasmid DNA (pUC19) was added to the reaction and the volume made up to 100 μL prior to extraction with 1 mL of 1-butanol and ethanol precipitation. For piperidine cleavage assays, DNA was resuspended in 100 μL of 1 M piperidine and heated to 95 °C for 20 min (*20*). For thermal cleavage assays, DNA was resuspended in 100 μL of buffer SSC (1.5 mM Na-citrate, pH 7.2, 15 mM NaCl) and heated to 90 °C for 30 min (*21*). Purified DNA cleavage products were resolved by high-resolution denaturing PAGE and visualized by autoradiography as previously described. Products from purine-specific chemical cleavage reactions were used as markers.

RESULTS

REPSA Selection of Preferred Covalent Binding Sequences. REPSA, like most PCR-based combinatorial methods, relies on multiple rounds of selection and amplification to identify DNA sequences bound with the highest affinity by a particular ligand. However, unlike conventional combinatorial methods, which require the physical separation of ligand-bound DNAs from free DNAs, REPSA achieves its selection enzymatically, through the inhibition of IISRE cleavage on ligand-bound DNAs. IISREs, unlike other type

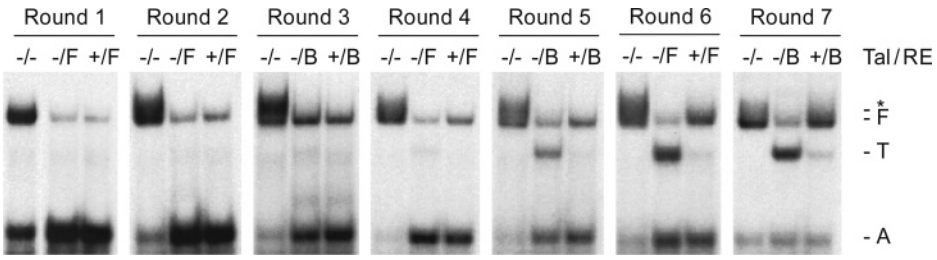

FIGURE 3: Emergence of a tallimustine-dependent, IISRE cleavage-resistant oligonucleotide population. Shown are autoradiograms of the PAGE-resolved products of PCR amplifications containing pools of oligonucleotides (2 ng) that were either untreated (−/−), *Fok*I-digested in the absence (−/F) or presence of tallimustine (+/F), or *Bpm*I digested in the absence (−/B) or presence of tallimustine (+/B). Indicated are the expected 63-bp full-length PCR product (F), a truncated 49-bp PCR product (T), the 24-nt radiolabeled 63AL amplimer (A), and partially annealed "bubble" ST-2 (*).

II restriction endonucleases, bind to specific sequences of DNA but then cleave duplex DNA at a fixed distance from these binding sites, irrespective of the sequences being cleaved. REPSA can be performed under physiological conditions with relatively uncharacterized ligands or mixed populations of different ligand types. REPSA has been successfully used with macromolecular DNA-binding ligands, including the purine motif triple-helical DNA-forming oligodeoxyribonucleotide ODN 1 and the TATA box-binding subunit of the human class II general transcription factor TFIID (*16*, *17*). More important for our present studies, REPSA has also been successfully used to identify the preferred duplex-DNA binding sites of small molecule ligands, including the antibiotics distamycin A and actinomycin D and the hairpin polyamides ImPyPyPy-γ-PyPy-PyPy-β-Dp and ImPyPyPy-γ-ImPyPyPy-β-Dp (*22*−*24*). To the best of our knowledge, REPSA is the only combinatorial method capable of identifying consensus sequences bound by unmodified small molecule ligands in their native state.

To adopt REPSA for the selection of DNA sequences preferentially bound by covalent-binding small molecule ligands, we incorporated an additional step in the selection process, one that removes noncovalently bound ligands. For our tallimustine study, we used multiple extractions with 1-butanol in a wash step that followed ligand−DNA binding (Figure 2B). We found that 1-butanol extraction was highly effective in removing the noncovalent-binding parent compound, distamycin A, from its preferred AT-rich DNA sequence (data not shown). Thus we expected that, following 1-butanol extraction, those DNAs that did not contain covalently bound tallimustine would be cleaved by a IISRE and not be PCR amplified, thereby being underrepresented in the selection pool (Figure 2B, column 3).

Combinatorial selection of covalently modified DNAs has its own unique complications. Primary among these is that covalently modified DNAs do not serve as good templates for PCR amplification. Cross-linked DNAs (Figure 2B, column 1) are protected from IISRE cleavage but are completely refractory to PCR amplification. Thus they are not represented in the resultant pool of DNAs following selection. To a lesser extent, even monoadducts are underrepresented in a selection pool, since one of the template strands cannot serve as a suitable template. However, so long as one intact and unmodified DNA strand is present, some amplification can occur, and these DNAs will be represented in the selection pool (Figure 2B, column 2). Therefore, combinatorial selection of covalently modified DNAs is feasible, though less efficient, than selection of noncovalently bound DNAs.

REPSA Selects Tallimustine-Dependent, FokI Cleavage-Resistant DNAs. The ST-2 REPSA template is a 63-bp DNA containing a central 14-bp random unit flanked by defined segments having nested IISRE sites (Figure 2A). In ST-2 DNA, the IISREs *Bpm*I and *Fok*I bind in the right flank of the 14-bp random region and cleave the template near the center of the 14-bp region. The 14-bp randomized site potentially offers over 134 million ($4^{14}/2$) possible sequences. Since 4 ng (96 fmol, 4.8×10^{10} DNA molecules) of ST-2 DNA is used in our REPSA reactions, and the expected binding site for tallimustine is only 6 bp, then the initial REPSA reaction most likely possessed a high representation of all possible tallimustine-binding sites with which to interact.

The progress toward the REPSA selection of a tallimustine-dependent, IISRE cleavage-resistant DNA population was gauged using a differential PCR amplification assay. For each round of REPSA, three parallel reactions were performed with the same starting material. One control reaction lacked tallimustine during the binding reaction and the IISRE during the cleavage step (−/−). The absence of these reagents allowed maximal PCR amplification during the final step, and the full-length (63 bp) products of this reaction served as our amplification control. Another control reaction lacked tallimustine but included a IISRE (−/F or −/B), thereby allowing maximal cleavage and minimal amplification. Finally the products of a standard REPSA selection, with tallimustine and IISRE both present (+/F or +/B), were compared with the two aforementioned controls. Evidence of a successful selection process would then be observed by an increase in amount of PCR product obtained when tallimustine and IISRE were both present. Once this amount approached that of the −/− amplification control, it would be indicative of a substantial emergent population and would be an appropriate time to characterize the selected DNAs.

In total, seven rounds of REPSA selection were carried out with tallimustine and either the IISRE *Fok*I (F) or *Bpm*I (B). *Fok*I was used in REPSA rounds 1, 2, 4, and 6, while *Bpm*I was used in rounds 3, 5, and 7. Different IISREs were used in alternating REPSA rounds to minimize the possible selection of IISRE-binding sites within the randomized cassette, as had been previously observed (*17*). A PCR amplification analysis of each REPSA round is shown in Figure 3. For rounds 1 through 3, there were no apparent

Table 1: REPSA-Selected Sequences Containing Tallimustine Binding Sites

Clone	Sequence (5′→3′)[a]	REPA Protection[b]	Clone	Sequence (5′→3′)[a]	REPA Protection[b]
3015	gaat CAGC*TTTTTTC*CTA ttac	++	3012	gaat CGTTTGATGGATTA ttac	+
3016	gaat CACC**GTTCTA**CCGG ttac	++	3032	gaat TAAGATTTCTTGTA ttac	+
3018	gaat GACGT**GTTCTA**GAA ttac	++	3033	gaat TTAGTCTACTTCTT ttac	+
3023	gaat CAA**GTTCTA**TGAAG ttac	++	3034	gaat TCTCCTTTTAAGTG ttac	+
3043	gaat TTCG**GTTCTA**ACTA ttac	++	3036	gaat GATAAAGAGTACTC ttac	+
3046	gaat CATTTC**TAGAAC**AG ttac	++	3053	gaat TGTTATGTCTACAC ttac	+
3054	gaat CG**TAGAAC**ATGGAA ttac	++	3056	gaat AGCGCACTACCAGG ttac	+
3055	gaat TTTTGT**GTTCTA**TC ttac	++	3057	gaat TAATTCAACTAGGG ttac	+
3060	gaat TGT**GTTCTA**ACCTT ttac	++	3061	gaat CAAAACCCTTGTAA ttac	+
3062	gaat GT**TAGAAC**TAGGTC ttac	++	3065	gaat AGTCTGTGTCCTAA ttac	+
3067	gaat CTGG*AAATTTC*CAT ttac	++	3068	gaat GGATTCACGAAAGC ttac	+
3074	gaat GTGCTTA*TTTTTTC* ttac	++	3069	gaat GGATTCACGAAAGC ttac	+
3075	gaat GG**GTTCTA**AATGGC ttac	++	3077	gaat CATACTGATAAGAT ttac	+
3087	gaat CTCGA*TTTTTTC*TC ttac	++	3078	gaat ATCTAGCGATCCTT ttac	+
3089	gaat ATAC**GTTCTA**AAGC ttac	++	3080	gaat AGCGCACTACCAGG ttac	+
3094	gaat **TAGAAC**CCATACTT ttac	++	3085	gaat AGTGTCGTAGTTCT ttac	+
3009	gaat TACTGATTTTGCTG ttac	+	3092	gaat AAAGTCTACGACTA ttac	+
3010	gaat TTACGATCCGTTTA ttac	+	3093	gaat GGAAGTTAATGTAG ttac	+

[a] Sequences present in the 14-bp randomized region are indicated in uppercase letters, while sequences present in the flanking regions are indicated in lowercase letters. Italic typeface indicates type 1 (distamycin-like) consensus binding sites. Bold typeface indicates type 2 tallimustine consensus binding sites. [b] Values correspond to the efficiency of *Fok*I cleavage inhibition observed when 50 μM tallimustine was present: 80% or less cleavage inhibition (+) and more than 80% cleavage inhibition (++).

differences between amount of product DNA obtained in the −/F(−/B) and +/F(+/B) reactions, while both of these yielded less than 10% of the DNA obtained in the corresponding −/− control. That the yields were comparable in the −/F(−/B) and +/F(+/B) reactions indicated that the tallimustine-binding sequences were a very small percentage of the DNAs present in these populations, while the differences between these and the −/− controls indicated the efficiency of selection in these rounds (approximately 10-fold). However, by round 4, an emergent tallimustine-dependent, IISRE cleavage-resistant population was apparent (compare lanes −/F and +/F). Subsequent REPSA rounds were performed to enrich for these tallimustine-binding DNAs, and by round 7 they were a majority of the DNAs present (compare lanes −/− and +/B). It should be noted that, after round 5, an increasing percentage of the DNAs obtained in the −/F and −/B reactions electrophoresed with an apparent mobility of 49 bp (T). These were found to be composed of DNAs lacking the central 14-bp random region (data not shown). The origin of this species is as yet unknown but is not the result of a tallimustine-dependent process since it was also observed in independent REPSA experiments lacking tallimustine (data not shown). In addition, while we have previously reported that *Fok*I was considerably better than other IISREs for REPSA selection (*16*, *22*, *23*), in our present experiment we found very little difference in the selection efficiencies obtained with either *Fok*I or *Bpm*I.

REPSA-Selected DNAs Contain Two Types of Consensus Tallimustine−DNA Binding Sites. After subcloning and purification of the pUC19 plasmid clones containing the tallimustine-selected ST-2 templates, the insert DNAs of 54 randomly selected clones were sequenced by the Sanger dideoxynucleotide sequencing method. Of these clones, nine contained mixed 14-bp sequences, indicative of the cloning of partially annealed ST-2 templates. These "bubble" DNAs contain ST-2 strands that are noncomplementary in their randomized regions. Their formation is promoted under PCR conditions having insufficient amplimers present (e.g., see Figure 3, round 3, lane −/−). An additional seven clones did not have full-length randomized regions, ranging instead from zero to 13 bp (median value, 11 bp). Finally, two of the remaining clones contained *Fok*I recognition sequences (5′-GGATG-3′/5′-CATCC-3′) within the 14-bp region. None contained the longer *Bpm*I recognition sequences (5′-CTGGAG-3′/5′-CTCCAG-3′). Thus only 36 of the sequenced clones were considered suitable for further analysis. Their sequences are reported in Table 1.

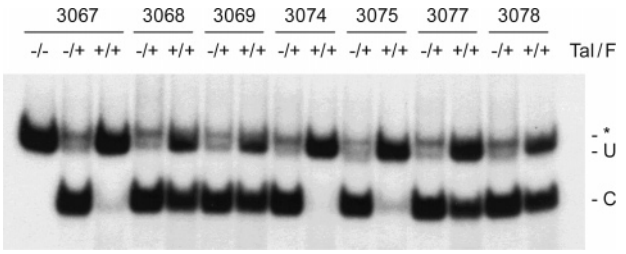

FIGURE 4: Characterization of REPSA-selected clones by REPA. Labeled DNA probes (63 bp) generated by PCR from different tallimustine REPSA-selected clones were incubated without (−) or with (+) 50 μM tallimustine before digestion with *Fok*I. An autoradiogram of their PAGE-resolved products, both uncleaved (U) and cleaved (C), is shown. An asterisk (*) indicates a PCR artifact. Undigested probe control (−/−).

Analysis of the nucleotide ratios in these sequences revealed no significant differences between those of the ST-2 starting material (27% A, 18% C, 16% G, and 39% T) and those of the tallimustine-selected DNAs (26% A, 20% C, 19% G, and 35% T) (*17*). Such similarities indicated that, under the binding conditions used in our REPSA selection, tallimustine did not merely bind to generally AT-rich sequences, which are usually favored by noncovalent DNA-binding *N*-methylpyrrole polyamides such as netropsin and distamycin (*25*). Likewise, using the minimal degenerate recognition sequence for these polyamides, $(A/T)_4$, we found that these AT-rich sequences were significantly underrepresented in the tallimustine-selected sequences than would be expected by chance (37 of 396 possible 4-bp sites, Fisher's exact test, two-sided $P = 0.0001$). Such was not true for some of the highest affinity distamycin binding sites, 5′-AAAAAA-3′/5′-TTTTTT-3′ and 5′-AAATTT-3′, which were comparable in frequency to the number expected by chance (4 of 324 possible 6-bp sites, $P = 0.69$). Interestingly, the previously identified tallimustine consensus sequences, 5′-TTTTGPu-3′/5′-PyCAAAA-3′ and 5′-TTTTAA-3′/5′-TTAAAA-3′, were not observed in our REPSA-selected sequences, though this is not significantly different than the frequency expected by chance (0 out of 324 possible 6-bp sites, $P = 0.50$). Taken together, our data suggest that sequences other than those normally recognized by the noncovalent binding portion of tallimustine or the previously defined tallimustine consensus sequences were preferentially selected under our REPSA conditions.

REPSA-Selected DNAs Contain High-Affinity Tallimustine−DNA Binding Sites. To identify the REPSA-selected clones containing high-affinity tallimustine binding sites, radiolabeled probes carrying the inserts were generated from the above clones by PCR amplification. These DNA probes were then screened for *Fok*I cleavage protection following incubation with 100 μM tallimustine at 37 °C for 5 h. Unfortunately, we observed complete protection of all probe DNAs under these conditions whenever tallimustine was present (data not shown). To distinguish among these sequences, even more stringent binding conditions were employed (50 μM tallimustine, 37 °C, 15 min). Note that these conditions are more rigorous than those previously reported in the literature (*12*) and those used in our REPSA selections (100 μM tallimustine, 37 °C, 16 h). Afterward, the products of the *Fok*I cleavage reaction were resolved by nondenaturing PAGE and visualized by autoradiography. Figure 4 shows a representative REPA on seven tallimustine-selected clones. Note that several probes demonstrated more than 80% cleavage protection under these conditions. REPSA-selected clones that exhibited this higher level of cleavage protection are indicated by ++ in Table 1.

Having identified those clones containing the highest affinity tallimustine binding sites, it was then possible to statistically determine consensus tallimustine binding sequences. Sequences corresponding to high affinity distamycin binding sites (5′-AAAAAA-3′/5′-TTTTTT-3′ and 5′-AAATTT-3′) were present in this population. However, their prevalence was not considered statistically significant with regard to either the percentage of clones involved (4 of 16, $P = 0.17$) or the total number of 6-bp sites present (4 of 144, $P = 0.37$). Interestingly, each of these distamycin-like sites was flanked by a 3′ C after its run of pyrimidines, suggesting that this terminal C was part of the consensus. Otherwise, the only other 6-bp sequences present in statistically significant numbers in the high affinity tallimustine-binding population were members of the complementary pair 5′-GTTCTA-3′/5′-TAGAAC-3′. Remarkably, these sequences were present in 12 of 16 clones ($P < 0.0001$) and in 12 of 144 6-bp sites ($P = 0.0004$). Both prevalences are considered extremely statistically significant. Interestingly, each of the identified high-affinity tallimustine-binding clones contained only one distamycin-like or novel tallimustine sequence, while none of the low affinity clones had either of these sequences. Thus we consider the distamycin-like 7-bp sequences 5′-GAAAAAA-3′/5′-TTTTTTC-3′ and 5′-AAATTTC-3′/5′-GAAATTT-3′ and the novel 6-bp sequence 5′-GTTCTA-3′/5′-TAGAAC-3′ to be the two types of consensus binding sequences for tallimustine.

Identification of High-Affinity Tallimustine−DNA Binding Sites by DNase I Footprinting. To determine the exact binding sites for tallimustine, DNase I cleavage protection assays (footprinting) were carried out on four high-affinity tallimustine-binding clones, two from each type of consensus binding sequence. Singly radiolabeled DNA fragments of the clones were generated by 3′ end-filling *Hin*dIII-cut plasmids, subsequent *Eco*RI digestion, and fragment purification. These were incubated in the absence or presence of 50 μM tallimustine for 15 min at 37 °C. Afterward, noncovalently bound tallimustine was removed by exhaustive 1-butanol extraction before the DNA was subjected to limiting DNase I cleavage. The cleavage products were subjected to high-resolution denaturing gel electrophoresis and visualized by autoradiography. Cleavage products from purine-specific chemical sequencing reactions were run in parallel as sequence markers. Since many of these sequences contained runs of consecutive As and Ts, which are not efficiently cleaved by DNase I, a relatively overexposed autoradiogram (Figure 5A) was used to determine the locations of the footprints. Their positions on the sequences of these four clones are shown in Figure 5B. As seen in these examples, DNase I cleavage protection corresponded with the expected tallimustine consensus sites present within the 14-bp region. Those protections occurring with type 1 consensus sequences (e.g., 3074 and 3078) were typically 7 to 8 bp long, centered on the consensus sequence, with a 1-bp offset toward the 3′ side as expected for right-handed helical B-form DNA (*26*). Protections covering type 2 consensus sequences (e.g., 3043 and 3046) were generally much larger (12 to 13 bp), but still centered on the consensus

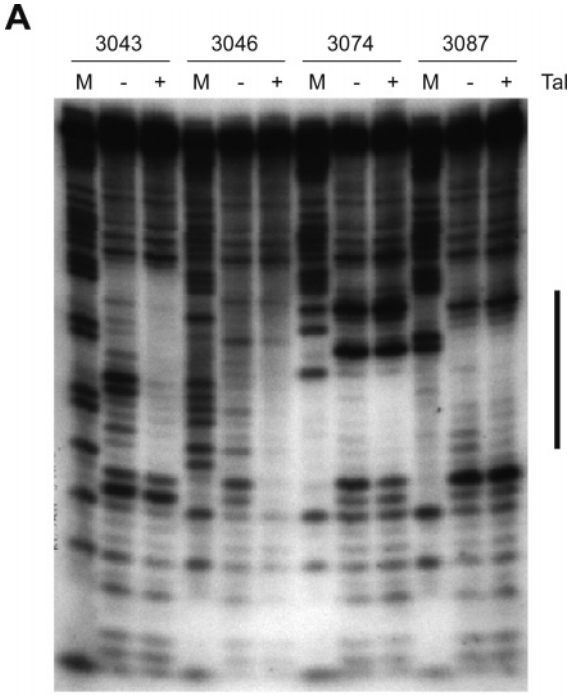

FIGURE 5: Identification of tallimustine binding sites by DNase I cleavage protection and high-resolution denaturing gel electrophoresis. (A) Autoradiogram of the resolved DNA fragments obtained after incubating singly 3′ end-labeled restriction fragment probes from the above indicated clones without (−) or with (+) 50 μM tallimustine and limiting DNase I cleavage. Purine-specific chemical sequencing reaction products (M) served as electrophoretic DNA length markers. The location of the randomized 14-bp region is indicated at the right. (B) Tallimustine footprints for each clone are indicated by a line above the sequence (shown 5′→3′). Sequences present in the 14-bp randomized region are indicated in uppercase letters, while sequences present in the flanking regions are indicated in lowercase letters. Italic and boldface fonts indicate type 1 and type 2 tallimustine consensus sequences, respectively.

sequence with a slight 1- to 2-bp 3′ offset. No tallimustine footprints were observed in either flanking region, indicative of the sequence specificity of tallimustine binding under these conditions.

Identification and Characterization of Covalent Tallimustine Binding Sites. While DNase I footprinting can identify DNA binding sites for ligands such as tallimustine, this method cannot precisely determine the exact site of the covalent adduct or identify the nature of the chemical bond. To determine this information, we used chemical agents that break the phosphodiester backbone of DNA at the site of modified nucleic acid bases. For example, the imidazole ring of alkylated N7-guanine is susceptible to opening at elevated pH. Heating DNA with the nucleophile piperidine converts these alkylated guanine sites into strand breaks. Likewise, N3 alkylation renders adenines and guanines labile at elevated temperature, with DNA cleavage then occurring at the abasic sites. Both these methods have been used previously to identify sites of covalent modification of DNA by various small molecule ligands (*20*, *21*).

Radiolabeled DNA probes were generated by PCR using the 5′ ^{32}P-63AR and 63AL primers. These probes were incubated with 6.25 μM tallimustine in either our standard binding buffer (buffer A: 20 mM Tris-acetate, pH 7.9, 50 mM K-acetate, 10 mM Mg-acetate, and 0.05% Nonidet P-40) or an alternative binding buffer (Buffer E: 20 mM K-phosphate, pH 6) for 16 h at 37 °C to allow covalent binding to occur. Noncovalently bound tallimustine was removed by extraction with 1-butanol. To identify nucleophile-labile DNA adducts under alkaline conditions, tallimustine-treated DNA was resuspended in 1 M piperidine and heated for 20 min at 95 °C. To identify heat-labile adducts under low ionic strength, neutral pH conditions, tallimustine-treated DNA was resuspended in buffer D (1.5 mM Na-citrate, pH 7.2, 15 mM NaCl) and heated for 30 min at 90 °C. Following 1-butanol precipitation, the cleavage products were subjected to high-resolution denaturing gel electrophoresis and visualized by autoradiography. Cleavage products from purine-specific chemical sequencing reactions were run in parallel as sequence markers.

Representative autoradiograms for experiments performed with probes containing type 1 or type 2 consensus sequences are shown in Figure 6. As seen in these examples, substantially more tallimustine-dependent cleavage products were observed when binding reactions were performed in a phosphate buffer than were observed when our standard Tris buffer was used. These bands were more apparent when piperidine was used to cleave tallimustine-modified DNA than when cleavages were performed in a neutral pH, low ionic strength buffer (SSC). Interestingly, the pattern observed with SSC cleavage was identical to that observed with piperidine, except that the relative mobilities of the SSC cleavage products were two nucleotides slower. Note that different clones had different cleavage patterns, and many of these cleavages were proximal to the consensus tallimustine binding sequence present in each clone. An intrinsic tallimustine cleavage site was also found in the flanking sequences (AR) of these clones, with different extents of cleavage evident for different clones. Significant cleavage only occurred when incubations were performed in the lower pH phosphate buffer; no tallimustine-dependent cleavages at this intrinsic site were observed with our standard basic pH Tris buffer. In addition to the faster mobility cleavage products, slower mobility species (+) were also observed in these autoradiograms. Their presence was tallimustine and clone specific, though their appearance was not always completely reproducible (compare Figures 6B and 7).

The locations of piperidine-induced cleavages for each clone are indicated in Figure 6E. As shown by these data, cleavages occurred predominantly at guanines and to a lesser extent at adenines. For type 1 tallimustine consensus sequences (clones 3015 and 3087), proximal cleavages occurred at guanines within the consensus sequence (5′-GAAAAAA-3′) and immediately upstream. For type 2 tallimustine consensus sequences (clone 3043), proximal cleavages were evident at guanines downstream of the consensus sequence (5′-TAGAAC-3′). The intrinsic tallimustine cleavage site within the flanking sequence occurred

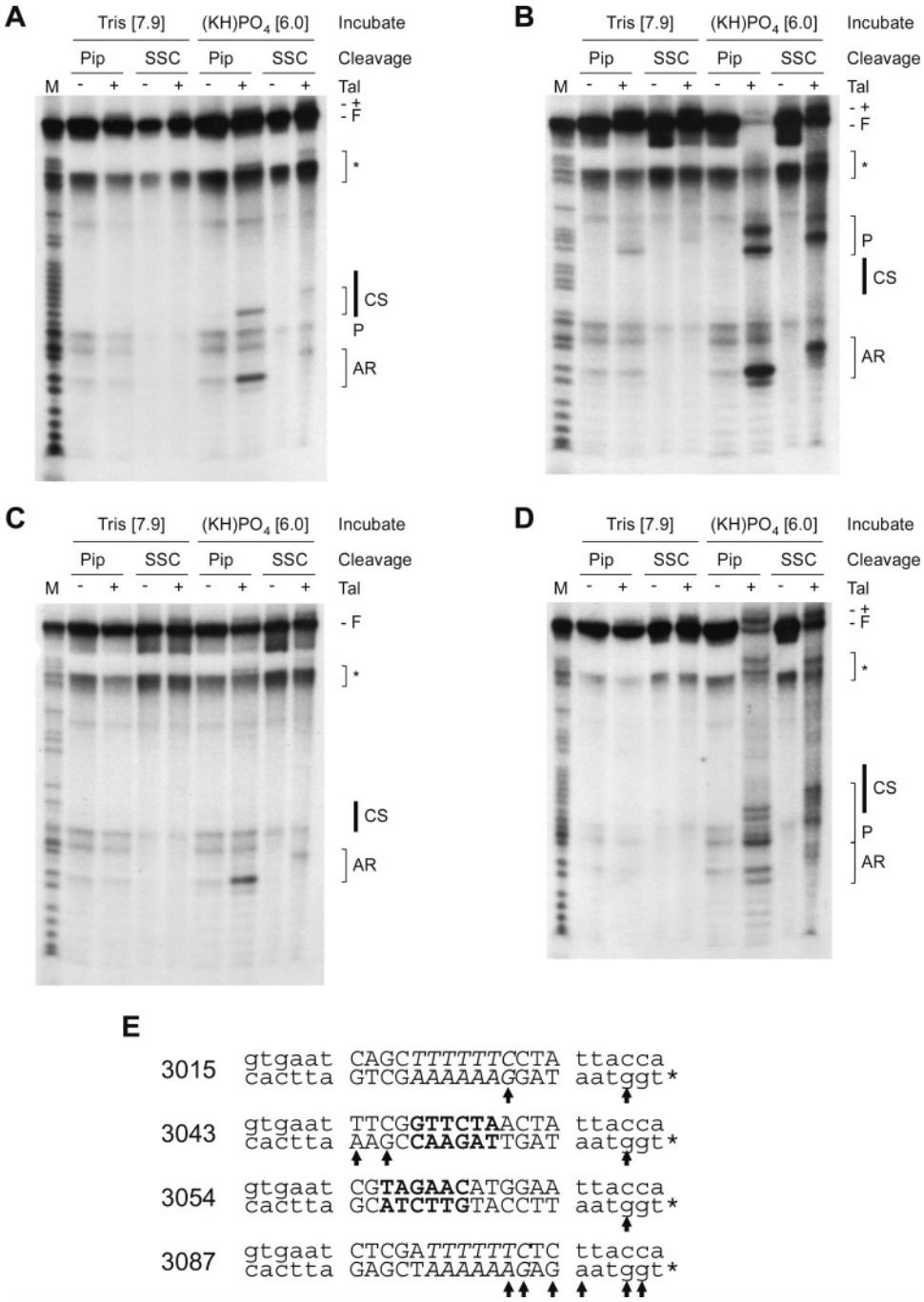

FIGURE 6: Identification of tallimustine covalent adducts by chemical probes and high-resolution denaturing gel electrophoresis. Singly 5′ end-labeled PCR probes were incubated at 37 °C for 16 h in either a Tris-HCl or potassium phosphate buffer, without (−) or with (+) 6.25 μM tallimustine, and then subjected to thermal cleavage in the presence of either a sodium citrate/sodium chloride (SSC) or piperidine (Pip) buffer. Shown are the autoradiograms obtained with probes from the clones (A) 3015, (B) 3043, (C) 3054, and (D) 3087. Purine-specific chemical sequencing reaction products (M) served as electrophoretic DNA length markers. The locations of the full length denatured probe (F), a slower mobility species (+), undenatured duplex probe (*), a tallimustine consensus sequence (CS), and cleavage sites within the randomized (P) and flanking regions (AR) are indicated at the right of each panel. (E) Sites of tallimustine covalent adducts on each clone are indicated by an arrowhead below each partial sequence of the 5′ end-labeled strand (*). Sequences present in the 14-bp randomized region are indicated in uppercase letters, while sequences present in the flanking regions are indicated in lowercase letters. Italic and boldface fonts indicate type 1 and type 2 tallimustine consensus sequences, respectively.

primarily at a unique guanine within the sequence 5′-GGGATGGTAA-3′ (G22, underlined) and to a lesser extent at an adjacent guanine on some clones. Since SSC visualized cleavages occurred two nucleotides further from the 5′ end label than those found by piperidine treatment, SSC cleavages occurred predominantly at adenines, and to a lesser extent at thymines and guanines.

To better understand the chemistry behind tallimustine−DNA covalent adduct formation, we investigated different physical and chemical parameters that could affect this process. As shown in Figure 7A, piperidine-induced cleavages were tallimustine concentration dependent, with cleavages first notable at 0.12 μM tallimustine for the proximal site and 0.5 μM for the AR flanking site. These data indicated

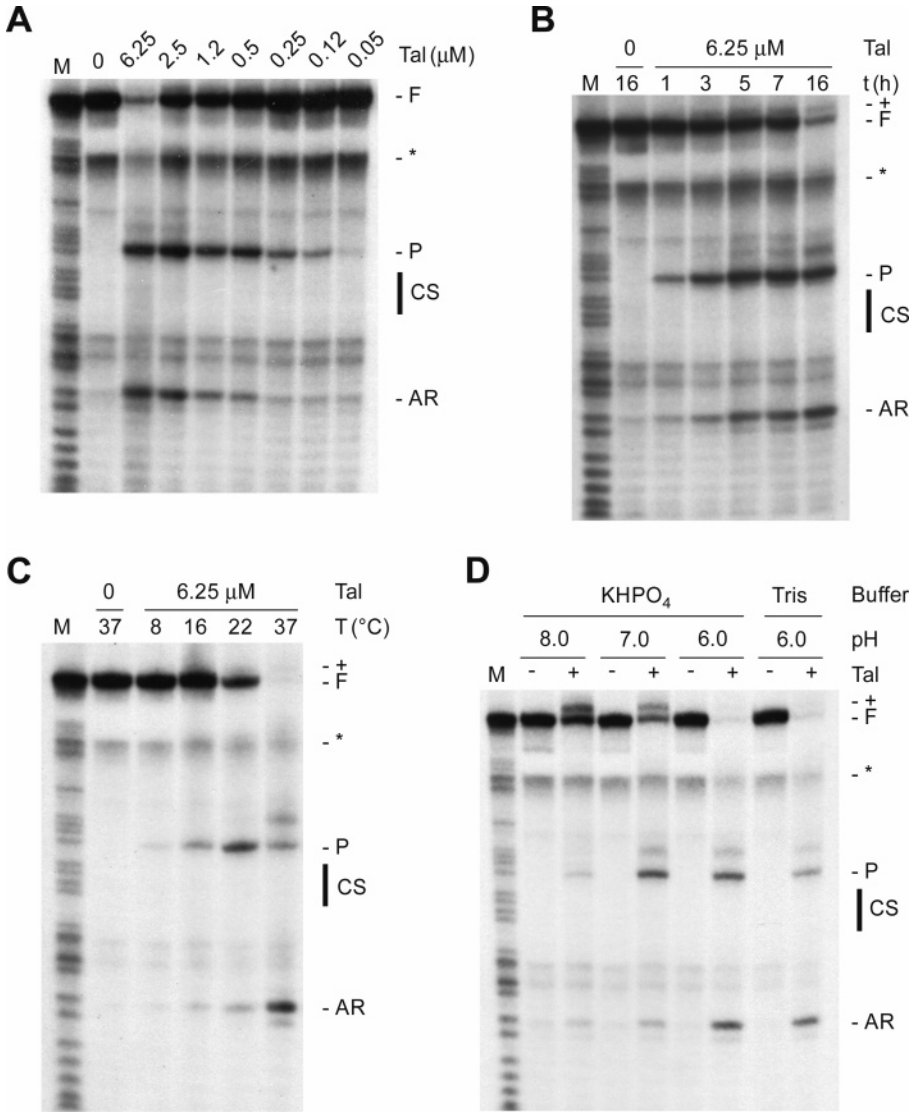

FIGURE 7: Characterization of physical parameters affecting tallimustine covalent adduct formation. A singly 5′ end-labeled 3043 PCR probe was incubated under our standard reaction conditions (i.e., 6.25 μM tallimustine, potassium phosphate buffer, pH 6.0, 16-h incubation at 22 °C), with the following exceptions: (A) different tallimustine concentrations, (B) different time periods, (C) different temperatures, and (D) different pHs, before treatment with piperidine and high-resolution denaturing gel electrophoresis. Purine-specific chemical sequencing reaction products (M) served as electrophoretic DNA length markers. The locations of the full length denatured probe (F), a slower mobility species (+), undenatured duplex probe (*), a tallimustine consensus sequence (CS), and cleavage sites within the randomized (P) and flanking regions (AR) are indicated at the right of each panel.

that the consensus sequence proximal site in clone 3043 is a preferred site for forming piperidine-evident, tallimustine covalent adducts. At 6.25 μM tallimustine, almost all DNA has at least one cleavage, at either the proximal or AR sites. Note that in this experiment only a single proximal cleavage product (5′-TAGAACCGAA-3′, underlined) was observed and no slower mobility products were found (compare with Figure 6B, lane 7). Tallimustine-induced cleavages were time dependent (Figure 7B), with appreciable cleavage occurring at the proximal site within a 1 h incubation. Comparable reaction kinetics were found with the AR site, though the extent of cleavage was less than those found for the proximal site. Tallimustine-induced cleavages were also temperature dependent (Figure 7C), with appreciable cleavages occurring at both sites at 16 °C. More cleavage was consistently observed at the proximal site, except in the case of incubation at 37 °C, which apparently was the result of multiple cleavages on nearly all probe DNA. Tallimustine-induced cleavages were also pH-dependent, with maximal cleavage occurring at more acidic pH (Figure 7D). Note that this phenomenon was not dependent on buffer composition, as both phosphate and Tris-based buffers yielded comparable cleavage amounts at pH 6.0. Interestingly, significant tallimustine-dependent lower mobility species were found under high pH conditions. This could reflect the minimal tallimustine-dependent cleavage that occurs under these conditions, which when present would normally eliminate this species.

Our cleavage data suggested that guanines might be the primary DNA bases that react with tallimustine (Figure 6E). To determine at which site on guanine tallimustine was reacting, we performed experiments with DNA probes that contained the guanine analogues inosine (I) or 7-deazaguanine (^{7deaza}G). Inosine lacks the exocyclic 2-amino group normally present in the DNA minor groove, while 7-deazaguanine lacks the major groove accessible heterocyclic 7-amine favored by most alkylators. As before, probes were

synthesized by PCR using a radiolabeled 63AR amplimer. However, in these reactions the nucleotide GTP was substituted with either ITP or ^{7deaza}GTP. This resulted in a probe containing uniform guanine substitutions on both DNA strands throughout the randomized cassette region, but with substitutions on only the amplimer complementary strands within the flanking sequences. Note that the presence of guanines within the flanking sequences served as internal controls for tallimustine-induced cleavages. As shown in Figure 8, cleavages at the proximal sites were enhanced on the inosine-substituted probe relative to the AR site when compared with a guanine-containing probe (Figure 6B). However, cleavages at the proximal sites were completely absent on the 7-deazaguanine-substituted probe. As observed with guanine-containing probes, qualitatively similar results were obtained with both piperidine and SSC-treated probes, with the SSC cleavage pattern again being shifted two nucleotides further from the 5′ labeled end. In addition, while not easily visible in the darker exposure with the 3043 inosine-substituted probe, both the inosine- and 7-deazaguanine-substituted probes demonstrated the appearance of slower mobility species when reacted with tallimustine.

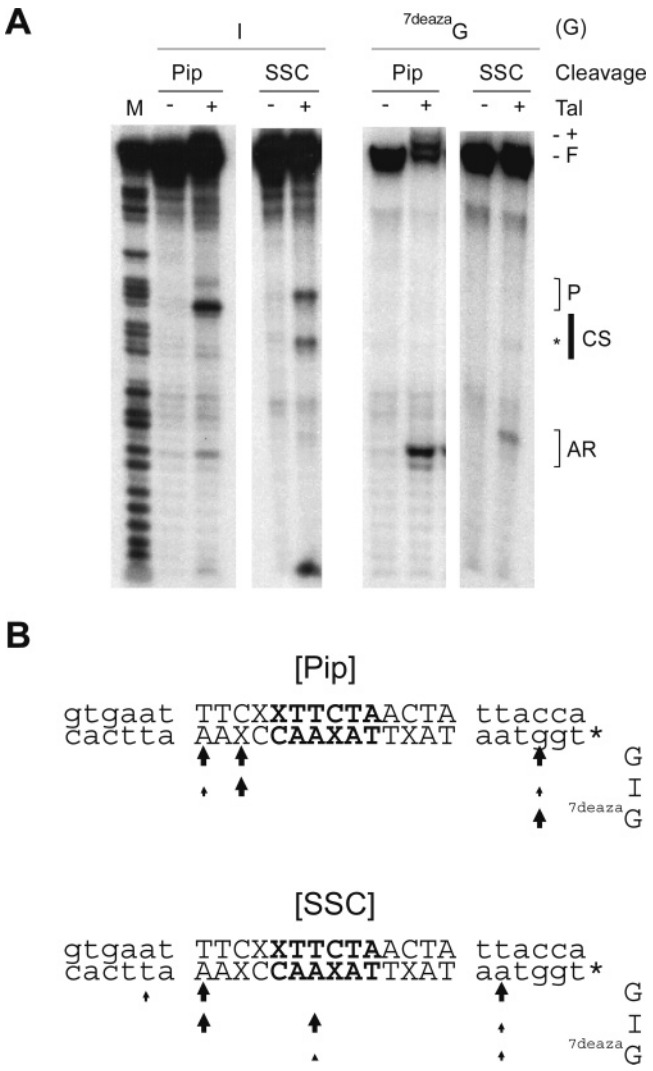

FIGURE 8: Identification of the guanine substituents responsible for tallimustine covalent adduct formation. Singly 5′ end-labeled 3043 PCR probes substituted with either inosine or 7-deazaguanine were incubated under our standard reaction conditions (i.e., 6.25 μM tallimustine, potassium phosphate buffer, pH 6.0, 16-h incubation at 22 °C) and the sites of tallimustine adduct formation determined by either piperidine or SSC treatment and high-resolution denaturing gel electrophoresis. (A) Autoradiograms of labeled DNA cleavage products. Purine-specific chemical sequencing reaction products (M) served as electrophoretic DNA length markers. The locations of the full length denatured probe (F), a slower mobility species (+), a tallimustine consensus sequence (CS), and cleavage sites within the randomized (P) and flanking regions (AR) are indicated at the right of each panel. A novel SSC cleavage site within the consensus sequence is also indicated (*). (B) Sites of tallimustine-dependent cleavages induced by piperidine (top) or SSC (bottom) treatment. Sequences present in the 14-bp randomized region are indicated in uppercase letters, while sequences present in the flanking regions are indicated in lowercase letters. Boldface fonts indicate a type 2 tallimustine consensus sequence. Sites of inosine or 7-deazaguanine substitution (X). Arrowheads indicate sites of cleavage, with arrow size reflecting relative cleavage extent.

DISCUSSION

Our laboratory has developed a combinatorial selection method, REPSA, which uses enzymatic rather than physical means to identify duplex DNA sequences specifically bound by a variety of ligands. Of the different combinatorial methods described to date, REPSA is unique in that it can be applied to native small molecule ligands under physiological conditions. We have successfully used REPSA to identify the preferred binding sites of a variety of DNA-binding small molecules, including distamycin A, actinomycin D, and the hairpin polyamides ImPyPyPy-γ-PyPyPyPy-β-Dp and ImPyPyPy-γ-ImPyPyPy-β-Dp (*22*−*24*). However, to the best of our knowledge, this is the first report of the application of combinatorial selection methods to identify the preferred binding sites of any covalent binding ligand bound to DNA.

Using a modification of REPSA, we surveyed over 134 million different 14-bp sequences and found a subset of these that significantly inhibited IISRE cleavage when tallimustine was present. Sequence analysis of these selected clones identified two types of consensus sequences: a minor variety (type 1, 5′-GAAAAAA-3′/5′-TTTTTTC-3′ and 5′-AAATTTC-3′/5′-GAAATTT-3′) and a major variety (type 2, 5′-GTTCTA-3′/5′-TAGAAC-3′). Type 1 tallimustine consensus sequences are quite similar to the high affinity binding sites of distamycin, with the addition of a terminal G·C base pair. Preferred binding to such sequences is quite reasonable and suggests that the additional G·C base pair may play an important role in tallimustine recognition. However, the type 2 tallimustine consensus sequence is quite different from the exclusively AT-rich sequences normally preferred by distamycin. Distamycin has been reported to preferentially interact in a 2:1 complex with AT-rich sequences containing a single G·C base pair, exhibiting modest equilibrium binding constants but extremely slow dissociation kinetics (*27*, *28*). This latter characteristic might greatly affect the selectability of DNA sequences by REPSA, especially since organic extractions only remove free ligands and not ligands noncovalently bound to DNA. Also it should be noted that our selected DNA sequences did not include any corresponding to the previously identified consensus sequences, 5′-TTTTGPu-3′/5′-PyCAAAA-3′ and 5′-TTTTAA-3′/5′-TTAAAA-3′, nor did these earlier investigators identify sequences corresponding to either our type 1 or type 2 consensus sequences (*12*, *29*). The latter observation is understandable, given the technology used and the limited variety of sequences

investigated. However, that REPSA did not select for these earlier consensus sequences suggests that these sequences may not be the preferential binding sites for tallimustine, at least under our experimental conditions.

As determined by DNase I footprinting, tallimustine formed specific complexes with both type 1 and type 2 consensus sequences that were resistant to exhaustive *n*-butanol extraction. Such data were consistent with, but did not prove, the formation of covalent tallimustine−DNA complexes. Attempts to identify the exact sites of tallimustine adduct formation by a *Taq* polymerase stop assay were not highly successful in our hands (data not shown). Thus we used thermal cleavage assays with either piperidine, which exposes alkylations at the N7 position of guanine, or SSC buffer, which exposes alkylations at the N3 positions of adenines and guanines, to identify sites of tallimustine adduct formation. Investigating several DNA probes, we found that tallimustine generally promoted piperidine-induced cleavages at guanines located at the terminus of the type 1 consensus sequence or located within two base pairs downstream of a type 2 consensus sequence. Probes containing type 1 consensus sequences typically provided the most consistent results, while those containing type 2 consensus sequences varied from no detectable proximal cleavages to a minority demonstrating intense cleavages. For both types of consensus sequences, SSC treatment caused the appearance of a similar pattern of cleavages as found with piperidine treatment, except that the apparent sites of cleavage were shifted two base pairs further from the probe's 5′ labeled end. In most cases the sites of SSC cleavage corresponded to adenines, though sites at guanines and thymines were also found. Thus while it is possible that tallimustine specifically labilized bases at these distal sites, the coincidence between the piperidine and SSC patterns and the knowledge that SSC-induced DNA cleavage products migrate slower than those generated by piperidine (*21*) strongly suggests that these bands actually involve guanine adducts. This contention was also supported by experiments with 7-deazaguanine-substituted DNA, which demonstrated a loss of both piperidine- and SSC-induced consensus sequence-proximal cleavages. This loss of cleavages was not the result of a loss of tallimustine binding, since we were able to observe tallimustine-dependent DNase I footprints on 7-deazaguanine-substituted DNA (data not shown). Taken together, our data strongly suggest that tallimustine primarily forms covalent adducts with guanines in or near its consensus sequence. We also observed the formation of a slower migrating species when DNA was treated with tallimustine. We believe that this species represents a covalent tallimustine complex with the intact probe DNA that is not completely labilized following thermal treatments with piperidine or SSC. Note that we do not believe that it involves adducts with the exocyclic amine at the N2 position of guanine, since similar slower mobility species were observed with inosine-substituted DNA, which lacks this functional group.

Characterization of piperidine-induced cleavages on one of our most reactive sequences (clone 3043) showed concentration, time, temperature, and pH dependencies comparable to those found for DNA alkylation by simple nitrogen mustards such as chlorambucil and melphalan (*30*, *31*). Our data are therefore consistent with the formation of an azaridinium intermediate but do not show any significantly increased reactivity given the presence of the distamycin DNA-binding moiety. Nitrogen mustards preferentially attack clusters of guanines at their N7 position (*31*, *32*). Using DNAs containing the guanine analogues inosine and 7-deazaguanine, we found that only DNAs lacking the nucleophilic N7 site were refractory to tallimustine-dependent cleavages, thus strongly suggesting that tallimustine specifically attacks the N7 position of guanine. Note that our data do not preclude the possibility that tallimustine can also form adducts with the N3 position of either adenine or guanine. In fact, we did find one cleavage within the type 1 consensus site of inosine-substituted 3043 DNA (Figure 8A, asterisk) that was more reactive with SSC-mediated thermal treatment than with piperidine, consistent with a N3 adduct. Such was not observed with 3043 DNA containing guanine (Figure 6B), suggesting that it may be a unique reactivity afforded by inosine-substituted DNA. Note that our results are quite different than those previously reported, which stated that tallimustine preferentially alkylates the N3 position of the 3′ terminal adenines in the consensus sequences 5′-TTTTGA-3′ and 5′-TTTTAA-3′ (*11*, *12*). These investigators also found no evidence for alkylation at the N7 position of guanines. As stated above, part of the reason for these different observations is the different consensus sequences investigated by the different groups. However, it should also be noted that N7 alkylation of guanine has been described for related compounds, though not at the aforementioned consensus sites (*29*, *33*).

In B-form, duplex DNA, the N7 position of guanine is located in the DNA major groove. However, guanine N7 alkylation is difficult to explain in light of the expected rapid binding of tallimustine in the DNA minor groove, as occurs with its parent compound distamycin. To explain our findings, we propose a model in which two tallimustine molecules are involved in the formation of a cleavage event (Figure 9). For a type 1 consensus site, we propose that a single tallimustine molecule binds in the narrow minor groove of an AT-rich sequence with its aniline mustard tail perturbing an adjacent G•C base pair. The perturbed guanine in this base pair then becomes more accessible to attack by a nitrogen mustard present in a second tallimustine. A similar mechanism is thought to occur with a type 2 consensus site, except that the region of perturbed DNA encompasses at least two base pairs. At present we do not know the exact nature of the guanine perturbation, though a placement of the guanine higher in the DNA major groove would be consistent with our model. Note, however, that we do not envision the second tallimustine to interact over a large surface of the DNA, e.g., binding within the DNA major groove. Rather we imagine the second tallimustine molecule to be best represented as a sterically hindered nitrogen mustard that can only efficiently form adducts with exposed guanines. Additionally, involvement of two tallimustine molecules in our model would seem quite reasonable given the vast molar excess of tallimustine to the target DNA used in our and others' experiments. Our model is supported by preliminary structural data, which find a single tallimustine molecule bound in the minor groove of a type 2 consensus sequence, with no evidence of adduct formation when equimolar concentrations of tallimustine and DNA are present (Kumaralal Kaluarachchi, unpublished observations).

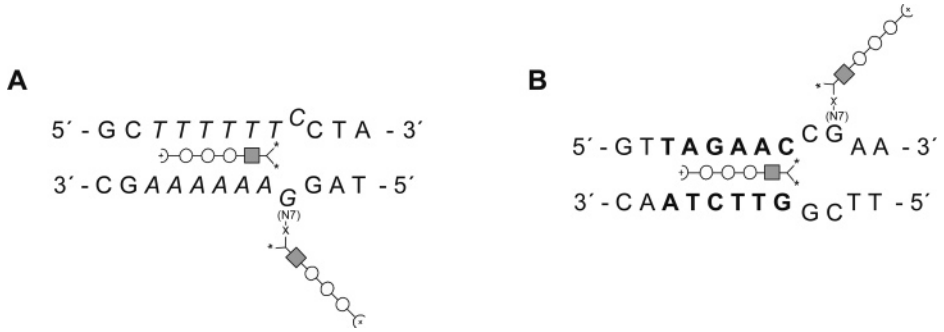

FIGURE 9: Schematic representations of tallimustine binding to (A) type 1 and (B) type 2 REPSA-identified consensus sequences. Polyamides situated between complementary DNA sequences are considered binding in the minor groove, while those above and below these sequences are considered to act on another side of the DNA. DNA bases situated outside the linear sequences are considered to have a structure significantly different from B-form duplex DNA. Polyamide subunits include 3-amino-propionamidine (⊃), 4-amino-1-methyl-1*H*-pyrrole-2-carboxylic acid (○), and 4-[bis(2-chloroethyl)amino]benzoic acid (■<). Reacted (X) and unreacted (*) 2-chloroethylamines, as well as the site of covalent attachment (e.g., N7 position on guanine), are indicated.

What is the significance of the intrinsic tallimustine adduct site situated in the AR flanking region? Many of REPSA-selected sequences that exhibited high tallimustine-dependent REPA protection also demonstrated cleavages at G22 following piperidine treatment (Figure 6), though this was not universally true (e.g., clones 3046 and 3047, data not shown). Under some circumstances and for certain sequences these cleavages were more pronounced than those observed at the sites proximal to the type 1 and type 2 consensus sequences. We do not know exactly what parameters dictate tallimustine adduction at G22, but have found no obvious correlation with the presence or the location of a tallimustine consensus site within the randomized region, or with any particular sequence present in the randomized region immediately adjacent to the AR flank. In fact, G22 appears to have a slight propensity to be cleaved by piperidine treatment even in the absence of tallimustine, suggestive of an intrinsically labile base. We do not believe significant adduct formation occurred at G22 during our REPSA selection rounds, since very little adduct formation occurs under our standard buffer conditions (i.e., incubation in pH 7.9 buffer). However, we cannot exclude the possibility that tallimustine binding in the AR flanking region might have had some effect on REPSA selection. Such a phenomenon could help explain the universal IISRE cleavage inhibition observed in our initial REPA characterization of the REPSA-selected clones using high tallimustine concentrations.

While some of the most widely used antineoplastic agents are those that alkylate DNA, they often exhibit a relatively low therapeutic index given their propensity to damage DNA nonspecifically in both neoplastic and normal proliferative cell types. Efforts to improve the efficacy of these drugs led to the discovery and development of bifunctional molecules possessing sequence-specific, noncovalent DNA binding and covalent DNA-modifying moieties. While these molecules have demonstrated increased specificity for their reaction with DNA, they have yet to show great clinical promise (*4*). Perhaps one limitation of these molecules is their relatively promiscuous reactivity in the context of an entire human genome. Ideally such DNA-targeted therapeutics should only react with unique disease-related sequences, e.g., the integrated HIV-1 genome in AIDS or the BCR/ABL translocation in chronic myelogenous leukemia. Bifunctional molecules based on hairpin polyamides have the potential of delivering DNA reactive moieties to almost any targeted sequence with very high specificity (*34*). It remains to be determined whether such molecules will eventually lead to improved treatments for chronic diseases such as cancer and AIDS.

ACKNOWLEDGMENT

We thank Dr. G. L. Devani at Pharmacia & Upjohn (Milan) for the generous gift of tallimustine, and Vashisht Gopal Yennu Nanda and Kumaralal Kaluarachchi for critical reading of the manuscript.

REFERENCES

1. Hurley, L. H., and Boyd, F. L. (1988) DNA as a Target for Drug Action, *Trends Pharmacol. Sci. 9*, 402−407.
2. Yang, X. L., and Wang, A. H. (1999) Structural Studies of Atom-Specific Anticancer Drugs Acting on DNA, *Pharmacol. Therapeut. 83*, 181−215.
3. Reddy, B. S. P., Sharma, S. K., and Lown, J. W. (2001) Recent Developments in Sequence Selective Minor Groove DNA Effectors, *Curr. Med. Chem. 8*, 475−508.
4. Marchini, S., Broggini, M., Sessa, C., and D'Incalci, M. (2001) Development of Distamycin-Related DNA Binding Anticancer Drugs, *Expert Opin. Invest. Drugs 10*, 1703−1714.
5. Sessa, C., Pagani, O., Zurlo, M. G., de Jong, J., Hofmann, C., Lassus, M., Marrari, P., Strolin-Benedetti, M., and Cavalli, F. (1994) Phase I Study of the Novel Distamycin Derivative Tallimustine (FCE 24517), *Ann. Oncol. 5*, 901−907.
6. Punt, C. J., Humblet, Y., Roca, E., Dirix, L. Y., Wainstein, R., Polli, A., and Corradino, I. (1996) Tallimustine in Advanced Previously Untreated Colorectal Cancer, a Phase II Study, *Br. J. Cancer 73*, 803−804.
7. Viallet, J., Stewart, D., Shepherd, F., Ayoub, J., Cormier, Y., DiPietro, N., and Steward, W. (1996) Tallimustine is Inactive in Patients with Previously Treated Small Cell Lung Cancer. A Phase II Trial of the National Cancer Institute of Canada Clinical Trials Group, *Lung Cancer 15*, 367−373.
8. Beran, M., Jeha, S., O'Brien, S., Estey, E., Vitek, L., Zurlo, M. G., Rios, M. B., Keating, M., and Kantarjian, H. (1997) Tallimustine, an Effective Antileukemic Agent in a Severe Combined Immunodeficient Mouse Model of Adult Myelogenous Leukemia, Induces Remissions in a Phase I Study, *Clin. Cancer Res. 3*, 2377−2384.
9. 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) A Phase I and Pharmacokinetic Study of Tallimustine [PNU 152241 (FCE 24517)] in Patients with Advanced Cancer, *Clin. Cancer Res. 4*, 53−59.
10. Arcamone, F., Animati, F., Barbieri, B., Configliacchi, E., D'Alessio, R., Geroni, C., Giuliani, F., Lazzari, E., Menozzi, M., Mongelli, N., Penco, S., and Verini, M. A. (1989) Synthesis, DNA-Binding Properties, and Antitumor Activity of Novel Distamycin Derivatives, *J. Med. Chem. 32*, 774−778.

11. Broggini, M., Erba, E., Ponti, M., Ballinari, D., Geroni, C., Spreafico, F., and D'Incalci, M. (1991) Selective DNA Interaction of the Novel Distamycin Derivative FCE 24517, *Cancer Res. 51*, 199−204.
12. Broggini, M., Coley, H. M., Mongelli, N., Pesenti, E., Wyatt, M. D., Hartley, J. A., and D'Incalci, M. (1995) DNA Sequence-Specific Adenine Alkylation by the Novel Antitumor Drug Tallimustine (FCE 24517), a Benzoyl Nitrogen Mustard Derivative of Distamycin, *Nucleic Acids Res. 23*, 81−87.
13. Broggini, M., Ponti, M., Ottolenghi, S., D′Incalci, M., Mongelli, N., and Mantovani, R. (1989) Distamycins Inhibit the Binding of OTF-1 and NFE-1 Transfactors to their Conserved DNA Elements, *Nucleic Acids Res. 17*, 1051−1059.
14. Brooks, N., Lee, M., Wright, S. R., Woo, S., Centioni, S., and Hartley, J. A. (1997) Synthesis, DNA Binding, Cytotoxicity and Sequence Specificity of a Series of Imidazole-Containing Analogs of the Benzoic Acid Mustard Distamycin Derivative Tallimustine Containing an Alkylating Group at the C-Terminus, *Anti-Cancer Drug Des. 12*, 591−606.
15. Herzig, M. C. S., Trevino, A. V., Arnett, B., and Woynarowski, J. M. (1999) Tallimustine Lesions in Cellular DNA are AT Sequence-Specific but Not Region-Specific, *Biochemistry 38*, 14045−14055.
16. Hardenbol, P., Wang, J. C., Van Dyke, M. W. (1997) Identification of Preferred hTBP-DNA Binding Sites by the Combinatorial Method REPSA, *Nucleic Acids Res. 25*, 3339−3344.
17. Hardenbol, P., and Van Dyke, M. W. (1996) Sequence Specificity of Triplex DNA Formation: Analysis by a Combinatorial Approach Restriction Endonuclease Protection Selection and Amplification, *Proc. Natl. Acad. Sci. U.S.A. 93*, 2811−2816.
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Cheng, A. J., and Van Dyke, M. W. (1994) Oligodeoxyribonucleotide Length and Sequence Effects on Intermolecular Purine-Purine-Pyrimidine Triple-Helix Formation, *Nucleic Acids Res. 22*, 4742−4747.
20. Mattes, W. B., Hartley, J. A., and Kohn, K. W. (1986) DNA Sequence Selectivity of Guanine-N7 Alkylation by Nitrogen Mustards, *Nucleic Acids Res. 14*, 2971−2987.
21. Reynolds, V. L. Molineux, I. J., Kaplan, D. J., Swenson, D. H., and Hurley, L. H. (1985) Reaction of the Antitumor Antibiotic CC-1065 with DNA Location of the Site of Thermally Induced Strand Breakage and Analysis of DNA Sequence Specificity, *Biochemistry 24*, 6228−6237.
22. Hardenbol, P., Wang, J. C., and Van Dyke, M. W. (1997) Identification of Preferred Distamycin-DNA Binding Sites by the Combinatorial Method REPSA, *Bioconjugate Chem. 8*, 617−620.
23. Shen, J., Wang, J. C., and Van Dyke, M. W. (2001) Identification of Preferred Actinomycin-DNA Binding Sites by the Combinatorial Method REPSA, *Bioorg. Med. Chem. 9*, 2285−2293.
24. Vashisht Gopal, Y. N., and Van Dyke, M. W. (2003) Combinatorial Determination of Sequence Specificity for Nanomolar DNA-Binding Hairpin Polyamides, *Biochemistry 42*, 6891−6903.
25. Abu-Daya, A., Brown, P. M., and Fox, K. R. (1995) DNA Sequence Preferences of Several AT-Selective Minor Groove Binding Ligands, *Nucleic Acids Res. 23*, 3385−3392.
26. Van Dyke, M. W., and Dervan, P. B. (1983) Methidiumpropyl-EDTA·Fe(II) and DNase I Footprinting Report Different Small Molecule Binding Site Sizes on DNA, *Nucleic Acids Res. 11*, 5555−5567.
27. Chen, F. M., and Sha, F. (1998) Circular Dichroic and Kinetic Differentiation of DNA Binding Modes of Distamycin, *Biochemistry 37*, 11143−11151.
28. Baliga, R., and Crothers, D. M. (2000) On the Kinetics of Distamycin Binding to its Target Sites on Duplex DNA, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7814−7818.
29. Wyatt, M. D., Lee, M., Gabiras, B. J., Souhami, R. L., and Hartley, J. A. (1995) Sequence Specificity of Alkylation for a Series of Nitrogen Mustard-Containing Analogues of Distamycin of Increasing Binding Site Size: Evidence for Increased Cytotoxicity with Enhanced Sequence Specificity, *Biochemistry 34*, 13034−13041.
30. Mattes, W. B., Hartley, J. A., and Kohn, K. W. (1986) Mechanism of DNA Strand Breakage by Piperidine at Sites of *N*7-Alkylguanines, *Biochim. Biophys. Acta 868*, 71−76.
31. Kohn, K. W., Hartley, J. A., and Mattes, W. B. (1987) Mechanism of DNA Sequence Selective Alkylation of Guanine-N7 Positions by Nitrogen Mustards, *Nucleic Acids Res. 15*, 10531−10549.
32. Sunters, A., Springer, C. J., Bagshawe, K. D., Souhami, R. L., and Hartley, J. A. (1992) The Cytotoxicity, DNA Crosslinking Ability and DNA Sequence Selectivity of the Aniline Mustards Melphalan, Chlorambucil and 4[Bis(2-Chloroethyl)Amino] Benzoic Acid, *Biochem. Pharmacol. 44*, 59−64.
33. Lee, M., Rhodes, A. L., Wyatt, M. D., Forrow, S., and Hartley, J. A. (1993) Design, Synthesis, and Biological Evaluation of DNA Sequence and Minor Groove Selective Alkylating Agents, *Anti-Cancer Drug Des. 8*, 173−192.
34. Wurtz, N. R., and Dervan, P. B. (2000) Sequence Specific Alkylation of DNA by Hairpin Pyrrole-Imidazole Polyamide Conjugates, *Chem. Biol. 7*, 153−161.

BI047877L