Pergamon

Bioorganic & Medicinal Chemistry 9 (2001) 2285–2293

BIOORGANIC & MEDICINAL CHEMISTRY

Identification of Preferred Actinomycin–DNA Binding Sites by the Combinatorial Method REPSA

Jing Shen, Jo C. Wang and Michael W. Van Dyke*

Department of Molecular and Cellular Oncology, Box 79, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA

Received 25 January 2001; accepted 23 March 2001

Abstract—An important question in the study of ligand–DNA interactions is the determination of binding specificity. Here, we used the combinatorial method restriction endonuclease protection, selection, and amplification (REPSA) to identify the preferred duplex DNA-binding sites of the antineoplastic agent actinomycin D. After 10 rounds of REPSA, over 95% of the cloned DNAs exhibited significantly reduced *Fok*I restriction endonuclease cleavage in the presence of 1 μM actinomycin. A χ^2 statistical analysis of their sequences found that 39 of the 45 clones contained one or more copies of the sequence 5′-(T/A)GC(A/T)-3′, giving a $p < 0.001$ for this consensus. A DNase I footprinting analysis of the cloned DNAs found that all possessed relatively high affinity actinomycin-binding sites with apparent dissociation constants between 12 and 258 nM (average 98 nM). The average footprint encompassed 7.6 bases and in most cases (90%) included one or more consensus sequences. Interestingly, several of the selected clones contained overlapping consensus sequences (e.g., 5′-TGCTGCT-3′), suggesting that such close proximity DNA-binding sites may actually be preferred by actinomycin under physiological conditions.

Introduction

Many drugs important in antibiotic chemotherapy avidly bind to double helical DNA with some sequence specificity.[1] Their mechanisms of action, while not completely known, are thought to rely on the strength and selectivity of these interactions. Considerable efforts have been made to identify the preferred binding sites of these small molecules, primarily employing either chemical or enzymatic cleavage protection methods such as MPE or DNase I footprinting.[2,3] While such methods are capable of defining a small molecule binding site with base pair resolution, they are only able to assay 100–200 bp of heterogeneous, native duplex DNA per experiment. This limitation would be inconsequential if a molecule only responded to a small region of DNA (e.g., 3–4 bp), since it would be likely that all possible sequence combinations would be encountered in a reasonable number of experiments. However, phenomena such as co-operative binding,[4,5] flanking sequence effects,[6,7] and the development of small molecules capable of recognizing longer spans of DNA[8,9] have now made it increasingly important to survey several million base pairs in order to identify the best binding sites of these molecules, those presumably recognized at physiological drug concentrations.

Combinatorial methods, those employing large libraries of different molecules and the means of screening them for certain desired properties, have proven especially powerful in the identification of preferred oligomeric receptors for a variety of different ligands. With regard to nucleic acids and the ligands that bind them, combinatorial methods such as in vitro genetics, CASTing and SELEX have been described that have allowed the identification of DNA or RNA sequences that interact with high affinity to many different types of ligands, ranging from small molecules such as nucleotides and organic dyes to macromolecules such as proteins and other nucleic acids.[10–13] Each of these methods utilizes (1) a library of oligonucleotides containing a stretch of randomized nucleotides, (2) a binding equilibration step, whereby those oligonucleotides possessing higher affinity binding sites are preferentially bound by the ligand, (3) a selection step in which ligand-bound oligonucleotides are separated from unbound oligonucleotides, and (4) an amplification step, typically involving PCR, which allows the generation of workable amounts of selected oligonucleotides. Also, since only a very small fraction of the original oligonucleotides contain

*Corresponding author. Tel.: +1-713-792-8954; fax: +1-713-794-0209; e-mail: mishko@odin.mdacc.tmc.edu

0968-0896/01/$ - see front matter
PII: S0968-0896(01)00146-8

high affinity-binding sites and the selection process has only limited efficiency, the cycle of binding, selection, and amplification is often repeated several times until a population of oligonucleotides possessing the desired range of affinities is finally obtained.

Conventional combinatorial methods all rely on the physical separation of ligand–nucleic acid complexes from free oligonucleotides.[10–13] This is typically achieved either as a result of altered chemical properties of the ligand-oligonucleotide complex (e.g., increased mass-to-charge ratio and reduced electrophoretic mobility or increased hydrophobicity and adsorption onto a matrix) or through the use of affinity methods (e.g., ligand-specific antibodies or matrix-immobilized ligands). However, such separation methods are not always amenable for small molecule ligands, which often do not appreciably change the physical properties of the oligonucleotide complex or have altered nucleic acid binding properties when simultaneously bound by a matrix or through an affinity tag. Ideally, what is needed is a combinatorial approach that works with intact, unmodified small molecule ligands and probes their binding specificity under physiological conditions and concentrations.

We have described a combinatorial method, restriction endonuclease protection selection and amplification (**REPSA**), that does not rely on the physical separation of ligand-bound from free DNAs but rather uses the inhibition of an enzymatic process, cleavage by a type IIS restriction endonuclease (**IISRE**), to select for duplex DNAs that are preferentially bound by a variety of ligands.[14] IISREs, unlike the more familiar type II restriction endonucleases such as *Eco*RI, bind to a specific sequence but cleave DNA at a fixed distance from their binding site with no sequence specificity.[15] These characteristics make IISREs the ideal probes for screening different DNA sequences to identify those that are specifically recognized by DNA-binding ligands, so long as the ligand–DNA complex effectively interferes with DNA cleavage by the IISRE. We have described the successful application of REPSA for the identification of preferred nucleic acid-, protein-, and small molecule-binding sites on duplex DNA and have found REPSA to be a general method applicable to most any ligand that interferes with endonuclease cleavage of DNA (e.g., ligands amenable to DNase I footprinting analysis).[14,16,17]

In this paper, we describe the use of REPSA to identify a consensus sequence for the highest affinity duplex DNA-binding sites of the natural product antibiotic actinomycin D. Actinomycin consists of a 2-aminophenoxazin-3-one chromophore to which is appended two pentapeptide lactone moieties having the sequence L-threonine, L-*N*-methyl-valine, sarcosine, L-proline, and D-valine. Actinomycin binds double-helical DNA as a 1:1 complex with the phenoxazinone chromophore intercalated between two base pairs and the cyclic peptide moieties residing in the DNA minor groove on either side of the intercalation site, occupying a total of four base pairs (Fig. 1).[18,19] MPE and DNase I footprinting analyses of actinomycin binding to heterogeneous DNAs found that the drug prefers to intercalate between a 5′-GC-3′ dinucleotide step and protects a site 4–9 bp long.[2,3] However, not all 5′-GC-3′ sequences are bound with equal affinity, since flanking sequences can affect actinomycin binding constants by an order of magnitude.[6,7,20,21] Using REPSA, we found that 5′-(A/T)GC(A/T)-3′ is a consensus binding site for actinomycin, with overlapping multiples of this sequence constituting the highest affinity binding sites for this small molecule ligand.

Results

REPSA selection of actinomycin-binding DNAs

An important first step in a REPSA combinatorial experiment is the design of a selection template. A suitable template requires a central randomized region and two defined flanks, which direct IISRE cleavage to the randomized region. Note that the randomized region needs to be long enough so that it could contain the anticipated highest affinity ligand-binding sites, but not so long that a reasonable quantity of the starting oligonucleotide library would not contain a good representation of all possible sequence combinations. For our REPSA selection of actinomycin-binding DNAs, we chose to use our 63-bp ST2 selection template (Fig. 2), which we had previously successfully used to identify preferred DNA binding sites for the human TATA-binding protein (hTBP).[16] ST2 has nested *Fok*I and *Bpm*I IISRE binding sequences on one flank, positioned so that they direct cleavage to the center of a 14-bp randomized region. This randomized region is more than long enough to simultaneously contain three non-overlapping 4-bp actinomycin-binding sites. Also, since

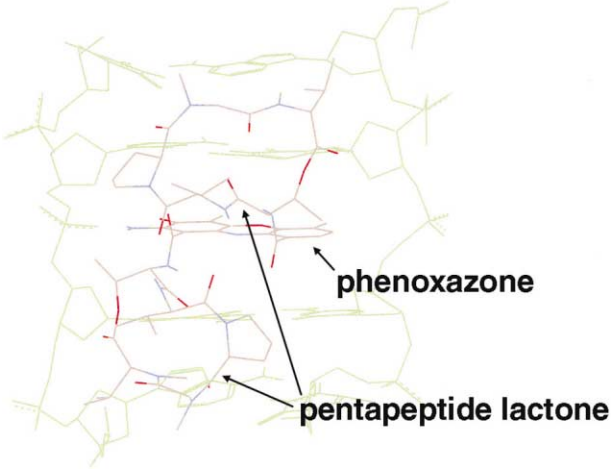

Figure 1. Structure of the DNA-binding antibiotic actinomycin D bound to the complementary tetramer d[AGCT]$_2$. View is looking into the minor groove of duplex DNA. The two DNA strands are shown in light green while a line representation of actinomycin is shown incorporating the standard elemental colors for carbon (gray), nitrogen (blue) and oxygen (red). The locations of the actinomycin phenoxazinone chromophore (intercalated between two G/C base pairs) and its two pentapeptide lactone moieties (residing in the DNA minor groove on either side of the intercalation site) are indicated.

a 14-bp randomized region has only 134 million different possible sequences ($4^{14}/2$, considering the redundancy inherent in duplex DNA), it is highly likely that most all possible sequences would be present in our standard starting REPSA reaction mixture (48 fmol double-stranded being equivalent to 2.9×10^{10} molecules).

A single round of REPSA consists of three steps: (1) ligand-DNA binding, (2) IISRE cleavage, and (3) PCR amplification. These steps are then sequentially repeated until a population of oligonucleotides emerges that demonstrates ligand-dependent IISRE cleavage inhibition. A flow diagram depicting this process is shown in Fig. 3. For our REPSA selection of actinomycin-binding DNAs, 1 μM actinomycin was incubated with 48 fmol of an ST2 library under conditions that allow

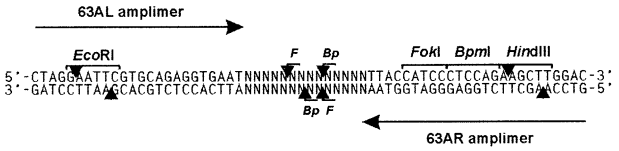

Figure 2. Schematic of the selection template, ST2, used for the identification of preferred actinomycin-DNA binding sites by REPSA. Locations of restriction endonuclease binding (brackets) and cleavage (arrows) sites are indicated. Long horizontal arrows correspond to the sequences of the PCR amplimers. N, random nucleotides. *F*, *Fok*I; *Bp*, *Bpm*I cleavage sites.

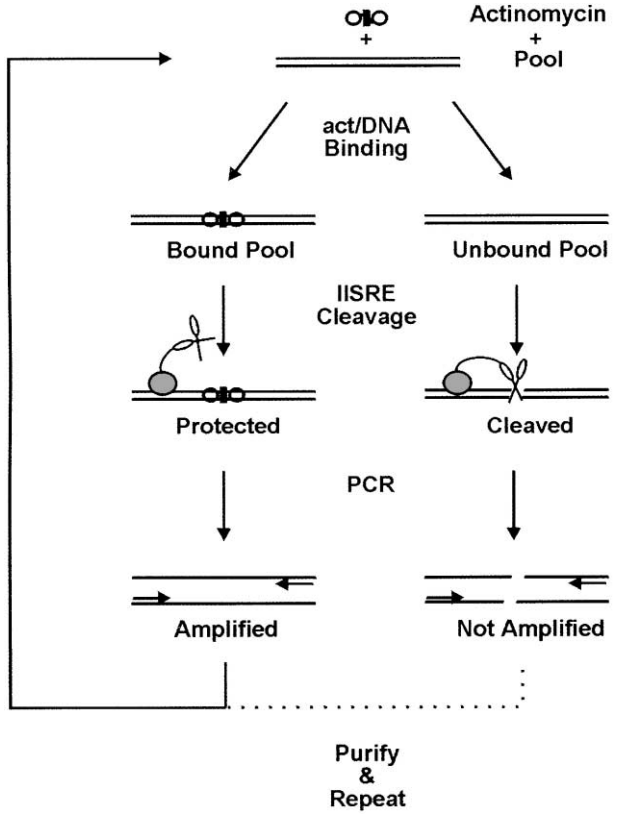

Figure 3. Flow chart for the combinatorial method restriction endonuclease protection, selection, and amplification (REPSA). Actinomycin is represented by a winged rectangle, while a type IIS restriction endonuclease is represented by a scissors (nonspecific cleaving domain) attached to a gray oval (sequence-specific DNA-binding domain).

sequence-specific DNA binding. This actinomycin concentration was chosen since it corresponded to the average affinity that actinomycin has demonstrated to 'strong' sites in heterogeneous DNA, as determined by DNase I footprinting.[6] *Fok*I (0.02 units), sufficient for cleaving >95% of the unbound template under these conditions, was then added, and the reaction mixture incubated for an additional 2 min at 37 °C. Note that these IISRE cleavage conditions were far less aggressive than those used in the REPSA selection of hTBP (2 units *Fok*I and 30 min digestion).[16] This difference reflects the nonspecific IISRE cleavage inhibition found with partially purified protein preparations. After IISRE challenge, the mixture was added directly to a PCR reaction buffer, including the 63AL and 63AR amplimers and *Taq* polymerase, and intact DNA amplified for six cycles. Afterwards the full-length selection templates were purified from IISRE-cleaved templates, unused amplimers, proteins, nucleotides, buffers, and salts by phenol extraction and spin filtration. These selected templates were then used to initiate a second round of REPSA, following the same procedures described above. A total of 10 rounds of REPSA were eventually performed. Rounds 1–9 used the IISRE *Fok*I, while round 10 used *Bpm*I. *Fok*I was chosen as the primary IISRE since it cleaved DNA more efficiently than *Bpm*I[16] and provided better discrimination between G/C-rich and A/T-rich test templates when actinomycin was present (Jing Shen, unpublished observation).

Rounds of REPSA selection are typically continued until evidence for a ligand-dependent, IISRE cleavage-resistant population is finally observed. This is usually obtained by performing semiquantitative PCR with REPSA-selected templates that have been either (1) incubated first with the ligand and then cleaved, (2) cleaved directly, or (3) left uncleaved. In our REPSA selection of actinomycin-binding DNAs, such assays were performed with templates obtained following round 5 selection and thereafter. No evidence for an actinomycin-dependent, cleavage-resistant population was observed with material from rounds 5 or 6, and only very slight evidence was seen with round 7 templates (data not shown). However, material from round 8 gave very strong evidence for an actinomycin-dependent, cleavage-resistant population (Fig. 4). Following a densitometric analysis of the nine cycle PCR products, reaction mixtures containing 1 μM actinomycin and then cleaved with *Fok*I demonstrated PCR product amounts midway between those found when actinomycin was absent (cleavage control) and those found when *Fok*I cleavage was omitted (DNA control). These data suggest that over 50% of the round 8 selection templates contained strong actinomycin-binding sites. The PCR assay with actinomycin and *Fok*I cleavage constituted a ninth round of REPSA. The material from this round was then subjected to a 10th round of REPSA with the IISRE *Bpm*I. This step reduced the possibility that templates containing only *Fok*I binding sites in the randomized region would appear, as was found previously when selections with only a single IISRE were performed.[14]

Actinomycin-selected oligonucleotide sequences

Oligonucleotides from the 10th round of REPSA selection were digested with *Eco*RI and *Hin*dIII, subcloned into like digested pUC19 and transformed into *Escherichia coli*. Plasmid DNA was isolated from individual clones and dideoxynucleotide sequenced. DNA sequences of these cloned inserts, corresponding to the top strand of ST2 (Fig. 2), are shown in Table 1. What was initially quite evident is the enrichment of G/C bases (51%) in these sequences compared to the original ST2 library (34%). Many of these G/C bases were present in 5′-GC-3′ dinucleotides, the expected preferred DNA-binding site for actinomycin.[2,3] In fact, every selected clone contained at least one 5′-GC-3′ dinucleotide, and many contained two or three. A χ^2 statistical analysis of 5′-GC-3′ showed it occurred far more often than expected randomly ($p < 0.001$), even given the enrichment of G/C bases in this population. Conversely, the other G/C-containing dinucleotides were underrepresented in the selected population, though their underrepresentation occurred with far lower statistical significance. Analyses of tetranucleotide sequences containing the 5′-GC-3′ dinucleotide (Table 2) showed that only 5′-TGCA-3′ and 5′-TGCT-3′ were significantly overrepresented in the selected clones ($p < 0.01$, where $p = 0.05$ is considered significant), though these analyses were not possible for most tetranucleotides given their low observed occurrence. However, the frequency of the degenerate sequence 5′-(A/T)GC(A/T)-3′ appeared in the selected clones easily reached extremely high significance ($p < 0.001$), suggesting that this sequence is a consensus. Additionally, taking into account the A/T-rich defined sequences immediately adjacent to the 14-bp randomized region, which could participate in the creation of an actinomycin binding site that could be selected by REPSA, fully 41 of the 45 selected clones contained at least one 5′-(A/T)GC(A/T)-3′ consensus sequence, and 14 contained multiple consensus sequences. Surprisingly, these multiples were most often present as two overlapping consensus sequences, that is 5′-(A/T)GC(A/T)GC(A/T)-3′. The frequency of this longer sequence compared with that expected by chance (8 vs 0.3 in 45 clones) was also found by a χ^2 statistical analysis to be very significant ($p < 0.0074$), suggesting this 7-bp sequence may also be a high affinity actinomycin-binding site consensus sequence.

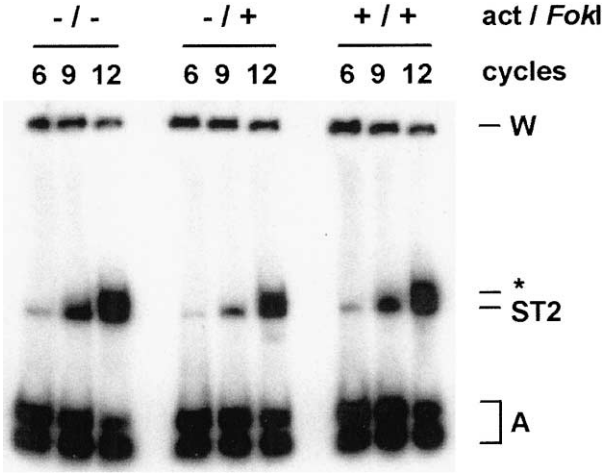

Figure 4. Evidence of an emergent, actinomycin-dependent, *Fok*I cleavage-resistant oligonucleotide population by semiquantitative PCR and nondenaturing PAGE. Shown is an autoradiogram of PCR products obtained using oligonucleotides selected after eight rounds of REPSA as templates. Oligonucleotides were either incubated in the presence of 1 μM actinomycin and then cleaved with *Fok*I (+/+), incubated in the absence of actinomycin and then cleaved with *Fok*I (−/+), or neither incubated with actinomycin nor cleaved with *Fok*I (−/−). The products were amplified by PCR for the number of temperature cycles indicated. The locations of the gel well (W), an improperly annealed or 'bubble'-containing ST2 DNA (*), properly annealed ST2 DNA, and the labeled 63AR amplimer (A) are indicated.

Identification and characterization of actinomycin binding sites

The clones selected by REPSA were individually assayed to determine their actinomycin binding affinity and exact binding sites. Two assays were employed: (1) a *Fok*I cleavage protection assay and (2) DNase I footprinting.[6,22] A IISRE cleavage protection assay provides a rapid, quantitative means of screening clones for their ability to inhibit cleavage in a ligand-dependent fashion. This provides an indirect measure of ligand-binding affinity, since the ability of a ligand to inhibit cleavage is also dependent on other factors independent of binding affinity (e.g., relative locations of ligand binding and IISRE cleavage sites). Typically, a IISRE cleavage protection assay is not capable of defining exact ligand binding sites nor of determining accurate binding affinities at discrete sites. However, both of these determinations can be accurately made with the more laborious, quantitative DNase I footprinting assay.[6]

For the IISRE cleavage protection assay, radiolabeled 105-bp probes containing the ST2 insert were generated by PCR amplification using the MSU and MSR amplimers. These probes were incubated with or without 1 μM actinomycin for 30 min at 37 °C to affect specific binding and then subjected to cleavage by *Fok*I. Cleavage products were resolved by nondenaturing gel electrophoresis and visualized by autoradiography. The results from a representative series of assays are shown in Fig. 5. As seen in this example, all the probes showed significant *Fok*I cleavage protection after incubation

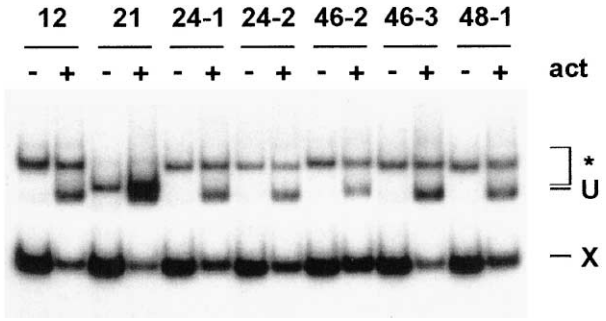

Figure 5. Analysis of REPSA-selected clones by *Fok*I cleavage protection and nondenaturing PAGE. Shown is an autoradiogram of the *Fok*I cleavage products produced after incubation without (−) or with (+) 1 μM actinomycin for a series of REPSA-selected clones. The locations of a PCR-artifact species (*), intact ST2 (U), and *Fok*I-cleaved ST2 (X) are indicated.

Table 1. REPSA-selected actinomycin D–DNA binding sites

Clone	Sequence (5′->3′)[a]	REP[b]	K_d (nM)[c]
27	gaa**tGCTGCTGCA**GGTTAttac	+ + +	12
9-1	gaatTTTAGTC**AGCAGCGt**tac	+ +	26
12	gaatA**TGCACCC**AAACATttac	+ +	26
36-1	gaatT**AGCTGCAGGT**ATTttac	+ + +	30
2/4	gaatTTAATGTA**TGCTGCt**tac	+ +	36
35-1	gaatCTGAAGA**TGCTATG**ttac	+	45
5-3	gaat**GCAGGGACAGCAG**Attac	+ + +	51
49-1	gaatCCTTA**GCTGCTTCG**ttac	+ +	51
18-1	gaatGTAA**GCTGTGTGCT**ttac	+	54
10-1	gaat**TGCAG**AGTCTCCAAttac	+	55
29-1	gaatTACCGT**TGCACTC**Tttac	+	59
1-1	gaatA**TGCAGCAA**ACCAGttac	+ +	68
26-1	gaatCGCACGCTATAATCttac	+ +	69
46-3	gaatCGGG**AGCTGCTTT**Attac	+	69
11-1	gaa**tGCTAGCATGCT**GTGttac	+ + +	70
1-3	gaatCC**TGCAGTA**CAGGAttac	+	71
33-1	gaatGGC**TGCATTAA**GAGttac	+	71
9-2	gaatACTTGTACCT**AGCAtta**c	+/−	82
31	gaatC**AGCTGGT**TTAGATttac	+ +	82
45-1	gaatGAGGA**GCGCAG**AAAttac	+ +	82
44-2	gaatTTAGCCCC**GCTCAC**ttac	+/−	83
15-1	gaat**GCGCTGA**AATT**AGCt**tac	+ +	87
23-2	gaatTTTTTTA**TGCACTt**tac	+ +	89
25-1	gaatTTGTCCTAA**TGCATtt**ac	+	90
47-1	gaa**tGCGTGG**TTCAATCGttac	+	90
42-3	GaatCACCTTTT**TGCTGCt**tac	+ + +	94
48-1	GaatGCCC**TGCAACA**GAGttac	+	98
39-1	GaatAGTT**AGCAAGCAT**Tttac	+ +	99
46-2	Gaa**tGCAA**ATCTCATATAttac	+/−	107
48-2	GaatTA**TGCACAA**ACTGGttac	+ +	110
3-2	GaatA**CGCTCAA**GGC**TGCt**tac	+ +	112
38	GaatAGTCTCAAGGC**TGCt**tac	+	121
14-1	Gaa**tGCTGCTTT**CTGGCAttac	+ + +	123
34	GaatCCATGGCA**CGCAT**Cttac	+ +	124
19	GaatTG**TGCTGT**ATTCCAttac	+ +	125
8	GaatCT**TGCAACC**TTAATttac	+	133
24-1	Gaat**TGCTAC**CT**TGCACTt**tac	+	133
21	GaatGCCCGCGCA**GCATAt**tac	+ +	141
13-3	GaatGACATTC**AGCATT**Tttac	+ +	147
24-2	GaatTTC**TGCTGGC**GACTttac	+	159
17-2	GaatGGCTT**CCGCAATGCt**tac	+ +	165
4-2	Gaat**TGCTACTC**TGCGTCttac	+	188
5-1	Gaa**tGCTTCC**TCAG**AGCT**ttac	+	205
11-3	Gaat**GCTT**CAGCGGTCGCttac	+ +	223
13-1	GaatGGGTCAA**TGCAGTA**ttac	+ +	258

[a]Sequences present in the original 14-bp randomized region are indicated by upper case letters, while those in the defined flanks are indicated by lower case letters. Sequences protected from DNase I cleavage at the apparent K_d of actinomycin D are indicated by boxes. Bold sequences are those that fit the consensus 5′-(A/T)GC(A/T)-3′.
[b]REP values correspond to the percentage of *Fok* I cleavage inhibition exhibited in the presence of 1 μM actinomycin. (+/−) corresponds to less than 20% cleavage inhibition, (+) 21–50% cleavage inhibition, (+ +) 51–80% cleavage inhibition, and (+ + +) greater than 80% cleavage inhibition.
[c]K_d values were determined from a one-site binding analysis of DNase I cleavage protection through a titration of actinomycin concentrations.

with actinomycin. The extent of this cleavage protection was quantitated by densitometry for each probe and the results shown in Table 1. For the representative series of probes shown in Fig. 5, cleavage protection ranged from 10% (clone 46-2) to 76% (clone 21). For all 45 clones investigated, cleavage inhibition ranged from a low of 10% to a high of 94%, with an average of 54%. In some cases, the low values of cleavage inhibition can be attributed to inefficient *Fok*I cleavage of the unbound probe (data not shown). However, in no case could this intrinsic cleavage resistance be attributed to the presence

Table 2. Appearance of 5′-XGCY-3′ tetranucleotides in REPSA-selected DNAs

X/Y	A	C	G	T
A	9	0	3	6
C	4	0	1	3
G	2	0	1	3
T	15	0	1	13

of *Fok*I binding sites, like those previously encountered in our REPSA selection of purine-motif triplex forming duplexes.[14]

For the DNase I footprinting assay, the identical radiolabeled 105-bp probes described above were incubated with either 0, 0.03, 0.1, 0.3, or 1 μM actinomycin for 30 min at 37 °C to affect specific binding and then subjected to cleavage by DNase I. Cleavage products were resolved by high-resolution denaturing gel electrophoresis and visualized by autoradiography. Cleavage products from either adenine- or guanine-specific chemical sequencing reactions were run in parallel as sequence markers. The results from a representative series of assays are shown in Fig. 6. As seen in this example, all the probes showed an area of substantial DNase I cleavage protection after incubation with 1 μM actinomycin in the region corresponding to the randomized nucleotide cassette, and several demonstrated substantial cleavage protection at lower actinomycin concentrations. The extent of this cleavage protection was quantitated by densitometry for each probe, using a DNase I cleavage band in each footprint as the test (indicated by an asterisk) and the band corresponding to T20 in the left flank of ST2 as a DNase I cleavage control. These data were then fit to a one-site binding equation and the derived equilibrium dissociation constant (K_d) shown in Table 1. For the representative series of probes shown in Fig. 6, actinomycin dissociation constants ranged from 12 nM (clone 27) to 133 nM (clone 8). For all 45 clones investigated, K_ds ranged from 12 to 258 nM, with an average of 98 nM. The exact nucleotides on the top strand of ST2, protected from DNase I cleavage with the actinomycin concentration tested greater than the K_d, is also shown for each clone sequence in Table 1. Every clone demonstrated at least one footprint, ranging in size from 5 to 15 nucleotides long (average 7.6). What is most striking is that each footprint sequence contained a 5′-GC-3′ dinucleotide, usually located on the 5′ edge of the footprint (e.g., 5′-TGCATTAA-3′ for clone 33-1). Such is readily understandable given the right-handed nature of B-form, double helical DNA. Longer footprints typically encompassed multiple 5′-GC-3′ dinucleotides, though not all 5′-GC-3′ dinucleotides were present in footprints (e.g., clone 11-3).

Discussion

Using REPSA, we screened almost 134 million different 14-bp sequences, to identify those that inhibit restriction *Fok*I endonuclease cleavage in an actinomycin concentration-dependent manner. Forty-five of these selected

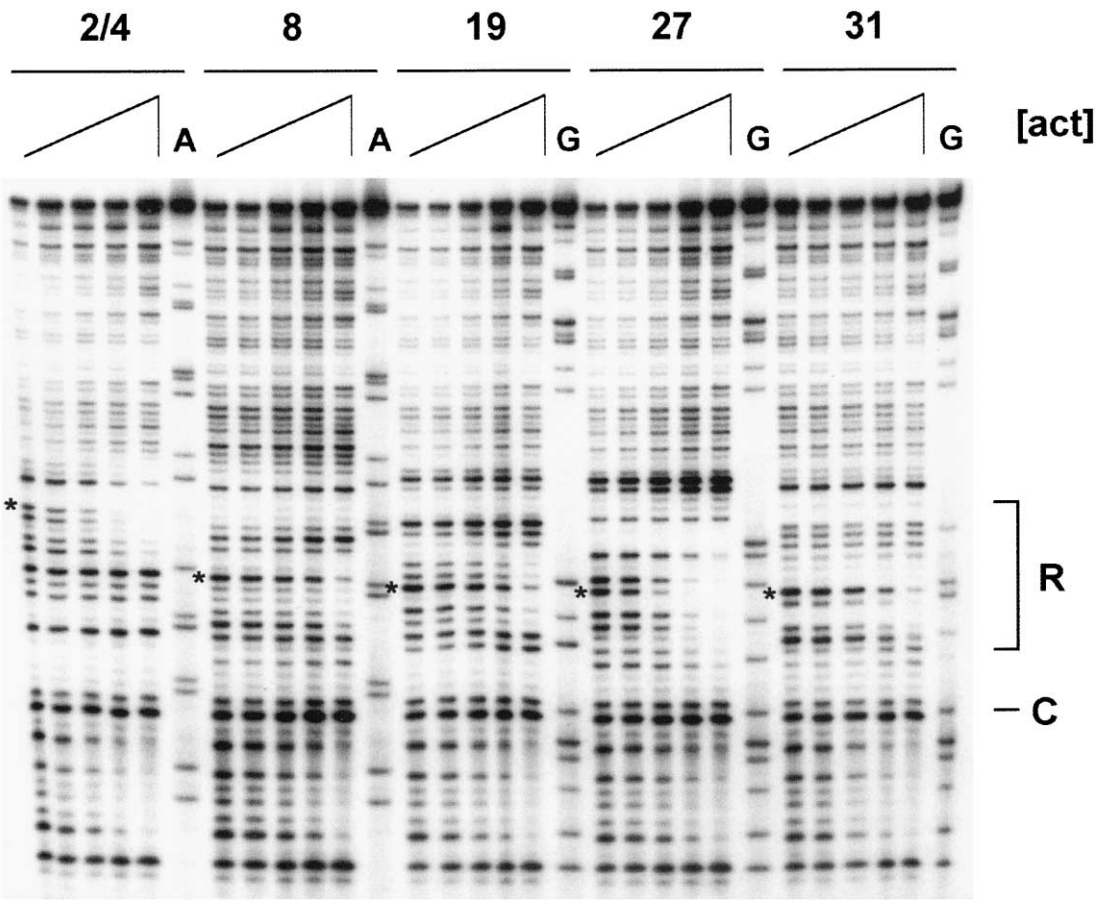

Figure 6. Identification of preferred actinomycin binding sites on REPSA-selected clones by DNase I cleavage protection and high resolution denaturing gel electrophoresis. Shown is an autoradiogram of the reaction products produced after DNase I cleavage in the presence of 0, 30, 100, 300, and 1000 nM actinomycin (actinomycin concentration indicated by the height of the wedge above each lane) for a series of REPSA-selected clones. Products of adenine- (A) or guanine- (G) specific chemical sequencing reactions, which serve as electrophoretic markers, are indicated at right of the DNase I footprinting lanes for each clone. The DNase I cleavage product that was used to quantify actinomycin binding affinity is indicated by an asterisk at left for each clone. The locations of the randomized cassette (R) and the DNase I cleavage product that served as an internal control (C) are also indicated.

sequences were subcloned and individually analyzed by quantitative DNase I footprinting. Each was found to contain at least one region of actinomycin-dependent DNase I cleavage protection, with apparent K_ds ranging from 12 to 258 nM. A statistical analysis of all selected sequences found the dinucleotide 5′-GC-3′ to be vastly overrepresented, with the degenerate tetranucleotide 5′-(A/T)GC(A/T)-3′ constituting a consensus. These sequences were present in each actinomycin-binding site defined by DNase I footprinting, though apparently not every 5′-GC-3′ was bound by actinomycin under these conditions. A longer sequence, 5′-(A/T)GC(A/T)GC(A/T)-3′, appeared in several of the highest affinity actinomycin-binding clones, suggesting it constitutes a preferred actinomycin-binding site consensus sequence.

In general, our findings are quite understandable, given our current knowledge of DNA recognition by actinomycin. By a number of methods, the sequences 5′-XGCY-3′, have been shown to have actinomycin-binding affinities in the order of A, T > C ≫ G for nucleotide X and A, T > G ≫ C for nucleotide Y.[6,7,20,21] These canonical actinomycin-binding sequences have binding affinities ranging from 100 to 400 nM, though these values are dependent on flanking sequences and the assay method used. Thus, our observed K_ds are consistent with those described for high-affinity actinomycin-binding sites. It should be noted that high-affinity, noncanonical actinomycin-binding sites have also been described in the literature (e.g., 5′-TGGGT-3′).[7,23,24] However, while similar sites were observed in some of our clones (e.g., clones 17, 5-3, 13-1), they often did not overlap with a DNase I footprint, nor were there significantly more of them than expected in a random distribution. Thus it appears high-affinity, canonical actinomycin binding sites were preferentially selected by REPSA.

The longer consensus sequence we identified, 5′-(A/T)GC(A/T)GC(A/T)-3′, is quite interesting for several reasons. Preferred binding to these overlapping sites is somewhat unexpected, given the minimal four base pair binding site size found for actinomycin by NMR spectroscopy and X-ray crystallography.[20,21,25,26] It has been reported that overlapping actinomycin binding sites can bind two actinomycin molecules, albeit

only at high actinomycin concentrations.[27] In the sequence 5′-GCGC-3′, the sites of intercalation for the phenoxazinone chromophore of the two actinomycins are immediately adjacent to one another. Nonetheless, simultaneous binding could be achieved through a significant conformational change in the DNA structure, including an unwinding of the DNA double helix and a pronounced kink between the adjacent binding sites towards the major groove, resulting in an opening of the minor groove. Our longer consensus should not require such significant alterations in the DNA structure, since the two actinomycin intercalation sites are separated by a single A-T base pair. While the high apparent affinity of this site suggests cooperative binding occurs, it remains unclear why the binding of one actinomycin molecule would facilitate the binding of a second so close by at low actinomycin concentrations. An alternative explanation is that only a single actinomycin molecule is actually bound to these sites, and the apparent high affinity is the sum of both partially occupied sites inhibiting DNase I cleavage. Such a model is supported by the observation that actinomycin 'shuffles' one-dimensionally along duplex DNA, eventually gravitating to sites having the longest residence time.[28,29] In the case of our longer consensus sequence, the presence of two adjacent high-affinity sites would create a 'sink' that may locally capture a diffusing actinomycin molecule, since half of all migration events would encounter the adjacent high affinity site. This model would also suggest that a site containing three adjacent, overlapping actinomycin-binding sites (e.g., 5′-TGCTGCTGCA-3′) might have an even greater affinity for actinomycin. In fact, this was observed with clone 27, which demonstrated the highest actinomycin binding affinity of all the sequences selected by REPSA (Table 1).

Since both *Fok*I cleavage protection and DNase I footprinting assays were performed on all our REPSA-selected clones, we were able to directly compare these two assays for their ability to identify high-affinity actinomycin-binding sites. The selected DNAs exhibited a wide range of *Fok*I cleavage inhibition in the presence of actinomycin. Comparing these values to those obtained by DNase I footprinting, we found no meaningful correlation between actinomycin-dependent *Fok*I cleavage protection and the K_d determined by DNase I footprinting for the highest affinity actinomycin-binding site. Similarly, no correlation was found between *Fok*I cleavage protection and the location of the actinomycin binding site within the 14-bp randomized cassette. However, a weak correlation was observed between *Fok*I cleavage protection and the absolute size of the DNase I footprint, suggesting multiple actinomycin binding sites are generally better able to inhibit *Fok*I cleavage. A IISRE cleavage-protection assay can be effectively used to quantitatively determine the dissociation constant of a single, large ligand–DNA complex, with a pyrimidine-motif triplex being a perfect example.[22] However, the accuracy of this method is dependent on a number of parameters, including the intrinsic effectiveness of cleavage inhibition by the ligand–DNA complex and the position of the complex relative to the sites of IISRE cleavage, as well as the number of complexes present and their individual dissociation constants. Thus, in practice for many small molecule ligands, quantitative DNase I footprinting is the superior method for determining ligand-binding affinities as well as binding sites. However, for small molecules that recognize and protect longer regions of DNA with high affinity (e.g., hairpin polyamides), IISRE cleavage protection assays can provide a powerful tool for screening REPSA-selected clones as well as for accurately determining individual clone binding affinities (Y. N. Vashisht Gopal, unpublished observations).

Conclusions

REPSA is the only combinatorial method that allows the identification of preferred duplex DNA-binding sites for unmodified small molecule ligands under physiological conditions. We have successfully used REPSA to identify high-affinity binding sites for the minor groove-binder distamycin A[17] and intercalator actinomycin D. In the former case we identified several large binding sites with nanomolar binding affinities, suggesting occupancy by multiple, cooperatively interacting distamycin molecules. In the latter case, we identified a consensus sequence for high-affinity actinomycin binding, though we also found that multiple, overlapping consensus sequences yielded actinomycin-binding sites with even higher apparent affinity. The common theme of these two studies is that multiple small molecule ligand binding sites tend to be those of highest affinity, and these are preferentially selected by REPSA. From our previous studies with REPSA, we have come to find that this combinatorial method can be used with mixtures of uncharacterized ligands. Thus, we believe REPSA is a powerful tool for screening libraries of small molecules to identify sets of preferred binding sites. These may in turn be back selected against the small molecule libraries to determine which molecules or combinations of molecules are responsible for high-affinity binding. Eventually, such two-dimensional combinatorial searches could be used to identify drugs that pharmacologically regulate targeted gene expression or catalyze replacement/repair of defective genes.

Experimental

Oligonucleotides

Phosphodiester oligodeoxyribonucleotides were prepared on a Millipore Cyclone DNA synthesizer. The nucleotide sequences of oligonucleotides used in this study were (5′ to 3′): 63AL, CTAGGAATTCGTGCAGAGGTGA; 63AR, GTCCAAGCTTCTGGAGGGATG; 63R14, CTAGGAATTCGTGCAGAGGTGAAT$(N)_{14}$TTACCATCCCTCCAGAAGCTTGGAC; MSR, AAACAGCTATGACCATG; and MSU, ACGACGTTGTAAAACGA. For oligonucleotide 63R14, the sites containing mixed bases (N) were synthesized using an equimolar mixture of each phosphoramidite. The distribution of nucleotides incorporated into the random cassette was 27% A, 18% C, 16% G, and 39% T

as determined by sequencing of eight individual clones derived from starting material.

REPSA

The double-stranded selection template ST2 was synthesized by four rounds of PCR using the oligonucleotide 63R14 as template and 63AL and 63AR as amplimers. To effect actinomycin binding, 2 ng (48 fmols) ST2 was incubated with 10 pmoles of actinomycin D in a 10-μL volume containing 20 mM Tris–acetate, pH 7.9, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol, 0.01% Triton X-100, and 1 μg bovine serum albumin ('binding buffer') for 30 min at 37 °C. Following actinomycin binding, 0.05 U of either *Fok*I or *Bpm*I (New England Biolabs) in a 1-μL volume containing the appropriate enzyme dilution buffer was added, and the incubation continued for an additional 2 min. To amplify the cleavage-resistant duplex DNA subpopulation, 100 ng 63AL, 95 ng 63AR, 5 ng 5′ ^{32}P end-labeled 63AR (50,000 cpm), 2.5 U *Taq* DNA polymerase (Promega), 0.25 mM concentrations of dATP, dCTP, dGTP, and dTTP, 10 mM Tris–Cl (pH 9.0), 50 mM KCl, 1 mM $MgCl_2$, and 0.1% Triton X-100 were added to each sample, to a final volume of 100 μL. The amplification profile used for PCR was 94 °C for 1 min followed by 50 °C for 3 min. Duplicate reactions were amplified for 6, 9, and 12 cycles. Following PCR amplification, each reaction mixture was phenol extracted and the aqueous phase concentrated on a Millipore Ultrafree-MC 5000 cellulose spin filter by centrifugation for 30 min at 15,000×*g*. Filters were washed 10 min with 200 μL 10 mM Tris–Cl (pH 7.4), 1 mM EDTA and centrifuged for 30 min. The retained template DNA was resuspended in 20 μL Tris–EDTA, and a 2-μL aliquot was analyzed by PAGE and autoradiography to determine relative levels of amplification. These steps—actinomycin binding, enzyme cleavage, PCR amplification, and filter purification—were repeated for a total of 10 times. In the first nine rounds, *Fok*I was the selection IISRE. In the final round, *Bpm*I was used, in part to reduce the selection of DNA containing *Fok*I binding sites in the randomized region.

Sequence determination and statistical analysis

The finally selected ST2 templates were digested with *Eco*RI and *Hin*dIII and cloned into similarly cut plasmid pUC19. Individual colonies were used to inoculate 5-mL overnight cultures in Luria broth medium containing 0.2 mg/mL ampicillin. Mini-plasmid preparations were made from the clones, and their inserts sequenced by Sanger dideoxynucleotide sequencing.

The significance of differences in experimentally determined consensus sequences was determined by a chi-squared (χ^2) comparison of actual distributions in REPSA-selected sequences to their expected frequency of appearance, given nucleotide distributions present in the total population of sequences isolated after the final REPSA round. *p* Values less than 0.05 were considered significant. In the actinomycin selection, the final nucleotide distribution was 22.2% A, 26.5% C, 24.4% G, and 26.8% T from a total of 630 nucleotides sequenced.

Binding affinity determination

The binding affinity of actinomycin to each REPSA-selected DNA was determined by a IISRE cleavage protection assay and by quantitative DNase I footprinting. DNase I footprinting also allowed a determination of the exact site of actinomycin binding on each DNA. For both assays, radiolabeled probes were generated by PCR amplification of each clone using 5′ end-labeled primer MSU and unlabeled primer MSR. To effect DNA binding by actinomycin, 10 fmol purified, labeled probe DNA was incubated with actinomycin (from 0 to 1 μM, as indicated in each figure legend) in 10 μL of binding buffer for 30 min at 37 °C. For the IISRE cleavage protection assay, 0.05 U *Fok*I was added and the incubation continued for an additional 2 min. Reaction products were analyzed directly by nondenaturing PAGE (8%), visualized by autoradiography, and quantitated by densitometry. *Fok*I cleavage inhibition was calculated with the formula $I = 1-(A/C)$, where A is the fraction of cleaved labeled probe when actinomycin was present and C is the fraction in the absence of actinomycin. For the DNase I footprinting assay, cleavage was initiated by the addition of 2 ng DNase I (Sigma, DN-EP grade) and allowed to continue for 30 s at room temperature before it was terminated by the addition of 3 μL stop buffer (3.0 M ammonium acetate, 0.5 M EDTA, 1.0 mg/mL tRNA). Products from adenine- or guanine-specific chemical cleavage reactions were used as markers. Samples were purified by phenol extraction and ethanol precipitation, dried, resuspended in a 95% formamide/10 mM NaOH buffer, and denatured by heating before loading onto a denaturing 8% polyacrylamide gel. Cleavage products were visualized by autoradiography and quantitated by densitometry. Actinomycin binding affinity was determined by measuring the reduction in DNase I cleavage for that band within the randomized region that exhibited the greatest change with increasing actinomycin concentration and normalizing this to a DNase I cleavage control located in a flanking region of ST2 (T20). Values for DNase I cleavage protection were subjected to a nonlinear regression analysis using a one-site binding hyperbolic equation ($I = I_{max}*[act]/[K_d + X]$, where I is cleavage inhibition, I_{max} is the maximal extent of cleavage inhibition observed, [act] is the actinomycin concentration, and K_d is the dissociation constant). DNase I footprints were defined by those bases that demonstrated appreciable cleavage inhibition when the actinomycin concentration tested was greater than the K_d.

Acknowledgements

The authors thank Y. N. Vashisht Gopal for critical reading of the manuscript. This research was supported by grants from the American Cancer Society (RPG-97-028-03-LBC) and the Robert A. Welch Foundation (G-1199).

References and Notes

1. Neidle, S.; Waring, M. J. *Molecular Aspects of Anticancer Drug–DNA Interactions*; Macmillan: London, 1993.
2. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470.
3. Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3260.
4. Winkle, S. A.; Krugh, T. R. *Nucleic Acids Res.* **1981**, *9*, 3175.
5. Samuelson, P.; Jansen, K.; Kubista, M. *J. Mol. Recognit.* **1994**, *7*, 233.
6. Goodisman, J.; Rehfuss, R.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1992**, *31*, 1046.
7. Bailey, S. A.; Graves, D. E.; Rill, R.; Marsch, G. *Biochemistry* **1993**, *32*, 5881.
8. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geirstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586.
9. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
10. Szostak, J. W. *Trends Biochem. Sci.* **1992**, *17*, 89.
11. Wright, W. E.; Funk, W. D. *Trends Biochem. Sci.* **1993**, *18*, 77.
12. Kenan, D. J.; Tsai, D. E.; Keene, J. D. *Trends Biochem. Sci.* **1994**, *19*, 57.
13. Gold, L.; Polisky, B.; Uhlenbeck, O.; Yarus, M. *Annu. Rev. Biochem.* **1995**, *64*, 763.
14. Hardenbol, P.; Van Dyke, M. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2811.
15. Szybalski, W.; Kim, S. C.; Hasan, N.; Podhajska, A. J. *Gene* **1991**, *100*, 13.
16. Hardenbol, P.; Wang, J. C.; Van Dyke, M. W. *Nucleic Acids Res.* **1997**, *25*, 3339.
17. Hardenbol, P.; Wang, J. C.; Van Dyke, M. W. *Bioconjugate Chem.* **1997**, *8*, 617.
18. Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445.
19. Shinomiya, M.; Chu, W.; Carlson, R. G.; Weaver, R. F.; Takusagawa, F. *Biochemistry* **1995**, *34*, 8481.
20. Chen, F. M. *Biochemistry* **1988**, *27*, 6393.
21. Rill, R. L.; Marsch, G. A.; Graves, D. E. *J. Biomol. Struct. Dyn.* **1989**, *7*, 591.
22. Ward, B. *Nucleic Acids Res.* **1996**, *24*, 2435.
23. Snyder, J. G.; Hartman, N. G.; D'Estantoit, B. L.; Kennard, O.; Remeta, D. P.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3968.
24. Bailey, S. A.; Graves, D. E.; Rill, R. *Biochemistry* **1994**, *33*, 11493.
25. Creighton, S.; Rudolph, B.; Lybrand, T.; Singh, U. C.; Shafer, R.; Brown, S.; Kollman, P.; Case, D. A.; Andrea, T. *J. Biomolec. Struct. Dyn.* **1989**, *6*, 929.
26. Lian, C. Y.; Robinson, H.; Wang, A. H. J. *J. Am. Chem. Soc.* **1996**, *118*, 8791.
27. Chen, H.; Liu, X.; Patel, D. J. *J. Mol. Biol.* **1996**, *258*, 457.
28. Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271.
29. Bailly, C.; Graves, D. E.; Ridge, G.; Waring, M. J. *Biochemistry* **1994**, *33*, 8736.