

Determination of Netropsin-DNA Binding Constants from Footprinting Data

Brian Ward, Robert Rehfsuss, Jerry Goodisman,* and James C. Dabrowiak*

Department of Chemistry, Syracuse University, Syracuse, New York 13244-1200

Received May 29, 1987; Revised Manuscript Received September 18, 1987

ABSTRACT: A theory for deriving drug-DNA site binding constants from footprinting data is presented. Plots of oligonucleotide concentration, as a function of drug concentration, for various cutting positions on DNA are required. It is assumed that the rate of cleavage at each nucleotide position is proportional to the concentration of enzyme at that nucleotide and to the probability that the nucleotide is not blocked by drug. The probability of a nucleotide position not being blocked is calculated by assuming a conventional binding equilibrium for each binding site with exclusions for overlapping sites. The theory has been used to evaluate individual site binding constants for the antiviral agent netropsin toward a 139 base pair restriction fragment of pBR-322 DNA. Drug binding constants, evaluated from footprinting data in the presence of calf thymus DNA and poly(dGdC) as carrier and in the absence of carrier DNA, were determined by obtaining the best fit between calculated and experimental footprinting data. Although the strong sites on the fragment were all of the type (T·A)₄, the value of the binding constant was strongly sequence dependent. Sites containing the dinucleotide sequence 5'-TA-3' were found to have significantly lower binding constants than those without this sequence, suggesting that an adenine-adenine clash produces a DNA structural alteration in the minor groove which discourages netropsin binding to DNA. The errors, scope, and limitations associated with the method are presented and discussed.

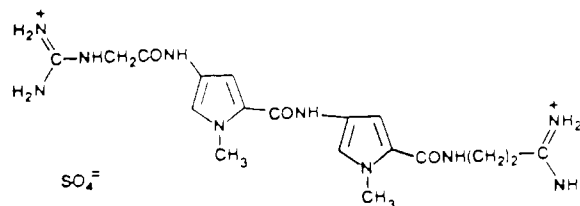
Footprinting analysis is widely used in studying ligand binding to DNA. The approach, which utilizes DNA sequencing technology, relies on the ability of a DNA-bound ligand to inhibit DNA cleavage at its binding site. Since the initial footprinting studies involving the *lac* repressor-DNA interaction (Galas & Schmitz, 1978; Schmitz & Galas 1979), the method has been used to study a variety of DNA binding proteins (Johnson et al., 1979; Sakonju & Brown, 1982) as well as drugs (Dabrowiak, 1983; Lane et al., 1983; Low et al., 1984; Scamrov & Beabealashvili, 1983; Van Dyke et al., 1982; Dervan, 1986) which are capable of binding to DNA in a sequence-specific manner. The DNA-hydrolyzing enzymes DNase I and DNase II (Drew, 1984), a variety of metal complexes (Dervan, 1986; Tullius & Dombroski, 1986; Kuwabara et al., 1986; Ward et al., 1987), alkylating agents (Ogata & Gilbert, 1979; Ward & Dabrowiak, 1987), and UV radiation (Becker & Wang, 1984) have been used as probes to identify ligand binding sites in the footprinting experiment.

Aside from simply revealing the sites of ligand binding, footprinting analysis has the potential for yielding thermodynamic information on ligand-DNA binding as a function of sequence. Unlike any other method, it allows evaluation of individual site binding isotherms and thus provides direct information on ligand binding constants as a function of the sequence of DNA bases. Although the bulk of the footprinting experiments reported to date has been qualitative in nature and has focused on identifying the interaction sites of DNA binding ligands, recent elegant work of Ackers and co-workers (Brenowitz et al., 1986a,b; Senear et al., 1986) demonstrated that footprinting data can be used to measure thermodynamic quantities associated with the binding of bacteriophage λ CI repressor to right and left operators.

Although the extension of quantitative footprinting analysis to small ligand-DNA studies appears straightforward, certain characteristics of these ligands complicate the extraction of thermodynamic data. Binding constants of drugs are in most cases several orders of magnitude less than those of DNA binding proteins. Thus, the DNA cleaving agents or probes that are used to score ligand sites may in fact disturb the

ligand-DNA interaction that is being measured. A second important factor associated with drugs concerns their small size and relatively low binding sequence specificities. Since they have low specificities, they bind to many sites on DNA, thereby displacing the probe from DNA. If either the ligand-DNA or the probe-DNA equilibrium were disturbed during the experiment, the concentrations of oligonucleotides produced in the footprinting digests would not reflect the true binding isotherms for the various site loading events taking place in the system. Finally, it has been shown (Ward et al., 1987, 1988) that binding of drugs to DNA causes an increase in DNase I cleavage away from the site of binding. This enhancement, explained by redistribution of the enzyme on DNA as binding sites load with drug, needs to be taken into account in determining binding constants from footprinting data.

Using the antiviral agent netropsin and a 139-base segment



Netropsin

of pBR-322 DNA, we earlier demonstrated that it is possible to uncover the individual site loading events on the restriction fragment with two probes, the enzyme DNase I and a cleaving cationic manganese porphyrin complex (Ward et al., 1987). Footprinting plots, showing the rate of cleavage of the probe as a function of netropsin concentration and sequence, revealed the sites of strong and weak drug binding and regions of enhancement on the fragment. In this paper we present the first quantitative analysis of DNase I footprinting titrations involving netropsin and show how values may be derived for site-specific drug binding constants on a DNA restriction fragment. The method does not require an independent de-

termination of the free drug concentration during the footprinting titration. This quantity, along with the individual site isotherms and the binding constants, is determined by a minimization approach involving only the measured footprinting autoradiographic spot intensities and the total drug concentration. In addition to defining the scope and limitations of the method, we present new evidence that netropsin can discriminate between high-affinity sites of the type (A·T)₄ in a sequence-dependent fashion.

MATERIALS AND METHODS

The various footprinting studies involving netropsin and the 139 base pair restriction fragment of pBR-322 DNA and data analysis using linear scanning microdensitometry were as earlier described (Ward et al., 1987, 1988; Dabrowiak et al., 1986). Briefly, the total DNA concentration in base pairs for the experiments involving poly(dGdC) and calf thymus DNA as carrier was $\sim 194 \mu\text{M}$ ($\sim 1 \mu\text{M}$ fragment and $193 \mu\text{M}$ carrier) in a buffer consisting of 50 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), 8 mM MgCl_2 , and 2 mM CaCl_2 at pH 7.5. The experiments involving only the labeled fragment (no carrier) were carried out at $\sim 1 \mu\text{M}$ base pairs. All digests, at 37°C , were terminated early such that $\sim 80\%$ of the 139-mer remained uncleaved during the various reactions. The concentration of DNase I in the experiment involving poly(dGdC) and calf thymus DNA as carrier was $\sim 2 \times 10^{-7} \text{ M}$ while the enzyme concentration for the no-carrier experiment was $\sim 2 \times 10^{-8} \text{ M}$.

Each footprinting titration yielded a series of footprinting plots showing oligonucleotide concentration, as a function of drug concentration, for 61 sites on the 139-mer. The autoradiographic data were scanned with a linear scanning microdensitometer computer system yielding cross-sectional areas directly proportioned to oligonucleotide concentration (Dabrowiak et al., 1986). The oligonucleotide concentrations (areas) a_i were corrected for differences in the total extent of digest for a given reaction. These corrections were usually <10% (Ward et al., 1987). The present analysis, aimed at identifying the binding constants for strong sites of netropsin-DNA interaction, considered only the early loading events on the 139-mer, so only the lower drug concentrations were used. Sequence analysis revealed all the strong sites to be of the type (A·T)₄.

Densitometric scanning of the autoradiographic data (Dabrowiak et al., 1986) also showed that the total amount of cleavage on the restriction fragment for all experiments was constant over the concentration range during which the strong sites on the fragment were accepting drug. This observation eliminated the need for correcting the observed footprinting plots for a shift in the enzyme-DNA equilibrium as drug was added to the system.

The sequence of the fragment studied along with the locations of strong netropsin binding is shown in Figure 1. Selected densitometric scans of autoradiographic data are given in Figure 2. In order to determine individual site binding constants when overlapping drug sites were involved, it was necessary to monitor the site loading events at as many points as possible throughout the overlap region. The observed footprinting plots and the calculated plots for the most highly overlapped region of the fragment, having the sequence 5'-CAGTTAAATT-3', positions 53-62, are shown in Figures 3 and 4.

In the case of calf thymus DNA as carrier, the sites at 45, 48, 50, 54, 55, 59, 62, 63, 85, 87, 90, and 154 were used to monitor strong binding to the fragment while those at 53, 71,

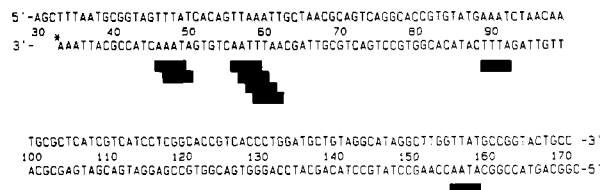


FIGURE 1: Sequence of the 139 base pair, *Hind*III/*Nci*I restriction fragment of pBR-322 DNA. The locations of netropsin binding are indicated with a solid rectangle (four nucleotides long) on the sequence. The fragment is end labeled (32 P) at position 33.

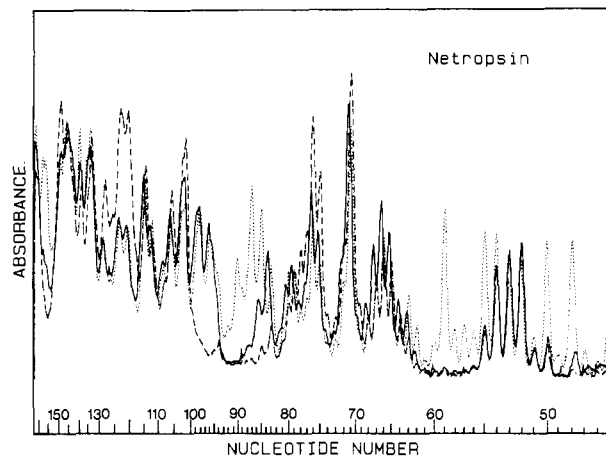


FIGURE 2: Densitometric scans of selected autoradiographic data involving netropsin and the 139-mer in the presence of calf thymus DNA (193 μ M base pairs). DNase I alone (---); 7.8 μ M netropsin (—), the strong sites; 39 μ M netropsin (---), the strong and weak sites.

114, and 102 measured the enhancement effect due to redistribution of enzyme on DNA (Ward et al., 1988). Sixteen netropsin concentrations C in the range $0 \leq C \leq 8.99 \times 10^{-6}$ M were used in this analysis. There were thus 256 intensities (oligonucleotide concentrations) to be fitted.

For the experiments using poly(dGdC) as carrier, the sites used for strong drug binding were 49, 50, 54, 55, 56, 57, 58, 59, 60, 62, 87, 154, and 90; data for sites 102 and 114 were used to monitor the enhancement. The number of sites, along with six drug concentrations, in the range $0 \leq C \leq 4.97 \times 10^{-7}$ M, gave rise to 90 concentrations to be fitted.

For the carrier-free experiments, the sites monitored were 48, 49, 50, 53, 54, 55, 57, 58, 59, 62, 71, 85, 87, 90, 92, and 98. All showed the effect of drug binding to strong sites except 53, 71, and 98, which monitored enhancement. Since there were 11 drug concentrations, in the range $0 \leq C \leq 1.95 \times 10^{-7}$ M, 176 intensities (oligonucleotide concentrations) were used in the determination of equilibrium constants. Due to a small amount of background cleavage observed in the control lanes of the gel, a "background" densitometric scan was subtracted from all scans in the carrier-free experiments.

THEORY

The rate of cleavage by DNase I at an exposed phosphodiester linkage of DNA, $(\text{rate})_i$, is governed by expression 1,

$$(\text{rate})_i = k'_i[\text{DNase I}]_i \quad (1)$$

where k'_i is the cleavage rate constant at the site and $[DNase I]_i$ is the concentration of enzyme at the site. A site on DNA becomes inaccessible to the enzyme when a drug molecule binds in such a way as to block the hydrolysis of the phosphodiester linkage. Assuming that $(rate)_i$ is then reduced to 0, we write, instead of (1), eq 2, where p_i is the probability

$$(\text{rate})_i = k'_i [\text{DNase I}]_i p_i \quad (2)$$

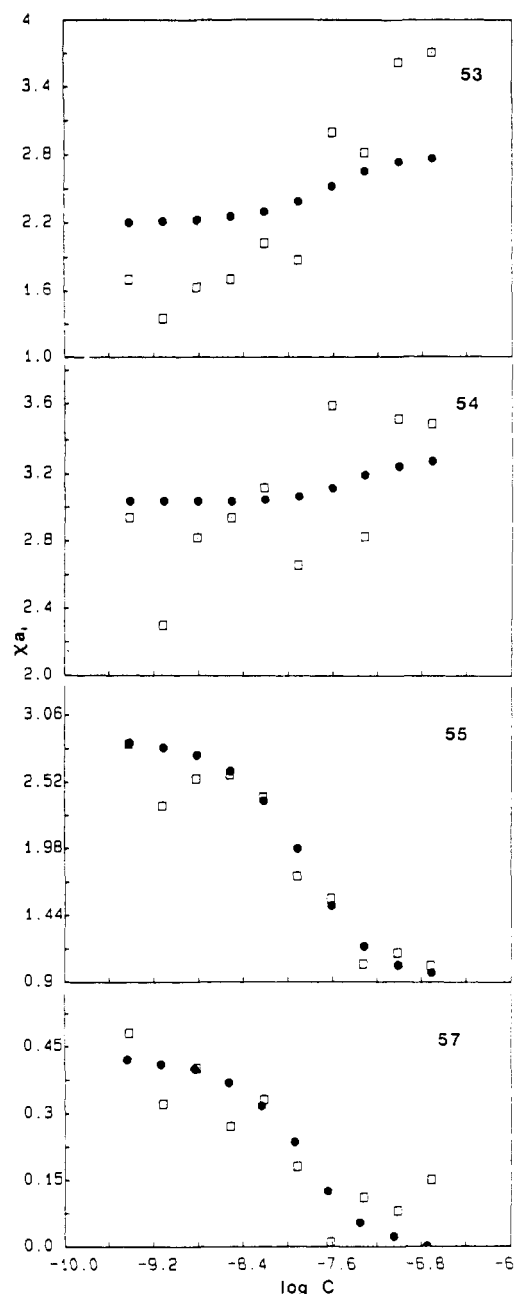


FIGURE 3: Observed (\square) and calculated (\bullet) oligonucleotide areas (concentrations) as a function of the logarithm of the total netropsin concentration, $\log C$, for the no carrier experiment. The quantity, χ , a correction factor for slight differences in the extent of digest, and a_i , the linearly scanned band area, were determined as earlier described (Ward et al., 1987; Dabrowiak et al., 1986).

that site i is not blocked to the enzyme by bound drug. As evidence that drug binding totally blocks cleavage, we note that, for some sites such as 59 and 90, the cleavage rate can be reduced essentially to 0 by addition of the drug.

Since the footprinting digests are terminated early enough so that one is in the "single-hit" regime, the concentrations of the various oligonucleotides produced by cleavage are proportional to the initial cutting rates at the corresponding sites. Thus, the observed oligonucleotide concentrations reflect the equilibrium drug binding situation over the sites which exists before cleavage of DNA takes place (Goodisman & Dabrowiak, 1985). To predict these concentrations for comparison to experiment, one has to calculate the $\{p_i\}$ as a function of total drug concentration. In order to do so, other factors that affect the rate must be considered.

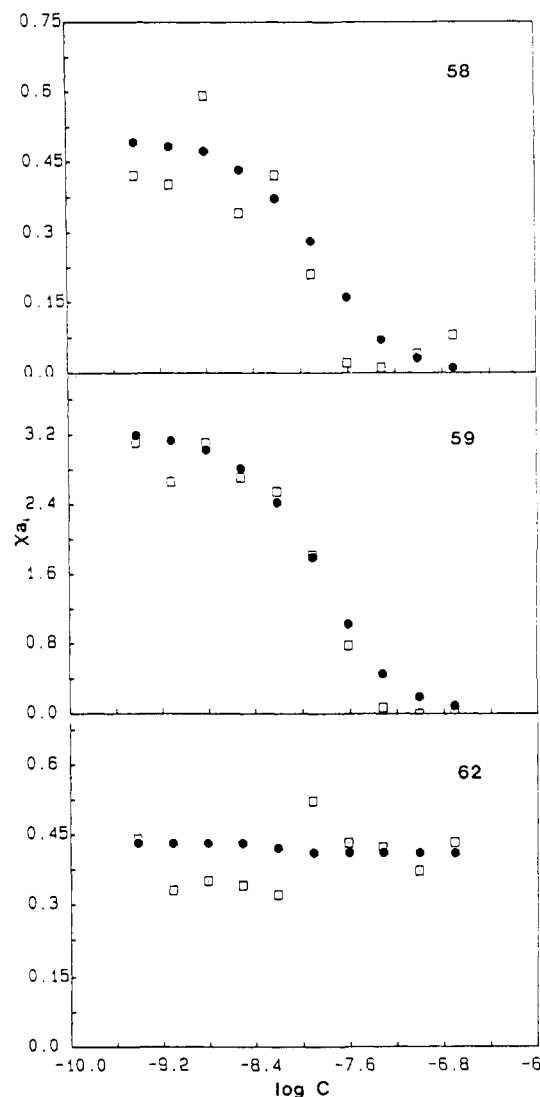


FIGURE 4: Observed (\square) and calculated (\bullet) oligonucleotide areas (concentrations) as a function of the logarithm of the total netropsin concentration, $\log C$, for the no-carrier experiment. Quantities χ and a_i are defined as in Figure 3.

If, for example, drug binding to one site produces a structural change in DNA within a binding region of another drug site, i.e., a change in k_i , the measured spot intensities associated with the second site will not reflect the true occupancy of the site. Although this is not likely to occur for netropsin, which is known to cause little distortion to DNA (Kopka et al., 1985), other effects are known to exist.

Correction of Redistribution of DNase I on DNA. The amount of cleavage at a particular site can be affected by changes in $[\text{DNase I}]_i$ as drug loading on the restriction fragment takes place. Previous studies with netropsin and the 139-mer revealed that the total amount of cleavage on the fragment remains constant while the strong sites are accepting drug (Ward et al., 1987, 1988). This shows that the amount of enzyme on the fragment is constant and that drug binding simply shifts enzyme to other sites on the fragment. Since drug binding at one site on a restriction fragment increases $[\text{DNase I}]_i$ at other unoccupied sites on the same fragment, the spot intensities associated with these sites must be corrected for this enhancement. Since fractional enhancements at almost all nonbinding sites were observed to be essentially the same, we take the enhancement effect into account by writing

$$[\text{DNase I}]_i = (1 + \alpha b)[\text{DNase I}]_{i0} \quad (3)$$

where α is a constant to be determined and b is the amount of bound drug on a fragment. The enhancement factor of $1 + \alpha b$ applies to binding as well as nonbinding sites.

If it is true that the enhancement arises because drug at one site shifts enzyme to other sites, without decreasing the amount of enzyme on the fragment, the factor $1 + \alpha b$ should be replaced by $(1 - \alpha b)^{-1}$. This would correspond to an even distribution of enzyme over all sites as would occur with a facilitated diffusion mechanism (Winter & von Hippel, 1981). As long as αb is not large compared to unity, $(1 - \alpha b)^{-1}$ is essentially the same as $1 + \alpha b$. In fact, we have found that replacing the latter by the former gives rise to only extremely small changes in the results of our calculations.

The effective concentration of DNase I at a particular site i in the absence of drug, $[\text{DNase I}]_0$, may be incorporated into the effective rate constant. Since the concentration of a particular oligonucleotide produced in the footprinting digest c_i is proportional to the initial cleavage rate constant and the measured area a_i is proportional to the oligonucleotide concentration, we have

$$a_i = k_i p_i (1 + \alpha b) \quad (4)$$

where $k_i = k'_i [\text{DNase I}]_0$ multiplied by the digest time.

Site Exclusion. The value of p_i obviously depends on the drug binding equilibrium constant for site i , K_i , on the drug concentration, and on the total concentration of restriction fragment present in the system. Consideration of the structure of DNase I and the length of netropsin on DNA suggested that drug binding to site i actually blocks seven or eight sites from DNase I cleavage (Suck & Oefner, 1986; Ward et al., 1988). Four sites are blocked due to the length of netropsin on DNA, while three or four additional sites are blocked due to the relative orientation between a loop on the enzyme and its cleavage site. In the following analysis of netropsin footprinting data, it will be evident that the exact size of the protection region, seven or eight nucleotides, may also depend on the specific sequence of DNA at which binding is taking place. However, for the purpose of discussing site exclusion a protection region of seven is assumed, implying that seven nucleotide positions can be used to measure K_i . Alternatively, p_i depends on all of the values of K_j , with j between $i - 3$ and $i + 3$. By "binding of a drug at site j ", we mean netropsin occupation of nucleotide positions $j, j + 1, j + 2$, and $j + 3$.

For an isolated site i , the binding equilibrium is

$$K_i = \frac{\nu_b}{D_0(1 - \nu_b)} \quad (5)$$

where D_0 is the free drug concentration and ν_b is the fraction of site bound by drug, $\nu_b = 1 - p_i$. However, the situation is complicated by the existence of overlapping binding sites. For example, $(A \cdot T)_5$ is composed of two overlapping netropsin sites which are each four nucleotides long. One involves base pairs from i to $i + 3$ while the other involves base pairs $i + 1$ to $i + 4$. Base pairs $i + 1$ to $i + 3$ are occluded by binding to either site. On a given DNA restriction fragment only one of these situations can occur, so that binding to one tetrameric sequence excludes binding at the other overlapping tetrameric sequence on the same fragment.

Although McGee and von Hippel (1974) and Epstein (1978) have presented a treatment of such exclusion for a lattice of identical binding sites, not all overlapping sites have the same binding constant on the restriction fragment studied. In the case of netropsin, the exclusion obtains for binding regions of the type $(A \cdot T)_n$, where $n = 5-7$. To treat these, let K_j be the binding constant for a particular nucleotide position j , s be the fragment concentration (which is also equal to the con-

centration of any base pair j), s_0 be the concentration of binding regions of the type $(A \cdot T)_n$ with no drug bound, and s_j be the concentration of binding regions of the type $(A \cdot T)_n$ with drug bound at site j within the binding region. Then

$$s = s_0 + \sum_j s_j \quad (6)$$

and

$$K_j = s_j / D_0 s_0 \quad (7)$$

since the exclusion means that only a totally empty binding region can bind drug. The probability of having drug bound at site j is s_j/s . From eq 6 and 7 one can show

$$\frac{s_j}{s} = K_j D_0 \frac{1}{1 + \sum_k K_k D_0} \quad (8)$$

where the sum is over all the binding sites: two for a 5-mer, three for a 6-mer, and four for a 7-mer.

Now consider cutting at a base pair i which can be blocked by drugs bound at different tetramers of a binding region. If drug at any of the 4-mers of the binding region blocks cutting at i , the fraction of that base pair which is free, $1 - \nu_b$, is given by $1 - \sum_j (s_j/s)$. However, if only drug binding at sites p and q , say, stops cutting (because of the position of site i relative to the binding region)

$$1 - \nu_b = 1 - \frac{K_p D_0 + K_q D_0}{1 + \sum_k K_k D_0} = \frac{1 + D_0(\sum_k K_k - K_p - K_q)}{1 + \sum_k K_k D_0} \quad (9)$$

so that, in the numerator, we have $1 + D_0 \sum' K_j$, with $\sum' K_j$ being over all base pairs in the binding region for which drug binding does not affect cutting.

Determination of the Free Drug Concentration. All the expressions involve the free drug concentration rather than the total drug concentration, and it is only the latter that is known in the footprinting experiment. For an experiment with no carrier, or an experiment with a carrier that does not bind drug, the free drug concentration D_0 is obtained from the total drug concentration D_t by subtracting the concentration of drug bound to the restriction fragment. If we consider all sites where drug may bind (including sites for which drug binding is monitored directly via the footprinting)

$$D_t = D_0 + b$$

and

$$b = \sum_i s_i = s D_0 \sum_i K_i / (1 + K_i D_0) \quad (10)$$

assuming all sites are isolated sites. Therefore, D_0 can be obtained by solving

$$D_t = D_0 + s D_0 \sum_i \frac{K_i}{1 + K_i D_0} \quad (11)$$

For a binding region for which exclusion is a factor, $D_0 K_i / (1 + K_i D_0)$ is replaced by $D_0 (\sum_k K_k) / (1 + \sum_k K_k D_0)$.

If a carrier that binds drug (e.g., calf thymus DNA) is present in large excess, the free drug concentration D_0 is largely determined by the carrier. Let c be the carrier site concentration in base pairs. Supposing all the carrier sites to have the same binding constant K_c , we have $D_t = D_0 + cx$ and

$$cx / D_0 (c - cx) = K_c \quad (12)$$

where x is the fraction of carrier site bound by drug. Since the amount of drug bound to fragment is small compared to the amount bound to carrier and since c is only an effective

Table I: Equilibrium Binding Constants for Netropsin as a Function of Sequence ($\times 10^{-6}$ M)^a

nucleotide position	sequence of site	carrier DNA		
		poly-(dGdC)	none	calf thymus DNA
46	5'-TTTA-3' AAAT	1.9	7.0	5.5
47	5'-TTAT-3' AATA	10	68	87
56	5'-TTAA-3' AATT	4.2	1.8	3.4
57	5'-TAAA-3' ATTT	6	39	14.5
58	5'-AAAT-3' TTTA	35	164	167
59	5'-AATT-3' TTAA	35	70	131
89	5'-AAAT-3' TTTA	34	100	145
156	5'-TTAT-3' AATA	3.6	12	82

^aThe position of the lowest numbered nucleotide of a tetramer binding site is given.

site concentration, exclusions for sites of the type (A·T)_n, when $n = 5-7$, need not be considered. Although one could consider several kinds of sites on the carrier, with different binding constants, it does not seem necessary with our present data.

The preceding discussion shows how spot intensities for various sites can be calculated for a given total drug concentration, given values for the constants k_i (one for each nucleotide where binding or enhancement are measured), K_i (one for each four-nucleotide site where drug binding occurs), α , and K_c (if carrier, rather than the fragment, controls the free drug concentration). We determine these constants by seeking the values for which calculated and experimental intensities differ as little as possible. This is done by minimizing the sum of squared deviations as follows. Assuming a set of K_i , we solve eq 11 or 12 to obtain D_0 and b corresponding to each given D_i . Then ν_b is calculated for each nucleotide and each concentration. The optimum values of the k_i are found by minimizing the mean-square deviation between the experimental and calculated (eq 4) spot intensities for each base pair, assuming a value for α . We then search, using the Fletcher-Davidon-Powell simplex method (Fletcher, 1980), for the set of K_i and α (and K_c if necessary) which makes the sum of the squared deviations of theoretical from experimental spot intensities as small as possible.

RESULTS

From the fitting procedure, the cleavage rate constants at various nucleotide positions, the drug equilibrium constants, and a redistribution constant associated with DNase I were obtained for each experiment. In the tables given in the supplementary material (see paragraph at end of paper regarding supplementary material), calculated and experimental intensities are compared point by point. The equilibrium constants are given in Table I. Some calculated plots are compared with experimental plots in Figures 3 and 4. The difference in plots as one goes from site to site through a binding region, and the way in which the theoretical curves reproduce the experimental, should be noted.

Table I shows that while the binding constants for the experiments involving no carrier DNA and calf thymus DNA as carrier are in good agreement with each other, the values

Table II: Effect of Change in Base Pair Concentration on Derived Equilibrium Constants ($\times 10^{-7}$ M)^a

constant	fragment concn		
	1.5×10^{-6} M	1×10^{-6} M	6.67×10^{-7} M
K_{46}	0.6	0.7	0.7
K_{47}	8.2	6.8	5.9
K_{56}	0.2	0.2	0.2
K_{57}	4.1	3.9	3.1
K_{58}	19.6	16.4	14.5
K_{59}	7.7	7.0	6.1
K_{89}	13.5	10.0	8.2
K_{156}	1.4	1.2	1.1

^aThe data used are from the experiment involving no carrier DNA.

obtained in the presence of poly(dGdC) are significantly lower than those obtained in the former cases. However, the relative binding constants agree for the three experiments, thus allowing establishment of the order of the affinities of the various sites. The order found, in decreasing values of K , was (all 5' \rightarrow 3') AAAT (58, 89) > AATT (59) \sim TTAT (47, 156) > TAAA (57) > (46) \sim TTAA (56). Since site 156 appeared in a poorly resolved region of the autoradiogram and the intensity of only one band was used to monitor the loading event, Figure 2, the value of K for this site is less certain than the remaining values in Table I.

To determine how the concentration of the fragment would affect the values of K for the various sites, a series of minimization calculations in which the value of the fragment concentration was varied was carried out. The results of these determinations for the experiment involving no carrier DNA are given in Table II. As is evident from the table, significant changes in the concentration of the fragment have only a modest effect on the value of K .

DISCUSSION

Netropsin-DNA Binding Constants. The data in Table I reveal striking new information about the sequence dependence of the netropsin-DNA interaction. Consideration of the sequences of the sites with the lowest binding constants, TAAA (46, 57), TTAT (47), and TTAA (56) revealed that they contain the dinucleotide sequence TA. As has been previously outlined by Calladine (1982), this sequence gives rise to a strong clash in the minor groove between adenine bases on opposite polynucleotide strands. This clash may be relieved by a number of mechanisms, ultimately giving rise to a local distortion in DNA away from the idealized B form of the helix. Since netropsin utilizes specific hydrogen bonds and van der Waals contacts with the floor of the minor groove (Kopka et al., 1985), the distortions produced by the TA step very likely reduce optimal drug-DNA contact and with it the binding constant. The tetrameric sequences having the highest binding constants of about $\sim 10^8$ M⁻¹, AAAT (58, 89) and AATT (59), do not possess the sequence TA. These sites would be expected to more closely conform to the B-helical structure adopted by the AT core sequence in the duplex d-(CGCGAATTCGCG)₂ (Wing et al., 1980) and present optimal hydrogen bonding and van der Waals contacts toward netropsin. Although variable-temperature footprinting studies will be necessary to determine the contributions made by ΔH and ΔS to the free energy of binding, recent studies suggest that ΔS for the netropsin-DNA interaction may be relatively sequence independent (Marky & Breslauer, 1987). In this case, enthalpic effects, which have their origins in the different structural features of the tetrameric binding sequences, would appear to be the controlling factor in the netropsin-DNA interaction.

The effect that flanking bases may have on the binding constant of netropsin is difficult to discern from the data presented in Table I. For example, the sequence TAAA which occurs at positions 46 and 57 possesses different flanking bases in the two locations of the fragment. Ascertaining whether or not these differences in the netropsin binding constants are real, and if or to what extent flanking bases can influence netropsin binding to a core sequence, will require additional quantitative footprinting studies.

Although comparison of the binding constants in Table I with values determined by other methods is not possible in most cases, netropsin binding to the sequence AATT has recently been studied by calorimetry (Marky & Breslauer, 1987). This sequence, when present in the duplex d(GCGAATTCGC)₂, was found to bind netropsin with a binding constant of $2.84 \times 10^8 \text{ M}^{-1}$, with $[\text{Na}^+] = 16 \text{ mM}$, and $\partial \log K / \partial \log [\text{Na}^+]$ of -1.55 . The main cationic components in the footprinting experiments are 10 mM Ca^{2+} plus Mg^{2+} . Taking this as equivalent to 40 mM Na^+ , we estimate K for our conditions as $6.9 \times 10^7 \text{ M}^{-1}$, which is in good agreement with the values given in Table I for the sequence AATT.

The values in Table I do not in general agree with the order of binding affinity predicted from footprinting studies using poly(*N*-methylpyrrole)-containing peptides equipped with a DNA cleaving agent (Youngquist & Dervan, 1985). These studies indicated that the preference of hydrogen bonds between the three amide groups of netropsin and adjacent bases on opposite helix strands decreases in the order $\text{AT} \gg \text{AA} > \text{TT}$ for sites having only adenine and thymine. As is evident from Table I, sites AATT and TTAA possess netropsin binding constants which differ by 2 orders of magnitude. Since the bridged-base possibilities within the sites are similar (AT, TT, and TA for AATT; TA, AA, and AT for TTAA), it appears that factors other than the number and type of bridged base pairs are important in determining the affinity of the antiviral agent. As stated earlier, structural changes in DNA, due to adenine-adenine clashes within the minor groove, are the most likely cause for the observed differences in the affinities associated with the AT tetramers.

Scope, Limitations, and Errors Associated with Quantitative Footprinting Analysis. A number of sources of error in the quantitative footprinting experiment need to be considered. Although the number of pieces of data is, in general, much greater than the number of parameters to be calculated, there are some cases for which constants cannot be determined accurately. For instance, K_{156} is derived from intensity data associated with a single band, located at position 154 on the autoradiogram, Figure 2. Because only a single band is involved and the site is in a poorly resolved region of the autoradiogram, the determined value of K is subject to considerable error.

The fragment concentration was not specifically determined in the analysis. Its concentration was estimated from an initial determination of the plasmid concentration (using optical methods) and an assumed total recovery of the fragment in the labeling and isolation procedures. To determine the sensitivity of the determined value of K for the various sites, trial calculations using altered values of the fragment base pair concentration for the experiment involving no carrier DNA were carried out. The value assumed for all experiments, 10^{-6} M in base pairs or $7.2 \times 10^{-9} \text{ M}$ in fragment concentration, was first increased by half and then decreased by one-third. The results, given in Table II, show that there are only very slight changes in equilibrium constants when the fragment concentration is altered. The change can only be small as long

as the maximum concentration of bound drug (fragment concentration times number of sites per fragment which can be occupied simultaneously) is not comparable to the total drug concentration. This is realized for our experiments.

One should also recall that, in the studies involving calf thymus DNA as carrier, the fragment concentration does not directly enter into the determination of the equilibrium constants associated with the fragment. In this case, the carrier, which is in a greater than 100-fold excess over the fragment, controls the concentration of free drug. In the studies involving calf thymus DNA, the concentration of strong netropsin sites on the carrier was assumed to be 10^{-6} M , and the effective binding constant K_c to a site on the carrier was taken as an additional parameter to be found in the minimization procedure. A value of K_c for the carrier of $1 \times 10^6 \text{ M}^{-1}$ was found to give the best fit between theory and experiment for the data. In fact, only the product of K_c and the number of binding sites on the carrier is significant.

For the experiments involving poly(dGdC) as carrier, it was assumed that no drug binding to the carrier takes place. However, recent studies have shown that the binding constant of netropsin toward poly(dGdC) is $\sim 5 \times 10^5 \text{ M}^{-1}$ (Marky & Breslauer, 1987). This value of K_c , along with the fact that the carrier concentration is about 2 orders of magnitude greater than that of the fragment, may mean that binding to the carrier cannot be ignored in the analysis. This would mean that the free drug concentrations were actually lower than what we calculated by assuming no binding to the carrier. The use of free drug concentrations higher than the actual ones would give low values for all of the binding constants. As Table I indicates, the constants found from poly(dGdC) experiments are in fact much lower than those from the other experiments. Assuming an effective binding constant of 10^5 M^{-1} for drug on poly(dGdC) and using the known concentration of sites on this carrier ($35 \mu\text{M}$), we used the experimental data in a fitting procedure to determine the binding constants for the fragment. We indeed found binding constants of the same size as those found from the no carrier and calf thymus DNA carrier experiments. We also attempted to use the effective binding constant to poly(dGdC) as a variational parameter and find the value which gave the best agreement between theoretical and experimental data by minimizing the deviation. The search was not successful, multiple minima being found. This is because, for a wide range of values for the effective binding constant, there exists a set of K_i which can produce agreement of theoretical binding curves with our experimental data, within the errors of the latter.

In the case of the experiments using calf thymus DNA as carrier, the minimization algorithm converged. We did not find solutions (sets of equilibrium binding constants K_i which lead to minimum deviation between theoretical and experimental intensities) for which the K_i were very different from values reported. Further, as shown in Table I, the K_i values are in good agreement with those obtained from the no carrier experiment. Apparently, the range of acceptable values for K_c , the effective binding constant to the carrier, is not so wide in this case, and/or there is a small correlation between K_c and K_i , the site binding equilibrium constants to the fragment, in this range of K_c . The latter would occur, for example, if K_c were small, so as to make D_0 similar to D_i for most of the drug concentrations investigated.

The possibility that drug binding to the fragment occurs in a cooperative manner was not addressed in the study. Unlike numerous intercalating agents, netropsin binds without greatly distorting the DNA helix (Kopka et al., 1985). This obser-

vation and the fact that the initial loading events on the 139-mer are generally well separated from each other strongly suggest that binding, at least for the strong sites, occurs in an independent noncooperative fashion, and thus the analysis presented here is valid.

There may also be an error connected with the correction for enhancement (Ward et al., 1987, 1988) made according to eq 3 in the determination of binding constants. While the enhancement was generally small and relatively uniform over long segments of DNA remote from drug binding, it appeared to be 2–3 times greater at sites adjacent to a netropsin binding site on DNA. The reason for this is unclear, but since netropsin does not greatly distort DNA, it does not appear to be structural in origin. The presence of this effect would also assign netropsin sites 56 and 46 a larger enhancement factor than, for example, that of site 89 and lead to an increase in the calculated binding constants of the weak sites at positions 46 and 56. However, the increase would be modest ($<1/3$), so the large differences found between the values of K for these sites and those of the strong sites appear to be due to sequence-dependent DNA structural effects rather than to artifacts associated with the enhancement.

In the analysis, seven nucleotide positions were assumed to be equally affected by the binding of a single netropsin molecule to DNA, except for the isolated site at 89–92, because the footprinting plots for the region 85–92 show that site 85 exhibits partial protection by the drug which is bound at 89–92. Thus, in this case the geometry of the enzyme may block cleavage four base pairs to the 3' side of a drug binding site rather than three, yielding a protection region eight nucleotides long rather than seven. If we assume the protection region is eight nucleotides long everywhere, we find significantly poorer agreement between the experimental and calculated intensities (sum of squared deviations). This is because, if the protection region were eight nucleotides in length, the intensities for site 55 would decrease with drug concentration in exactly the same way as for sites 57, 58, and 59, and this is clearly not borne out by the data (see supplementary material). Thus one must assume seven sites are protected here. These observations suggest that the protection region may be controlled not only by the size of the drug and the geometry of the cleavage agent but, to a certain extent, also by the sequence at which binding is taking place.

The data for site 85 suggest that the protection is less than that observed at sites 87, 90, and 92. This may be taken into account by replacing eq 2 by

$$(\text{rate})_i = k_i[\text{DNase I}]_i p_i + k''_i[\text{DNase I}]_i (1 - p_i)$$

where k''_i is a cleavage rate constant for a blocked site, as was done by Brenowitz et al. (1986a). However, this would significantly increase the number of variable parameters leading to a decrease in the significance of the values obtained.

The method described is independent of the magnitudes of the association/dissociation rates of the ligand and the cleavage rate of the probe (Goodisman & Dabrowiak, 1986). If the ligand–DNA interaction is at equilibrium before the addition of the probe and remains so during the time of the digest, the concentrations of oligonucleotide products, after correction for any alterations in the probe–DNA equilibrium, reflect the fraction of a particular site which is occupied by ligand. Since a DNA cleaving probe is being used to report the amount of a particular site which is occupied by ligand, the possibility exists that the probe itself shifts the ligand–DNA equilibrium and that a competitive equilibrium is being measured. At least for DNase I and the strong netropsin binding sites on the fragment this does not appear to be the case. Although the

enzyme/DNA ratios were different in the calf thymus DNA and no carrier experiments, both experiments returned the same binding constants, indicating that the enzyme is not disturbing the drug–DNA equilibrium.

CONCLUSIONS

In this paper we have presented the theory for quantitative analysis of a drug–DNA footprinting experiment. The approach incorporates enhancements due to redistribution of DNase I on DNA, as well as exclusion associated with overlapping binding sites on DNA. By use of the theory with the autoradiographic spot intensities obtained from three sets of footprinting experiments involving a 139 base pair restriction fragment of pBR-322 DNA, binding constants for the antiviral agent netropsin as a function of sequence were determined. The analysis revealed that the presence of the dinucleotide sequences TA within sites of the type (A·T)₄ greatly discourages netropsin binding to DNA. This effect very likely has its origins in a strong clash between adenines on opposite polynucleotide strands, which produce an altered DNA structure not optimal for netropsin binding. The scope, limitations, and errors associated with the new method were also presented and discussed.

SUPPLEMENTARY MATERIAL AVAILABLE

Tables I–III giving the experimentally determined and calculated concentration (areas) a_i values for the various experiments (3 pages). Ordering information is given on any current masthead page.

REFERENCES

- Becker, M. M., & Wang, J. C. (1984) *Nature (London)* 309, 682.
- Brenowitz, M., Senear, D. F., Shea, M. A., & Ackers, G. K. (1986a) *Methods Enzymol.* 130, 132–181.
- Brenowitz, M., Senear, D. F., Shea, M. A., & Ackers, G. K. (1986b) *Proc. Natl. Acad. Sci. U.S.A.* 83, 8462–8466.
- Calladine, C. R. (1982) *J. Mol. Biol.* 161, 343–352.
- Dabrowiak, J. C. (1983) *Life Sci.* 32, 2915–2931.
- Dabrowiak, J. C., Skorobogaty, A., Rich, N., Vary, C. P. H., & Vournakis, J. N. (1986) *Nucleic Acids Res.* 14, 489–499.
- Dervan, P. B. (1986) *Science (Washington, D.C.)* 232, 464–471.
- Drew, H. R. (1984) *J. Mol. Biol.* 176, 535–557.
- Epstein, I. R. (1978) *Biophys. Chem.* 8, 327–339.
- Fletcher, R. (1980) *Practical Methods of Optimization*, pp 14–15, Wiley, Chichester, U.K.
- Galas, D. J., & Schmitz, A. (1978) *Nucleic Acids Res.* 5, 3157–3170.
- Goodisman, J., & Dabrowiak, J. C. (1985) *J. Biomol. Struct. Dyn.* 2, 967–980.
- Johnson, A., Meyer, B., & Ptasne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76, 5061–5065.
- Kopka, M. J., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 1376–1380.
- Kuwabara, M., Yoon, C., Goynes, T., Thederahn, T., & Sigman, D. S. (1986) *Biochemistry* 25, 7401–7408.
- Lane, M., Dabrowiak, J. C., & Vournakis, J. N. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 3260–3264.
- Low, C. M. Y., Drew, H. R., & Waring, M. J. (1984) *Nucleic Acids Res.* 12, 4865.
- Marky, L. A., & Breslauer, K. J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 4359–4363.
- McGhee, J. D., & von Hippel, P. H. (1974) *J. Mol. Biol.* 86, 469–489.

- Ogata, R. T., & Gilbert, W. (1979) *J. Mol. Biol.* 132, 709.
- Sakonju, S., & Brown, D. (1982) *Cell (Cambridge, Mass.)* 31, 394-405.
- Scamrov, A. V., & Beabealashvili, R. S. (1983) *FEBS Lett.* 164, 97-99.
- Schmitz, A., & Galas, D. (1979) *Nucleic Acids Res.* 6, 111-137.
- Senear, D. F., Brenowitz, M., Shea, M. A., & Ackers, G. K. (1986) *Biochemistry* 25, 7344-7354.
- Suck, D., & Oefner, C. (1986) *Nature (London)* 321, 620-625.
- Tullius, T. D., & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 5469-5473.
- Van Dyke, M. W., Hertzberg, R. P., & Dervan, P. B. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 5470-5474.
- Ward, B., & Dabrowiak, J. C. (1987) *J. Am. Chem. Soc.* 109, 3810-3811.
- Ward, B., Rehfuess, R., & Dabrowiak, J. C. (1987) *J. Biomol. Struct. Dyn.* 4, 685-695.
- Ward, B., Rehfuess, R., Goodisman, J., & Dabrowiak, J. C. (1988) *Nucleic Acids Res.* (in press).
- Wing, R., Drew, A., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London)* 287, 755-758.
- Winter, R. B., & von Hippel, P. H. (1981) *Biochemistry* 20, 6948-6960.
- Youngquist, R. S., & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 2565-2569.

Properties of the High-Affinity Single-Stranded DNA Binding State of the *Escherichia coli* RecA Protein[†]

Joseph P. Menetski, Abraham Varghese, and Stephen C. Kowalczykowski*

Department of Molecular Biology, Northwestern University Medical School, Chicago, Illinois 60611

Received July 28, 1987; Revised Manuscript Received October 21, 1987

ABSTRACT: The properties of the high-affinity single-stranded DNA (ssDNA) binding state of *Escherichia coli* recA protein have been studied. We find that all of the nucleoside triphosphates that are hydrolyzed by recA protein induce a high-affinity ssDNA binding state. The effect of ATP binding to recA protein was partially separated from the ATP hydrolytic event by substituting calcium chloride for magnesium chloride in the binding buffer. Under these conditions, the rate of ATP hydrolysis is greatly inhibited. ATP increases the affinity of recA protein for ssDNA in a concentration-dependent manner in the presence of both calcium and magnesium chloride with apparent K_d values of 375 and 500 μ M ATP, respectively. Under nonhydrolytic conditions, the molar ratio of ATP to ADP has an effect on the recA protein ssDNA binding affinity. Over an ATP/ADP molar ratio of 2-3, the affinity of recA protein for ssDNA shifts cooperatively from a low- to a high-affinity state.

The recA protein has been shown to bind to single-stranded DNA (ssDNA)¹ (McEntee et al., 1981; Silver & Fersht, 1982; Menetski & Kowalczykowski, 1985). The affinity of recA protein for ssDNA is influenced by the presence of ATP, ADP, and NaCl (Menetski & Kowalczykowski, 1985). Increasing the concentration of ADP and NaCl decreases the affinity, while the presence of ATP increases the affinity of recA protein for ssDNA. Weinstock et al. (1981) have shown that recA protein can hydrolyze other nucleoside triphosphates in addition to ATP. In this paper, we have extended our previous analysis of the modulation of recA protein ssDNA binding affinity by nucleotide cofactors to include those nucleotides having significant concentrations in vivo. The data show that all nucleotides which are hydrolyzed by recA protein induce a high ssDNA binding affinity state similar to that induced by ATP.

The ATP-induced high-affinity state was first observed (Silver & Fersht, 1982) and characterized (Menetski &

Kowalczykowski, 1985) under conditions which support the ATPase activity of recA protein. Therefore, the data reflect only the steady-state effect of ATP on recA protein ssDNA binding affinity. Under these conditions, the individual effects of binding and hydrolysis of ATP cannot be clearly separated. Ideally, the two events could be separated under conditions that prevent hydrolysis but not nucleotide binding. Weinstock et al. (1981) have reported that in the presence of CaCl_2 , the rate of ATP hydrolysis is less than 10% of that observed in MgCl_2 . We have found that ATP is indeed bound to recA protein in the presence of CaCl_2 and that ATP induces a high ssDNA binding affinity state similar to the one observed in MgCl_2 . This result shows that only ATP binding, and not hydrolysis, is required to induce the high-affinity state. Under conditions of reduced hydrolysis, the affinity of all recA protein molecules for ssDNA is sensitive to the ATP/ADP molar ratio in the buffer. As the ATP/ADP molar ratio decreases, the

[†] This research was funded, in part, by a grant from the National Institutes of Health (AI-18987) and an American Cancer Society Junior Faculty research award (JFRA-70) to S.C.K. and by a predoctoral training grant from the National Institutes of Health (GM 08061) to J.P.M.

¹ Abbreviations: etheno M13 DNA, modified ssDNA containing 1,*N*⁶-ethenoadenosine and 3,*N*⁴-ethenocytosine; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; ATP_γS, adenosine 5'-*O*-(3-thio-triphosphate); poly(dA), poly(deoxyadenylic acid); RFI, relative fluorescence increase; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; SSB protein, *E. coli* single-stranded DNA binding protein.