SPECTROSCOPIC ANALYSIS OF DRUG-NUCLEIC ACID INTERACTIONS

Authors: **Geoffrey Dougherty***
School of Physics
Universiti Sains Malaysia
Penang, Malaysia

William J. Pigram
Department of Physics
University of Keele
Keele, England

Referee: Michael J. Waring
Department of Pharmacology
University of Cambridge Medical School
Cambridge, England

I. INTRODUCTION

There exists a family of chemical compounds of biomedical importance, commonly referred to as drugs, which possesses extended heterocyclic aromatic chromophores. In general, these drugs interact with both deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) and interfere with nucleic acid synthesis.[1] Among the numerous members of this family, one may include the phenanthridine drugs dimidium and ethidium bromide,[2] the acridines such as proflavine,[3] the antitumour antibiotics daunomycin, adriamycin, and nogalamycin,[4,5] the schistosomicidal drug miracil D,[6,7] and the antimalarial drug chloroquine.[8]

There is strong evidence that many of these planar, aromatic drugs interact with DNA by the intercalation of their chromophores between adjacent base pairs of the double helical DNA molecule. Stereochemical models have been proposed for the intercalation of ethidium bromide,[2] actinomycin D,[9,39] and daunomycin.[10] In addition, there is mounting evidence that these drugs may also bind to nucleic acids in a nonintercalative manner and that the two binding modes, i.e., intercalation and external binding, may coexist.[11-15]

The visible absorption spectra of drugs which possess extended heterocyclic chromophores alter when they are bound to nucleic acids. In particular, the bound drug spectrum is shifted to longer wavelengths and its maximum absorption is reduced in magnitude. This behavior is not diagnostic for a particular type of binding, since both intercalated and externally bound drug molecules display the same characteristics,[16] but can be used to explore the affinity and specificity of the drug binding.

The total concentration of bound drug in a drug-nucleic acid mixture can be estimated by spectrophotometric methods,[17,17a] equilibrium dialysis,[18] ultracentrifugation,[19] or the solute-enhanced partition analysis method.[20] Solution spectra taken at a series of binding ratios, r (the number of drugs bound per base pair[18]), can be analyzed in most instances as a linear combination of two extreme or library spectra. These correspond to the cases of a completely free drug and a drug completely bound to the DNA, and the concentrations of

* Manuscript prepared while at Institut für Zellbiologie, ETH — Hönggerberg, Zurich, Switzerland.

the two components may be computed using a weighted least-squares minimization procedure applied to the digitized absorbances.[17,21,22] These data can then be interpreted using a particular binding model to yield the binding parameters of the drug-nucleic acid interaction, such as the association constant and the number of binding sites associated with each base pair.

This review deals with the treatment and analysis of drug binding to nucleic acids. Although certain drugs are cited to illustrate these ideas, we do not attempt to survey individual drug properties. Our aim is to provide a perspective on the problems associated with the analysis and subsequent interpretation of drug binding data.

II. THE ABSORPTION SPECTRA OF DRUG-NUCLEIC ACID COMPLEXES

Spectral series of drug-nucleic acid complexes in solution are recorded at different binding ratios. Different mixing schemes[23-28] have been used in the past, and the authors have proposed[29] a novel method which incorporates all the advantages of the various earlier schemes. It is generally convenient to keep the total drug concentration constant, and this is achieved either experimentally or by normalization of the spectra.

Within each series there is a progressive shift in the absorbance maximum,* generally towards longer wavelengths with increasing nucleic acid content, and the absorption at the maximum is depressed. These spectral changes provide a convenient means of assessing and quantitating complex formation.[11,12,17] In many cases, all the spectra in a series pass through a common point called an isosbestic point. The most common circumstance under which this obtains is when there are only two forms of the drug, in this case "free" and "bound" components. If these have the same absorption at some wavelength then shifts in the relative proportions of each will produce no change in absorbance at that wavelength, for a constant total drug concentration (if the DNA has negligible absorptivity in this wavelength region). However, the existence of isosbestism is not proof that only two forms are present. There could be more than one form of bound drug (e.g., intercalated and externally bound) having different spectra, but if these are always present in the same ratio an isosbestic point will be observed. Alternatively, two different bound species may be present, each producing shifted spectra so similar that they are indistinguishable within experimental error.

The converse behavior allows a more definite conclusion. If there is not an isosbestic point, even though the spectra of complexed and free drug cross, then there must be more than one form of binding. This is the case for proflavine,[14] where is is interpreted as demonstrating intercalative binding at low binding ratios and additional external stacking on the sugar-phosphate backbone when the drug concentration is high.

A. Analysis of Data

Figure 1 shows schematically a two-component case (i.e., a free and a single bound species) in which the free drug spectrum is denoted "library 1", the bound spectrum is called "library 2", and spectra taken at intermediate values of r are the "mixture" spectra.

At any wavelength λ,

$$A = a_1 c_1 + a_2 c_2 \qquad (1)$$

* The spectral changes can be explained[30] on the basis of excitation interaction between the transition dipole corresponding to the visible band of the drug and that associated with the UV band of a polynucleotide base.

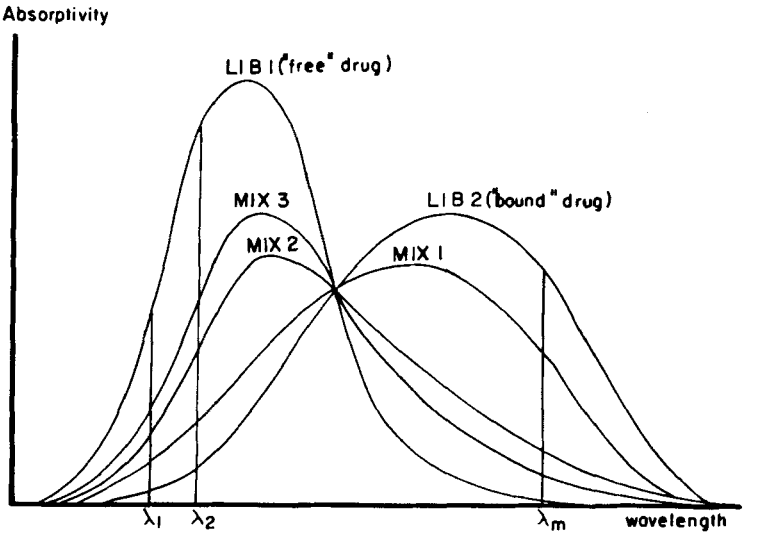


FIGURE 1. Schematic representation of a mixture spectrum in terms of two library spectra.

where A is the absorbance of the mixture, a_1, a_2 are the absorptivities of the free and bound drugs, and c_1, c_2 are the concentrations of free and bound components in the mixture. The concentrations c_1 and c_2 can be found by considering the spectra at any one wavelength, λ, since the total drug concentration, c, is constant and known.[17]

By proportions,

$$\text{the fraction bound } \left(= \frac{c_2}{c}\right) = \frac{|A_1 - A|}{|A_1 - A_2|} \qquad (2)$$

where A_1, A_2, and A are the absorbances of library 1, library 2, and the mixture spectrum, respectively, at that wavelength. The best wavelength at which to perform these measurements would be that at which the proportional difference between A_1 and A_2 $\left(\text{viz. } \frac{|A_1 - A_2|}{A_1 + A_2}\right)$ is greatest, since this will minimize the effect of experimental error. The computation will be most reliable for the mixture whose spectrum is midway between the two library spectra, since in that case the numerator and denominator of Equation 2 will be of similar magnitude.

An extension of this analysis would be to use the entire spectra, rather than readings at just one wavelength, to improve the reliability of the values for c_1 and c_2.

Thus, using m wavelengths,*

$$\begin{aligned} A^O(\lambda_1) &= a_{11}c_1 + a_{21}c_2 + \delta_1 \\ A^O(\lambda_2) &= a_{12}c_1 + a_{22}c_2 + \delta_2 \\ &\vdots \quad\quad \vdots \quad\quad \vdots \quad\quad \vdots \\ A^O(\lambda_m) &= a_{1m}c_1 + a_{2m}c_2 + \delta_m \end{aligned} \qquad (3)$$

* Note that a reading at the isosbestic wavelength gives no information on c_1 and c_2 since

$$\begin{aligned} a_{1i} &= a_{2i} \\ A(\lambda_1) &= a_{1i}(c_1 + c_2) \\ &= a_{1i}c \end{aligned}$$

where $\delta_1, \delta_2 \ldots \delta_m$ are the experimental errors and the superscript "O" denotes "observed" (i.e., experimental) absorbances. The problem becomes one of selecting the best values for c_1 and c_2, and the criterion adopted is to minimize

$$\phi = \sum_{j=1}^{m} \delta_j^2 = \sum_{j=1}^{m} (A_j^o - A_j^c)^2 \tag{4}$$

where the superscript "c" refers to "calculated" absorbances, using the formula:

$$A_j^c = a_{1j}c_1 + a_{2j}c_2$$

As mentioned previously, values of c_1 and c_2 derived at wavelengths where the two library spectra differ by a large amount will be more reliable so that, in this generalized scheme, it is judicious to weight contributions to the final solution by a term proportional to the magnitude of this difference,

$$w_j = |a_{1j} - a_{2j}|^N \tag{5}$$

We consider a first-order dependence appropriate, i.e.,

$$w_j = |a_{1j} - a_{2j}| \tag{5a}$$

because of the form of the dependence of the concentrations on the absorptivities (Equation 2), and because there is an approximately constant random error associated with each observed and digitized absorbance.

A standard least-squares minimization treatment results in the following solution:

$$c_1 = \frac{(\Sigma w_j A_j a_{1j})(\Sigma w_j a_{2j}^2) - (\Sigma w_j a_{1j} a_{2j})(\Sigma w_j A_j a_{2j})}{\Delta} \tag{6}$$

$$c_2 = \frac{(\Sigma w_j a_{1j}^2)(\Sigma w_j A_j a_{2j}) - (\Sigma w_j A_j a_{1j})(\Sigma w_j a_{1j} a_{2j})}{\Delta} \tag{7}$$

where $\Delta = (\Sigma w_j a_{1j}^2)(\Sigma w_j a_{2j}^2) - (\Sigma w_j a_{1j} a_{2j})^2$, and the summations are performed from $j = 1$ to m.

A computer program was written to find the concentrations c_1 and c_2, using the library and mixture spectra. The program is an extension of one used by Porumb[21] with added options including weighting of the data in accordance with Equation 5 and computer graphic output.

For each mixture the program plots the digitized observed and calculated spectra, and determines the concentrations, c_1 and c_2. Results can be obtained by utilizing the entire wavelength range, or a selected part of it. The difference between the total drug concentration, c, and the sum of the computed concentrations, $c_1 + c_2$, will reflect the total experimental errors. The discrepancies, $\Delta A (= A_j^o - A_j^o)$, for each mixture spectrum are plotted against wavelength. Inspection of these plots can be used to give useful information on the error structure of the data.[31] In particular they are helpful in detecting the presence of a baseline shift (i.e., a systematic error) in the mixture spectra. The presence of an additional species in the system (either a second bound mode, or a contribution from DNA absorption) would be evident from persistent systematic effects in a spectral series in the wavelength region where this species absorbs.

The program was tested rigorously using Gaussian library spectra and calculated

"observed" mixture spectra modified by random and systematic errors in excess of those likely to be encountered experimentally. The weighted scheme deals with random errors slightly better, and with baseline (systematic) errors a little worse, than the unweighted scheme, although the differences are small. Since baseline errors can be recognized from the discrepancy plots,[22,31] and compensated for if necessary, the weighted program was adopted for all further work.[22,29] Different interpretations of the binding process predict various relationships between the binding parameters (see following section), and the program produces graphs appropriate to these relationships.

III. INTERPRETATION OF THE DATA

A. Simple Single-Species Binding

If there is only one bound species the reaction may be described by

$$\underset{\text{(free sites)}}{\text{DNA}} + \underset{\text{(free drug)}}{\text{DRUG}} \rightleftharpoons \underset{\text{(bound drug)}}{\text{COMPLEX}}$$

Assuming that the binding sites are independent, the law of mass action gives

$$k_a = \frac{[\text{bound drug}]}{[\text{free sites}][\text{free drug}]}$$

where k_a is the association constant (an equilibrium constant, not a rate constant) and the square brackets denote concentrations.

Defining the following quantities, c_1 = free drug concentration; c_2 = bound drug concentration; $c = c_1 + c_2$ = total drug concentration; (BP) = total concentration of DNA base pairs; and n = number of binding sites per base pair, then:

$$k_a = \frac{c_2}{(n(BP) - c_2)c_1}$$

The bound drug concentration, c_2, can be replaced by r(BP), where r is the binding ratio, giving

$$k_a = \frac{r}{(n - r)c_1}$$

Rearranging,

$$\frac{r}{c_1} = k_a(n - r) \qquad (8)$$

For the evaluation of the parameters n and k_a, several linear transformations of this equation have been used (Klotz and Hunston;[32] Scatchard[33]) and three of these are shown in Figure 2. With transformation (i), $1/r$ vs. $1/c_1$, the value of $1/n$ corresponds to the value of $1/r$ when $1/c_1 = 0$. The extrapolation required is from the largest experimental value of c_1 to $c_1 = \infty$, which is a much larger extrapolation than is apparent from the graphical plot. Transformation (ii), c_1/r vs. c_1, has been used rarely and transformation (iii), r/c_1 vs. r, leads to the so-called Scatchard plot which has enjoyed widespread usage. Despite being commonly employed in binding data analysis, the

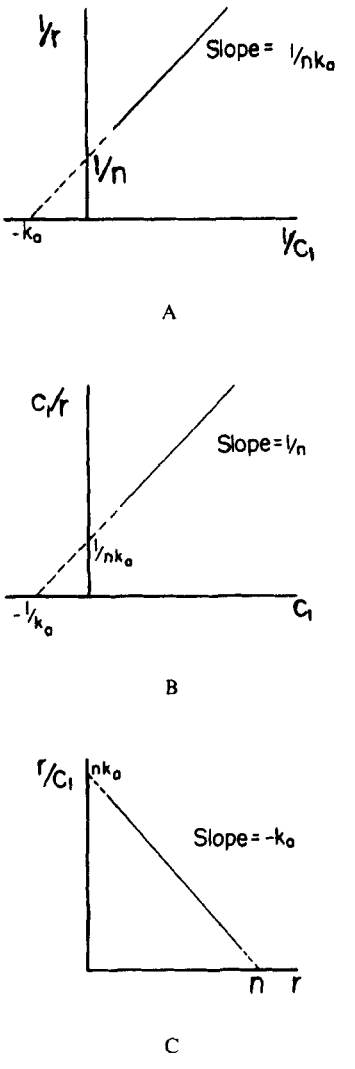


FIGURE 2. Graphical representation of three possible linear transformations of the binding equation for a single binding species (see Equation 8).

Scatchard plot has been widely misinterpreted in the literature as discussed by Nørby et al.[34]

The association constant, k_a, is related to the standard free energy of the binding by the expression

$$k_a = e^{-\frac{\Delta G}{RT}}$$

where ΔG is the change in the standard free energy, and is a negative quantity for binding, R is the gas constant, and T is the temperature (in K).

An energetically strong reaction will therefore result in k_a being large, leading to a steep negative gradient on the Scatchard plot, and a weaker reaction will result in a shallow gradient. A typical case of strong binding, such as the binding of acridines at low binding ratios to DNA, results in a value for ΔG between -6 and -9 kcal/mol.[14]

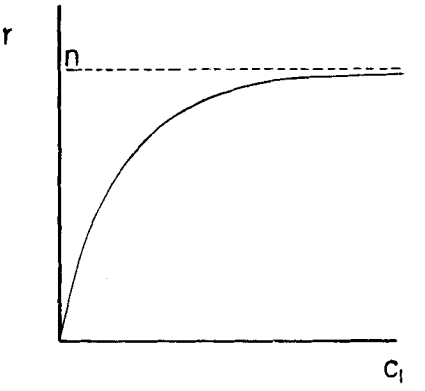


FIGURE 3. A schematic binding curve for single-species binding.

A binding isotherm (i.e., a plot of r against c_1) is another useful graph, especially since it presumes no special relationship between r and c_1. For the case of simple single-species binding its form is shown in Figure 3. The initial slope is large for a strong binding reaction and the value of r asymptotically approaches n, the maximum number of binding sites available per base pair at large values of c_1.

At low binding ratios, errors in r/c_1 will be prominent and at high binding ratios errors in r are more important. In practice many points are required at the extremities of the plot if accurate values of the association constant, k_a, and the number of binding sites per base pair, n, are to be determined. A plot of r vs. c_1 will typically have the form shown in Figure 3, with r increasing to a maximum, in the case of a single binding mode, approaching n. The quantity r rises as the bound drug occupies sites on the DNA until eventually it approaches a limit, when all the available binding sites on the DNA become filled. The estimation of c_1 will tend to be inaccurate when it is small, i.e., for mixture spectra which are difficult to differentiate from library 2. Similarly, c_2 (and hence r) is more inaccurate for mixtures that are close to library 1.

In an experimental situation the Scatchard plot is often curved when r is moderately large. The value of n obtained by extrapolation of the linear region is frequently less than unity even for homopolymers. For synthetic DNA containing a single type of base pair only, at each base pair either binding (of an integral number of drugs) or no binding should occur. For the homopolymers which bind the drug the value of n obtained from the Scatchard plot should be integral or 1/integral. Other values of n imply that this simple theory is inadequate in describing the prevailing physical situation.

Various more complex theories have been advanced to account for the curvature of experimental Scatchard plots. There may be more than one type of binding operating, resulting in more than one bound species. Interactions between drugs could be envisaged which could either favor the binding of additional drugs or discourage further binding. Binding at a particular base pair might preclude binding at a neighboring site, perhaps due to the steric hindrance of a side group of the initially bound drug molecule. Binding specificity could operate with a particular drug whereby the drug preferentially binds to a particular base pair or to a particular sequence of base pairs. These possibilities are discussed in the following sections.

B. More Than One Independent Bound Species

Equation 8 applies for only one species of bound drug and can be rearranged as

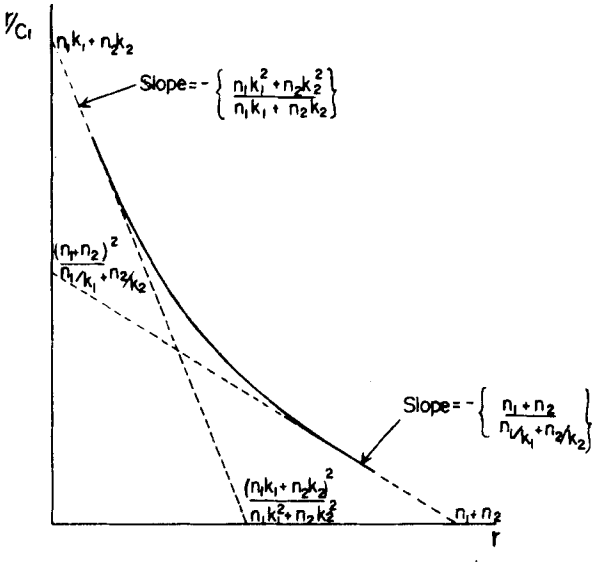


FIGURE 4. Schematic curve of r/c_1, vs. r for two independent binding modes, 1 and 2, showing the intercepts and limiting slopes.

$$r = \frac{nk_a c_1}{1 + k_a c_1}$$

If the drug can bind by m spectrally identical different binding modes at each base pair, and the number of binding sites per base pair for mode i is given by n_i with a corresponding association constant k_i, then the binding ratio is given by

$$r = \sum_{i=1}^{m} \frac{n_i k_i c_1}{1 + k_i c_1} \tag{9}$$

The total number of sites per base-pair is then given by

$$n = \sum_{i=1}^{m} n_i$$

As m increases, the number of binding parameters increases as well. Given that the experimental data are not perfect it becomes increasingly difficult to obtain precise solutions for these parameters. Considering the case where $m = 2$, a Scatchard plot would result in a shape similar to that shown in Figure 4. This can be treated as being comprised of two approximately linear regions separated by a curved portion. The data in these two linear regions can be analyzed to yield two gradients and four intercepts, all of which are generally complex functions of n_i and k_i. The only parameter directly obtainable is $n(= n_1 + n_2)$, the total number of binding sites per base pair, although experimentally this is often difficult to determine accurately. Reliable data at large r are difficult to obtain because of aggregation or precipitation of the drug, and in this region the plot tends to approach the axis asymptotically.

It should be noted that the apparent intercepts and gradients associated with the two linear regions are not what one might expect from an intuitive generalization of the single binding site case. In fact, the two regions are only separable as such when the binding parameters involved, k_1 and k_2, are appreciably different. Consider the case where

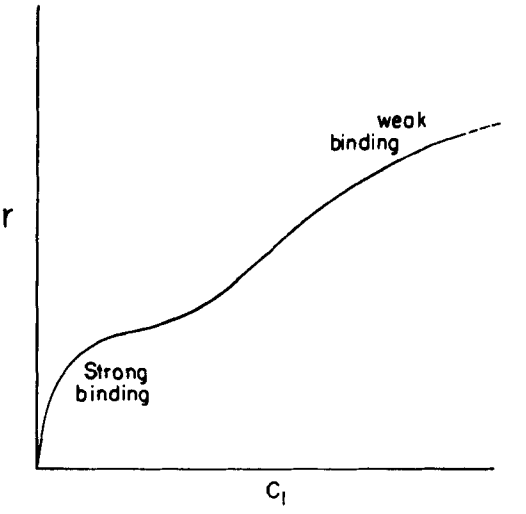


FIGURE 5. The binding curve when two independent binding modes are present.

$n_1 = n_2 = 1$. If $k_1/k_2 = 10$ then an 8% error would result if the slope of the stronger binding region were taken as $-k_1$, and there would be an error of 9% in taking the slope of the weaker binding region as $-k_2$. If $k_1/k_2 = 100$, the slope of the strong binding region could be taken as $-k_1$ and result in only 1% error, but if the slope of the weak binding region were taken as $-k_2$ a 98% error would result. This latter case would hardly arise in practice since if k_2 were so small, the analysis would have been carried out presuming only single-species binding. The situation is more complicated if n_1 and n_2 are not known explicitly, which is usually the case.

Spectroscopic data obtained from proflavine binding to DNA have been analyzed according to a model assuming two independent bound species.[17] Peacocke and Skerrett took the gradients of the two linear portions as being equal to $-k_1$ and $-k_2$, and obtained values corresponding to a ΔG of between -6 and -9 kcal/mol for the strong binding mode, and between -2 and -3 kcal/mol for the weak binding mode. The stronger binding mode was later associated with the intercalation mechanism when related to results obtained from X-ray diffraction studies.[2,3] The weaker binding was considered to be an external binding scheme with a strong electrostatic character, since it was found to be more dependent on the ionic environment of the complex.

In a particular analysis if the intercept, Σn_i, is less than m, and the individual values n_i do not seem to correlate with any structural features of possible DNA-drug models, then an alternative binding model may have to be considered. The binding curve, r vs. c, is often helpful in deciding whether or not a model with more than one bound species is appropriate. For the case of two bound species, in which k_1 and k_2 are well separated, the binding curve will exhibit the characteristic shape illustrated in Figure 5. A region of near saturation binding, associated with an initially dominant strong binding mode, is followed by a slower but definite rise in r at higher c, associated with a weaker binding mode.

C. Cooperatively and Anticooperativity

Electrostatic interactions between bound cationic drug molecules may have the effect of discouraging further binding at nearby sites (anticooperativity). However, in some cases it seems that attractive Van der Waal forces may overshadow electrostatic effects

and favor the attachment of further drug molecules at adjacent sites (cooperativity).

For the very simple case in which interactions of bound drugs involving only nearest neighbors are considered, and neither the cooperativity nor the binding constant is affected by any change in the DNA secondary structure caused by the binding of the drugs, the following relationship has been proposed (after Fowler and Guggenheim[35]):

$$\lambda\chi = \left(\frac{\beta - 1 + 2\theta}{\beta + 1 - 2\theta}\right) e^{\frac{W}{kT}} \tag{10}$$

where λ = the absolute activity of the drug; χ = the partition function for a molecule of bound drug; θ = the fraction of occupied sites (i.e., r/n); W = the free energy of interaction of a nearest neighbor pair of bound drug molecules; k = Boltzmann's constant; and $\beta = \{1 - 4\theta\,(1 - \theta)\,(1 - e^{-W/kt})\}^{1/2}$

For dilute solutions the left-hand side of Equation 10 can be replaced by $k_a c_1$ to give

$$k_a c_1 = \left(\frac{\beta - 1 + 2\theta}{\beta + 1 - 2\theta}\right) e^{\frac{W}{kT}} \tag{11}$$

If nearest neighbor interactions are negligible ($W \approx 0$) this reduces simply to Equation 8, as expected.

If the forces of interaction are purely electrostatic in origin, further simplications to Equation 11 lead to a binding equation of the form

$$\frac{r}{c_1} = k' k_a (n - r) \tag{12}$$

where k' is an interaction coefficient which is related to the ionic charges and the dielectric constant of the medium, but it can be more conveniently treated as an adjustable empirical parameter. For instance, if

$$k' = e^{-\frac{\Delta G^\circ(r)}{RT}}$$

then $\Delta G^\circ(r)$ can be taken as an electrostatic free energy term[36] which depends on the binding ratio, r.

In cases where nonelectrostatic interactions may be important, the more exact approach of Equation 10 is preferable. A theoretical plot of $\theta / k_a c_1$ against θ (equivalent to a Scatchard plot with both k_a and n normalized to unity) for several values of $e^{-W/kT}$, as computed from Equation 10, is shown in Figure 6.

When W is nonzero, pronounced curvature is obtained. For positive values of W (i.e., anticooperativity due to a repulsive interaction) the plots are curved downwards below the simple linear case, and for negative values of W (i.e., cooperativity resulting from attractive forces) the curvature is opposite and the plot lies above the classical Scatchard line. It will be noticed that the anticooperative case could be difficult to distinguish from the curvature arising from the two independent binding species model. Cooperative binding is easier to recognize because of the characteristic direction of the curvature of the Scatchard plot, and the sigmoidal nature of the binding curves as shown in Figure 7.

The acridines, being simple planar molecules, exhibit a strong tendency to aggregate by stacking in concentrated solutions and the aggregate spectrum is blue-shifted relative to that of the free monomer. A cooperative effect has been identified in acridine molecules, thought to be externally bound to DNA, at r values close to unity.[17,37] This stacking behavior along the DNA is not normally observed with ethidium bromide

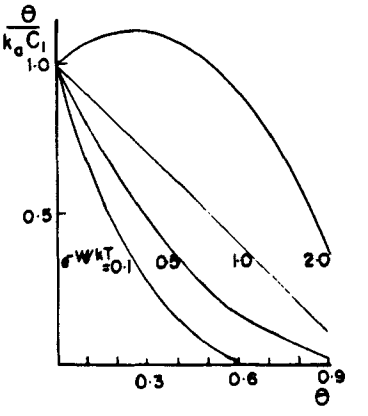


FIGURE 6. A normalized Scatchard plot for various interaction energies, W.

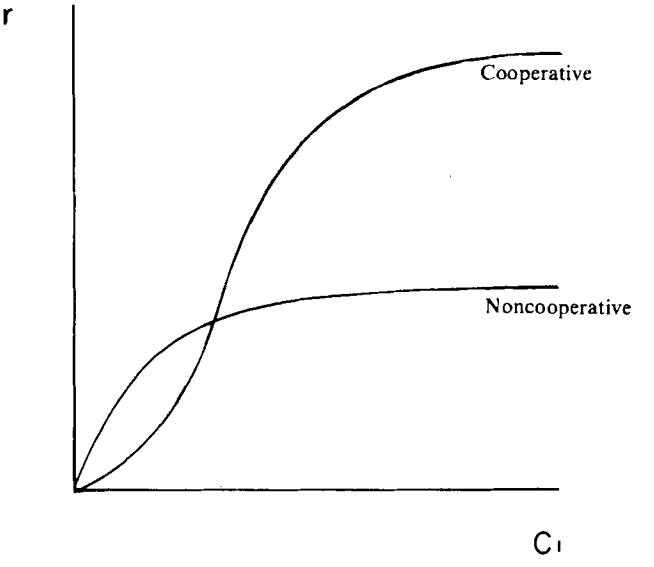


FIGURE 7. Schematic binding curves to illustrate the difference in shape between data for noncooperative and cooperative binding.

because the phenyl side group prevents close contact, but anticooperativity could operate either because of this steric restriction or because of changes in the DNA secondary structure due to local distortion at the strong binding sites (now believed to be intercalative).[37a]

D. The Excluded Site Model

1. The Rationale

It is possible to conceive of a situation where a bound drug occupies more than one binding site. The drug does not need to physically cover more than one site: all that is required is that binding at a particular site precludes binding at neighboring sites. This could arise from

1. Steric hindrance. For example, the phenyl substituent on ethidium and dimidium

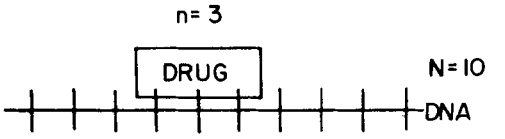


FIGURE 8. An example to illustrate the terms n and N (see text) for the excluded site model.

bromide, and the phenyl group and the pyrimidine moiety on prothidium dibromide, could prevent the approach, within a certain distance, of another potential binding drug. A detailed discussion of these effects will be discussed elsewhere.[29,38]

2. An anticooperative effect, e.g., due to the local electrostatic shielding of the lattice phosphate by a bound drug. It will be seen later in this section that this anticooperative character which we have already considered separately can be conveniently included in the excluded site model. Indeed it is difficult to separate it out as a term independent from the steric hindrance and secondary distortion effects.
3. Distortion of the DNA secondary structure. Binding of a drug could cause stereomodification of the DNA at or near a binding site. If its effect spreads to adjacent sites then it may not prove feasible to implement the same modification at the neighboring sites. According to the intercalative binding model of Sobell et al.,[39] DNA kinks or bends prior to accepting an intercalative drug. This is made possible by altering the normal C2′ endo sugar ring puckering in B DNA to a mixed puckering pattern of the type C3′ endo (3′, 5′) C2′ endo and partially unstacking the bases. This mixed pucker system necessarily limits intercalation to every other base pair (i.e., nearest neighbor exclusion[40]).

2. The Binding Analysis

In the treatment of the excluded site model by McGhee and von Hippel[41] the drug covers n base pairs* on a DNA molecule whose length is N base pairs (Figure 8). At any degree of binding the number of free binding sites depends not only on the number of drug molecules already bound, but also on their distribution along the DNA lattice. This is a major departure from the classical Scatchard analysis. Figure 9 shows some of the immediate differences. At complete saturation, N/n drug molecules are bound. The classical analysis presumes that at any degree of saturation, the numbers of bound and free sites should add up to this number. Figure 9A shows that this is not the case. For this particular example, although the lattice is not completely saturated, there are no acceptable vacant binding sites because the bound drugs have not been constrained to bind n base pairs apart. Figure 9B shows that the number of binding sites on a naked lattice is $(N - n + 1)$. A consequence at low binding ratios (Figure 9C) is that the classical analysis would underestimate the number of free sites (in this instance, as four instead of seven) and thus overestimate k_a (see Table 1).

By incorporating these features and using simple conditional probabilities, McGhee and von Hippel were able to begin with the law of mass action and develop it to obtain an equation analogous to Equation 8:

$$\frac{r}{c_1} = k(1 - nr)\left(\frac{1 - nr}{1 - (n-1)r}\right)^{n-1} \qquad (13)$$

* Note that this differs from the n used earlier. The current n is the number of base pairs occluded per bound drug molecule, whereas the previous n was the number of binding sites per base pair. They are related by their inverses.

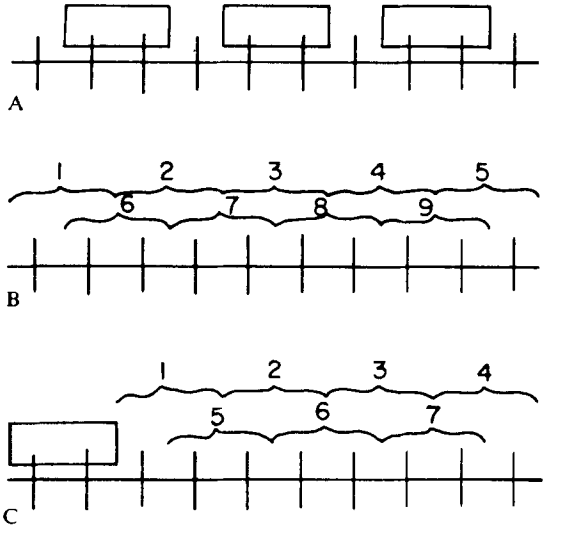


FIGURE 9. Diagrams to illustrate some of the special features, described in the text, of the excluded site model. For these examples, n = 2 and N = 10. (A) No. of bound sites = 3; No. of free sites = 0. (B) No. of binding sites on a naked lattice = N − n + 1 = 9. (C) In this example of low binding density, the number of free sites is 7.

Table 1
COMPARISON OF BINDING PARAMETERS

	Literature values		Computer-fitted values	
	n	k	n	k
Ethidium bromide	2.76	5.03×10^{4}[a]	2.923	1.838×10^{4}
	(2.00	1.83×10^{4})[b]		
Actinomycin C3 (low salt)	11.11	2.4×10^{6}[c]	6.230	2.003×10^{5}
Actinomycin C3 (high salt)	12.50	1.2×10^{7}[c]	7.349	8.055×10^{5}

[a] Values obtained from the classical Scatchard analysis (Equation 1, applied to the linear portion of the data of Müller and Crothers[18]).
[b] Values obtained by Bresloff and Crothers[15] using the analysis of Crothers,[40] which is equivalent to Equation 6 with n constrained to equal two.
[c] Values obtained by Müller and Crothers[18] applying Equation 1 to their data.

where n, c_1, and r have been previously defined, and k is the association constant of an isolated site.

When n = 1 the last factor in this expression becomes equal to unity, and the equation reduces to

$$\frac{r}{c} = k(1 - r)$$

This is identical to the classical Equation 8, except for the factor 1 inside the brackets. Its appearance, rather than Scatchard's n, results from the assumption in this treatment that every base pair is a potential binding site for a single drug. If the

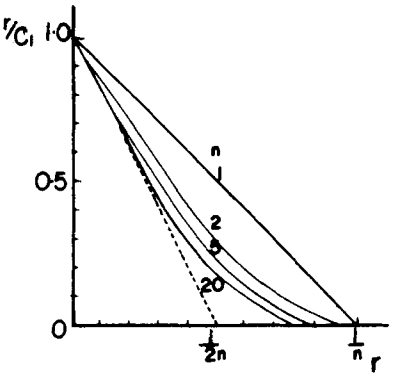


FIGURE 10. Theoretical plots calculated from Equation 13 for the binding of drugs ranging in size from n = 1 to n = 20, for k = 1 M^{-1}. (The abscissa is scaled in units of 1/n, and the dashed line represents the limiting slope of the n = 20 curve) (After McGhee, J. D. and vonHippel, P. H., *J. Mol. Biol.*, 86, 469, 1974. With permission.)

bound drug molecule occupies more than one base pair on the DNA, the last factor in Equation 13 is less than unity and is a function of n and r. The resulting Scatchard plot is curved and falls below the classical straight line plot (Figure 10).

The intercept on the abscissa is 1/n as required for complete saturation, though the plots become very shallow at large r for even moderately sized drug molecules. Crothers[40] noted that this implied a reluctance of the drugs to bind, and resulted from the accumulation of gaps smaller than n base pairs in length.

The slope of Equation 13 at low r is $-k(2n-1)$ so that extrapolation of data from this portion of the curve leads to an intercept of about $1/(2n-1)$. This results in an overestimation of the binding site size (see Table 2). Many early DNA-drug binding studies produced extrapolated intercepts of around 0.3 drugs per base pair, which were difficult to explain in terms of the classical analysis. It can now be seen that these values are consistent with an exclusion site model in which n = 2 (viz. nearest neighbor exclusion).

Equation 13 is the most convenient, but not the sole, formulation of the solution for the excluded site model. The results of Zasedatelev et al.,[42] obtained by a combinatorial method, can be arranged to give this expression, and both Crothers,[40] using a Monte Carlo method, and Schellman,[43] using sequence-generating functions, have proposed similar equations. In addition, there have been a number of more restricted analyses which have produced less general solutions.[44-47]

Recently, good quality X-ray diffraction patterns[48] from a platinum-substituted intercalator-DNA complex have strongly supported the nearest neighbor exclusion theory by demonstrating a 10.2 Å periodicity along the helix axis. The interpretation is that this spacing corresponds to the distance along the helix axis between adjacent drug molecules which are separated by two base pairs. The presence of the meridional suggests that this spacing is regular and implies that the drug cannot bind with only one base pair between them. Unfortunately, however, no estimate of the pitch was made from the diffraction data and because the 10.2 Å meridional is considerably less intense than that corresponding to the 3.4 Å base pair repeat, there are doubts[49] as to whether saturation of the binding had been reached.

Ethidium bromide binding to DNA is a particularly difficult case to analyze spectroscopically. If an external binding mode is present, it is likely that it causes a similar

Table 2
THE INTERCEPT, r_{max}, FOR VARIOUS f VALUES (% G-C CONTENT OF DNA) WHEN n = 2[a]

DNA type	f	r_{max}		
		$P = 2f - f^2$	$P = f$	$P = f^2$
Poly dG – poly dC	1.00	0.50	0.50	0.50
M. lysodeikticus	0.72	0.48	0.42	0.34
Calf thymus	0.42	0.40	0.30	0.15
Cl. perfringens	0.30	0.34	0.23	0.08

[a] See text and Figure 20.

shift to the free drug spectrum as is caused by intercalative binding.[21] Previously the curved Scatchard plot had been analyzed according to a two binding mode scheme,[12] but more recently the data have been shown to accurately fit a single binding species model in which nearest neighbor exclusion effects are operating.[15]

Theoretical calculations have been quoted[50,51] to support a saturation of binding at less than r = 0.5 for the nearest neighbor exclusion model. It has been suggested that the maximum occupancy under this scheme would be $(1 - e^{-2})/2 = 0.436$. The original paper[52] considered the random selection of adjacent pairs of points from a line of N points, but the theory is nonapplicable to drug binding. Once pairs of points had been selected they were not considered again, whereas in drug binding a dynamic situation prevails in which a rearrangement of drugs along the DNA is continuously occurring.

The excluded site model can be generalized further to include more than one binding species and/or to include drug interactions, but in most such extensions the number of unknown parameters proliferates at a rate greater than do the experimental means of distinguishing the different effects. The introduction[41] of a drug interaction parameter w* modifies the theory to produce the plots of Figure 11. Values of w greater than unity correspond to cooperativity, and values less than unity to anticooperativity. Anticooperativity and binding site overlap are seen to result in the same effect on the plot (viz. a curving below the classical straight line), and for most applications it will suffice to absorb the interaction term within the largely steric overlap factor to produce an "apparent site size" which includes both effects.

3. The Computer Programs

A computer program written by Dr. James McGhee uses the Regula falsi iterative method to minimize (with respect to n and k) the sum of the squares of the differences between the two sides of Equation 13, in order to fit experimental binding data to that expression. This program, modified for accelerated convergence by using the Simplex method, has been used extensively by the authors.

The program was tested using previously published data for ethidium bromide[15] and actinomycin C3.[53] The values obtained for n and k are compared with those quoted in the original papers in Table 1. As previously noted, use of Equation 1 (applicable to the simple single binding species model) rather than Equation 13 (resulting from the excluded site theory) leads to an overestimation of both n and k. Figures 12 and 13 show

* The drug interaction parameter, w, is the (unitless) equilibrium constant for the process of moving a bound drug from an isolated site to a site next to another bound drug, or from this latter site to a site between two bound drugs.[41]

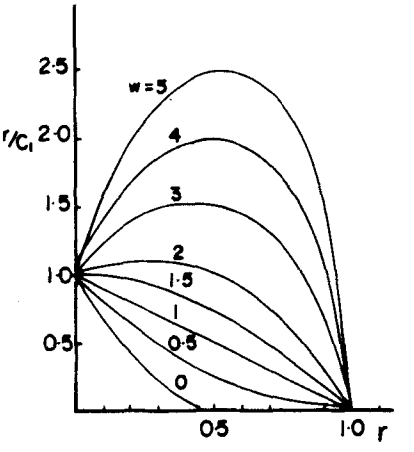


FIGURE 11. Theoretical plots for the cooperative binding of a ligand of n = 1 (for k = 1 M^{-1}) with w ranging from 0 to 5. (After McGhee, J. D. and von Hippel, P. H., *J. Mol. Biol.*, 86, 469, 1974. With permission.)

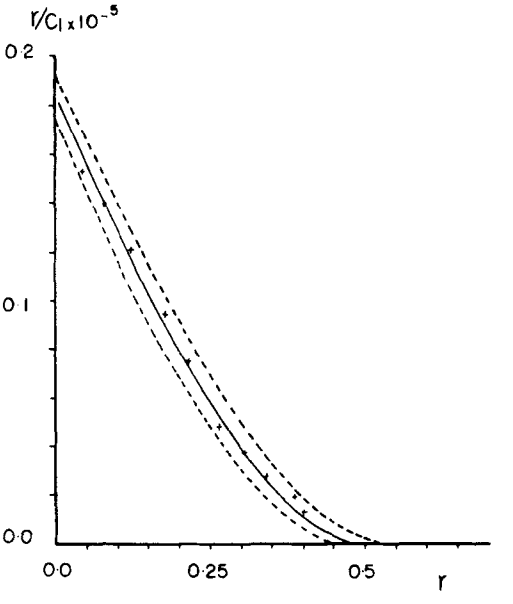
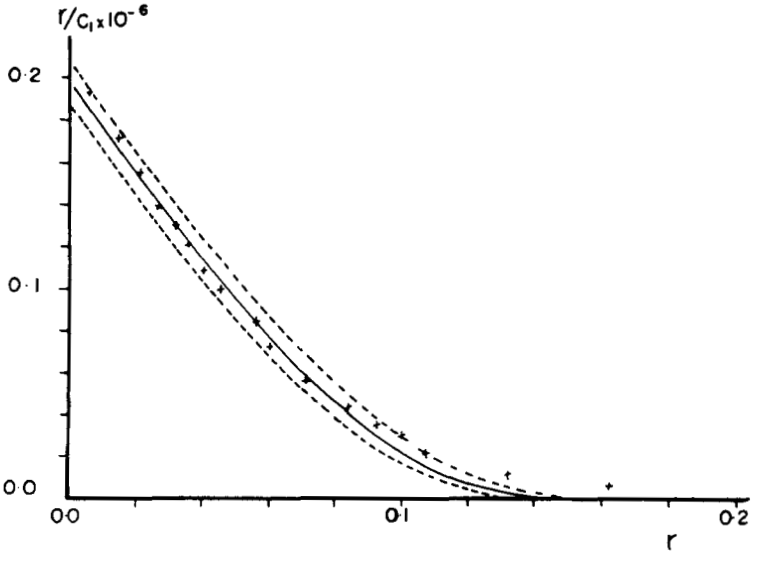
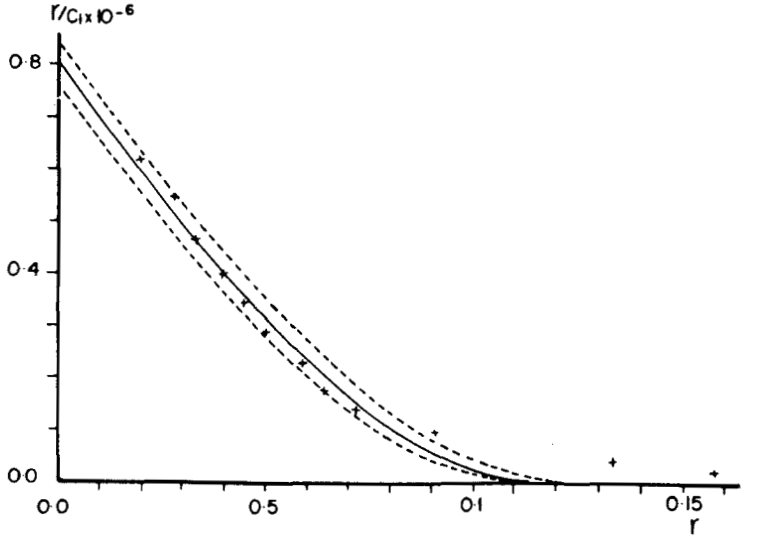


FIGURE 12. Ethidium bromide-DNA binding data fitted to Equation 13 using the program "SCATFIT". The dotted lines indicate the goodness of fit (see text). (After Bresloff, J. L. and Crothers, D. M., *J. Mol. Biol.*, 95, 103, 1975. With permission.)

the generated best-fit curves for these sets of data and two other curves generated by taking values of n 5% lower and k 5% higher than the best-fit values (the upper dotted curve), and then n 5% higher and k 5% lower than the best values (the lower dotted curve). This is used to give an idea of how well the best-fit curve fits the experimental data. The ethidium data are seen to be well described by Equation 13 so that the ± 5% margins for n and k are too conservative. The actinomycin data deviate markedly from the theoretical



A



B

FIGURE 13. Actinomycin C3-DNA binding data fitted to Equation 13 using "SCATFIT". (A) Low ionic strength. (B) High ionic strength. (After Müller, W. and Crothers, D. M., *J. Mol. Biol.*, 35, 251, 1968. With permission.)

fitted curve at large r (as previously noted by McGhee and von Hippel[41]). This implies that an additional complication is operating in the actinomycin binding, and this is probably that actinomycin C3 prefers to bind to a G(3′, 5′) C sequence.[54] Equation 6 is sufficiently inflexible that variations from it due to binding site heterogeneity (discussed fully in the following section) can be detected by a poor fitting of the generated curve to the data.

The log of the sum of the residuals from Equation 13 can be plotted against a range of values of n and k enabling the region where the minimum occurs to be located, and

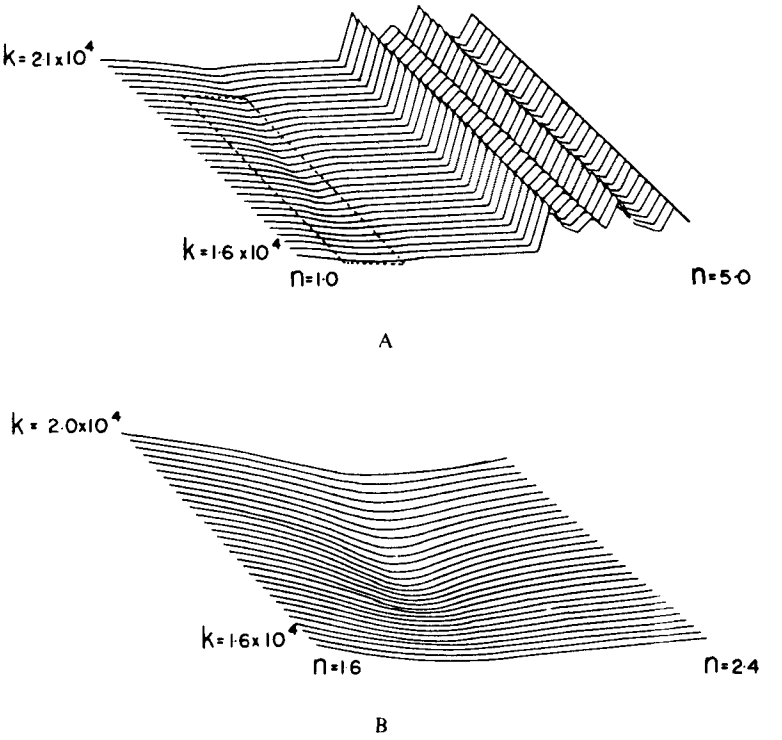


FIGURE 14. Surface plots produced by "SURFPLOT" from DNA-ethidium bromide binding data. The dotted portion of (A) is shown, expanded in scale, as (B). (Data from Bresloff, J. L. and Crothers, D. M., *J. Mol. Biol.*, 95, 103, 1975. With permission.)

hence starting values of n and k to be obtained. Figures 14 and 15 show these plots for the ethidium and actinomycin data. The steps at the right side of the plots show regions in which a minimum will not be found. The absence of readily identifiable minima in Figure 15 for actinomycin indicates that Equation 13 may be a poor description of the experimental situation.

E. Heterogeneous Binding Sites

1. Base-Pair Specificity

A curve Scatchard plot could result from either base-pair or base-sequence specificity in the binding reaction. Present knowledge of the existence of base-pair specificities is restricted largely to the actinomycin-DNA system,[53,55] to the interaction of several acridine dyes[56] and analogs,[18] and to the antibiotics netropsin[57] and echinomycin.[58]

Crothers[40] produced a nonlinear Scatchard plot for a heterogenous polynucleotide using a Monte Carlo method. For a drug which can bind between two base pairs (Figure 16) in which at least one member is G-C, and for which n = 4, he obtained the plots shown in Figure 17. Crothers considered the binding to be intercalative, but this is not necessary for the theory. All that is required is that the drug show a preference for a G-C base pair. For this simulation Crothers[40] considered a DNA molecule comprising 5000 base pairs, a value which permitted advance specification of the base-pair composition of the random sequence to within, roughly, 2%. If f is the fractional G-C content, then the probability of one base pair chosen at random being G-C is f. The probability of choosing two adjacent base pairs with at least one of them being G-C is $(1 - (1 - f)^2)$, i.e., $(2f - f^2)$. This probability holds no matter what the binding site size, n, of the drug,

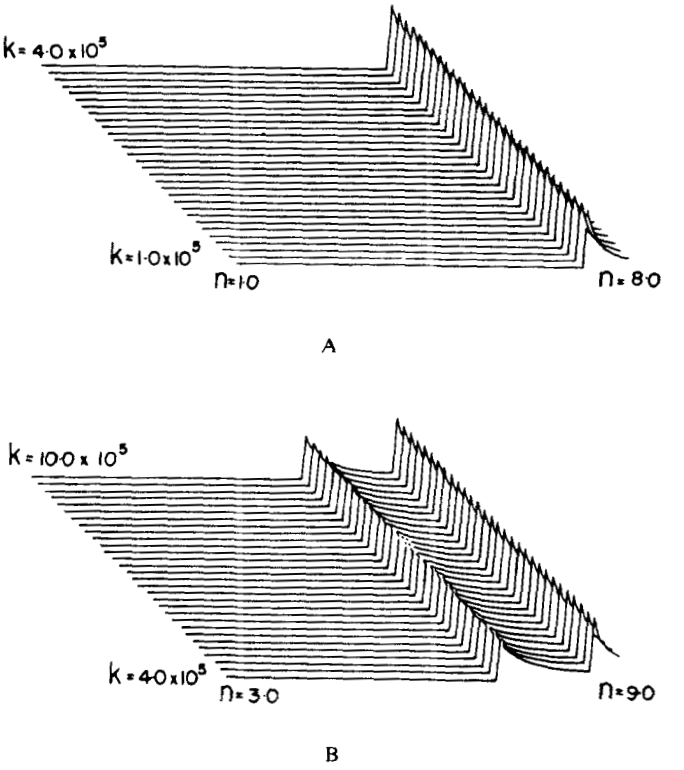


FIGURE 15. Surface plots produced by "SURFPLOT" from DNA-actinomycin C3 binding data at (A) low ionic concentration and (B) high ionic concentration. (Data from Müller, W. and Crothers, D. M., *J. Mol. Biol.*, 35, 251, 1968. With permission.)

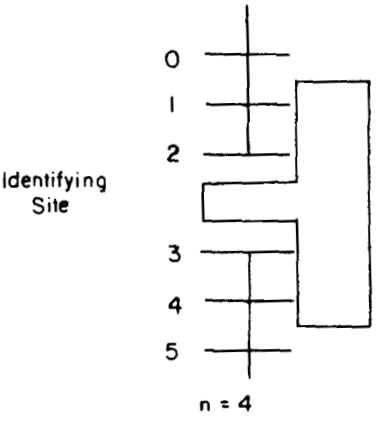


FIGURE 16. A schematic representation of the conditions applicable to the plots of Figure 21. Either base pair 2 or 3 is G-C, or both of them are G-Cs.

so long as the identifying site (Figure 18) on the drug (i.e., the chemical group which actually binds) is fixed in a certain position. This would be the case for most binding schemes, although an exception would be the bifunctional intercalators such as echinomycin.[58]

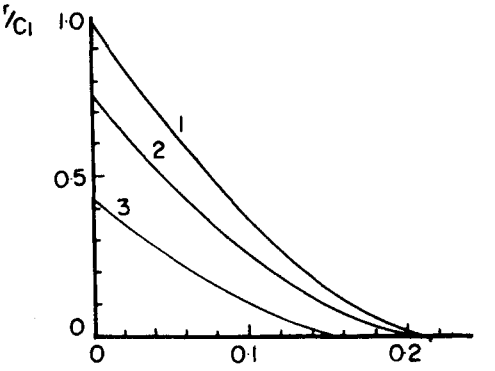


FIGURE 17. Theoretical Scatchard plot for a model in which a drug (n = 4) binds between two base pairs of which at least one is G-C(i.e., $P = 2f - f^2$). k has been set to 1 M^{-1}. (After Crothers, D. M., *Biopolymers*, 6, 575, 1968. With permission.)

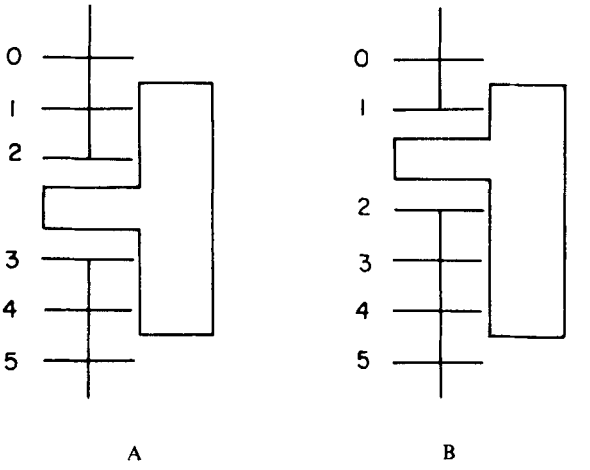


FIGURE 18. Two different identifying sites for the case where n = 4. (A) The identifying site is 2 (or 3). (B) The identifying site is 1 (or 4).

In the simpler case of a homogeneous lattice the intercept on the r/c_1 axis of the Scatchard plot is k, the intrinsic association constant of each binding site. Because of the need to satisfy the base-pair preference in a heterogeneous lattice system, the level of binding will always be less for a given total drug concentration. The r/c_1 intercept will be reduced by the probability factor, P, to Pk. P depends on the fractional G-C content, f, according to the details of the binding preference. For the scheme considered by Crothers, the dependence of P on f is shown below:

f	$P(= 2f - f^2)$
1.0	1.0
0.5	0.755
0.257	0.448

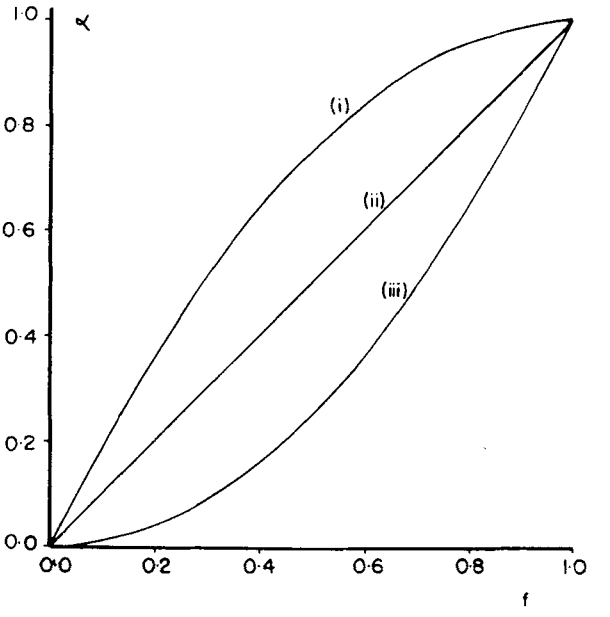


FIGURE 19. The theoretical variation of the affinity, α (normalized to unity), with the fractional G-C content, f. P is the probability factor (see text) and equals (i) $2f - f^2$, (ii) f, and (iii) f^2, respectively.

If binding were to occur exclusively on only one particular side of a G-C pair then the probability factor would equal the fractional G-C content (i.e., P = f), and if it required two adjacent G-C pairs then the probability, P, would be equal to f^2. Thus a plot of the r/c_1 intercept (called the affinity, α, by Müller and Crothers[19]), determined for a variety of DNA's, against a specific expression for P in terms of f, the G-C content, could be useful in testing different models for the binding site specificity requirements. Figure 19 shows the theoretically expected variation of the affinity, α, with the fractional G-C content, f, for the cases where $P = 2f - f^2$, $P = f$ and $P = f^2$.

These relationships hold only for a heteropolymer with random base-pair sequence, whereas in DNA the sequence is generally nonrandom. Josse et al.[59] showed that the base-pair sequence in DNA from odd-numbered type T phages is close to random if sufficiently long segments of DNA are examined, while in calf thymus and bacterial DNA there is a correlation in the arrangement of immediate neighbors. Shugalin et al.[60] have concluded that these latter types can be considered as blocks of quasirandom sequences, although each block contains a different G-C content. For calf thymus DNA the mean molecular weight of the block is 10 to 15 million with a mean square deviation of the content of the blocks from the average of an entire genome as 9.5%, and for *E. coli* the respective figures are 5 to 10 million and 4%. However, Gurskii et al.,[61] using these values for calf thymus DNA, have shown that the resulting value of the binding site size, n, for actinomycin binding to DNA differs by only 5% from the estimate made on the assumption of a completely random heteropolymer. This deviation is lower than the normal experimental error in these investigations. In any case, the molecular weights of the DNA fragments used in the experiments described in this work are likely to be less than the mean value of a block so that the sequence within each fragment can be taken as being completely random.

For the situation where $n > 2$ the concept of an identifying site[61] becomes important.

The affinity is independent of the position of the identifying site within the drug, but the saturation value of r, r_{max}, for a random heteropolymer does depend on the position of this site. For the case where binding requires a G-C pair adjacent to the identifying site, r_{max} is given by

$$r_{max} = \frac{f}{1 + f(n-1) + (1-f)\{1-(1-f)^{t-1}\}} \quad (14)$$

where t denotes the position of the identifying site. In Figure 18A, t could be taken as either 2 or 3, and in Figure 18B it could be taken as either 1 or 4. For use with Equation 14, t should be taken such that $1 \leqslant t \leqslant n/2$.

The equation can be generalized by replacing f by the probability factor P, so that it will describe other base-pair specificity schemes:

$$r_{max} = \frac{P}{1 + P(n-1) + (1-P)\{1-(1-P)^{t-1}\}} \quad (15)$$

For n = 2, when the position of the identifying site is unimportant, this reduces to

$$r_{max} = \frac{P}{1+P} \quad (16)$$

Table 2 illustrates the dependence of r_{max} on f for random-sequence DNA samples of differing G-C content. The relationship between r_{max} and f is shown fully in Figure 20, for the three different probability factors considered previously.

For n = 4, the case considered by Crothers,[40] Equation 15 reduces to

$$r_{max} = \frac{P}{1+3P} \text{ for } t = 1 \quad (17)$$

and

$$r_{max} = \frac{P}{1+4P-P^2} \text{ for } t = 2 \quad (18)$$

The intercepts, r_{max}, for these two distinct cases are tabulated in Table 3 for the G-C contents considered by Crothers, and the relationships between r_{max} and f are plotted in Figure 21. Crothers considered the case where the identifying site requires at least one G-C pair (i.e., $P = 2f - f^2$). However, it is difficult to tell from his results whether his example corresponded to having the identifying site at t = 1 or t = 2.

Reduced binding due to base-pair specificities gives an intercept r_{max} less than the value otherwise expected for that particular binding site size, n. The inverse of the intercept gives a measure of the binding site size so that the apparent binding site size, n_{app} ($= 1/r_{max}$), will be larger than the actual site size. This is illustrated, for the different types of specificity considered, in Figure 22 for an actual site size of two (n = 2) and in Figure 23 for an actual site size of four (n = 4).

To recap on the significance of the intercepts measured from the Scatchard plots:

1. The r/c intercept (the affinity, α) is independent of the size of the binding site and of the position of the identifying site within a drug molecule. Its variation with the fractional G-C content depends on the particular type of base-pair specificity operating, and its value is always less than in the case where there is no specificity in the binding (Figure 19).

Table 3
THE INTERCEPT, r_{max}, FOR VARIOUS f VALUES (% G-C CONTENT OF DNA) WHEN n = 4[a]

	r_{max} for t = 1		r_{max} for t = 2	
f	$P = 2f - f^2$	P = f	$P = 2f - f^2$	P = f
1.00	0.250	0.250	0.250	0.250
0.50	0.231	0.201	0.219	0.183
0.257	0.191	0.145	0.173	0.131

[a] See text and Figure 21.

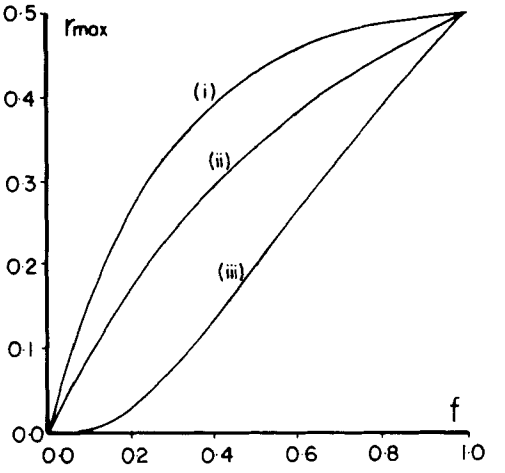


FIGURE 20. The r intercept, r_{max}, from a Scatchard plot shown as a function of the G-C content, f, for n = 2. P has the same values as in Figure 19.

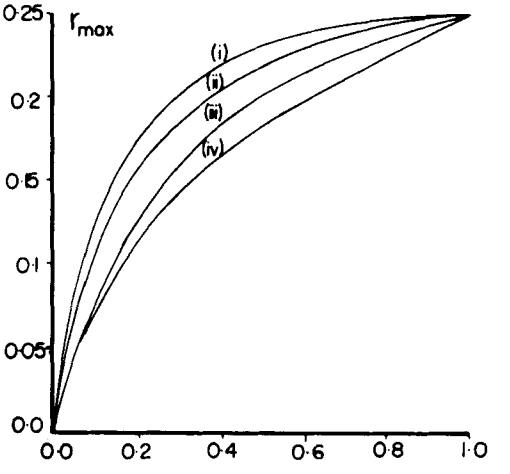


FIGURE 21. The r intercept, r_{max}, as a function of G-C content, f, for n = 4: (i) t = 1, $P = 2f - f^2$; (ii) t = 2, $P = 2f - f^2$; (iii) t = 1, P = f; and (iv) t = 2, P = f.

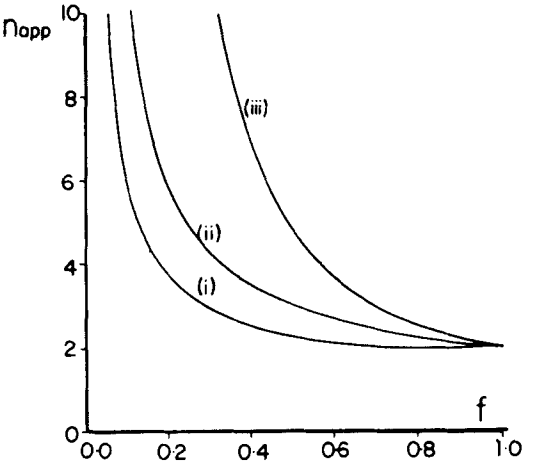


FIGURE 22. Apparent binding site size, n_{app}, as a function of the fractional G-C content, f, for n = 2: (i) P = $2f - f^2$; (ii) P = f; and (iii) P = f^2.

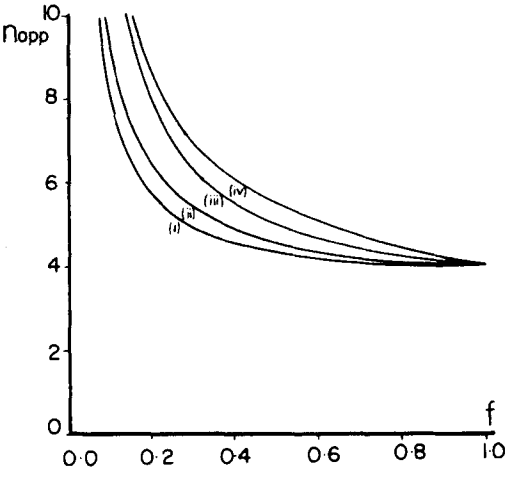


FIGURE 23. Apparent binding site size, n_{app}, as a function of the fractional G-C content, f, for n = 4: (i) t = 1, P = $2f - f^2$; (ii) t = 2, P = $2f - f^2$; (iii) t = 1, P = f; and (iv) t = 2, P = f.

2. The r intercept (r_{max}) depends both on the binding site size and, when $n > 2$, on the position of the identifying site (Figures 20 and 21). Its value is always smaller than the expected value when no base-pair specificity is present. Although it depends on more variables than does the affinity, it may prove useful to consider its variation with G-C content if r_{max} is known more reliably than α.

A further complication would be that binding occurs to both A-T and G-C pairs but with differing affinities (k_{AT} and k_{GC}), a circumstance that could hardly be detected by these methods.

Müller and Crothers[18] studied the base-pair specificities of a series of proflavine and acridine orange analogs by differential dialysis of the drugs against DNA samples of

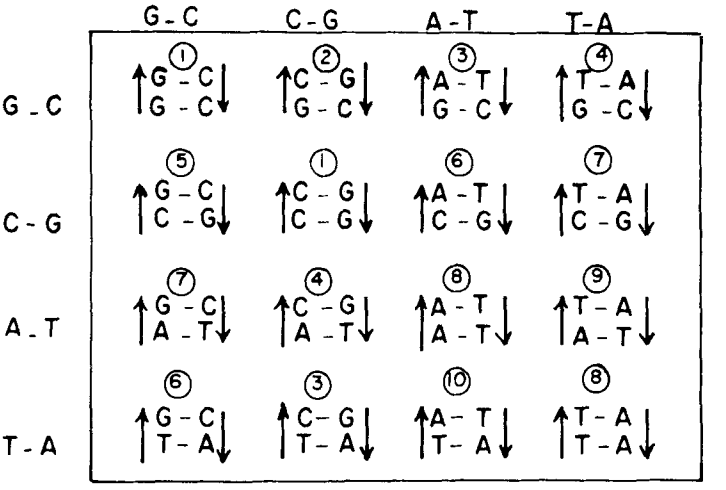


FIGURE 24. Schematic drawing of the possible intercalating sites in DNA. A degeneracy exists so that of the 16 possibilities, only 10 are distinctly different.

differing G-C content. For the bis(dimethyl-amino)phenazoxonium ion, the results closely fitted a linear relationship between the affinity, α, and the fractional G-C content indicating that binding requires one G-C pair.

2. Base-Sequence Specificity

Analyzing the dependence of α and r_{max} on the G-C content, as described above, may provide information on the preference of a drug for a particular sequence of bases. The situation where a drug is found to prefer binding to a site containing two adjacent G-Cs (i.e., $P = f^2$) is equivalent to a sequence specificity for GpG, GpC, CpC, or CpG.

It can be shown that, as a consequence of the right-handed nature of the Watson-Crick DNA double helix with its strands of opposite polarity, there are 10 distinctly different dinucleotide binding sites (Figure 24). Although the experimental evidence is conflicting,[62-65] each of these sites could theoretically possess a different association constant ($k_1, k_2, k_3 \ldots . k_{10}$) and the differences among them would be more apparent at low binding ratios. These 10 sites could be built into a double-helical model of 11 base pairs, or into a circular double helix of 10 base pairs, for inspection purposes (Figure 25). It is clear, for example, that sites 1, 2, and 3 in Figure 25 are not equivalent and that they present different environments to a binding drug molecule.

In a random arrangement of base pairs in a DNA molecule, the frequency of any nearest neighbor pair (viz. a sequence of two) should be predictable as the product of the frequencies of its constituent monomers (e.g., $F(ApT) = F(TpA) = F(Ap) \times F(Tp)$, where F indicates "frequency of"). Josse et al.[59] observed that while most sequence frequencies fell within these predictions, for several DNAs they differed sharply. In particular, the sequence C(3′, 5′)G seems to occur with a frequency which is invariably less than the random value in animal and plant cells (0.016 in calf thymus, instead of an expected $0.21 \times 0.21 = 0.044$); and, by contrast, its isomeric sequence G(3′, 5′)C differs little from the random expectation (in calf thymus, 0.044 experimentally). In bacteria, the reverse of this tendency appears to be the case. Such departures from random-sequence frequencies would distort to some extent any conclusions on base-sequence specificity which were derived from drug binding analysis on the assumption of a random-base sequence for a DNA of known G-C content.

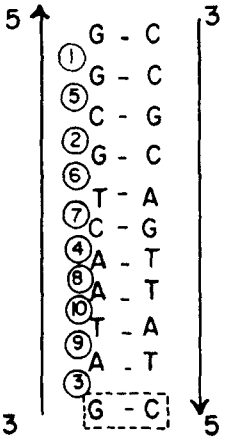


FIGURE 25. The 10 distinguishable sites may be built into a linear DNA model 11 base-pairs long, or into a circular DNA model 10 base-pairs long with 5 G-Cs and 5 A-Ts.

The base-pair specificity of drugs (netropsin,[57] echinomycin[58]) has been determined through binding experiments using DNA samples of varying base-pair composition. Direct determination of sequence specificity[66] is facilitated by using synthetic double-helical polynucleotides, but these are limited in their availability.

Krugh and Reinhardt[67] have investigated the strong (intercalative) binding of ethidium bromide to dinucleotides using visible spectrophotometry, circular dichroism, and NMR techniques and have reported a base-sequence specificity. A definite preference for binding to a pyrimidine (3′, 5′) purine sequence (CpG) compared with the isomeric purine (3′, 5′) pyrimidine sequence (GpC) has been detected. They predict the existence of different equilibrium constants for the ten different intercalation sites based on the importance of stabilizing stacking interactions between the base pair and the intercalated phenanthridium ring. Patel and Canuel[68] have extended sequence studies to the tetranucleotide level, and have observed stronger binding of ethidium to the self-complementary CpGpCpG (i.e., two CpG intercalation sites) and CpCpGpGp (i.e., one CpG site) duplexes compared to the GpGpCpC (i.e., no CpG sites) by monitoring the UV absorbance at 480 nm of various mixtures.

A recent theoretical analysis[69] of the origins of the base-sequence specificity in the intercalative binding of ethidium to DNA has considered the binding in two stages. In the first stage, the double helix changes conformation (unwinds) to accept the intercalating molecule, and in the second stage, the drug is bound and its interaction energy (primarily with the bases) compensates for the energy required for unwinding. It is concluded that the sequence specificity of ethidium is more readily explained in terms of the conformational energy changes than by preferential stacking interactions. NMR studies[70] of the binding of proflavine and propidium diodide have shown that both exhibit specificity for C(3′, 5′)G sites, similar to that shown by ethidium, and it is probable[69] that all drugs which unwind DNA in the same manner as ethidium would exhibit this particular sequence specificity.*

*Actinomycin D shows a strong specificity for C(3′, 5′)G sequences, which has been explained in terms of specific hydrogen bonding by the lactone rings.[9] It is interesting, however, that 9-amino-acridine[71] shows a preference for a purine (3′,5′) pyrimidine sequence, UpA.

IV. CONCLUDING REMARKS

The change in the free energy when drugs bind to nucleic acids results from a complex mixture of electrostatic effects, π-electron interactions, hydrophobic bonding, and dipole/polarizability factors, all constrained by steric considerations. Estimation of the relative importance of each of these effects would not be easy even with detailed thermodynamic and kinetic studies. Free energy calculations (e.g., Gersch and Jordan[63] on the aminoacridines) remain at a formative stage and, because of the complexity of the situation, must continue to function only in a supporting capacity to empirical observations.

Useful information can be obtained from a careful interpretation of a Scatchard plot of the binding parameters. It is important to realize that much (if not all) of the nonlinearity in such a plot can be attributed to a binding overlap effect (i.e., an exclusion site model). The physical nature of the exclusion will not be revealed directly, since it will probably be a combination of the actual size of the drug and distortion of the local DNA conformation. Experimental data should first be analyzed according to this model, and only if this proves inadequate should one consider a multibinding species or a cooperativity treatment. Analysis using nucleic acid samples of different G-C content will reveal any base-pair specificity of a drug, and may indicate a base-sequence specificity in its binding.

Drug binding studies may be carried out either to investigate the properties of the drug or to probe the nucleic acid conformation. The presence of different binding sites may prove to be of great importance in the recognition process involved in protein-nucleic acid interactions and in DNA replication and transcription, as well as in chemotherapeutic considerations. Simple electrostatic and hydrogen bonding interactions cannot, by themselves, be very selective since there are many such sites on the helix, but selectivity due to stacking/steric/electronic/unwinding factors would allow a sequence within the nucleic acid molecule to be recognized. Highly specific antibiotic and antitumour drugs might then be designed to bind at or adjacent to these sequences.

REFERENCES

1. **Hawking, F.,** *Experimental Chemotherapy,* Schnitzer, R. J. and Hawking, F., Eds., Academic Press, New York, 1963, 129.
2. **Fuller, W. and Waring, M. J.,** A molecular model for the interaction of ethidium bromide with deoxyribonucleic acid, *Ber. Bunsenges. Phys. Chem.,* 68, 805, 1964.
3. **Lerman, L. S.,** The structure of the DNA-acridine complex, *Proc. Natl. Acad. Sci. U.S.A.,* 49, 94, 1963.
4. **Kersten, W., Kersten, H., and Szybalski, W.,** Physico-chemical properties of complexes between DNA and antibiotics which affect RNA synthesis, *Biochemistry,* 5, 236, 1966.
5. **DiMarco, A. and Arcamone, F.,** DNA-complexing antibiotics. Daunomycin, adriamycin and their derivatives, *Arzneim.-Forsch,* 25, 368, 1975.
6. **Rosi, D., Peruzzotti, G., Dennis, E. W., Berberian, D. A., Freele, H., Tullar, B. F., and Archer, S.,** Hycanthone, a new active metabolite of lucanthone, *J. Med. Chem.,* 10, 867, 1967.
7. **Hirschberg, E., Weinstein, I. B., Gersten, N., Marner, E., Finkelstein, T., and Carchman, R.,** Structure-activity studies on the mechanism of action of Miracil D, *Cancer Res.,* 28, 601, 1968.
8. **O'Brien, R. L., Allison, J. L., and Hahn, F. E.,** Evidence for intercalation of chloroquine into DNA, *Biochim. Biophys. Acta,* 129, 622, 1966.
9. **Jain, S. C. and Sobell, H. M.,** Stereochemistry of actinomycin binding to DNA. I. Refinement and further structural details of the actinomycin-deoxyguanosine crystalline complex. II. Detailed molecular model of actinomycin-DNA complex and its implications, *J. Mol. Biol.,* 68, (I:1; II:21), 1972.
10. **Pigram, W. J., Fuller, W., and Hamilton, L. D.,** Stereochemistry of intercalation: interaction of daunomycin with DNA, *Nature New Biol.,* 53, 17, 1972.

11. **Le Pecq, J. B. and Paoletti, C.,** A fluorescent complex between ethidium bromide and nucleic acids. Physical-chemical characterization, *J. Mol. Biol.*, 27, 87, 1967.
12. **Waring, M. J.,** Complex formation between ethidium bromide and nucleic acids, *J. Mol. Biol.*, 13, 269, 1965.
13. **Zunino, F., Gambetta, R., DiMarco, A., and Zaccara, A.,** Interaction of daunomycin and its derivatives with DNA, *Biochim. Biophys. Acta*, 277, 489, 1972.
14. **Blake, A. and Peacocke, A. R.,** Perspectives report: the interaction of aminoacridines with nucleic acids, *Biopolymers*, 6, 1225, 1968.
15. **Bresloff, J. L. and Crothers, D. M.,** DNA-ethidium bromide reaction kinetics: demonstration of direct ligand transfer between DNA and binding sites, *J. Mol. Biol.*, 95, 103, 1975.
16. **Li, H. J. and Crothers, D. M.,** Relaxation studies of the proflavine-DNA complex: the kinetics of an intercalation reaction, *J. Mol. Biol.*, 39, 461, 1969.
17. **Peacocke, A. R. and Skerrett, J. N. H.,** The interaction of aminoacridines with nucleic acids, *Trans. Far. Soc.*, 52, 261, 1956.

17a. **Dougherty, G.,** A single-beam microspectrophotometer suitable for investigating the linear dichroism of DNA-drug fibers, *Anal. Biochem.*, 115, 52, 1981.

18. **Müller, W. and Crothers, D. M.,** Interactions of heteroaromatic compounds with nucleic acids. I. The influence of heteroatoms and polarizability on the base-pair specificity of intercalating ligands, *Eur. J. Biochem.*, 54, 267, 1975.
19. **Lloyd, P. H., Prutton, R. N., and Peacocke, A. R.,** Sedimentation studies on the interaction of proflavine with deoxyribonucleic acid, *Biochem. J.*, 107, 353, 1968.
20. **Krugh, T. R., Winkle, S. A., and Graves, D. E.,** Solute enhanced partition analysis — a novel method for measuring the binding of drugs to DNA, *Biochem. Biophys. Res. Commun.*, 98, 317, 1981.
21. **Porumb, H.,** Optical and X-Ray Diffraction Studies of Polynucleotides and Their Interaction with Drugs, Ph.D. thesis, University of Keele, England, 1976.
22. **Dougherty, G.,** Quantitative Microspectroscopic Studies of DNA-Drug Complexes, Ph.D. thesis, University of Keele, England, 1979.
23. **Waring, M. J.,** Variation of the supercoils in closed circular DNA by binding of antibiotics and drugs: evidence for molecular models involving intercalation, *J. Mol. Biol.*, 54, 247, 1970.
24. **Angerer, L. M. and Moudrianakis, E. N.,** Interaction of ethidium bromide with whole and selectively deproteinized DNA from calf thymus, *J. Mol. Biol.*, 63, 505, 1972.
25. **Douthart, R. J., Burnett, J. P., Beasley, F. W., and Frank, B. H.,** Binding of ethidium bromide to double-stranded RNA, *Biochemistry*, 12, 214, 1973.
26. **Patel, D. J. and Shen, C.,** Sugar pucker geometries at the intercalation site of propidium diiodide into miniature RNA and DNA duplexes in solution, *Proc. Natl. Acad. Sci. U.S.A.*, 75, 2553, 1978.
27. **Plumbridge, T. W. and Brown, J. R.,** Spectrophotometric and fluorescence polarization studies of the binding of ethidium, daunomycin and mepacrine to DNA and to poly (I-C), *Biochim. Biophys. Acta*, 479, 441, 1977.
28. **Dalgleish, D. G., Peacocke, A. R., Fey, G., and Harvey, C.,** The CD in the UV of aminoacridines and ethidium bromide bound to DNA, *Biopolymers*, 10, 1853, 1971.
29. **Dougherty, G.,** A comparison of the base-pair specificities of three phenanthridine drugs using solution spectroscopy, *Int. J. Biochem.*, in press.
30. **Philpott, M. R.,** Exciton theory of the electronic states of dye-polymer complexes. I. Linear polymers, *J. Chem. Phys.*, 53, 968, 1970.
31. **Ellis, K. J. and Duggleby, R. G.,** What happens when data are fitted to the wrong equation?, *Biochem. J.*, 171, 513, 1978.
32. **Klotz, I. M. and Hunston, D. L.,** Properties of graphical representations of multiple classes of binding sites, *Biochemistry*, 10, 3065, 1971.
33. **Scatchard, G.,** The attractions of proteins for small molecules and ions, *Ann. N.Y. Acad. Sci.*, 51, 660, 1949.
34. **Nørby, J. G., Ottolenghi, P., and Jensen, J.,** Scatchard plot: common misinterpretation of binding experiments, *Anal. Biochem.*, 102, 318, 1980.
35. **Fowler, R. and Guggenheim, E.,** *Statistical Thermodynamics*, Cambridge University Press, London, 1952, chap. 4.
36. **Ramstein, J. and Leng, M.,** Effect of DNA base composition on the intercalation of proflavine. A kinetic study, *Biophys. Chem.*, 3, 234, 1975.
37. **Bradley, D. F. and Felsenfeld, G.,** Aggregation of an acridine dye on native and denatured deoxyribonucleates, *Nature*, 184, 1920, 1959.

37a. **Dougherty, G.,** Intercalation — fact or fiction?, *Comments Mol. Cell Biophys.*, in press.

38. **Dougherty, G. and Waring, M. J.,** The interaction between prothidium dibromide and DNA at the molecular level, *Biophys. Chem.*, in press.
39. **Sobell, H. M., Tsai, C-C., Jain, S. C., and Gilbert, S. G.,** Visualization of drug-nucleic acid interactions at atomic resolution. III. Unifying structural concepts in understanding drug-DNA interactions and their broader implications in understanding protein-DNA interactions, *J. Mol. Biol.*, 114, 333, 1977.
40. **Crothers, D. M.,** Calculation of binding isotherms for heterogeneous polymers, *Biopolymers*, 6, 575, 1968.
41. **McGhee, J. D. and von Hippel, P. H.,** Theoretical aspects of DNA-protein interactions: cooperative and noncooperative binding of large ligands to a one-dimensional homogeneous lattice, *J. Mol. Biol.*, 86, 469, 1974.
42. **Zasedatelev, A. S., Gurskii, G. V., and Vol'kenshtein, M. V.,** Theory of one-dimensional adsorption. I. Adsorption of small molecules on a homopolymer. *Mol. Biol. (Moscow)*, 5, 194, 1971.
43. **Schellman, J. A.,** Cooperative multi-site binding to DNA, *Isr. J. Chem.*, 12, 219, 1974.
44. **Hill, T. L.,** Some statistical problems concerning linear macromolecules, *J. Polymer Sci.*, 23, 549, 1957.
45. **Schwarz, G.,** Cooperative binding to linear biopolymers. I. Fundamental static and dynamic properties, *Eur. J. Biochem.*, 12, 442, 1970.
46. **Bauer, W. and Vinograd, J.,** The interaction of closed circular DNA with intercalative dyes. II. The free energy of superhelix formation in SV40 DNA, *Mol. Biol.*, 47, 419, 1970.
47. **Armstrong, R. W., Kurucsev, T., and Strauss, U. P.,** The interaction between acridine dyes and DNA, *J. Am. Chem. Soc.*, 92, 3174, 1970.
48. **Bond, P. J., Langridge, R., Jennette, K. W., and Lippard, S. J.,** X-ray fiber diffraction evidence for neighbor exclusion binding of a platinum metallointercalation reagent to DNA, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 4285, 1975.
49. **Goodwin, D. C.,** Studies on DNA Polymorphism, Ph.D. thesis, University of Keele, England, 1977.
50. **Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H., and Waring, M. J.,** *The Molecular Basis of Antibiotic Action*, 2nd eds., John Wiley & Sons, London, 1981, chap. 5.
51. **Cairns, J.,** The application of autoradiography to the study of DNA viruses, *Cold Spring Harbor Symp.*, 27, 311, 1962.
52. **Page, E. S.,** The distribution of vacancies on a line, *J. R. Stat. Soc. (Ser. B)*, 21, 364, 1959.
53. **Müller, W. and Crothers, D. M.,** Studies of the binding of actinomycin and related compounds to DNA, *J. Mol. Biol.*, 35, 251, 1968.
54. **Sobell, H. M.,** *Progress in Nucleic Acid Research and Molecular Biology*, Academic Press, New York, 1973, 153.
55. **Sobell, H. M., Jain, S. C., and Sakore, T. D.,** Stereochemistry of actinomycin-DNA binding, *Nature New Biol.*, 231, 200, 1971.
56. **Ramstein, J., Dourlent, M., and Leng, M.,** Interaction between proflavine and DNA: influence of DNA base composition, *Biochem. Biophys. Res. Commun.*, 47, 874, 1972.
57. **Wartell, R. M., Larson, J. E., and Wells, R. D.,** Netropsin: a specific probe for A-T regions of duplex DNA, *J. Biol. Chem.*, 249, 6717, 1974.
58. **Wakelin, L. P. G. and Waring, M. J.,** The binding of echinomycin to deoxyribonucleic acid, *Biochem. J.*, 157, 721, 1976.
59. **Josse, J., Kaiser, A. D., and Kornberg, A.,** Enzymatic synthesis of DNA. VIII. Frequencies of nearest neighbor base sequences in DNA, *J. Biol. Chem.*, 236, 864, 1961.
60. **Shugalin, A. V., Frank-Kamenetskii, M. D., and Lazurkin, Yu.S.,** Block heterogeneity of DNA from calf thymus to *E. Coli*, *Mol. Biol. (Moscow)*, 5, 613, 1971.
61. **Gurskii, G. V., Zasedatelev, A. S., and Vol'kenshtein, M. V.,** Theory of single-dimensional adsorption. II. Adsorption of small molecules on a heteropolymer, *Mol. Biol. (Moscow)*, 6, 385, 1972.
62. **Chan, L. M. and van Winkle, Q.,** Interaction of acriflavine with DNA and RNA, *J. Mol. Biol.*, 40, 491, 1969.
63. **Gersch, N. F. and Jordan, D. O.,** Interaction of DNA with aminoacridines. *J. Mol. Biol.*, 13, 138, 1965.
64. **Kleinwächter, V., Balcarova, Z., and Bohaček, J.,** Thermal stability of complexes of diaminoacridines with DNAs of varying base content, *Biochim. Biophys. Acta*, 174, 188, 1969.
65. **Bidet, R., Chambron, J., and Weill, G.,** Heterogeneity of proflavine binding sites on DNA, *Biopolymers*, 10, 225, 1971.
66. **Müller, W., Crothers, D. M., and Waring, M. J.,** A nonintercalating proflavine derivative, *Eur. J. Biochem.*, 39, 223, 1973.
67. **Krugh, T. R. and Reinhardt, C. G.,** Evidence for sequence preferences in the intercalative binding of ethidium bromide to dinucleoside monophosphates, *J. Mol. Biol.*, 97, 133, 1975.

68. **Patel, D. J. and Canuel, L. L.,** Ethidium bromide-(dC-dG-dC-dG)$_2$ complex in solution: intercalation and sequence specificity of drug binding at the tetranucleotide duplex level, *Proc. Natl. Acad. Sci. U.S.A.,* 73, 3343, 1976.
69. **Pack, G. R. and Loew, G.,** Origins of the specificity in the intercalation of ethidium into nucleic acids. A theoretical analysis, *Biochim. Biophys. Acta,* 519, 163, 1978.
70. **Patel, D. J. and Canuel, L. L.,** Biphasic helix-coil transition of eithidium bromide-poly (dA.dT) and propidium diiodide-poly (dA.dT) complexes. Stabilization of base-pair regions centered about the intercalation site, *Biopolymers,* 16, 857, 1977.
71. **Seeman, N. C., Day, R. O., and Rich, A.,** Nucleic acid-mutagen interactions. Crystal structure of adenylyl-3′-5′-uridine plus 9-aminoacridine, *Nature,* 253, 324, 1975.