

DETERMINING THE BINDING CONSTANTS OF BIOLOGICALLY
ACTIVE COMPOUNDS TO NUCLEIC ACIDS:
A NONLINEAR REGRESSION APPROACH.

David E. Booth
Administrative Sciences Department
Kent State University
Kent, OH
Telephone: 216-672-2750
Internet: DBooth@KentVM.Kent.Edu
Fax: 216-672-2448

Abstract

The study of the strength of binding of drugs and other biologically active compounds to nucleic acids is an important area of research. The present paper shows that using a nonlinear regression model can provide advantages to the standard linear statistical model approach for determining such constants. The necessary methods are developed and discussed. The method is illustrated using a real data set.

Introduction

The binding of biologically active compounds to nucleic acids is an important biological event^{1,3,4,12,13,14,15,18,24,27,30,32}. Many of the compounds that bind nucleic acids are anti-tumor drugs^{3,9,14,18,24}, antibiotics^{14,18,19,36}, mutagens^{2,12,27}, and carcinogens^{16,24} as well as other compounds of biological interest^{22,28,30,32,35}. In addition, it is known that the binding of some drugs to nucleic acids can increase the drugs pharmaceutical effectiveness¹³. Thus, it is important to study drug-DNA interactions.

There are many different experimental techniques that can be used to study the interaction of these compounds and nucleic acids^{3,10,12,15,16,17,20,21,27,28,30,31}. However, an analysis technique that can be used with many different types of experimental study is the Scatchard plot^{12,20,37}. The purpose of this plot is to determine the association constant for complex formation, K , and the maximum number of binding sites per mononucleotide, n . The constant K gives the strength of binding of the ligand to the nucleic acid.

For DNA binding the Scatchard plot is biphasic because there are basically two modes of binding that the ligand can take, intercalation or binding to the DNA sugar-phosphate backbone. Thus, one observes a Scatchard plot of the form of Figure 1. The standard method of determining the two binding constants (K_I , for intercalation and K_{II} for sugar-phosphate binding) is by calculation from the slopes of the two lines on the plot in Figure 1 (described in detail later). What this paper proposes is that a more accurate and more realistic approach to the determination of

K_I and K_{II} can be obtained by considering the plot of Fig. 1 to be a single curve. Then one determines the binding constants from the application of nonlinear regression models. In the following sections, the standard approach and the proposed approach will be described. Then the nonlinear regression method will be demonstrated by considering the binding of phenazine methosulfate with double stranded DNA. Finally, the advantages of the nonlinear regression method will be discussed.

The Standard Scatchard Plot

We follow the descriptions given by Ishizu et al.¹² for the use of the plot with optical spectrophotometry. The Scatchard plot itself is a plot of $\bar{B}/F$ versus $\bar{B}$ where $\bar{B}$ is the average number of bound ligands per nucleotide and F is the molar concentration of free ligand. If all binding sites are equivalent and noninteracting such plots are approximately straight lines according to the expression, $\bar{B}/F = K(n-\bar{B})$. If optical spectrophotometry is used to study the complex formation then the fraction of total ligand bound, α , is given by $\alpha = (A_f - A)/(A_f - A_b)$ where A_f and A_b are the absorbance of ligand when it is present entirely in the free or bound form respectively, and A is the observed absorbance of the mixture. $\bar{B}$ and F can then be determined from the relations $\bar{B} = (T-F)/P$ and $F = T(1-\alpha)$ where T is the total molar concentration of ligand and P is the molar concentration of DNA nucleotide. Once $\bar{B}/F$ and $\bar{B}$ are determined from the above relations, a Scatchard plot of the form of Fig 1 is made and then

K_I and K_{II} are computed from the slopes of the two lines using standard least squares regression.

Place Figure 1 about here

The Nonlinear Regression Approach

The nonlinear regression approach is now considered. It follows the usual Scatchard Plot approach through formation of the plot itself. The determination of the equilibrium constants, K_I and K_{II} , is different. We first recall that a linear statistical model is one that is linear in the regression coefficients while a nonlinear model is one that is not linear in the regression coefficients^{5,6,7,8,11,23,25,26,33}. Full details on dealing with such models are given in the references^{8,26,33}. In this paper, we assume that the Scatchard plot data is not perfectly linear and hence the data graph shown in Fig.1 is really a curve rather than two intersecting straight lines. We propose to model the curve, compute the two constants, K_I and K_{II} , with only one nonlinear regression run and then determine the exact point of intersection of the two binding regimes.

For a solution to this type of problem two models have been proposed³⁴. These models are proposed for use with data in the form of Figure 1, under the assumptions given in the last paragraph.

In dealing with the behavior of stagnant band height data Watts and Bacon³⁴ considered data which appear to behave according to two different essentially straight-line relationships on opposite sides of an undetermined join point. Such a situation

could be treated by a model of the form

(1) $Y = \alpha_0 + \alpha_1(X - X_0) + \alpha_2(X - X_0) \tanh \{ (X - X_0) / \gamma \} + Z$ where Z is a random error term. By considerations of differential geometry, Watts and Bacon were able to show that X_0 and γ determine, respectively, the location of the join point and the radius of curvature of the model at the join point. Further, in the limit, as γ approaches 0, (1) devolves to two straight lines with slopes $(\alpha_1 - \alpha_2)$ and $(\alpha_1 + \alpha_2)$, intersecting at the coordinates (X_0, α_0) . Thus by fitting the nonlinear regression model (1), one can compute the binding constants K_I and K_{II} as well as the location of the join point.

Watts and Bacon³⁴ also proposed a second approach based on a hyperbolic model for the data in Fig. 1. They considered a hyperbola for which

- (i) Y is a single valued function of X
- (ii) the left asymptote has slope θ_1
- (iii) the right asymptote has slope θ_2
- (iv) the asymptotes intersect at (X_0, β_0) and
- (V) the radius of curvature at $X = X_0$ is proportional to a quantity δ . Such a hyperbola can be written as

$$(2) \quad Y = \beta_0 + \beta_1(X - X_0) + \beta_2[(X - X_0)^2 + \delta^2/4]^{1/2} + z$$

where $\beta_1 = (\theta_1 + \theta_2)/2$ and

$$\beta_2 = (\theta_2 - \theta_1)/2$$

Thus the binding constants in this model are θ_1 , and θ_2 and they can be determined simply by fitting model (2) to the data. Watts and Bacon note the parameterization for model (2) is chosen to correspond directly to that in model (1). We will now examine the use of model (1) and (2) to determine binding constants

Data

The data for this study were determined based on a binding study of the methylphenazinium cation with DNA authored by Ishizu et al¹². The data were taken from their Scatchard plot (Figure 2 of the reference). This data is summarized in Table 1 and plotted in Figure 2. Figure 2 is thus the same plot as shown in Ishizu et al¹² and it compares favorably with theirs.

Table 1 & Figure 2 about here

Nonlinear Regression

The nonlinear regression of the Table 1 data with models (1) and (2) was accomplished with SAS²⁵ Proc NLIN using the approach described in Wadsworth³³. The DUD^{25,33} method of computation was used because it does not require the specification of derivative values. The program for Model 1 is given in Fig.3.

Figure 3 about here

The program requires the specification of initial starting values of the model parameters. The starting values used in this study were the values reported by Ishizu et al¹². Because the computation can be sensitive to the choice of starting values I recommend that these values be determined from the standard Scatchard plot. Should outlying observations be a potential problem, robust (Outlier resistant) versions of these procedures can be used^{5,6,7,8,23}.

Results

The results for model (1) are given in Table 2 and those for Model (2) are given in Table 3. The binding constants are computed as described by Watts and Bacon³⁴ and the previous discussion.

Tables 2 & 3 about here

We first consider the Table 2 results obtained with Model (1). We observe that K_I is quite reasonable in considering the results reported by Ishizu et al and K_{II} is as well when one considers that the slope of a mostly horizontal line is quite difficult to measure. In any case, K_I , the intercalation constant, is the most important based on the currently held theories of ligand-DNA binding and we seem to have determined it well.

We now consider the Table 3 results obtained with Model (2). These are much less satisfactory. It is especially important to notice that the Jacobian Matrix used in the computation was singular. Such a situation is discussed by Seber²⁶, p. 623. In the case of a singular Jacobian, it is a signal of a lack of identifiability in the parameters in the nonlinear regression model. This is quite like the situation of multicollinearity in a linear statistical model. In such a case, the regression coefficients are often uninterpretable. In fact as we can see in Table 3 K_{II} even has the wrong sign. Thus, we can not recommend the use of model (2) in studying the binding of ligands to DNA.

Conclusion

It is clear that the study of the binding of drugs and other biologically active compounds to nucleic acids is an important endeavor. This research shows that more information (such as the join point) can be obtained using a nonlinear regression approach along with a standard linear model method for the initial parameter estimates. Further, if outliers are expected in the data more robust methods can be used to accomodate them. These methods help to determine the binding constants to the fullest accuracy allowed by the data. Thus, model (1) seems to provide a very reasonable approach to the analysis of binding constant data.

However, a cautionary note must be added. Model (2) does not seem to provide a good method for use in a ligand-DNA binding study since it provides unreliable parameter estimates based on a singular Jacobian Matrix. I therefore recommend only Model (1) be applied in these studies.

References

1. Axton, R.A. and Brock, D.J.H. (1994), "Use of stable dye - DNA intercalating complexes to detect systic fibrosis mutations, Molecular and Cellular Probes 8, 245-250.
2. Berry, R.J. (1972), Genetics, London: StPaul's House.
3. Delepierre, M, Delbarre, A, Langlois D'Estaintont, B., Igolen, J. and Roques, B.P. (1987), ¹H-NMR Studies of a Monointercalating Drug into a d[CpGpCpG]₂ Minihelix", Biopolymers, 26, 981-1000.
4. Drake, J.W., Glickman, B.W. and Ripley, L.S.. (1983), "Updating the Theory of Mutation", American Scientist 71, 621-631.
5. Dutter, R. (1977), Numerical solution of robust regression problems: Computational aspects, a comparison, J. Stat. Comput. Simulation, 5, 207-238.
6. Dutter, R. (1983a), Computer Program BLINWDR for Robust and Bounded Influence Regression, Research Report No. 8, Institute for Statistics, Technische Universitat Graz, Austria.
7. Dutter, R. (1983b), CONVINTER: A computer Program for Computing Robust Covariances and for Plotting Tolerance Ellipses, Research Report No. 10, Institute for Statistics, Technische Universitat Graz, Austria.
8. Dutter, R., and Huber, P.J. (1981), Numerical methods for the nonlinear robust regression problem, J. Stat. Comput. Simulation, 13, 79-114.
9. Foster, N., Singhal, A.K., Smith, M.W. Marcos, N.G. and Schray, K.J. (1988), "Interaction of Porphyrins and transfer RNA", Biochimica et Biophysica Acta 950, 118-131.
10. Hard, T. (1987), "Out-of-Plane Mobility in the Ethidium/DNA Complex, " Biopolymers 26, 613-618.
11. Hogg, R.V. (1979), "An Introduction to Robust Estimation" Launer, R.L. and Wilkinson, G.N., Eds, Robustness in Statistics, New York: Academic Press.
12. Ishizu, K., Dearman, H.H., Huang, M.T. and White, J.R. (1969), "Interaction of the 5-Methylphenazinium Cation Radical with Deoxyribonucleic Acid", Biochemistry 8, 1238-1246.
13. Jamison, JM, Krabill, K, Allen K.A., Stuart, S.H. and Tsai, C. C. (1990), "RNA-Intercalating agent interactions: In vitro antiviral activity studies, Antiviral Chemistry and Chemotherapy 1 (6), 333-347.

14. James, M.B., Hollstein, U. and, Allen, F.S., (1987), " Site Specificity of Binding of Antitumor Antibiotics to DNA," Biopolymers 26, 121-135.
15. Kozelka, J., Archer, S., Petsko, G.A., Cippard, S.T. and Quigley, G.J., "(1987), "Molecular Mechanics Modeling of Oligo nucleotide Adducts of the Antitumor Drug cis-Diamminedichloroplatinum(II)", Biopolymers 26, 1245-1271.
16. Lying, R., Hard, T. and Norden, B. (1987), "Induced CD of DNA Intercalators: Electric Dipole Allowed Transitions", Biopolymers 26, 1327-1345.
17. MacGregor, Jr, R.B. and Clegg, R.M., (1987), "Ethidium-DNA Binding Kinetics in an Agarose Gel", Biopolymers, 26, 2103-2106.
18. Nandi, R., Debnath, D. and Maiti, M. (1990), "Interactions of berberine with Poly (A) and tRNA", Biochimica et Biophysica Acta 1049, 339-342.
19. Nyssen, E., DiGiambattista, and Cocito, C. (1989), "Analysis of the reversible binding of Virginiamycin M to ribosome and particle functions after removal of the antibiotic", Biochimica et Biophysica Acta 1009, 39-46.
20. Price, N.C. and Swick, R.A. (1979), Principles and Problems in Physical Chemistry for Biochemists, Oxford: Clarendon Press.
21. Richardson, C.L. and Schulman, G.E. (1981), "Competitive Binding Studies of Compounds that interact with DNA utilizing Fluorescence Polarization", Biochimica et Biophysica Acta 652, 55-63.
22. Ross, W.E. Bradley, M.O. (1981), "DNA Double-Strand Breaks in Mammalian Cells after Exposure to Intercalating Agents", Biochimica et Biophysica Acta 654, 129-134.
23. Rousseeuw, P.J. and LeRoy, A.N. (1987), Robust Regression and Outlier Detection, New York: John Wiley and Sons.
24. Ruggiero, J., Xodo, L.E., Ciana, A., Manzini, G. and Quadrifoglio (1992), " Charge effect in the interaction of doxorubicin and derivatives with polydeoxynucleotides", Biochimica et Biophysica Acta 1129, 294-302.
25. SAS Institute Inc. (1990), SAS/STAT User's Guide, version 6, Fourth Ed. Volume 2, Cary N.C: SAS Institute.
26. Seber, G.A.F. and Wild, C.J. (1989), Nonlinear Regression. New York: John Wiley and Sons.
27. Smith, C.A. and Wood, E.J. (1991), Molecular Biology and Biotechnology, NY: Chapman and Hall.

28. Stradowski, C., Gorner, H., Currell, G.J. and Schulte-Frohlinde, D. (1987), "Number of Binding Sites of DNA and Polynucleotides with Tris (2, 2'-bipyridil) ruthenium (II) Cations", Biopolymers 26, 189-201. (bo[urodo;)
29. Tada, M., Ikeda, S., Suzuki, M., Minoura, Y, Kojima, M. and Morita, T. (1991), "Spectrophotometrical and immunochemical studies of the conformational changes in Poly (dG-dC). Poly (dG-dC) after Modification by 4- hydroxyaminoquinoline 1-oxide", Biochemica et Biophysica Acta 1090, 29-37.
30. Tornaletti, S, Russo, P, Parodi, S., and Pedrini, A.M. (1989), "Studies on DNA binding of Caffeine and Derivatives: evidence of intercalation by DNA unwinding experiments, Biochimica et Biophysica Acta. 1007, 112-118.
31. Van Holde, K.E. (1971), Physical Biochemistry, Englewood Cliffs, NJ: Prentice-Hall.
32. Vertino, P.M., Bergeron, R.J., Cavanaugh, P.F. and Porter, C.W., (1987), "Structural Determinants of Spermidine-DNA Interactions", Biopolymers 26, 691-703.
33. Wadsworth, Jr., H.M. (1990), Handbook of Statistical Methods for Engineers and Scientists, New York: McGraw-Hill.
34. Watts, D.G. and Bacon, D.W. (1974), "Using An Hyperbola as a Transition Model to Fit Two-Regime Straight-Line Data", Technometrics 16, 369-373.
35. Zhu, J., Sun, X. and Wang, F. (1991), "The DNA intercalator, ethidium bromide, alters the pattern of DNA case I hypersensitive sites of the B^A-globin gene in chicken erythrocytes", Biochemica et Biophysica Acta 1089, 158-169.
36. Zunino, F., Dimarco, A., Zaccara, A. and Cambetta, R.A. (1980), "The Interaction of Daunorubicin and Doxorubicin with DNA and Chromatin", Biochimica et Biophysica Acta 607, 206-214.
37. Lai, T.L. and Zhang, L. (1994), "Statistical Analysis of Ligand-Binding Experiments," Biometrics 50, 782-797.

Table 1
Binding Data From Ishizu et al.

Observation	$\bar{B}/F \times 10^{-4}$	$\bar{B}$
1	1.73	.088
2	1.30	.10
3	1.17	.104
4	.91	.11
5	.757	.116
6	.63	.119
7	.56	.123
8	.54	.131
9	.48	.141
10	.476	.162

Table 2
Model (1) Results

Parameter	Estimate	Binding Constants
α_0	5982.4408	
α_1	-186712.8028	
α_2	155632.5020	
x_0	0.1213	$K_I +3.42 \times 10^5$
γ	0.0134	$K_{II} +3.1 \times 10^4$

Table 3
Model (2) Results*

Parameter	Estimates	Binding Constants
β_0	4082.079	$K_I +3.92 \times 10^6$
β_1	28938.152	$K_{II} -3.98 \times 10^6$
β_2	3951100.05	
x_0	0.150	
δ	-0.128	

*In this analysis the Jacobian Matrix was singular.

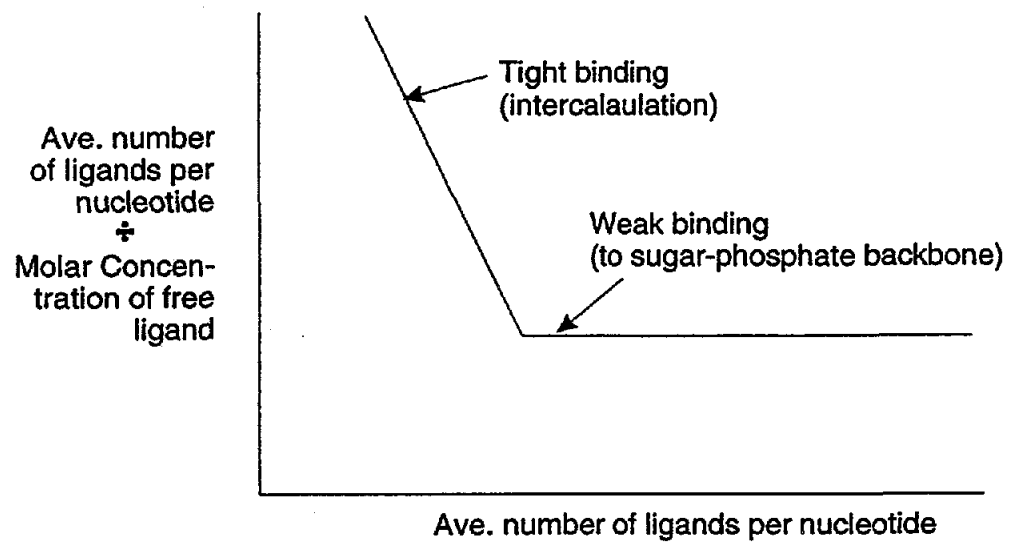


Figure 1. Biphasic Scatchard Plot for Ligand-DNA Interaction

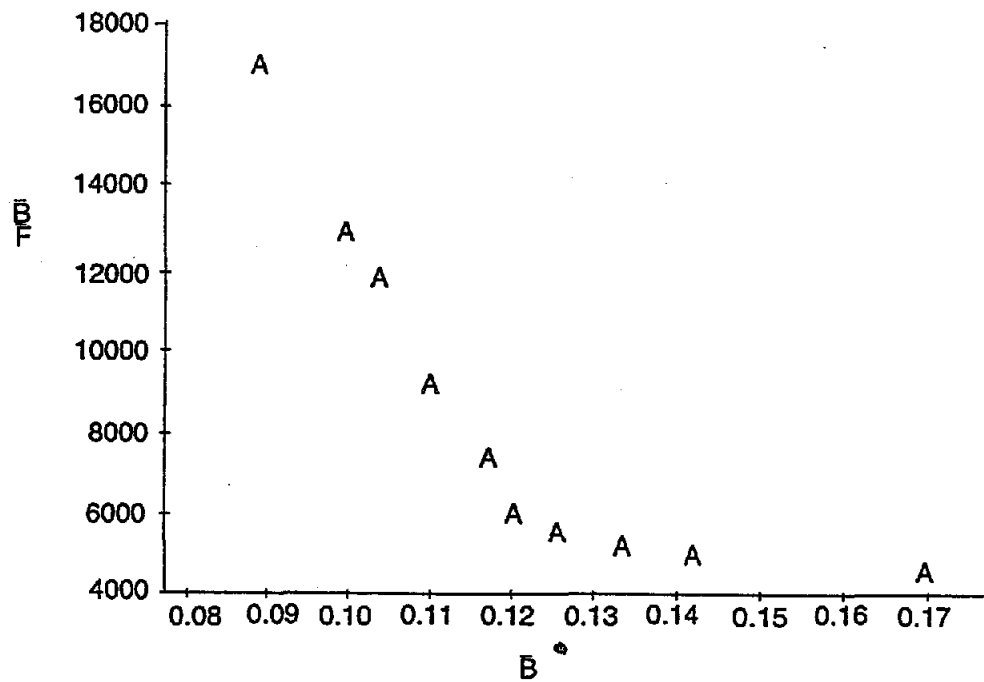


Figure 2. Plot of Table 1 Data

```

*REF EX 3 P 17. 14 WADSWORTH, HBOOK OF STAT METH FOR ENG & SCIENT. 1989;
TITLE 'HHD DATA COMPUTATIONS';
* DATA FROM BIOCHEM 8,1238 (1969), MODEL TECHNOMETRICS 16,369 (1974);
* START VALUES FROM GRAPH;
* USING ACTUAL DATA 6/13/94;
DATA A;
INPUT X Y @@;
CARDS;
.088 1.73e4
.1 1.3e4
.104 1.17e4
.11 .91e4
.116 .757e4
.119 .63e4
.123 .56e4
.131 .54e4
.141 .48e4
.162 .476e4
;PROC PLOT; PLOT Y*X;
PROC NLIN METHOD = DUD;
PARMS AO=.59e4 XO=.122 A1=-2.03E5 A2=1.97e5 GAM=.1;
MODEL Y=AO + A1*(X-XO) + A2*(X-XO)*TANH(X-XO)/GAM);

```

Fig. 3. SAS Program for the Analysis