ANALYSIS OF DNA-RNA HYBRIDIZATION DATA USING THE SCATCHARD PLOT

J. Lawrence Marsh and Brian J. McCarthy*

Departments of Biochemistry and Genetics
University of Washington
Seattle, Washington 98195

Received October 15,1973

SUMMARY: DNA-RNA hybridization experiments, especially in eukaryotic systems, are complicated by such factors as large genome size and the difficulty of separating RNA fractions. Current analytical methods can lead to erroneous conclusions since heterogeneity is often obscured and data in the most useful concentration ranges are not readily obtainable. Some of these difficulties can be overcome by using a Scatchard plot which is superior to other methods in preventing common misinterpretations of data resulting from heterogeneous RNA samples or non-specific reaction.

In principle, the titration of single-stranded DNA with excess RNA provides a quantitative estimate of the fraction of the genome coding for that particular species or group of RNA molecules. Ideally the binding should be measured over a broad range of RNA and DNA concentrations. In practice, however, the experimental design is limited by such factors as the specific radioactivity of the RNA or DNA, the amount of RNA available and its solubility. It is often difficult to achieve saturating RNA concentrations. Decreasing the amount of DNA in the reaction is not always practical since detection of very small amounts of binding requires rather high specific radioactivities or excessively long periods of incubation. The approach to saturation in such reactions is asymptotic with increasing RNA concentration, presenting obvious difficulties in extrapolation (1-3). Indeed, in order to determine the final saturation values, experiments are often performed at the highest possible RNA concentrations. It is exactly these data which tend to maximize the contribution of dilute contaminants and non-specific reaction.

A method of anlaysis allowing a linear extrapolation to infinite RNA

---

*Present address: Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94122.

concentration is therefore desirable. Although the double reciprocal plot, as applied by Bishop to DNA/RNA hybridization, partially fills this need (4), Scatchard's analysis of the binding of small ions by proteins (5) serves this purpose in a manner superior to other available treatments. The function is bounded on both ends by meaningful intercepts; any range of concentrations can be conveniently and usefully displayed; and the points are spread more evenly along the line. It is particularly advantageous for distinguishing between specific and non-specific binding and for analyzing the binding of an enriched component in the presence of heterogeneous contaminants.

THEORETICAL:

The equation describing the function can be derived as follows: In a DNA-RNA hybridization reaction, self reaction of the DNA is prevented either by imobilization of the DNA on filters or by dilution. Reaction mixtures containing increasing concentrations of RNA relative to DNA are allowed to come to equilibrium and the fraction of DNA in heteroduplex is measured. The reaction equation can be written as:

$$[D] + [R] \rightleftharpoons [DR] \qquad (1)$$

It should be noted here that the reaction described in equation (1) considers only those sequences of DNA which can react with RNA, hence the term [D] or unreacted DNA refers only to that unreacted DNA which could eventually react with RNA. [R] refers to the concentration of free RNA. [DR] is the concentration of hybridized DNA-RNA complex, all concentrations being in terms of moles of nucleotides per liter. The association constant K_a is given by:

$$K_a = [DR]/[D][R] \qquad (2)$$

The maximum possible extent of hybridization $[DR]_{max}$ is by definition equal to the sum of the bound DNA [DR] and the unreacted DNA [D], i.e., $[DR]_{max} = [DR] + [D]$. Multiplying equation (2) by [D] and substituting $[DR_{max} - DR]$ for D yields:

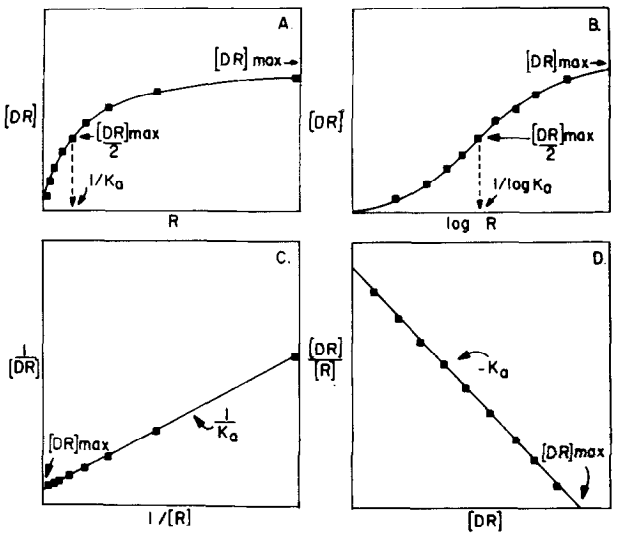

Figure 1. Binding of a single RNA species to a single DNA site is depicted in 4 ways. (A) direct plot, (B) semi log plot, (C) double reciprocal plot, (D) Scatchard plot.

$$[DR]/[R] = K_a([DR]_{max} - [DR]) \qquad (3)$$

A plot of DNA hybridized/free RNA against DNA hybridized is a straight line, whose x intercept is $[DR]_{max}$ the value for the maximum amount of DNA in hybrid.

DISCUSSION:

There are four methods of plotting DNA-RNA binding data to obtain the association constant K_a and the maximum extent of hybridization. These are [DR] vs. [R], [DR] vs. log [R], 1/[DR] vs. 1/[R], and [DR]/[R] vs. [DR] (Fig. 1). Of these four, the last, as described in this paper, is to be prefered for both theoretical (6) and practical reasons. The direct plot [DR] vs. [R] suffers from the fact that the curve approaches the limiting value of $[DR]_{max}$ asymptotically. Prohibitively high RNA concentrations are sometimes needed to approach saturation and extrapolation is difficult in all but the simplest cases. Heterogeneity in a sample is not readily apparent in either the linear or logarithmic presentation. The double reciprocal plot is useful in that it gives a straight line if all the binding sites are equivalent. However, it is difficult to plot all of the

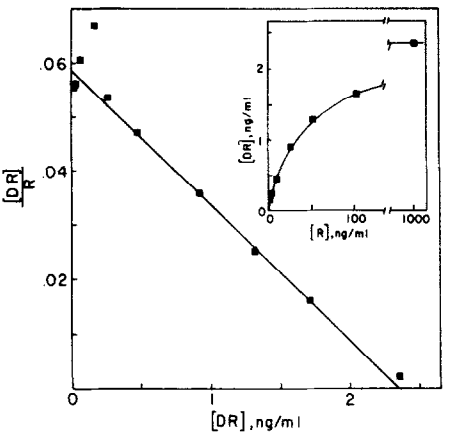

Figure 2. Hybridization of chicken 28s rRNA to chicken DNA immobilized on nitro-cellulose filters. RNA was labeled with ^{125}I to 2.64 x 10^7 cpm μg^{-1} (7). Filters containing 12.5 µg DNA were incubated in .1 ml 2xSSC at 70°C.

data in a meaningful way since the scale of both axes is not limited by a finite number. This results in low values of [R] being displaced far to the right where they have a disproportionately large influence on the slope of the line. Conversely, high values of [R] are so close to the ordinate as to obscure curvature.

A Scatchard plot on the other hand weights all of the theoretically obtainable data evenly and displays it in a useful manner. Furthermore, the function is bounded on both ends by meaningful intercepts: the y intercept is $K_a[DR]_{max}$, the x intercept is $[DR]_{max}$, maximum extent of reaction, and the slope is $-K_a$. Any other linear function goes to infinity on at least one axis, thus making extrapolation difficult.

In the case of multicomponent systems the curve described will be the sum of the straight lines for the individual components, a feature which makes heterogeneity readily apparent and facilitates analysis of complex reactions. For purposes of illustration, consider experiments designed to titrate the number of genes coding for "28s" rRNA in the chicken. Using a pure fraction of highly labeled "28s" RNA, increasing amounts of RNA were incubated with filters containing chicken DNA (7). The difficulty of accurately determining the final saturation value by extrapolation of the

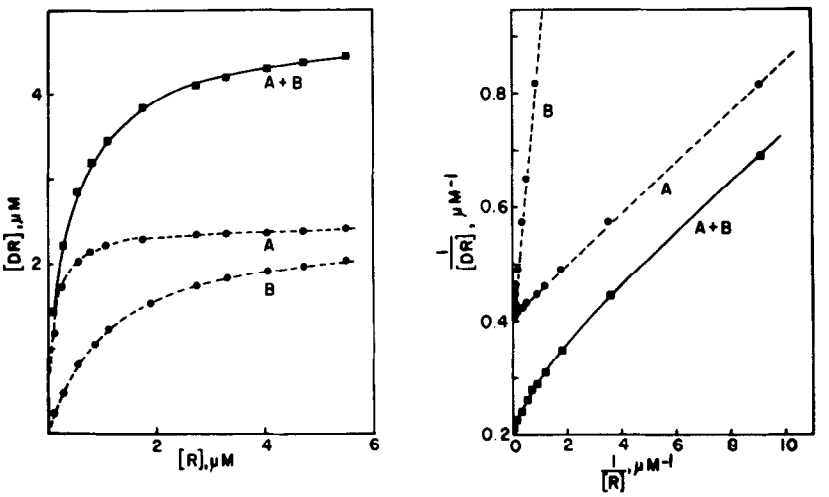

Figure 3. Binding behavior was calculated for a sample containing 2 RNA species, A and B, in 10:1 relative concentration. The K_a for both species was taken to be 10^7. The total concentration of DNA was taken to be 50 μg/ml 0.1% of which was complementary to each site A and B. A single RNA species binding to two DNA sites with different K_a in a 10:1 ratio produces the same result. -■- binding of the mixture, --●-- binding of the individual components as a function of total $[R] = [R_A] + [R_B]$.

binding curve (inset, Fig. 2) is clear. A ten-fold increase in RNA concentration was necessary to increase binding from 75% to 99% of saturation. The extrapolation to infinite [R] is facilitated by plotting these data according to equation 3 (Fig. 2). From the slope of the line the association constant is $8.6 \times 10^6\ M^{-1}$, a value consistent with the values of $2.6 \times 10^6\ M^{-1}$ reported for E. coli RNA/DNA hybrids (8,9).

Several factors, such as GC content, base pair mismatching, non-specific association and differing relative concentrations influence the value of the binding constant. With a mixture of several RNA species, the apparent binding constant for each species depends on its relative concentration in the mixture. This allows one to distinguish between an enriched component and a number of dilute contaminants. Similarly, locus specific hybridization can be distinguished from non-specific reaction. In both cases the curve will break into two sections, one displaying a decreased slope reflecting the binding of dilute contaminants or non-specific association. Two such cases will be illustrated below.

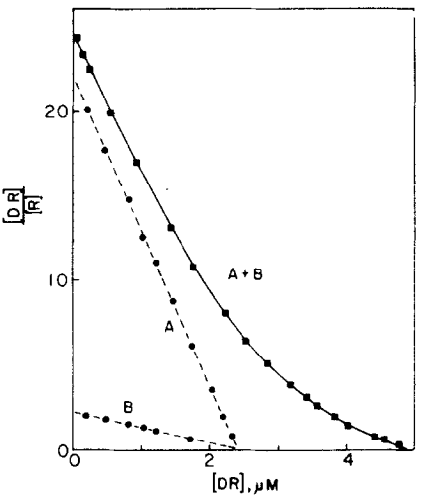

Figure 4. Data from Fig. 3 analyzed by the Scatchard method.

In the first, the binding behavior of a single RNA species which binds to two DNA sites with different association constants was calculated. This situation resembles non-specific binding. Secondly, the behavior of a sample containing a 10:1 ratio of two RNA species binding to two DNA sites with the same binding constant was calculated to simulate a 9% contamination of an RNA sample. The binding behavior of these two cases is identical. It is displayed as direct and double reciprocal plots in Fig. 3 and as a Scatchard plot in Fig. 4. Only by considering other parameters of the reactions such as the thermal stability of the product can these two cases be distinguished (10). In either case the direct plot offers no evidence for heterogeneity and the values for saturation of (A) the component of interest are grossly overestimated. The Scatchard plot of Fig. 4 deviates strikingly from linearity indicating that the sample is heterogenous and must be analyzed as a multiple binding mixture. The advantages of the Scatchard plot can be seen by comparing Fig. 4 to a double reciprocal plot of the same data (Fig. 3).

The Scatchard analysis makes maximum use of the data obtained between 20 and 80% saturation in determining the final saturation values and association constants. Rather than maximizing the effect of dilute contaminants and

non-specific reaction, these data make it possible to identify the binding of the component of interest despite the concomitant binding of multiple unknown contaminants. Moreover, these data are subject to the least percentage error and are theoretically most significant (6). The fact that this plot is bounded at both ends allows one to determine immediately by visual inspection what portion of the total titrant has been performed. The best fit to the data can be determined by any of several published methods (11,12).

Acknowledgment

We are greatly indebted to Dr. A. Tereba for the use of his unpublished data (Fig. 2), and to Dr. David A. Deranleau for his stimulating discussions. This research was supported by a predoctoral fellowship (USPHS Grant #GM00052) and USPHS Grant #GM20287 to B. J. McCarthy.

REFERENCES:

1. Bishop, J. O. (1972) Karolinska Symposia on Research Methods in Reproductive Endocrinology, 5th Symposium, Gene Transcription in Reproductive Tissue, pp. 247-273.
2. Walker, P. M. B. (1969) Prog. Nucleic Acid Res. Mol. Biol. 9, 301-326.
3. Birnstiel, M. L., Chipchase, M. and Speirs, J. (1971) Prog. Nucleic Acid Res. Mol. Biol. 11, 351-389.
4. Bishop, J. O. (1969) Nature 224, 600-603.
5. Scatchard, G. (1949) Ann. N.Y. Acad. Sci. 51, 660-672.
6. Deranleau, D. A. (1969) J. Amer. Chem. Soc. 91, 4044-4049.
7. Tereba, A. and McCarthy, B. J. (1973) Biochemistry, in press.
8. Bishop, J. O. (1970) Biochem. J. 116, 223-228.
9. Kennell, D. E. (1971) Prog. Nucleic Acid Res. Mol. Biol. 11, 259-301.
10. Farquhar, M. N. and McCarthy, B. J. (1973) Biochemistry (in press).
11. Deranleau, D. A. (1969) J. Amer. Chem. Soc. 91, 4050-4054.
12. Feldman, H. A. (1972) Anal. Biochem. 48, 317-338.