Reprint from *B. I. F. Futura* (1997) **12**, 20-26

Analysis of protein-DNA binding at equilibrium

Karsten Rippe

Deutsches Krebsforschungszentrum, Abt. Biophysik der Makromoleküle, Im Neuenheimer Feld 280, D-69120 Heidelberg.

Tel.: 06221-423392
Fax: 06221-423391
e-mail: Karsten.Rippe@dkfz-heidelberg.de

Key words: protein-DNA interactions, binding curves, equilibrium binding, binding constant, dissociation constant, cooperativity, stoichiometry

Summary

The analysis of protein-DNA binding is an important aspect of many processes in the field of molecular biology. In the present article the principles for a quantitative analysis of equilibrium protein-DNA interaction are reviewed. The simplest system involves the binding of a protein to a single site on a DNA molecule. This can be analyzed conveniently by the equations developed in the text. In many cases, however, multiple protein ligands bind to a certain DNA target *in vivo*. This requires an extension of the analysis to include the occurrence of multiple binding sites that can have different affinities for a given protein (heterogeneity) and/or interactions between the binding of multiple proteins (cooperativity). Binding interactions of this type are found frequently *in vivo* as they offer an increased flexibility in terms of regulation. To illustrate the effect of cooperative binding an example is discussed for the binding of a protein to two identical binding sites comparing either independent or cooperative binding. Several graphical representations of the corresponding binding curves are presented, and it is shown how their shape reflects the different binding parameters.

Introduction

The binding of proteins to DNA plays an important role in virtually all biological processes that involve DNA, as for example the transcription of DNA to make messenger RNA. Obviously a protein complex like RNA polymerase has to bind to DNA to be able to perform a DNA template directed RNA synthesis. In addition the activity of RNA polymerase is modulated by protein activators and repressors that also bind to the DNA and are able to control the initiation, elongation or termination phase of transcription [for a review see for example (von Hippel *et al*., 1996)]. A defined temporal and spatial expression pattern of the RNA is required for developmental regulation and tissue specific metabolism. It is created by a large number of gene-specific transcription factors, the majority of which is involved in DNA binding.

A very simple regulation mechanism would be that of a repressor binding at the transcription start site of a specific gene and thereby preventing transcription initiation by RNA polymerase. In this case the DNA binding activity of the repressor is directly related to its ability to regulate transcription. From a mechanistic point of view it is now important to understand under which conditions the transcription initiation site is fully or partially occupied by the repressor and how the binding affinity is modulated. To pursue these questions a quantitative analysis of the binding interaction is necessary. From *in vitro* experiments that analyze the binding process a number of parameters can be determined such as: (i) how many protein ligands bind per DNA molecule (the stochiometry of binding), (ii) the binding affinity or equilibrium binding constants for a given site as determined by the DNA sequence, (iii) the specificity of the binding interaction, which can be deduced from comparison with the binding affinity to other DNA sequences, and (iv) the effect of already bound protein on the binding of additional proteins (cooperativity of binding).

Here I would like to cover some general aspects of the quantitative analysis of binding curves using a few specific examples. A thorough review of this subject is beyond the scope of the present paper and the reader is referred to Refs. (Cantor & Schimmel, 1980; Weber, 1992; Wyman & Gill, 1990) for further information.

General aspects of binding analysis

For the reversible binding reaction of a protein ligand that is binding to a DNA molecule with a single site to form a protein-DNA complex we can write the binding reaction as shown in Eq. 1.

$$D_{\text{free}} + P_{\text{free}} \rightleftarrows DP \qquad K_1 = \frac{D_{\text{free}} \cdot P_{\text{free}}}{DP} \qquad [1]$$

D_{free} denotes the concentration of free DNA strands , P_{free} is the concentration of free protein ligand in the solution and DP the concentration of the protein-DNA complex with one protein molecule bound. The corresponding equilibrium dissociation constant for the reaction is denoted here as K_1 and is given in Eq. 1 according to the mass-action law. The equilibrium association constant of the reaction would equal $1/K_1$.

If the DNA molecule has not only one but two binding sites for the protein, an additional binding reaction can occur, which is shown in Eq. 2. Here DP_2 is the concentration of the protein-DNA complex with two protein molecules bound and a dissociation constant K_2. Corresponding expressions can be derived for the binding of additional protein ligands.

$$DP + P_{\text{free}} \rightleftarrows DP_2 \qquad K_2 = \frac{DP \cdot P_{\text{free}}}{DP_2} \qquad [2]$$

What is usually determined experimentally is the fractional degree of saturation of the available protein binding sites on the DNA at increasing protein concentrations. This can be expressed as the concentration of bound protein ligand P_{bound} to the total concentration of DNA molecules D_{tot}.

$$\nu = \frac{P_{\text{bound}}}{D_{\text{tot}}} \qquad [3]$$

In the case of one binding site (n=1) ν is given by Eq. 4 or for n=2 by Eq. 5 and so on. A value of ν= 0 would mean that no protein is bound whereas a value of ν= 1 (for n=1) or ν= 2 (for n=2) would indicate that all binding sites on the DNA are occupied with protein.

$$\nu_1 = \frac{DP}{D_{free} + DP} \qquad [4]$$

$$\nu_2 = \frac{DP + 2 \cdot DP_2}{D_{free} + DP + DP_2} \qquad [5]$$

Alternatively, the degree of binding can be expressed in terms of the fraction θ of sites occupied with respect to the total number of binding sites. For a DNA with *n* binding sites the total number of sites on the DNA is given by $n \cdot D_{\mathrm{tot}}$ so that

$$\theta = \frac{P_{\mathrm{bound}}}{n \cdot D_{\mathrm{tot}}} \qquad [6]$$

which is equivalent to:

$$\theta = \frac{\nu}{n} \qquad [7]$$

Binding to a single site

Let us now examine how binding parameters are determined. In the simplest case there is only a single protein binding site present per DNA molecule. A typical experiment would consist of adding various amounts of protein to a solution with DNA molecules and to measure the amount of protein or DNA bound by some technique. This could be done for example by electrophoresis mobility shift experiments (Fried, 1989; Revzin, 1989), nitrocellulose filter binding (Wong & Lohman, 1993), equilibrium dialysis (Takahashi *et al.*, 1989), or some type of fluorescence assay (Perez-Howard *et al.*, 1995). Any technique that allows it to quantitate the amount of protein-DNA complexes in relation to the free DNA and/or the free protein at various concentrations of the reactants can be used. The resulting data set is termed a binding curve or titration isotherm. Its shape will depend on the concentration of DNA and protein and the affinity to the given binding site as given by the dissociation constant. The latter is usually a function of the solution conditions (salt, pH,

temperature, etc.). Examples of binding curves for the binding of a protein to a single DNA binding site are shown in Fig. 1 and 2.

If the DNA concentration in the solution is much larger than the dissociation constant ($D_{tot} >> K_{dis}$) virtually every protein molecule added will be bound until all sites are occupied. This results in a linear increase of ν until the point is reached at which all the binding sites are saturated. An example is shown in Fig. 1. Such a titration is termed a stochiometric titration because only the stochiometry of the complex, i. e. how many protein ligands are bound to a DNA molecule can be determined, provided that the DNA concentration and the concentration of active protein are known. If on the other hand the number of binding sites is known, a stochiometric titration can be used to determine the concentration of active protein. This parameter can be derived from the equivalence point that indicates complete occupancy of all the binding sites and the known DNA concentration (Fig. 1).

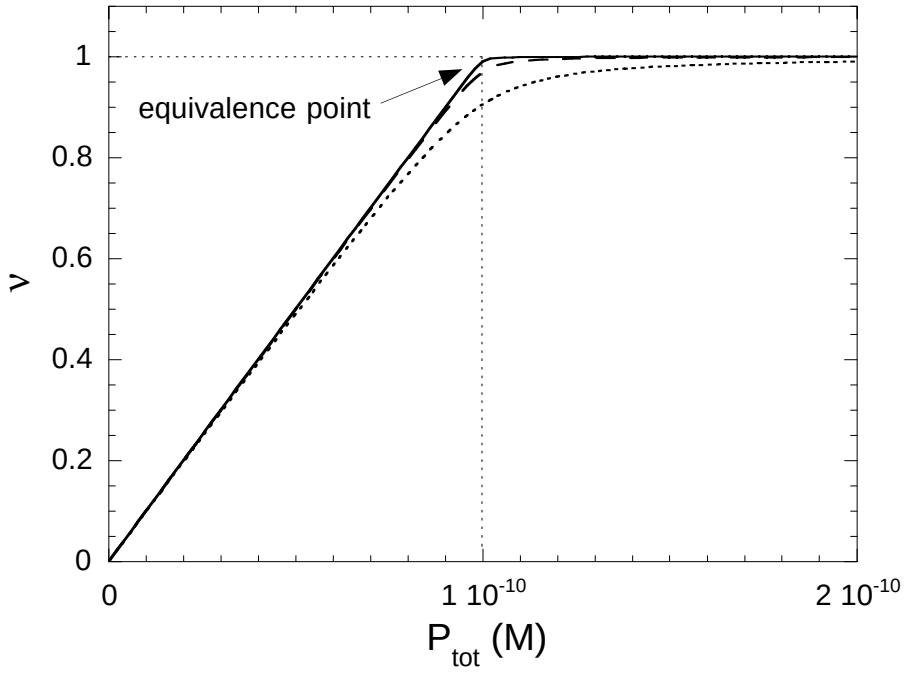

Figure 1: Stoichiometric binding of a protein to a DNA with a single binding site. The concentration of DNA molecules is $D_{tot} = 10^{-10}$ M. The dissociation constant for the three binding curves shown are $K_{dis} = 10^{-14}$ (———), $K_{dis} = 10^{-13}$ (- - - -), and $K_{dis} = 10^{-12}$ (·····). Only in the very small region around the equivalence point the difference in the dissociation constant can be detected and above a certain binding affinity at about $K_{dis} \leq 10^{-13}$ the curves become indistinguishable.

From an experiment under conditions of stochiometric binding K_{dis} cannot be determined accurately. Only a limit for the magnitude of the dissociation constant can be given. To derive the dissociation constant from a binding curve, it is required to have a significant amount of free DNA, free protein, and the complex to be present in the solution. This is because the initial linear increase of binding with added protein and the end part of the curve where all sites are occupied are independent of K_{dis}. Only from the middle region of the curve at which free DNA, free protein, and the complex are present the dissociation constant can be determined. In general it is required that $D_{tot} \leq K_{dis}$ (as opposed to the situation that $D_{tot} >> K_{dis}$ described above) to obtain a sufficiently large region of the binding curve from which the dissociation constant can be obtained.

Now in order to determine the binding constant from the experimental data we need to express the observed degree of binding (i. e. ν or θ) as a function of the known parameters (protein concentration and the DNA concentration) plus the unknown dissociation constant, and then derive the binding constant by fitting the experimental data to the resulting function.

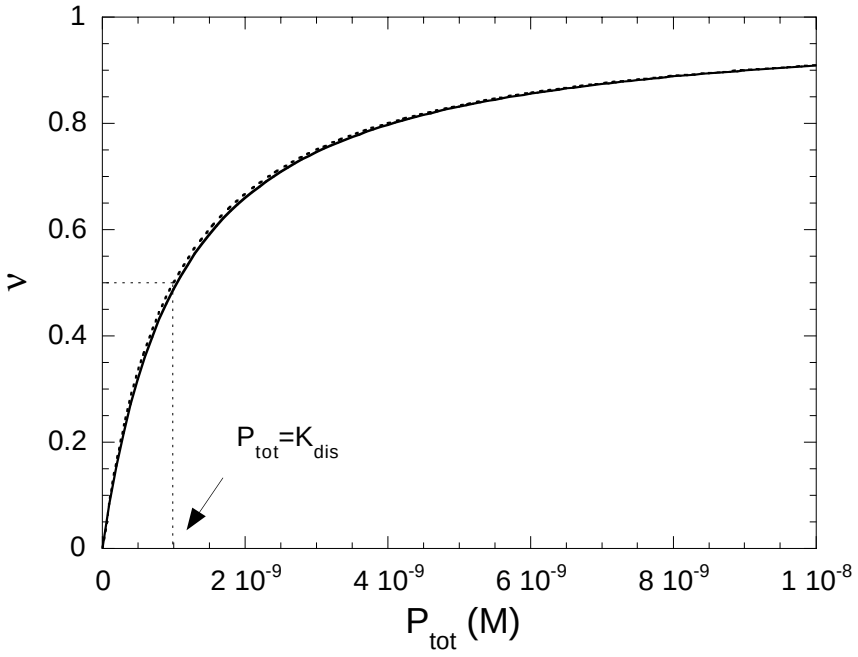

Figure 2: Theoretical binding curves according to Eqs. 9 (———) and 10 (·····) for a titration of a DNA with one binding site at a concentration of $D_{tot} = 10^{-10}$ M. The two curves correspond to a dissociation constant of $K_{dis} = 10^{-9}$. The dotted curve has been generated with the assumption that $P_{free} \approx P_{tot}$. The dissociation constant in this case is defined as the protein concentration at which half of binding sites on the DNA are occupied by a protein, i. e. $\nu = 0.5$.

In Eq. 1, *DP* can be expressed in terms of D_{free}, P_{free} and K_1 (= K_{dis}). Then *DP* is substituted in Eq. 4 to obtain Eq. 8.

$$\nu_1 = \theta_1 = \frac{P_{free}}{K_{dis} + P_{free}} \qquad [8]$$

Thus from a plot of ν versus the concentration of free protein P_{free} one can determine the dissociation constant by fitting the data points to Eq. 8. However, frequently the concentration of free protein cannot be measured directly and only the total protein concentration is known. There are two solutions to this problem. One is to express P_{free} in terms of P_{tot} and D_{tot}. By using Eq. 3 we can write $P_{free} = P_{tot} - \nu \cdot D_{tot}$. The resulting expression for ν_1 is shown in Eq. 9.

$$\nu_1 = \frac{D_{tot} + P_{tot} + K_{dis} - \sqrt{\left(D_{tot} + P_{tot} + K_{dis}\right)^2 - 4 \cdot D_{tot} \cdot P_{tot}}}{2 \cdot D_{tot}} \qquad [9]$$

The second option is to choose the experimental conditions so that $P_{free} \approx P_{tot}$, and Eq. 8 can be rewritten as:

$$\nu_1 = \frac{P_{tot}}{P_{tot} + K_{dis}} \quad \text{if} \quad P_{free} \approx P_{tot} \qquad [10]$$

By comparing Eq. 9 and 10 it is evident that already for the simplest type of binding equilibrium the analysis becomes much easier if experimental conditions can be found that allow measurements under the condition that $P_{free} \approx P_{tot}$. In addition the exact concentration of the DNA needs not to be known as it does not enter Eq. 10. If the reaction becomes more complex (multiple binding sites and/or cooperative binding, see below) it is even more important to find conditions where this is true because the derivation of an expression corresponding to Eq. 9 is not straightforward.

When is the approximation that $P_{free} \approx P_{tot}$ valid? We can relate this question to the dissociation constant. If a certain amount of protein is added to the DNA solution and only a small fraction binds, then we can assume $P_{free} \approx P_{tot}$. For all practical purposes this means that we can use the simplified expression for ν or θ, if the dissociation constant K_{dis} is at least 10 times higher than the total DNA concentration in the solution. As it is shown for this

case in Fig. 2 (D_{tot} = 10^{-10} M, K_{dis} = 10^{-9} M) Eqs. 9 and 10 yield almost identical results. The practical problem with lowering the DNA concentration significantly below the value of the dissociation constant is that the measured signal is often proportional to the DNA concentration. It will get noisier the lower the DNA concentration is. Typically we need nanomolar concentration for measurements that involve a fluorescence signal (10^{-8} to 10^{-10} M). Using a radioactive label on the DNA the detection limit is in the range of 10^{-10} to 10^{-13} M. This means that for very tightly binding proteins (i. e. $K_{dis} \leq 10^{-10}$ M) only with a radioactive probe it is possible to measure under conditions where the approximation $P_{free} \approx P_{tot}$ can be used to simplify the analysis, if P_{free} cannot be determined.

Complex binding interactions

In the previous paragraphs we have dealt with the binding of a protein ligand to a single binding site. This is the simplest case which describes the binding of some proteins to a short DNA fragment with their respective binding site adequately. However, in biological systems we often find that DNA binding of a certain protein involves more than one binding site.

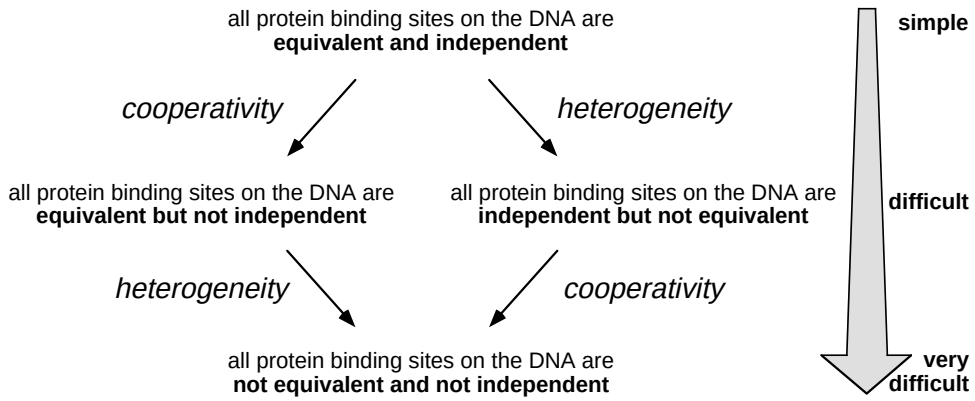

Figure 3: Increasing complexity of binding interactions if multiple sites are present. The complexity of the system increases from the top to the bottom. The figure is adapted from a scheme by Peter H. von Hippel.

In addition the different binding sites can vary in their binding affinity (heterogeneity) or binding of proteins can be influenced by already bound proteins (cooperativity), either

facilitating or impeding the binding of additional proteins (see Ref. (Ackers *et al*., 1982) for an example). This leads to an increasingly complex system of binding interactions as shown in the scheme in Fig. 3. A detailed treatment of the various scenarios is beyond the scope of this review. In practice it turns out that even for good data it is usually very difficult to distinguish between heterogeneity and cooperativity, if more than two sites are present. Furthermore, it is almost impossible to derive accurate binding parameters from real data, if binding is cooperative and the binding sites are not equivalent. Accordingly, I will focus in the following on the binding of a protein to two equivalent sites and introduce the most important concepts for binding to multiple sites.

If more than one binding site is present, it becomes important to distinguish between intrinsic or microscopic binding constants, denoted here by a small letter *k*, and the macroscopic binding constants, denoted here by a large letter *K* as they are defined in Eq. 1 and 2. The microscopic binding constant *k* refers to the binding to a specific site, out of the *n* sites present on every DNA strand. In contrast the macroscopic binding constant refers to the formation of a certain species. For example, on the macroscopic level the species with one protein bound can consist of a variety of microspecies, with protein being bound to site 1 or to site 2, etc. What is usually measured is the macroscopic binding constant, because experimentally the different microspecies cannot be distinguished. In the case of a DNA with a single protein binding site, the microscopic and macroscopic binding constants are the same, because only one species can be formed.

For *n* independent and identical binding sites the expression for ν can be developed from Eq. 8 by simply treating each binding site as independent and summation over all binding sites.

$$\nu_{\mathrm{n}} = \frac{n \cdot P_{\mathrm{free}}}{k_{\mathrm{dis}} + P_{\mathrm{free}}} \qquad [11]$$

In this expression k_{dis} is the intrinsic or microscopic dissociation constant for one site. If on the other hand the binding occurs with infinite cooperativity, that is only the free and the fully occupied species appear in the reaction equilibrium, we can write the reaction as:

$$D + n \cdot P_{\mathrm{free}} \rightleftarrows DP_{\mathrm{n}} \qquad K_n = \frac{D \cdot P_{\mathrm{free}}^{\mathrm{n}}}{DP_{\mathrm{n}}} \qquad [12]$$

In Eq. 12, DP_n can be expressed in terms of D_{free}, P_{free} and K_n, the apparent dissociation reaction. Substitution of DP_n in the corresponding expression for ν as defined in Eq. 3 yields Eq. 13.

$$\nu_n = \frac{n \cdot P_{free}^n}{K_n + P_{free}^n} \qquad [13]$$

In terms of cooperativity Eq. 11 and 13 define the two extreme cases, i. e. either completely independent binding according to Eq. 11 or binding with infinite cooperativity as described by Eq. 13. By analogy to Eq. 11 and 13 a semiempirical description is often used (Hill, 1910) for the cooperative binding to n equivalent sites over a part of the saturation range (typically 25 to 75 %) in the form given below.

$$\nu_n = \frac{n \cdot P_{free}^{\alpha_H}}{K^{\alpha_H} + P_{free}^{\alpha_H}} \qquad [14]$$

The parameter α_H is known as the Hill constant or coefficient with $1 \le \alpha_H \le n$. This means when $\alpha_H = 1$, there is no cooperativity between the sites and we would obtain Eq. 11. If on the other hand $\alpha_H = n$ we would have completely cooperative binding (Eq. 13).

Exact expressions that make no assumption on the type of binding and are valid over the whole range of ligand concentrations can be derived from the various steps of the reaction. For the case of the binding of two proteins this is done by using Eq. 1 and 2 to express DP and DP_2 in terms of D_{free}, P_{free} and K_1 and K_2 and then substitute DP and DP_2 in Eq. 5 to obtain:

$$\nu_2 = \frac{K_2 \cdot P_{free} + 2 \cdot P_{free}^2}{K_1 \cdot K_2 + K_2 \cdot P_{free} + P_{free}^2} \qquad [15]$$

As defined in Eqs. 1 and 2, K_1 and K_2 are the macroscopic dissociation constants for the binding of the first and the second protein, respectively. If the intrinsic affinity for the two sites is the same, i. e. the microscopic dissociation constant k_{dis} is identical, and binding to the two sites is independent (no cooperativity), one would nevertheless observe a difference in the magnitude of K_1 and K_2 which is given by Eq. 16.

$$\frac{K_1}{K_2} = \frac{k_{\mathrm{dis}}/2}{2 \cdot k_{\mathrm{dis}}} = \frac{1}{4} \qquad [16]$$

This indicates that the binding of the first protein occurs with a four times lower macroscopic dissociation constant as compared to binding of a second protein to the other sites, although the sites are identical and the intrinsic or microscopic binding site k_{dis} is the same. This is a statistical consequence of the presence of multiple sites. For the binding of the first protein there are two sites available, which can be removed in only one way. This leads to a factor of 2 for the formation of the complex *DP* corresponding to a factor of 1/2 for its dissociation. In contrast there is only one way to add a second protein to the *DP* complex, but two ways to remove a protein from the DP_2 conformation, leading to the statistical factor of $K_2 = 2 \cdot k_{\mathrm{dis}}$.

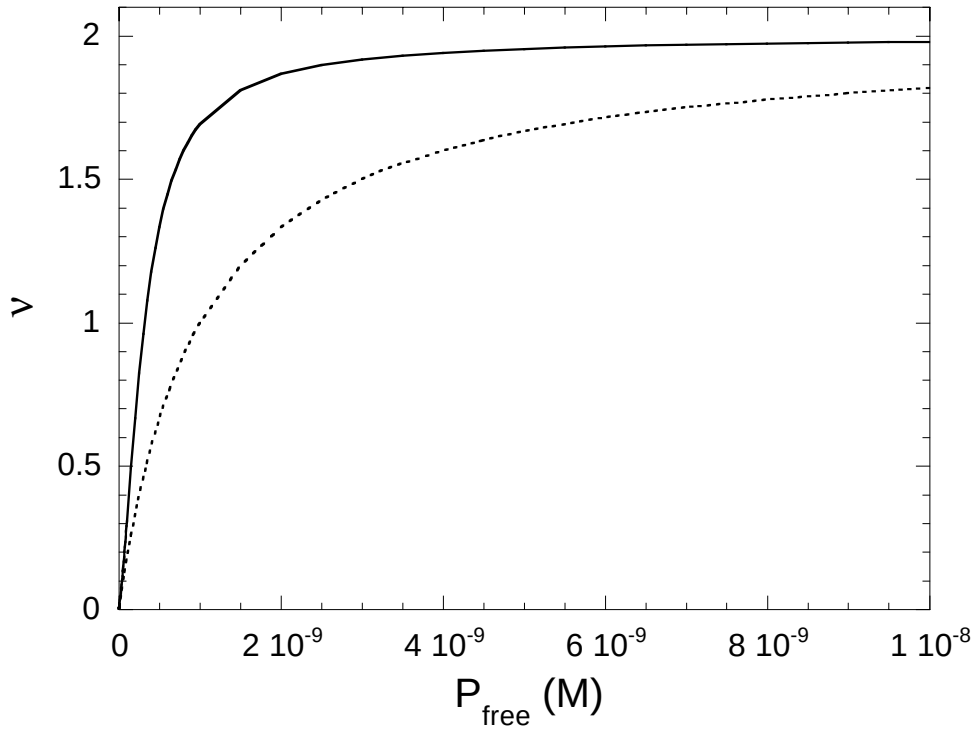

Figure 4: Binding of a protein to a DNA with two equivalent binding sites. The microscopic dissociation constant for one site is $k_{\mathrm{dis}} = 10^{-9}$ M. Two curves are shown: The lower curve (·····) represents independent binding where $K_1 = k_{\mathrm{dis}}/2 = 5 \cdot 10^{-10}$, $K_2 = 2 \cdot k_{\mathrm{dis}} = 2 \cdot 10^{-9}$ and $\omega = 0.25$. The upper curve (———) displays cooperative binding where binding of the second protein occurs with a ten fold higher affinity as compared to the case of independent binding with $K_1 = 5 \cdot 10^{-10}$ and $K_2 = 2 \cdot 10^{-10}$ and $\omega = 2.5$.

If binding to the two sites is not independent, we have to introduce an additional parameter, namely the cooperativity constant ω to describe the binding. We define:

$$\frac{K_1}{K_2} = \omega \qquad [17]$$

According to the previous considerations a value of ω = 0.25 would mean that the microscopic binding constant k_{dis} is the same for the two sites, whereas a larger value of ω would indicate that binding of the second protein is facilitated and a lower value that it is inhibited (anticooperativity). In the following the binding of a protein to two identical sites with a microscopic dissociation constant $k_{dis} = 10^{-9}$ M is analyzed for two examples: In the first case, binding to the two sites is independent with $K_1 = k_{dis} / 2 = 5 \cdot 10^{-10}$ and $K_2 = 2 \cdot k_{dis} = 2 \cdot 10^{-9}$. In the second case binding occurs with a cooperativity factor of ω = 2.5, and the macroscopic dissociation constants $K_1 = 5 \cdot 10^{-10}$ and $K_2 = 2 \cdot 10^{-10}$.

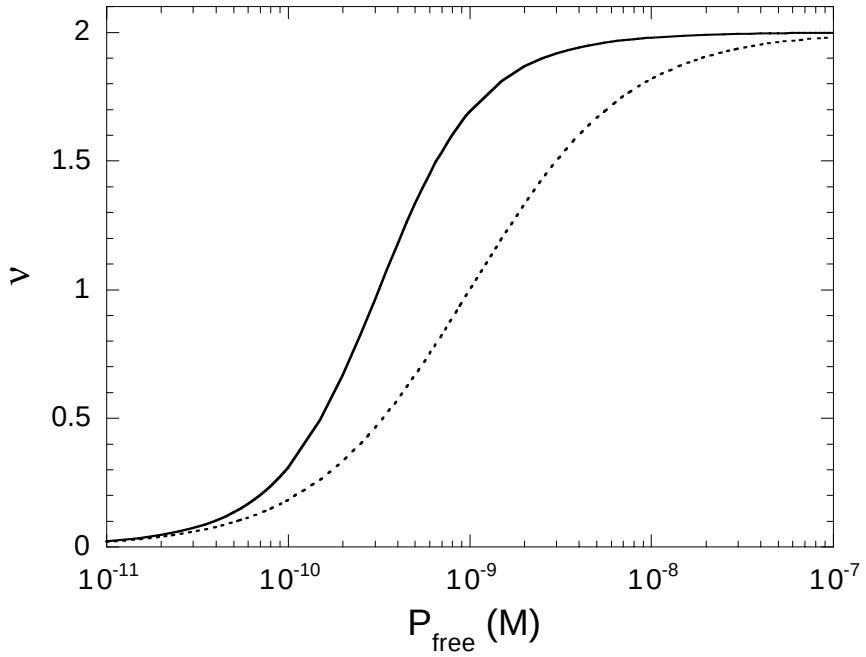

Figure 5: Logarithmic plot of binding curves for a two-site reaction. The two binding curves shown correspond to those described in the legend to Fig. 4.

It is evident from Fig. 4 that the cooperativity of binding leads to a faster saturation of the two sites. The same two curves are displayed in Fig. 5 with a logarithmic scale of the protein concentration of free protein. This allows the display over a larger range of protein

concentrations and makes also sense from a thermodynamic point of view, as the logarithm of the ligand activity is proportional to the chemical potential of the ligand.

In order to determine whether cooperative interactions or heterogeneity exist, it is often helpful to plot the data in a different way. Obviously the data of the binding curves are independent from the way they are plotted, but it turns out that some graphical methods show the occurrence of cooperative binding or heterogeneity more clearly than others. However, it should be noted that the binding parameters are best determined from a fit to the original data as given in Fig. 4, because the linearization described below leads to large changes of the standard error across the curve, especially for the very high and very low protein concentrations. The two most popular methods for the visualization of complex binding interactions are the so called Scatchard plot (Scatchard, 1949) and the Hill plot (Hill, 1910) [see also (Dahlquist, 1978; Wyman & Gill, 1990)]. For the same two examples that have been already shown in Fig. 4 and Fig. 5 the data are plotted also as a Scatchard plot (Fig. 6) or as a Hill plot (Fig. 7).

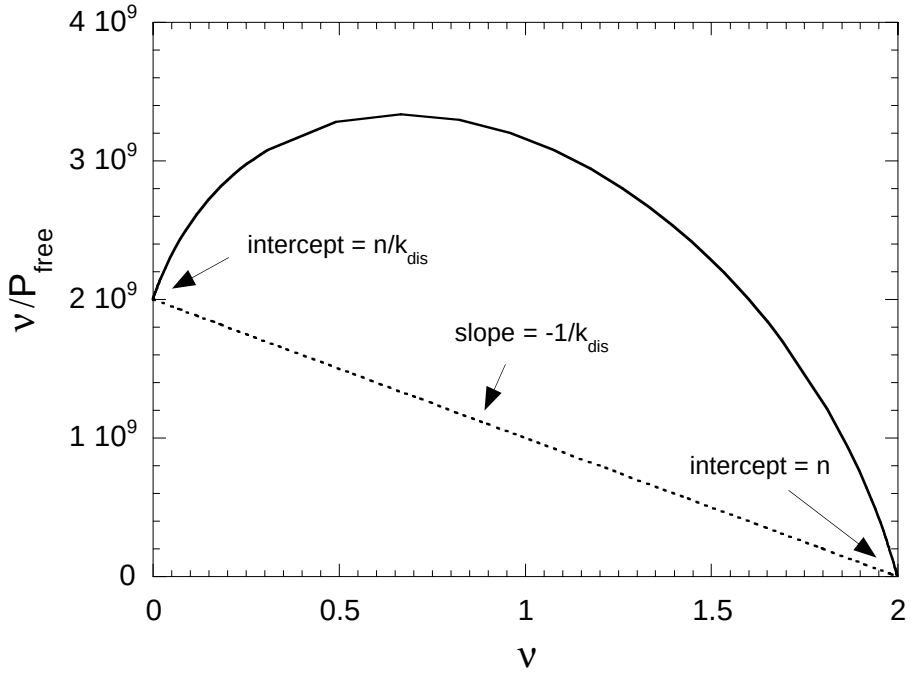

Figure 6: Scatchard plot of ν/P_{free} versus ν for binding of a protein to a DNA with two identical binding sites. The two curves correspond to independent binding (·····) or cooperative binding (———) to the two sites with the binding constants given in the legend to Fig. 4. The value of the maximum is at $\nu = 0.67$ leading to a Hill coefficient of $\alpha_H = 1.5$ as calculated according to Eq. 19.

For the Scatchard plot, Eq. 11 can be rearranged to Eq. 18:

$$\frac{\nu_n}{P_{free}} = \frac{n}{k_{dis}} - \frac{\nu_n}{k_{dis}} \qquad [18]$$

According to this expression a plot of ν/P_{free} versus ν results in a linear line if all binding sites are equivalent and independent. The slope of the line is $-1/k_{dis}$, the intercept with the y-axis gives n/k_{dis} and the intercept with the x-axis yields n. Cooperative binding is characterized by a Scatchard plot that is curved and shows a maximum. The position of the maximum is related to the observed degree of cooperativity. With increasing cooperativity it will shift along the x-axis to a higher saturation value. As discussed in (Dahlquist, 1978), the position of the maximum is related to the Hill coefficient α_H (see also Eq. 14) by

$$\alpha_H = \frac{n}{n - \nu_{max}} \qquad [19]$$

where ν_{max} is the value of ν at which the maximum is observed. For infinite cooperativity the maximum of the Scatchard plot would be located at $\nu_{max} = n - 1$. Although cooperative binding can be analyzed as shown by using a Scatchard plot, the preferred method to analyze this type of binding is the Hill plot (Hill, 1910). By rearranging Eq. 15 we can see that the ratio of occupied sites ν to unoccupied sites 2-ν is given by:

$$\frac{\nu_2}{2-\nu_2} = \frac{\theta}{1-\theta} = \frac{K_2 \cdot P_{free} + 2 \cdot P_{free}^2}{2 \cdot K_1 \cdot K_2 + K_2 \cdot P_{free}} \qquad [20]$$

If we now plot the log $[\nu/(2-\nu)]$ versus the log (P_{free}) we obtain the Hill plot from which several parameters of the cooperative binding interaction can be determined and which is shown in Fig. 7 for our example. It is most instructive to separate the above Hill plot for the cooperative binding (solid line) into three regions to understand how the type of binding is reflected in the shape of the curve. The graph approaches a slope of unity at the extremes where cooperativity becomes neglectable. If there are hardly any ligands bound, the graph will approach the limit of binding the first ligand with a dissociation constant K_1. On the other hand, if there are only very few sites unoccupied the binding will approach that of a single site reaction which is characterized by the dissociation constant K_2. It is only in the region where the binding sites are about half saturated, i. e. log $[\nu/(2-\nu)] = 0$ in which the cooperative binding becomes important which is reflected by the slope of the curve α_H being larger than one.

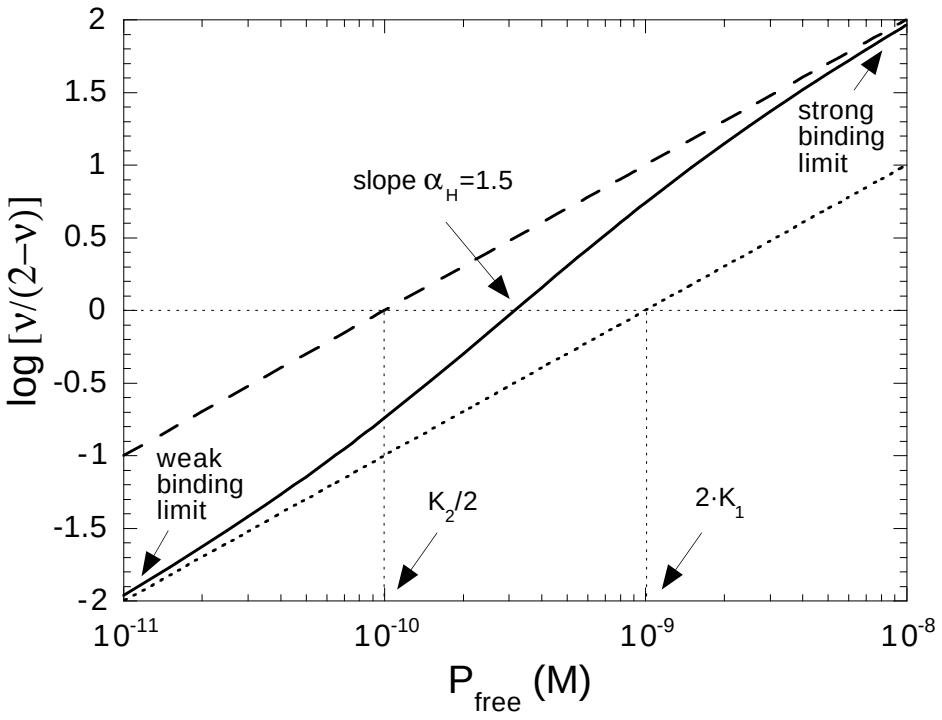

Figure 7: Hill plot of log [ν/(2-ν)] versus the log (P_{free}) for binding of a protein to a DNA with two identical binding sites. The two curves correspond to independent binding (·····) or cooperative binding (———) to the two sites with the binding constants given in the legend to Fig. 4. The Hill coefficient is determined for the cooperative binding from the slope at half saturation yielding $\alpha_H = 1.5$. The line for the strong binding limit (- - - -) corresponds to that of single site binding with a dissociation constant of 10^{-10}.

In particular we can derive the following information from the Hill plot: From examining the situation that P_{free} approaches 0 in Eq. 20 it can be seen that for the binding approaching that of the weak binding site in the Hill plot we obtain

$$P_{\text{free}} \to 0 \;\Rightarrow\; \log\left(\frac{\nu_2}{2-\nu_2}\right) = \log\left(P_{\text{free}}\right) - \log\left(2K_1\right) \qquad [21]$$

whereas for the limit of the strong binding we get:

$$P_{\text{free}} \to \infty \;\Rightarrow\; \log\left(\frac{\nu_2}{2-\nu_2}\right) = \log\left(P_{\text{free}}\right) - \log\left(\frac{K_2}{2}\right) \qquad [22]$$

In the intermediate region where the cooperative binding is observed the plot can be described according to Eq. 14 yielding a line with the slope of α_H.

$$P_{free} \approx K \Rightarrow \log\left(\frac{\nu_2}{2-\nu_2}\right) = \alpha_H \cdot \log(P_{free}) - \alpha_H \cdot \log(K) \quad [23]$$

In the case of the independent binding to the two sites (dotted line in Fig. 7) we obtain a linear plot with a slope of $\alpha_H = 1$ which also constitutes the lower limit for the curve with cooperative binding.

Concluding remarks

In the present article I have reviewed the basic principles for a quantitative description of equilibrium protein-DNA interaction, trying to provide the most important concept as a starting point for the analysis. Kinetic aspects of the binding process have not been included here, but it should be noted that these are equally important for a detailed understanding of a given protein-DNA binding process.

Almost certainly the evaluation of a given experiment will require some modifications of the equations presented here, due to the specifics of the technique used and the particular system studied. Nevertheless by proceeding along the lines discussed above it should be possible in most cases to derive an appropriate expression relating the observed experimental quantities to the binding parameters.

Acknowledgments

I am grateful to Peter H. von Hippel, Jörg Langowski, Franz W. Sevenich, and Alexandra Schulz for critical reading of the manuscript. In addition I would like to thank the Boehringer Ingelheim Fonds for a travel grant allowing me to attend the meeting "DNA-Recognition by Proteins and other Ligands", Aug. 16-21, 1993 in Copenhagen, Denmark which initiated this work.

References

Ackers, G. K., Johnson, A. D. & Shea, M. A. (1982). Quantitative model for gene regulation by λ phage repressor. *Proceedings of the National Academy of Sciences of the USA* **79**, 1129-1133.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry. Part III: The behavior of biological macromolecules*, 3. 3 vols, Freeman, New York.

Dahlquist, F. W. (1978). The meaning of Scatchard and Hill plots. *Methods in Enzymology* **48**, 270-299.

Fried, M. G. (1989). Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* **10**, 366-376.

Hill, A. V. (1910). The possible effects of the aggreagation of the molecules hemoglobin on its dissociation curve. *J. Physiol. (London)* **40**, iv-vii.

Perez-Howard, G. M., Weil, P. A. & Beechem, J. M. (1995). Yeast TATA binding protein interaction with DNA: fluorescence determination of oligomeric state, equilibrium binding, on-rate, and dissociation kinetics. *Biochemistry* **34**, 8005-8017.

Revzin, A. (1989). Gel electrophoresis assays for DNA-protein interactions. *Biotechniques* **7**, 346-355.

Scatchard, G. (1949). The attraction of proteins for small molecules and ions. *Ann. N. Y. Acad. Sci.*, 660-672.

Takahashi, M., Blazy, B., Baudras, A. & Hillen, W. (1989). Ligand-modulated binding of a gene regulatory protein to DNA. Quantitative analysis of cyclic-AMP induced binding of CRP from Escherichia coli to non-specific and specific DNA targets. *Journal of Molecular Biology* **207**, 783-796.

von Hippel, P. H., Rees, W. A., Rippe, K. & Wilson, K. S. (1996). Specificity mechanisms in the control of transcription. *Biophysical Chemistry* **59**, 231-246.

Weber, G. (1992). *Protein interactions*, Chapman and Hall, New York.

Wong, I. & Lohman, T. M. (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proceedings of the National Academy of Sciences of the USA* **90**, 5428-5432.

Wyman, J. & Gill, S. J. (1990). *Binding and linkage. Functional chemistry of biological macromolecules*, University Science Books, Mill Valley, CA.