Published in final edited form as:

Methods Enzymol. 2015 ; 562: 305–330. doi:10.1016/bs.mie.2015.04.009.

Analytical Ultracentrifugation as a Tool to Study Nonspecific Protein–DNA Interactions

Teng-Chieh Yang^{*}, Carlos Enrique Catalano[†], and Nasib Karl Maluf^{‡,1}

^{*}Pfizer, Pharmaceutical Research and Development, Chesterfield, Missouri, USA

[†]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado, Anschutz Medical Campus, Aurora, Colorado, USA

[‡]Alliance Protein Laboratories, San Diego, California, USA

Abstract

Analytical ultracentrifugation (AUC) is a powerful tool that can provide thermodynamic information on associating systems. Here, we discuss how to use the two fundamental AUC applications, sedimentation velocity (SV), and sedimentation equilibrium (SE), to study nonspecific protein–nucleic acid interactions, with a special emphasis on how to analyze the experimental data to extract thermodynamic information. We discuss three specific applications of this approach: (i) determination of nonspecific binding stoichiometry of *E. coli* integration host factor protein to dsDNA, (ii) characterization of nonspecific binding properties of Adenoviral IVa2 protein to dsDNA using SE-AUC, and (iii) analysis of the competition between specific and nonspecific DNA-binding interactions observed for *E. coli* integration host factor protein assembly on dsDNA. These approaches provide powerful tools that allow thermodynamic interrogation and thus a mechanistic understanding of how proteins bind nucleic acids by both specific and nonspecific interactions.

1. INTRODUCTION

Fundamental biological processes such as DNA replication, repair, recombination, transcription, and translation are regulated via protein–nucleic acid interactions. At any time during the cell cycle, a nucleic acid will be covered with a variety of proteins that compete for binding sites, and this competition dictates how the cell responds to its environment. These proteins bind the nucleic acid both specifically—where a protein recognizes a certain sequence of bases with much higher affinity than it does with other sequences, and nonspecifically—where a protein binds indiscriminately to any sequence. Any attempt to model this competition—and therefore develop a predictive understanding of how a cell responds appropriately to its environment—requires a mechanistic understanding of how proteins bind nucleic acids by both specific and nonspecific interactions.

Proteins that bind specifically to a site on a nucleic acid (a unique base sequence) must be able to select this sequence from what is essentially a vast excess of random base sequences

¹Corresponding author: karl.maluf@ap-lab.com.

(i.e., nonspecific sites). Thus, while the specific protein–nucleic acid interaction is important for controlling cellular functions, the nonspecific interaction is equally important in consideration of the biological outcome. So how does the protein molecule find a specific site buried within the vast nonspecific background, especially when this site is obscured with nonspecifically bound proteins? Do nonspecific sequences serve as a reservoir for sequestering specific binding proteins close to their sites to minimize their search space and thus optimize specific binding rates? How does one evaluate a protein whose biological role requires both specific and nonspecific DNA-binding interactions? To answer these fundamental questions, we need rigorous experimental approaches to measure these interactions and sophisticated computational tools to appropriately analyze the data within the context of meaningful molecular models.

Analytical ultracentrifugation (AUC) has proved to be a reliable approach for studying macromolecular interactions. With modern computational power and the newly developed fluorescence detection system, AUC allows one to conduct experiments over a wide concentration range and to directly apply first-order physical principles to analyze the data. More importantly, different AUC experimental approaches—sedimentation velocity (SV-AUC) and sedimentation equilibrium (SE-AUC)—provide rigorous hydrodynamic (molecular shape distributions) and thermodynamic (strengths of interactions at equilibrium) information (Demeler & van Holde, 2004; Laue & Stafford, 1999; Lebowitz, Lewis, & Schuck, 2002; Philo, 2011; Ucci & Cole, 2004; Yphantis, 1964). Here, we discuss how to study nonspecific protein–DNA interactions using AUC approaches and emphasize in particular how to model AUC data via nonlinear least squares (NLLS) approaches using models that account for nonspecific nucleic acid binding. We have organized this review as follows: First we discuss the elegant theory, originally developed by McGhee and von Hippel (McGhee & von Hippel, 1974), which describes the binding of a ligand to a one-dimensional lattice and application of the model to interrogate nonspecific protein–nucleic acid interactions. Next we discuss a more complex scenario, where specific and nonspecific sites compete for protein binding. Then we review four different case studies originally developed by Record and coworkers (Tsodikov, Holbrook, Shkel, & Record, 2001) that deal with the competition between specific and nonspecific binding and discuss how we have applied these models to the analysis of AUC data collected in our laboratories.

2. NONSPECIFIC INTERACTION MODEL

2.1 Nonspecific Binding of Proteins to Linear Nucleic Acids (Non-Cooperative Versus Cooperative)

A DNA-binding protein that interacts with a nonspecific DNA sequence is characterized by three thermodynamic parameters: (1) the intrinsic binding affinity, K , (2) the binding (occluded) site size, n , and (3) the cooperativity parameter, ω . In this case, the intrinsic binding affinity quantifies the strength with which a protein binds to a nonspecific site on the nucleic acid lattice and the binding site size represents the space that single protein occupies, or occludes, on the lattice. For a nucleic acid molecule that can accommodate more than one protein, the binding of the first protein may influence binding of the second protein adjacent to it, and this tendency is quantified by the cooperativity parameter. Positive cooperativity

refers to the case where binding of the first protein increases the probability of binding of the second, while negative cooperativity implies that the first binding event decreases the *probability* of the second event.

A crucial aspect of the nonspecific interaction is that, if a given space on the nucleic acid is occupied by one protein, it is assumed to be physically impossible for another protein to interact with any of the occluded site. Moreover, if two proteins are bound close enough to each other such that the number of bases between them is less than the occluded site size, then this segment of the nucleic acid is unavailable for binding by additional proteins (the so-called overlapping effect). This creates a conundrum for developing a model to quantify the number of proteins bound to the nucleic acid since the concentration of available nucleic acid binding sites depends not only on the number of proteins bound to the lattice, but also on the configuration with which the proteins are bound—the same number of proteins can be bound to the nucleic acid in multiple configurations, and these configurations in general have differing numbers of remaining sites for the next protein-binding event (Fig. 1).

To resolve this complication, McGhee and von Hippel (MvH) used an approach based on conditional probabilities to account for all the possible configurations for protein bound nonspecifically to an infinite, one-dimensional lattice, all in terms of the intrinsic binding affinity and the occluded binding site size. They also derived expressions that incorporate “nearest-neighbor” cooperativity. This concept assumes that the only possibility for cooperative interactions occurs when two ligands are bound immediately next to each other; if any two proteins have at least one free (unbound) nucleic acid base between them, there can be no cooperativity.

The MvH model was later extended by Epstein, who developed exact expressions that accounted for a protein binding to a nucleic acid of finite length using a combinatorial approach (Epstein, 1978). While this approach takes into account all possible nonspecifically bound protein states, it does not produce a simple closed-form expression convenient for NLLS modeling of binding data. Since many studies have been carried out using rather short nucleic acids, and since the MvH formalism provides a straightforward method for NLLS data analysis, Record and coworkers modified the MvH model, using a weighted conditional probability approach, to provide approximate (although highly accurate), closed-form expressions suitable for analysis of data collected on short nucleic acid lattices of finite length (Tsodikov et al., 2001). In addition to the MvH and Epstein models, which consider only one mode of nonspecific interaction, more complex binding models including multiple nonspecific binding modes have been considered (Bujalowski, 2006; Bujalowski, Lohman, & Anderson, 1989; Bujalowski, Overman, & Lohman, 1988). Several groups have adapted these models to analyze data from electrophoretic mobility shift (EMS) (Aeling et al., 2006; Sanyal, Yang, & Catalano, 2014), calorimetry (Holbrook, Tsodikov, Saecker, & Record, 2001) and filter binding assays (Bujalowski, 2006; Lohman & Bujalowski, 1991) with success. In this review, we will discuss how to analyze AUC data to interrogate nonspecific binding of proteins to nucleic acid lattices of finite length, by adapting the modified MvH model presented by Record and coworkers.

2.2 MvH Model

In the following discussion, we will assume that the nucleic acid lattice is dsDNA. The MvH model assumes a dynamic equilibrium exists between the free protein and the free protein-binding site on a one-dimensional duplex, described by the following reaction scheme:

where K_{ns} is the intrinsic, nonspecific association binding constant. Applying mass conservation to this reaction, we obtain:

$$\frac{\overline{P}_b}{P_f} = K_{ns} \cdot \overline{D}_f \quad (1)$$

where $\overline{P}_b$ is the average number of proteins bound per DNA, P_f is the concentration of free protein, and $\overline{D}_f$ is the average number of free protein-binding sites per DNA. This is essentially a Scatchard formulation that takes into account the nonspecific binding of multiple copies of a ligand (e.g., protein) to a one-dimensional lattice of finite length (e.g., DNA).

Consider a DNA molecule that is N -bp long and where each bound protein occupies (occludes) n bp. According to the MvH formalism, the general equation for the nonspecific binding isotherm is given by:

$$\frac{v}{K_{ns} \cdot P_f} = (1 - nv) (ff)^{n-1} \left(\frac{N - n + 1}{N} \right) \quad (2)$$

where v represents the protein binding density, i.e., moles of protein bound per mole of *base pair* of DNA. Logically, $\overline{P}_b$, from Eq. (1), is a function of v , N , and the concentration of DNA molecules ($[DNA_t]$), and can be expressed as:

$$\overline{P}_b = vN [DNA_t] \quad (3)$$

According to mass conservation, $P_f = P_t - \overline{P}_b = P_t - vN [DNA_t]$, where P_t is the total protein concentration. The term $(ff)^{n-1}$ in Eq. (2) is the conditional probability of randomly selecting $(n - 1)$ consecutive bp, that are adjacent to any given free bp, but that are not bound by protein. Since the probability of finding one free bp is $(1 - nv)$, the product $(1 - nv)(ff)^{n-1}$ represents the probability of finding n consecutive, unoccupied base pairs (i.e., a stretch of n free base pairs). The term $\left(\frac{N - n + 1}{N} \right)$ is an approximation that has been shown by Record and coworkers to adequately account for the finite length of the DNA lattice (i.e., end effects) (Tsodikov et al., 2001).

The term $(ff)^{n-1}$ depends upon different binding scenarios. In the case of non-cooperative binding, $(ff)^{n-1}$ is given by $\left(\frac{1-nv}{1-(n-1)v}\right)^{n-1}$. For proteins that display nearest-neighbor cooperativity (either favorable or unfavorable), we have to consider different scenarios for how different “protein clusters” are found on the DNA. A cooperativity parameter, ω , is introduced to describe this likelihood of having one free protein bound immediately adjacent to another protein. Thus, $(ff)^{n-1}$ must depend on ω and is expressed as

$$(ff)^{n-1} = \left(\frac{(2\omega - 1)(1 - nv) + v - R}{2(\omega - 1)(1 - nv)} \right)^{n-1} \cdot \left(\frac{1 - (n+1)v + R}{2(1 - nv)} \right)^2 \quad (4a)$$

where $R = \sqrt{[1 - (n+1)v]^2 + 4\omega v(1 - nv)}$. Note that the above equation as written is undefined for the non-cooperative case, i.e., when $\omega = 1$. In this case, Bujalowski et al. derived a single, closed-form expression of the equation,

$$(ff)^{n-1} = \left(\frac{2\omega(1 - nv)}{(2\omega - 1)(1 - nv) + v + R} \right)^{n-1} \cdot \left(\frac{1 - (n+1)v + R}{2(1 - nv)} \right)^2 \quad (4b)$$

which is applicable for either non-cooperative or cooperative binding. This expression is easier to incorporate into NLLS fitting algorithms since a single equation can be used in the fitting algorithm for analysis of the experimental data over the range $0 < \omega < \infty$ (Bujalowski et al., 1989). It is important to keep in mind that as written, the binding density v is an implicit variable (i.e., it is not possible to solve Eq. 2 for v). For any NLLS data analysis algorithm, v must be determined at each value of P_f and this is accomplished by an iterative procedure that examines successive values of v within the range $0 < v < \frac{1}{n}$, where $\frac{1}{n}$ represents the maximum value of v (when the DNA is completely saturated with ligand). This is done until the right- and left-hand sides of Eq. (2) are equal to each other within a defined tolerance. In all the cases we consider here, we have used the built-in implicit solver available in the commercially available Scientist software (MicroMath, St. Louis MO).

2.3 Specific and Nonspecific Competitive Binding of Proteins to Linear Nucleic Acids

In general, DNA-binding proteins can interact with DNA both specifically and nonspecifically. In some cases, specific binding is so much stronger than nonspecific binding that the latter can be ignored during data analysis. However, in many cases nonspecific interactions cannot be ignored. This can occur in cases where discrimination between specific and nonspecific binding sites is modest (i.e., $K_{NS} < K_{SP} < \sim 10 \times K_{NS}$). In other cases, the concentration of nonspecific sites is large enough such that nonspecific binding can compete significantly with specific binding. This can occur experimentally (i.e., *in vitro*) if the duplex used in the experiment is long relative to the site-specific site size; indeed, this is commonly encountered *in vivo*, where the number of specific binding sites are always vastly outnumbered by the total number of nonspecific sites in the genome. For example, in the case of the *Escherichia coli lac* operon, it has been estimated that there are 10–30 lac repressor proteins that must find one to four copies of the specific lac operon out of a total of

~10⁷ nonspecific DNA-binding sites in the *E. coli* genome (von Hippel & Berg, 1986). In either case, competitive nonspecific binding must be accounted for to properly analyze the binding isotherms. Record and coworkers have developed specific models for this purpose, and we discuss two particularly useful cases below (Tsodikov et al., 2001).

In the first case (case 1), they consider the binding of a protein to a relatively short DNA molecule that contains a single specific DNA-binding sequence. For this model, it is important to keep in mind that it is common for specific DNA-binding proteins to occlude more DNA upon binding to their specific site than they do when they bind nonspecifically, i.e., it is usually the case that the specific binding site size is larger than the nonspecific site size. Thus, in this model they consider the case where a DNA molecule is bound to its protein either entirely specifically or entirely nonspecifically—there are no DNA molecules that have both specific and nonspecifically bound proteins since there is not enough space on the short duplex to simultaneously accommodate both protein binding modes. Thus, if a protein is bound to the specific DNA site, the binding stoichiometry is 1:1. However, if the protein is bound nonspecifically, then it is possible to have multiple proteins bound, up to the saturation limit (Fig. 1C).

The second case (case 2) stipulates that the DNA contains a single specific site with flanking regions that are much longer than the length of the specific site so that proteins can simultaneously interact with both the specific and nonspecific DNA-binding sites. Note that if a protein is bound nonspecifically to a site that overlaps with the specific site, it is assumed that this physically prevents another protein from binding to the site. Thus, the entire specific site must be available for a protein to bind in a specific binding mode.

In these two cases, we must consider three states of the protein ligand: protein that is free in solution (P_f), protein that is bound to DNA in a nonspecific manner ($P_{b,ns}$), and protein that is bound to a specific site in the duplex ($P_{b,sp}$). According to Eq. (2), P_f can be expressed as:

$$P_f = \frac{Nv}{K_{ns}(N - n + 1)(ff)^{n-1}(1 - nv)} \quad (5)$$

For case 1, $P_{b,ns}$ can be expressed as (Tsodikov et al., 2001):

$$P_{b,ns} = vN [DNA_t] \left(\frac{K_{ns}(N - n + 1)}{K_{ns}(N - n + 1) + NK_{sp}v(ff)^{N-n}} \right) \quad (6)$$

where K_{sp} is the association constant for the specific binding of the protein to the DNA and the expression in parenthesis accounts for the competition of protein binding to nonspecific sites by the specific site. Similarly, the expression for $P_{b,sp}$ is given by:

$$P_{b,sp} = vN [DNA_t] \left(\frac{K_{sp}(ff)^{N-n}}{K_{ns}(N - n + 1) + NK_{sp}v(ff)^{N-n}} \right) \quad (7)$$

The total protein concentration is then obtained by combining Eqs. (5)–(7), as shown in Eq. (8):

$$P_t = \frac{Nv}{K_{ns}(N-n+1)(ff)^{n-1}(1-nv)} + vN[DNA_t] \left(\frac{K_{ns}(N-n+1) + K_{sp}(ff)^{N-n}}{K_{ns}(N-n+1) + N K_{sp}v(ff)^{N-n}} \right) \quad (8)$$

For case 2, since there are many more nonspecific binding sites than the single, specific site, the expression for $P_{b,ns}$ is simply Eq. (3). However, $P_{b,sp}$ is more complicated since nonspecifically bound protein can overlap with the specific site, preventing specific binding. Therefore, $P_{b,sp}$ also depends on the location of the specific binding sequence within the DNA duplex. With these considerations, $P_{b,sp}$ can be expressed as:

$$P_{b,sp} = \frac{K_{sp}\lambda P_f [DNA_t]}{1 + K_{sp}\lambda P_f}, \quad \text{with} \quad \lambda = \frac{1}{N} (1 - nv) \frac{2 - (ff)^{N_{f1}} - (ff)^{N_{f2}}}{1 - (ff)} \quad (9)$$

Where N_{f1} and N_{f2} are the known lengths of the flanking regions (in bp) on either side of the specific DNA site. Combining Eqs. (3), (5), and (8) affords an expression for the total protein concentration as shown in Eq. (10):

$$P_t \cong \frac{Nv}{K_{ns}(N-n+1)(ff)^{n-1}(1-nv)} + [DNA_t] \left(\frac{Nv + (Nv+1)\lambda K_{sp}P_f}{1 + \lambda K_{sp}P_f} \right) \quad (10)$$

In many experimental situations, what is actually measured is not the moles of ligand bound per bp, but rather the fraction of DNA that is bound (e.g., EMS data). For these cases, it is necessary to obtain expressions for the fraction of DNA that is free and the fraction of DNA that has at least one bound protein (i.e., the fraction of DNA that is bound). These expressions are given below in Eqs. (11) and (12), originally derived in the following references (Tsodikov et al., 2001; Yang & Maluf, 2014):

$$F_f = (1 - nv)(ff)^{N-1} \quad (11)$$

$$F_b = 1 - (1 - nv)(ff)^{N-1} \quad (12)$$

where F_f is the fraction of free DNA and F_b is the fraction of DNA that has at least one protein bound. Note that Eq. (11) represents the probability of finding a stretch of $(N-1)$ free bp next to a free DNA bp; i.e., it represents the probability of finding a DNA with N free bp, i.e., a free DNA molecule.

For the data-fitting procedures, one must solve for two implicit variables: P_f and v . We obtain P_f over the range $0 < P_f < P_b$ and v over the range $0 < v < \frac{1}{n}$. The independent variables are $[P_t]$ and $[DNA_t]$. For case 1, we use Eqs. (2), (4a), (4b), and (8) to calculate v and P_f at any $[P_t]$ and $[DNA_t]$. For case 2, we use Eqs. (2), (4a), (4b), and (10). Then, depending on the type of data to be analyzed, we use Eqs. (11) and (12) to calculate the fraction of free or bound DNA. This procedure allows us to determine the best-fit parameters K_{ns} , K_{sp} , ω , and n .

3. CASE STUDIES

3.1 Determination of Binding Stoichiometry Using SE-AUC

A key measurement in the study of any protein–DNA binding reaction is the binding stoichiometry, which is defined as the number of protein ligands bound per DNA molecule at saturation. SE-AUC is an ideal technique for this measurement since it reports directly the average molar mass of the protein–DNA complex. In our laboratories, we have successfully carried out these experiments using DNA molecules that have been conjugated with a Cy3 chromophore. This allows us to selectively monitor the absorbance of the chromophore at 522 nm, a wavelength where protein molecules do not typically absorb light. At this wavelength, we detect all the DNA species—the free DNA along with the ensemble of protein-bound DNA species—without pollution of the signal with any free protein species. This experimental design simplifies the data analysis.

E. coli integration host factor (IHF) is a basic, heterodimeric DNA-binding protein that belongs to a class of histone-like proteins capable of bending and wrapping DNA. As such, it plays an important role in condensation of the bacterial nucleoid, binding nonspecifically to duplex DNA. In addition, it serves as a regulatory protein associated with a number of cellular processes, including transcription, DNA replication, and site-specific recombination. In this role, IHF binds to a specific sequence with high affinity to introduce a strong bend ($>140^\circ$) into the duplex. This provides a duplex architecture conducive to the cooperative assembly of additional regulatory proteins at the site. Thus, IHF must interact with DNA in both specific and nonspecific binding modes to exert distinct biological functions.

In Fig. 2, we show an application of the approach described above to determine the binding stoichiometry of IHF to a 10 bp nonspecific DNA duplex using sedimentation equilibrium approaches. Each experiment (i.e., data point) was carried out at three rotor speeds (22, 27, and 33 K RPM) and then globally fit to an ideal single-species model,

$$\begin{aligned} A(r) &= A_{ref} \cdot \exp\left(\sigma \cdot \left(\frac{r^2 - r_{ref}^2}{2}\right)\right) + b, \\ \sigma &= \frac{M_{b,app}}{RT} \omega^2 \end{aligned} \quad (13)$$

where $A(r)$ and A_{ref} are the total absorbance of all DNA species (free and protein bound) at radial position, r , and reference radial position, r_{ref} , respectively; b is the baseline offset, R is the gas constant, T is the absolute temperature, ω is the angular speed of the rotor, and $M_{b,app}$ is the apparent buoyant molecular weight, given by:

$$M_{b,app} = M_{app} (1 - \bar{v}_{app} \rho) \quad (14)$$

where M_{app} is the apparent weight-average molecular weight of all species (Yphantis, 1964) and $\bar{v}_{app}$ is the apparent weight-average partial specific volume of all species, and ρ is the buffer density. The fitted $M_{b,app}$ is then used to calculate the approximate average number of IHF, L , bound per DNA duplex (i.e., the degree of binding) according to Eq. (15):

$$M_{b,app} = M_{b,DNA} + L \cdot M_{b,IHF} \quad (15)$$

Note that this approach assumes there is no significant volume change associated with formation of the protein–DNA complexes. $M_{b,DNA}$ and $M_{b,IHF}$ can either be calculated from their respective sequences or can be measured experimentally. In this case, we calculate $M_{b,IHF} = 5.6$ kDa using the program SEDNTERP (Laue, Shah, Ridgeway, & Pelletier, 1992), and utilized SE-AUC to measure $M_{b,DNA} = 2.7(\pm 0.1)$ kDa.

As shown in Fig. 2, a plot of the degree of binding as a function of [IHF]/[DNA] increases linearly and then reaches a plateau at about 1.9. This feature indicates that the experiment is carried out under stoichiometric binding conditions (i.e., the total DNA concentration is at least 10 times greater than the apparent dissociation constant) and that this 10 bp duplex DNA can accommodate two IHF proteins at saturation. This suggests that each IHF occupies about 5 bp of the DNA ($n=5$), which is in agreement with previous ITC studies (carried out under similar buffer conditions) (Holbrook et al., 2001).

A common problem with the interpretation of a stoichiometric binding curve is that the linear-plateau breakpoint is dependent upon not only the stoichiometry of the interaction but also the *activity* of the purified protein. For instance, a breakpoint of two can result from (i) a binding stoichiometry of 2:1 if the protein is 100% active or (ii) a stoichiometry of 1:1 if the protein is only 50% active. A significant advantage when using SE-AUC to measure protein–DNA binding stoichiometries is that the plateau provides a *direct* measure of the buoyant mass of the complex. This allows the researcher to make a direct comparison between the binding stoichiometry estimated from the plateau value with that implied by the breakpoint. If 100% of the protein molecules are active for DNA binding, then these two measurements should provide identical values for the binding stoichiometry. If, however, the breakpoint yields an estimate of the binding stoichiometry that is larger than that measured from the plateau, this most likely indicates that the protein preparation is not 100% active. In Fig. 2, comparison of the break point with the plateau indicates that this preparation of IHF is 83% active for nonspecific DNA binding.

3.2 Evaluation of Protein Nonspecific Binding Parameters Using SE-AUC

In the above model-independent analysis, we have essentially ignored the rich information content of the SE-AUC data to provide a simple method to estimate binding stoichiometries. It is also possible to analyze the primary data directly from first principles to obtain information on the nonspecific binding parameters. This requires the construction of a fitting

model to directly analyze the concentration distributions. Here, we discuss how we analyzed SE-AUC data collected for the nonspecific binding of the Adenoviral IVa2 protein to a Cy3-labeled, duplex DNA, and compare these results with orthogonal data collected using the filter binding technique. Other examples are available that rigorously analyze nonspecific protein–DNA interactions using SE-AUC using the Epstein model (Ucci & Cole, 2004). In the approach presented here, using the MvH model, we can treat n as a continuous fitting parameter, which significantly simplifies the data analysis procedure.

To directly analyze the SE concentration distributions using the MvH nonspecific binding model, we substituted the expression for an ideal sedimenting free protein into Eq. (2), which yields Eq. (16):

$$v(r) = K_{ns} \cdot \left[P_{f,m} \cdot \exp \left(\frac{M_p (1 - \bar{v}_p \rho) \cdot \omega^2}{RT} \cdot \frac{r^2 - r_m^2}{2} \right) \right] \cdot (1 - nv(r)) \cdot (ff(r))^{(n-1)} \cdot \left(\frac{N - n + 1}{N} \right) \quad (16)$$

where $P_{f,m}$ denotes the concentration of free protein at meniscus position, r_m , M_p is the molar mass of free protein, and $\bar{v}_p$ is the partial specific volume of free protein. The binding density, $v(r)$ and $(ff(r))$, which are functions of the radial position, is calculated implicitly at each radial position, over the range $0 < v(r) < \frac{1}{n}$.

Next, in order to express the total DNA concentration in terms of the radial position, we multiplied Eq. (11) by $[DNA_t]$, which provides the free DNA concentration, $[DNA_f]$. Solving for $[DNA_t]$ yields:

$$[DNA_t] = \frac{[DNA_f]}{(1 - nv(r)) (ff(r))^{N-1}} \quad (17)$$

Note that Eq. (17) is expressed in terms of the independent variables $[DNA_f]$ and $[P]$. To obtain an expression for the total DNA concentration $[DNA_t](r)$, as a function of radial position, r , we substituted Eq. (18), which is the expression for an ideal free DNA species at sedimentation equilibrium into Eq. (16), which yields Eq. (19):

$$[DNA_f](r) = [DNA_{f,ref}] \cdot \exp \left(\frac{M_D (1 - \bar{v}_D \rho) \omega^2}{RT} \cdot \frac{r^2 - r_{ref}^2}{2} \right) \quad (18)$$

$$[DNA_t](r) = \frac{[DNA_{f,ref}] \cdot \exp \left(\frac{M_D (1 - \bar{v}_D \rho) \omega^2}{RT} \cdot \frac{r^2 - r_{ref}^2}{2} \right)}{(1 - nv(r)) (ff(r))^{N-1}} \quad (19)$$

where $[DNA_{f,ref}]$ denotes the concentration of free DNA at the reference radial position (usually taken as the meniscus position), M_D is the molecular weight of the free DNA, and $\bar{v}_D$ is the partial specific volume of the free DNA. To directly analyze the absorbance data, we substituted Eq. (15) into Beer's law, $Abs(r) = [DNA_f](r) \cdot \epsilon_\lambda \cdot l$, to obtain Eq. (20):

$$Abs(r) = \frac{[DNA_{f,ref}] \cdot \exp\left(\frac{M_D(1-\bar{v}_D\rho)\bar{\omega}^2}{RT} \cdot \frac{r^2 - r_{ref}^2}{2}\right)}{(1 - nv(r))(ff(r))^{N-1}} \cdot \epsilon_\lambda \cdot l + b \quad (20)$$

where b is the baseline offset, l is the path length; and ϵ_λ is the extinction coefficient of the Cy3-labeled DNA at 522 nm. During the data-fitting procedure, Eqs. (16), (4a), and (4b) are used to implicitly solve for $v(r)$ at each radial position. Then each pair of $v(r)$ and r is substituted into Eq. (20) to compare with the experimental data using NLLS.

For this system, the total DNA concentration was high enough to ensure the binding reaction occurred under stoichiometric conditions. However, the distribution of bound states is affected by the parameters n and ω . In general, proteins that bind DNA with positive cooperativity ($\omega > 1$) tend to produce longer clusters of bound protein, and therefore more of the DNA sediments with a higher average molar mass than for systems that display lower cooperativity. On the other hand, larger values of n tend to disfavor formation of longer clusters, since it becomes more difficult to load the next protein onto the DNA. For this reason, even under stoichiometric conditions, the shapes of the isotherms will provide information of the relative values of ω and n . To test for consistency, we fitted the SE-AUC data using the values determined for K_{ns} , ω , and n from an independent filter binding study (Yang & Maluf, 2014), while allowing only the free protein and DNA concentrations at reference radial positions and baseline offsets to float. The excellent agreement between the model and the data indicate the values obtained for K_{ns} , ω , and n from filter binding are consistent with the SE-AUC data (Fig. 3), and thus, the two orthogonal approaches are in agreement.

3.3 Examination of Specific and Nonspecific Competitive Binding Using SV-AUC

Sedimentation equilibrium approaches were used in the above two examples to determine the stoichiometry, affinity, and cooperativity of proteins binding to nonspecific duplex DNA. While powerful, SE approaches require that the nucleoprotein complexes under investigation are stable during the time course of the experiment, which can be up to 10 days in length. In contrast, a sedimentation velocity experiment can be completed in a matter of hours. Here, we demonstrate that SV-AUC, by resolving different interacting species, can be used to dissect the competition between specific and nonspecific DNA-binding interactions. Again, *E. coli* IHF serves as an excellent example because this protein binds to DNA in distinct specific (bent) and nonspecific (condensed) complexes (*vide supra*, Section 3.1).

IHF binds to its specific recognition element with nanomolar affinity (Aeling et al., 2006; Holbrook et al., 2001; Ortega & Catalano, 2006; Sanyal et al., 2014). By labeling the DNA molecule with fluorescein, we can use the AU-FDS (Fluorescence Detection System) to monitor the sedimentation of the IHF-DNA complexes at sufficiently low DNA

concentrations (4 nM) such that equilibrium binding conditions are ensured. In this case, the data can be analyzed to afford equilibrium association constants. SV-AUC data were collected for a series of samples containing labeled DNA in the presence of the indicated concentrations of IHF. The raw data were analyzed by Sedfit (Schuck, 2003) to obtain $c(s)$ sedimentation coefficient distributions and the weight-average sedimentation coefficient, $\langle s \rangle$, for each mixture of IHF and DNA. This analysis yields a sedimentation coefficient for 1.96 S for both the specific and nonspecific 27 bp duplexes. When these DNA molecules are titrated with IHF, $\langle s \rangle$ increases, indicating the formation of IHF–DNA complexes (Fig. 4A and B). Each value of $\langle s \rangle$ reflects the ensemble distribution of IHF–DNA complexes at a given total IHF concentration.

A plot of $\langle s \rangle$ as a function of IHF concentration is presented in Fig. 4C; close inspection of the data reveals that while IHF binds to the nonspecific DNA in a monotonic manner, it appears that the binding curve for the specific DNA is biphasic. This strongly suggests that a high affinity (likely specific) binding interaction has been detected by SV-AUC. Ultimately, the specific and nonspecific binding curves reach similar plateau values. We interpret this to indicate that at high enough IHF concentration, nonspecific binding out-competes the specific binding interaction and dominates the sedimentation behavior of the complex. To address this feature, we can express $\langle s \rangle$ for either the specific (Eq. 21) or the nonspecific (Eq. 22) DNA as follows:

$$\langle s \rangle = s_{\text{free}} F_{\text{free}} + \langle s_{b,ns} \rangle F_{b,ns} + s_{b,sp} F_{b,sp} \quad (21)$$

$$\langle s \rangle = s_{\text{free}} F_{\text{free}} + \langle s_{b,ns} \rangle F_{b,ns} \quad (22)$$

where s_{free} is the sedimentation coefficient of free DNA, $\langle s_{b,ns} \rangle$ is the weight-average sedimentation coefficient the ensemble of nonspecific protein–DNA complexes, $s_{b,sp}$ is the sedimentation coefficient for the specific protein–DNA complex, F_{free} is the fraction of free DNA (given by Eq. 11), $F_{b,ns}$ is the fraction of protein nonspecifically bound DNA, and $F_{b,sp}$ is the fraction of protein specifically bound DNA, at a given protein concentration. Note $F_{b,ns}$ represents the DNA with at least one protein bound in a nonspecific manner. Thus, when fitting data collected for the DNA with a specific site, one uses Eq. (21), and when fitting data collected for a fully nonspecific DNA, one uses Eq. (22).

The specific DNA duplex in our experiment is short (27 bp) and only contains a single specific binding site, such that it can accommodate either one specifically bound IHF protein or multiple IHF nonspecific interactions. Hence, the most suitable model to analyze the specific binding isotherm is the specific/nonspecific competition model (case 1, Section 2.2), which requires that only one IHF can bind in a specific mode, while multiple IHF molecules can bind in a nonspecific mode, *but not both* (i.e., there is competition between specific and nonspecific binding). Therefore, in this case the concentration of specifically bound IHF is the same as the concentration of DNA that is specifically bound (since the *specific* binding

stoichiometry is 1:1, and is usually different from the *nonspecific* binding stoichiometry; see Section 3.1). Next, we can rearrange Eq. (7) to get an expression for $F_{b,sp}$:

$$F_{b,sp} = vN \left(\frac{K_{sp}(ff)^{N-n}}{K_{ns}(N-n+1) + NK_{sp}v(ff)^{N-n}} \right) \quad (23)$$

and $F_{b,ns}$ can be calculated according to:

$$F_{b,ns} = 1 - (F_{free} + F_{b,sp}) \quad (24)$$

where F_f is given by Eq. (11). To fit data collected for titration of the *specific* DNA, Eqs. (23) and (24) were substituted into Eq. (21) to yield an expression that describes $\langle s \rangle$ as a function of IHF concentration. To fit data collected for titration of the *nonspecific* DNA, Eq. (24) (with $F_{b,sp} = 0$) is substituted into Eq. (22). Importantly, it is not possible to extract both the specific and nonspecific parameters by carrying out experiments using only the specific DNA—there is insufficient information to constrain the NLLS analysis. However, when titrations are carried out using both specific and nonspecific DNA, as was done in this case, the ensemble of datasets can be analyzed globally (i.e., simultaneously); the nonspecific DNA titration provides information exclusively on the nonspecific DNA-binding parameters (Eq. 23) while the specific DNA titration provides information on both the specific- and nonspecific binding parameters (Eq. 21). The global fitting approach provides adequate constraints in the data analysis to accurately resolve all of the fitting parameters. In this fitting routine, the nonspecific binding site size (n), DNA length (N), and experimentally determined s for free DNA were fixed as global constants; $s_{b,sp}$ and $\langle s_{b,ns} \rangle$ were local variables that were allowed to float for the specific and nonspecific datasets, respectively, and K_{ns} , K_{sp} , and ω were global floating parameters for the combined datasets. The implicit variables P_f and v were calculated as described in Section 2.2 and the results of this analysis are shown as the solid curves in Fig. 4C. This analysis returns the local parameters: $s_{b,sp} = 2.59(\pm 0.12)S$ and $\langle s_{b,ns} \rangle = 3.37(\pm 0.04)S$, and the global parameters: $K_{ns} = 1.7(\pm 0.6) \times 10^6 M^{-1}$, $K_{sp} = 2.0(\pm 0.6) \times 10^8 M^{-1}$, and $\omega = 10(\pm 4)$. Moreover, according to our analysis, we calculate a binding specificity ($\frac{K_{sp}}{K_{ns}}$) for IHF of 117. Notably, these parameters agree well with those determined previously using ITC to interrogate IHF–DNA binding interactions (under slightly different solution conditions) (Holbrook et al., 2001).

3.4 Identification of a Conformational Change Upon Nonspecific Protein–DNA Binding Using SV-AUC

SV-AUC is a hydrodynamic technique and as such is not only useful to measure the mass change upon protein–DNA binding, but it is also sensitive to the overall shape of the complexes. In this way, it is able to detect conformational differences in the sedimenting species (Correia & Stafford, 2015; Stafford et al., 2001). IHF provides an exceptional model system with which to study protein-induced conformational changes in DNA molecules, as follows. IHF binds to its cognate element and introduces a severe bend into the duplex (Khrapunov, Brenowitz, Rice, & Catalano, 2006). In contrast, the protein weakly bends and

wraps the duplex when bound in a nonspecific mode. While these distinct binding modes are not evident using the minimal length duplexes described above, a different picture emerges when a longer DNA is employed.

In this study, two longer 274 bp substrates were examined; (i) a specific DNA (cos-274) that contains a single, centrally positioned 27 bp specific IHF recognition element and (ii) a 274 bp DNA of random sequence (Sanyal et al., 2014). EMS assays have demonstrated that IHF binds to the specific site in cos-274 with a low nanomolar affinity to afford a specific, strongly bent nucleoprotein complex (Ortega & Catalano, 2006; Sanyal et al., 2014). In contrast, the protein binds weakly to ns-274 to afford a smeared complex on the gel. Here, we demonstrate by AUC that addition of IHF to the specific cos-274 substrate results in a progressive, concentration-dependent increase in $\langle s \rangle$ from 4.2 S for free DNA to a maximal value of ~ 10 S at apparent saturation (Fig. 5A). A plot of $\langle s \rangle$ as a function of IHF concentration is presented in Fig. 5C. The concentration dependence and the magnitude of the shift suggest that in contrast to what is observed in the EMS studies, multiple IHF proteins weakly assemble on the longer duplex. Further, we note that $\langle s \rangle$ increases up to 1.6 μ M IHF, but subsequently decreases slightly as the concentration of IHF is further increased. This is discussed further below.

The concentration of nonspecific flanking sequence (247 bp) is much greater than the unique 27 bp specific site in the cos-274 DNA, therefore, we hypothesized that nonspecific binding would dominate the binding reaction observed in the AUC. We therefore examined IHF binding to the nonspecific 274 bp DNA. In contrast to the relatively simple binding curve observed with the specific DNA, binding of IHF to ns-274 is remarkable. An increase in $\langle s \rangle$ is initially observed similar to that for the cos-274 DNA (Fig. 5B); however, a 2.5 S species emerges as the protein concentration is further increased, at the expense of the larger s species. We note that the appearance of two distinct peaks in the $c(s)$ distributions collected for the nonspecific DNA suggests that the kinetics of interconversion between these two populations is slow relative to the timescale of SV-AUC experiment (a few hours). We further note that this new species has a sedimentation coefficient *smaller* than that of the free DNA (4.2 S). This remarkable transition is quite apparent in the plot of $\langle s \rangle$ as a function of IHF concentration presented in Fig. 5C, which further shows that a similar, though less striking trend is observed with the specific cos-274 DNA. How can this observation be rationalized?

EMS studies confirm that IHF binds to both duplexes in this concentration range to afford complexes of increasing mass (Sanyal et al., 2014). This is consistent with the initial increase in $\langle s \rangle$; however, s is also a function of particle shape and *decreases* as the complex becomes more elongated in solution (Correia & Stafford, 2015; Stafford et al., 2001). Thus, the most parsimonious explanation is that while both species represent IHF–DNA complexes of increasing buoyant mass, they ultimately afford complexes that differ significantly in shape. Specifically, the 2.5 S species must be elongated, even relative to the free, unbound duplex (4.2 S). Within this context, the contour length of a 274 bp DNA is about 93 nm, which is larger than the persistence length of DNA (35–50 nm depending upon solution conditions) (Bustamante, Bryant, & Smith, 2003; Kratky & Porod, 1949; Wang, Yin,

Landick, Gelles, & Block, 1997). This means that the free 274 bp DNA molecule is flexible during sedimentation and it will behave like a floppy polymer in solution. Hence, we propose that the transition from the ~10 to the 2.5 S “species” is due to a dramatic conformational change in the DNA resulting from a substantial increase in the stiffness of nucleoprotein complex as the duplex becomes saturated with IHF. Thus, even though the molar mass of a nearly saturated DNA species is approximately four to fivefold higher than the molar mass of the free DNA, the sedimentation coefficient is significantly smaller, due to a dramatic increase in the frictional coefficient of the “rigid rod” relative to that of free DNA.

We developed a quantitative model to analyze IHF binding to both specific and nonspecific 274 bp duplex substrates and to determine the energetics associated with such a conformational transition (Fig. 6; Sanyal et al., 2014). Briefly, the model takes into account cooperative, nonspecific assembly of IHF on DNA such that depending on the number of IHF molecules bound, the DNA will eventually transition from a flexible strand to an extended rigid, rod-like nucleoprotein complex. The model incorporates three DNA species: free DNA, flexible IHF–DNA complexes, and rigid, extended IHF–DNA complexes. At low IHF concentrations, IHF binds weakly and cooperatively to bend, wrap, and condense the DNA (i.e., at least one IHF bound); the increase in the buoyant mass and the decrease in the frictional ratio results in a significant increase in the sedimentation coefficient of the complex. Importantly, the occluded site size for IHF binding in a bent complex (~35 bp) (Holbrook et al., 2001) is significantly larger than that of a nonspecific, linear complex (~8 bp) (Holbrook et al., 2001) and the DNA must therefore be “unbent” in order to fully saturate the duplex. Thus, above a threshold value, the ensemble of IHF–DNA complexes is “linearized” or “unbent” so that all of the available nonspecific sites can be bound by IHF. This increases the stiffness and extends the shape of the nucleoprotein complex, even relative to free DNA. The model assumes this “unbending” is driven by the binding of additional IHF molecules, and this reaction is described by a single equilibrium constant (essentially a Hill-binding model). Details of the mathematical model and fitting procedures are presented in Sanyal et al. (2014).

The best-fit parameters obtained from this analysis are shown in Table 1 and the results of the fit are shown as the smooth curves in Fig. 5C. The model well describes the experimental data and implies that the unbending (linearization) energetic barrier is at least 1.4 kcal/mol. In other words, it is at least 13 times easier to linearize a nonspecific duplex with IHF bound than a specific duplex with IHF bound. This reflects the stability of the severe bend introduced by IHF bound to its cognate site. In addition, the model suggests that four IHF dimers are required to trigger the unbending transition. Importantly, the AUC studies discussed here are consistent with electron microscopy images (Sarkar et al., 2009) that show rigid, extended IHF–DNA complexes and we have provided a solution-based, thermodynamic analysis of the transition from condensed to extended nucleoprotein complexes.

4. SUMMARY

In this review we have described the power of sedimentation equilibrium and sedimentation velocity approaches to interrogate protein–DNA binding interactions. We have paid special attention to how these data can be analyzed using appropriate mathematical models to obtain biologically relevant, solution-based thermodynamic information on the interactions. These approaches are particularly well suited to these studies because they are founded on well-understood physical principles and consequently provide rigorous quantitative information on the interactions, largely free from simplifying assumptions. In addition, the physical responses to protein–DNA interactions—the increase in the molar mass (detected by sedimentation equilibrium and velocity) and any associated changes in nucleoprotein complex conformation (detected by sedimentation velocity), can be probed and quantified in a rigorous manner. The mechanism of specific recognition of a DNA-binding element within the context of a vast excess of nonspecific DNA can be addressed, which is essential since this is the situation *in vivo*; regulatory proteins must not only compete with nonspecific binding proteins for their cognate element, but they must also overcome their own nonspecific binding interactions to avoid sequestration by the cellular genome. Importantly, the tools described here are general and can be applied to any experimental system to provide mechanistic insight into biological function of proteins within a cell.

ACKNOWLEDGMENTS

We thank Dr. Saurarshi J. Sanyal for generating the data in Fig. 2 and Dr. John Sumida for guidance on the AUC fluorescence detection system.

REFERENCES

- Aeling KA, Opel ML, Steffen NR, Tretyachenko-Ladokhina V, Hatfield GW, Lathrop RH, et al. Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: The role of DNA deformation energy. *The Journal of Biological Chemistry*. 2006; 281(51):39236–39248. [PubMed: 17035240]
- Bujalowski W. Thermodynamic and kinetic methods of analyses of protein-nucleic acid interactions. From simpler to more complex systems. *Chemical Reviews*. 2006; 106(2):556–606. [PubMed: 16464018]
- Bujalowski W, Lohman TM, Anderson CF. On the cooperative binding of large ligands to a one-dimensional homogeneous lattice: The generalized three-state lattice model. *Biopolymers*. 1989; 28(9):1637–1643. [PubMed: 2775853]
- Bujalowski W, Overman LB, Lohman TM. Binding mode transitions of Escherichia coli single strand binding protein-single-stranded DNA complexes. Cation, anion, pH, and binding density effects. *The Journal of Biological Chemistry*. 1988; 263(10):4629–4640. [PubMed: 3280566]
- Bustamante C, Bryant Z, Smith SB. Ten years of tension: Single-molecule DNA mechanics. *Nature*. 2003; 421(6921):423–427. [PubMed: 12540915]
- Correia JJ, Stafford W. Sedimentation velocity: A classical perspective. *Methods in Enzymology*. 2015; 562:49–80. [PubMed: 26412647]
- Demeler B, van Holde KE. Sedimentation velocity analysis of highly heterogeneous systems. *Analytical Biochemistry*. 2004; 335(2):279–288. [PubMed: 15556567]
- Epstein IR. Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions. *Biophysical Chemistry*. 1978; 8(4):327–339. [PubMed: 728537]
- Holbrook JA, Tsodikov OV, Saecker RM, Record MT Jr. Specific and non-specific interactions of integration host factor with DNA: Thermodynamic evidence for disruption of multiple IHF surface

- salt-bridges coupled to DNA binding. *Journal of Molecular Biology*. 2001; 310(2):379–401. [PubMed: 11428896]
- Khrapunov S, Brenowitz M, Rice PA, Catalano CE. Binding then bending: A mechanism for wrapping DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(51):19217–19218. [PubMed: 17159146]
- Kratky O, Porod G. Diffuse small-angle scattering of X-rays in colloid systems. *Journal of Colloid Science*. 1949; 4(1):35–70. [PubMed: 18110601]
- Laue, TM.; Shah, BD.; Ridgeway, TM.; Pelletier, SL. Computer-aided interpretation of analytical sedimentation data for proteins. Royal Society of Chemistry; Cambridge: 1992.
- Laue TM, Stafford WF 3rd. Modern applications of analytical ultracentrifugation. *Annual Review of Biophysics and Biomolecular Structure*. 1999; 28:75–100.
- Lebowitz J, Lewis MS, Schuck P. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Science*. 2002; 11(9):2067–2079. [PubMed: 12192063]
- Lohman TM, Bujalowski W. Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein-DNA interactions: Spectroscopic approaches to monitor binding. *Methods in Enzymology*. 1991; 208:258–290. [PubMed: 1779838]
- McGhee JD, von Hippel PH. Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *Journal of Molecular Biology*. 1974; 86(2):469–489. [PubMed: 4416620]
- Ortega ME, Catalano CE. Bacteriophage lambda gpNu1 and Escherichia coli IHF proteins cooperatively bind and bend viral DNA: Implications for the assembly of a genome-packaging motor. *Biochemistry*. 2006; 45(16):5180–5189. [PubMed: 16618107]
- Philo JS. Limiting the sedimentation coefficient for sedimentation velocity data analysis: Partial boundary modeling and g(s) approaches revisited. *Analytical Biochemistry*. 2011; 412(2):189–202. [PubMed: 21284932]
- Sanyal SJ, Yang TC, Catalano CE. Integration host factor assembly at the cohesive end site of the bacteriophage lambda genome: Implications for viral DNA packaging and bacterial gene regulation. *Biochemistry*. 2014; 53(48):7459–7470. [PubMed: 25335823]
- Sarkar T, Petrov AS, Vitko JR, Santai CT, Harvey SC, Mukerji I, et al. Integration host factor (IHF) dictates the structure of polyamine-DNA condensates: Implications for the role of IHF in the compaction of bacterial chromatin. *Biochemistry*. 2009; 48(4):667–675. [PubMed: 19132923]
- Schuck P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Analytical Biochemistry*. 2003; 320(1):104–124. [PubMed: 12895474]
- Stafford WF, Jacobsen MP, Woodhead J, Craig R, O'Neill-Hennessey E, Szent-Gyorgyi AG. Calcium-dependent structural changes in scallop heavy meromyosin. *Journal of Molecular Biology*. 2001; 307(1):137–147. [PubMed: 11243809]
- Tsodikov OV, Holbrook JA, Shkel IA, Record MT Jr. Analytic binding isotherms describing competitive interactions of a protein ligand with specific and non-specific sites on the same DNA oligomer. *Biophysical Journal*. 2001; 81(4):1960–1969. [PubMed: 11566770]
- Ucci JW, Cole JL. Global analysis of non-specific protein-nucleic interactions by sedimentation equilibrium. *Biophysical Chemistry*. 2004; 108(1–3):127–140. [PubMed: 15043926]
- von Hippel PH, Berg OG. On the specificity of DNA-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1986; 83(6):1608–1612. [PubMed: 3456604]
- Wang MD, Yin H, Landick R, Gelles J, Block SM. Stretching DNA with optical tweezers. *Biophysical Journal*. 1997; 72(3):1335–1346. [PubMed: 9138579]
- Yang TC, Maluf NK. Characterization of the non-specific DNA binding properties of the adenoviral IVa2 protein. *Biophysical Chemistry*. 2014; 193–194:1–8.
- Yphantis DA. Equilibrium ultracentrifugation of dilute solutions. *Biochemistry*. 1964; 3:297–317. [PubMed: 14155091]

Figure 1.

The McGhee and von Hippel approach is used to resolve the nonspecific binding properties of protein–DNA interactions. This approach assumes that the protein interacts with a one-dimensional DNA lattice and uses conditional probabilities to analyze three nonspecific binding properties: (A) the occluded site size (number of bases occluded by the bound protein), (B) the overlapping effect, and (D) the nearest-neighbor cooperativity. Because of the overlapping effect, under this nonspecific binding scenario, it is difficult to fully saturate the duplex (C).

Figure 2.

Stoichiometry of IHF binding to a 10 bp nonspecific duplex determined by SE-AUC.

Samples were prepared by mixing various concentrations of IHF with 400 nM DNA; SE-AUC experiments were performed at three different rotor speeds (22, 27, and 33 K RPM) at 4 °C. The data were analyzed by an ideal single-species model (Eq. 13) to approximate the average molecular weight of all species, from which the degree of binding was calculated using Eq. (15). The error bar indicates 68.3% confidence intervals. *Data provided by Dr. Saurarshi J. Sanyal.*

Figure 3.

NLLS analysis of the primary sedimentation equilibrium data using the MvH model. The data in this figure were analyzed by NLLS according to the MvH model. The smooth curves are the result of this analysis fixing K_{ns} , ω , and n at the values determined from other approaches: $K_{ns} = 3.9 \times 10^4 M^{-1}$, $\omega = 125$, and $n = 18.8 \text{ bp}$ (Yang & Maluf, 2014). In this analysis, the concentration of free IVa2 protein and free DNA at the meniscus position were allowed to float. The insets show the predicted concentration of the total IVa2 protein calculated at 5, 7.5, and 10 K RPM, using the best-fit parameters from the NLLS analysis (Yang & Maluf, 2014).

Figure 4.

Interrogation of binding of IHF to minimal DNA duplexes using sedimentation velocity analytical ultracentrifugation (SV-AUC). Increasing concentrations of IHF were added to 27 bp specific and nonspecific DNA model duplexes, and their sedimentation behavior was monitored by SV-AUC. The $c(s)$ distribution for each binding experiment was calculated using Sedfit. (A) Normalized $c(s)$ profiles for the specific DNA. (B) Normalized $c(s)$ profiles for the nonspecific DNA. (C) Weight-average sedimentation coefficients for each of the $c(s)$ distributions shown in panels A (triangles) and B (triangles) were calculated using Sedfit and are plotted as a function of IHF concentration. The dotted line represents the best fit of the specific binding data to the nonspecific finite lattice DNA binding model, which does not adequately describe the data. The solid lines represent the best fit of global analysis of the nonspecific and specific binding data to the (i) nonspecific finite lattice DNA binding and (ii) case 1 models, respectively.

Figure 5.

Interrogation of IHF assembly on specific and nonspecific 274 bp DNAs using SV-AUC. The specific duplex has a 27 bp IHF recognition element embedded at the center of the duplex while the nonspecific duplex is of random sequence. Increasing concentrations of IHF were added to each duplex, and their sedimentation behavior was monitored by SV-AUC. The $c(s)$ distribution for each binding experiment were calculated using Sedfit. (A) Normalized $c(s)$ profiles for the specific 274 bp DNA. (B) Normalized $c(s)$ profiles for the nonspecific 274 bp DNA. (C) Weight-average sedimentation coefficients for each of the $c(s)$ distributions shown in panel A (circles, specific DNA) and panel B (circles, nonspecific DNA) were calculated using Sedfit and are plotted as a function of IHF concentration. The solid lines represent the best fits of global analysis of the ensemble of binding data to the DNA unbending model (Fig. 6).

Figure 6.

Model for IHF–DNA complexes. Details are provided in the text and in Sanyal et al. (2014).

Table 1

Analysis of the ns-274 and cos-274 AUC Data to Approximate DNA Unbending Energy

	m^a	$\langle s \rangle$ (S)	K_L (M^{-4})	Apparent $K_{L,app}^b$ (M^{-1})	G^c (kcal/mol)
ns-274	4 ± 2	(10.0 ± 0.4)	$K_{L,ns} = (1.2 \pm 0.8) \times 10^{36}$	$K_{L,app-ns} = (1.0 \pm 0.2) \times 10^9$	-11.5
cos-274	4 (2, N/A) ^d	(3.7 ± 0.4)	$K_{L,sp} \quad 3.7 \times 10^{31}$	$K_{L,app-sp} \quad 7.8 \times 10^7$	-10.1

^a m is the number of IHF required to unbent the DNA.^b $K_{L,app}$ values were calculated from $(K_L)^{1/m}$.^cThe free energy change was calculated from $G = -RT \ln(K_{L,app})$.^dThe upper limit of m for cos-274 DNA is not available due to the insufficient unbending of specific DNA.