Biophysical Chemistry 154 (2011) 82–89

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Thermodynamic linkage of large-scale ligand aggregation with receptor binding

Nasib Karl Maluf *, Teng-Chieh Yang

University of Colorado Denver, Department of Pharmaceutical Sciences, School of Pharmacy, C238-P15, 12700 E. 19th Ave, Aurora, CO 80045, United States

ARTICLE INFO

Article history:
Received 6 December 2010
Received in revised form 13 January 2011
Accepted 13 January 2011
Available online 22 January 2011

Keywords:
Aggregation
Thermodynamic linkage
Self association
Indefinite aggregation
Protein–DNA interaction

ABSTRACT

There are many examples in the literature that deal explicitly with the coupling of ligand oligomerization with receptor binding. For example, many transcription factors dimerize and this plays a fundamental role in sequence specific DNA recognition. However, many biological macromolecules undergo reversible, large scale aggregation processes, some of which are indefinite. The thermodynamic coupling of these aggregation processes to other processes, such as protein–protein and protein–DNA interactions, has not been explored in depth. Here we consider the thermodynamic consequences of large scale ligand aggregation on the determination of fundamental thermodynamic parameters, such as equilibrium binding constants and ligand–receptor stoichiometries. We find that a fundamental consequence of an aggregating ligand is that the free ligand concentration (ligand that is not found in aggregates) is buffered over a wide total ligand concentration range. In general, the larger the size of the aggregates, the wider the range over which the free ligand concentration is buffered. An additional consequence of this observation is that an upper limit is set on the fractional occupancy of the ligand's receptor, such that even if the ligand is over-expressed to very high levels in the cell, this will not necessarily ensure that 100% of the ligand's receptors will be occupied. The implications of these results for sequence specific DNA binding proteins will be discussed.

1. Introduction

Many proteins are difficult to study *in vitro* due to their propensity to self aggregate (1). In most cases, investigators attempt to limit this aggregation experimentally, typically by studying truncated proteins. The potential role that reversible protein aggregation may play in the mechanisms of action of these proteins is often ignored, or only studied when the protein forms relatively small, well defined aggregates, such as dimers [2,3]. For example, several DNA binding proteins have been shown to reversibly assemble into large aggregates [4–8], and many transcription factors have been shown to posses one of the most common protein domains observed in nature, Sterile Alpha Motif (SAM) [9], which has been shown in many cases to confer the ability to polymerize in a head to tail fashion [5,6,10–12].

A fundamental consequence of large scale protein aggregation is that the free concentration of the smallest polymerizing unit (e.g. the monomer), will be buffered over a wide total protein concentration range [8]. This observation has significant consequences when considering a wide range of biological processes, since the fractional occupancy of the receptor is determined solely by the free ligand concentration [13], where the free ligand is defined as the smallest polymerizing protein unit. Thus, cells could conceivably overexpress a self-assembling protein ligand to high levels, while simultaneously maintaining a constant free ligand concentration, and guarantee a fixed fractional saturation of the ligand's receptor. Importantly, the fractional saturation may be set at a value less than 100%, even when a large excess of total ligand is produced. Previous seminal works that consider the linkage of large scale ligand aggregation with receptor binding have not explicitly dealt with this phenomena [4,14]. We show here that if a ligand aggregates reversibly, the linkage of ligand aggregation with receptor binding must be accounted for to deduce the equilibrium binding mechanism. Failure to do so in model development will result in incorrect measurements of equilibrium binding constants, cooperativities and ligand–receptor stoichiometries. Furthermore, consideration of this linkage suggests novel regulation mechanisms for control of cellular processes.

2. Theory and results

2.1. Large aggregation processes buffer the free monomer concentration

Our objective here is to study the dependence of the free monomer concentration, or equivalently, the free concentration of the smallest polymerizing unit, on the total monomer concentration for reversible, large scale aggregation processes. To begin this investigation, we have used an isodesmic assembly model (Scheme 1), where the affinity for addition of each monomer to the growing aggregate is described by the same equilibrium constant, L [15,16].

* Corresponding author. Tel.: +1 303 724 4036; fax: +1 303 724 2627.
E-mail address: karl.maluf@ucdenver.edu (N.K. Maluf).

0301-4622/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2011.01.004

$$(N\text{-}1)A + A \underset{L}{\rightleftharpoons} (N\text{-}2)A + A_2 \underset{L}{\rightleftharpoons} (N\text{-}3)A + A_3 \underset{L}{\rightleftharpoons} \cdots \underset{L}{\rightleftharpoons} A_N$$

Scheme 1. Isodesmic assembly which terminates at a stoichiometry of *N*.

We use the term "large scale aggregation processes" to describe the formation of aggregates with stoichiometries larger than around 30. Using the isodesmic model, we can obtain an expression that relates the total monomer concentration, $[A_T]$, to the free monomer concentration, $[A_1]$. For an aggregation process that terminates at a stoichiometry of N, the total monomer concentration (in mole/L) is given by:

$$[A_T] = [A_1] + 2[A_2] + 3[A_3] + \cdots + N[A_N] \tag{1}$$

The equilibrium constant for the formation of each stoichiometric species is given by:

$$L = \frac{[A_j]}{[A_{j-1}][A_1]} \tag{2}$$

Substitution of Eq. (2) into Eq. (1) yields:

$$[A_T] = [A_1] + 2L[A_1]^2 + 3L^2[A_1]^3 + \cdots + NL^{N-1}[A_1]^N \tag{3}$$

Multiplying Eq. (3) by L, and defining $L[A_1] \equiv x$ yields:

$$L[A_T] = x + 2x^2 + 3x^3 + \cdots + Nx^N \tag{4}$$

Dividing Eq. (4) by x yields:

$$\frac{L[A_T]}{x} = 1 + 2x + 3x^2 + \cdots + Nx^{N-1} \tag{5}$$

Subtracting Eq. (4) from Eq. (5) yields:

$$\frac{L[A_T]}{x} - L[A_T] + Nx^N = 1 + x + x^2 + \cdots + x^{N-1} \tag{6}$$

Now we multiply the right hand side (RHS) of Eq. (6) by x, and then subtract this result from the RHS of Eq. (6):

$$\text{RHS} - \text{RHS}\cdot x = 1 - x^N \tag{7}$$

Solving Eq. (7) for RHS yields:

$$\text{RHS} = \frac{1-x^N}{1-x} \tag{8}$$

Substituting Eq. (8) into the RHS of Eq. (6), inserting $L[A_1]$ in for x, and finally solving for $[A_T]$ yields:

$$[A_T] = \frac{[A_1]\left(1-(L[A_1])^N\right)}{(1-L[A_1])^2} - \frac{N(L[A_1])^{N+1}}{L(1-L[A_1])} \tag{9}$$

Taking the limit of Eq. (9), when N approaches infinity and $L[A_1] < 1$ yields:

$$[A_T] = \frac{[A_1]}{(1-L[A_1])^2} \tag{10}$$

which is the well known result for the indefinite, isodesmic aggregation case [15,16], i.e. when there is no upper limit on the stoichiometry of the highest aggregation state (and Scheme 1 represents an infinite series). For the indefinite case, it can be seen that the product $L[A_1]$ must be less than one; otherwise the infinite sum shown in Eq. (4) will be guaranteed to diverge. For the definite case, i.e. when N is equal to a finite value (the aggregation terminates at a stoichiometry of N), there is no such requirement. However, inspection of Eq. (9) reveals that it is undefined when $L[A_1] = 1$, or equivalently, when $[A_1] = 1/L$. To determine the value of $[A_T]$ when $L[A_1] = 1$, we evaluated Eq. (4) when $x = 1$ (i.e. $L[A_1] = 1$), which yields:

$$[A_T] = \frac{N(N+1)}{2L} \tag{11}$$

Thus, the function $[A_T]$ is correctly defined as piecewise continuous, where it takes the form of Eq. (9) for the domain $0 < [A_1] < 1/L$ and $[A_1] > 1/L$, and it takes on the form of Eq. (11) for $[A_1] = 1/L$.

Fig. 1A shows simulations using Eqs. 9–11. In these simulations, L was fixed at 100 μM^{-1}, and $[A_T]$ was calculated over a wide range of $[A_1]$. In Fig. 1, we have plotted $[A_1]$ as a function of $[A_T]$ to illustrate the "buffering capacity" of the large aggregates with respect to the free monomer concentration. Fig. 1A shows that for an isodesmic aggregation process that terminates at a stoichiometry of 30 or greater, the free monomer concentration changes very little between 0.5 and 10 μM total monomer. Even for smaller aggregation processes, e.g. when $N = 10$, we see that $[A_1]$ only increases from ~10 nM to ~14 nM upon increasing the total monomer concentration, $[A_T]$ from 0.5 μM to 10 μM. In other words, a 20 fold increase in the total monomer concentration only results in a 1.4 fold increase in the free monomer concentration. Fig. 1B shows the same simulations plotted on a log scale. For aggregation processes that terminate at $N = 30$ or greater, we see that the dependence of $[A_1]$ on $[A_T]$ is weak over a very wide total monomer concentration range. For example, for $N = 30$, it takes a four order of magnitude increase in $[A_T]$ to increase $[A_1]$ from ~9.2 nM to 14 nM. Thus, for most practical considerations, we see that large aggregation processes behave similarly to indefinite aggregation processes ($N \to \infty$) in terms of placing an upper limit on the free monomer concentration that is practically achievable in solution. For the isodesmic case, this upper limit is given to a very good approximation by $1/L$. We conclude that over physiologically relevant concentration ranges, the free monomer concentration can be buffered at around $1/L$, if the monomer participates in a reversible, isodesmic large scale aggregation process.

2.2. Linkage of ligand aggregation with receptor binding for large aggregation processes

To investigate the effect of free ligand buffering on the fractional saturation of a ligand's receptor, we have performed a series of simulations according to Scheme 2. In this scheme, a ligand monomer can bind reversibly to its receptor, or it can reversibly aggregate up to a maximal stoichiometry of N. The total ligand monomer concentration, for Scheme 2, is given by:

$$[A_T] = [AR] + [A_1] + 2[A_2] + 3[A_3] + \cdots + N[A_N] \tag{12}$$

The equilibrium association constant for free monomer binding to receptor is given by:

$$K = \frac{[AR]}{[A_1][R_f]} \tag{13}$$

where $[R_f]$ is the free receptor concentration. The total receptor concentration is given by:

$$[R_T] = [R_f] + [AR] = [R_f] + K[A_1][R_f] = [R_f](1 + K[A_1]) \tag{14}$$

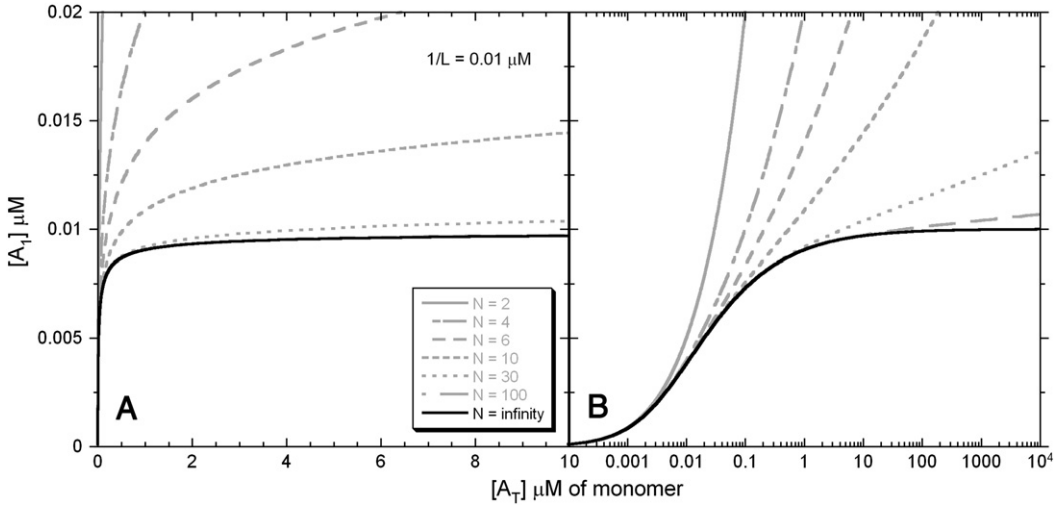


Fig. 1. Dependence of the free monomer concentration, $[A_1]$, on the total monomer concentration, $[A_T]$, as a function of aggregation termination stoichiometry, N. Eqs. (9)–(11) were used to generate these simulations, which were derived from Scheme 1. Panel A shows the simulations plotted on a linear scale, while Panel B is plotted on a log-scale. These simulations show that as N increases, the variation in $[A_1]$ with $[A_T]$ decreases markedly. Even for aggregation stoichiometries as small as 30, the $[A_1]$ is buffered over a wide range of $[A_T]$. Note that the shape of these simulations is not affected by the choice of L; i.e. the exact same figure can be drawn for any value of L by multiplying the x and y axis by a constant. This can be seen by multiplying both sides of Eq. (9) by L, then plotting $L[A_1]$ as a function of $L[A_T]$.

Solving Eq. (14) for $[R_f]$ yields:

$$[R_f] = \frac{[R_T]}{1 + K[A_1]} \tag{15}$$

Substituting Eqs. (2), (13) and (15) into Eq. (12) yields:

$$[A_T] - \frac{K[A_1][R_T]}{1 + K[A_1]} = [A_1] + 2L[A_1]^2 + 3L^2[A_1]^3 + \cdots + NL^{N-1}[A_1]^N \tag{16}$$

Recognizing that the RHS of Eq. (16) is identical to the RHS of Eq. (9), we arrive at:

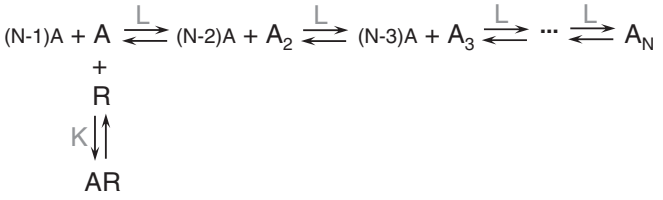
$$[A_T] = \frac{K[A_1][R_T]}{1 + K[A_1]} + \frac{[A_1]\left(1-(L[A_1])^N\right)}{(1-L[A_1])^2} - \frac{N(L[A_1])^{N+1}}{L(1-L[A_1])} \tag{17}$$

For the indefinite case (i.e. when $N \to \infty$, and $[A_1] < 1/L$), we obtain:

$$[A_T] = \frac{K[A_1][R_T]}{1 + K[A_1]} + \frac{[A_1]}{(1-L[A_1])^2} \tag{18}$$

For the definite case (when N takes on a finite value), we again see that Eq. (17) is undefined for $[A_1] = 1/L$. To determine the value of $[A_T]$ when $[A_1] = 1/L$, we multiplied Eq. (16) by L, then set $x = 1$. This resulted in:

$$[A_T] = \frac{N(N+1)}{2L} + \frac{K[R_T]}{L + K} \tag{19}$$



Scheme 2. Linkage of large scale ligand aggregation with receptor binding, when the ligand monomer is the active binding species.

Thus, we see that for finite N, $[A_T]$ is again a piecewise continuous function, which takes the form of Eq. (17) for the domain $0 < [A_1] < 1/L$ and $[A_1] > 1/L$, and takes the form of Eq. (19) for $[A_1] = 1/L$.

For Scheme 2, the fractional saturation of the ligand's receptor is given by:

$$f_{AR} = \frac{[AR]}{[R_T]} = \frac{K[A_1][R_f]}{[R_f] + K[A_1][R_f]} = \frac{K[A_1]}{1 + K[A_1]} \tag{20}$$

We imagine that the signal we are experimentally monitoring originates from the receptor. For example, if the receptor is a DNA molecule, and the ligand is a sequence specific DNA binding protein, the signal may correspond to the fraction of radiolabeled DNA that is retained on a nitrocellulose filter as a function of the total protein concentration. Another possibility is that the total signal may correspond to the change in the fluorescence anisotropy that originates from a fluorescently labeled DNA molecule upon binding to its protein ligand. To quantify the change in the intrinsic receptor signal that is brought about due to ligand binding, we write [13]:

$$S_{obs} = S_R(1 - f_{AR}) + S_{AR}f_{AR} \tag{21}$$

where S_{obs} is the total observed signal, S_R is the intrinsic signal that originates from the unligated receptor and S_{AR} is the intrinsic signal that originates from the receptor–ligand complex. Substituting Eq. (20) into Eq. (21), we obtain:

$$(S_{obs} - S_R) = (S_{AR} - S_R)\frac{K[A_1]}{1 + K[A_1]} \tag{22}$$

Defining $\Delta S_{obs} \equiv S_{obs} - S_R$, and $\Delta S_{AR} \equiv S_{AR} - S_R$, i.e. as the change in receptor signal upon ligand binding, we obtain:

$$\Delta S_{obs} = \Delta S_{AR}\frac{K[A_1]}{1 + K[A_1]} \tag{23}$$

Determination of S_R simply requires measuring the total signal before any ligand has been added to solution. Determination of S_{AR} is also straightforward, for the case where no significant ligand aggregation occurs. In this case, S_{AR} is simply taken as the value of the total signal upon complete saturation of all the receptor binding sites in solution, so that the only signal that is measured originates

from the receptor–ligand complex. However, in the case of significant ligand aggregation, we show that S_{AR} is not reliably determined from the plateau value of the observed signal change.

Fig. 2 shows a series of simulations for ΔS_{obs} as a function of $[A_T]$, carried out for two different total receptor concentrations, over a range of ligand aggregation termination stoichiometries, with $\Delta S_{AR}=1$, $L=100\ \mu M^{-1}$ and $K=200\ \mu M^{-1}$. For simulations carried out at $[R_T]=1$ pM, the free ligand concentration is unaffected by receptor binding, since $[A_1]>>[R_T]$. For simulations carried out at $[R_T]=10\ \mu M$, the free ligand concentration is much smaller than the ligand–receptor concentration, since the vast majority of added $[A_1]$ binds the receptor.

First we discuss the simulations carried out at $[R_T]=1$ pM (Fig. 2A). When these simulations were carried out as a function of the aggregation termination stoichiometry, N, we see that for increasing N, more total ligand must be added to solution to drive full saturation of the receptor. However, for the indefinite case (when $N\rightarrow\infty$), we see that no matter how much total ligand is added, full saturation of the receptor can never be achieved, at a constant receptor concentration. This is due to the requirement that $[A_1]<1/L$. Thus, due to the particular values chosen for L and K, the maximal saturation of the receptor, ΔS^*_{obs}, will be given by substituting $1/L$ in for $[A_1]$ in Eq. (23), which yields:

$$\Delta S^*_{obs}=\Delta S_{AR}\frac{K}{K+L} \tag{24}$$

For the values chosen for L and K, Eq. (24) predicts an apparent maximal signal of $\Delta S^*_{obs}=200/(100+200)=2/3$ (where $\Delta S_{AR}\equiv 1$). However, even though an apparent plateau has been reached for the indefinite case, this does not mean that all the receptor molecules are bound by ligand. In fact, this means that a fraction of the receptor molecules are guaranteed to be unoccupied by ligand, regardless of the quantity of total ligand that is added to solution. This fraction of free receptor is simply $1-2/3=1/3$. Eq. (24) also points out that if $L\ll K$, ligand aggregation will not affect quantitatively the plateau value.

While these conclusions are strictly valid only for the indefinite case, we can see that for large values of N (≥~30), an apparent plateau value is obtained that is essentially identical to what is observed for the indefinite case, i.e. it is well approximated by Eq. (24). For example, for $N=30$, we see that for total ligand concentrations from 1 to 100 μM, very little change is observed in ΔS_{obs}. This is due to the free ligand buffering effect discussed in the preceding section. It is clear that a small amount of experimental uncertainty could easily mask a slowly increasing plateau value.

A second consequence of the ligand buffering effect is that the ligand concentration at 50% of the total observed signal change cannot be strictly interpreted as the equilibrium dissociation constant, K_D. To see this quantitatively, we solved Eq. (23) for $[A_1]$ when $\Delta S_{obs}=0.5^*\Delta S^*_{obs}$:

$$[A_1]_{0.5}=\frac{1}{K+2L} \tag{25}$$

where $[A_1]_{0.5}$ refers to the value of the free ligand concentration at 50% of the observed plateau in ΔS_{obs}. It is important to point out that $[A_1]_{0.5}$ is not the free ligand concentration that is required to achieve 50% receptor saturation; it is simply the $[A_1]$ at 50% of the observed total signal change upon receptor binding. Since experimentally we have direct access to the total ligand monomer concentration, it is most useful to evaluate $[A_T]_{0.5}$, which is analogously defined as the value of $[A_T]$ at 50% of the observed plateau in ΔS_{obs}. To evaluate $[A_T]_{0.5}$ we substitute Eq. (25) into Eq. (10), which yields:

$$[A_T]_{0.5}=\frac{K+2L}{(K+L)^2} \tag{26}$$

Note that when $L\ll K$ (i.e. we can ignore the linkage between ligand aggregation and ligand–receptor binding), Eqs. (25) and (26) both reduce to $1/K$, where $1/K\equiv K_D$. Thus, in the absence of ligand assembly, and under conditions where the total receptor concentration, $[R_T]$, is much less than the K_D, we see that the free ligand concentration is essentially equal to the total ligand concentration, and the point where we observe 50% of the total signal change is identical to the point where 50% of the receptors are bound by ligand. However, when significant ligand aggregation occurs (i.e. when K is not much greater than L), the free ligand concentration is not equal to the total ligand concentration, and we cannot interpret the total ligand concentration that is required to achieve 50% of the total signal change as the equilibrium dissociation constant for the ligand–receptor binding reaction. For the specific case of isodesmic ligand

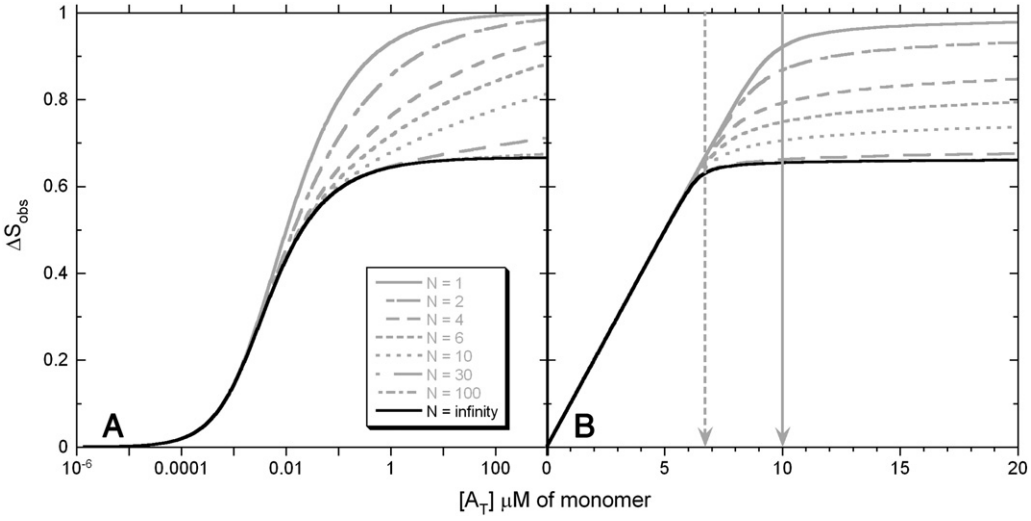


Fig. 2. Effect of free ligand buffering on receptor fractional saturation. Eqs. (17)–(19) and (23), which were derived from Scheme 2, were used in these simulations, with $\Delta S_{AR}=1$, $K=200\ \mu M^{-1}$ and $L=100\ \mu M^{-1}$. Panel A shows simulations performed at $[R_T]=1$ pM, while Panel B shows simulations performed at $[R_T]=10\ \mu M$. In Panel A, we see that for increasing N, more total ligand must be added to reach 100% fractional saturation of the receptor. However, for $N=$ infinity, we see that 100% receptor saturation cannot be achieved, due to the requirement that $[A_1]<1/L$. In fact, for $N>$~30, we see that impractical concentrations of $[A_T]$ are required to increase receptor saturation by only modest amounts. In Panel B we see that under stoichiometric conditions, the apparent breakpoint is shifted to a value smaller than 1:1 for large N. The solid gray arrow indicates the true reaction stoichiometry, while the dashed gray line indicates the apparent stoichiometry, calculated using Eq. (28).

aggregation, we interpret the apparent K_D using Eq. (26), and note that Eq. (26) is a good approximation for ligand termination stoichiometries ≥~30.

We next investigated the free ligand buffering effect on the determination of ligand binding stoichiometries. We performed a series of simulations at high total receptor concentrations ($[R_T]=10\,\mu M$), i.e. when $[R_T]>>1/K$ (Fig. 2B). The apparent breakpoints in these simulations are generally interpreted as quantifying the amount of ligand that is required to bind all the receptor; in this way, the ligand–receptor stoichiometry can be measured. For example, in the case of $N=1$, i.e. no ligand aggregation, we see that 10 μM of added ligand is required to completely bind all 10 μM of receptor; thus, the stoichiometry of this interaction is 1:1, exactly as predicted according to Scheme 2. However, when ligand aggregation is taken into account, we see that the apparent stoichiometric breakpoint shifts to smaller total ligand concentrations, while the initial regions of these simulations are all linear, with identical slopes. Based on this observation, we can write:

$$\Delta S_{obs} = \Delta S_{AR}\frac{[A_T]}{[R_T]} \tag{27}$$

for the linear part of Fig. 2B. To evaluate the breakpoint, we simply need to extrapolate Eq. (27) to the observed plateau value, which we have already determined is given by Eq. (24) for the indefinite case. This extrapolation yields:

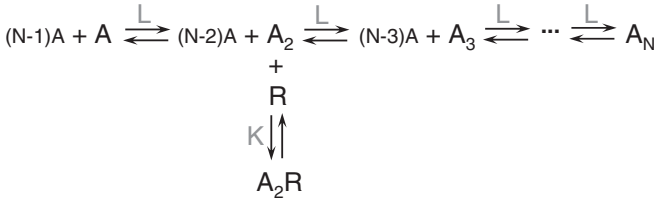
$$\left(\frac{[A_T]}{[R_T]}\right)_{App} = \frac{K}{K+L} \tag{28}$$

where $([A_T]/[R_T])_{App}$ is the apparent stoichiometric breakpoint. From Fig. 2B, we see that this extrapolation is valid for $N>\sim30$. The vertical dashed arrow shows the calculated break point using Eq. (28), with $K=200\,\mu M^{-1}$ and $L=100\,\mu M^{-1}$. Using these values, Eq. (28) yields $([A_T]/[R_T])_{App}=2/3$. These simulations show that if significant ligand aggregation occurs, stoichiometric breakpoints cannot be reliably interpreted as reaction stoichiometries, since they will always underestimate the true reaction stoichiometry.

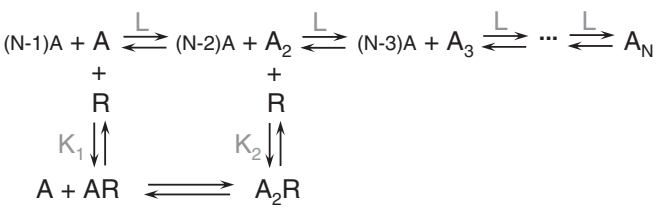
To illustrate this further, we have carried out simulations for the cases shown in Schemes 3 and 4, where ligand aggregation has been modeled using the indefinite isodesmic case (i.e. when $N=$ infinity). As we have found above, the indefinite case is a good model for large ligand termination stoichiometries. In Scheme 3, we consider the case where a preformed ligand dimer is required to bind to its receptor (e.g. a dimeric transcription factor binding to its cognate DNA sequence). In this case, we consider that the binding is highly cooperative; i.e. we do not see significant levels of monomers bound to the receptor. Following an analogous approach to that presented above, we obtain the following expressions for the maximal observed signal change, ΔS^*_{obs}, and the apparent stoichiometry, $([A_T]/[R_T])_{App}$:

$$\Delta S^*_{obs} = \Delta S_{A_2R}\frac{K}{K+L} \tag{29}$$

$$\left(\frac{[A_T]}{[R_T]}\right)_{App} = \frac{2K}{K+L} \tag{30}$$



Scheme 3. Linkage of large scale ligand aggregation with receptor binding, when the ligand dimer is the active binding species.



Scheme 4. Linkage of large scale ligand aggregation with receptor binding, when both the ligand monomer and dimer can bind the receptor.

Fig. 3A shows a stoichiometric binding simulation carried out using Scheme 3, with $L=100\,\mu M^{-1}$, $K=200\,\mu M^{-1}$, $\Delta S_{A2R}=1$ and $[R_T]=10\,\mu M$. The apparent ligand binding stoichiometry was calculated using Eq. (30), and is shown by the vertical dashed arrow.

Scheme 4 considers a case where both ligand monomers and dimers can bind their receptor, each generating a different signal change upon binding, given by ΔS_{AR} and ΔS_{A2R}, respectively. In this scheme, K_1 is the equilibrium association constant for the monomer–

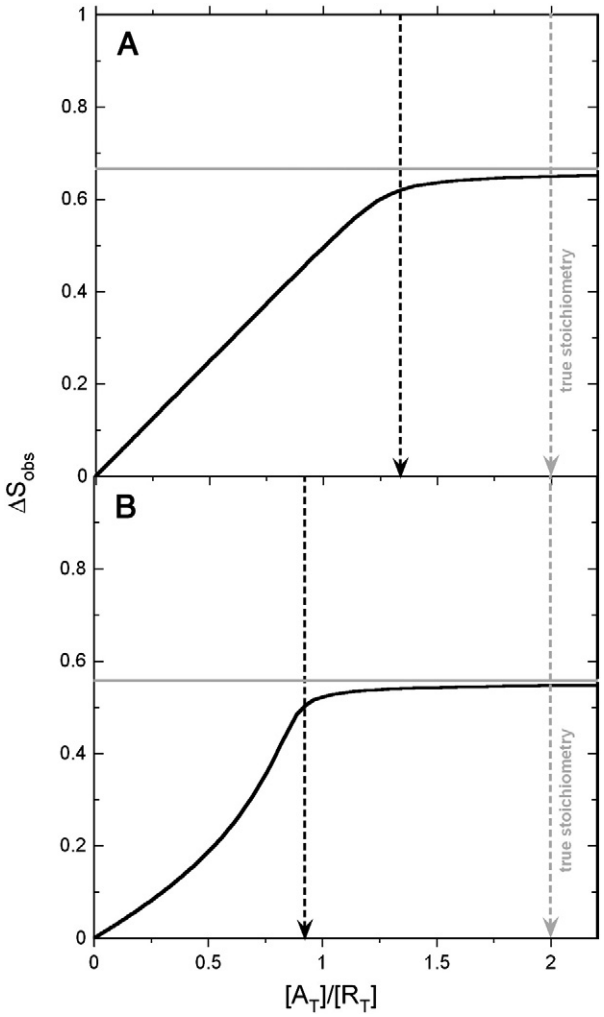


Fig. 3. Effect of free ligand buffering on determination of reaction stoichiometries. Panel A shows stoichiometric simulations using Scheme 3, while Panel B shows simulations using Scheme 4. In both cases, N was set at infinity. For Scheme 3, the values used in the simulation were $[R_T]=10\,\mu M$, $\Delta S_{A2R}=1$, $K=200\,\mu M^{-1}$ and $L=100\,\mu M^{-1}$. The horizontal gray line corresponds to the $\Delta S_{obs}{}^*$calculated using Eq. (29). The vertical dashed gray arrow is the true reaction stoichiometry from Scheme 3, while the black dashed arrow corresponds to the apparent stoichiometry calculated using Eq. (30). For Scheme 4, the values used in the simulation were $[R_T]=10\,\mu M$, $L=100\,\mu M^{-1}$, $K_1=450\,\mu M^{-1}$, $K_2=50\,\mu M^{-1}$, $\Delta S_{AR}=0.3$ and $\Delta S_{A2R}=4$. The horizontal gray line corresponds to the $\Delta S_{obs}{}^*$calculated using Eq. (31), while the dashed black arrow corresponds to the apparent reaction stoichiometry calculated using Eq. (32).

receptor binding reaction, while K_2 is the equilibrium association constant for the dimer–receptor binding reaction. The expressions for ΔS^*_{obs} and $([A_T]/[R_T])_{App}$ are:

$$\Delta S^*_{obs} = \frac{\Delta S_{AR}K_1 + \Delta S_{A_2R}K_2}{L + K_1 + K_2} \tag{31}$$

$$\left(\frac{[A_T]}{[R_T]}\right)_{App} = \frac{K_1 + 2K_2}{L + K_1 + K_2} \tag{32}$$

Fig. 3B shows a stoichiometric binding simulation carried out using Scheme 4, with $L = 100\ \mu M^{-1}$, $K_1 = 450\ \mu M^{-1}$, $K_2 = 50\ \mu M^{-1}$, $\Delta S_{AR} = 0.3$, $\Delta S_{A2R} = 4$ and $[R_T] = 10\ \mu M$. The apparent ligand binding stoichiometry was calculated using Eq. (32), and is shown by the vertical dashed arrow. The non-linear nature of this stoichiometric simulation is due to the assignment of different signal changes for monomer and dimer binding [13].

It can be seen from these simulations that it is impossible to determine the true stoichiometry if knowledge of the ligand aggregation state is not taken into account. For example, in the case of Scheme 3, even though the true ligand binding stoichiometry is known to be 2:1, the apparent value that would be measured from an experiment is intermediate between 1:1 and 2:1. For the particular values chosen for the equilibrium constants and signal changes used to simulate Scheme 4, we see that the apparent stoichiometry is close to 1:1, even though the true stoichiometry is 2:1. In all cases, we see that the true reaction stoichiometry is in fact the upper limit for the observed break point; thus if the affinity of ligand aggregation is comparable to the affinity of receptor binding, this will result in an underestimate of the true reaction stoichiometry.

2.3. Consequences of large scale ligand aggregation for the interpretation of ligand binding competition reactions

It is often observed that transcriptional activators and repressors compete in a mutually exclusive manner for the same DNA binding site [17]. Here we consider a simple system where two proteins compete for the same DNA binding site, where one of the protein ligands undergoes large scale aggregation, while the other does not (Scheme 5). In this scheme, the ligand A undergoes aggregation, while the ligand B is a stable monomer, and D represents the specific DNA sequence for which the two ligands compete. Based on Scheme 5, the total concentrations of A, B and D are given by:

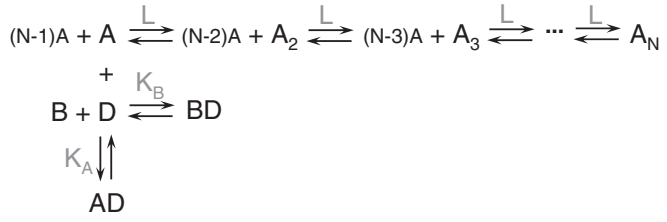
$$[A_T] = [AD] + [A_1] + 2[A_2] + 3[A_3] + \cdots + N[A_N] \tag{33}$$

$$[B_T] = [B_f] + [BD] \tag{34}$$

$$[D_T] = [D_f] + [AD] + [BD] \tag{35}$$

The equilibrium association constants for A and B binding to D are given by:

$$K_A = \frac{[AD]}{[A_1][D_f]} \tag{36}$$



Scheme 5. Effect of large scale ligand aggregation on competition for receptor binding.

$$K_B = \frac{[BD]}{[B_f][D_f]} \tag{37}$$

where $[B_f]$ is the free concentration of B. Substituting Eq. (36) into Eq. (33), Eq. (37) into Eq. (34), and Eqs. (36) and (37) into Eq. (35) yields:

$$[A_T] - K_A[A_1][D_f] = [A_1] + 2[A_2] + 3[A_3] + \cdots + N[A_N] \tag{38}$$

$$[B_f] = \frac{[B_T]}{1 + K_B[D_f]} \tag{39}$$

$$[D_f] = \frac{[D_T]}{1 + K_A[A_1] + K_B[B_f]} \tag{40}$$

Substituting Eqs. (39) and (40) into Eq. (38), and recognizing that for the indefinite case ($N \to \infty$), the RHS of Eq. (38) is given by Eq. (10), we arrive at:

$$[A_T] = \frac{K_A[A_1][D_T]\left(1 + K_B[D_f]\right)}{1 + K_B\left([D_f] + [B_T]\right) + K_A[A_1]\left(1 + K_B[D_f]\right)} + \frac{[A_1]}{(1 - L[A_1])^2} \tag{41}$$

Finally, combining Eqs. (39) and (40) provides a relationship between $[D_f]$ and $[A_1]$:

$$[D_f]^2 + \frac{(1 + K_A[A_1] + K_B([B_T] - [D_T]))}{(K_AK_B[A_1] + K_B)}[D_f] - \frac{[D_T]}{K_AK_B[A_1] + K_B} = 0 \tag{42}$$

Solving the quadratic and recognizing that the positive root guarantees that $[D_f]$ is positive yields:

$$[D_f] = \frac{1}{2}\sqrt{\left(\frac{1 + K_A[A_1] + K_B([B_T] - [D_T])}{K_AK_B[A_1] + K_B}\right)^2 + \frac{4[D_T]}{K_AK_B[A_1] + K_B}} - \left(\frac{1 + K_A[A_1] + K_B([B_T] - [D_T])}{2(K_AK_B[A_1] + K_B)}\right) \tag{43}$$

Note that the $[B_f]$ variable does not appear in Eqs. (42) and (43).

The fractional occupancy of the DNA site by either ligand A or B is defined as follows:

$$f_{AD} = \frac{[AD]}{[D_T]} = \frac{K_A[A_1](1 + K_B[D_f])}{1 + K_B([D_f] + [B_T]) + K_A[A_1](1 + K_B[D_f])} \tag{44}$$

$$f_{BD} = \frac{[BD]}{[D_T]} = \frac{K_B[B_T]}{1 + K_B([D_f] + [B_T]) + K_A[A_1](1 + K_B[D_f])} \tag{45}$$

To simulate the fractional occupancy of the DNA site by either ligand A or B, Eq. (43) was substituted into Eq. (41), and the $[A_1]$ was calculated implicitly (for any particular set of parameters K_A, K_B, $[D_T]$, $[B_T]$ and $[A_T]$) using the constraint $0 < [A_1] < 1/L$. Next, the calculated value of $[A_1]$ was substituted into Eq. (43) to calculate $[D_f]$. Finally, the entire set of values was substituted into Eqs. (44) and (45) to calculate f_{AD} and f_{BD}.

Fig. 4 shows the results of a representative set of simulations with $L = 100\ \mu M^{-1}$, $K_A = 200\ \mu M^{-1}$, $K_B = 20\ \mu M^{-1}$. In these simulations we set the affinity for ligand B binding to be 10-fold weaker than the affinity for ligand A. In Fig. 4A, we first simulated the fractional occupancy of the DNA upon titration with ligand A, in the absence of ligand B. This simulation shows that the maximal fractional saturation observed occurs at $f_{AD} = 2/3$, as expected based on Eq. (24), since Scheme 5 reduces to Scheme 2 for $[B_T] = 0$. Thus, due to the free ligand buffering effect, we see that 1/3 of the DNA sites are not bound by ligand A. For the indefinite case, as $[A_T] \to \infty$, $f_{AD} \to 2/3$. For large enough finite aggregation termination stoichiometries ($N \geq \sim 30$),

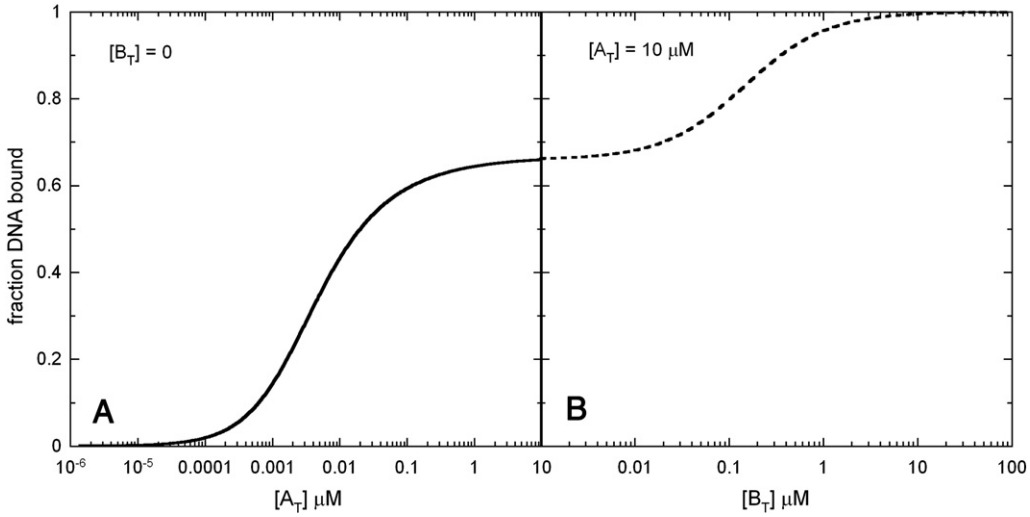


Fig. 4. Effect of free ligand buffering on competitive binding to a specific DNA sequence. The simulations were carried out as described in the text according to Scheme 5, with N = infinity, $K_A = 200\ \mu M^{-1}$, $K_B = 20\ \mu M^{-1}$, and $L = 100\ \mu M^{-1}$. In Panel A, $[B_T] = 0$, and the simulation was carried out as a function of $[A_T]$ up to 10 μM. In Panel B, the $[A_T]$ was kept fixed at 10 μM, while the simulation was carried out as a function of $[B_T]$. In Panel A, we see that a fraction of DNA remains unoccupied by ligand A, due to the free ligand buffering effect. In Panel B, we see that ligand B can bind the remaining DNA, even though ligand A is in a large excess and binds the DNA with 10 fold higher affinity.

the apparent plateau value of f_{AD} is well approximated by Eq. (24), and is stable over an increase in $[A_T]$ of greater than four orders of magnitude (see Figs. 1 and 2). The simulation in Fig. 4A was carried out up to $[A_T] = 10$ μM. Next, we continued this simulation by keeping $[A_T]$ constant at 10 μM, and increasing $[B_T]$ from 0 to 100 μM (Fig. 4B). This simulation shows that even though ligand A binds DNA with 10 fold stronger affinity than ligand B, and even though 10 μM total A is present to compete with ligand B for binding to the DNA, the remaining 1/3 of the DNA is titrated with ligand B with a saturation midpoint of $[B_T]_{0.5} = 0.15$ μM. Thus, since ligand A participates in a large aggregation process, this ensures a fraction of DNA sites will be apportioned for ligand B binding, even if A binds with higher affinity than B and is present in an excess over B. This provides the cell with a potential mechanism to apportion a fraction of receptor sites for a particular competitive ligand, regardless of the ratio of affinities for the competitors.

3. Discussion

In this work we have shown that failure to account for the linkage of receptor binding with large scale ligand aggregation processes can result in incorrect measurements of fundamental equilibrium thermodynamic parameters. In addition to this practical concern, the realization that large scale ligand aggregation buffers the free ligand concentration suggests novel control mechanisms to explain biological function. A fundamental observation is that if the ligand polymerization equilibrium constant rivals the intrinsic affinity of the ligand (or a finite number of its smaller aggregates) for its receptor, then even overexpression of the ligand will not completely saturate all the receptors in the cell. This provides the cell with a mechanism to apportion a fraction of the receptor sites to remain free of the ligand, which would have major implications for a variety of biological processes. For example, this consequence may play a role in the competitive binding of transcription factors, since it is often observed that transcriptional activators and repressors will compete for identical DNA binding sequences [17].

A specific example that is relevant to this discussion are the *Drosophila melanogaster* transcription factors Pnt-P2 (an activator), and Yan (a repressor) [9,10]. As members of the ETS family of transcription factors, they both bind the consensus DNA binding motif 5′-GGA(A/T)-3′ [9]. In addition, they both contain the so-called SAM domains (Sterile Alpha Motif). SAM domains are ubiquitous and often (but not always) confer a head-to-tail polymerization function to proteins [9]. The Yan protein has been shown to polymerize reversibly *in vitro* and also to polymerize *in vivo* [11,12], while the Pnt-P2 protein has been shown to be monomeric [18]. Thus, depending upon the relative affinities for DNA binding and polymerization (aggregation), a possible consequence of Yan polymerization is that it will not be able to saturate all DNA binding sites, even if the protein is overexpressed in the cell. This would imply that complete transcriptional repression would not be guaranteed by simply increasing levels of the repressor, unless the aggregates retain repressor activity. Thus, a fraction of the DNA binding sites would be available for other ETS family transcription factors, including the Pnt-P2 activator. In this way, some transcriptional activation could occur even in the presence of a large background concentration of the Yan repressor.

Another example where large scale aggregation of a sequence specific DNA binding protein may play a fundamental regulatory role is the Adenoviral L4-22 K protein [8]. Two viral proteins, L4-22 K and IVa2, bind to conserved sequences, called A repeats, in the viral genome, and these reactions are required for viral genome encapsulation [19–22]. IVa2 exists as a monomer in solution [23], while L4-22 K aggregates reversibly via an indefinite, isodesmic mechanism [8]. While by itself IVa2 can bind with high affinity to DNA [19,21,23], sequence specific DNA binding of L4-22 K requires IVa2 [19]. As pointed out by us previously [8], a consequence of the fact that the L4-22 K protein assembles indefinitely is that saturation of the A repeats is not guaranteed by solely increasing cellular levels of the L4-22 K protein.

In general viral systems must ensure the production of optimal levels of structural components for the construction of infectious viral particles, otherwise entry into the virus assembly stage of the lifecycle will result in mis-assembled, non-infectious viral particles [22,24–27]. Consequently, the entry into virus assembly must be tightly regulated. Since during the late phase of the viral lifecycle, the L4-22 K protein is expressed to high levels [28], and likely undergoes a large scale aggregation process [8], this implies that IVa2 levels will control fractional occupancy of the A repeats. This simplifies a multi-component molecular switch by assigning a single ligand, in this case the IVa2 protein monomer, to be the primary decision maker for entry into the virus assembly stage of the infection.

The observations presented here may have implications for a number of other large scale aggregating systems, such as human RAD52 [7], FOXL2 [29] and p-bodies [30], to name just a few. In

addition, it has been pointed out that drug self association equilibria must be accounted for to properly quantify drug–DNA binding interactions [31], and the framework provide here will be useful in this regard. In general, a careful study of the aggregation properties of all components in a particular system is necessary to properly deduce its equilibrium binding mechanism.

References

[1] E.Y. Chi, S. Krishnan, T.W. Randolph, J.F. Carpenter, Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation, Pharm. Res. 20 (2003) 1325–1336.
[2] D.L. Bain, A.F. Heneghan, K.D. Connaghan-Jones, M.T. Miura, Nuclear receptor structure: implications for function, Annu. Rev. Physiol. 69 (2007) 201–220.
[3] I. Wong, T.M. Lohman, Linkage of protein assembly to protein–DNA binding, Methods Enzymol. 259 (1995) 95–127.
[4] M. Takahashi, Analysis of DNA–RecA protein interactions involving the protein self-association reaction, J. Biol. Chem. 264 (1989) 288–295.
[5] C.A. Kim, M.L. Phillips, W. Kim, M. Gingery, H.H. Tran, M.A. Robinson, S. Faham, J.U. Bowie, Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression, EMBO J. 20 (2001) 4173–4182.
[6] R. Ramachander, C.A. Kim, M.L. Phillips, C.D. Mackereth, C.D. Thanos, L.P. McIntosh, J.U. Bowie, Oligomerization-dependent association of the SAM domains from Schizosaccharomyces pombe Byr2 and Ste4, J. Biol. Chem. 277 (2002) 39585–39593.
[7] E. Van Dyck, A.Z. Stasiak, A. Stasiak, S.C. West, Binding of double-strand breaks in DNA by human Rad52 protein, Nature 398 (1999) 728–731.
[8] T.C. Yang, N.K. Maluf, Self-association of the adenoviral L4-22 K protein, Biochemistry 49 (2010) 9830–9838.
[9] F. Qiao, J.U. Bowie, The many faces of SAM, Sci. STKE (2005), re7.
[10] F. Qiao, B. Harada, H. Song, J. Whitelegge, A.J. Courey, J.U. Bowie, Mae inhibits pointed-P2 transcriptional activity by blocking its MAPK docking site, EMBO J. 25 (2006) 70–79.
[11] F. Qiao, H. Song, C.A. Kim, M.R. Sawaya, J.B. Hunter, M. Gingery, I. Rebay, A.J. Courey, J.U. Bowie, Derepression by depolymerization; structural insights into the regulation of Yan by Mae, Cell 118 (2004) 163–173.
[12] J. Zhang, T.G. Graham, P. Vivekanand, L. Cote, M. Cetera, I. Rebay, Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the *Drosophila* ETS family transcriptional repressor Yan, Mol. Cell. Biol. 30 (2010) 1158–1170.
[13] T.M. Lohman, W. Bujalowski, Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein–DNA interactions: spectroscopic approaches to monitor binding, Methods Enzymol. 208 (1991) 258–290.
[14] M. Takahashi, C. Strazielle, J. Pouyet, M. Daune, Co-operativity value of DNA RecA protein interaction. Influence of the protein quaternary structure on the binding analysis, J. Mol. Biol. 189 (1986) 711–714.
[15] R.B. Martin, Comparisons of indefinite self-association models, Chem. Rev. 96 (1996) 3043–3064.
[16] F. Oosawa, M. Kasai, A theory of linear and helical aggregations of macromolecules, J. Mol. Biol. 4 (1962) 10–21.
[17] F.M. Rossi, A.M. Kringstein, A. Spicher, O.M. Guicherit, H.M. Blau, Transcriptional control: rheostat converted to on/off switch, Mol. Cell 6 (2000) 723–728.
[18] C.D. Mackereth, M. Scharpf, L.N. Gentile, S.E. MacIntosh, C.M. Slupsky, L.P. McIntosh, Diversity in structure and function of the Ets family PNT domains, J. Mol. Biol. 342 (2004) 1249–1264.
[19] S.G. Ewing, S.A. Byrd, J.B. Christensen, R.E. Tyler, M.J. Imperiale, Ternary complex formation on the adenovirus packaging sequence by the IVa2 and L4 22-kDa proteins, J. Virol. 81 (2007) 12450–12457.
[20] P. Ostapchuk, M.E. Anderson, S. Chandrasekhar, P. Hearing, The L4 22-kDa protein plays a role in packaging of the adenovirus genome, J. Virol. 80 (2006) 6973–6981.
[21] R.E. Tyler, S.G. Ewing, M.J. Imperiale, Formation of a multiple protein complex on the adenovirus packaging sequence by the IVa2 protein, J. Virol. 81 (2007) 3447–3454.
[22] W. Zhang, M.J. Imperiale, Requirement of the adenovirus IVa2 protein for virus assembly, J. Virol. 77 (2003) 3586–3594.
[23] T.C. Yang, Q. Yang, N.K. Maluf, Interaction of the adenoviral IVa2 protein with a truncated viral DNA packaging sequence, Biophys. Chem. 140 (2009) 78–90.
[24] H. Gaussier, Q. Yang, C.E. Catalano, Building a virus from scratch: assembly of an infectious virus using purified components in a rigorously defined biochemical assay system, J. Mol. Biol. 357 (2006) 1154–1166.
[25] P. Lei, S.T. Andreadis, Stoichiometric limitations in assembly of active recombinant retrovirus, Biotechnol. Bioeng. 90 (2005) 781–792.
[26] K.N. Parent, A. Zlotnick, C.M. Teschke, Quantitative analysis of multi-component spherical virus assembly: scaffolding protein contributes to the global stability of phage P22 procapsids, J. Mol. Biol. 359 (2006) 1097–1106.
[27] H. Tsukamoto, M.A. Kawano, T. Inoue, T. Enomoto, R.U. Takahashi, N. Yokoyama, N. Yamamoto, T. Imai, K. Kataoka, Y. Yamaguchi, H. Handa, Evidence that SV40 VP1–DNA interactions contribute to the assembly of 40-nm spherical viral particles, Genes Cells 12 (2007) 1267–1279.
[28] S.J. Morris, G.E. Scott, K.N. Leppard, Adenovirus late-phase infection is controlled by a novel L4 promoter, J. Virol. 84 (2010) 7096–7104.
[29] L. Moumne, A. Dipietromaria, F. Batista, A. Kocer, M. Fellous, E. Pailhoux, R.A. Veitia, Differential aggregation and functional impairment induced by polyalanine expansions in FOXL2, a transcription factor involved in cranio-facial and ovarian development, Hum. Mol. Genet. 17 (2008) 1010–1019.
[30] M. Kulkarni, S. Ozgur, G. Stoecklin, On track with P-bodies, Biochem. Soc. Trans. 38 (2010) 242–251.
[31] V.A. Bloomfield, D.M. Crothers, I. Tinoco, Nucleic Acids, Structures, Properties and Functions, University Science Books, Sausalito, 2000.