

ASSOCIATION KINETICS WITH COUPLED DIFFUSION. AN EXTENSION TO COILED-CHAIN MACROMOLECULES APPLIED TO THE LAC REPRESSOR–OPERATOR SYSTEM

Otto G. BERG and Clas BLOMBERG

Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received 6 January 1977

Revised manuscript received 14 March 1977

The association of a molecule onto a specified binding site on a large chain-like macromolecule is described in the “sliding” model, where the molecule is allowed to move along the chain in a one-dimensional diffusion which is coupled to the three-dimensional diffusion in solution. The present work extends a previous one by treating the chains more generally as coiled instead of straight. The model is applied to the lac repressor–operator association. A general expression for the rate of unspecific attachment to a chain-like macromolecule is also derived.

1. Introduction

The so-called “sliding” model has been used to describe the association of a molecule onto a specific site on a large chain-like macromolecule. In this model the molecule is allowed to associate unspecifically to any site on the chain, and then slide along the chain in a one-dimensional diffusion until it either finds its specific site or falls off, only to be reassociated to the chain somewhere else to try anew. This way of facilitating the search for a specified site can greatly enhance the association rate to an extent that is determined by the strength of the unspecific binding and the one-dimensional mobility. The model was first applied by Richter and Eigen [1] in particular to explain the high association rate of the lac repressor onto its specific site, the operator, on a large DNA molecule.

In a previous paper [2] we treated this model under the simplifying assumptions that the chains were straight, and distributed in a regular manner in solution. The influence on the result of these rather unsatisfactory assumptions was difficult to assess. In this paper it is shown that this influence is surprisingly small. By a slight revision of the previous calculations, the model has been extended to the case of coiled chains. This extension is important as it gives results of general validity;

also it shows how the methods of the very idealized system can be used to treat the complex case. Each chain is assumed to be coiled up in a limited and separate region. This is evidently the case for the low DNA concentrations that have been used for the determination of the association rate of the lac repressor and operator [3].

In the following we shall talk about repressor, operator and DNA chain, though the model is not limited to this case. Other DNA-binding proteins can be expected to behave in a similar manner, i.e. RNA polymerase, restriction endonucleases etc. In section 2 the particulars of the model are explained. The calculations rely necessarily on our previous work and are not explained in detail. The association rate to the specific site is derived and discussed in section 3 together with some properties of the unspecific association. In section 4 the general association rate onto a coiled macromolecule is determined with a simple method. Finally, the conclusions regarding the lac system are discussed in section 5.

2. The model

Introduce a coordinate z along the length $2L$ of the DNA chain, $-L < z < L$. Let the repressor move along

the chain with the diffusion constant D_1 when it is unspecifically bound, and dissociate from the chain with the rate λ per second. Assume that the operator is situated at $z = 0$ in the middle of the chain. Then, for symmetry reasons it suffices to consider only one half, $0 < z < L$, of the chain. Let $u(z, t)$ be the probability density of finding the repressor at the point z on the chain at time t . It was shown [2] that the diffusion equation along the chain can be expressed quite generally as

$$\partial u / \partial t = D_1 \partial^2 u / \partial z^2 - \lambda u + \lambda \int_0^L dz' \int_0^t dt' F(z, z', t-t') u(z', t') + G(z, t). \quad (1)$$

$F(z, z', t-t')$ expresses the probability of return to the chain at z for a repressor that dissociated from z' $t-t'$ seconds earlier. $G(z, t)$ is the first arrival of the repressor to the chain and will in general be independent of z . Eq. (1) is quite general, the particulars of the assumptions about the spatial organization of the chains will enter in the specific choices of F and G .

In contrast to our previous treatment we will now consider an inhomogeneous chain distribution. Assume each chain (length $2L$) coiled up into an approximately spherical region with the radius r_g . If there are n_0 chains per cm^3 , each can be considered alone and independent of the others in a volume $n_0^{-1} \text{ cm}^3$. (The usefulness and limitations of this "closed cell" approach are discussed in [2].) This volume can be taken as spherical with the radius R_0 , fig. 1.

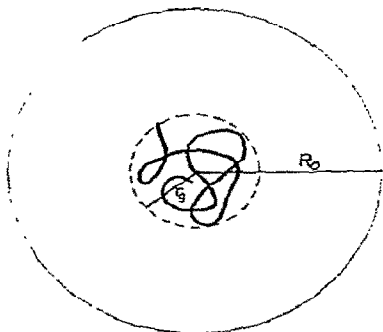


Fig. 1. The two spherical regions described in the text.

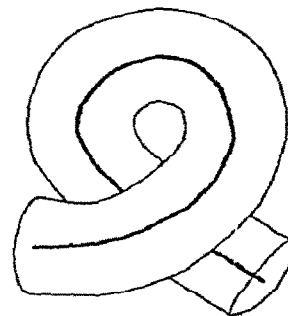


Fig. 2. The imaginary cylinder that follows the chain and fills the inner sphere of fig. 1.

$$\frac{4}{3} \pi R_0^3 = 1/n_0. \quad (2)$$

Construct a cylinder around the chain such that it fills the inner sphere, fig. 2. This cylinder will on the average have the radius R defined by

$$2\pi L R^2 = \frac{4}{3} \pi r_g^3. \quad (3)$$

In the internal z -coordinate the DNA chain can be pictured as straight. As DNA has a persistence length of 10^{-5} cm and no sharp turns, the cylinder of fig. 2 can also well be pictured as straight and circular with the DNA chain at its center, fig. 3. Let this chain have the radius b equal to the reaction radius for the encounter between the DNA and the repressor.

In ref. [2] a similar cylinder was considered as a

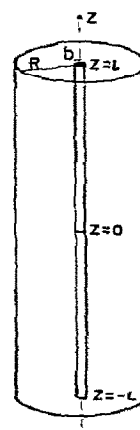


Fig. 3. The chain and the imaginary cylinder as pictured in the internal z -coordinate along the chain.

closed system which introduced artificial correlations along the chain. Here we shall treat the cylinder as open, and in particular take into account that a repressor diffusing across the cylinder wall will lose its correlations, and enter a region with an entirely different z -coordinate (fig. 2), or even leave the cylinder altogether.

The return probability F of eq. (1) can be calculated by considering the diffusion of the repressor in the cylinder just as was done previously [2]; only now we shall separate all cases where the repressor has crossed the cylinder wall and thereby lost its z -coordinate. This can be done by using an absorbing boundary condition at the wall and effectively removing all those repressors that cross. Particle conservation is achieved by reintroducing them with an unspecified z -coordinate; the function $\psi(r)$ below. The previous calculations [2] can be followed step by step after the replacement of eq. (3.11) with the absorbing boundary condition. Again the Laplace transformation simplifies the calculations. The Laplace transform $\tilde{f}(s)$ of a function $f(t)$ is defined as

$$\tilde{f}(s) = \int_0^{\infty} e^{-st} f(t) dt.$$

One finds for the return probability

$$\begin{aligned} \tilde{F}(z, z', s) = & (1/L)[\tilde{\varphi}(s) + \tilde{\psi}(s)] \\ & + (2/L) \sum_{n=1}^{\infty} \tilde{\varphi}(s + Dn^2\pi^2/L^2) \\ & \times \cos(n\pi z'/L) \cos(n\pi z/L), \end{aligned} \quad (4)$$

where $\tilde{\varphi}$ is an expression involving modified Bessel functions

$$\begin{aligned} \tilde{\varphi}(s) = & [kK_0(qR)I_0(qb) - kI_0(qR)K_0(qb)] \\ & \times \{K_0(qR)[kI_0(qb) - qI_1(qb)] \\ & - I_0(qR)[kK_0(qb) + qK_1(qb)]\}^{-1}, \end{aligned} \quad (5)$$

$$q \equiv (s/D)^{1/2}.$$

The parameter a of ref. [2] is taken $a = b$ without loss of generality. As before, D is the regular diffusion constant for the repressor in solution and k is a parameter such that $2\pi Dbk$ is the reaction rate per unit length of the chain for a homogeneous distribution of repressors

[cf. eq. (36)]. The undetermined function $\tilde{\psi}(s)$ is the return to the chain of those repressors that crossed the cylinder walls; they have to return sometime but without a specified z -coordinate.

When eq. (4) is entered into the Laplace transformed version of eq. (1), one finds the probability flux of the repressor into the operator

$$\begin{aligned} \tilde{\Phi}(s) = & D_1 (\partial u / \partial z)_{z=0} \\ = & L\tilde{G}(s) \left[1 + 2 \{ \lambda [1 - \tilde{\varphi}(s) - \tilde{\psi}(s)] + s \} \right. \\ & \times \left. \sum_{n=1}^{\infty} \{ D_1 n^2 \pi^2 / L^2 + s + \lambda [1 - \tilde{\varphi}(s + Dn^2 \pi^2 / L^2)] \}^{-1} \right]^{-1}. \end{aligned} \quad (6)$$

This expression differs from our previous result only in the explicit form of the functions \tilde{G} and $\tilde{\varphi}$, and in the insertion of the extra term $\tilde{\psi}$. In principle there remains to determine \tilde{G} and $\tilde{\psi}$, which must be quite complicated. As will be shown in the next section, they do not have to be known explicitly if we are only interested in the mean time of arrival.

3. The rate of association

The following relations hold

$$\tilde{\Phi}(0) = \lim_{s \rightarrow 0} \int_0^{\infty} e^{-st} \Phi(t) dt = \int_0^{\infty} \Phi(t) dt = 1, \quad (7)$$

$$\begin{aligned} \tilde{\Phi}'(0) = & \lim_{s \rightarrow 0} \frac{d\tilde{\Phi}}{ds} = \lim_{s \rightarrow 0} \frac{d}{ds} \int_0^{\infty} e^{-st} \Phi(t) dt \\ = & - \int_0^{\infty} t \Phi(t) dt = -\tau, \end{aligned} \quad (8)$$

where τ is the mean time of association. Eq. (7) holds under the assumption that there is one repressor per operator and implies that the repressor eventually will be bound to the operator. In the same way also

$$L\tilde{G}(0) = 1, \quad (9)$$

$$\tilde{\varphi}(0) + \tilde{\psi}(0) = 1. \quad (10)$$

Eq. (9) implies that the first unspecific association will certainly take place, and eq. (10) implies that the repressor will certainly return after a dissociation from the chain. The mean time of association is again found

to be a sum of mean times on the general form

$$\tau = -L\tilde{G}'(0) + N[-\tilde{\varphi}'(0) - \tilde{\psi}'(0)] + N/\lambda$$

$$\equiv \tau_1 + N\tau_2 + N\tau_3. \quad (11)$$

The number N can be interpreted as the mean number of unspecific dissociations before the final association to the operator.

$$N = \sum_{n=1}^{\infty} \frac{2\lambda}{D_1 n^2 \pi^2 / L^2 + \lambda[1 - \tilde{\varphi}(D n^2 \pi^2 / L^2)]}. \quad (12)$$

The mean time for the first unspecific association, $\tau_1 = -L\tilde{G}'(0)$, will be determined in the following section. In general it is much smaller than the other contributions to τ , and to a good approximation it can be neglected. $\tau_3 = 1/\lambda$ is the mean time the repressor stays unspecifically bound to the chain each time it is bound. $\tau_2 = -\tilde{\varphi}'(0) - \tilde{\psi}'(0)$ is the mean time the repressor is free in solution between two subsequent associations. It can be determined without explicit knowledge of $\tilde{\psi}$ in the following manner.

We showed previously [2] that the unspecific dissociation constant K_c per unit length of the chain can be expressed as

$$K_c \equiv n_0 \tau_2 / \tau_3 = \lambda / 4\pi L D b k \quad [\text{units cm}^{-3}], \quad (13)$$

where n_0 is the concentration of chains. The first part of eq. (13) is by definition; K_c/n_0 is the ratio of the probabilities of finding the repressor in solution and unspecifically bound. This must be equal to the ratio of the mean times the repressor spends in the respective ways.

The equilibrium constant K_c must, however, be independent of the spatial configuration of the chains. As $\tau_3 = 1/\lambda$ obviously is independent of chain configuration, also τ_2 has to be the same as our previous result, i.e.

$$\tau_2 = 1/4\pi L D b k n_0. \quad (14)$$

It is noteworthy that τ_2 is completely independent of the diffusion; by the definition of the parameter k , τ_2 is just the mean time of unspecific association in the reaction limited case, i.e. when the diffusion is infinitely fast [cf. eq. (36)]. This is in complete agreement with the discussion of the equilibrium constant, but perhaps a little unexpected in view of the complicated diffusion-association process that is actually involved.

Hence, the total mean time of eq. (11) is

$$\tau = N(1 + n_0/K_c)/4\pi L D b k n_0. \quad (15)$$

This expression is formally the same as our previous result; so is the expression for N of eq. (12). What is different is the expression for $\tilde{\varphi}$ of eq. (5), which enters the result through N . Also the artificial cylinder radius, R , is quite different. Under quite general conditions (see appendix) N can be approximated as

$$N \approx L \{ [b k \ln(R/b) + 1] \lambda / D_1 \}^{1/2}. \quad (16)$$

Hence, the total association rate is (expressed in cm^3/s)

$$k_a = \frac{1}{n_0 \tau} = \frac{4\pi D b k}{(1 + n_0/K_c) \{ [b k \ln(R/b) + 1] \lambda / D_1 \}^{1/2}} \quad (17)$$

The factor $\ln(R/b) + 1$ is a measure of the correlations along the chain. It takes a more general form if we introduce the volumes $V = 2\pi L R^2 = \frac{4}{3}\pi r_g^3$ which is the volume the coiled chain occupies in solution, and $v = 2\pi L b^2$ which is the real volume of the chain. Then the correlation factor can be expressed without the artificial cylinder radius R :

$$b k \ln(R/b) + 1 = \frac{1}{2} b k \ln(V/v) + 1.$$

If the repressor does not remember from which part of the chain it comes after each dissociation, the correlation factor will be just 1; this also corresponds to the case when the unspecific association is strongly reaction controlled, $b k \ln(V/v) \ll 1$. For a diffusion controlled unspecific association $b k \ln(V/v) \gg 1$.

In units that are more convenient for the experimental data the DNA concentration can be expressed as $D_1 = 2L n_0 \times 3 \times 10^{-8} \mu\text{g}/\text{cm}^3$, and the unspecific dissociation constant $K_{RD} = 2L K_c \times 3 \times 10^{-8} \mu\text{g}/\text{cm}^3$ (if DNA weighs $3 \times 10^{-8} \mu\text{g}$ per cm of its length). When the unspecific association is diffusion limited, eq. (17) gives

$$k_a = \frac{1}{1 + D_1/K_{RD}} \left[\frac{16\pi D D_1 \times 3 \times 10^{-8}}{K_{RD} \ln(V/v)} \right]^{1/2}.$$

The exact form of the association rate of eq. (17) is dependent on the approximations in the appendix. For the lac repressor-operator association they are justified, but for other systems the full expression for N might be needed.

4. The rate of unspecific association

Consider again fig. 1. Let the repressor start with equal probability at any point in the sphere $0 \leq r < R_0$ at time zero. Let $c(r, t)$ be the probability distribution for the repressor at later times. This will obey the diffusion equation

$$\partial c / \partial t = D \nabla^2 c - k_{\text{ass}} n c, \quad 0 \leq r < r_g; \quad (18)$$

$$= D \nabla^2 c, \quad r_g < r < R_0, \quad (19)$$

with the initial condition

$$c(r, 0) = c_0 = 3/4\pi R_0^3, \quad (20)$$

and the boundary conditions

$$(\partial c / \partial r)_{r=R_0} = 0, \quad (21)$$

$$(\partial c / \partial r)_{r=r_g+0} = (\partial c / \partial r)_{r=r_g-0},$$

$$c(r_g + 0, t) = c(r_g - 0, t) \quad (22)$$

In eq. (18) n is the concentration of the chain in the inner region, i.e. (if it is expressed in cm^{-3})

$$n = 3/4\pi r_g^3 \quad (23)$$

and k_{ass} is the unspecific association rate in this region. The boundary condition (21) is the "closed cell" approach discussed earlier [2], and (22) is just the condition of continuity. As we are only interested in the first arrival the dissociation from the chain does not have to be included. The flow of arrival is

$$\Phi(t) = - \int_0^{R_0} 4\pi r^2 (\partial c / \partial t) dr, \quad (24)$$

with the Laplace transform

$$\tilde{\Phi}(s) = - \int_0^{R_0} 4\pi r^2 [s\tilde{c}(r, s) - c_0] dr. \quad (25)$$

This is the same as $L\tilde{G}(s)$ of the previous sections. Hence, the mean time of the first arrival is

$$\tau_1 = -\tilde{\Phi}'(0) = \int_0^{R_0} 4\pi r^2 \tilde{c}(r, 0) dr. \quad (26)$$

Thus, only the limit $s \rightarrow 0$ is needed. Letting $s \rightarrow 0$ in the Laplace transform of the diffusion equations (18)

and (19) considerably simplifies their solution with the boundary conditions. One finds

$$\begin{aligned} \tau_1 = & \frac{1}{k_{\text{ass}} n} \left[2 - \left(\frac{r_g}{R_0} \right)^3 \right] \\ & + \frac{4\pi c_0}{9D} \frac{R_0^6}{r_g} \left[1 - \frac{9}{5} \frac{r_g}{R_0} + \left(\frac{r_g}{R_0} \right)^3 - \frac{1}{5} \left(\frac{r_g}{R_0} \right)^6 \right] \\ & + \frac{4\pi c_0}{9D} \frac{R_0^6}{r_g} \left[1 - \left(\frac{r_g}{R_0} \right)^3 \right]^2 \frac{1}{\mu r_g \coth(\mu r_g) - 1}, \end{aligned} \quad (27)$$

where μ is defined as

$$\mu \equiv (k_{\text{ass}} n / D)^{1/2}. \quad (28)$$

The expression for τ_1 does not appear very enticing but its limiting cases are simple and important.

k_{ass} has been derived for the homogeneous case [2] (eq. (4.11) of ref. [2] by mistake contains the diffusion limited approximation $bk \ln(V/v) \gg 1$, which was not introduced until eq. (5.1)):

$$k_{\text{ass}} = 4\pi L D b k / [\frac{1}{2} b k \ln(V/v) + 1]. \quad (29)$$

If $r_g \rightarrow R_0$ in eq. (27) one gets

$$\tau_1 = 1/k_{\text{ass}} n = 1/k_{\text{ass}} n_0, \quad (30)$$

which is the result for a homogeneous distribution of the chains. If the chains are strongly localized, $R_0 \gg r_g$, one gets

$$\tau_1 = \frac{1}{3D} \frac{R_0^3}{r_g} \left[1 + \frac{1}{\mu r_g \coth(\mu r_g) - 1} \right]. \quad (31)$$

For a diffusion limited reaction, $bk \ln(V/v) \gg 1$ and $(\mu r_g)^2 \approx L/r_g \gg 1$ and τ_1 of eq. (31)

$$\tau_1 \approx R_0^3 / 3D r_g = 1/4\pi D r_g n_0, \quad (32)$$

which is just the mean time to reach the inner region. To check the consistency of expression (31) we also investigate the strongly reaction limited case when $\mu r_g \ll 1$. Then

$$\tau_1 \approx \frac{R_0^3}{D r_g (\mu r_g)^2} = \frac{R_0^3 [\frac{1}{2} b k \ln(V/v) + 1]}{3D L b k} \approx \frac{R_0^3}{3D L b k}. \quad (33)$$

The total unspecific association rate for $r_g \ll R_0$ is

$$k_A \equiv \frac{1}{n_0 \tau_1} = 4\pi D r_g \left[1 + \frac{1}{\mu r_g \coth(\mu r_g) - 1} \right]^{-1}, \quad (34)$$

which in the diffusion limited case becomes

$$k_A = 4\pi D r_g, \quad (35)$$

and in the strongly reaction limited case

$$k_A = 4\pi D L b k. \quad (36)$$

Eq. (35) is the traditional von Smoluchowski result for the association rate onto a sphere and eq. (36) is also the expected result in view of the definition of the parameter k [2]. Eq. (34), however, contains the information about the spatial arrangement of the chain.

5. Discussion and application to the lac system

If the unspecific association is assumed to be diffusion limited, $b k \ln(V/v) \gg 1$, the same identification of the parameters for the lac repressor-operator association can be made as previously [2]. The only difference will be in the estimate of the one-dimensional diffusion constant D_1 , which will be a factor $\frac{1}{2} \ln(V/v) / [\ln(2L/\pi b) - 0.577]$ smaller. With the size of the coil $V \approx 0.5 \times 10^{-12} \text{ cm}^3$ [3] and the other data as before [2], it will be about half. This confirms our guess that the influence of a coiled chain would be small, but to what extent was not obvious. If this model is applied to other systems, however, the differences could be important. It is not the coiling per se that increases the association rate; rather it is the removal of the artificial correlations along the chain which were inherent in our previous treatment. Thus, this result is quite different from the transfer model as proposed by von Hippel et al. [3] which has coiling and chain movement as essential prerequisites for the association mechanism.

The present result clearly distinguishes the reaction limited and the diffusion limited cases. If the unspecific association is reaction limited, $b k \ln(V/v) \lesssim 1$, we will be left with an extra parameter, $b k$, which cannot be determined without actual knowledge of the rate of unspecific association. This limiting case could very well be dominant in the high concentrations of a cell.

Regarding the first unspecific association, the experimental conditions for the lac repressor-operator association leads to the dominance of the strongly diffusion limited case, eq. (32). This would give τ_1 of the order of one second, which is much smaller than the total mean time of association.

All calculations above have been made under the

tacit assumption that the coils themselves are fairly homogeneous. As a strongly localized coil by itself does not create any large differences, there is no reason to believe that inhomogeneities in the coil will be important unless they involve stationary structures. A DNA chain in solution has a gaussian density distribution which can quite well be approximated as a localized homogeneous density with a cut-off radius.

The trick of considering each DNA chain as independent of the others in a closed volume (i.e. the spherical region $0 \leq r < R_0$), and deriving the association rate from this case (the "closed cell" approach) was discussed at length in ref. [2]. The approach will be good, in principle, whenever the reaction can be described by a bimolecular reaction rate.

The result relies heavily on the fact that the unspecific equilibrium constant K_c is independent of the chain distribution. This must be so as there can be no inhomogeneity in the repressor concentration in equilibrium, and consequently the repressor cannot "feel" the chain configuration. Only if the chain is folded so tightly as to virtually seal off certain parts from association would there be differences.

Acknowledgement

This work has been supported by the Swedish Natural Science Research Council.

Appendix

$$N = \sum_{n=1}^{\infty} \frac{2\lambda}{D_1 n^2 \pi^2 / L^2 + \lambda [1 - \tilde{\varphi}(D n^2 \pi^2 / L^2)]},$$

where

$$\begin{aligned} \tilde{\varphi}(D n^2 \pi^2 / L^2) = & [K_0(n\alpha)I_0(n\beta) - I_0(n\alpha)K_0(n\beta)] \\ & \times \{K_0(n\alpha)[I_0(n\beta) - (n\beta/bk)I_1(n\beta)] \\ & - I_0(n\alpha)[K_0(n\beta) + (n\beta/bk)K_1(n\beta)]\}^{-1}, \end{aligned}$$

$$\alpha \equiv \pi R / L, \quad \beta \equiv \pi b / L.$$

If $b \ll R \ll L$ and terms of order $(n\beta)^2 = (n\pi b/L)^2$ and higher are excluded in the expansion of the Bessel func-

tions, one finds

$$1 - \tilde{\varphi} \approx \{bk \ln(R/b) + 1\}$$

$$- bk [K_0(n\alpha)/I_0(n\alpha) + \ln(n\alpha/2) + \gamma]^{-1},$$

where $\gamma = 0.577\dots$ is Euler's constant. The expression in square brackets can also be written as

$$[2/I_0(n\alpha)] \sum_{l=1}^{\infty} I_{2l}(n\alpha)/l$$

and is of the order $(n\alpha)^2 = (n\pi R/L)^2$. As long as the sum is dominated by the terms for which $n < L/R$ it can be expressed as (with $1 - \tilde{\varphi} \approx \{bk \ln(R/b) + 1\}^{-1}$):

$$N \approx \frac{2\lambda}{D_1} \left(\frac{L}{\pi}\right)^2 \sum_{n=1}^{\infty} \frac{1}{n^2 + \lambda L^2/D_1 \pi^2 \{bk \ln(R/b) + 1\}}$$

$$= L \{[bk \ln(R/b) + 1] \lambda/D_1\}^{1/2}$$

$$\times \coth \{\lambda L^2/D_1 [bk \ln(R/b) + 1]\}^{1/2}$$

$$- bk \ln(R/b) + 1 \approx L \{[bk \ln(R/b) + 1] \lambda/D_1\}^{1/2}.$$

The last approximation is under the assumption that $N \gg 1$. $1 - \tilde{\varphi}$ is a slowly varying function of n . For larger n it goes over into the previously used approximation [2]:

$$1 - \tilde{\varphi} \approx \{bk [\ln(2L/\pi b) - \gamma - \ln(n)] + 1\}^{-1}.$$

References

- [1] P.H. Richter and M. Eigen, *Biophys. Chem.* 2 (1974) 255.
- [2] O.G. Berg and C. Blomberg, *Biophys. Chem.* 4 (1976) 367.
- [3] P.H. von Hippel, A. Revzin, C.A. Gross and A.C. Wang, in: *Protein-ligand interactions*, eds. H. Sund and G. Blauer (de Gruyter, Berlin, 1975) p. 270.