analytical chemistry

Article

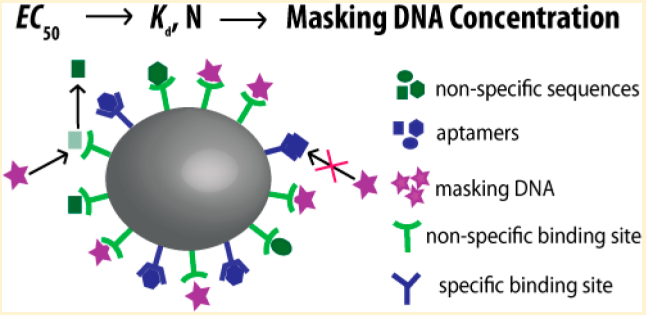
pubs.acs.org/ac

Theoretical Modeling of Masking DNA Application in Aptamer-Facilitated Biomarker Discovery

Leonid T. Cherney, Natalia M. Obrecht, and Sergey N. Krylov*

Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada

ABSTRACT: In aptamer-facilitated biomarker discovery (AptaBiD), aptamers are selected from a library of random DNA (or RNA) sequences for their ability to specifically bind cell-surface biomarkers. The library is incubated with intact cells, and cell-bound DNA molecules are separated from those unbound and amplified by the polymerase chain reaction (PCR). The partitioning/amplification cycle is repeated multiple times while alternating target cells and control cells. Efficient aptamer selection in AptaBiD relies on the inclusion of masking DNA within the cell and library mixture. Masking DNA lacks primer regions for PCR amplification and is typically taken in excess to the library. The role of masking DNA within the selection mixture is to outcompete any nonspecific binding sequences within the initial library, thus allowing specific DNA sequences (i.e., aptamers) to be selected more efficiently. Efficient AptaBiD requires an optimum ratio of masking DNA to library DNA, at which aptamers still bind specific binding sites but nonaptamers within the library do not bind nonspecific binding sites. Here, we have developed a mathematical model that describes the binding processes taking place within the equilibrium mixture of masking DNA, library DNA, and target cells. An obtained mathematical solution allows one to estimate the concentration of masking DNA that is required to outcompete the library DNA at a desirable ratio of bound masking DNA to bound library DNA. The required concentration depends on concentrations of the library and cells as well as on unknown cell characteristics. These characteristics include the concentration of total binding sites on the cell surface, N, and equilibrium dissociation constants, K_{nsL} and K_{nsM}, for nonspecific binding of the library DNA and masking DNA, respectively. We developed a theory that allows the determination of N, K_{nsL}, and K_{nsM} based on measurements of EC_{50} values for cells mixed separately with the library and masking DNA (EC_{50} is the concentration of fluorescently labeled DNA at which half of the maximum fluorescence signal from DNA-bound cells is reached). We also obtained expressions for signals from bound DNA (measured by flow cytometry) in terms of N, K_{nsL}, and K_{nsM}. These expressions can be used for the verification of N, K_{nsL}, and K_{nsM} values found from EC_{50} measurements. The developed procedure was applied to MCF-7 breast cancer cells, and corresponding values of N, K_{nsL}, and K_{nsM} were established for the first time. The concentration of masking DNA required for AptaBiD with MCF-7 breast cancer cells was also estimated.



Cell biomarkers are molecules with distinct characteristics, such as their state, quantity, or localization both on and within cells, which are specific to cell type. Changes in biomarker characteristics can serve as quantitative signs associated with cellular state. As a result, they can be used in disease identification and treatment.[1] The discovery of biomarkers is a slow and difficult process. Conventional approaches in biomarker discovery include Western blotting, mRNA screening using hybridization arrays or quantitative polymerase chain reaction (PCR), and two-dimensional gel electrophoresis in combination with mass spectrometry.[2−4] Unfortunately, these techniques are labor intensive and often produce false positive or false negative results.[5−9] To overcome these limitations, our group developed the technology of aptamer-facilitated biomarker discovery (AptaBiD).[10] Aptamers are short oligonucleotide sequences that can be selected to bind with high affinity and specificity to a variety targets. Aptamers are typically selected through the systemic evolution of ligands by exponential enrichment (SELEX).[11,12] Traditionally, they have been selected for purified targets including proteins and small molecules,[13−15] but recent developments have led to the ability to select aptamers for more complex targets including whole living cells.[16−18] Several issues arise when whole cells are used as the targets of aptamer selection mainly related to nonspecific oligonucleotide binding to the diverse cell surface. Dead cells present in suspension have also been shown to nonspecifically take up a large amount of oligonucleotide sequences resulting in a loss of potential aptamers in solution.[19]

The use of masking DNA has been suggested to meet the challenges commonly seen with traditional cell SELEX methods.[18,20] Masking DNA is an oligonucleotide sequence typically identical in length to the random region of the

Received: February 4, 2013
Accepted: March 12, 2013
Published: March 12, 2013

dx.doi.org/10.1021/ac400385v | *Anal. Chem.* 2013, 85, 4157−4164

selection library without primer to avoid amplification during SELEX. It is selected at random, contains an approximately 1:1 CG to AT ratio (C, G, A, and T stand for cytidine, guanosine, adenosine, and thymidine), and cannot have significant binding to library primer regions to eliminate amplification during the PCR stage of cell SELEX. The addition of masking DNA along with the DNA library during the aptamer selection process leads to increased efficiency, sensitivity, stability, and stringency.[10,18,20] Masking DNA present within the SELEX suspension will compete with low-affinity binders and nonspecific binders from the library for their ability to bind cell-surface targets. By doing this, it will allow high-affinity sequences from the library to bind to their unique targets resulting in increased stringency in the selection process. The increased stringency, in turn, leads to improved efficiency of selection by decreasing the number of SELEX rounds needed to select high-affinity aptamers. Masking DNA also improves the efficiency of selection through improving the efficiency of PCR amplification. Indeed, in the presence of masking DNA, the aptamer pool will be less heterogeneous while the masking DNA bound to the target cells will not be amplified due to lack of primer regions. In addition, masking DNA in solution will act as a substrate for any nuclease activity present within the suspension and will, thus, decrease the number of library sequences being digested.

To effectively use masking DNA in AptaBiD, one needs to determine the concentration of masking DNA required to outcompete nonspecific library DNA sequences by a desirable number of times. Generally speaking, appropriate concentrations of masking DNA could be determined experimentally by testing a range of masking DNA concentrations for each new experimental setup. In this way, DNA-cell association can be controlled and quantified using flow cytometry.[21−23] Unfortunately, such an approach would be very time-consuming since experimental conditions depend on many parameters (e.g., the concentrations of library DNA and cells, K_d values for binding of masking DNA and library to the cell surface, and cell characteristics). Here, K_d is the equilibrium dissociation constant for the interaction of DNA with the cell surface. Therefore, it is essential to develop a theoretical model for the estimation of masking DNA concentrations required to outcompete library DNA under any given experimental condition.

Analysis of oligonucleotide's binding to the cell surface is a complex multifaceted problem that depends on the properties of biomolecules localized on the cell surface which act as potential DNA binding sites.[13−18,24] Characteristics of such sites and thermodynamics of DNA binding reactions can be studied using biochemical and structural methods and calorimetric measurements.[25−30] By their nature, such studies are very specific, require significant efforts and, therefore, cannot be employed universally in masking DNA application to AptaBiD. Theoretical works on this issue are scarce and, in our view, use overly sophisticated mathematics and excessively detailed process descriptions.[31−34] As a result, they require a large amount of experimental data that is not available in AptaBiD applications.

In this study, we develop a simple mathematical model describing simultaneous binding of masking DNA and library DNA (including possible aptamers among the latter) to the cell surface. This model allows one to estimate the concentration of masking DNA that is required to outcompete library DNA by any given number of times, κ, in binding to the cell surface. The found expression for masking DNA concentration uses two types of experimental data. Data of the first type (initial concentrations of the library DNA and cells and the cell surface area) can be readily obtained from the experimental setup. Data of the second type (K_d values for binding of masking DNA and library DNA to the cell surface and the concentration of nonspecific binding sites at the cell surface) is calculated on the basis of a series of EC_{50} measurements using flow cytometry. For this purpose, we developed an additional mathematical procedure. A value of EC_{50} is defined as the concentration of fluorescently labeled DNA at which half of the maximum fluorescence signal from cells binding such DNA is reached. We applied our model to the MCF-7 breast cancer cell line.[35] In particular, we determined K_d values (for masking DNA and library DNA) and surface concentration of binding sites for these cells. We also determined the concentration of masking DNA required to outcompete library DNA's binding to cells' binding sites under varying conditions.

RESULTS AND DISCUSSION

Basic Equations. Since stoichiometry of DNA binding to the cell surface is usually unknown, we have to use a simplified approach to building a mathematical model. We assume that specific and nonspecific binding of DNA can be described as DNA adsorption to the cell surface. Accordingly, the surface presumably contains two types of binding sites: specific sites, which can provide high-affinity binding of aptamers, and nonspecific binding sites responsible for low-affinity binding of both the library DNA and masking DNA. Generally speaking, specific binding sites can also bind nonaptamers from the library and masking DNA but with low affinity. The two types of binding sites should be considered as some effective sites that describe an average behavior of real binding sites. The latter, of course, can be of many types with different stoichiometry of DNA binding to them. In such a simplified approach, basic equations describing DNA-cell binding in the state of equilibrium have the following form:

$$A(N_s - l_s - m_s) = K_{sA}a_s \quad (1)$$

$$(L - A)(N_s - l_s - m_s) = K_{nsL}(l_s - a_s) \quad (2)$$

$$M(N_s - l_s - m_s) = K_{nsM}m_s \quad (3)$$

$$L(N_{ns} - l_{ns} - m_{ns}) = K_{nsL}l_{ns} \quad (4)$$

$$M(N_{ns} - l_{ns} - m_{ns}) = K_{nsM}m_{ns} \quad (5)$$

Here, A, L, and M are, respectively, the volume concentrations of aptamers, library DNA, and masking DNA (present in the unbound state in the mixture containing cells, library DNA, and masking DNA); a_s, l_s, and m_s are, respectively, the surface concentrations of aptamers, library DNA, and masking DNA (bound to the specific binding sites on the cell surface); l_{ns} and m_{ns} are the surface concentrations of library DNA and masking DNA (bound to the nonspecific binding sites on the cell surface); N_s and N_{ns} are the surface concentrations of specific and nonspecific binding sites on the cell surface; K_{sA} is the equilibrium dissociation constant for high-affinity binding of aptamers to specific binding sites; K_{nsL} and K_{nsM} are the equilibrium dissociation constants for nonspecific binding of nonaptamers from the library and masking DNA, respectively. The term "library DNA" describes both aptamers and nonaptamers. Surface concentrations are measured in mols

dx.doi.org/10.1021/ac400385v | Anal. Chem. 2013, 85, 4157−4164

per unit area. We assume for simplicity that all aptamers bind with high affinity and the same value of K_{sA} to the specific binding sites and that all nonaptamers from the library bind with low affinity and approximately the same value of K_{nsL} to both specific and nonspecific binding sites. Similarly, masking DNA binds with low affinity and approximately the same value of K_{nsM} to all binding sites. Such assumptions follow from fundamental differences between DNA aptamers and non-aptamer DNA (i.e., nonaptamers from the library and masking DNA). Aptamers are considered to be DNA molecules that can bind specific sites with high affinity. The nonaptamer DNA binds with low affinity and low selectivity a great variety of sites. It follows from definitions of two types of binding sites that $K_{sA} \ll K_{nsL}$ and $K_{sA} \ll K_{nsM}$.

Equation 1 describes the high-affinity aptamer binding to the specific binding sites. Equations 2 and 3 govern the low-affinity binding of nonaptamer DNA from the library and masking DNA to the specific binding sites. Finally, eqs 4 and 5 represent binding laws for low-affinity binding of library DNA and masking DNA to the nonspecific binding sites. In eqs 1−3, we also take into account the binding to specific sites of the following DNA: (i) all nonaptamers from the DNA library and (ii) masking DNA. Indeed, the concentrations of these DNA can be much higher than that of aptamers. As a result, nonaptamers from the DNA library and masking DNA can be bound to specific binding sites in amounts comparable to aptamers bound there even though these types of DNA bind to such sites with a much lower affinity than aptamers.

Equations 1−5 should be supplemented by relations resulting from conservation of all three types of DNA during the mixture equilibration:

$$A + nS(a_s + a_{ns}) = A_0 \tag{6}$$

$$L + nS(l_s + l_{ns}) = L_0 \tag{7}$$

$$M + nS(m_s + m_{ns}) = M_0 \tag{8}$$

Here, n is the volume concentration of cells measured in number of cells per liter, S is the area of the cell surface, A_0, L_0, and M_0 are, respectively, the volume concentrations of the aptamers, the library DNA, and the masking DNA in the initial mixture before equilibration (i.e., before binding starts). If units M = mol/L were used for n, we should additionally multiply n by the Avogadro's number N_A in all relations starting from eq 6. Equations 1−8 allow one to determine the volume and surface concentrations of aptamers, library DNA, and masking DNA in the state of equilibrium if all other parameters (A_0, L_0, M_0, K_{sA}, K_{nsL}, K_{nsM}, N_s, and N_{ns}) are known. Indeed, values of L_0 and M_0 are usually known since they are defined by the experimental conditions. Values of K_{sA}, K_{nsL}, K_{nsM}, N_s, and N_{ns} can be determined from measurements of EC_{50} for aptamers, library DNA, and masking DNA at several different values of cell concentrations. However, the initial concentration of aptamers in the library, A_0, remains unknown until these aptamers are actually selected and identified.

The full system, eqs 1−8, is too complicated because of the presence of unknown aptamer characteristics. However, for the purpose of studying competition between masking DNA and library DNA's binding to nonspecific binding sites, we can simplify the system, eqs 1−8. Since aptamers usually constitute only a small part of the library, we conclude that $A \ll L$ and $a_s \ll l_s + l_{ns}$. By taking into account these inequalities and by adding eq 2 to eq 4 and eq 3 to eq 5, we obtain the following approximate relations:

$$L(N - l - m) = K_{nsL}l, \quad M(N - l - m) = K_{nsM}m \tag{9}$$

Here, N, l, and m are the total concentrations of binding sites, library DNA, and masking DNA on the cell surface, respectively. These concentrations are defined as follows:

$$N = N_s + N_{ns}, \quad l = l_s + l_{ns}, \quad m = m_s + m_{ns} \tag{10}$$

Equations 9 should be supplemented by relations 7 and 8 that can be rewritten in the form

$$L + nSl = L_0, \quad M + nSm = M_0 \tag{11}$$

Equations 9 and 11 allow one to study competition between library DNA and masking DNA for nonspecific binding to both specific and nonspecific binding sites. If condition $a_s \ll l_s + l_{ns}$ is not satisfied, eqs 9 can still be used to study competition between masking DNA and hypothetical library that does not contain aptamers. If masking DNA can outcompete such a library, then it should also outcompete the real library (for nonspecific binding) since high-affinity binding of aptamers impede nonspecific binding to specific binding sites in the same way for both nonaptamers in the DNA library and masking DNA.

Equations 9 and 11 can be simplified further by introducing the following dimensionless variables:

$$L^* = \frac{L}{nSN}, \quad L_0^* = \frac{L_0}{nSN}, \quad l^* = \frac{l}{N}, \quad \lambda = \frac{K_{nsL}}{nSN} \tag{12}$$

$$M^* = \frac{M}{nSN}, \quad M_0^* = \frac{M_0}{nSN}, \quad m^* = \frac{m}{N}, \quad \mu = \frac{K_{nsM}}{nSN} \tag{13}$$

With these variables, eqs 9 and 11 take a form

$$L^*(1 - l^* - m^*) = \lambda l^*, \quad M^*(1 - l^* - m^*) = \mu m^* \tag{14}$$

$$L^* + l^* = L_0^*, \quad M^* + m^* = M_0^* \tag{15}$$

Mathematical Solution and Estimation of the Required Concentration of Masking DNA. Equations 14 and 15 represent a nonlinear system of four algebraic equations. This system does not have a simple analytical solution. However, studying the competition between masking DNA and library DNA does not require a full solution of eqs 14 and 15. Let us say, we want to find the initial concentration of masking DNA that allows it to outcompete library DNA by κ times in binding to cells:

$$m = \kappa l, \quad m^* = \kappa l^* \tag{16}$$

Using relations 15 and 16 to exclude unknown variables L^*, M^*, and m^* from eqs 14, we can rewrite eqs 14 in the form:

$$(L_0^* - l^*)(1 - l^* - \kappa l^*) = \lambda l^* \tag{17}$$

$$(M_0^* - \kappa l^*)(1 - l^* - \kappa l^*) = \kappa\mu l^* \tag{18}$$

Excluding $(1 - l^* - \kappa l^*)$ from these relations, we have

$$\kappa\mu(L_0^* - l^*) = \lambda(M_0^* - \kappa l^*) \tag{19}$$

Solving eq 17 with respect to l^* and eq 19 with respect to M_0^*, we obtain the following expressions:

dx.doi.org/10.1021/ac400385v | Anal. Chem. 2013, 85, 4157−4164

$$l^* = \frac{1}{2}\left(L_0^* + \frac{1+\lambda}{1+\kappa}\right) - \sqrt{\frac{1}{4}\left(L_0^* + \frac{1+\lambda}{1+\kappa}\right)^2 - \frac{L_0^*}{1+\kappa}} \quad (20)$$

$$M_0^* = \frac{\kappa\mu L_0^*}{\lambda} + \kappa l^*\left(1 - \frac{\mu}{\lambda}\right) \quad (21)$$

Our choice of a minus sign before the radical in relation 20 follows from a physical condition $l^* = 0$ at $L_0^* = 0$. Finally, the substitution of expression 20 for l^* into eq 21 gives an expression for M_0^* that we sought:

$$M_0^* = \frac{\kappa\mu L_0^*}{\lambda} + \kappa\left(1 - \frac{\mu}{\lambda}\right)\left[\frac{1}{2}\left(L_0^* + \frac{1+\lambda}{1+\kappa}\right) - \sqrt{\frac{1}{4}\left(L_0^* + \frac{1+\lambda}{1+\kappa}\right)^2 - \frac{L_0^*}{1+\kappa}}\right] \quad (22)$$

According to definition 16, κ shows by how many times masking DNA outcompetes library DNA in binding to cells. Usually, we are interested in a significant outcompeting, which means $\kappa \gg 1$. In this case, the expression in square brackets in 22 can be expanded in a small parameter $1/\kappa$ (if $L_0^* \gg 1/\kappa$). As a result, we obtain a simpler approximate expression for M_0^*:

$$M_0^* = \frac{\kappa\mu L_0^*}{\lambda} + \left(1 - \frac{\mu}{\lambda}\right)\frac{\kappa L_0^*}{\lambda + \kappa L_0^*}, \quad \kappa >> 1, \quad L_0^* >> \frac{1}{\kappa} \quad (23)$$

Figures 1 and 2 show exact (solid lines) and approximate (dashed lines) dependencies $M_0^*(L_0^*)$ determined by expressions 22 and 23, respectively, at two values for each of the parameters κ and μ/λ ($\kappa = 2$ and 10, $\mu/\lambda = 0.1$ and 10) and at various values of parameter λ (0.1, 1, 5, 10, and 20). It is obvious from Figure 1 that approximate expression 23 can be used even at a moderate value of outcompeting ratio $\kappa = 2$ if $\lambda > 10$. Figure 2 shows that approximations based on relation 23 significantly improve at larger $\kappa = 10$. In general, this approximation works better when $\mu/\lambda \gg 1$ rather than $\mu/\lambda \ll 1$ (Figures 1 and 2). However, approximate expression 23 does not work at all at small values of L_0^* (it gives negative values for M_0^*) if $\mu/\lambda = 10$ and $\lambda = 0.1$ (Figures 1 and 2). This is not surprising since the last inequality 23 is not satisfied in this case.

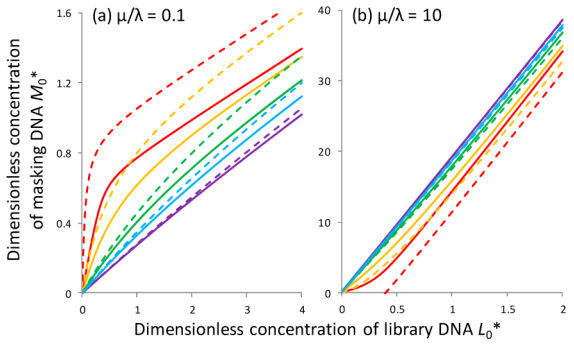


Figure 1. Dependences $M_0^*(L_0^*)$ at $\kappa = 2$ and various values of ratio μ/λ and parameter λ. Solid and dashed lines show dependences obtained from relations 22 and 23, respectively. Red, yellow, green, blue, and magenta lines correspond to $\lambda = 0.1$, 1, 5, 10, and 20, respectively.

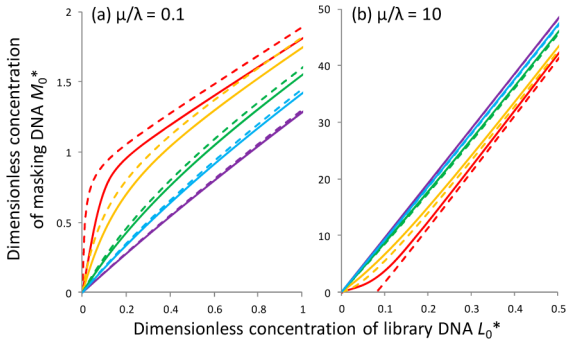


Figure 2. Dependences $M_0^*(L_0^*)$ at $\kappa = 10$ and various values of ratio μ/λ and parameter λ. Solid and dashed lines show dependences obtained from relations 22 and 23, respectively. The color coding is the same as in Figure 1.

Given definitions 12 and 13 for L_0^*, λ, M_0^*, and μ, relations 22 and 23 determine the concentration M_0 of masking DNA that allows it to outcompete the library DNA by a factor of κ at given values of L_0, n, S, N, K_{nsL}, and K_{nsM}. These relations can be rewritten in the dimensional form as follows:

$$M_0 = \frac{\kappa K_{nsM} L_0}{K_{nsL}} + \kappa\left(1 - \frac{K_{nsM}}{K_{nsL}}\right) \times \left[\frac{1}{2}\left(L_0 + \frac{nSN + K_{nsL}}{1+\kappa}\right) - \sqrt{\frac{1}{4}\left(L_0 + \frac{nSN + K_{nsL}}{1+\kappa}\right)^2 - \frac{nSNL_0}{1+\kappa}}\right] \quad (24)$$

$$M_0 = \frac{\kappa K_{nsM} L_0}{K_{nsL}} + \left(1 - \frac{K_{nsM}}{K_{nsL}}\right)\frac{\kappa nSNL_0}{K_{nsL} + \kappa L_0}, \quad \kappa >> 1, \quad \kappa L_0 >> nSN \quad (25)$$

If equilibrium dissociation constants for the binding of masking DNA and library DNA (to nonspecific binding sites) are identical, then both dependencies 24 and 25 degenerate to the following obvious relation:

$$M_0 = \kappa L_0 \quad (K_{nsM} = K_{nsL}) \quad (26)$$

Parameters L_0, n, and S are usually known in AptaBiD experiments. A value of κ is defined by a degree of desirable outcompeting. The remaining parameters N, K_{nsL}, and K_{nsM} can be determined from measurements of EC_{50} for masking DNA and library DNA as shown below.

Determination of Parameters N, K_{nsL}, and K_{nsM}. For a mixture that contains only masking DNA and cells, the second eq 9 gives (at $l = 0$):

$$M(N - m) = K_{nsM} m \quad (27)$$

By definition of EC_{50}, its value for masking DNA, EC_{50M}, is equal to the initial concentration M_0 at which the bound masking DNA molecules occupy a half of all binding sites:

$$m = \frac{1}{2}N, \quad M_0 = EC_{50M} \quad (28)$$

The substitution of these expressions into eqs 11 and 27 gives:

dx.doi.org/10.1021/ac400385v | Anal. Chem. 2013, 85, 4157−4164

$$M + \frac{1}{2}nSN = EC_{50M}, \quad M = K_{nsM} \tag{29}$$

Excluding M from relations 29, we finally obtain the first relation of the next two:

$$EC_{50M} = \frac{1}{2}nSN + K_{nsM}, \quad EC_{50L} = \frac{1}{2}nSN + K_{nsL} \tag{30}$$

Here, EC_{50L} is the EC_{50} value for library DNA. The second relation in 30 is derived similarly for a mixture that contains only library DNA and cells. It can be done simply by replacing M_0, M, m, K_{nsM}, and EC_{50M} with L_0, L, l, K_{nsL}, and EC_{50L}, respectively, in relations 27−29 and by excluding L from modified relations 29. Obviously, unknown values of N, K_{nsM}, and K_{nsL} can be determined by fitting linear plots 30 for $EC_{50M}(n)$ and $EC_{50L}(n)$ at various N, K_{nsM}, and K_{nsL} into experimentally measured dependences EC_{50M} and EC_{50L} on the cell concentration n.

Solving a system of eqs 11 and 27 with respect to the surface concentration m of bound DNA, we have:

$$m = N\left(\frac{1}{2}(1 + \mu + M_0^*) - \sqrt{\frac{1}{4}(1 + \mu + M_0^*)^2 - M_0^*}\right) \tag{31}$$

The minus sign before the square root was chosen on the basis of a physical condition that $m = 0$ when $M_0^* = 0$. At small and large values of the dimensionless concentration $M_0^* = M_0/(nSN)$, expression 31 can be simplified by expansions in M_0^* and $1/M_0^*$, respectively. As a result, we obtain:

$$m = \alpha_M M_0, \quad \alpha_M \equiv \frac{N}{K_{nsM} + nSN} \quad (M_0 \ll nSN) \tag{32}$$

$$m = \frac{NM_0}{K_{nsM} + M_0}, \quad \max_{M_0}(m) = N \quad (M_0 \gg nSN) \tag{33}$$

If a signal $F(M_0)$ from DNA bound to the cells is measured at some initial volume concentration M_0 of masking DNA, then such a signal is usually proportional to the surface concentration m:

$$F(M_0) = gm(M_0) \tag{34}$$

When signals are measured from isolated cells, the coefficient g does not depend on the cell concentration n. Nevertheless, the signal itself depends on n since the function $m(M_0)$ also depends on n (through μ and M_0^* according to eq 31). Using relations 32−34, we can conclude that the slope $\alpha_{FM} = g\alpha_M$ of the dependence $F(M_0)$ at small values of M_0 is related to the saturated value of signal $F_{max,M} = g\max(m)$ reached at large values of M_0 by the first of the following two expressions:

$$\frac{\alpha_{FM}}{F_{max,M}} = \frac{1}{K_{nsM} + nSN}, \quad \frac{\alpha_{FL}}{F_{max,L}} = \frac{1}{K_{nsL} + nSN} \tag{35}$$

The second relation in 35 can be derived similarly for the slope α_{FL} of the signal and its saturated value $F_{max,L}$ corresponding to library DNA bound to the cells (in the absence of masking DNA). Relations 35 can be used to verify values of N, K_{nsM}, and K_{nsL} found from EC_{50} measurements.

Application to MCF-7 Breast Cancer Cells. We tested the developed approach on the MCF-7 breast cancer cell line. EC_{50} values for masking DNA and library DNA (i.e., EC_{50M} and EC_{50L}) were measured separately for masking DNA and library DNA bound to MCF-7 cells. The measurements were performed at various cell concentration values of n = 1, 2, 3, 4, and 5 nL^{-1}. Figure 3 demonstrates the obtained experimental

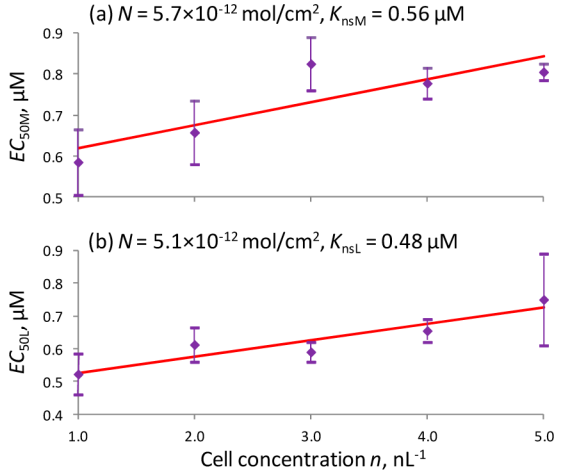


Figure 3. Dependencies of EC_{50} on the cell concentration n for masking DNA (a) and library DNA (b). Experimental data (mean values with error bars) are shown in magenta color. The best fits of theoretical dependencies $EC_{50M}(n)$ and $EC_{50L}(n)$ determined by relations 30 are depicted by red lines. The shown values of N, K_{nsM}, and K_{nsL} correspond to the best fits.

dependencies of $EC_{50M}(n)$ and $EC_{50L}(n)$ (the magenta lines) and the best fit of the linear plots of eqs 30 (the red lines) into these experimental data. The cell surface area S present in relations 30 was estimated as $S = 2 \times 10^{-5}$ cm^2 based on a measured value of the average cell diameter of approximately $d = 2.5 \times 10^{-3}$ cm (see Materials and Methods). Though the magenta and red lines look significantly deviated from each other, such an effect is mainly a result of magnification along the vertical axis. If the vertical axis starts from the point of $EC_{50} = 0$, the magenta and red lines will appear reasonably close. A quality of the best fit can be estimated by normalized root mean square deviation (NRMSD).[36] We have NRMSD = 5.9% and 3.6% for Figure 3a,b, respectively. The best-fit lines were obtained by varying unknown parameters N, K_{nsM}, and K_{nsL}. Their determined values are shown in Figure 3. Values of N found from measurements EC_{50M} and EC_{50L} were determined to be fairly close to each other. This fact confirms that our simple model of nonspecific binding to the same binding sites for both masking and library DNAs is reasonable. We can estimate the cell surface area S_{ns} per one nonspecific biding site as $S_{ns} = (NN_A)^{-1} = 3.1 \times 10^3$ Å^2 where we used the average value of $N = 5.4 \times 10^{-12}$ mol/cm^2 found from EC_{50} measurements for both library and masking DNAs and N_A is the Avogadro's number. The obtained value of S_{ns} is sufficiently high to allow both library DNA and masking DNA to bind the cell surface with all their nucleotides. In this case, the 80-nt long library ssDNA would occupy ~3×10^3 Å^2 and the 40-nt long masking ssDNA would occupy half of this area. Of course, such full binding is not required and probably not realized in experiments. To verify the validity of the obtained values of N, K_{nsM}, and K_{nsL}, we substituted them into relations 35 and calculated ratios $\alpha_{FM}/F_{max,M}$ and $\alpha_{FL}/F_{max,L}$. We also measured these ratios using fluorescence signals from fluorescently labeled DNA bound to MCF-7 cells obtained via flow cytometry. Theoretical and experimental results are shown in Figure 4 by red and magenta lines, respectively. A deviation

dx.doi.org/10.1021/ac400385v | Anal. Chem. 2013, 85, 4157−4164

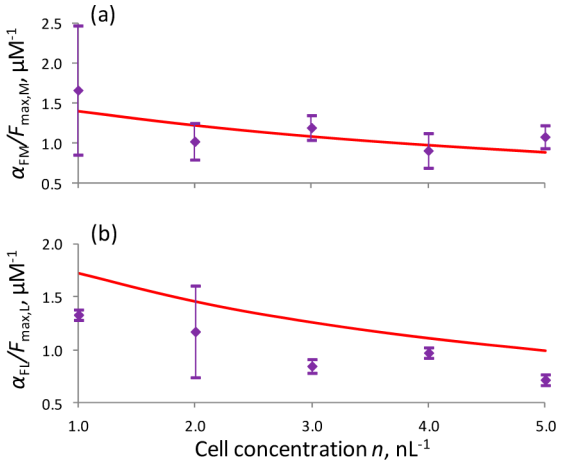


Figure 4. Dependencies of the ratios of the signal slope α_F (observed at small DNA concentrations) to the saturated signal F_{max} (reached at large DNA concentrations) for masking DNA (a) and library DNA (b). Experimental data (mean values with error bars) are shown in magenta color. Theoretical dependencies $\alpha_{FM}/F_{max,M}(n)$ and $\alpha_{FL}/F_{max,L}(n)$ determined by relations 35 are depicted by red lines.

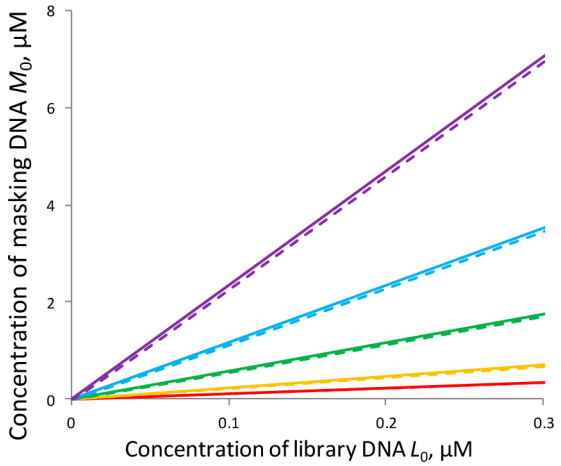


Figure 5. Dependences $M_0(L_0)$ for MCF-7 cells at various values of the outcompeting ratio κ and cell concentration n. Red, yellow, green, blue, and magenta lines correspond to κ = 1, 2, 5, 10, and 20, respectively. Solid and dashed lines show dependences for n = 1 nL^{-1} and n = 5 nL^{-1}, respectively.

between them can be quantitatively characterized by NRMSDs of 10.9% and 23.5% for Figure 4a,b, respectively. We again used an average value of $N = 5.4 \times 10^{-12}$ mol/cm^2 determined from Figure 3. Given the approximate nature of the theoretical model employed, these results should be considered as a satisfactory confirmation of the validity of obtained values of N, K_{nsM}, and K_{nsL} and the model itself.

On the basis of previous cell-SELEX experiments, it was determined that cell-specific aptamers typically have an EC_{50} value of less than 200 nM (with approximately 1−2 target cells in 1 nL of solution).[16−18,20,37] As a result, the selection progress is usually monitored using flow cytometry and selection is deemed to be complete when the binding affinity of the ssDNA for MCF-7 cells reaches an EC_{50} of 200 nM or less based on the typical binding affinity seen previously.[38] In order to minimize nonspecific binding and maximize binding of high-affinity aptamers, we need to determine the amount of masking DNA required to outcompete library DNA at a given ratio equal to κ. Figure 5 shows dependencies of the required concentration M_0 of masking DNA on the concentration L_0 of library DNA for MCF-7 cells at various values of the outcompeting ratio κ and at two different cell concentrations. Calculations were based on expression 24 in which we used values of $S = 2 \times 10^{-5}$ cm^2, $N = 5.4 \times 10^{-12}$ mol/cm^2, K_{nsM} = 0.56 μM, and K_{nsL} = 0.48 μM obtained for MCF-7 cells. Obviously, these $M_0(L_0)$ dependences are practically linear and only weakly affected by the cell concentration. Such behavior is explained by the fact that expression 24 depends on the cell concentration n through the second term that is relatively small at the considered concentrations of cells (n < 5 nL^{-1}) and at values of N, K_{nsM}, and K_{nsL} found by us for MCF-7 cells.

MATERIALS AND METHODS

Masking DNA and Library DNA. All oligonucleotides were purchased from IDT (Coralville, IA, USA). All DNA used was 6-carboxyfluorescein (FAM)-labeled at the 5′ end. The masking DNA sequence was generated at random and modified to ensure no binding with the library primer regions (5′-FAM/AA GGG TCC TGT GCT ATA ACT GTG GGT CTA GTG GTA TTT AG-3′). ssDNA library consisted of a 40-nt random region flanked by 20-nt-long primer regions (5′-FAM/AGC CTA ACG CAG AAC AAT GG-random region-CGA TGC CAG GTT AAA GCA CT-3′).

Cell Cultures and Suspension Preparation. MCF-7 cells were grown in the Dulbecco's Modified Eagle medium with high glucose content (DMEM/HG, catalog #D5796, Sigma-Aldrich, St. Louis, MO, USA) with 10% fetal bovine serum (FBS, catalog #SH30396.03, HyCyclone Laboratoried, Logan, UT, USA) and 1% 100 U/mL penicillin−streptomycin (penicillin−streptomycin solution catalog #P4333, Sigma-Aldrich, Oakville, ON, Canada) at 37 °C under humidified 5% CO_2. Cell diameter was estimated using hematocytometer markings and was in agreement with previously determined values for this cell line.[39] Plates were rinsed twice with Dulbecco's phosphate buffered saline (PBS, catalog #D8537, Sigma-Aldrich, Oakville, ON, Canada), and cells were detached using 0.05% Trypsin (Trypsin-EDTA Solution 1×, catalog #59417C, Sigma-Aldrich, Oakville, ON, Canada) at 37 °C. Cells were then resuspended in DMEM/HG, counted via hematocytometer, and spun down for 5 min at 300g and 4 °C. Medium was discarded, and cellular pellet was resuspended in PBS + 5 mM $MgCl_2$ to the desired concentration (1 to 5 cells per 1 nL).

Determination of EC_{50}. Cell suspension was alliquoted into 500 μL samples and incubated with increasing concentrations of masking DNA or library DNA for 30 to 60 min at 37 °C while gently shaking. Cells were then analyzed with flow cytometry for 30 000 events (except 1 cell/nL suspension which was analyzed for 10 000 events). Mean fluorescence signal was obtained at each masking DNA or library DNA concentration. A binding affinity curve was generated by plotting DNA concentration versus mean fluorescence signal, and the EC_{50} was determined from the obtained binding curve by locating the DNA concentration at which half of the maximum fluorescence signal was reached. Values of the standard deviation for experimental data presented in Figure 3 lie in intervals 0.01−0.09 and 0.02−0.14 μM for masking DNA and library DNA, respectively. The corresponding coefficients of variation fall into ranges 0.6−16% and 3.4−18%.

dx.doi.org/10.1021/ac400385v | Anal. Chem. 2013, 85, 4157−4164

Determination of α_{FM} and α_{FL}. For each cell concentration (1, 2, 3, 4, and 5 cells per 1 nL), α_{FM} was obtained by first plotting (masking or library) DNA concentration versus mean fluorescence signal for five experimental data points (DNA concentrations of 0, 10, 20, 30, and 50 nM). A best fit straight line was then drawn through these points, and α_{FM} was determined as the slope of this line. Values of the standard deviation for experimental data presented in Figure 4 lie in intervals 0.14−0.89 and 0.03−0.43 μM^{-1} for masking DNA and library DNA, respectively. The corresponding coefficients of variation fall into ranges 13−24% and 3.0−36% (with exception of 50% for the first experimental point in Figure 4a).

CONCLUSIONS

Here, we introduced a simple mathematical model, eqs 1−5, that describes the binding of library DNA (containing potential aptamers) and masking DNA to the cell surface. The model takes into account the high-affinity binding of aptamers to specific binding sites on the cell surface and the low-affinity binding of library DNA and masking DNA to nonspecific sites on the cell surface. On the basis of this model, we developed a simple method for estimating the concentration of masking DNA required to outcompete library DNA at any desirable ratio κ describing binding to nonspecific sites. The required concentration of masking DNA, M_0^*, is given by a dimensionless expression 22. This expression was obtained as a mathematical solution of the original model, eqs 1−5, and is illustrated in Figures 1 and 2. The corresponding dimensional expression 24 for M_0 depends on the outcompeting ratio κ, concentrations L_0 and n of library DNA and cells, respectively, and on the following cell characteristics: S, N, K_{nsL}, and K_{nsM}. As a result, expression 24 can be used to study the dependence of M_0 on L_0 at a given κ. Such dependence allows one to estimate the efficiency of masking DNA outcompeting the library DNA. Parameters κ, L, n, and S are known from experimental conditions whereas characteristics N, K_{nsL}, and K_{nsM} are not. We obtained theoretical relations 30 allowing the determination of N, K_{nsL}, and K_{nsM} based on measurements of EC_{50} values for cells mixed separately with the library and masking DNAs. We also obtained expressions 35 for signal characteristics ($\alpha_{FM}/F_{max,M}$ and $\alpha_{FL}/F_{max,L}$) from cells bound to DNA (measured with flow cytometry) in terms of N, K_{nsL}, and K_{nsM}. These expressions can be used for the verification of values of N, K_{nsL}, and K_{nsM} found from EC_{50} measurements. The developed procedure of cell characteristics determination was successfully applied to the MCF-7 breast cancer cell line (Figures 3 and 4), and corresponding values of N, K_{nsL}, and K_{nsM} were established for the first time. Finally, we estimated the concentration of masking DNA required to outcompete library DNA in nonspecific binding to MCF-7 breast cancer cells (Figure 5). The obtained results will allow one to employ masking DNA more efficiently in AptaBiD and cell-SELEX procedures.

AUTHOR INFORMATION

Corresponding Author

*E-mail: skrylov@yorku.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was funded by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

(1) Etzioni, R.; Urban, N.; Ramsey, S.; McIntosh, M.; Schwartz, S.; Reid, B.; Radich, J.; Anderson, G.; Hartwell, L. *Nat. Rev. Cancer* **2003**, *3*, 243−252.
(2) Meyer, H. E.; Stuhler, K. *Proteomics* **2007**, *7*, 18−26.
(3) Kelly, D. J.; Ghosh, S. *Dis. Markers* **2005**, *21*, 43−48.
(4) Lescuyer, P.; Hochstrasser, D.; Rabilloud, T. *J. Proteome Res.* **2007**, *6*, 3371−3376.
(5) Diamandis, E. P. *Mol. Cell. Proteomics* **2004**, *3*, 367−378.
(6) Horvatovich, P.; Govorukhina, N.; Bischoff, R. *Analyst* **2006**, *131*, 1193−1196.
(7) Reid, J. D.; Parker, C. E.; Borchers, C. H. *Curr. Opin. Mol. Ther.* **2007**, *9*, 216−221.
(8) Norton, S. M.; Huyn, P.; Hastings, C. A.; Heller, J. C. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 325−331.
(9) Rifai, N.; Gillette, M. A.; Carr, S. A. *Nat. Biotechnol.* **2006**, *24*, 971−983.
(10) Berezovski, M. V.; Lechmann, M.; Musheev, M. U.; Mak, T. W.; Krylov, S. N. *J. Am. Chem. Soc.* **2008**, *130*, 9137−9143.
(11) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818−822.
(12) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505−510.
(13) Osborne, S. E.; Ellington, A. D. *Chem. Rev.* **1997**, *97*, 349−370.
(14) Nutiu, R.; Li, Y. *Angew. Chem., Int. Ed.* **2005**, *44*, 1061−1065.
(15) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611−647.
(16) Chen, H. W.; Medley, C. D.; Sefah, K.; Shangguan, D.; Tang, Z.; Meng, L.; Smith, J. E.; Tan, W. *Chem. Med. Chem.* **2008**, *3*, 991−1001.
(17) Dwivedi, H. P.; Smiley, R. D.; Jaykus, L. *Appl. Microbiol. Biotechnol.* **2010**, *87*, 2323−2334.
(18) Shangguan, D.; Li, Y.; Tang, Z.; Cao, Z. C.; Chen, H. W.; Mallikaratchy, P.; Sefah, K.; Yang, C. J.; Tan, W. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11838−11843.
(19) Avci-Adali, M.; Metzger, M.; Perle, N.; Ziemer, G.; Wendel, H. P. *Oligonucleotides* **2010**, *20*, 317−323.
(20) Tang, Z.; Shangguan, D.; Wang, K.; Shi, H.; Sefah, K.; Mallikratchy, P.; Chen, H. W.; Li, Y.; Tan, W. *Anal. Chem.* **2007**, *79*, 4900−4907.
(21) Chatelier, R. C.; Ashcroft, R. G.; Lloyd, C. J.; Nice, E. C.; Whitehead, R. H.; Sawyer, W. H.; Burgess, A. W. *EMBO J.* **1986**, *5*, 1181−1186.
(22) McCoy, S. L.; Hausman, F. A.; Deffebach, M. E.; Bakke, A.; Merkens, L. S.; Bennett, R. M.; Hefeneider, S. H. *J. Immunol. Meth.* **2000**, *241*, 141−146.
(23) Volná, P.; Jarjour, J.; Baxter, S.; Roffler, S. R.; Monnat, R. J., Jr; Stoddard, B. L.; Scharenberg, A. M. *Nucleic Acids Res.* **2007**, *35*, 2748−2758.
(24) Christopoulos, A. *Nat. Rev. Drug Discovery* **2002**, *1*, 198−210.
(25) Hefeneider, S. H.; Cornell, K. A.; Brown, L. E.; Bakke, A. C.; McCoy, S. L.; Bennett, R. M. *Clin. Immunol. Immunopathol.* **1992**, *63*, 245−251.
(26) Künne, A. G. E.; Sieber, M.; Meierhans, D.; Allemann, R. K. *Biochemistry* **1998**, *37*, 4217−4223.
(27) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. *Structure* **2000**, *8*, 1015−1023.
(28) Jen-Jacobson, L.; Engler, L. E.; Ames, J. T.; Kurpiewski, M. R.; Grigorescu, A. *Supramol. Chem.* **2000**, *12*, 143−160.
(29) Peters, W. B.; Edmondson, S. P.; Shriver, J. W. *J. Mol. Biol.* **2004**, *343*, 339−360.
(30) Biswas-Fiss, E. E.; Kukiratirat, J.; Biswas, S. B. *BMC Biochem.* **2012**, *13*, 10_1−10_16.
(31) Berg, O. G.; Ehrenberg, M. *Biophys. Chem.* **1982**, *15*, 41−51.
(32) Berg, O. G. *Biophys. J.* **1985**, *47*, 1−14.
(33) Zhou, H.-X.; Szabo, A. *Phys. Rev. Lett.* **2004**, *93*, 178101_1−178101_4.

(34) Gopich, I. V.; Szabo, A. *J. Chem. Phys.* **2002**, *117*, 507−517.
(35) Soundararajan, S.; Wang, L.; Sridharan, V.; Chen, W.; Courtenay-Luck, N.; Jones, D.; Spicer, E. K.; Fernandes, D. J. *Mol. Pharmacol.* **2009**, *76*, 984−991.
(36) Kanoatov, M.; Retif, C.; Cherney, L. T.; Krylov, S. N. *Anal. Chem.* **2012**, *84*, 149−154.
(37) Bayrac, A. T.; Sefah, K.; Parekh, P.; Bayrac, C.; Gulbakan, B.; Oktem, H. A.; Tan, W. *ACS Chem. Neurosci.* **2011**, *2*, 175−181.
(38) Ai, J.; Li, T.; Li, B.; Xu, Y.; Li, D.; Liu, Z.; Wang, E. *Anal. Chim. Acta* **2012**, *741*, 93−99.
(39) Gobert, G. N.; Schatten, H. *J. Electron Microsc.* **2000**, *49*, 539−544.