Quantitative Characterization of Partitioning in Selection of DNA Aptamers for Protein Targets by Capillary Electrophoresis

An T. H. Le, Tong Ye Wang, Svetlana M. Krylova, Stanislav S. Beloborodov, and Sergey N. Krylov*

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

*Corresponding author's email: skrylov@yorku.ca

ABSTRACT: Partitioning of protein–DNA complexes from protein-unbound DNA is a key step in selection of DNA aptamers. Conceptually, the partitioning step is characterized by two parameters: transmittance for protein-bound DNA (binders) and transmittance for unbound DNA (nonbinders). Here we present the first study to reveal how these transmittances depend on experimental conditions; such studies are pivotal to the effective planning and control of selection. Our focus was capillary electrophoresis (CE) which is a partitioning approach of high efficiency. By combining a theoretical model and experimental data, we evaluated the dependence of transmittances of binders and nonbinders on the molecular weight of protein target in two modes of CE-based partitioning: nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) and ideal-filter capillary electrophoresis (IFCE). Our data suggest that as the molecular weight of the protein target decreases: (i) the transmittance for binders remains close to unity in NECEEM but decreases drastically in IFCE and (ii) the transmittance for nonbinders increases orders of magnitude in NECEEM but remains relatively stable at a very low level in IFCE. To determine the optimal CE conditions for a given size of protein target, a balance between transmittances of binders and nonbinder must be reached; such a balance would ensure the collection of binders of sufficient purity and quantity. We conclude that, as a rule of thumb, IFCE is preferable for large-size protein targets while NECEEM should be the method of choice for small-size protein targets.

Aptamers are single-strand DNA (or RNA) capable of tightly and specifically binding to targets for which they have been selected.^{1,2} Aptamers can serve as affinity probes and therapeutic agents.³⁻¹² They are selected from a random-sequence DNA library using their ability to bind to the target as a selection criterion. A typical aptamer-selection procedure is systematic evolution of ligands by exponential enrichment (SELEX).^{13,14} SELEX involves repetitive rounds of three steps: 1) reacting the library with the target to form target–DNA complexes (target–binder complexes), 2) partitioning the complexes from the unbound DNA (nonbinders), and 3) amplifying the collected DNA by PCR to obtain a binder-enriched library for the next round of selection. Another approach for aptamer selection is non-SELEX, in which consecutive rounds of two initial steps — reacting and partitioning — are conducted without PCR amplification between the rounds.¹⁵⁻¹⁷ Non-SELEX is faster than SELEX, but the maximum number of rounds in non-SELEX is limited due to unavoidable losses of binders in partitioning and between the selection rounds.

Partitioning is evidently a key step in aptamer selection — increasing the efficiency of partitioning allows completion of aptamer selection in fewer rounds and can help avoid selection failures.¹⁸ Partitioning can be conceptually presented as a physical filter that lets binders through but stops nonbinders (Figure 1a). It can then be described quantitatively using a term of “transmittance” utilized for the characterization of optical filters. Transmittance of partitioning for binders (B), k_B , is defined as the ratio between quantities of binders at the output, B_{out} , and input, B_{in} , of partitioning, respectively:

$$k_B = B_{out} / B_{in} \quad (1)$$

Accordingly, transmittance of partitioning for nonbinders (N), k_N , is defined as the ratio between the quantities of nonbinders at the output, N_{out} , and input, N_{in} , of partitioning, respectively:

$$k_N = N_{out} / N_{in} \quad (2)$$

Both values theoretically range between 0 and 1, and k_B must be

greater than k_N . Ideal partitioning is the one with $k_B = 1$ and $k_N = 0$, while in reality $k_B < 1$ and $k_N > 0$. The passage of nonbinders through partitioning creates the nonbinder background in the selection process and contaminates binders at the exit of partitioning. The value of k_N can be used as a quantitative measure of the nonbinder background. Collectively, the values of k_B and k_N are sufficient to characterize partitioning quantitatively. Their ratio, k_B/k_N , is the efficiency of partitioning which ranges between 0 and ∞ and links the binder-to-nonbinder ratio at the output of partitioning to that at the input of partitioning:

$$\frac{B_{out}}{N_{out}} = \frac{B_{in}}{N_{in}} \frac{k_B}{k_N} \quad (3)$$

When optimizing partitioning, one needs to minimize k_N without proportionally decreasing k_B . Knowing how k_N and k_B depend on multiple experimental parameters for a specific method of partitioning is, therefore, pivotal to planning efficient aptamer selection and to controlling it. This knowledge can also help validate the results of single-round aptamer selection.¹⁸ We found no reports on quantitative study dedicated to what and how k_N and k_B depend on. The lack of such studies is arguably the major reason for aptamer selection still being more an art than science. This work results from our effort to initiate such studies and make quantitative characterization of partitioning a foundation for technological advancement of aptamer selection.

Many different partitioning methods are utilized in aptamer selection. They can be divided into two major approaches: surface-based partitioning and solution-based partitioning. In surface-based partitioning, the target is immobilized on the surface (e.g., of magnetic beads) to capture the binders and facilitate relatively easy removal of nonbinders by simply rinsing the surface.^{1,2,19-23} The efficiency of partitioning (k_B/k_N) for surface-based methods is greatly affected by non-specific binding of nonbinders to the surface which creates high nonbinder background. As a result, surface-based partitioning is characterized by relatively high k_N values typically exceeding

10^{-3} .^{19,21,22} In solution-based partitioning, the target–binder complexes are formed in solution and separated from nonbinders due to different mobilities of the complexes and nonbinders in a force field — typically, electric field in electrophoresis.^{24–30} In this case, the non-specific binding of nonbinders to the surface may not affect the efficiency of partitioning; the sources of nonbinder background are different from this surface-associated effect. The nonbinder background in solution-based partitioning is lower than in surface-based partitioning, and k_N values are typically well below 10^{-3} .^{24,25}

Solution-based partitioning of aptamers for protein targets by capillary electrophoresis (CE) is characterized by the lowest k_N values on record: $k_N < 10^{-5}$.^{24,25} CE-based partitioning was used for both SELEX and non-SELEX.^{15–17,28–32} Different modes of CE-based partitioning have been proposed and successfully used.^{24,25,28} The small k_N values of CE-based partitioning allow single-round aptamer selection for some protein targets.²⁵ The high efficiency and versatility of CE-based partitioning made it a worthy subject for this first study on quantitative characterization of partitioning in aptamer selection. We combined theory and experiment to understand the underlying concepts and obtain empirical information required for this study to be instructive for practical users.

In essence, here we study the dependence of k_N and k_B on the molecular weight of protein target for varying pH and ionic strength of the running buffer. For a given protein size, the optimal running buffer conditions should guarantee low k_N and high k_B for a higher chance of successful selection. When the running buffer has lower than physiological ionic strength and/or higher than physiological pH, the target–binder complexes and nonbinders move in the same direction; these mode of partitioning is known as nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM).²⁴ We found that in NECEEM, k_N increases by several orders of magnitude while k_B remains close to unity when the molecular weight of the protein target decreases. When the running buffer has both ionic strength and pH near physiological, the target–binder complexes and nonbinders move in the opposite directions. Such conditions facilitate so-called ideal-filter capillary electrophoresis (IFCE). IFCE is characterized by the lowest k_N values on record (10^{-9})²⁵ achieved at the expense of large decrease in k_B . With decreasing molecular weight of the protein target, k_N does not change much while k_B decreases by as much as orders of magnitude in IFCE. The decrease in k_B discourages from using IFCE conditions for small-size protein targets. Our results suggest that IFCE conditions are most suitable for large-size protein targets to obtain high affinity binders in a minimal number of partitioning rounds. When the non-SELEX approach is used (Figure 1b), losses of binders are significant between partitioning rounds, and these losses cannot be compensated as there is no PCR amplification in between the rounds. Therefore, in non-SELEX, NECEEM (in which k_B is close to unity) is preferred over IFCE to retain a sufficient quantity of binders for the next rounds of partitioning and the concluding PCR amplification.

MATERIALS AND METHODS

Chemicals and Materials. All chemicals were from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries with inner and outer diameters of 75 and 360 μm , respectively, were obtained from Molex Polymicro (Phoenix, AZ, USA). All DNA molecules were custom synthesized by Integrated DNA Technologies (Coralville, IA, USA). Bodipy (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) was purchased from Life Technologies Inc. (Burlington, ON, Canada).

The CE running buffers were 50 mM Tris-Acetate at pH 8.2 without NaCl for NECEEM and 50 mM Tris-HCl at pH 7.0 supplemented with 100 mM NaCl for IFCE, resulting in ionic strength of the running buffer (I) of 25 and 146 mM, respectively. The sample buffer was always the same as the running buffer to

Figure 1. Schematic representation of a) partitioning of binders (B) from nonbinders (N), b) non-SELEX selection of binders and c) SELEX selection of binders. See text for details.

prevent adverse effects of buffer mismatch. Accordingly, all dilutions of sample components used in CE experiments were done by adding the corresponding buffer.

We used a synthetic FAM-labeled DNA library (N40) with a 40-nt random region: 5'-FAM-CT ACG GTA AAT CGG CAG TCA-(N40)-AT CTG AAG CAT AGT CCA GGC-3'. The nucleotide sequence of the forward primer (uF) was 5'-CTA CGG TAA ATC GGC AGT CA-3', and the sequence of the reverse primer (uR) was 5'-GCC TGG ACT ATG CTT CAG AT-3'. All solutions were prepared in deionized water filtered through a 0.22- μm Milipore filter membrane (Nepean, ON).

Capillary Electrophoresis. All CE experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detection system. Fluorescence was excited with a blue line (488 nm) of a solid-state laser and detected at 520 nm using a spectrally optimized emission filter system.³³

All capillaries were 50-cm long (40 cm to the detector) and had an inner diameter of 75 μm and an outer diameter of 360 μm . The poly(vinyl alcohol) (PVA)-coated capillary was prepared as described elsewhere.³⁴ Prior to every fraction-collection experiment, a new capillary was installed and conditioned successively with MeOH at 20 psi for 10 min, 0.1 M HCl at 20 psi for 3 min, 0.1 M NaOH at 10 psi for 6 min, water at 20 psi for 3 min and a running buffer at 40 psi for 40 min. Prior to every run, the capillary was rinsed successively with 0.1 M HCl, 0.1 M NaOH, deionized H₂O, and a running buffer for 3 min each. Conditioning steps were not required for PVA-coated capillaries; such capillaries were rinsed with the running buffer only at 20 psi for 10 min prior to the fraction-collection experiment.

The sample contained 10 μM annealed oligonucleotides (melted at 90 $^{\circ}\text{C}$ for 2 min and gradually cooled down to 20 $^{\circ}\text{C}$ at a rate of 0.5 $^{\circ}\text{C}/\text{s}$) and 150 nM Bodipy (an electrically neutral molecule, EOF marker). The sample mixture was injected with a pressure pulse of 0.5 psi (~ 3.5 kPa) $\times 10$ s to yield a 10-mm long sample plug. The injected sample plug was propagated through the uncooled part of the capillary at the inlet by injecting a 5.7-cm long plug of the buffer with a pressure pulse of 0.3 psi (~ 2.1 kPa) $\times 90$ s.

CE was carried out at an electric field of 200 V/cm (10 kV over 50 cm). CE-run duration was 34 min for NECEEM conditions and 128 min for IFCE. For uncoated capillaries, CE was carried out with the positive electrode at the injection end of the capillary; for PVA-coated capillaries, the polarity was reversed. Collection vials contained 20 μL of the running buffer

each and were switched every 4 min for IFCE and every 1 min for NECEEM.

Quantitative PCR. DNA in the collected fractions was amplified and quantitated by qPCR using a CFX Connect instrument from Bio-Rad (Mississauga, ON, Canada). A qPCR reagent mixture was prepared by combining IQ SYBR Green Supermix from Bio-Rad with unlabeled DNA primers at final concentrations of $1 \times$ SYBR Green Supermix, 100 nM uF, and 100 nM uR. A qPCR reaction mixture was prepared by adding 2- μ L aliquot of each fraction into 18 μ L of the qPCR reagent mixture immediately before thermocycling. The thermocycling protocol was: 95 °C (initialization) for 3 min, 95 °C for 10 s (denaturation), 56 °C for 10 s (annealing), 72 °C for 10 s (extension), followed by a plate read at 72 °C and a return to the denaturation step (bypassing the 95 °C $\times$ 3 min initialization step) for a total of 43 cycles. All qPCR reactions were performed in duplicates.

THEORETICAL AND EXPERIMENTAL CONSIDERATION

Dependence of Number of Partitioning Rounds on k_N & k_B .

This analysis is general and does not depend on the type of partitioning method. Let us consider major conditions to be satisfied to ensure successful selection. PCR amplification of collected DNA is always used as a final step of aptamer selection before DNA sequencing. If selection is done in a single round (Figure 1a), then the quantity of binders at the output of partitioning, B_{out} , must exceed the level of PCR noise (N_{PCR}), produced during the PCR amplification of the selected binders, by a set number $Q_1 > 1$:

$$B_{out} > Q_1 N_{PCR} \quad (3)$$

Another condition is that the partitioning supports the removal of nonbinders sufficiently well to exceed a certain level of binder purity Q_2 which depends on the specifics of selection and would usually be set at or near unity (e.g., 0.1, 1, 10):

$$B_{out} / N_{out} > Q_2 \quad (4)$$

A single round of selection is rarely sufficient to satisfy the inequality in eq 4. If multiple consecutive rounds of selection are conducted without PCR amplification between them (applicable to non-SELEX), then the number of rounds m should, in turn, satisfy two conditions. Firstly, m should be not too high to prevent the excessive loss of binders, i.e., satisfy inequality in eq 3:

$$m \leq \left\lfloor \frac{\log(Q_1 N_{PCR} / B_{in})}{\log(k_B)} \right\rfloor \quad (5)$$

where $\lfloor x \rfloor$ represent a mathematical function that rounds x down to the nearest integer (see Note S1 for derivation of eq 5). Secondly, m should be high enough to satisfy inequality in eq 4:

$$m \geq \left\lceil \frac{\log(Q_2 N_{in} / B_{in})}{\log(k_B / k_N)} \right\rceil \quad (6)$$

where $\lceil x \rceil$ represent a mathematical function that rounds x up to the nearest integer (see Note S1 for derivation of eq 6). The last two inequalities establish the range of acceptable values of m :

$$\left\lceil \frac{\log(Q_1 N_{PCR} / B_{in})}{\log(k_B)} \right\rceil \geq m \geq \left\lceil \frac{\log(Q_2 N_{in} / B_{in})}{\log(k_B / k_N)} \right\rceil \quad (7)$$

The values of Q_1 and Q_2 may be set taking into account secondary considerations; for Q_2 it may be, for instance, the cost of post-selection screening.¹⁸ The values of k_B and k_N depend on the partitioning method of choice and can be estimated or determined experimentally. The values of B_{in} and N_{in} are not known for real selections (unlike mock selections in which known binders are spiked controllably into known nonbinders), but different scenarios can be considered for them conclusively when analyzing any specific selection case.

In the case of classic aptamer selection by SELEX, PCR is used between the consecutive rounds to maintain the quantity of binders (Figure 1c). Therefore, m does not have an upper limit,

but there is still a lower limit of m :

$$m \geq \left\lceil \frac{\log(Q_2 N_{in} / B_{in})}{\log((k_B / k_N)(Z_B / Z_N)^n)} \right\rceil \quad (8)$$

where Z_B and Z_N are the bases of the exponent describing PCR amplification of binders and nonbinders, respectively. In an unbiased amplification, binders and nonbinders are amplified with the same efficiency, i.e., $Z_B = Z_N$. It is likely, however, that $Z_B < Z_N$ (due to the more folded structure of aptamers), which imposes an upper limit for the number of PCR cycles:^{35,36}

$$n < \left\lfloor \frac{\log(k_N / k_B)}{\log(Z_B / Z_N)} \right\rfloor \quad (9)$$

Thus, for SELEX, there is a lower limit for the number of selection rounds and upper limit for the number of PCR cycles in a single round.

Accurately assessing the limits for m and n using eqs 7, 8, and 9 *a priori* is impossible due to the uncertainties in B_{in} , N_{in} , Z_B , and Z_N . However, some quantitative analysis of these limits can be conducted upon reasonable assumptions for the values of B_{in} , N_{in} , Z_B , and Z_N , and instructive conclusions can be made.

Major Modes of CE-based Partitioning. In CE-based partitioning, the zone of the protein–binder complexes (also denoted as P–DNA) is separated from the zone of the nonbinders (also denoted as unbound DNA) based on the difference between electrophoretic mobility of P–DNA (μ_{P-DNA}) and that of DNA (μ_{DNA}). If the running buffer does not contain the protein, then P–DNA starts dissociating as soon as it has been separated from the zone of unbound DNA (in a matter of seconds). Accordingly, there are three features in a CE separation profile: 1) a peak corresponding to intact P–DNA (contains binders), 2) a peak corresponding to DNA that was unbound in the equilibrium mixture (contains nonbinders), and 3) a “bridge” between the two peaks that corresponds to DNA dissociating from the complexes during CE separation (contains binders).

We distinguish two major modes of CE-based partitioning: 1) NECEEM, in which P–DNA and DNA move in the same direction and 2) IFCE, in which P–DNA and DNA move in the opposite directions. There are two sub-modes of NECEEM: “complex first” in which P–DNA moves faster than DNA (Figure 2a) and “complex last” in which DNA moves ahead of P–DNA (Figure 2c). In both sub-modes, NECEEM electropherograms contain all three features (the two peaks and the bridge between them); only the order of the peaks and the direction of the bridge change. In IFCE, P–DNA moves towards the collection end of the capillary while all unbound DNA (including binders dissociated from P–DNA during partitioning) moves in the opposite direction resulting in an electropherogram which contains only the peak of intact complex (Figure 2b).

The purpose of partitioning is to collect binders and reject nonbinders. In CE-based partitioning, this is achieved by collecting a sample fraction at the capillary outlet in a specific binder-collection window. In NECEEM, the binder-collection window can cover both intact P–DNA and DNA dissociated from P–DNA during CE separation (see Figure 2a,c). In IFCE, the binder-collection window can only cover the intact P–DNA (see Figure 2b) as the bridge moves along with the peak of nonbinders towards the injection end of the capillary.

The binder-collection window in NECEEM includes a tail of the unbound DNA peak which constitutes the nonbinder background (DNA background). In contrast, IFCE appears to be free of this effect based on the fundamentals of CE separation. In reality, both NECEEM and IFCE partitioning always have DNA background in the binder-collection window due to a phenomenon of non-uniform migration of DNA in a uniform electric field.²⁴ The phenomenon is hypothetically caused by the effect of the electric field on very stable complexes of DNA with counterions.³⁷ There is no quantitative theory of this effect that

Figure 2. Schematic representation of different modes of CE-based partitioning: a) “complex first” NECEEM, b) IFCE and c) “complex last” NECEEM. The EOF bar indicates migration time of an EOF marker (a neutral molecule). See text for details.

could help to predict the level of the DNA background in CE-based partitioning; therefore, this background should be studied empirically.

Sequential transition from complex-first NECEEM to IFCE and then to complex-last NECEEM is achieved by reducing the mobility of the electroosmotic flow (EOF, μ_{EOF}). The top part of Figure 2 schematically shows directions of velocities of P-DNA and DNA along with relations between $\mu_{\text{P-DNA}}$, μ_{DNA} , and μ_{EOF} for these three cases. The value of μ_{EOF} for a bare fused-silica capillary depends on pH of the running buffer and its ionic strength I : lowering pH and/or increasing I of the running buffer lead to decreasing μ_{EOF} . Coating the inner wall of the capillary with a non-ionizable layer suppresses EOF and can lead to $\mu_{\text{EOF}} \ll \mu_{\text{P-DNA}}$, μ_{DNA} and, thus, to complex-last NECEEM for a broad spectrum of values of running buffer pH and I . The following two sections consider what k_N and k_B depend on in CE-based partitioning. The sole purpose of this consideration is to assist in rationally designing our empirical study.

Parameters Influencing k_N and k_B in CE-Based Partitioning.

The value of k_N is a function of analytical resolution R of the peaks of P-DNA and DNA (k_N decreases with increasing R).³⁸ The value of R , in turn, depends on $\mu_{\text{P-DNA}}$, μ_{DNA} , and time of separation (or elution time) t . The value of $\mu_{\text{P-DNA}}$ greatly depends on 1) the size of the protein, which is linked with its molecular weight (MW_P), 2) the length of DNA (L_{DNA}), which is the same for binders and nonbinders, and 3) pH and ionic strength I of the running buffer. Notably, the dependence of $\mu_{\text{P-DNA}}$ on the charge of the protein is negligible in the first approximation because of the much higher charge density on DNA.³⁹ The value of μ_{DNA} depends on I , and slightly depends on L_{DNA} in gel-free CE.⁴⁰ As we described in the previous section, the value of k_N is greatly influenced by the nonbinder background (DNA background) caused by the nonuniform mobility of DNA in CE.³⁷ The background itself depends (directly or indirectly) on pH, I , MW_P , and t . Cumulatively, the value of k_N is a function of five parameters

$$k_N = F(R) = F(\mu_{\text{P-DNA}}, \mu_{\text{DNA}}, t) = F(MW_P, L_{\text{DNA}}, t, \text{pH}, I) \quad (10)$$

There may be some cross-influence of the parameters, *e.g.*, the choice of t may depend on MW_P , but such nuances do not affect the essence of eq 10, and, therefore, are not a subject of this conceptual consideration.

The value of k_B depends on whether the aptamer-collection window covers the entire span of binders in the profile (Figure 2). NECEEM and IFCE are radically different with regards to k_B . In theory, k_B decreases with time exponentially with a rate constant k_{off} of dissociation of P-DNA complexes. In NECEEM, both intact P-DNA complexes and DNA dissociated from P-DNA during CE migrate in the same direction. Therefore, nearly all binders can be collected, and k_B in NECEEM can be assumed to be close to unity. In IFCE, P-DNA moves towards the collection

end of the capillary, but the unbound DNA migrates in the opposite direction. As a result, only the intact complexes are collected while binders dissociated from the complexes during CE are not. P-DNA dissociates following the monomolecular decay. The time t during which P-DNA is allowed to dissociate before elution (*i.e.*, elution time) is defined by $\mu_{\text{P-DNA}}$ and, thus, depends on MW_P and L_{DNA} . A mixture of binders with different k_{off} values is not characterized by a specific k_{off} value; therefore, we use k_{off} here as a loose term. Thus, we can write for k_B in NECEEM and IFCE, respectively:

$$k_{B, \text{NECEEM}} = \text{const} \approx 1 \quad (11)$$

$$k_{B, \text{IFCE}} = F(k_{\text{off}}, MW_P, L_{\text{DNA}}, \text{pH}, I, t)$$

As seen from eqs 10 and 11, k_N and k_B in CE partitioning are defined by a total of 6 parameters: MW_P , L_{DNA} , k_{off} , t , pH, and I . Rational design of CE-based aptamer selection, thus, requires an experimental study that would lead to understanding how these parameters affect k_N and k_B .

Rational for Experimental Design. Studying experimentally the influence of all five parameters on k_N and all six parameters on k_B is not needed as these parameters have different roles and not all of them need to be varied. In all of our studies, L_{DNA} is typically 80-nt long (a 40-nt long random region flanked by 20-nt long PCR primer regions). The value of k_{off} is only defined for one aptamer and cannot be defined for a heterogeneous pool of aptamers. Moreover, it is a parameter that cannot be controlled, and, therefore, it is also a parameter not to be changed in this study. A hypothetical bulk value of k_{off} can still be considered for qualitative characterization of selection provided that no attempts are made to derive solid quantitative guidance from such consideration. MW_P is a parameter which is imposed by the target and is a major parameter for which selection conditions, *i.e.*, pH and I of the running buffer, should be selected rationally to achieve the highest efficiency of selection. Therefore, the characterization of CE partitioning can be reduced to studying how MW_P affects the values of k_N and k_B for varying values of pH and I .

Advantageously, the study of how MW_P affects the values of k_N and k_B for varying pH and I can be conducted without using proteins as there is a recently published empirical function that links the mobility of the protein-DNA complex with the molecular weight of the complex ($MW_{\text{P-DNA}}$):³⁹

$$\mu_{\text{P-DNA}} = A + B \mu_{\text{DNA}} L_{\text{DNA}}^{0.68} MW_{\text{P-DNA}}^{-1/3} \quad (12)$$

where electrophoretic mobilities are expressed in $\text{mm}^2/(\text{kVs})$, L_{DNA} is expressed in the number of nucleotides, $MW_{\text{P-DNA}}$ is expressed in kDa, while A and B are empirical constants. For a running buffer with $I < 50$ mM, these constants are $A = -9.95 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.0929 \text{ kDa}^{1/3}$. For a running buffer with $I = 146$ mM, the constants are $A = 10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.2365 \text{ kDa}^{1/3}$ (see Note S2). As $MW_{\text{P-DNA}} = MW_P + MW_{\text{DNA}}$, the predicted mobility values obtained with eq 12 are used to

estimate the associated velocity and the elution time of protein–DNA complex for given value of MW_P (see Note S3).

RESULTS AND DISCUSSION

Migration profiles of DNA in different modes of CE-based partitioning. To evaluate DNA background in the binder-collection window, we experimentally obtained DNA migration profiles in different modes of CE-based partitioning: both sub-modes of NECEEM and IFCE. The high value of EOF required for complex-first NECEEM can be achieved in the buffer system with low I and/or high pH (typical $I < 50$ mM) in a bare fused-silica capillary. In IFCE, the EOF is reduced by using buffer with high I and/or low pH (typical $I > 100$ mM) in a bare fused-silica capillary. For complex-last NECEEM, the EOF is suppressed via coating the inner wall of capillary (e.g., with PVA). In this study, we chose two previously published NECEEM and IFCE running buffers to generate qualitatively distinct migration profiles of DNA.^{24,25} The first buffer was 50 mM Tris-Acetate pH 8.2 ($I = 21$ mM), corresponding to NECEEM, in which P–DNA and DNA move in the same direction towards the capillary outlet. For consistency, we used the same buffer for both complex-first NECEEM in a bare-silica capillary and complex-last NECEEM in a PVA-coated capillary. The second buffer was 50 mM Tris-HCl pH 7.0 ($I = 146$ mM), corresponding to IFCE, in which P–DNA moves to the outlet while DNA moves to the inlet in a bare-silica capillary.

The sample of 2.8×10^{11} molecules of 80-nt DNA was subjected to both NECEEM and IFCE. One- and four-minute fractions were collected for NECEEM and IFCE, respectively; all collected fractions were analyzed by qPCR to build a “DNA quantity versus migration time to the capillary outlet” electropherogram. The results are shown in Figure 3.

In complex-first NECEEM, the main DNA peak eluted in 23 min after the start of CE (Figure 3a). A small part of DNA sample migrated in front of the main DNA peak and created the nonbinder background of a total of approximately 10^7 molecules in the zone of NECEEM binder-collection windows. As I increased to reach the IFCE condition, the migration direction of the main DNA peak switched from the direction towards the outlet end of the capillary to the opposite direction (towards the inlet end) while the nonbinder background became stretched out. In IFCE, the nonbinder background was reduced to below the limit of detection of qPCR for a 2-h binder-collection window (Figure 3c). In complex-last NECEEM, the main DNA peak moved ahead of the binder-collection window and eluted in 17 min, while the nonbinder background of a total of approximately 10^7 molecules tailed behind the main DNA zone in the binder-collection window (Figure 3b).

Although the quantities of background DNA were similar (10^7) in the two sub-modes of NECEEM, their background profiles were quantitatively different. In complex-first NECEEM, the nonbinder background was solely caused by the heterogeneity of electrophoretic velocity of DNA. This DNA background emerged above the LOQ in PCR along with the EOF marker and increased drastically (multiple orders of magnitude) within the zone of binder-collection windows with time progressing to that of elution of the main DNA peak. In complex-last NECEEM, the zone of binder-collection windows was behind the main DNA peak: in addition to the nonbinder background induced by the nonuniform electrophoretic mobility of DNA, the collection of protein–DNA complex also suffered from the contamination of residual DNA on the inner capillary wall and the outer surface of capillary outlet after the elution of the main DNA peak. In complex-first NECEEM, the nonbinder background decreased drastically (as low as 10^3 molecules) when the binder-collection window was located further away from the main DNA peak (the left boundary of the zone of binder-collection windows for complex-first sub-mode). However, in complex-last NECEEM, the nonbinder background in the regions away from the main

Figure 3. DNA background profiles under conditions of NECEEM (a and b) and IFCE (c). A sample of 2.8×10^{11} molecules of 80-nt DNA was subjected to CE-based partitioning. Fractions were collected every 1 min in NECEEM and every 4 min in IFCE. Concentration of DNA in every fraction was quantitated using qPCR, and these quantities are shown on the y-axis in the graph. The double-arrow indicates the appropriate zone of binder-collection windows in corresponding mode of CE-based partitioning.

DNA peak still remained relatively high at more than 10^6 molecules (the right boundary of the zone of binder-collection windows for complex-last NECEEM). As such, the associated nonbinder background values for protein–DNA complexes with different MW_P in complex-last NECEEM are expected to be quantitatively higher than those in complex-first NECEEM. A detailed analysis of the effect of MW_P on the nonbinder background for all modes of CE-based partitioning will be presented in the next section.

Influence of MW_P on k_N . Knowing the predicted mobility of protein–DNA complex (eq 12) allows one to calculate the binder-collection window in which P–DNA should elute from the capillary (see Note S3). Knowing this time window, in turn, allows the determination of the transmittance of partitioning for nonbinders k_N . The value of k_N was calculated based on eq 2 as the total number of background DNA (N_{out}) collected within the binder-collection window divided by the total number of DNA injected into the capillary (N_{in}).

The resulting dependence of k_N on MW_P ranging between 25 to 150 kDa is shown in Figure 4. In both NECEEM sub-modes, as MW_P decreased from 150 to 25 kDa, k_N increases approximately 3 orders of magnitude. As expected, k_N values for complex-last NECEEM are higher than those for complex-first NECEEM due to the elevated DNA background in complex-last NECEEM (Figure 3). In IFCE, the background profile is stable with k_N remaining near 10^{-9} throughout the 2-h run; thus, k_N is similar for target proteins with different MW_P . In general, the IFCE running buffer with higher ionic strength and/or lower pH leads to lower k_N values as well as weaker dependence of k_N on MW_P . However, the low EOF obtained with such a running buffer increases the predicted complex migration time to over 3 h for smaller proteins with $MW_P < 25$ kDa (Note S4). Note that the effect of MW_P on k_N cannot be measured experimentally for target–binder complexes with $MW_P < 25$ kDa under IFCE conditions due to an unreasonably long CE run. Given such a high stringency of partitioning (very long separation), dissociation of protein–DNA complex might reduce the level of intact complex to below the noise of PCR. Therefore, one has to adjust stringency of partitioning to achieve reasonably low DNA background while

Figure 4. The effect of the molecular weight of protein (MW_P) on the transmittance of CE-based partitioning for nonbinders (k_N) under the conditions of NECEEM (colored lines) and IFCE (black line).

maintaining a sufficient quantity of intact protein–DNA complexes. The extent of P–DNA dissociation in different CE-based partitioning modes will be evaluated in the next section.

Influence of MW_P on k_B . During CE-based partitioning, protein–DNA complexes dissociate at a certain rate; thus, in principle, k_B is governed mainly by the dissociation rate constant k_{off} of protein–DNA complexes and elution time t :

$$k_B = \frac{B_{out}}{B_{in}} = \frac{B_{in} e^{-k_{off} t}}{B_{in}} = e^{-k_{off} t} \quad (13)$$

The values of t as well as the binder-collection windows for complexes of 80-nt DNA and protein targets of different MW_P were estimated using the predicted mobility values of protein–DNA complexes obtained with eq 12. The transmittance of partitioning for binders as defined by eq 13 was estimated for two values of k_{off} , 10^{-3} and 10^{-4} s $^{-1}$, for which the characteristic complex-dissociation times ($\tau = 1/k_{off}$) are 20 min and ~ 3 h, respectively. The values of k_{off} outside of the 10^{-4} – 10^{-3} s $^{-1}$ range are more likely to be unpreferred for therapeutic targets since the complex either dissociate too quickly ($\tau = 1$ min for $k_{off} = 10^{-2}$ s $^{-1}$) or remain stable for too long ($\tau = 28$ h for $k_{off} = 10^{-5}$ s $^{-1}$).⁴¹

The predicted dependence of k_B on MW_P is shown in Figure 5. For $k_{off} = 10^{-3}$ s $^{-1}$, decreasing MW_P from 150 to 25 kDa leads to decreasing the value of k_B by up to 3 orders of magnitude in IFCE. In contrast, in NECEEM, k_B is not affected by changing MW_P . Notably, in IFCE, k_B is predicted to be lower than 10^{-3} when $MW_P < 30$ kDa, meaning that less than 0.1% of the total quantity of protein–DNA complex in the equilibrium mixture would survive separation until elution.

For $k_{off} = 10^{-4}$ s $^{-1}$, the decrease in k_B is much less pronounced for both IFCE and NECEEM. In both NECEEM sub-modes, k_B remains relatively stable over the specified range of MW_P with more than 80% of protein–DNA complexes reaching the capillary end intact. This finding agrees with eq 11, in which the k_B value in NECEEM is assumed to be constant. In IFCE, our prediction shows that up to 60% of protein–DNA complexes would dissociate as the values of MW_P decrease to 25 kDa. Overall, the dissociation of protein–DNA complexes is much less pronounced in NECEEM than in IFCE over the specified range of MW_P . On the other hand, k_N value in IFCE is orders of magnitudes lower than in NECEEM. A balance between k_B and k_N must be achieved in order to obtain a target level of binder purity Q_2 after partitioning.

Influence of MW_P on the number of partitioning rounds.

The odds of successful selections depend on two major conditions. The first condition is a high value of efficiency of partitioning (k_B/k_N) to enrich the initial library with a low level of binder abundance (B_{in}/N_{in}) to a desirable level of binder purity at the output ($B_{out}/N_{out} > Q_2$). The second condition is a sufficient quantity of binders at the input (B_{in}) so that the output quantity of

Figure 5. The dependency of the transmittance of CE-based partitioning for binders (k_B) on the molecular weight of target protein (MW_P) under the conditions of NECEEM (colored lines) and IFCE (black lines). The value of k_B was estimated based on two k_{off} values: 10^{-3} s $^{-1}$ and 10^{-4} s $^{-1}$. In the graph with k_{off} of 10^{-4} s $^{-1}$, the inset shows the same data but with a linear y-scale for k_B .

binder (B_{out}) can exceed the PCR noise. The efficiency of partitioning for IFCE and NECEEM can be derived from the k_N and k_B values for the two preferred k_{off} values considered above: 10^{-3} and 10^{-4} s $^{-1}$. To estimate B_{in}/N_{in} , we hypothesized different scenarios for aptamer selection based on the affinity of random DNA library to the target protein: (i) high abundance of binders ($B_{in}/N_{in} = 10^{-5}$ or approximately 1 binder per 10^5 nonbinders), (ii) moderate abundance of binders ($B_{in}/N_{in} = 10^{-6.5}$ or ~ 1 binder per 3×10^6 nonbinders), and (iii) low abundance of binders ($B_{in}/N_{in} < 10^{-8}$ or less than 1 binder per 10^8 nonbinders).

We then estimated the range of required partitioning rounds (m) to obtain a binder-enriched pool at the output of non-SELEX (without PCR amplification of the collected pools between the rounds) with each of NECEEM sub-modes and IFCE. The upper limit of m (m_{max}) and the lower limit of m (m_{min}) were calculated using eqs 5 and 6, respectively, for $Q_1 = 100$ (i.e., B_{out} exceeds PCR noise of 120 molecules of DNA by a factor of 100) and $Q_2 = 1$ (i.e., binders constitute 50% of the final DNA pool). In principle, a high value of m_{max} indicates a low level of binder losses throughout the selection process; therefore, many rounds of partitioning can be conducted to further enrich the pool without detrimental losses of binders. On the other hand, a low value of m_{min} is preferable to minimize the number of partitioning rounds required to reach a target level of binder purity in the resulting pool.

In the case of classic aptamer selection by SELEX, the estimation of m_{min} requires the knowledge of PCR bias (Z_B/Z_N) and the number of PCR cycles in a single round (eq 8). These two parameters vary greatly depending on which type of PCR method is used and how well the PCR procedure is done. Optimal PCR conditions (e.g., when Z_B/Z_N is close to 1) can be achieved by using an unbiased PCR procedure (e.g., emulsion PCR) with an optimal number of PCR cycles.³⁶ Values of m_{min} for non-SELEX presented in this section are also applicable to SELEX under optimal PCR conditions. Note that when there is a PCR bias, m_{min} in SELEX will be higher than in non-SELEX.

Figure 6 shows the predicted dependence of acceptable numbers of partitioning rounds in IFCE and NECEEM (both sub-modes) on MW_P for 3 different values of binder abundance in the initial library (B_{in}/N_{in}). Suitability of the partitioning method for selection can be assessed based on the following criteria: (i) a range of m is wide (m_{max} and m_{min} points are far apart on the graph), (ii) m_{max} is greater than m_{min} (m_{max} point is not lower than m_{min} point on the graph) and (iii) m_{max} is not smaller than 1 (the otherwise case indicates insufficient quantity of binders at the

a) High binder abundance: ≈ 1 binder per 10^5 nonbinders

b) Moderate binder abundance: ≈ 1 binder per 3×10^6 nonbinders

c) Low binder abundance: < 1 binder per 10^8 nonbinders

— $m_{\max}$ (upper limit of m)
— $m_{\min}$ (lower limit of m)

Figure 6. The dependence of required ($m_{\min}$) and allowed ($m_{\max}$) numbers of partitioning rounds on the molecular weight of target protein in non-SELEX selection of aptamers using different modes of CE-based partitioning. See text for details.

output). Based on Figure 6, CE-based selection of aptamers can be successful in the following scenarios: (i) the input library has high binder abundance ($\sim 10^{-5}$) or (ii) the input library has moderate binder abundance ($> 10^{-8}$) with a bulk value of $k_{\text{off}} \leq 10^{-4}$. In these scenarios, as MW_P increases from 25 to 150 kDa, the range of m increases in IFCE and complex-first NECEEM. This means that the extent of binder losses is smaller (higher $m_{\max}$) and fewer partitioning rounds are required (smaller $m_{\min}$) for larger protein targets in IFCE and complex-first NECEEM. On the other hand, in complex-last NECEEM, both $m_{\max}$ and $m_{\min}$ decrease as MW_P increases from 25 to 150 kDa. In complex-last NECEEM, the elution time of protein–DNA complex increases with increasing MW_P ; thus, the extent of binder losses is higher while fewer partitioning rounds are allowed for larger protein targets. Moreover, as both NECEEM sub-modes have higher range of k_N values (Figure 4), their range of $m_{\min}$ values is also higher than that in IFCE. This relation means that more partitioning rounds in NECEEM would be required to enrich the pool to a certain level of binder purity as compared to IFCE (particularly, complex-last NECEEM with the highest range of $m_{\min}$ values requires more partitioning rounds than complex-first NECEEM).

Despite the low value of k_B for small protein targets, the extremely low k_N values in IFCE suggest that IFCE could support the enrichment of binders in a single step of partitioning (50% binders in the resulting pool) over the whole range of MW_P . However, in IFCE, low values of $m_{\max}$ were observed for small-size protein targets, indicating that excessive losses of binders can potentially hinder the success of IFCE-based selection of aptamers for such targets. In some cases, the $m_{\max}$ value in IFCE were unacceptable ($m_{\max} \leq m_{\min}$ and/or $m_{\max} \leq 1$) for protein targets with the following ranges of molecular weights: (i) $MW_P < 60$ kDa for the input library with high binder abundance and $k_{\text{off}} = 10^{-3} \text{ s}^{-1}$, and (ii) $MW_P < 20$ kDa for the input library with higher binder abundance and $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$ and (iii) $MW_P < 25$ kDa for the input library with moderate binder

abundance and $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$. The latter two were extrapolated by assuming a constant k_N in the range of 10^{-9} for protein–DNA complex with MW_P ranging from 15 to 25 kDa in IFCE (see Note S5).

To ensure the collection of a sufficient quantity of binders at the output, NECEEM is the method of choice for selection of binders for small-size protein targets. This statement is especially true in non-SELEX, in which no PCR amplification is used between the rounds of partitioning to compensate for the dilution-induced losses of binders between the rounds. Our results suggest that IFCE is not preferred for small-size protein targets due to the excessive binder losses within every round of partitioning owing to complex dissociation. However, IFCE is the most suitable method for large-size protein targets to obtain high affinity binders in a minimal number of partitioning rounds (as few as a single round as our data suggest). Another important conclusion from data shown in Figure 6 is that both NECEEM and IFCE fail to retain a sufficient quantity of binders after one round of partitioning when the binder abundance in the initial library is as low as 10^{-8} or 1 binder per one hundred million nonbinders ($m_{\max} < 1$ in all cases in Figure 6C). When the bulk value of k_{off} is high ($k_{\text{off}} = 10^{-3} \text{ s}^{-1}$), this lower limit of binder abundance in the initial library for successful CE-based selection increases to $10^{-6.5}$ or $\sim 3 \times 10^{-7}$ ($m_{\max} \leq m_{\min}$ in all cases with $k_{\text{off}} = 10^{-3} \text{ s}^{-1}$ in Figure 6B).

Single-round IFCE-based selection and multi-round NECEEM-based selection will certainly fail when there is not enough B_{in} for PCR to reliably detect and amplify B_{out} . Thus, for less “aptagenic” target proteins, efforts must be made to increase B_{in} via using an initial library with higher binder abundance and/or increasing the input quantity of the initial library. The latter is limited by the maximum concentration of DNA library and the length of the injected sample plug. In our CE experiments, we used the highest possible concentration of DNA library in the final equilibrium mixture with the sample plug length of 1 cm. Our preliminary data suggest that increasing the sample-plug

length by an order of magnitude (from 1 to 10 cm) increases k_N by multiple orders of magnitude, resulting in insufficient separation of DNA nonbinders from P-DNA complexes. Due to such an inherent limitation on the size of the injected sample in CE-based partitioning, the ultimate solution to improve the success rate of aptamer selection for less aptogenic target proteins is to use DNA libraries with higher binder abundance, such as modified oligonucleotide libraries with functionalized protein-like groups.⁴²⁻⁴⁵ Selection of aptamers from modified DNA libraries have yielded high-affinity aptamers to many difficult-to-select-for proteins that had repeatedly failed SELEX with unmodified DNA libraries.⁴⁴ Application of CE-based partitioning to selection of aptamers from modified oligonucleotide libraries is a promising direction of further development of this partitioning approach.

CONCLUSION

In summary, the size of protein target dictates the choice of the mode of CE-based partitioning in aptamer selection. In NECEEM, as the size of protein target decreases, k_N increases by several orders of magnitude while the k_B is relatively stable and close to unity. On the other hand, IFCE improves k_N values (which are as low as 10^{-9} and do not change much with varying size of protein target) at the expense of sacrificing k_B . The k_B values in IFCE decrease drastically when the size of protein target decreases, thus, hindering the collection of sufficient quantity of intact complexes for small-size protein targets. To ensure obtaining a pool of binder of sufficient purity and quantity, one must find a balance between k_N and k_B to determine the most suitable mode of CE-based partitioning and corresponding running buffer. For large-size protein targets, IFCE is preferred in order to obtain high affinity aptamers in fewer rounds of partitioning (single-round selection could be possible as our data suggest). We recommend that NECEEM be used in selection of aptamers for small-size protein targets, especially in non-SELEX, in which there is no PCR amplification of the collected pools between the rounds of partitioning. Between the two sub-modes of NECEEM, the complex-first sub-mode is proven to facilitate selection of aptamer in fewer rounds due to lower k_N values. However, the use of complex-last NECEEM is still beneficial when the adsorption of some protein targets to the uncoated inner capillary surface is severe and detrimental for the selection; the coating of the walls can suppress such adsorption. For more difficult protein targets (e.g., those that had repeatedly failed SELEX), CE-based selection fails when the binder abundance in the initial library is as low as 10^{-8} . Under such circumstances, using a better library, such as chemically modified DNA library with a greater abundance of binders, could be a promising direction for future development in the field of highly efficient CE-based aptamer selection.

REFERENCES

- (1) Ellington, A. D.; Szostak, J. W. *In Vitro* Selection of RNA Molecules that Bind Specific Ligands. *Nature* **1990**, *346*, 818–822.
- (2) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment - RNA Ligands to Bacteriophage-T4 DNA-Polymerase. *Science* **1990**, *249*, 505–510.
- (3) Quang, N. N.; Miodok, A.; Cibiel, A.; Duconge, F. Selection of Aptamers Against Whole Living Cells: From Cell-SELEX to Identification of Biomarkers. *Methods Mol. Biol.* **2017**, *1575*, 253–272.
- (4) Navani, N. K.; Mok, W. K.; Yingfu, L. In Vitro Selection of Protein-binding DNA Aptamers as Ligands for Biosensing Applications. *Methods Mol. Biol.* **2009**, *504*, 399–415.
- (5) Li, L.; Jiang, Y.; Cui, C.; Yang, Y.; Zhang, P.; Stewart, K.; Pan, X.; Li, X.; Yang, L.; Qiu, L.; Tan, W. Modulating Aptamer Specificity with pH-Responsive DNA Bonds. *J. Am. Chem. Soc.* **2018**, *140*, 13335–13339.
- (6) Durand, G.; Lisi, S.; Ravelet, C.; Dausse, E.; Peyrin, E.; Toulme, J. J. Riboswitches Based on Kissing Complexes for the

- Detection of Small Ligands. *Angew. Chem. Int. Edit.* **2014**, *53*, 6942–6945.
- (7) Keefe, A. D.; Pai, S.; Ellington, A. Aptamers as Therapeutics. *Nat. Rev. Drug Discov.* **2010**, *9*, 537–550.
- (8) German, I.; Buchanan, D. D.; Kennedy, R. T. Aptamers as Ligands in Affinity Probe Capillary Electrophoresis. *Anal. Chem.* **1998**, *70*, 4540–4545.
- (9) Zhang, H. Q.; Li, F.; Dever, B.; Li, X. F.; Le, X. C. DNA-Mediated Homogeneous Binding Assays for Nucleic Acids and Proteins. *Chem. Rev.* **2013**, *113*, 2812–2841.
- (10) Zhang, H. Q.; Wang, Z. W.; Li, X. F.; Le, X. C. Ultrasensitive Detection of Proteins by Amplification of Affinity Aptamers. *Angew. Chem. Int. Edit.* **2006**, *45*, 1576–1580.
- (11) Yi, M.; Yang, S.; Peng, Z.; Liu, C.; Li, J.; Zhong, W.; Yang, R.; Tan, W. Two-photon Graphene Oxide/Aptamer Nanosensing Conjugate for *In Vitro* or *In Vivo* Molecular Probing. *Anal. Chem.* **2014**, *86*, 3548–3554.
- (12) Zhu, C.; Li, L.; Wang, Z.; Irfan, M.; Qu, F. Recent Advances of Aptasensors for Exosomes Detection. *Biosens. Bioelectron.* **2020**, *160*, 112213.
- (13) Yuce, M.; Ullah, N.; Budak, H. Trends in Aptamer Selection Methods and Applications. *Analyst* **2015**, *140*, 5379–5399.
- (14) Darmostuk, M.; Rimpelova, S.; Gbelcova, H.; Ruml, T. Current Approaches in SELEX: An Update to Aptamer Selection Technology. *Biotechnol. Adv.* **2015**, *33*, 1141–1161.
- (15) Berezovski, M.; Musheev, M.; Drabovich, A.; Krylov, S. N. Non-SELEX Selection of Aptamers. *J. Am. Chem. Soc.* **2006**, *128*, 1410–1411.
- (16) Berezovski, M. V.; Musheev, M. U.; Drabovich, A. P.; Jitkova, J. V.; Krylov, S. N. Non-SELEX: Selection of Aptamers without Intermediate Amplification of Candidate Oligonucleotides. *Nat. Protoc.* **2006**, *1*, 1359–1369.
- (17) Lisi, S.; Fiore, E.; Scarano, S.; Pascale, E.; Boehman, Y.; Ducongé, F.; Chierici, S.; Minunni, M.; Peyrin, E.; Ravelet, C. Non-SELEX Isolation of DNA Aptamers for the Homogeneous-phase Fluorescence Anisotropy Sensing of tau Proteins. *Anal. Chim. Acta* **2018**, *1038*, 173–181.
- (18) Le, A. T. H.; Krylova, S. M.; Beloborodov, S. S.; Wang, T. Y.; Hili, R.; Johnson, P. E.; Li, F.; Veedu, R. N.; Belyanskaya, S.; Krylov, S. N. How to Develop and Prove High-Efficiency Selection of Ligands from Oligonucleotide Libraries: A Universal Framework for Aptamers and DNA-Encoded Small-Molecule Ligands. *Anal. Chem.* **2021**, *93*, 5343–5354.
- (19) Irvine, D.; Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment with Integrated Optimization by Non-Linear Analysis. *J. Mol. Biol.* **1991**, *222*, 739–761.
- (20) Ciesiolka, J.; Illangasekare, M.; Majerfeld, I.; Nickles, T.; Welch, M.; Yarus, M.; Zinnen, S. Affinity Selection-amplification from Randomized Ribooligonucleotide Pools. *Methods Enzymol.* **1996**, *267*, 315–335.
- (21) Papoulas, O. Rapid Separation of Protein-Bound DNA from Free DNA Using Nitrocellulose Filters. *Curr. Protoc. Mol. Biol.* **2001**, *36*, 12.8.1–12.8.9.
- (22) Wang, J.; Rudzinski, J. F.; Gong, Q.; Soh, H. T.; Atzberger, P. J. Influence of Target Concentration and Background Binding on *In Vitro* Selection of Affinity Reagents. *PLoS One* **2012**, *7*, No. e43940.
- (23) Hamula, C. L. A.; Zhang, H. Q.; Guan, L. L.; Li, X. F.; Le, X. C. Selection of Aptamers against Live Bacterial Cells. *Anal. Chem.* **2008**, *80*, 7812–7819.
- (24) Berezovski, M.; Drabovich, A.; Krylova, S. M.; Musheev, M.; Okhonin, V.; Petrov, A.; Krylov, S. N. Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures: A Universal Tool for Development of Aptamers. *J. Am. Chem. Soc.* **2005**, *127*, 3165–3171.
- (25) Le, A. T. H.; Krylova, S. M.; Kanoatov, M.; Desai, S.; Krylov, S. N. Ideal-Filter Capillary Electrophoresis (IFCE) Facilitates the One-Step Selection of Aptamers. *Angew. Chem. Int. Edit.* **2019**, *58*, 2739–2743.
- (26) Zhu, C.; Li, L.; Yang, G.; Irfan, M.; Wang, Z.; Fang, S.; Qu, F. High-efficiency Selection of Aptamers for Bovine Lactoferrin by Capillary Electrophoresis and its Aptasensor Application in Milk Powder. *Talanta* **2019**, *205*, 120088.
- (27) Riley, K. R.; Saito, S.; Gagliano, J.; Colyer, C. L. Facilitating Aptamer Selection and Collection by Capillary Transient

Isotachopheresis with Laser-induced Fluorescence Detection. *J. Chromatogr. A* **2014**, *1368*, 183–189.

(28) Mendonsa, S. D.; Bowser, M. T. *In Vitro* Evolution of Functional DNA Using Capillary Electrophoresis. *J. Am. Chem. Soc.* **2004**, *126*, 20–21.

(29) Mendonsa, S. D.; Bowser, M. T. *In Vitro* Selection of High-Affinity DNA Ligands for Human IgE Using Capillary Electrophoresis. *Anal. Chem.* **2004**, *76*, 5387–5392.

(30) Mosing, R. K.; Mendonsa, S. D.; Bowser, M. T. Capillary Electrophoresis-SELEX Selection of Aptamers with Affinity for HIV-1 Reverse Transcriptase. *Anal. Chem.* **2005**, *77*, 6107–6112.

(31) Drabovich, A. P.; Berezovski, M.; Okhonin, V.; Krylov, S. N. Selection of Smart Aptamers by Methods of Kinetic Capillary Electrophoresis. *Anal. Chem.* **2006**, *78*, 3171–3178.

(32) Krylova, S. M.; Karkhanina, A. A.; Musheev, M. U.; Bagg, E. A. L.; Schofield, C. J.; Krylov, S. N. DNA Aptamers for as Analytical Tools for the Quantitative Analysis of DNA-dealkylating Enzymes. *Anal. Biochem.* **2011**, *414*, 261–265.

(33) Galievsky, V. A.; Stasheuski, A. S.; Krylov, S. N. Improvement of LOD in Fluorescence Detection with Spectrally Nonuniform Background by Optimization of Emission Filtering. *Anal. Chem.* **2017**, *89*, 11122–11128.

(34) de Jong, S.; Krylov, S. N. Pressure-Based Approach for the Analysis of Protein Adsorption in Capillary Electrophoresis. *Anal. Chem.* **2012**, *84*, 453–458.

(35) Warnecke, P. M.; Stirzaker, C.; Melki, J. R.; Millar, D. S.; Paul, C. L.; Clark, S. J. Detection and Measurement of PCR Bias in Quantitative Methylation Analysis of Bisulphite-treated DNA. *Nucleic Acids Res.* **1997**, *25*, 4422–4426.

(36) Takahashi, M.; Wu, X. W.; Ho, M.; Chomchan, P.; Rossi, J. J.; Burnett, J. C.; Zhou, J. H. High Throughput Sequencing Analysis of RNA Libraries Reveals the Influences of Initial Library and PCR Methods on SELEX Efficiency. *Sci. Rep.* **2016**, *6*, 33697.

(37) Musheev, M. U.; Kanoatov, M.; Krylov, S. N. Non-Uniform Velocity of Homogeneous DNA in a Uniform Electric Field: Consequence of Electric-Field-Induced Slow Dissociation of Highly Stable DNA-Counterion Complexes. *J. Am. Chem. Soc.* **2013**, *135*, 8041–8046.

(38) Kochmann, S.; Le, A. T. H.; Hili, R.; Krylov, S. N. Predicting Efficiency of NECEEM-Based Partitioning of Protein Binders from Nonbinders in DNA-Encoded Libraries. *Electrophoresis* **2018**, *39*, 2991–2996.

(39) Beloborodov, S. S.; Krylova, S. M.; Krylov, S. N. Spherical-Shape Assumption for Protein–Aptamer Complexes Facilitates Prediction of Their Electrophoretic Mobility. *Anal. Chem.* **2019**, *91*, 12680–12687.

(40) Desruisseaux, C.; Long, D.; Drouin, G.; Slater, G. W. Electrophoresis of Composite Molecular Objects. 1. Relation between Friction, Charge, and Ionic Strength in Free Solution. *Macromolecules.* **2001**, *34*, 44–52.

(41) Dahl, G.; Akerud, T. Pharmacokinetics and the drug–target residence time concept. *Drug Discov.* **2013**, *18*, 697–707.

(42) Eaton, B. E. The Joys of *In Vitro* Selection: Chemically Dressing Oligonucleotides to Siate Protein Targets. *Curr. Opin. Chem. Biol.* **1997**, *1*, 10–16.

(43) Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E. Expanding the Chemistry of DNA for *In Vitro* Selection. *J. Am. Chem. Soc.* **2010**, *132*, 4141–4151.

(44) Gold, L.; *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **2010**, *5*, No. e15004.

(45) Kong, D. H.; Yeung, W.; Hili, R. *In Vitro* Selection of Diversely Functionalized Aptamers. *J. Am. Chem. Soc.* **2017**, *139*, 13977–13980.

SUPPORTING INFORMATION

Quantitative Characterization of Partitioning in Selection of DNA Aptamers for Protein Targets by Capillary Electrophoresis

An T. H. Le, Tong Ye Wang, Svetlana Krylova, Stanislav S. Beloborodov, and Sergey N. Krylov*

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

Table of Contents

Note S1: Dependence of Number of Partitioning Rounds on k_N & k_B S2

Note S2: Empirical Mathematical Model to Predict the Electrophoretic Mobility of Protein–DNA Complexes..... S3

Figure S1. Line of the best fit for the electrophoretic mobility of protein–ssDNA complex as a function of X : $\mu_{P-DNA} = A + BX$ S3

Note S3: Determination of Elution Time of Protein–DNA complex, the Binder-Collection Window, and k_N for a Given Value of MW_P S4

Figure S2. An example of binder-collection time window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM S4

Note S4: The Predicted Elution Time of Protein–DNA Complex with MW_P Ranging from 15 to 150 kDa in CE-Based Partitioning..... S5

Figure S3. The predicted elution time for the protein–DNA complex as a function of MW_P under conditions of NECEEM and IFCE S5

Note S5: The Dependence of m_{min} and m_{max} on MW_P Ranging from 15 to 150 kDa in IFCE for DNA Library with Moderate to High Binder Abundance and Bulk $k_{off} = 10^{-4} \text{ s}^{-1}$ S6

Figure S4. The dependence of m_{min} and m_{max} on MW_P in IFCE for DNA library with moderate and high binder abundance..... S6

Additional supplementary file

File name

Excel file with raw data and calculation.zip

Description

Raw data and calculations

Note S1: Dependence of Number of Partitioning Rounds on k_N & k_B

Multi-round selection without PCR amplification (non-SELEX):

$$\begin{aligned}
 B_{\text{out}} &= B_{\text{in}} k_B^m \\
 B_{\text{out}} &> Q_1 N_{\text{PCR}} \\
 B_{\text{in}} k_B^m &> Q_1 N_{\text{PCR}} \\
 k_B^m &> Q_1 N_{\text{PCR}} / B_{\text{in}} \\
 m \log(k_B) &> \log(Q_1 N_{\text{PCR}} / B_{\text{in}}) \\
 \because k_B < 1 \text{ and } (Q_1 N_{\text{PCR}} / B_{\text{in}}) < 1; \log(k_B) < 0 \text{ and } \log(Q_1 N_{\text{PCR}} / B_{\text{in}}) < 0 \\
 m &< \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log(k_B)}
 \end{aligned}$$

$$\begin{aligned}
 B_{\text{out}} / N_{\text{out}} &> Q_2 \\
 \because B_{\text{out}} &= B_{\text{in}} k_B^m \text{ and } N_{\text{out}} = N_{\text{in}} k_N^m \\
 B_{\text{in}} k_B^m / (N_{\text{in}} k_N^m) &> Q_2 \\
 (k_B / k_N)^m &> Q_2 N_{\text{in}} / B_{\text{in}} \\
 m \log(k_B / k_N) &> \log(Q_2 N_{\text{in}} / B_{\text{in}}) \\
 \because k_B / k_N > 1 \text{ and } Q_2 N_{\text{in}} / B_{\text{in}} < 1; \log(k_B / k_N) > 0 \text{ and } \log(Q_2 N_{\text{in}} / B_{\text{in}}) < 0 \\
 m &> \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log(k_B / k_N)}
 \end{aligned}$$

Multi-round selection with PCR amplification (SELEX):

$$\begin{aligned}
 B_{\text{out}} &= B_{\text{in}} \left(k_B (Z_B)^n \right)^m \\
 B_{\text{out}} &> Q_1 N_{\text{PCR}} \\
 B_{\text{in}} \left(k_B (Z_B)^n \right)^m &> Q_1 N_{\text{PCR}} \\
 \left(k_B (Z_B)^n \right)^m &> Q_1 N_{\text{PCR}} / B_{\text{in}} \\
 m \log \left(k_B (Z_B)^n \right) &> \log(Q_1 N_{\text{PCR}} / B_{\text{in}}), \\
 \because k_B (Z_B)^n > 1 \text{ and } (Q_1 N_{\text{PCR}} / B_{\text{in}}) < 1; \log \left(k_B (Z_B)^n \right) > 0 \text{ and } \log(Q_1 N_{\text{PCR}} / B_{\text{in}}) < 0 \\
 m &> \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log \left(k_B (Z_B)^n \right)}
 \end{aligned}$$

$$\begin{aligned}
 (k_B / k_N) (Z_B / Z_N)^n &> 1 \\
 (Z_B / Z_N)^n &> k_N / k_B \\
 \log \left((Z_B / Z_N)^n \right) &> \log(k_N / k_B) \\
 n \log(Z_B / Z_N) &> \log(k_N / k_B), \\
 \because Z_B / Z_N < 1 \text{ and } k_N / k_B < 1; \log(Z_B / Z_N) < 0 \text{ and } \log(k_N / k_B) < 0 \\
 n &< \left\lfloor \frac{\log(k_N / k_B)}{\log(Z_B / Z_N)} \right\rfloor
 \end{aligned}$$

Note S2: Empirical Mathematical Model to Predict the Electrophoretic Mobility of Protein–DNA Complexes

The mobility of the protein–DNA complex is linked with the molecular weight of the complex (MW_{P-DNA}) based on the following empirical equation (Beloborodov, S. S.; Krylova, S. M.; Krylov, S. N. Spherical-Shape Assumption for Protein–Aptamer Complexes Facilitates Prediction of Their Electrophoretic Mobility. *Anal. Chem.* 2019, *91*, 12680–12687):

$$\mu_{P-DNA} = A + B\mu_{DNA}L_{DNA}^{0.68}MW_{P-DNA}^{-1/3} \quad (S1)$$

where electrophoretic mobilities are expressed in $\text{mm}^2\text{kV}^{-1}\text{s}^{-1}$, L_{DNA} is expressed in the number of nucleotides, MW_{P-DNA} (sum of MW_P and MW_{DNA}) is expressed in kDa, while A and B are empirical constants. The constant A and B for running buffer with $I < 50$ mM were published previously: $A = -9.95 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.0929 \text{ kDa}^{1/3}$.

Since μ_{DNA} is dependent on I , the new empirical constants A and B were re-established for running buffer with $I = 146$ mM. Linear fitting of experimental mobility data for six protein–DNA complexes with eq S1 resulted in $A = 10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.2365 \text{ kDa}^{1/3}$ with a correlation coefficient (R^2) of 0.946 (Figure S1).

Figure S1. Line of the best fit for the electrophoretic mobility of protein–ssDNA complex as a function of X : $\mu_{P-DNA} = A + BX$, where $X = \mu_{P-DNA}L_{DNA}^{0.68}MW_{P-DNA}^{-1/3}$. Calculated values for A and B were $10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and 0.2365 , respectively. The correlation coefficient was $R^2 = 0.946$.

Note 3: Determination of Elution Time of Protein–DNA complex, the Binder-Collection Window, and k_N for a Given Value of MW_P

The time of separation (or elution time) of protein–DNA complex (t) for a given value of MW_P was estimated using the mobilities value obtained from eq S1. For both NECEEM sub-modes ($I < 50$ mM), the A and B constants are $-9.95 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $0.0929 \text{ kDa}^{1/3}$ respectively. For IFCE ($I = 146$ mM), the A and B constants are $10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $0.2365 \text{ kDa}^{1/3}$, respectively. The binder-collection time window for a given value of MW_P was calculated as the elution time $\pm 5\%$. An example of determining the binder-collection time window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM is given below (Figure S2). Detailed calculations can be found from the attached Excel file.

Figure S2. An example of binder-collection time window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM. The black trace indicates the DNA background profile (DNA quantity versus elution time under qPCR detection) in complex-first NECEEM. The predicted elution time (predicted t) for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM was estimated to be 16.5 min. The binder-collection window was defined as $(16.5 \pm 5\%)$ min or 15.7 to 17.3 min. The double-headed arrow indicates the defined binder-collection window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM. The associated k_N for this protein–DNA complex was calculated as the integral under the DNA-background-profile curve within the binder-collection time window divided by the total quantity of DNA sampled into the capillary.

Note S4: The Predicted Elution Time of Protein–DNA Complex with MW_P Ranging from 15 to 150 kDa in CE-Based Partitioning

The predicted elution times of protein–DNA complexes with MW_P ranging from 15 to 150 kDa are shown in **Figure S3**. For both NECEEM sub-modes, the predicted elution times of protein–DNA complexes are within 25 min over the whole specified range of MW_P . For IFCE, the elution times of the complexes are highly sensitive to the variation in MW_P . When $MW_P < 25$ kDa, the predicted elution time of protein–DNA complexes is beyond 3 h in IFCE.

Figure S3. The predicted elution time for the protein–DNA complex as a function of MW_P under conditions of NECEEM and IFCE.

Note 5: The Dependence of $m_{\min}$ and $m_{\max}$ on MW_P Ranging from 15 to 150 kDa in IFCE for DNA Library with Moderate to High Binder Abundance and Bulk $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$

The values of $m_{\min}$ and $m_{\max}$ were calculated using eqs Error! Reference source not found. and Error! Reference source not found. in the main-text, respectively, for $Q_1 = 100$ (*i.e.*, B_{out} exceeds PCR noise of 120 molecules of DNA by a factor of 100) and $Q_2 = 1$. As k_N cannot be measured experimentally for target–binder complexes with $MW_P < 25$ kDa under IFCE conditions due to an unreasonably long CE run, the values of k_N for $15 \text{ kDa} < MW_P < 25 \text{ kDa}$ were assumed to be constant (in the range of 10^{-9}) and equal to the experimental k_N value obtained for protein–DNA complex with $MW_P = 25 \text{ kDa}$ (see the attached Excel file).

Figure S4. The dependence of $m_{\min}$ and $m_{\max}$ on MW_P in IFCE for DNA library with moderate (a) and high binder abundance (b). The values of k_B were estimated based on a bulk k_{off} of 10^{-4} s^{-1} .