analytical chemistry

Article

Cite This: *Anal. Chem.* 2019, 91, 12680−12687

pubs.acs.org/ac

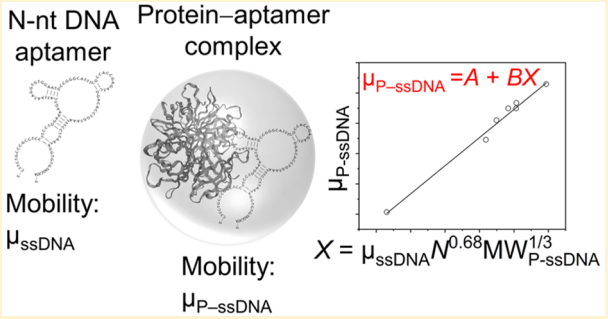
Spherical-Shape Assumption for Protein−Aptamer Complexes Facilitates Prediction of Their Electrophoretic Mobility

Stanislav S. Beloborodov, Svetlana M. Krylova, and Sergey N. Krylov*

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

Supporting Information

ABSTRACT: DNA aptamers are single-strand DNA (ssDNA) capable of selectively and tightly binding a target molecule. Capillary electrophoresis-based selection of aptamers for protein targets requires the knowledge of electrophoretic mobilities of protein−aptamer complexes, while measuring these mobilities requires having the aptamers. Here, we report on breaking this vicious circle. We introduce a mathematical model that allows prediction of protein−aptamer complex mobility, while requiring only three easy-to-determine input parameters: the number *N* of nucleotides in the aptamer, electrophoretic mobility of *N*-nucleotide-long ssDNA, and a sum molecular weight of the protein−aptamer complex. The model was derived upon simplifying assumptions of a spherical shape of the protein−aptamer complex. According to this model, the protein−aptamer complex mobility is a linear function of a combination of the three input parameters with empirically determined line's intercept and slope. The intercept and slope were determined using experimental data for seven complexes. The model was then cross-validated with the leave-one-out approach revealing only 2% residual standard deviations for both the slope and the intercept. Such a precise determination of these constants allowed accurate mobility prediction for the excluded complexes with only a 3% maximum deviation from the experimentally determined mobilities. The model was tested by applying it to three protein−aptamer complexes that were not a part of the training/cross-validation set; deviations of the predicted mobilities from the experimentally determined ones were within 5% of the latter. To complete this study, the model was fine-tuned using the 10 complexes. Our results strongly suggest the validity of the spherical-shape assumption for the protein−aptamer complexes when considering complex mobility. The developed model will make it possible to rationally design capillary electrophoresis-based selection of DNA aptamers for protein targets.



DNA (or RNA) aptamers are capable of binding their protein targets with high selectivity and affinity;[1] therefore, they are used to develop diagnostic probes, drug-delivery vehicles, and drugs.[2−4] DNA aptamers are selected from random-sequence ssDNA libraries by incubating the library with the target and then partitioning protein-bound DNA from unbound DNA. More than 70% of attempts to select aptamers fail;[5] inefficient partitioning is among the major culprits for these failures.[6]

The partitioning of protein-bound DNA from unbound DNA is typically performed with the protein immobilized on the surface of magnetic beads.[7] The efficiency of surface-based partitioning is typically below 10^2.[8] Such a low efficiency of partitioning makes it virtually impossible to select aptamers in a single round. Capillary electrophoresis (CE) facilitates homogeneous separation and, thus, provides a powerful alternative for surface-based partitioning. In classical CE-based partitioning, for example, by nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM), protein−DNA complexes move in the same direction as unbound DNA and the efficiency of partitioning is approximately 10^5.[8] In recently introduced ideal-filter CE (IFCE), the complexes move toward the capillary outlet while unbound DNA moves backward to the inlet.[9] The efficiency of IFCE-based partitioning reaches 10^9, which is presumed to be sufficient for reliable selection of aptamers in a single step of partitioning.

Both, NECEEM and IFCE require collection of fraction(s) of protein−ssDNA complexes, separated from unbound ssDNA by CE, followed by isolation and amplification of the collected ssDNA. Precise fraction collection benefits from the knowledge of the complex position in the capillary. If the complex can be optically detected, then collecting fractions containing the complex is simple.[8] However, if the complex is optically undetectable (due to its small amount) or if its optical signal interferes with signals from other species in the mixture, then the task becomes much more complicated. In principle, the position of the complex in the capillary can be predicted theoretically if the value of its electrophoretic mobility is known. The mobility of protein−aptamer complexes changes

Received: April 28, 2019
Accepted: September 17, 2019
Published: September 17, 2019

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

significantly with varying sizes of the protein and DNA, and it should be determined *de novo* for every new combination of the protein and the DNA library. Experimentally measuring the mobility of the protein−aptamer complex requires having the aptamers selected, which creates the vicious circle: selecting aptamers requires knowing the mobility, while measuring the mobility requires having the aptamers selected. A logical approach to breaking the circle is to find the mobility of the protein−aptamer complex theoretically by using a reliable mobility model. A suitable model is, however, not available.

Two major existing mobility models for protein−DNA complexes assume that DNA is either a rigid rod that has minor flexibility or a long free-draining polymer chain that undergoes no structural changes while moving through the solvent.[10−12] Further, they assume a single-point attachment of the end of the DNA molecule to the protein (Figure 1a,b).

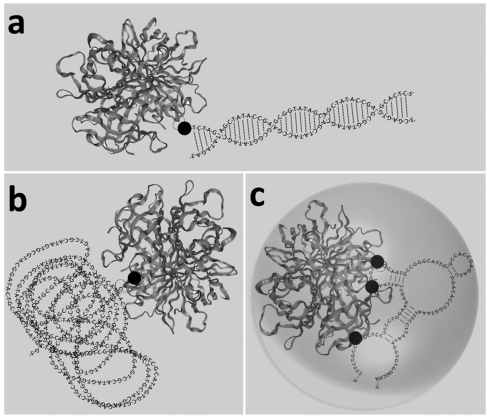


Figure 1. Schematic illustration of protein−DNA complexes with respect to three mobility models: rigid-rod DNA model (a), free-DNA model (b), and aptamer model (c). Black dots indicate the attachment points of DNA to the protein. The image of the protein (streptavidin) was generated using SWISS-MODEL software.[13] See text for details.

These assumptions are inapplicable to protein−aptamer complexes, in which ssDNA is linked to the protein via multiple bonds and retains a specific aptamer conformation (Figure 1c). Hence, the existing mobility models for protein−DNA complexes are inapplicable to protein−aptamer complexes.

Here, we present a mobility model that is designed to satisfy the close multipoint contact between a protein and an aptamer (Figure 1c). Our model assumes a spherical shape for such tight protein−aptamer complexes. By using basic rules governing migration of molecules in gel-free electrophoresis, we derived the dependence of the mobility of a spherical protein−ssDNA complex ($\mu_{P-ssDNA}$) on the mobility of ssDNA (μ_{ssDNA}), the number of nucleotides in ssDNA (N), the molecular weight of the complex ($MW_{P-ssDNA}$), and two empirical parameters (A and B)

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

where 0.68, the power of N, was previously determined by Slater and coauthors.[14] The values of A and B were found by fitting eq 1 into the experimental dependence of $\mu_{P-ssDNA}$ on $\{\mu_{ssDNA} N^{0.68} MW_{P-ssDNA}{}^{-1/3}\}$ obtained for seven different protein−ssDNA complexes.

The empirical model was successfully cross-validated by excluding one protein−DNA complex from the set of seven and finding pairs of A and B for seven subsets of six protein−DNA complexes (leave-one-out cross validation, LOOCV). The resulting seven pairs of A and B were close to each other, and residual standard deviations of A and B were only 2.2% and 1.7%, respectively. Such a precise determination of these constants allowed accurate mobility prediction for the excluded protein−DNA complexes with only a 3.0% maximum deviation from the experimentally determined mobility values. The model was tested on a set of three protein−aptamer complexes which were not part of the training/cross-validation set: human α-thrombin with its 29-nt aptamer, ABH2 with its 80-nt aptamer, and AlkB with its 80-nt aptamer (the complex with AlkB was analyzed at a pH value different from that used for the initial set of experiments with this protein). In all cases, the difference between the predicted mobility and the experimental one was below 5%, which confirmed model validity. The developed model will be indispensable in rational design of CE-based selection of DNA aptamers for protein targets. Moreover, it strongly suggests the validity of the spherical-shape assumption for the protein−aptamer complexes when considering complex mobility.

MATERIALS AND METHODS

Chemicals and Materials. All chemicals were purchased from Sigma-Aldrich (Oakville, ON), unless otherwise stated. A fused silica capillary with inner and outer diameters of 75 and 360 μm, respectively, was purchased from Polymicro (Phoenix, AZ). Chromeo P503 pyrylium dye was purchased from Active Motif (Burlington, ON). MutS and AlkB proteins were expressed and purified by us according to the published procedures.[15] SSB was purchased from Promega (Madison, WI). Streptavidin was purchased from New England Biolabs (Whitby, ON). ABH2 protein was kindly provided by Professor Christopher J. Schofield (University of Oxford, England). All DNA sequences were custom-synthesized by IDT (Oralville, IA). The sequences of ssDNA used in experiments with SSB were as follows: 5′-/Alex488N/TAA TAT TAT TGC AAA TAA ATT TAA A-3′ (25-nt); 5′-/Alex488N/AAA TTA AAG GGA ACG TAT ATA CAA CGC AAA GAA GCT GGA AAA TTG GCG AGA GAA TCT TCT TTC TGT CTA TCG AAG-3′ (75-nt); 5′-/56-FAM/CTT CTG CCC GCC TCC TTC CTG GTA AAG TCA TTA ATA GGT GTG GGG TGC CGG GCA TTT CGG AGA CGA GAT AGG CGG ACA CT-3′ (80-nt); 5′-/56-FAM/CTC CTC TGA CTG TAA CCA CGT GCC TAG CGT TTC ATT GTC CCT TCT TAT TAG GTG ATA ATA GCA TAG GTA GTC CAG AAG CCG ACA ACT G-3′ (88-nt). The sequence of MutS aptamer was exactly the same as 80-nt ssDNA sequence used for experiments with SSB. The sequence of 80-nt AlkB aptamer was 5′-/56-FAM/CTC CTC TGA CTG TAA CCA CGT GCC TAG CGT TTC ATT GTC CCT TCT TAT TAG GTG ATA ATA GCA TAG GTA GTC CAG AAG CC-3′. The sequence of the 80-nt ABH2 aptamer was 5′-/56-FAM/CTC CTC TGA CTG TAA CCA CGT GAC CGG TTC AGA TGG ATG GGT CCT TCG CAT AGA CCT ACA GCA TAG GTA GTC CAG AAG CC-3′. The aptamer pool for experiments with streptavidin was generated from the ssDNA library that contained a random region (N) of 40 nucleobases, using IFCE.[9] The sequence of the library was as follows: 5′-/Alex488N/CTA CGG TAA ATC GGC AGT CAN NNN NNN NNN NNN NNN NNN

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

NNN NNN NNN NNN NNN NNN NNN ATC TGA AGC ATA GTC CAG GC-3′. The pool was used instead of an individual aptamer as no individual aptamers had been identified and characterized from this pool before our experiments. IFCE-based partitioning leads to very pure aptamer pools. Therefore, the peak of the protein−ssDNA complex was narrow for the pool, and complex mobility could be determined accurately. Hence, there was no adverse effect of using the aptamer pool instead of an individual aptamer. Tris-HCl (50 mM final concentration, pH 8.0) was used as both a running buffer and a dilution buffer for all CE experiments with SSB. In a case of CE experiments with MutS, AlkB, streptavidin, and ABH2, Tris-acetate (50 mM final concentration, pH 8.2) was used as both a running buffer and a dilution buffer. The reason for the buffer change from Tris-HCl to Tris-acetate for protein−aptamer complexes was a noticeably better reproducibility of the these experiments conducted in Tris-acetate. Borax (25 mM, pH 9.2) was used as both a running buffer and a dilution buffer for the additional experiment with the AlkB−aptamer complex. Bodipy (difluoro-{2-[1-(3,5-dimethyl-2H-pyrrol-2-ylidene-N)ethyl]-3,5-dimethyl-1H-pyrrolato-N}boron) and fluorescein were used as internal standards for accurate mobility calculations. All solutions were prepared in deionized water filtered through a 0.22 μm Milipore filter membrane (Nepean, ON).

Capillary Electrophoresis. All CE experiments were done with a P/ACE MDQ instrument from Sciex (Brea, CA) using LIF detection with excitation at 488 nm. Fluorescence of FAM- and Alex488N-labeled species was detected at 520 nm; fluorescence of Chromeo-labeled proteins was detected at 630 nm. An uncoated fused silica capillary of an 82.2 cm total length and a 72.0 cm distance from the inlet to the detection point was used for all CE experiments with SSB and SSB-ssDNA complexes. For the rest of the proteins and their complexes with aptamers, an uncoated fused silica capillary of a 49.5 cm total length and a 39.5 cm distance from the inlet to the detection point was used. A shorter capillary was chosen for protein−aptamer complexes to ensure that they were intact by the time they reached the detector. New capillaries were preconditioned by washing them with 10 capillary volumes of methanol. In the beginning of each run, the capillary was rinsed with 8 capillary volumes of each of 100 mM HCl, 100 mM NaOH, deionized water, and the running buffer. The samples were injected into the capillary by pressure of 0.5 psi (3.45 kPa) applied for 10 s. Capillary coolant temperature was kept at 25 °C. In the case of an 82.2 cm-long capillary, an electric field of 304 V/cm with a positive electrode at the inlet was applied to carry out the electrophoresis. In the case of 49.5 cm-long capillary, the electric field strength was 505 V/cm. The electric field strength was 404 V/cm in the additional experiments with the AlkB−aptamer complex conducted at pH 9.2.

Sample Preparation. MutS, AlkB, SSB, and streptavidin (Table S1) were labeled with the Chromeo P503 dye according to the dye manufacture's procedure and were further analyzed with CE separately from ssDNA. The final concentrations of the proteins were 0.6, 1.9, 3.8, and 2.0 μM for MutS, AlkB, streptavidin, and SSB (per tetramer), respectively. For CE analysis of protein−ssDNA complexes, an unlabeled protein was mixed with ssDNA and incubated for 1 h at room temperature. A marker of the electroosmotic flow (bodipy) was added to all samples prior to their injection into the capillary. Fluorescein was added as an additional internal standard in experiments with SSB. In our experiments with AlkB, ABH2, and streptavidin, the peak of fluorescein partially overlapped with the peaks of protein−aptamer complexes due to their similar electrophoretic mobilities (−26.64 $mm^2\ kV^{-1}\ s^{-1}$ for fluorescein and −25.63, −24.58, and −25.96 $mm^2\ kV^{-1}\ s^{-1}$ for AlkB at pH 9.2 and 8.2 and ABH2 at pH 8.2, respectively), which made us exclude fluorescein from these experiments and from all experiments that were conducted in a shorter capillary, for consistency. For CE experiments with Chromeo-labeled MutS, streptavidin, and AlkB, the concentration of bodipy was 1 μM. The concentration of bodipy in the mixtures of unlabeled MutS, AlkB, and strepatavidin with their aptamers was 100 nM. Bodipy concentration was 500 nM and fluorescein concentration was 50 nM for all CE experiments with SSB protein and SSB−ssDNA mixtures. The concentration of any ssDNA (including aptamers) in a mixture with a protein was 100 nM, except for the streptavidin−aptamer pool mixture, where the concentration of the aptamer pool was 20 nM. Concentrations of the unlabeled protein in the mixtures with ssDNA (including aptamers) were 280, 250, 317, and 100 nM for MutS, AlkB, streptavidin, and SSB (per tetramer), respectively. In the experiments with ABH2 and AlkB at pH 9.2, concentrations of the protein in the mixture with the corresponding aptamer were 270 nM for ABH2 and 200 nM for AlkB. The concentration of 80-nt ssDNA in the experiment with AlkB at pH 9.2 was 200 nM. Migration times of proteins, ssDNA, and their complexes along with migration time of bodipy were used to find μ_{ssDNA} and $\mu_{P-ssDNA}$. All experiments were performed in triplicate.

RESULTS AND DISCUSSION

Development of the Model. In the following consideration, subscripts P, ssDNA, and P−ssDNA assign parameters to protein, ssDNA, and the protein−ssDNA complex, respectively. No distinction between aptamers and other ssDNA is made. The charge of the protein−ssDNA complex is a superposition of charges of the protein and ssDNA

$$q_{P-ssDNA} = aq_P + bq_{ssDNA} \quad (2)$$

where all charges take into consideration the ionic atmosphere around the molecule, and coefficients a and b adjust contributions of the individual charges to the resulting charge in consideration of rearrangements of the ionic atmosphere for each component as a result of complex formation.

A ratio between the charge in eq 2 and a corresponding frictional coefficient of the species represents its electrophoretic mobility (μ)

$$\mu = q/f \quad (3)$$

Expressing charges in eq 2 through electrophoretic mobilities and frictional coefficients of the components leads to the following equation for the electrophoretic mobility for the complex

$$\mu_{P-ssDNA} = \frac{a\mu_P f_P + b\mu_{ssDNA} f_{ssDNA}}{f_{P-ssDNA}} \quad (4)$$

In eq 4, electrophoretic mobilities of the protein and ssDNA can be easily determined experimentally, while frictional coefficients of the protein, ssDNA, and the complex cannot be easily measured. However, the frictional coefficients can be expressed through known size characteristics, such as

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

molecular weights of the species and the numbers of nucleotides in ssDNA.

In order to express the frictional coefficients for the protein and the complex through their molecular weights, we use the assumption of spherical shapes for both the protein and the protein−ssDNA complex (Figure 1c). The volume of the protein can be expressed as a volume of the sphere that encloses the protein[16,17]

$$V_P = \frac{4}{3}\pi {R_P}^3 \tag{5}$$

Here, R_P is the protein's hydrodynamic radius which also defines its frictional coefficient according to Stokes' law

$$f_P = 6\pi\eta R_P \tag{6}$$

where η is the dynamic viscosity of solution.

On the other hand, the volume of a globular biopolymer can be expressed as a ratio between its molecular weight (MW) and its density (ρ). Accordingly, we can write for the volume of the protein molecule[17]

$$V_P = \frac{MW_P}{\rho_P} \tag{7}$$

The combination of eqs 5−7 links the frictional coefficient of the protein with its molecular weight and density

$$f_P = 6\pi\eta\left(\frac{3}{4\pi\rho_P}\right)^{1/3} {MW_P}^{1/3} \tag{8}$$

In the first approximation, the density of the protein and the viscosity of the solution are constant. Equation 8 can then be rewritten as

$$f_P = \alpha {MW_P}^{1/3},\ \alpha = 6\pi\eta\left(\frac{3}{4\pi\rho_P}\right)^{1/3} \tag{9}$$

where α is a constant. Since we assume the protein−ssDNA complex to have a spherical shape, eqs 5−9 can also be applied to the complex, and the final equation for f of the complex will be similar to eq 9 but with a different constant β, which includes the density of the complex

$$f_{P-ssDNA} = \beta {MW_{P-ssDNA}}^{1/3},\ \beta = 6\pi\eta\left(\frac{3}{4\pi\rho_{P-ssDNA}}\right)^{1/3} \tag{10}$$

The last unknown parameter—the frictional coefficient of the ssDNA—can be expressed through the number of nucleotides (N) in ssDNA[14]

$$f_{ssDNA} = \gamma N^{0.68} \tag{11}$$

where γ is a constant coefficient of proportionality.

At this point, all frictional coefficients in eq 4 can be replaced with the corresponding right-hand sides of eqs 9−11, and the equation for $\mu_{P-ssDNA}$ will be

$$\mu_{P-ssDNA} = \frac{\alpha a \mu_P {MW^{1/3}}_P + \gamma b \mu_{ssDNA} N^{0.68}}{\beta {MW^{1/3}}_{P-ssDNA}} \tag{12}$$

For simplicity, $\alpha a/\beta$ and $\gamma b/\beta$ can be replaced with empirical constants C and B, respectively, and ($MW_{P-ssDNA}$) can be represented as a sum of MW_P and MW_{ssDNA}

$$\mu_{P-ssDNA} = \frac{C\mu_P {MW_P}^{1/3}}{(MW_P + MW_{ssDNA})^{1/3}} + \frac{B\mu_{ssDNA} N^{0.68}}{(MW_P + MW_{ssDNA})^{1/3}} \tag{13}$$

The empirical constants C and B can be obtained from eq 13, using experimental data for electrophoretic mobilities of the species, along with molecular weights of the protein and ssDNA and the number of nucleotides in ssDNA. In such a case, the known parameters in eq 13 can be grouped as X and Y, and the resulting $\mu_{P-ssDNA}$ will be a linear combination of X and Y

$$\mu_{P-ssDNA} = BX + CY$$

$$Y = \frac{\mu_P {MW^{1/3}}_P}{(MW_P + MW_{ssDNA})^{1/3}}$$

$$X = \frac{\mu_{ssDNA} N^{0.68}}{(MW_P + MW_{ssDNA})^{1/3}} \tag{14}$$

In a case of n different protein−ssDNA complexes, eq 14 turns into a group of equations, which can be written in a matrix form

$$M\cdot\Phi = \Psi,\ M = \begin{pmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \cdots & \cdots \\ X_n & Y_n \end{pmatrix},\ \Phi = \begin{pmatrix} B \\ C \end{pmatrix} \text{ and } \Psi = \begin{pmatrix} \mu_{P-ssDNA1} \\ \mu_{P-ssDNA2} \\ \cdots\cdots \\ \mu_{P-ssDNAn} \end{pmatrix} \tag{15}$$

Φ is a desired vector of solutions, which can be found as

$$\Phi = [(M^T M)^{-1} M^T]\cdot\Psi \tag{16}$$

Here, the upper indexes "T" and "−1" corresponds to the transpose and inverse of the corresponding matrices, respectively.

Finding μ_P, μ_{ssDNA}, and $\mu_{P-ssDNA}$ for a Set of Proteins and ssDNAs. For our initial study, we have chosen seven protein−ssDNA complexes. Among them, we had three protein−aptamer complexes for three proteins—MutS, streptavidin, and AlkB—and their 80-nucleotide-long aptamers. The remaining four complexes are SSB−ssDNA complexes with ssDNA of different lengths: 88, 80, 75, and 25 nucleotides. ssDNA wraps the SSB molecule and creates a tight complex similar to protein−aptamer complexes (Figure 1c). Using SSB allowed us to vary the length of ssDNA and ensure that the model is suitable for a range of aptamer lengths. The main three reasons for choosing MutS, AlkB, streptavidin, and SSB for the training set in model development were as follows. First, the range of molecular weights of the chosen proteins (24−95 kDa with a median value of 60 kDa) covers above 80% of proteins found in the eukaryotic proteome (49 ± 48 kDa).[18,19] Second, the range of isoelectric point values (pI) of the selected four proteins (starting from 5.61 for MutS and ending up with 8.36 for streptavidin) also covers most of pI values of proteins in the eukaryotic proteome.[19] Third, the chosen proteins are suitable for CE analysis in uncoated capillaries since they do not adsorb significantly onto bare fused silica.

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

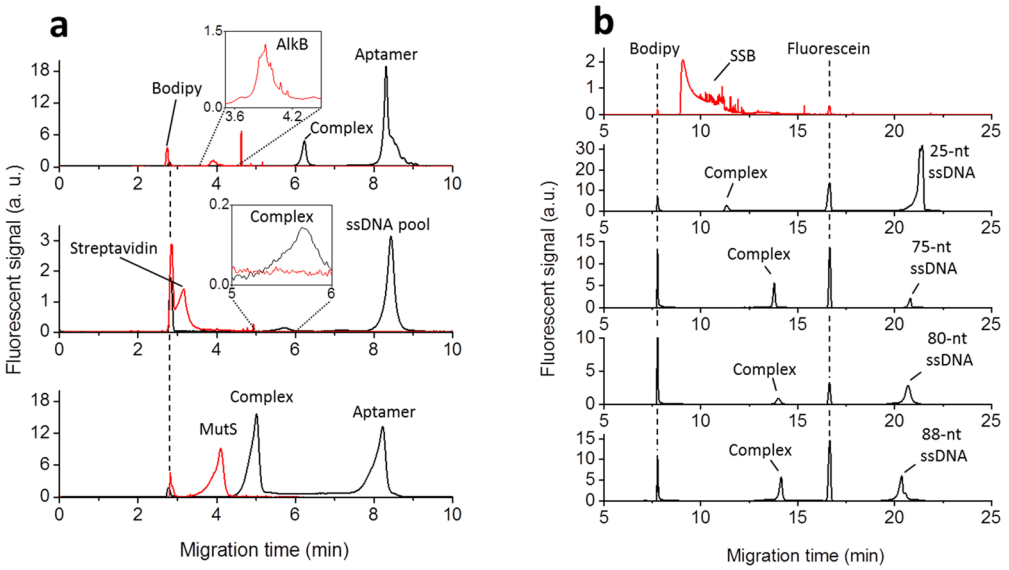


Figure 2. Electropherograms utilized for the determination of electrophoretic mobilities of four proteins (red) and seven protein−aptamer complexes and ssDNAs (black). Panel a shows electropherograms obtained for AlkB, MutS, and streptavidin as well as their aptamers and protein−aptamer complexes. Panel b shows electropherograms for SSB, four ssDNAs of different lengths, and SSB−ssDNA complexes. Bodipy was used as an electroosmotic flow marker and internal standard in all experiments; fluorescein was used as an additional internal standard in experiments with SSB. See Materials and Methods for other experimental conditions. Electropherograms were aligned with each other using positions of the internal standards as references. Electrophoretic mobilities were determined using experimental migrations times, capillary lengths, and electric field strengths as described in the text.

Table 1. Parameters of the Mobility-Prediction Model[a]

		Cross-validation for $\mu_{P-ssDNA} = BX + CY$							
		Excluded complexes							
Constants	All complexes	SSB−25-nt ssDNA	SSB−75-nt ssDNA	SSB−80-nt ssDNA	SSB−88-nt ssDNA	AlkB−Apt	Strepta-vidin−Apt	MutS−Apt	RSD%
C	0.236	0.001	0.245	0.245	0.230	0.284	0.332	0.470	55
B, kDa$^{1/3}$	0.151	0.164	0.152	0.152	0.154	0.154	0.145	0.142	4.6
		Cross-validation for $\mu_{P-ssDNA} = A + BX$							
Constants	All complexes	SSB−25-nt ssDNA	SSB−75-nt ssDNA	SSB−80-nt ssDNA	SSB−88-nt ssDNA	AlkB−Apt	Strepta-vidin−Apt	MutS−Apt	RSD%
A, mm^2 kV^{-1} s^{-1}	−9.95	−9.44	−9.93	−9.97	−9.93	−9.79	−10.00	−10.11	2.2
B, kDa$^{1/3}$	0.0929	0.0966	0.0929	0.0923	0.0932	0.0944	0.0922	0.0924	1.7
R^2	0.988	0.923	0.989	0.991	0.988	0.987	0.989	0.996	
$\Delta\mu$, %		1.9%	0.9%	1.8%	0.5%	1.0%	1.3%	3.0%	

[a]Top: Calculated constants C and B for protein−ssDNA complexes, presented in Figure 2. The constants were calculated with eq 14 for seven protein−ssDNA complexes, as well as for the cases when one protein−DNA complex was excluded from the complete set of seven. Bottom: Constants A and B derived from eq 19 by linear regression analysis. Constants A and B were calculated for seven protein−ssDNA complexes, as well as for the cases when one protein−DNA complex was excluded from the complete set of seven. The row "$\Delta\mu$, %" shows deviations of the calculated mobility from the experimental values for the excluded protein−ssDNA complexes.

We carried out CE experiments to find μ_P, μ_{ssDNA}, and $\mu_{P-ssDNA}$. In one set of experiments, we found μ_{ssDNA} and $\mu_{P-ssDNA}$. In these experiments, a mixture of an unlabeled protein with a corresponding fluorescently labeled ssDNA was subjected to CE with LIF detection (excitation at 488 nm and emission at 520 nm). Each electropherogram contained two peaks that corresponded to fluorescently labeled molecules: protein−ssDNA complex and ssDNA (Figure 2). In another set of experiments, we labeled proteins with a small electrically neutral fluorescent dye, ChromeoP503, that did not affect their mobility significanlty.[20,21] The four ChromeoP503-labeled proteins were then run in CE one by one and separately from DNA using LIF detection (excitation at 488 nm, emission at 630 nm) (Figure 2). We did not mix the labeled proteins with ssDNA to avoid any possible effects of the ChromeoP503 label on complex formation. The obtained electropherograms for the labeled proteins and the electropherograms for the mixtures of unlabeled proteins and ssDNA were overlaid with each other and aligned according to the positions of internal standards. In all CE experiments that were conducted in an 80.2 cm-long capillary, we used peaks of two internal standards (bodipy and fluorescein) for such an alignment. In our CE experiments with a 49.5 cm-long capillary, only the peak of bodipy was used for such an alignment (see Materials and Methods). Being an electrically neutral molecule, bodipy also served as an EOF marker, and its migration time (t_{bodipy}) was used to calculate the mobility of EOF (μ_{EOF})

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

$$\mu_{EOF} = \frac{L}{Et_{bodipy}} \quad (17)$$

Here, L is a capillary length from the inlet to the detection point, and E is the electric field strength.

The obtained values for μ_{EOF} along with the migration times (t) of the proteins, ssDNA, and the complexes were then used to find electrophoretic mobilities of the species (Table S2) using the following equation

$$\mu = \frac{L}{Et} - \mu_{EOF} \quad (18)$$

The experimentally found mobility values were used in the determination of empirical constants as explained in the next section.

Finding Empirical Constants. Empirical constants B and C for a training/cross-validation set of seven protein−ssDNA complexes (n = 7) were calculated by solving eq 15 using the experimentally obtained values of μ_P, μ_{ssDNA}, and $\mu_{P-ssDNA}$, along with known $MW_{P-ssDNA}$, MW_{ssDNA}, and N as input parameters (Table 1, top). Similar calculations were conducted for seven groups of six protein−ssDNA complexes (n = 6). These groups were obtained by excluding one protein−ssDNA complex from the full set of seven. Pairs of B and C for the seven subsets of six protein−ssDNA complexes were determined in order to perform LOOCV of the empirical model (Table 1, bottom). It was found that constant B did not vary much across the seven subsets; its residual standard deviation (RSD) was only 4.6%. However, constant C varied significantly resulting in RSD of 55%; such a large RSD indicated that parameter C could not be considered as a constant. Our comparison of variables X and Y showed that variable X was greater than Y by more than an order of magnitude (Table S3). The large variation of C along with relative smallness of Y suggested that BX dominated over CY in defining the value of $\mu_{P-ssDNA}$ in eq 14. Accordingly, we replaced CY in eq 14 with a constant A and obtained the following simplified equation

$$\mu_{P-ssDNA} = A + BX \quad (19)$$

Linear fitting of mobility data for seven protein−ssDNA complexes with eq 19 resulted in B = 0.0929 $kDa^{1/3}$ and A = −9.95 $mm^2kV^{-1}s^{-1}$ with a correlation coefficient (R^2) of 0.989 (Figure 3). The LOOCV approach that we previously used with eq 14 was also applied to eq 19, resulting in a very small RSD values of 2.2% and 1.7% for constants A and B, respectively (Table 1, bottom). Such a precise determination of A and B allowed accurate mobility prediction for the excluded protein−ssDNA complexes with only a 3.0% maximum deviation from the experimentally determined mobility (Table 1, bottom). Considering the obtained values of A and B for seven protein−ssDNA complexes, eq 17 can be rewritten as follows

$$\mu_{P-ssDNA} = -9.95 + 0.0929\mu_{ssDNA}N^{0.68}MW_{P-ssDNA}{}^{-1/3} \quad (20)$$

Equation 20 allows the prediction of $\mu_{P-ssDNA}$, using only the molecular weight of the complex, the number of nucleotides in ssDNA, and μ_{ssDNA} as input parameters. This equation resembles equations previously obtained for predicting electrophoretic mobilities of proteins, based on the assumption of their spherical shapes.[16,22] The main difference between eq 20 and the equations for protein mobilities is in the component

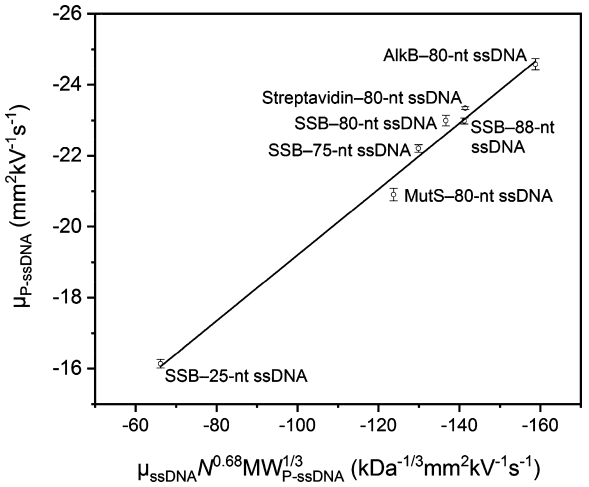


Figure 3. Line of the best fit for the electrophoretic mobility of protein−ssDNA complex as a function of X: $\mu_{P-ssDNA} = A + BX$, where $X = \{\mu_{P-ssDNA}N^{0.68}MW^{1/3}{}_{P-ssDNA}\}$. Calculated values for A and B were −9.95 mm^2 kV^{-1} s^{-1} and 0.0929 $kDa^{1/3}$, respectively. The correlation coefficient was R^2 = 0.988. Error bars correspond to the standard deviations of the experimentally determined electrophoretic mobilities.

that corresponds to the charge of the molecule. In eq 20, the charge of the P−ssDNA complex appears to be mainly defined by the charge of ssDNA which depends on N, where as in the mobility models for proteins, the charge of the protein is considered to be equal to the net charge of its amino acid sequence.[22]

As a side note, we would like to mention that eq 20 can be further simplified by assuming μ_{ssDNA} to be constant for buffers with low ionic strengths. The validity of such an assumption is based on a weak dependence of μ_{ssDNA} on the ionic strength (I) of the running buffer for I ranging between 0 and 0.05 M according to the following equation[10]

$$\mu_{ssDNA} = -31.6 + 53.2I - 29.1I^2(mm^2kV^{-1}s^{-1}) \quad (21)$$

As follows from eq 21, a 2-fold change in I changes μ_{ssDNA} by only a factor of ≈1.04. In our experiments (I = 0.027 M for Tris-HCl and I = 0.021 M for Tris-acetate), the average experimentally determined value of μ_{ssDNA} was −31.4 mm^2 kV^{-1} s^{-1} with RSD of 3.2%. The values of μ_{ssDNA} calculated with eq 21 for I = 0.021 and 0.027 M are −30.5 and −30.2 mm^2 kV^{-1} s^{-1}, respectively. Using either of these values (−31.4 or −30.5 and −30.2 mm^2 kV^{-1} s^{-1}) in eq 20 instead of experimentally determined μ_{ssDNA} does not cause a significant increase in deviations of calculated $\mu_{P-ssDNA}$ from the experimentally determined ones (Table S4).

In order to test our model and further improve it, we used a set of three additional protein−aptamer complexes (Table S5): a complex of human α-thrombin with its 29-nt aptamer at pH 7.5, a complex of ABH2 with its 80-nt aptamer at pH 8.2, and a complex of AlkB with its 80-nt aptamer at pH 9.2. The mobility value for the α-thrombin−aptamer complex was calculated from the literature data on migration times of the complex and an EOF marker.[23] The mobility values for the other two complexes were obtained by us experimentally (Figure S1). AlkB with the 80-nt aptamer was analyzed at a higher pH value than the pH value used for the training/cross-validation set in order to change the calculated charge of the protein (from −1 at pH 8.2 to −5.3 at pH 9.2). The three

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

predicted mobilities were different from the experimental ones by only 1.8%, 4.6%, and 0.1% for human α-thrombin, ABH2, and AlkB, respectively (Table S6). We further added the three new data points to our original set of seven points and recalculated empirical constants A and B (Figure S2). The new constants were $A = -9.47$ mm^2 kV^{-1} s^{-1} and $B = 0.0973$ kDa$^{1/3}$ and differed from the original ones by only 4.8% and 4.7%, respectively. The correlation coefficient was $R^2 = 0.987$. Cumulatively, these results confirm the validity of our mobility predictor.

CONCLUSIONS

To conclude, we developed an empirical mathematical model for accurate prediction of the electrophoretic mobility of protein−aptamer complexes in the approximation of their spherical shapes. The relationship between the mobility of a protein−aptamer complex and the three known parameters (mobility of ssDNA, length of ssDNA, and molecular weight of the complex) was derived using fundamental principles of molecular physics. In this respect, our model differs from predictive models based solely on the empirical correlation between parameters.[24] Empirical models require very large training sets to establish the correlation. In our case, the correlation is deterministic, and a much smaller training set should suffice finding the slope and intercept of the straight line. Our model was built upon mobility data for seven protein−ssDNA complexes, including four complexes of SSB with ssDNA of different lengths wrapped around SSB and three protein−aptamer complexes for MutS, streptavidin, and AlkB with their 80-nucleotide-long aptamers. The model suggests that the mobility of the protein−ssDNA complex depends linearly on a multiplication product of three terms that can be easily found: the electrophoretic mobility of ssDNA identical in length to the aptamer, the number of nucleotides in the aptamer to the power of 0.68, and a sum molecular weight of the protein and the aptamer to the power of negative 1/3. The linear dependence includes two empirical parameters (line's intercept A and line's slope B) which were determined by fitting the model into a data set obtained for the seven protein−ssDNA complexes. The best-fit linear function had a correlation coefficient of 0.988. The developed model will serve as a reliable predictor of protein−aptamer complex mobility. We expect that this model will be applicable to a wide range of pH values because the only pH-dependent parameters in the model are the unknown mobility of the complex and the experimentally found mobility of DNA. When pH changes, the mobility of DNA changes, and this change leads to a change in the predicted mobility of the complex as was confirmed with the AlkB−aptamer complex studied at pH 8.2 and 9.2 (Tables S4, S6). We do not exclude the possibility that the extremely big charge of the protein (similar to the one of DNA) may result in a noticeable difference between the experimental and predicted mobilities of the complex. However, the impact of protein charge on model performance can be minimized by reducing the protein charge via adjusting pH of the running buffer to the value of the protein's pI. The a priori knowledge of complex mobility will allow accurate blind setting of the time window for complex collection in aptamer partitioning by NECEEM.[8] It will also facilitate rational selection of the number of fractions to be blindly collected in the IFCE-based partitioning.[9] Overall, this mobility predictor will be an indispensable tool in rational design of CE-based aptamer selection. We expect that our model will be applicable to proteins with molecular weights outside the investigated range; however, it is likely that the accuracy of prediction will be lower for such proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.9b02019.

Table S1: Properties of the proteins. Table S2: Electrophoretic mobilities and molecular weights of the species. Table S3: Calculated variables X and Y for protein−ssDNA complexes. Table S4: Comparison of electrophoretic mobilities calculated for different μ_{ssDNA}. Table S5: Properties of proteins used for testing the predictive model. Table S6: Experimental and predicted electrophoretic mobilities for complexes of human α-thrombin, ABH2, and AlkB. Figure S1: Electropherograms used for the determination of electrophoretic mobilities for complexes of AlkB with its 80-nt aptamer at pH 9.2 and ABH2 with its 80-nt aptamer. Figure S2: Line of the best fit for electrophoretic mobilities of 10 protein−ssDNA complexes as a function of X: $\mu_{P-ssDNA} = A + BX$, where $X = \{\mu_{P-ssDNA} N^{0.68} MW^{1/3}_{P-ssDNA}\}$. (PDF)

Raw data for the electropherograms that are presented in Figure 2. (XLSX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: skrylov@yorku.ca.

ORCID

Stanislav S. Beloborodov: 0000-0003-2377-4845
Svetlana M. Krylova: 0000-0002-3291-6721
Sergey N. Krylov: 0000-0003-3270-2130

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the NSERC Canada (Grant STPG-P 521331-2018). The authors thank Dr. Sven Kochmann for technical help with manuscript preparation.

REFERENCES

(1) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818−822.
(2) Ciancio, D. R.; Vargas, M. R.; Thiel, W. H.; Bruno, M. A.; Giangrande, P. H.; Mestre, M. B. *Pharmaceuticals* **2018**, *11*, 86.
(3) Kruspe, S.; Mittelberger, F.; Szameit, K.; Hahn, U. *ChemMedChem* **2014**, *9*, 1998−2011.
(4) Proske, D.; Blank, M.; Buhmann, R.; Resch, A. *Appl. Microbiol. Biotechnol.* **2005**, *69*, 367−374.
(5) Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T. *PLoS One* **2010**, *5*, e15004.
(6) Vant-Hull, B.; Payano-Baez, A.; Davis, R. H.; Gold, L. *J. Mol. Biol.* **1998**, *278*, 579−597.
(7) Gopinath, S. C. B. *Anal. Bioanal. Chem.* **2006**, *387*, 171−182.
(8) Berezovski, M.; Drabovich, A.; Krylova, S. M.; Musheev, M.; Okhonin, V.; Petrov, A.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 3165−3171.
(9) Le, A. T. H.; Krylova, S. M.; Kanoatov, M.; Desai, S.; Krylov, S. N. *Angew. Chem., Int. Ed.* **2019**, *58*, 2739−2743.

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687

(10) Desruisseaux, C.; Long, D.; Drouin, G.; Slater, G. W. *Macromolecules* **2001**, *34*, 44−52.
(11) Meagher, R. J.; Won, J.-I.; McCormick, L. C.; Nedelcu, S.; Bertrand, M. M.; Bertram, J. L.; Drouin, G.; Barron, A. E.; Slater, G. W. *Electrophoresis* **2005**, *26*, 331−350.
(12) Bao, J.; Krylova, S. M.; Cherney, L. T.; Hale, R. L.; Belyanskaya, S. L.; Chiu, C. H.; Shaginian, A.; Arico-Muendel, C. C.; Krylov, S. N. *Anal. Chem.* **2016**, *88*, 5498−5506.
(13) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T; de Beer, T. A P; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. *Nucleic Acids Res.* **2018**, *46*, W296−W303.
(14) Nkodo, A. E.; Garnier, J. M.; Tinland, B.; Ren, H.; Desruisseaux, C.; McCormick, L. C.; Drouin, G.; Slater, G. W. *Electrophoresis* **2001**, *22*, 2424−2432.
(15) Beloborodov, S. S.; Bao, J.; Krylova, S. M.; Shala-Lawrence, A.; Johnson, P. E.; Krylov, S. N. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2018**, *1073*, 201−206.
(16) Erickson, H. P. *Biol. Proced. Online* **2009**, *11*, 32−51.
(17) Rickard, E. C.; Strohl, M. M.; Nielsen, R. G. *Anal. Biochem.* **1991**, *197*, 197−207.
(18) Brocchieri, L.; Karlin, S. *Nucleic Acids Res.* **2005**, *33*, 3390−3400.
(19) Kozlowski, L. P. *Nucleic Acids Res.* **2017**, *45*, D1112−D1116.
(20) Wojcik, R.; Swearingen, K. E.; Dickerson, J. A.; Turner, E. H.; Ramsay, L. M.; Dovichi, N. J. *J. Chromatogr. A* **2008**, *1194*, 243−248.
(21) Swearingen, K. E.; Dickerson, J. A.; Turner, E. H.; Ramsay, L. M.; Wojcik, R.; Dovichi, N. J. *J. Chromatogr. A* **2008**, *1194*, 249−252.
(22) Basak, S. K.; Ladisch, M. R. *Anal. Biochem.* **1995**, *226*, 51−58.
(23) Bai, Y.; Zhao, Q. *Anal. Methods* **2017**, *9*, 5684−5690.
(24) Rosenblatt, R.; Halámková, L.; Doty, K. C.; Costa de Oliveira, A. E.; Lednev, I. K. *Forensic Chem.* **2019**, *16*, 100175.

DOI: 10.1021/acs.analchem.9b02019
Anal. Chem. 2019, 91, 12680−12687