analytical chemistry

Article

pubs.acs.org/ac

Predicting Electrophoretic Mobility of Protein−Ligand Complexes for Ligands from DNA-Encoded Libraries of Small Molecules

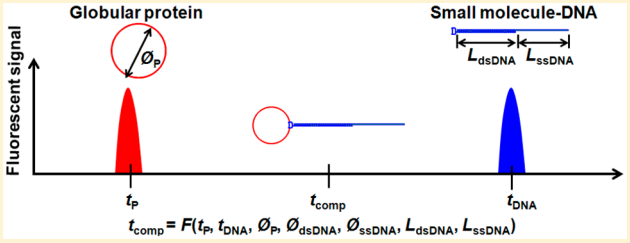
Jiayin Bao,[†] Svetlana M. Krylova,[†] Leonid T. Cherney,[†] Robert L. Hale,[‡] Svetlana L. Belyanskaya,[‡] Cynthia H. Chiu,[‡] Alex Shaginian,[‡] Christopher C. Arico-Muendel,[‡] and Sergey N. Krylov*[,†]

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

[‡]GlaxoSmithKline, 830 Winter Street, Waltham, Massachusetts 02451-8714, United States

Supporting Information

ABSTRACT: Selection of target-binding ligands from DNA-encoded libraries of small molecules (DELSMs) is a rapidly developing approach in drug-lead discovery. Methods of kinetic capillary electrophoresis (KCE) may facilitate highly efficient homogeneous selection of ligands from DELSMs. However, KCE methods require accurate prediction of electrophoretic mobilities of protein−ligand complexes. Such prediction, in turn, requires a theory that would be applicable to DNA tags of different structures used in different DELSMs. Here we present such a theory. It utilizes a model of a globular protein connected, through a single point (small molecule), to a linear DNA tag containing a combination of alternating double-stranded and single-stranded DNA (dsDNA and ssDNA) regions of varying lengths. The theory links the unknown electrophoretic mobility of protein−DNA complex with experimentally determined electrophoretic mobilities of the protein and DNA. Mobility prediction was initially tested by using a protein interacting with 18 ligands of various combinations of dsDNA and ssDNA regions, which mimicked different DELSMs. For all studied ligands, deviation of the predicted mobility from the experimentally determined value was within 11%. Finally, the prediction was tested for two proteins and two ligands with a DNA tag identical to those of DELSM manufactured by GlaxoSmithKline. Deviation between the predicted and experimentally determined mobilities did not exceed 5%. These results confirm the accuracy and robustness of our model, which makes KCE methods one step closer to their practical use in selection of drug leads, and diagnostic probes from DELSMs.



Finding molecules that can selectively bind therapeutic targets is the initial step in most mainstream approaches of modern drug development.[1−3] Selection of protein binders (ligands) from DNA-encoded librares of small molecules (DELSMs) is one such approach.[4,5] DELSMs provide a solution for the main dilemma of selection of ligands from highly diverse mixtures of molecules. On one hand, the probability of finding ligands increases with increasing diversity of the mixture. On the other hand, the increasing diversity decreases the number of copies of unique molecules in the mixture, making their identification impossible by classical structure-analysis methods. In DELSMs, the structure of every small molecule is encoded in its DNA tag and can thus be revealed by amplifying and sequencing the tag. The efficiencies of polymerase chain reaction (PCR) and DNA sequencing are so high[6] that selecting a few copies of each ligand from a DELSM is sufficient for identification of its structures. As a result, DELSMs with diversities of more than 1 billion structures are synthesized and used for drug-lead selection.[7]

The concept of DELSM was introduced in 1992,[4] and since then a number of synthetic approaches to the generation of DELSMs have been developed (Figure 1).[8,9] Different synthetic approaches lead to different structures of DNA tags. In general, DNA tags are linear DNA of two types: pure double-stranded DNA (dsDNA) and ds-ssDNA chimeras composed of dsDNA and single-stranded DNA (ssDNA) fragments.

Selection of ligands from DELSMs involves three major steps: (i) mixing and incubating the target protein with the DELSM to facilitate protein binding to the ligands, (ii) partitioning of protein-bound ligands from the rest of the library, and (iii) PCR amplification and sequencing of DNA tags of the collected ligands. High efficiency of the partitioning step is critical for successful selection from DELSMs. The collected ligands contain small molecules that cannot be amplified by tools of molecular biology. Therefore, the selection is typically achieved within 2−4 rounds and requires highly efficient partitioning.

The majority of partitioning methods are surface-based; they use either surface-immobilized protein target or partitioning on filters.[10,11] Nonspecific binding to the surface leads to high background and low partitioning efficiency. For instance, up to 15% of DNA can bind to a nitrocellulose filter via nonspecific

Received: March 11, 2016
Accepted: April 27, 2016
Published: April 27, 2016

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498−5506

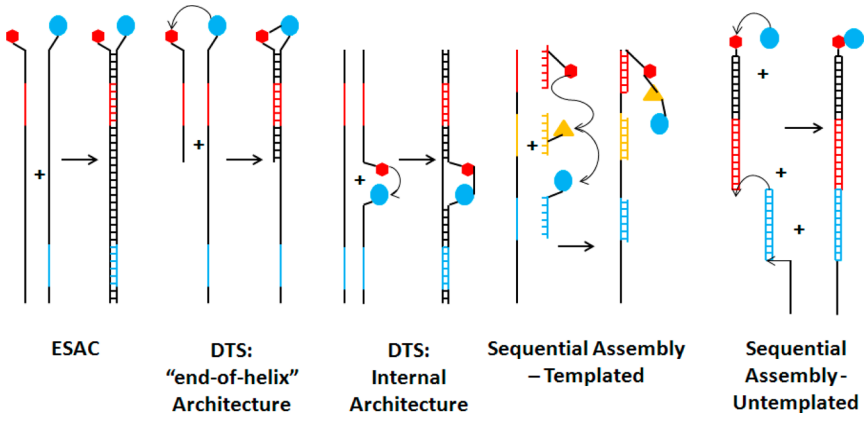


Figure 1. Schematic representation of assembly routes and corresponding structures of various DELSMs: (left to right) encoded self-assembling chemical (ESAC) library, DNA-templated synthesis (DTS) "end-of-helix" architecture, DTS internal architecture, sequential assembly templated, and sequential assembly untemplated. Building blocks of the small-molecule head and DNA fragments encoding them are shown by the same color.

interactions,[12] which is detrimental in selection. Failure to successfully select protein binders from DELSMs can be caused by low partitioning efficiency of surface-based methods.

Gel-free capillary electrophoresis (CE) is a solution-based alternative to classical surface-based partitioning methods. CE was successfully applied to selection of DNA aptamers from random ssDNA libraries with a background level of less than 0.01%.[13] We expect that CE can potentially be used for high-efficiency ligand selection from DELSM. This suggestion is based on the following logic. In all DELSMs, the DNA tag is much larger and has a much greater charge than the small-molecule head. As a result, the physical properties of DELSMs, such as size and charge, are mainly defined by the DNA tag. A single DELSM has a single structure of the DNA tags, and therefore the DELSM will migrate in CE as a single zone. Protein–ligand complexes will also migrate as a single zone with a mobility intermediate between those of the protein and the DELSM. Thus, the migration pattern of the three zones is similar to that in aptamer selection (Figure 2).[14]

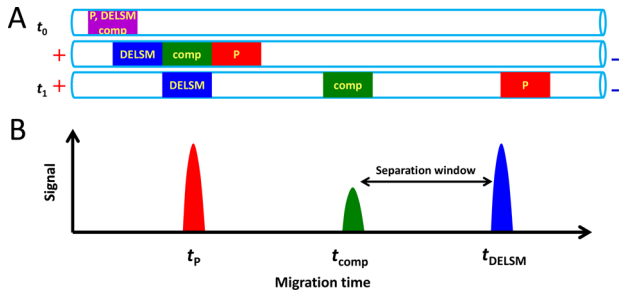


Figure 2. Conceptual depiction of migration patterns of DELSM, protein, and protein–ligand complex in capillary electrophoresis (CE) with strong electroosmotic flow (EOF). (A) Positions of the three corresponding electrophoretic zones at different times. The equilibrium mixture that contains DELSM, protein, and protein–ligand complexes is injected into the capillary at time t_0. When high voltage is applied, the three zones start moving toward the outlet with different velocities dependent on electrophoretic mobilities of the three components and the velocity of EOF. (B) Times at which the three zones reach the outlet and the separation window between them. The goal of selection is to collect the complex (green fraction) as accurately as possible. The objective of this work is to predict the mobility of the complex.

Efficient selection of ligands from DELSMs requires knowledge of the time at which the complex exits the capillary. Finding this time experimentally is very difficult, as the concentration of complex is below the limit of detection (LOD) even for highly sensitive fluorescence detection. Therefore, a theoretical model that can predict this time is required. In our previous study, we developed a model for predicting the electrophoretic mobility of protein–dsDNA complex. However, this model can be used only for DELSMs with pure dsDNA tags, such as encoded self-assembling chemical libraries, but not for other DELSMs that have ds-ssDNA tags (see Figure 1). Therefore, a more general solution for predicting protein–ligand complex migration for varying DELSMs is needed for practical use of CE in selection from such libraries.

Here we present a general theoretical model that can be equally applied to DELSMs with different structures of DNA tags. The model considers a globular protein attached to the DNA tag at a single point. The thin double layer model is used to find mobilities of protein, dsDNA, and ssDNA. Given these mobilities, effective electric and hydrodynamic forces acting upon protein, dsDNA, and ssDNA are determined. Then the mobility of protein–dsDNA–ssDNA complex is obtained from the equation of balance of all forces acting upon the complex. Finally, complex mobility is expressed in terms of experimentally measurable mobilities of protein and dsDNA–ssDNA chimera.

We derived an expression that links the unknown electrophoretic mobility of the protein–ligand complex with empirical data for electrophoretic mobilities of the protein and library. To test the developed mathematical model, we used binding of streptavidin (SA) to biotin-labeled dsDNA or ds-ssDNA with varying lengths of dsDNA and ssDNA regions. The predicted electrophoretic mobilities and migration times deviated from the experimentally measured ones by less than 11%. We also assessed our model by using two proteins, SA and carbonic anhydrase II (CAII), and two ligands with tag structures identical to those in actual GlaxoSmithKline (GSK) libraries. Deviation of predicted electrophoretic mobility from the experimental measured value did not exceed 5% for CAII and 3% for SA. We conclude that the model is adequate and can aid selection of protein binders from DELSMs and advance the use of such libraries in identifying drug leads and diagnostic probes.

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498–5506

EXPERIMENTAL SECTION

Chemicals and Materials. Fused-silica capillary was purchased from Polymicro (Phoenix, AZ). All reagents were dissolved in 50 mM Tris-HCl at pH 8.0 (unless otherwise specified); the same buffer was used as the CE run buffer. All DNA sequences used for constructing ds-ssDNA chimeras were purchased from IDT DNA Technology Inc. (Coralville, IA). The sequences were as follows: alexa80, 5′-alexa-TGA CTC CCA AAT CGA TGT GTT CCG CAA GAA GCC TGG TAA GCG GAG AAA GGT CGT TTT ACT GCC CGG TCT ACC TGA TGG CG-3′; alexa60, 5′-alexa-TCC GCA AGA AGC CTG GTA AGC GGA GAA AGG TCG TTT TAC TGC CCG GTC TAC CTG ATG GCG-3′; alexa40, 5′-alexa-CGG AGA AAG GTC GTT TTA CTG CCC GGT CTA CCT GAT GGC G-3′; alexa20, 5′-alexa-GCC CGG TCT ACC TGA TGG CG-3′; bioTEG-anti20, 5′-bioTEG-CGC CAT CAG GTA GAC CGG GC-3′; c1ss10, 5′-AAC GAC CTT T-3′; c2ss10, 5′-CAG GCT TCT T-3′; c3ss10, 5′-TCG ATT TGG G-3′. Alexa is the fluorophore used to label DNA; bioTEG indicates biotin linked to triethyleneglycol; and 80, 60, 40, and 20 indicate the number of nucleotides in each DNA sequence. The DNA sequences are annealed together to make different ds-ssDNA chimeras, detailed structures of which are shown in Figure S1 in Supporting Information. Annealing was achieved by incubating corresponding sequences of DNA at 90 °C for 10 min and then gradually cooling them down to the room temperature. Bodipy (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) was purchased from Life Technologies Inc. (Burlington, ON, Canada).

SA and CAII were labeled with a fluorogenic dye, chromeo P503 (Active Motif, Carlsbad, CA); the chromeo-labeled proteins will be referred to as chromeo-SA and chromeo-CAII. Briefly, 10 μL of protein solution (100 μM in 100 mM sodium bicarbonate, pH 8.3) was mixed with 6.6 μL of chromeo solution (1 mM in 100 mM sodium bicarbonate, pH 8.3), and then incubated at 4 °C overnight in the dark.

Biotin and Gly-(L)Leu-4-carboxybenzene sulfonamide (GLCBS-L-leucine) were used as small-molecule heads for binding to SA and CAII, respectively. The DNA-tagged small molecules will be referred to as biotin ligand and GLCBS-L-leucine ligand. Detailed synthetic procedures for these ligands were previously described with a modification of the closing primer ligation method.[9] Klenow polymerization was eliminated and the longer oligo strand was changed to the top, leaving a 31-nucleotide 3′ overhang to provide a non-competitive priming site for more efficient PCR amplification. The purification procedures for intermediates (produced during biotin ligand synthesis) are described in Supporting Information. All other reagents were purchased from Sigma–Aldrich (Oakville, ON, Canada). All solutions were made in deionized water filtered through a 0.22 μm filter (Millipore, Nepan, ON, Canada).

Instrumentation and Capillary Electrophoresis Conditions. All CE experiments were carried out on a MDQ-PACE instrument (Sciex, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detector. LIF signal was recorded at 520 nm for fluorescein, alexa, and bodipy detection and at 610 nm for detection of chromeo-SA and chromeo-CAII. Signal acquisition rate was 4 Hz. Inner diameter of the capillary was 75 μm. Total capillary length was 84.3 cm, with 74.2 cm from the injection end to the detection window. The capillary was flushed prior to each CE run with 20% bleach, 0.1 M HCl, 0.1 M NaOH, deionized H_2O, and run buffer. Sample was injected into the capillary at 0.5 psi for 10 s. The ends of the capillary were inserted into inlet and outlet reservoirs, and an electric field of 297 V/cm with a positive electrode at the injection end was applied to carry out electrophoresis. Temperature of the capillary was maintained at 15 °C. All experiments were performed in triplicate.

Migration Study of Protein–Ligand Complexes for Mock Ligands. For each binding mixture, 100 nM biotinylated DNA (either dsDNA or ds-ssDNA) was incubated with 1 μM chromeo-SA, 10 nM fluorescein (internal standard), and 5 μM bodipy (neutral marker) at room temperature for 30 min. For the control mixture, 100 nM ds-ssDNA was incubated with 10 nM fluorescein (internal standard) and 5 μM bodipy (neutral marker), at room temperature for 30 min.

Peak Identification of Biotin Ligand. The following synthetic intermediates were individually tested: biotin ligand head piece (native), biotin ligand head piece (denatured), splint with oligo alexa, and oligo alexa. The injected sample in each experiment contained 100 nM analyte with 10 nM fluorescein (internal standard).

Migration Studies of Protein–Ligand Complexes for GSK Ligands. Two binding systems were tested in this study: SA with biotin ligand and CAII with GLCBS-L-leucine ligand. Both ligands contain the same DNA structure, shown in Figure S2 in Supporting Information, a combination of two dsDNA (total of 94 bp) and two ssDNA (total of 23 nt) regions. For SA experiments, the binding mixture was made by incubating 1 μM chromeo-SA, 100 nM biotin ligand, 10 nM fluorescein (internal standard), and 5 μM bodipy (neutral marker), at 20 °C for 30 min; control mixture was the same as the binding mixture but without protein. For CAII experiments, the binding mixture was made by incubating 5 μM chromeo-CAII, 1 μM GLCBS-L-leucine ligand, 10 nM fluorescein (internal standard), and 5 μM bodipy (neutral marker) at 20 °C for 30 min; control mixture was the same as the binding mixture but without protein.

RESULTS AND DISCUSSION

Mathematical Model. In this work, we consider mobility of a protein–DNA complex in which the DNA is linear and is either pure dsDNA or a combination of dsDNA and ssDNA. The dsDNA regions are shorter than 72 base pairs, and the ssDNA regions are shorter than 50 nucleotides. This case describes a class of actual DELSMs used by pharmaceutical companies in selection of drug leads.[15] We assume that the protein is attached to one end of the dsDNA region as shown in Figure 3. This assumption excludes from consideration the "internal architecture" DELSMs (see Figure 1).

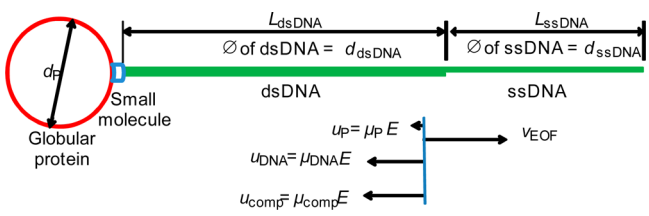


Figure 3. Schematic representation of one example of a complex between a globular protein and a ligand from DELSM. Binding is achieved via protein–small molecule interaction. The DNA tag in this example is composed of one dsDNA and one ssDNA region. The lower part illustrates relative values of velocities of EOF, protein, ligand, and protein–ligand complex.

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498–5506

Proteins have been earlier suggested as tags in DNA sequencing based on electrophoretic mobility shift of DNA; the approach is called end-labeled free-solution electrophoresis (ELFSE) of DNA. Although ELFSE-based DNA sequencing has never been advanced beyond proof of principle, the development of ELFSE helped to make significant progress in mobility theory for protein−DNA complexes.[16−22] Such models typically use the blob theory, which is applicable to DNA that is sufficiently long to be considered a semiflexible random coil and has a length significantly greater than the diameter of the protein.[20,22] The polymer can be considered as a semiflexible random coil if its contour length L is much greater (one or more orders of magnitude) than the Kuhn length b_K characterizing the polymer stiffness.[23,24] This assumption is not satisfied for dsDNA of fewer than 72 base pairs, for which $L_{dsDNA} < 24$ nm while $b_{K,dsDNA} > 100$ nm. Here and below, dsDNA and ssDNA in the subscript indicate that the corresponding parameters describe dsDNA or ssDNA. Thus, the complex of a protein linked to dsDNA can be considered as a rigid object with a diameter of more than 10 nm (for dsDNA at least 20 base pairs long and protein diameter ∼4 nm). For the contour length of ssDNA of fewer than 50 nt, we have $L_{ssDNA} < 21$ nm. Thus, L_{ssDNA} is of the same order of magnitude as the diameter of the protein−dsDNA complex (>10 nm). In this case, blob theory is not applicable. Moreover, ssDNA itself cannot be considered as a semiflexible random coil since its length is only three times larger than its Kuhn length, $b_{K,ssDNA}$ ∼6 nm.[20] Thus, for DNA of the lengths considered here, the protein−dsDNA complex is a rigid object and ssDNA cannot be treated as semiflexible random coil.

We will study the electrophoretic mobility of a complex formed by a globular protein attached to the end of a stretch of DNA that contains at least one dsDNA and one ssDNA regions (see Figure 3). This model does not describe the case of DTS internal architecture DELSM (see Figure 1). We consider globular proteins with a molecular weight of ≥30 kDa. Their average diameter can be estimated as $d_P \geq 4$ nm.[25] Thus, d_P is larger than the Debye length for the buffer, $\lambda_D \sim 1$ nm. In this case, the electrophoretic mobility μ_p of the protein can be estimated by an expression used in thin double layer theory:[16,20]

$$\mu_P = \frac{\varepsilon_0\varepsilon_r\zeta_P}{\eta} \approx \frac{-\sigma_P\lambda_D}{\eta} \qquad \zeta_P \approx \frac{-\sigma_P\lambda_D}{\varepsilon_0\varepsilon_r} \tag{1}$$

Here ε_0 is vacuum permittivity, ε_r is relative permittivity of the buffer, ζ_P is ζ potential of the globular protein, σ_P is average surface density of electric charge in the diffuse part of the double layer around the protein (i.e., excluding the Stern layer), and η is dynamic viscosity of the buffer. Equation 1 can be rewritten as follows:

$$\mu_P = \frac{-\sigma_P\lambda_D}{\eta} = \frac{Q_P\lambda_D}{\pi\eta d_P^2} \qquad \sigma_P = -\frac{Q_P}{\pi d_P^2} \tag{2}$$

where Q_P is electric charge of the protein (including the Stern layer charge). Note that protein mobility can have both positive and negative values (for positively and negatively charged proteins, respectively).

Equation 2 for μ_P can be also obtained from the balance of electric and hydrodynamic forces, $F_{E,P}$ and $F_{H,P}$, acting upon the protein molecule:

$$F_{E,P} + F_{H,P} = 0 \tag{3}$$

if the following effective values for these forces are assumed:

$$F_{E,P} = Q_P E \qquad F_{H,P} = -\frac{\pi\eta d_P^2}{\lambda_D} u_P \tag{4}$$

Here, E is electric field strength, u_P is relative velocity of the protein with respect to buffer, and $\mu_P = u_P/E$. Hereafter, we use a coordinate system in which both electric and hydrodynamic forces have only x-components. We will use eq 4 in the equation of balance of all forces acting upon the complex (see eq 14) to find the complex mobility.

Taking into account that L_{dsDNA} is smaller than $b_{K,dsDNA}$, we assume that dsDNA (shorter than 72 bp) behaves like a rigid rod. The dsDNA diameter, d_{dsDNA}, can be estimated as 2 nm,[26,27] which is larger than λ_D, while L_{dsDNA} is many times larger than λ_D. Thus, we can assume that the electrophoretic mobility of dsDNA, μ_{dsDNA}, is determined by an expression used in thin double layer theory:[16,20]

$$\mu_{dsDNA} = \frac{\varepsilon_0\varepsilon_r\zeta_{dsDNA}}{\eta} \approx \frac{-\sigma_{dsDNA}\lambda_D}{\eta}$$
$$\zeta_{dsDNA} \approx \frac{-\sigma_{dsDNA}\lambda_D}{\varepsilon_0\varepsilon_r} \tag{5}$$

Here, ζ_{dsDNA} is ζ potential of dsDNA and σ_{dsDNA} is surface density of the electric charge in the diffuse part of the double layer around dsDNA (i.e., excluding the Stern layer). Equation 5 can be rewritten as follows:

$$\mu_{dsDNA} = \frac{-\sigma_{dsDNA}\lambda_D}{\eta} = \frac{q_{dsDNA}\lambda_D}{\pi\eta d_{dsDNA}}$$
$$\sigma_{dsDNA} = -\frac{q_{dsDNA}}{\pi d_{dsDNA}} \tag{6}$$

where q_{dsDNA} is charge per unit length of dsDNA. In calculations of q_{dsDNA}, we should take into account the condensation of counterions on dsDNA.[28−32] The condensation takes place for cylindrical objects with linear density electric charge, q, satisfying the following relationship:[28]

$$|q| \geq q_{eff} \qquad q_{eff} = \frac{e}{z_i\lambda_B} \qquad \lambda_B = \frac{e^2}{4\pi\varepsilon_0\varepsilon_r k_B T} \tag{7}$$

Here e is proton charge, z_i is the valence of counterions, λ_B is Bjerrum length, k_B is the Boltzmann constant, and T is absolute temperature of the buffer. Usually, dsDNA has two negative charges per 0.34 nm of its length[20] and $\lambda_B = 0.7$ nm for water solutions at room temperature.[19,30] Thus, eq 7 is always satisfied for dsDNA, and condensation of counterions reduces the density of DNA charge q_{DNA} (excluding the Stern layer) to the effective value $-q_{eff}$, determined by the second relationship in eq 7.[28] Since we consider the Stern layer as a part of the condensed counterion layer, $|q_{DNA}|$ will be even less than q_{eff}. In this case, q_{DNA} can be considered as an adjustable parameter. We should note that dsDNA mobility has negative values since dsDNA is negatively charged.

Equation 6 for μ_{dsDNA} can be also obtained from the balance of electric and hydrodynamic forces, $F_{E,dsDNA}$ and $F_{H,dsDNA}$, acting upon dsDNA:

$$F_{E,dsDNA} + F_{H,dsDNA} = 0 \tag{8}$$

if we assume the following effective values for these forces:

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498−5506

$$F_{E,dsDNA} = q_{dsDNA} L_{dsDNA} E$$
$$F_{H,dsDNA} = -\frac{\pi\eta d_{dsDNA} L_{dsDNA}}{\lambda_D} u_{dsDNA} \quad (9)$$

Here u_{dsDNA} is relative velocity of dsDNA with respect to buffer. We will use eq 9 (and similar expressions obtained for ssDNA) in the equation of balance of all forces acting upon the complex to find the complex mobility.

Similarly to eq 5, we can determine the electrophoretic mobility of ssDNA using an expression from thin double layer theory:

$$\mu_{ssDNA} = \frac{\varepsilon_0 \varepsilon_r \zeta_{ssDNA}}{\eta} \approx \frac{-\sigma_{ssDNA} \lambda_D}{\eta}$$
$$\zeta_{ssDNA} \approx \frac{-\sigma_{ssDNA} \lambda_D}{\varepsilon_0 \varepsilon_r} \quad (10)$$

Here, ζ_{ssDNA} is ζ potential of ssDNA and σ_{ssDNA} is surface density of electric charge in the diffuse part of the double layer around ssDNA. Equation 10 can be rewritten as follows:

$$\mu_{ssDNA} = \frac{-\sigma_{ssDNA} \lambda_D}{\eta} = \frac{q_{ssDNA} \lambda_D}{\pi\eta d_{ssDNA}}$$
$$\sigma_{ssDNA} = -\frac{q_{ssDNA}}{\pi d_{ssDNA}} \quad (11)$$

where d_{ssDNA} is ssDNA diameter and q_{ssDNA} is charge per unit length of ssDNA. To find q_{ssDNA}, we also have to take into account condensation of counterions on ssDNA and the Stern layer charge.[28−32]

Equation 11 can be obtained from the balance of all effective forces acting upon ssDNA:

$$F_{E,ssDNA} + F_{H,ssDNA} = 0 \quad (12)$$

where $F_{E,ssDNA}$ and $F_{H,ssDNA}$ are effective electric and hydrodynamic forces acting upon ssDNA. They are determined by the following relationships similar to eq 9:

$$F_{E,ssDNA} = q_{ssDNA} L_{ssDNA} E$$
$$F_{H,ssDNA} = -\frac{\pi\eta d_{ssDNA} L_{dsDNA}}{\lambda_D} u_{ssDNA} \quad (13)$$

Here u_{ssDNA} is relative velocity of ssDNA with respect to buffer. Equation 13 will be used in eq 14.

The electrophoretic mobility of a globular protein attached to the end of dsDNA, the other end of which is linked to ssDNA, can be found from the balance of all effective forces acting upon such a complex:

$$F_{E,P} + F_{E,dsDNA} + F_{E,ssDNA} + F_{H,P} + F_{H,dsDNA} + F_{H,ssDNA} = 0 \quad (14)$$

Substitution of eq 4, eq 9, and eq 13 into eq 14 gives

$$(Q_P + q_{dsDNA} L_{dsDNA} + q_{ssDNA} L_{ssDNA}) E = \left(\frac{\pi\eta d_P^2}{\lambda_D} + \frac{\pi\eta d_{dsDNA} L_{dsDNA}}{\lambda_D} + \frac{\pi\eta d_{ssDNA} L_{ssDNA}}{\lambda_D} \right) u_{comp} \quad (15)$$

By solving this equation with respect to u_{comp} and taking into account that $u_{comp} = \mu_{comp} E$, we obtain the electrophoretic mobility of the complex, μ_{comp}:

$$\mu_{comp} = \frac{Q_P + q_{dsDNA} L_{dsDNA} + q_{ssDNA} L_{ssDNA}}{\frac{\pi\eta d_P^2}{\lambda_D} + \frac{\pi\eta d_{dsDNA} L_{dsDNA}}{\lambda_D} + \frac{\pi\eta d_{ssDNA} L_{ssDNA}}{\lambda_D}} \quad (16)$$

Taking into account eq 2, eq 6, and eq 11 for electrophoretic mobilities of globular protein, dsDNA, and ssDNA, we rewrite eq 16 as follows:

$$\mu_{comp} = \frac{d_P^2 \mu_P + d_{dsDNA} L_{dsDNA} \mu_{dsDNA} + d_{ssDNA} L_{ssDNA} \mu_{ssDNA}}{d_P^2 + d_{dsDNA} L_{dsDNA} + d_{ssDNA} L_{ssDNA}} \quad (17)$$

In the absence of protein, eq 17 reduces to the expression for electrophoretic mobility of ds-ssDNA chimera, $\mu_{ds\text{-}ssDNA}$:

$$\mu_{ds\text{-}ssDNA} = \frac{d_{dsDNA} L_{dsDNA} \mu_{dsDNA} + d_{ssDNA} L_{ssDNA} \mu_{ssDNA}}{d_{dsDNA} L_{dsDNA} + d_{ssDNA} L_{ssDNA}} \quad (18)$$

Given eq 18, we can express the electrophoretic mobility of complex in terms of the electrophoretic mobilities of protein and ds-ssDNA chimera:

$$\mu_{comp} = \frac{d_P^2 \mu_P + (d_{dsDNA} L_{dsDNA} + d_{ssDNA} L_{ssDNA}) \mu_{ds\text{-}ssDNA}}{d_P^2 + d_{dsDNA} L_{dsDNA} + d_{ssDNA} L_{ssDNA}} \quad (19)$$

By using eq 19 for the electrophoretic mobility of complex, we can readily find the complex migration time to the detector, t_{comp}:

$$t_{comp} = \frac{L_{capillary}}{v_{EOF} + \mu_{comp} E} \quad (20)$$

Here, $L_{capillary}$ is distance from beginning of the capillary to the detector and v_{EOF} is velocity of EOF in the capillary.

Derivation of eq 19 for the electrophoretic mobility of complex can be readily generalized for the case of ds-ssDNA molecules containing more than one dsDNA region and more than one ssDNA section (see Supporting Information). In this case, eq 19 will be still valid if we define L_{dsDNA} as total contour length of all dsDNA sections and L_{ssDNA} as total contour length of all ssDNA sections.

It is important to emphasize that eq 19 does not contain any empirical parameters except for the diameter of protein, which can typically be found from independent studies or from the literature, and the diameter and length of DNA, which are known. Therefore, no "training set" is required for making eq 19 eligible, and its general validity can be tested with a limited set of experimental data. If experimental systems that are poorly described by this expression are ever found, this would mean that at least one of the following assumptions is not fulfilled: (1) the protein is globular, (2) the protein diameter is greater than Debye length (which requires that its molecular weight be ≥30 kDa), or (3) the DNA tag is rodlike. However, since the utility of the model is to predict an approximate complex mobility for selection of binders, even if the prediction has a systematic error, it can still be useful.

Experimental Validation of Mathematical Model. We used the interaction between SA and biotinylated ds-ssDNA to test our model expressed by eq 19 and eq 20. Biotin played the role of small molecule. The interaction between SA and biotin is renowned for its exceptionally high affinity ($K_d \approx 10^{-14}$ M).

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498−5506

We have tested 14 different constructs of ds-ssDNA together with four dsDNA to ensure the ruggedness of our model. Detailed structural information for DNA tags used is shown in Figure S1 in Supporting Information.

As can be seen in eq 19 and eq 20, finding the mobility and migration time of the protein−DNA complex requires knowledge of the hydrodynamic sizes of protein (d_p), which is SA in this specific example, and DNA (d_{dsDNA}, d_{ssDNA}, L_{dsDNA}, and L_{ssDNA}). We used values of d_{dsDNA} = 2.6 nm and d_{ssDNA} = 1.6 nm, which include the hydration shells around dsDNA and ssDNA,[33] and a value for SA molecule diameter of d_p = 5.3 nm determined from crystallographic studies.[34] The dsDNA and ssDNA contour lengths were calculated as $L_{dsDNA} = b_{dsDNA}N_{dsDNA}$ and $L_{ssDNA} = b_{ssDNA}N_{ssDNA}$, where b_{dsDNA} = 0.34 nm and b_{ssDNA} = 0.43 nm, the lengths of dsDNA and ssDNA monomers.[20] It is worth recalling that the mathematical model was developed with no assumptions on protein or DNA sizes except for the assumption that a protein diameter is larger than a Debye length, which is satisfied for proteins larger than 30 kDa. Accordingly, the model is applicable to a wide range of molecular sizes provided that the preceding assumption and assumptions of a globular protein and a rodlike DNA are satisfied. In general, shorter DNA tags are beneficial, as they would allow small proteins to introduce great mobility shifts for the ligands

In addition to the sizes of protein and DNA, we need to experimentally find electrophoretic mobilities and velocities for the protein and DNA tag. Finding these mobilities requires, in turn, the knowledge of v_{EOF}. To facilitate finding v_{EOF}, a neutral marker (NM) was added to the protein−DNA mixture in each experiment. An internal standard (IS) was added for correcting migration time variation between trials. Neither NM nor IS interacted with the ligand or the protein.

SA was labeled with chromeo, a fluorogenic dye that does not change the mobility of protein;[35] we also confirmed that the labeling did not significantly affect protein binding to biotinylated DNA. The protein could, thus, be detected with laser-induced fluorescence (LIF) at 610 nm. The biotinylated DNAs were end-labeled with the alexa dye for LIF detection at 520 nm. Protein−ligand complexes exhibited fluorescence at both wavelengths. Examples of migration patterns of protein, ds-ssDNA, and their complex are shown in Figure 4. SA is a homotetramer that can bind up to four molecules of biotin, depending on the SA/biotin concentration ratio. The peak next to IS at the right corresponds to a complex of one tetrameric SA with two molecules of biotin-containing ligand. However, in the present study we focus only on the complex with 1:1 stoichiometry.

Electrophoretic mobilities of both free DNA (in the absence of protein) and free protein (in the absence of DNA) were found to be negative, which indicated that they were both negatively charged. As a result, the complex was also negatively charged and its experimentally measured electrophoretic mobility was negative. The absolute value of protein's electrophoretic mobility was found to be significantly less than that of dsDNA. Using the current model, we calculated electrophoretic mobilities (Table 1) and migration times (Table S1 in Supporting Information) of complexes for all DNA tags. The presence of two markers was essential to ensure the precision of measured migration times and calculated mobilities. In our case, RSD was 1% for both mobility and migration times.

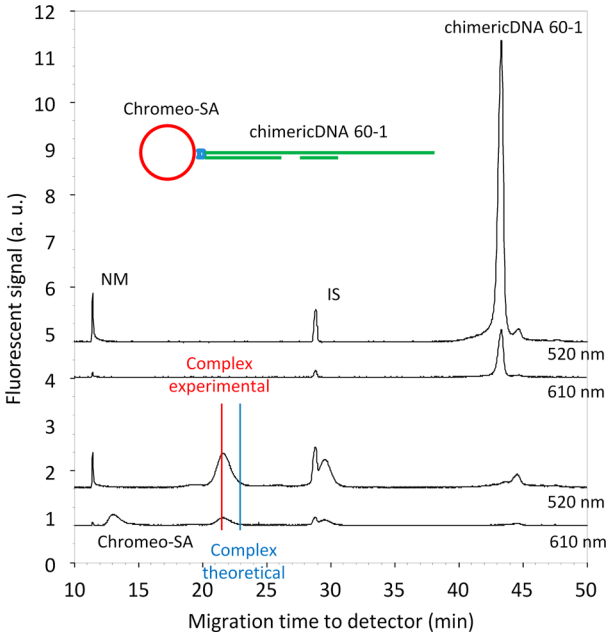


Figure 4. Migration analysis of complex between chromeo-SA and ds-ssDNA chimera. The top two traces represent control experiment with different detection wavelengths. The control contains 100 nM ds-ssDNA (60-1), neutral marker (NM), and internal standard (IS). The bottom two traces represent binding, which has the same composition as control plus 1 μM chromeo-SA. Experimental and theoretical positions of the complex are highlighted with red and blue lines, respectively. Traces are offset vertically for clarity. All experiments were performed in triplicate, and representative traces are shown. A schematic illustration of complex used in these experiments is shown in the top panel.

Electrophoretic mobility of complex increased with the overall contour length, which is reasonable, as both DNA and protein are negatively charged; also, as the major contributor to the charge, DNA has the major influence on complex mobility. The dsDNA, however, has higher electrophoretic mobility than ds-ssDNA with similar contour lengths, which is also anticipated as dsDNA has a more rigid rodlike shape and hence experiences less friction. On the other hand, ds-ssDNA has patches of more flexible ssDNA, which can form random-coil-like structure and hence experiences greater friction. This is the likely explanation of systematic overestimation for theoretical complex mobility with ds-ssDNA (Table 1) as the model was based on the assumption of DNA with rigid rod shape; thus, the presence of ssDNA patches introduces flexibility and decreases experimental mobility. By comparing the predicted and experimental values of complex mobility and migration time, we found the accuracy of prediction. For all tested DNA tags, deviation of predicted from experimental values did not exceed 11%. The same data have also been tested by using the previously developed model, where DNA is considered as a rigid rod shape with constant diameter. The old model resulted in approximately doubled errors for ds-ssDNA.

Finally, we tested two ligands, biotin and GLCBS-L-leucine, which were synthesized with DNA tags identical to those used in DELSMs by GlaxoSmithKline. Both ligands had the same DNA structure: a combination of two dsDNA (total of 94 bp) and two ssDNA (total of 23 nt) regions. DNA tags were labeled with alexa to facilitate LIF detection at 520 nm. SA and CAII were both labeled with chromeo for their LIF detection at 610

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498−5506

Table 1. Electrophoretic Mobilities of Complexes between SA and ds-ssDNA Chimeras of Different Structures[a]

ds-ssDNA name	Structure	Experimental complex mobility, $mm^2(kV \cdot s)^{-1}$	Theoretical complex mobility, $mm^2(kV \cdot s)^{-1}$	Difference between theoretical and experimental complex mobility
20ds		13.39 ± 0.36	12.63 ± 0.39	6%
40ds		16.44 ± 0.08	16.70 ± 0.01	2%
80ds		20.07 ± 0.27	20.80 ± 0.13	4%
120ds		21.95 ± 0.08	22.81 ± 0.01	4%
40-0		15.15 ± 0.16	14.03 ± 0.70	7%
60-0		16.74 ± 0.08	17.48 ± 0.12	4%
80-0		17.76 ± 0.06	18.59 ± 0.07	5%
40-1		15.82 ± 0.02	16.37 ± 0.05	3%
60-1		17.05 ± 0.09	18.38 ± 0.11	8%
60-2		16.87 ± 0.09	18.42 ± 0.04	9%
80-1		17.64 ± 0.11	19.28 ± 0.07	9%
80-2		17.73 ± 0.04	19.30 ± 0.05	9%
80-3		17.63 ± 0.12	19.29 ± 0.07	9%
60-1-2		16.97 ± 0.03	18.66 ± 0.09	10%
80-1-2		17.75 ± 0.03	19.35 ± 0.11	9%
80-2-3		17.65 ± 0.07	19.36 ± 0.08	10%
80-1-3		17.70 ± 0.05	19.49 ± 0.06	10%
80-1-2-3		17.61 ± 0.14	19.50 ± 0.03	11%

[a]See Figure S1 in Supporting Information for detailed sequences and structures of the chimeras. Precision of experimental and theoretical complex mobility is presented as one standard deviation of results from the mean value, based on three experiments.

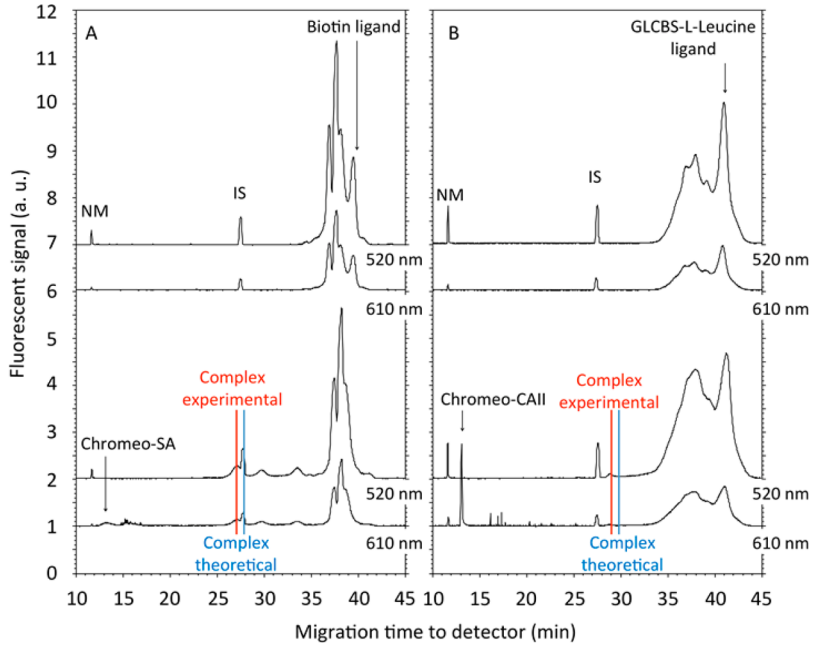


Figure 5. Migration study for protein–ligand complex between (A) 1 μM chromeo-SA and 100 nM biotin ligand and (B) 5 μM chromeo-CAII and 1 μM GLCBS-L-leucine ligand. In each panel, the top two traces represent the no-protein control, which contained 100 nM ligand, neutral marker (NM), and internal standard (IS). The bottom two traces correspond to the protein–ligand binding experiment, which had the same composition as the control plus 1 μM protein. Experimental and theoretical positions of the complexes are highlighted with red and blue lines, respectively. Traces are offset vertically for clarity. All experiments were performed in triplicate, and representative traces are shown.

nm. Protein–ligand complexes, thus, contained both fluorophores and could be detected at both 520 and 610 nm.

When sampled in CE without proteins, the unbound ligands revealed several peaks, suggesting that, in addition to the ligands, the samples contained impurities. The impurities were identified as the starting material and the intermediates from each step of synthesis used for manufacturing of DELSMs (Figure S2 in Supporting Information). The full-length ligand contains the most negative charge and bears the highest electrophoretic mobility. Accordingly, it was identified as the rightmost peak in the electropherogram. In this study, we focused on migration patterns of full-length ligand and the corresponding protein–ligand complex.

Electropherograms for protein–ligand binding experiments are shown in Figure 5. In each panel, the top two traces are the no-protein control and the bottom two traces correspond to sampling the protein–ligand mixture. SA is built of four subunits and can bind up to four biotin molecules. Accordingly,

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498–5506

complexes with different binding stoichiometries are seen in panel A. In this study, we considered only 1:1 binding; the corresponding complex is indicated by the red line in panel A. By using our mathematical model, we found electrophoretic mobilities of protein−ligand complexes of 21.97 ± 0.09 $mm^2(kV{\cdot}s)^{-1}$ for biotin ligand and 23.35 ± 0.04 $mm^2(kV{\cdot}s)^{-1}$ for GLCBS-L-leucine ligand. Deviations between experimentally and theoretically determined complex mobilities were found to be 3% for biotin ligand and 5% for GLCBS-L-leucine ligand. Such accurate prediction will guarantee accurate collection of protein−ligand fraction in selection experiments.

In conclusion, we have developed a versatile approach for accurate prediction of electrophoretic mobility and migration time of protein−ligand complexes for selection of protein binders from DELSMs. We consider a globular protein attached to a DNA tag at a single point and use the thin double layer model to find mobilities of protein, dsDNA, and ssDNA. We then determine effective electric and hydrodynamic forces acting upon protein, dsDNA, and ssDNA and express the complex mobility in terms of experimentally measurable mobilities of protein and DNA-tagged ligand. The model for complex mobility was tested through studying the mobilities of protein−ligand complexes for ligands with varying structures of DNA tags: 4 dsDNAs and 14 ds-ssDNAs. It was also validated by use of two small molecules with DNA tags identical to those used by GlaxoSmithKline in their DELSMs. The accuracy and ruggedness of our model were confirmed by comparing predicted complex mobility and migration time with experimentally measured values. The model is feasible for analyzing DELSMs with various lengths and composition of DNA tags. In addition, the model is generic and expected to be applicable to all proteins with near-globular shapes and molecular weights of 30 kDa or more and any DELSMs with a rodlike DNA part and a ligand attached to the end of DNA. We foresee that this approach will help to advance kinetic capillary electrophoresis methods to their practical use in selection of drug leads from DELSMs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b00980.

Additional text describing synthesis of biotin ligand and GLCBS-L-leucine ligand and purification of intermediates; two figures showing structural details of dsDNA and chimeric-DNA and migration studies of biotin-DEL; one table listing migration times of complexes of streptavidin with different ds-ssDNA chimeras; and text and 16 equations showing mathematical model for chimeras with multiple dsDNA-ssDNA sections (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail skrylov@yorku.ca.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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

■ REFERENCES

(1) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. *Br. J. Pharmacol.* **2011**, *162*, 1239−1249.
(2) Jorgensen, W. L. *Acc. Chem. Res.* **2009**, *42*, 724−733.
(3) Keserű, G. M.; Makara, G. M. *Drug Discovery Today* **2006**, *11*, 741−748.
(4) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5381−5383.
(5) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotechnol.* **2004**, *22*, 568−574.
(6) Pfaffl, M. W.; Hageleit, M. *Biotechnol. Lett.* **2001**, *23*, 275−282.
(7) Franzini, R. M.; Neri, D.; Scheuermann, J. *Acc. Chem. Res.* **2014**, *47*, 1247−1255.
(8) Li, X.; Liu, D. R. *Angew. Chem., Int. Ed.* **2004**, *43*, 4848−4870.
(9) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. *Nat. Chem. Biol.* **2009**, *5*, 647−654.
(10) Cooper, A. M. *Nat. Rev. Drug Discovery* **2002**, *1*, 515−528.
(11) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505−510.
(12) Papoulas, O. Rapid separation of protein-bound DNA from free DNA using nitrocellulose filters. *Current Protocols in Molecular Biology*; Wiley: Hoboken, NJ, 2001; Unit 12.8.10.1002/0471142727.mb1208s36
(13) Musheev, M. U.; Kanoatov, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2013**, *135*, 8041−8046.
(14) Bao, J.; Krylova, S. M.; Cherney, L. T.; Hale, R. L.; Belyanskaya, S. L.; Chiu, C. H.; Arico-Muendel, C. C.; Krylov, S. N. *Anal. Chem.* **2015**, *87*, 2474−2479.
(15) Kleiner, R. E.; Dumelin, C. E.; Liu, D. R. *Chem. Soc. Rev.* **2011**, *40*, 5707−5717.
(16) Viovy, J. L. *Rev. Mod. Phys.* **2000**, *72*, 813−872.
(17) Mayer, P.; Slater, G. W.; Drouin, G. *Anal. Chem.* **1994**, *66*, 1777−1780.
(18) Hubert, S. J.; Slater, G. W. *Electrophoresis* **1995**, *16*, 2137−2142.
(19) Desruisseaux, C.; Long, D.; Drouin, G.; Slater, G. W. *Macromolecules* **2001**, *34*, 44−52.
(20) 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.
(21) Meagher, R. J.; McCormick, L. C.; Haynes, R. D.; Won, J. I.; Lin, J. S.; Slater, G. W.; Barron, A. E. *Electrophoresis* **2006**, *27*, 1702−1712.
(22) Long, D.; Dobrynin, A. V.; Rubinstein, M.; Ajdari, A. *J. Chem. Phys.* **1998**, *108*, 1234−1244.
(23) Teraoka, I. *Polymer Solutions: An Introduction to Physical Properties*; John Wiley & Sons: New York, 2002; DOI: 10.1002/0471224510.
(24) Strobl, G. R. *the Physics of Polymers: Concepts for Understanding Their Structures and Behavior*; Springer: Berlin, 2007; DOI: 10.1007/978-3-540-68411-4.
(25) Protein size calculator: www.calctool.org/CALC/prof/bio/protein_size.
(26) Simmel, F. C.; Dittmer, W. U. *Small* **2005**, *1*, 284−299.
(27) Krishnan, Y.; Simmel, F. C. *Angew. Chem., Int. Ed.* **2011**, *50*, 3124−3156.
(28) Manning, G. S. *J. Chem. Phys.* **1969**, *51*, 924−933.
(29) Manning, G. S. *J. Phys. Chem.* **1981**, *85*, 1506−1515.
(30) Barrat, J. L.; Joanny, J. F. *Adv. Chem. Phys.* **1996**, *94*, 1−66.
(31) Anik, N.; Airiau, M.; Labeau, M.-P.; Vuong, C.-T.; Reboul, J.; Lacroix-Desmazes, P.; Gérardin, C.; Cottet, H. *Macromolecules* **2009**, *42*, 2767−2774.
(32) Ibrahim, A.; Koval, D.; Kašička, V.; Faye, C.; Cottet, H. *Macromolecules* **2013**, *46*, 533−540.

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498−5506

(33) Schneider, B.; Patel, K.; Berman, H. M. *Biophys. J.* **1998**, *75*, 2422−2434.
(34) Hendrickson, W. A.; Pähler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 2190−2194.
(35) de Jong, S.; Krylov, S. N. *Anal. Chem.* **2011**, *83*, 6330−6335.

DOI: 10.1021/acs.analchem.6b00980
Anal. Chem. 2016, 88, 5498−5506

SUPPORTING INFORMATION

Predicting Electrophoretic Mobility of Protein-Ligand Complexes for Ligands from DNA-Encoded Librares of Small Molecules

Jiayin Bao,[1] Svetlana M. Krylova,[1] Leonid T. Cherney,[1] Robert L. Hale,[2] Svetlana L. Belyanskaya,[2] Cynthia H. Chiu,[2] Alex Shaginian,[2] Christopher C. Arico-Muendel,[2] and Sergey N. Krylov[1*]

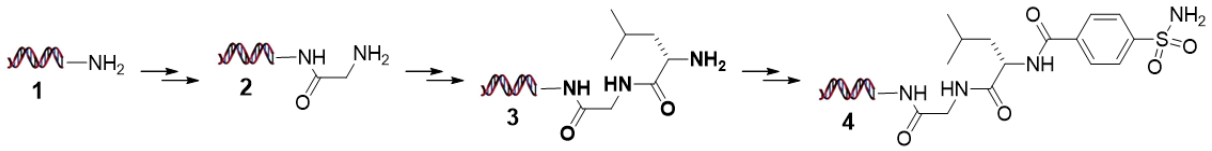
[1]*Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada*
[2]*GlaxoSmithKline, 830 Winter St., Waltham, MA 02451-8714, USA*

Table of Contents

1. Supporting Materials and Methods

1.1. Synthetic of biotin ligand and GLCBS-L-Leucine ligand

The following are the reaction steps:



Detailed synthetic procedures for DNA headpiece and on-DNA chemistry have been previously reported [1].

Intermediate **2**. DNA headpiece (1 mM DNA headpiece solution; 100 µL, 100 nmol) in pH 9.4 sodium borate buffer (250 mM) was treated with a solution of Fmoc-Gly-OH/HATU-active ester (see below for preparation) in DMF (60 eq). The mixture was vortexed, kept at room temperature for 12 h, and subjected to ethanol precipitation. The resultant DNA pellet was collected, dissolved in water (100 µL), and treated with piperidine (50 µL). The mixture was vortexed, kept at room temperature for 30 min, and subjected to ethanol precipitation. The resultant DNA pellet **2** was collected and dried under vacuum.

Intermediate **3**. DNA pellet **2** was dissolved in 100 µL 250 mM sodium borate buffer at pH 9.4 and treated with a solution of Fmoc-Leu-OH/HATU-active ester (see below for preparation) in DMF (60 eq). The mixture was vortexed, kept at room temperature for 12 h, and subjected to ethanol precipitation. The resultant DNA pellet was collected, dissolved in 100 µL of water, and treated with piperidine (50 µL). The mixture was vortexed, kept at room temperature for 30 min, and subjected to ethanol precipitation. The resultant DNA pellet **3** was collected and dried under vacuum.

Product **4**. DNA pellet **3** was dissolved in 250 mM pH 9.4 sodium borate buffer (100 µL) and treated with a solution of benzoic acid 4-sulfonamide/HATU-active ester (see below for preparation) in DMF (60 eq). The mixture was vortexed, kept at room temperature for 12 h, and subjected to ethanol precipitation. The resultant DNA pellet **4** was collected, dried under vacuum, and dissolved in water (100 µL) to afford a 1 mM solution of product **4** that was used without further purification.

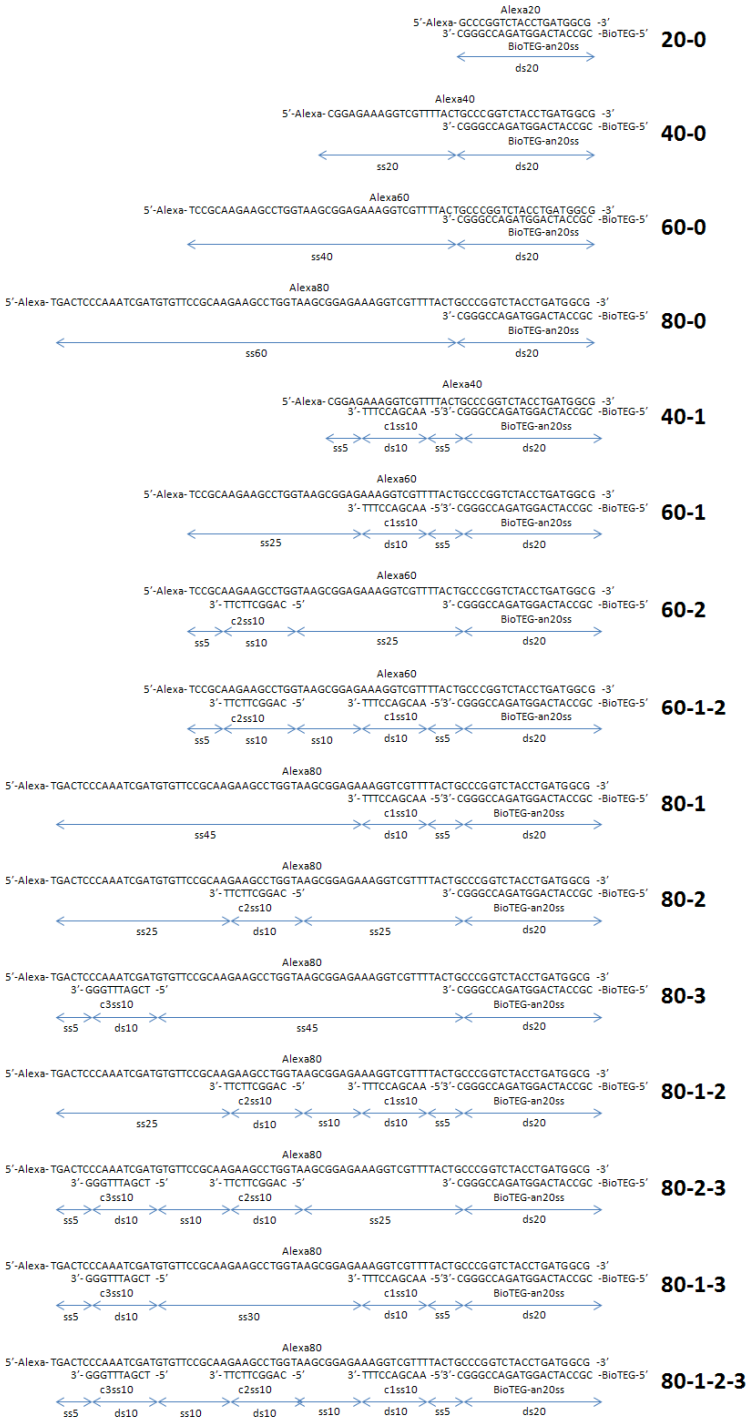
Preparation of HATU active esters. Carboxylic acid (200 mM in DMF, 30 µL, 6000 nmol) was combined with DIEA (200 mM in DMF, 30 µL, 6000 nmol), and HATU (200 mM in DMF, 30 µL, 6000 nmol) at 4°C. The mixture was vortexed, kept at 4°C for 20 min, and transferred into the tube containing the DNA starting material.

1.2. Purification of intermediates

Possible components that could contribute to signal were gel purified. Gel reagents and SYBR Gold were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Constructs containing the non-covalently attached splint were run on a 6% DNA retardation gel according to the protocol recommended by the manufacturer. Removal of the splint was attained by subjecting samples to electrophoresis on 10% TBE-urea gels according to the protocol recommended by the manufacturer. Gels were stained with SYBR Gold diluted 1:1000 in 30 mL TBE buffer (89 mM Tris base, 89 mM Boric acid, 2 mM EDTA, pH 8.0) for 20 min and visualized by 254 nm transillumination. Corresponding bands were excised by razor and placed in 1.5 mL tubes. Gel slices were crushed and eluted in 0.5 mL passive elution buffer (300 mM NaOAc, 20 mM

EDTA) overnight with shaking at room temperature. Liquid was separated from gel slices using Ultafree-MC 0.22 μm pore size spin filers (Millapore, Billerica, MA, USA) by centrifugation for 1 min at 13,000 rpm. The eluate was mixed with 1 mL isopropanol (–20°C) and centrifuged for 30 min at 4°C and the supernatant was removed from the final product.

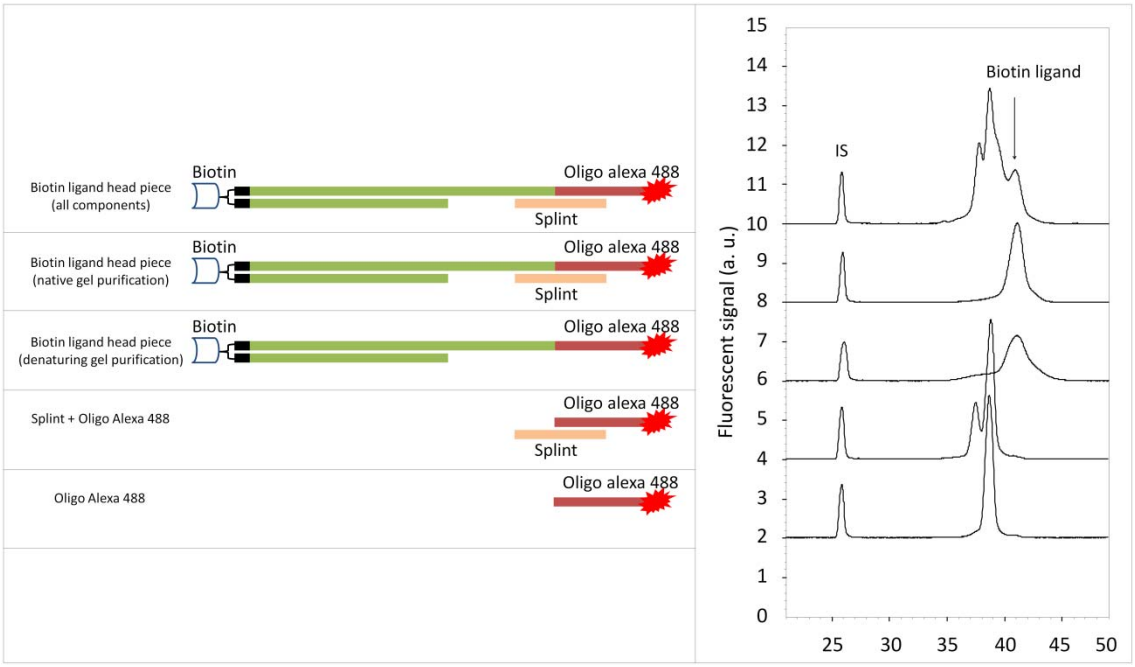
1.3. Structural details of dsDNA and chimeric-DNA



Supporting Figure 1. Structural details of dsDNA and chimeric-DNA.

2. Supporting Results

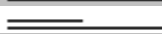
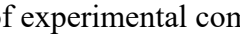
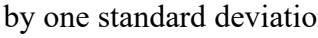
2.1. Migration studies of biotin-DEL



Supporting Figure 2. Peak identification of biotin-DEL. The schematic structures of components are illustrated on the left panel; the corresponding electropherograms are on the right panel. Each sample contained 100 nM analyzed component with 10 nM internal standard (IS). All experiments were done in triplicates.

2.2. Migration times of complexes of streptavidin with different ds-ssDNA chimeras

Supporting Table 1. Migration times of complexes of streptavidin with ds-ssDNA chimeras of different structures. See **Supporting Figure 1** for detailed sequences and structures of the chimeras.

Chimeric DNA Name	Structure	Experimental complex migration time, s	Theoretical complex migration time, s	Deviation of theoretical migration time from experimental
20ds		999.8 ± 3.7	970.6 ± 6.1	6%
40ds		961.0 ± 0.4	971.6 ± 2.2	2%
80ds		1128.7 ± 10.1	1170.3 ± 3.6	4%
120ds		1377.7 ± 42.9	1453.1 ± 52.3	4%
40-0		1100.6 ± 18.4	1049.7 ± 41.0	7%
60-0		1176.3 ± 33.9	1218.2 ± 32.6	4%
80-0		1305.9 ± 15.9	1364.7 ± 17.9	5%
40-1		1169.3 ± 7.1	1200.1 ± 3.7	3%
60-1		1291.2 ± 3.7	1386.4 ± 7.0	8%
60-2		1283.0 ± 1.3	1394.5 ± 1.6	9%
80-1		1323.2 ± 1.6	1448.6 ± 7.1	9%
80-2		1332.3 ± 3.9	1453.8 ± 4.5	9%
80-3		1340.3 ± 12.4	1471.2 ± 7.6	9%
60-1-2		1276.5 ± 1.6	1397.5 ± 9.2	10%
80-1-2		1324.1 ± 2.2	1446.6 ± 10.0	9%
80-2-3		1336.8 ± 1.9	1470.4 ± 3.4	10%
80-1-3		1314.1 ± 5.6	1450.6 ± 14.1	10%
80-1-2-3		1315.2 ± 3.7	1460.4 ± 8.2	11%

Precisions of experimental complex migration time and theoretical complex migration time are represented by one standard deviation of results from the mean of three repeated experiments.

3. Supporting Mathematics

Mathematical model for chimeras with multiple dsDNA-ssDNA sections

Let us consider the mobility of protein-DNA complex in the case of chimeric DNA that has the following structure dsDNA-ssDNA-…-dsDNA-ssDNA. The mobility of each dsDNA section can be determined by expressions (5) from the main text:

$$\mu_{\mathrm{dsDNA},i} = \frac{\varepsilon_0 \varepsilon_\mathrm{r} \zeta_{\mathrm{dsDNA},i}}{\eta} \approx \frac{-\sigma_{\mathrm{dsDNA},i} \lambda_\mathrm{D}}{\eta}, \quad \zeta_{\mathrm{dsDNA},i} \approx \frac{-\sigma_{\mathrm{dsDNA},i} \lambda_\mathrm{D}}{\varepsilon_0 \varepsilon_\mathrm{r}} \tag{S1}$$

where subscript $i = 1, \ldots, N$ is the number of the dsDNA section under consideration. Equation (S1) can be rewritten similarly to relation (6):

$$\mu_{\mathrm{dsDNA},i} = \frac{-\sigma_{\mathrm{dsDNA},i} \lambda_\mathrm{D}}{\eta} = \frac{q_{\mathrm{dsDNA},i} \lambda_\mathrm{D}}{\pi \eta d_{\mathrm{dsDNA}}}, \quad \sigma_{\mathrm{dsDNA},i} = -\frac{q_{\mathrm{dsDNA},i}}{\pi d_{\mathrm{dsDNA}}} \tag{S2}$$

Expression (S2) for $\mu_{\mathrm{dsDNA},i}$ can be also obtained from the balance of electric and hydrodynamic forces, $F_{\mathrm{E,dsDNA},i}$ and $F_{\mathrm{H,dsDNA},i}$, acting upon the i-th dsDNA section:

$$F_{\mathrm{E,dsDNA},i} + F_{\mathrm{H,dsDNA},i} = 0 \tag{S3}$$

if we assume the following effective values for these forces:

$$F_{\mathrm{E,dsDNA},i} = q_{\mathrm{dsDNA},i} L_{\mathrm{dsDNA},i} E, \quad F_{\mathrm{H,dsDNA},i} = -\frac{\pi \eta d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i}}{\lambda_\mathrm{D}} u_{\mathrm{dsDNA},i} \tag{S4}$$

Here, $L_{\mathrm{dsDNA},i}$ is the contour length of the i-th dsDNA section, $u_{\mathrm{dsDNA},i}$ is a relative velocity of this dsDNA section with respect to the buffer.

The mobility of each ssDNA section can be determined from expressions (10) in the main text:

$$\mu_{\mathrm{ssDNA},i} = \frac{\varepsilon_0 \varepsilon_\mathrm{r} \zeta_{\mathrm{ssDNA},i}}{\eta} \approx \frac{-\sigma_{\mathrm{ssDNA},i} \lambda_\mathrm{D}}{\eta}, \quad \zeta_{\mathrm{ssDNA},i} \approx \frac{-\sigma_{\mathrm{ssDNA},i} \lambda_\mathrm{D}}{\varepsilon_0 \varepsilon_\mathrm{r}} \tag{S5}$$

where subscript $i = 1, \ldots, N$ is the number of ssDNA section. Equation (S5) can be rewritten similarly to relation (11):

$$\mu_{\mathrm{ssDNA},i} = \frac{-\sigma_{\mathrm{ssDNA},i} \lambda_\mathrm{D}}{\eta} = \frac{q_{\mathrm{ssDNA},i} \lambda_\mathrm{D}}{\pi \eta d_{\mathrm{ssDNA}}}, \quad \sigma_{\mathrm{ssDNA},i} = -\frac{q_{\mathrm{ssDNA},i}}{\pi d_{\mathrm{ssDNA}}} \tag{S6}$$

Expression (S6) for $\mu_{\mathrm{ssDNA},i}$ can be also obtained from the balance of electric and hydrodynamic forces, $F_{\mathrm{E,ssDNA},i}$ and $F_{\mathrm{H,ssDNA},i}$, acting upon the i-th ssDNA section:

$$F_{\mathrm{E,ssDNA},i} + F_{\mathrm{H,ssDNA},i} = 0 \tag{S7}$$

if we assume the following effective values for these forces:

$$F_{\mathrm{E,ssDNA},i} = q_{\mathrm{ssDNA},i} L_{\mathrm{ssDNA},i} E, \quad F_{\mathrm{H,ssDNA},i} = -\frac{\pi \eta d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i}}{\lambda_\mathrm{D}} u_{\mathrm{ssDNA},i} \tag{S8}$$

Here, $L_{\mathrm{ssDNA},i}$ is the contour length of the i-th ssDNA section, $u_{\mathrm{ssDNA},i}$ is a relative velocity of this ssDNA section with respect to the buffer.

The electrophoretic mobility of a globular protein attached to the end of the dsDNA-ssDNA chimera can be found from a balance of all effective forces acting upon such a complex:

$$F_{\mathrm{E,P}} + \sum_{i=1}^{N} F_{\mathrm{E,dsDNA},i} + \sum_{i=1}^{N} F_{\mathrm{E,ssDNA},i} + F_{\mathrm{H,P}} + \sum_{i=1}^{N} F_{\mathrm{H,dsDNA},i} + \sum_{i=1}^{N} F_{\mathrm{H,ssDNA},i} = 0 \tag{S9}$$

Substitution of expressions (4), (S4), and (S8) into equation (S9) gives:

$$\left(Q_{\mathrm{P}} + \sum_{i=1}^{N} q_{\mathrm{dsDNA},i} L_{\mathrm{dsDNA},i} + \sum_{i=1}^{N} q_{\mathrm{ssDNA},i} L_{\mathrm{ssDNA},i} \right) E = \left(\frac{\pi \eta d_{\mathrm{P}}^2}{\lambda_{\mathrm{D}}} + \sum_{i=1}^{N} \frac{\pi \eta d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i}}{\lambda_{\mathrm{D}}} + \sum_{i=1}^{N} \frac{\pi \eta d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i}}{\lambda_{\mathrm{D}}} \right) u_{\mathrm{comp}} \tag{S10}$$

Here, we assume that:

$$u_{\mathrm{P}} = u_{\mathrm{comp}}, \quad u_{\mathrm{dsDNA},i} = u_{\mathrm{comp}}, \quad u_{\mathrm{ssDNA},i} = u_{\mathrm{comp}} \tag{S11}$$

Solving equation (S10) with respect to u_{comp} and taking into account that $u_{\mathrm{comp}} = \mu_{\mathrm{comp}} E$ we obtain the electrophoretic mobility of the complex μ_{comp}:

$$\mu_{\mathrm{comp}} = \frac{Q_{\mathrm{P}} + \sum_{i=1}^{N} q_{\mathrm{dsDNA},i} L_{\mathrm{dsDNA},i} + \sum_{i=1}^{N} q_{\mathrm{ssDNA},i} L_{\mathrm{ssDNA},i}}{\frac{\pi \eta d_{\mathrm{P}}^2}{\lambda_{\mathrm{D}}} + \sum_{i=1}^{N} \frac{\pi \eta d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i}}{\lambda_{\mathrm{D}}} + \sum_{i=1}^{N} \frac{\pi \eta d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i}}{\lambda_{\mathrm{D}}}} \tag{S12}$$

Taking into account expressions (2), (S2), and (S6) for the electrophoretic mobilities of the globular protein, the i-th dsDNA section, and i-th ssDNA section we can rewrite relation (S12) as follows:

$$\mu_{\mathrm{comp}} = \frac{d_{\mathrm{P}}^2 \mu_{\mathrm{P}} + \sum_{i=1}^{N} d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i} \mu_{\mathrm{dsDNA},i} + \sum_{i=1}^{N} d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i} \mu_{\mathrm{ssDNA},i}}{d_{\mathrm{P}}^2 + \sum_{i=1}^{N} d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i} + \sum_{i=1}^{N} d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i}} \tag{S13}$$

In the absence of the protein, relation (S13) reduces to expression for the electrophoretic mobility of the dsDNA-ssDNA chimera, μ_{chim}:

$$\mu_{\mathrm{chim}} = \frac{\sum_{i=1}^{N} d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i} \mu_{\mathrm{dsDNA},i} + \sum_{i=1}^{N} d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i} \mu_{\mathrm{ssDNA},i}}{\sum_{i=1}^{N} d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA},i} + \sum_{i=1}^{N} d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA},i}} \tag{S14}$$

Given expression (18), we can express the electrophoretic mobility of complex in terms of the electrophoretic mobilities of the protein and the DNA chimera:

$$\mu_{\mathrm{comp}} = \frac{d_{\mathrm{P}}^2 \mu_{\mathrm{P}} + \left(d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA}} + d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA}} \right) \mu_{\mathrm{chim}}}{d_{\mathrm{P}}^2 + d_{\mathrm{dsDNA}} L_{\mathrm{dsDNA}} + d_{\mathrm{ssDNA}} L_{\mathrm{ssDNA}}} \tag{S15}$$

where L_{dsDNA} is the total contour length of all dsDNA sections and L_{ssDNA} is the total contour length of all ssDNA sections. They are determined by expressions:

$$L_{\mathrm{dsDNA}} = \sum_{i=1}^{N} L_{\mathrm{dsDNA},i}, \quad L_{\mathrm{ssDNA}} = \sum_{i=1}^{N} L_{\mathrm{ssDNA},i} \tag{S16}$$

4. Reference

[1] Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol*. **2009**, *5*, 647-654.