

Using Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM) for Simultaneous Determination of Concentration and Equilibrium Constant

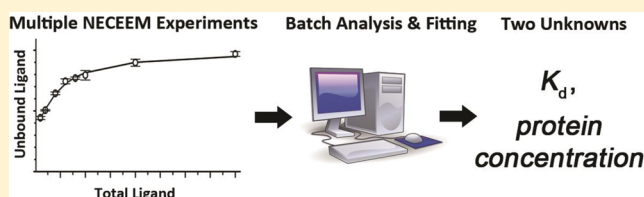
Mirzo Kanoatov,[†] Victor A. Galievsky,[†] Svetlana M. Krylova,[†] Leonid T. Cherney,[†] Hanna K. Jankowski,[‡] and Sergey N. Krylov^{*,†}

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

[‡]Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada

S Supporting Information

ABSTRACT: Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) is a versatile tool for studying affinity binding. Here we describe a NECEEM-based approach for simultaneous determination of both the equilibrium constant, K_d , and the unknown concentration of a binder that we call a target, T. In essence, NECEEM is used to measure the unbound equilibrium fraction, R , for the binder with a known concentration that we call a ligand, L. The first set of experiments is performed at varying concentrations of T, prepared by serial dilution of the stock solution, but at a constant concentration of L, which is as low as its reliable quantitation allows. The value of R is plotted as a function of the dilution coefficient, and dilution corresponding to $R = 0.5$ is determined. This dilution of T is used in the second set of experiments in which the concentration of T is fixed but the concentration of L is varied. The experimental dependence of R on the concentration of L is fitted with a function describing their theoretical dependence. Both K_d and the concentration of T are used as fitting parameters, and their sought values are determined as the ones that generate the best fit. We have fully validated this approach *in silico* by using computer-simulated NECEEM electropherograms and then applied it to experimental determination of the unknown concentration of MutS protein and K_d of its interactions with a DNA aptamer. The general approach described here is applicable not only to NECEEM but also to any other method that can determine a fraction of unbound molecules at equilibrium.



Noncovalent (reversible) affinity binding of biomolecules plays important roles in cellular regulation and in biomedical diagnostics.¹ The binding reaction of a target (T) and a ligand (L) with the formation of an affinity complex (TL) can be described by a simple chemical equation:



where k_{on} and k_{off} are rate constants of binding and dissociation, respectively. Complex stability is typically characterized by the equilibrium dissociation constant:

$$K_d = \frac{[T]_{\text{eq}}[L]_{\text{eq}}}{[TL]_{\text{eq}}} \quad (2)$$

where $[T]_{\text{eq}}$, $[L]_{\text{eq}}$, and $[TL]_{\text{eq}}$ are concentrations of the three components at equilibrium.

Affinity methods constitute a broad class of techniques, some of which can facilitate the determination of k_{on} , k_{off} , and K_d , as well as finding of an unknown concentration of the target by detecting its complex with a ligand of known concentration.^{2,3} Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) is an affinity method which relies on

separation of T, L, and TL by CE. In particular, it has been used to (i) determine K_d if the concentrations of both T and L are known or (ii) determine the unknown concentration of T if K_d and the concentration of L are known.^{4,5} In some cases, however, both the K_d and the concentration of T are unknown—an example is difficult-to-purify proteins. Not knowing the concentration of T prevents NECEEM with the previously used mathematical treatment from finding K_d , and, in turn, from finding the concentration of T in the sample. Here we describe an approach that allows NECEEM to be used for simultaneous determination of both K_d and the concentration of T. This approach requires NECEEM experiments to be performed at varying concentrations of L but at a constant unknown concentration of T. The experimental results are presented as a fraction of unbound L, denoted by R , versus concentration of L and fitted with a theoretical dependence of these variables in which K_d and the concentration of T are used as fitting parameters. These parameters are determined as those

Received: January 13, 2015

Accepted: February 10, 2015

Published: February 10, 2015



resulting in the best fit. Our approach requires finding R from multiple NECEEM electropherograms. To make finding R more accurate, unbiased, and fast we have developed software that deconvolutes areas of NECEEM peaks and calculates R based on a recently published theory,⁶ for sets of NECEEM electropherograms in batch. To test the suggested approach for finding K_d and the target concentration, as well as to validate the software serving it, COMSOL multiphysics was utilized to accurately model the NECEEM processes and simulate NECEEM electropherograms at various sets of initial conditions *in silico*. The simulated electropherograms were processed as experimental ones, and the returned values of K_d and concentration of T were confirmed to be within 10% of the values programmed into the simulations over a wide range of initial conditions, demonstrating the remarkable stability of the developed approach. Lastly, the *in silico* validated method was used to determine the protein concentration of in-house expressed and purified MutS protein and the K_d of its interaction with a DNA aptamer. The results suggest that the previously measured K_d value for this interaction was significantly overestimated by a single-point NECEEM approach.⁷ The determined concentration of the protein was in perfect agreement with measurements performed by an independent method.

■ EXPERIMENTAL SECTION

Materials. All chemicals and buffer components were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. The pETMutS plasmid was obtained from Addgene (MA, USA). The *E. coli* Rossetta-gami 2(DE3) competent cells were purchased from EMD Millipore (PA, USA). Fused-silica capillaries were purchased from Molex (Phoenix, AZ). *Thermus aquaticus* MutS recombinant protein was expressed in Rossetta-gami 2 (DE3) competent *E. coli* cells transformed by the pETMutS plasmid. The protein was purified by an aptamer-facilitated affinity pulldown method, similar to that recently described.⁸ The stock concentration of the purified MutS protein was measured using a bicinchoninic acid (BCA) assay kit (Thermo Scientific, Wilmington, DE), according to manufacturer's instructions. A solution of Bovine Serum Albumin (BSA) of known concentration was used to make calibration curve standards. As BCA assay detects peptide bonds, the final concentration values were normalized according to the ratio of peptide bonds between MutS (816 bonds) and BSA standard (582 bonds). A DNA aptamer (clone ID: 2-06) with affinity toward MutS protein was selected previously^{7,9} and was custom synthesized by Integrated DNA Technologies (Coralville, IA). The nucleotide sequence of the fluorescently labeled, single-stranded DNA MutS aptamer was 5'-/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'. A NanoDrop-1000 spectrometer (Thermo Scientific, Wilmington, DE) was used to verify the concentration of the DNA stock by measuring light absorbance at 260 nm and dividing the absorbance by a manufacturer-provided extinction coefficient.

Capillary Electrophoresis. Analyses were carried out using both a commercial and a custom-made CE instruments. The commercial instrument, P/ACE MDQ (Beckman Coulter, Mississauga, ON, Canada), was equipped with a standard fluorescence detector and a 488 nm line of continuous Wave Solid-State laser (JDSU, Santa Rosa, CA) for fluorescence excitation. Runs were performed in a 30-cm-long uncoated

capillary, with on-column detection 20 cm away from the capillary inlet. The fused-silica capillary had an inner diameter of 50 μm and an outer diameter of 360 μm . The temperature of the capillary coolant was set to 25 $^{\circ}\text{C}$.

The custom instrument was described in detail previously.^{10,11} The instrument was equipped with a highly sensitive off-column fluorescence detection scheme, with a sheath-flow hydrodynamic focusing cell. A few modifications were incorporated into the existing design to facilitate fluorescent measurements in the required picomolar range of aptamer concentrations. A diode-pumped solid state laser (AixiZ, Houston, TX) was applied for excitation of fluorescence. The laser beam had a diameter of about 2 mm and a power of 60 mW (attenuated to 40 mW) at 473 nm. An optical bandpass filter of 525 ± 25 nm (Semrock, Rochester, NY) was used to select for the fluorescent signal. A 20 Hz low-pass electronic filter was introduced between the photomultiplier tube R1477 (Hamamatsu) and the analog-to-digital signal converter PCI-6036E (National Instruments), to reduce electronic noise. To eliminate random spikes in the signal due to laser-light scattering by gas bubbles and dust particles, a Millex liquid filter unit (Millipore, Billerica, MA) with a 0.22 μm pore size was incorporated into a sheath-flow fluidic line. To further attenuate random signal spikes, the digitized signal was treated by a three-point-median digital filter. Electrophoresis high-voltage electronics were segregated from the detection electronics to reduce the effects of electromagnetic interference on the recorded signal. Runs were performed in a 40-cm-long uncoated capillary, with off-column detection at the end of the capillary. The fused-silica capillary inner diameter was 50 μm , and the outer diameter, 150 μm . The measurements were carried out at room temperature (23 $^{\circ}\text{C}$).

For both instruments, CE experiments were conducted in the following manner. Both the inlet and the outlet reservoirs contained the electrophoresis run buffer (50 mM Tris-HCl at pH 8.5). Prior to every run, the capillary was rinsed with the run buffer solution at 30 psi pressure for 1 min. At the end of each run, the capillary was rinsed with 100 mM HCl, 100 mM NaOH, and deionized water, with the same pressure for 1 min each. The samples were injected into the capillary, prefilled with the run buffer, by a 1 psi pressure pulse of 3.7 and 5.0 s for commercial and custom-made instruments, respectively. The length of the sample plugs in both cases was calculated to be 6.8 mm. Electrophoresis was carried out with a positive electrode at the injection end of the capillary, with electric field strengths of 600 and 400 V/cm for the commercial and custom-made instruments, respectively.

NECEEM. NECEEM experiments were performed to study the interaction between MutS protein and a DNA aptamer (Clone ID: 2-06). All dilutions were prepared with electrophoresis run buffer, to minimize issues from electrolyte mismatch between the sample plug and the run buffer. For all experiments, protein stock solution concentration was treated as unknown, and samples were expressed as a dilution relative-to-stock. The DNA stock solution was prepared at a concentration of 1 μM and had Rhodamine 110 mixed in at a 7 μM concentration, to act as an internal standard. Mixtures of various concentration combinations were prepared, as described in the following sections. For experiments with the commercial instrument, performed at nanomolar concentrations, mixtures were incubated at room temperature for 10 min to achieve equilibration. For experiments with the custom-made instrument, performed at picomolar concentrations, mixtures

were incubated at room temperature for 60 min, as equilibration proceeded more slowly. After incubation, the equilibrium mixtures were subjected to CE as described above. Each experimental point was measured in triplicates, with a fresh equilibrium mixture prepared for each replicate. R coefficients were extracted from resultant electropherograms by the NAAP program (described below) in batch. In the experiments performed on the custom-made instrument, peak “shoulders” were observed between the TL and L peaks (Figure S13), which were not observed in experiments performed on the commercial instrument (Figure S15). We identified these shoulders to be the products of nuclease-facilitated detachment of the fluorophore from the DNA aptamer and excluded them when calculating the R values (see the description of Figure S13 for additional justification). Values of K_d and $[T]_0$ parameters were obtained by fitting the dependence of R on $[L]_0$ into eq 7, using Microsoft Excel 2010 Solver Add-in.

RESULTS AND DISCUSSION

The Concept of NECEEM. In NECEEM, T and L are mixed and reacted to approach equilibrium in reaction 1. A short plug of the equilibrium mixture is injected into a capillary, and the three components, T, TL, and L, are separated by electrophoresis. L is typically labeled with a fluorophore which facilitates detection of both L and TL, while T is not detected. After the initial zones of T, TL, and L are separated, TL starts to dissociate and release L, which is continuously separated from TL. Both intact TL and L that was nonbound in the equilibrium mixture migrate as narrow zones and generate peaks in “signal versus time-to-detector” dependence (Figure 1A, green and blue). In contrast, L, which was gradually

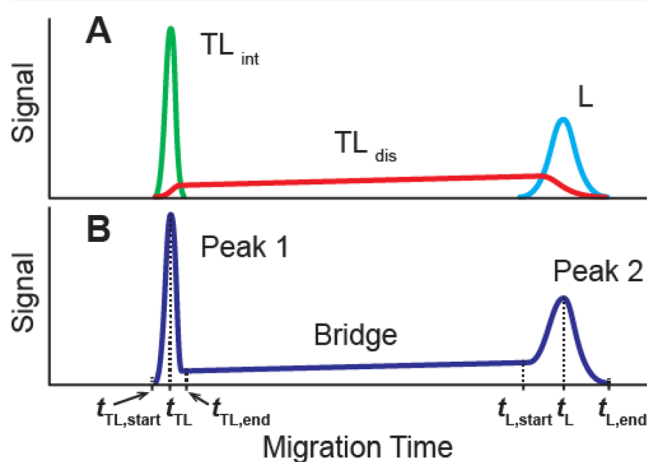


Figure 1. Example of NECEEM data: (A) signal separation into three components, and (B) combined signal in a typical NECEEM electropherogram. Time labels corresponding to the beginnings ($t_{TL,start}$ and $t_{L,start}$), the ends ($t_{TL,end}$ and $t_{L,end}$), and the maxima (t_{TL} and t_L) of the TL_{int} and L peaks, respectively. TL_{int} stands for intact complex which has reached the detector, while TL_{dis} stands for complex which dissociated before it has reached the detector.

released from TL, while it was dissociating during its migration to the detector, generates a “bridge” (Figure 1A, red) merging with each of the two peaks. A fraction of unbound L in the equilibrium mixture can be determined from three areas under the curves, designated by letter A:

$$R = \frac{A_L}{A_L + \varphi A_{TL_{int}} + A_{TL_{dis}}} \quad (3)$$

where subscripts correspond to curve designations in Figure 1A, and φ is the quantum yield of fluorescence of bound L relative to that of unbound L.^{12,13} The value of K_d can then be calculated using the following equation:¹⁴

$$K_d = \frac{[T]_0 - [L]_0(1 - R)}{(1/R) - 1} \quad (4)$$

where $[T]_0$ and $[L]_0$ are total concentrations of the target and ligand, respectively. A rearranged eq 4 can be used to find unknown $[T]_0$ if K_d and $[L]_0$ are known:

$$[T]_0 = K_d((1/R) - 1) + [L]_0(1 - R) \quad (5)$$

In some cases, for example when the target is a difficult-to-purify protein, $[T]_0$ is unknown. As a result, K_d cannot be determined with eq 4 and without knowing K_d we cannot determine $[T]_0$ with eq 5. The goal of this work was to develop an approach for simultaneous determination of K_d and unknown $[T]_0$.

Solution for Simultaneous Determination of K_d and $[T]_0$. The mathematical solution for simultaneous determination of K_d and $[T]_0$ is simple. By multiplying eq 4 by its denominator we can transform this equation into a quadratic equation with respect to R :

$$R^2 + \frac{K_d + [T]_0 - [L]_0}{[L]_0}R - \frac{K_d}{[L]_0} = 0 \quad (6)$$

A physically meaningful positive solution of eq 6 has a form:

$$R = -\frac{K_d + [T]_0 - [L]_0}{2[L]_0} + \sqrt{\left(\frac{K_d + [T]_0 - [L]_0}{2[L]_0}\right)^2 + \frac{K_d}{[L]_0}} \quad (7)$$

Equation 7 explicitly relates R with K_d , $[T]_0$, and $[L]_0$. It can be used to find unknown K_d and $[T]_0$ if it is applied to treat data from multiple NECEEM experiments performed with varying known $[L]_0$ but a fixed unknown $[T]_0$ (i.e., by titration with L).

In a NECEEM electropherogram, the three features shown by different colors in Figure 1A are added and form a single curve shown in Figure 1B. To deconvolute the areas, the three features are separated by two vertical boundaries, with approximate or calculated positions, and their areas are determined with lesser or greater accuracy. R is calculated with eq 3 for multiple NECEEM electropherograms and plotted as a function of $[L]_0$. This experimental curve is then fitted with eq 7 in which R and $[L]_0$ are dependent and independent variables, respectively, and K_d and $[T]_0$ are fitting parameters. A required nonlinear regression procedure can be performed with readily available software such as Microsoft Excel (see Section 4 in the Supporting Information for the description and a link to an Excel file used for our calculations).

The described mathematical procedure is simple, but utilizing it in practice requires some additional considerations. There are three questions that will concern an experimenter: (i) in which range should the $[L]_0$ be varied during its titration (i.e., what should be the values of $[L]_{0,first}$ and $[L]_{0,last}$?); (ii) how many points should be measured between $[L]_{0,first}$ and $[L]_{0,last}$?; and (iii) what relative-to-stock-solution concentration

of T (i.e., dilution of $[T]_{\text{stock}}$) should be used during the titration with L? Answers to these questions are important, as an improper choice of $[L]_0$ and dilution of $[T]_{\text{stock}}$ can significantly diminish the accuracy of measurements.

When determining K_d , the experimental error in the measurement of R can become amplified if the experiments are performed at $[L]_0$ significantly larger than K_d . This can be made apparent by examining the derivative of eq 4 with respect to R (see Section 1 in Supporting Information for differentiation), which relates small variations in experimental measurements (ΔR) to variations in the calculated final result (ΔK_d):

$$\Delta K_d \approx \left(-[L]_0 + \frac{[T]_0}{(1-R)^2} \right) \Delta R \quad (8)$$

From eq 8, increasing the value of $[L]_0$ will, at a certain point, result in a ΔK_d value which is larger than K_d itself. If the sign of ΔR is positive, this can lead to 0 or negative values of K_d , which are physically meaningless for reversible interactions. The dependence of $\Delta K_d/K_d$ on increasing values of $[L]_0/K_d$ at constant values of R and ΔR , is plotted in Figure S1.A. As apparent from this graph, the values of ΔK_d are consistently small as long as $[L]_0$ is smaller than or equal to $2 \times K_d$, which represents the point at which concentrations of equilibrium components are perfectly balanced (i.e., $[L]_{\text{eq}} = [T]_{\text{eq}} = [TL]_{\text{eq}} = K_d$). As $[L]_0$ increases over $2 \times K_d$, however, the values of ΔK_d begin to increase exponentially. Thus, when the K_d value is not known *a priori*, it is practical to keep the value of $[L]_{0,\text{first}}$ as the lowest concentration of L which produces reliably measurable areas of all peaks in a NECEEM electropherogram. The exact value is determined by the limit of quantitation (LOQ) of the NECEEM method, which depends on both the sensitivity of the instrument used for the measurements and the accuracy of the data-treatment approach. In this work we defined the LOQ of NECEEM solely through the sensitivity of fluorescence detection to facilitate this consideration. In CE, the LOQ is defined as the concentration of an analyte which results in a signal with a peak height that exceeds the background noise by a certain threshold value. To ensure that all three areas characteristic of NECEEM electropherograms (described in Figure 1) are accurately quantified in our experiments, we have chosen the threshold value to equal 100. It should be noted that if an experiment performed at $[L]_{0,\text{first}} = \text{LOQ}$ results in an apparent K_d value which is smaller than $2 \times [L]_{0,\text{first}}$, then the produced results may not be accurate. This criterion will be discussed later.

Once the value of $[L]_{0,\text{first}}$ is determined, the concentration of $[L]_0$ should be increased until some maximum value $[L]_{0,\text{last}}$ is reached. At a constant $[T]_0$, increasing $[L]_0$ will result in R asymptotically approaching the value of 1. According to eq 8, small variations in R measured close to the asymptote will yield large errors in K_d . The dependence of ΔK_d on increasing values of R, at constant $[T]_0$ and ΔR , is plotted in Figure S1.B. According to this plot, with ΔR set to 1% of K_d , a relative error typical for NECEEM experiments, the value of ΔK_d surpasses the value of K_d at $R \approx 0.9$. This means that points above $R = 0.9$ carry little information on the value of K_d . While the precise value at which ΔK_d overcomes K_d is strongly dependent on the precision of experimental measurements, we suggest, as a general rule, that the value of $[L]_{0,\text{last}}$ should not result in R values above 0.9.

As a fitting-based method, the accuracy of the results produced by our approach relies on the statistical significance of the available data. The least-squares technique, used in the fitting of the experimental data, requires that at least one degree of statistical freedom be available to the nonlinear regression algorithm.¹⁵ When experimental replicate values are averaged and their standard deviation is calculated, 1 degree of freedom is lost. During fitting, the adjustment of two parameters, K_d and $[T]_0$, results in the loss of another 2 degrees of freedom, for a total of 3 degrees of freedom lost. Thus, a minimum of four statistically significant experimental points (i.e., points with nonoverlapping standard deviation ranges), including $[L]_{0,\text{first}}$ and $[L]_{0,\text{last}}$, are required for nonlinear regression analysis. Increasing the number of points will further strengthen the confidence levels associated with the resulting values.

Lastly, accurate determination of K_d requires that the value of R is sensitive to the changes in K_d . By differentiating eq 7 with respect to K_d under an assumption that $[L]_0 \ll [T]_0$, a condition which should be satisfied if $[L]_0 \ll K_d$, we obtain that R is most sensitive to K_d at $[T]_0 \approx K_d$ (see Section 2 in Supporting Information). Although neither K_d nor the concentration of T is known, satisfying this condition is possible by performing one set of preliminary experiments. If $[L]_0 \ll [T]_0$, which should be satisfied by setting $[L]_0$ at LOQ of the instrument, then $[T]_0 \approx K_d$ when $R \approx 0.5$. Thus, a series of NECEEM experiments must be performed at $[L]_{0,\text{first}}$ held constant, and varying unknown concentrations of $[T]_0$, achieved by dilution of a stock solution of T. One of the approaches is serial dilution of the stock solution until two neighboring dilution factors are found for one of which $R < 0.5$ while for the other $R > 0.5$. Intermediate dilution between these two can then be used to reasonably estimate the optimal dilution factor, a , that corresponds to $R \approx 0.5$. This dilution factor can then be used in NECEEM experiments with varying $[L]_0$ described above.

In some cases, users will observe that even the smallest possible dilution of T yields values of R above 50%. These results suggest that $[T]_{\text{stock}}$ is smaller than the value of K_d . In such cases, it might become more difficult, or even impossible, to measure the minimum of four statistically significant experimental points due to the narrowing range of possible R values. If the shape of the R-curve appears to be flat, we suggest that users attempt to increase $[T]_{\text{stock}}$ through one of the available methods of analyte preconcentration.

The approach, thus, requires two sets of NECEEM experiments. In the first set, R is determined as a function of dilution factor of $[T]_{\text{stock}}$ at the lowest concentration of $[L]_{0,\text{first}}$ to find the dilution factor a that results in $R \approx 0.5$. In the second set, R is determined for varying $[L]_0$, but at a constant $[T]_0$ obtained by diluting the stock solution by dilution factor a .

Method for Accurate, Unbiased, And Fast Determination of R. The two sets of NECEEM experiments described above and performed in replicates would produce a large number of NECEEM electropherograms for which accurate and unbiased R values need to be determined in an efficient manner. Accurate, unbiased, and fast determination of R, in turn, requires accurate, unbiased, and fast determination of the three areas in NECEEM electropherograms (Figure 1B). Finding these areas is not straightforward, since the two peaks overlap with the bridge between them. Because of this overlap, visual judgment in finding the boundary between the areas is sometimes difficult to apply and it will never give an accurate solution. We have recently conducted a thorough

theoretical investigation and suggested an algorithm for accurately finding the areas without any human bias.⁶ The algorithm uses an assumption of axially symmetrical peaks of L and TL_{int} to account for their overlaps with the TL_{diss} bridge. Equations for calculating the areas require knowing four time points in a NECEEM electropherogram corresponding to the beginning of the TL_{int} peak ($t_{TL,start}$), the end of the L peak ($t_{L,end}$), and the maxima of these peaks (t_{TL} and t_L , respectively) (Figure 1B). One can find these points manually by using data-processing software (e.g., software provided with CE instrument), but this approach is time-consuming and can be justified only for the analysis of a few electropherograms. For processing a large number of electropherograms, a specialized user-friendly computer program allowing batch data treatment is necessary. We created such a program, referred to as the NECEEM Area Analysis Program (NAAP), in Microsoft Excel with VBA procedures for data input and analysis. NAAP works with text data files that are obtained by an ASCII-export procedure in a capillary electrophoresis instrument or created by an enclosed Excel template from any data in ASCII-format. The detailed explanation of program operation, user instructions, and the download link can be found in Section 3 in the Supporting Information.

In Silico Validation. The validation of the described approach, including the software tools, was conducted *in silico*. NECEEM experiments were simulated with COMSOL Multiphysics software, version 4.4, by using a “Transport of Diluted Species” module in a one-dimensional form. Molecular concentrations of T, L, and TL along the capillary were computed taking into account the interaction between T and L described by reaction 1 and the translational movement of the three species in electrophoresis. For the most adequate modeling, we used realistic values for all experimental parameters including capillary length, position, and size of detection window, shape of injected plug of the equilibrium mixture, diffusion coefficients and electrophoretic mobilities of the species, and an electric field. Simulated electropherograms were plotted as a time dependence of the sum of concentrations of L and TL averaged over the detection volume with a constant level of noise. The noise was generated as a random signal with the standard deviation of the amplitude corresponding to 1% of the lowest simulated signal observed at $[L]_0 = 10$ nM and $[T]_0 = 480$ nM. Additional technical details for our COMSOL modeling can be found in Section 5 in the Supporting Information.

The first set of NECEEM experiments was simulated to find the dilution factor, $a = [T]_{stock}/[T]_0$, by which the stock solution of T of concentration $[T]_{stock}$ needs to be diluted to satisfy $[T]_0 \approx K_d$. We chose $[L]_0 = 10$ nM as a typical concentration of a fluorescently labeled molecule that can generate a signal suitable for NECEEM measurements. We chose $K_d = 100$ nM as a typical value for affinity binding of biological molecules. The “unknown” concentration of the target stock solution, $[T]_{stock}$, was chosen to be $10 \mu\text{M}$. NECEEM electropherograms were simulated for the stock concentration of T and a number of two-time dilutions (Figure 2A). The three areas required for calculation of R in eq 3 were found in an unbiased way by using NAAP. The relative quantum yield, ϕ , of the bound ligand was set to equal 1. R values were plotted as a function of the dilution factor a (Figure 2B). The value of a that resulted in $R \approx 0.5$ was estimated from the graph ($a = 100$ in this example). At this point, we could calculate $[T]_0$ that corresponded to $R \approx 0.5$ and compare it

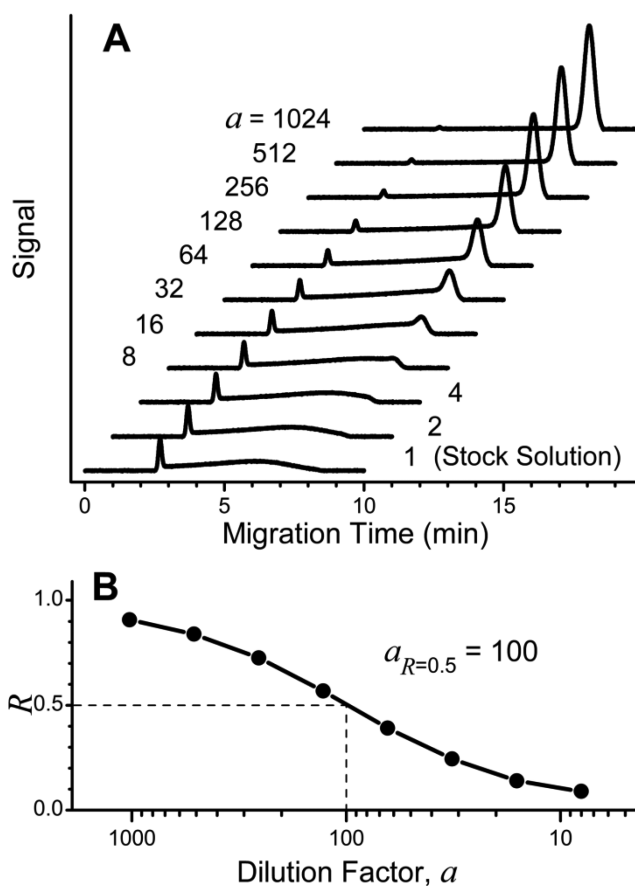


Figure 2. Procedure for the determination of optimal dilution factor a . (A) Set of simulated NECEEM electropherograms reveals the effect of serial dilutions of the stock solution of T, whereas concentration $[L]_0$ is fixed to 10 nM. A dilution factor a is changed from 1 to 1024. (B) Dependence of R (fraction of unbound L) on a . Values of R are calculated from the simulated electropherograms using NAAP. The optimal factor $a = 100$ corresponds to a dilution at which $R = 0.5$.

with K_d specified as 100 nM: $[T]_0 = [T]_{stock}/a = 10 \mu\text{M}/100 = 100$ nM. The equality between $[T]_0$ and K_d confirms that this part of the procedure works correctly.

The second set of NECEEM experiments was simulated to find “unknown” values of K_d and $[T]_0$. In this set, we used $[T]_0 = [T]_{stock}/a = 100$ nM and varied $[L]_0$ between 10 and 400 nM. The resulting simulated NECEEM electropherograms are shown in Figure S10 and in the inset to Figure 3. R values were calculated from these electropherograms and plotted as a function of $[L]_0$ (open circles in Figure 3). The solid curve in Figure 3 is a fitting of the R values with the theoretical dependence of R versus $[L]_0$ presented by eq 7 using Microsoft Excel (see Section 4 in Supporting Information). The best fit was obtained for $K_d = 98.9$ nM and $[T]_0 = 101.1$ nM. These values are equal (with 99% accuracy) to the values used in COMSOL simulations of NECEEM, which confirms that the entire procedure (the mathematics, the algorithms, and the software) works correctly.

To assess the robustness of the approach used in the second set of NECEEM experiments, we simulated this set of NECEEM electropherograms at a wide range of $[L]_{0,first}$ and $[T]_0$. In essence, we calculated R values at increasing $[L]_0$ for a series of $[T]_0$, both greater and lesser than K_d . For every graph, we found K_d and $[T]_0$ and compared them with the correct ones (preset in COMSOL calculations to simulate the

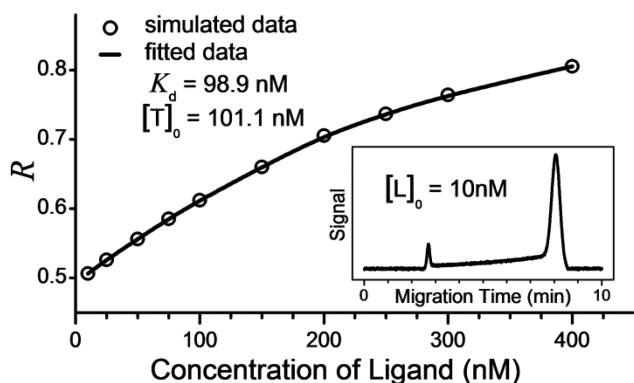


Figure 3. Dependence of R (fraction of unbound L) on concentration $[L]_0$ at constant concentration $[T]_0$ obtained from the stock solution by diluting with factor $a = 100$. Open circles represent calculated R -values; solid curve is a fitting with eq 7. The insert demonstrates the typical simulated NECEEM electropherogram used for the R calculation.

electropherograms). The error was calculated as the absolute difference between the determined and correct values relative to the correct value (Figure S11). Slices of these contour plots at $[L]_0/K_d = 0.2$ appear in Figure 4 and illustrate the general

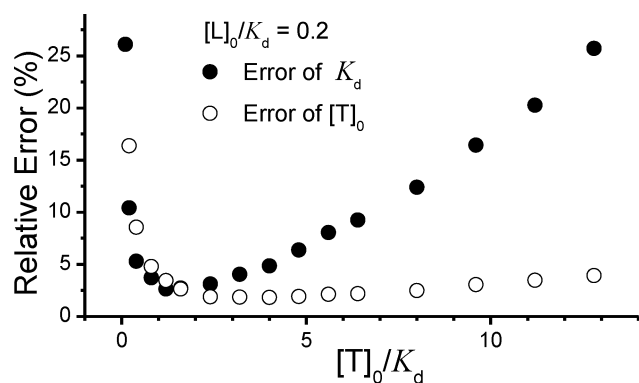


Figure 4. Accuracy of determination of K_d and $[T]_0$ estimated *in silico* at dimensionless ligand concentration $[L]_0/K_d = 0.2$. Closed and open circles represent relative errors of K_d and $[T]_0$, respectively.

trends in dependence of the errors on $[T]_0/K_d$. In general, a three-time overstepping of $[T]_0$ above K_d did not cause the error to surpass more than 5%, which suggests that an overestimate of $[T]_0$ from the first set of NECEEM experiments (Figure 2B) can be suitable for the second set of NECEEM experiments (Figure 3). As expected, using $[L]_{0,first}$ values which were significantly above $2 \times K_d$ led to unacceptably high errors. Interestingly, these results also suggest that the accuracy of determination of $[T]_0$ is not affected by the improper choice of $[L]_{0,first}$ and $[T]_0$, as severely as the accuracy of K_d determination.

Experimental Determination of K_d and $[T]_0$. As an experimental model we used a pair of binding molecules: MutS protein from *Thermus aquaticus* and a previously selected DNA aptamer for this protein.⁷ The concentration of the DNA aptamer stock solution was determined by measuring the optical density at 260 nm and dividing the absorbance by the manufacturer-provided extinction coefficient. MutS protein from *Thermus aquaticus* is not commercially available, and therefore, we expressed a corresponding gene in *E. coli* and

purified it in-house. In the NECEEM experiments, we assumed that the concentration of the stock solution of MutS was unknown. MutS was, thus, the target, and the DNA aptamer was the ligand in our experiments.

In the first set of NECEEM experiments, we determined the dilution of the stock solution of MutS which resulted in $R \approx 0.5$. Using a custom-made CE instrument, the concentration of L which resulted in a peak height 100 times larger than that of the noise of the background was 100 pM and we used it as $[L]_{0,first}$. Various dilutions of the protein were used to obtain a set of NECEEM electropherograms, and the three areas were determined using NAAP in an unbiased manner. The value of $\varphi \approx 1$ was determined experimentally by comparing the signal of the free aptamer with the signal of the fully saturated complex (Figure S12). The values of R were plotted as a function of dilution factors used to prepare the protein solutions from the stock solution (Figure 5A). The $a = 1000$ value was chosen as being the closest to $R \approx 0.5$ and was used in the second set of experiments.

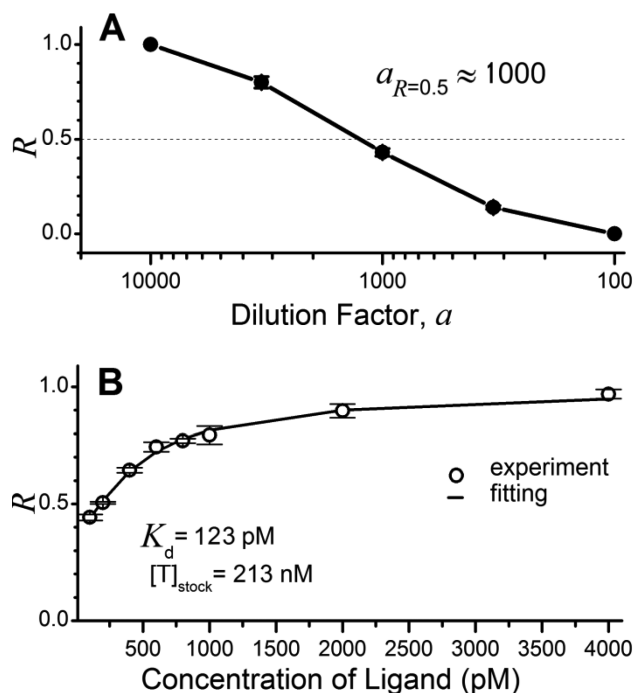


Figure 5. Determination of K_d and $[T]_0$ from experimental NECEEM data: (A) Estimation of optimal dilution factor a from the dependence of R (fraction of unbound L) on a . (B) Dependences of R on concentration $[L]_0$ at constant concentration $[T]_0$ obtained from the stock solution by diluting with factor $a = 1000$. Open circles represent calculated R -values; solid curve is a fitting with eq 7.

In the second set of NECEEM experiments, we used $[T]_0 = [T]_{stock}/1000$ and $[L]_0$ varying from $[L]_{0,first} = 100$ pM to $[L]_{0,last} = 4$ nM, with six points in between. The obtained electropherograms are shown in Figure S13. The values of R were calculated and plotted as a function of $[L]_0$ (Figure 5B). The experimental points in Figure 5B were fitted with the theoretical dependence of R versus $[L]_0$ presented by eq 7. The best fit was obtained at $K_d = 123$ pM and $[T]_0 = 213$ pM, meaning that $[T]_{stock} = 213$ pM \times 1000 = 213 nM. Experimental variability of R translates into errors in the parameters obtained by fitting and can be estimated by constructing a covariance matrix based on each of the

parameters (see Section 4 in the Supporting Information).¹⁶ We determined the relative errors to be 9% and 4% for K_d and $[T]_0$, respectively. As the resultant K_d is larger than the value of $[L]_{0,first}$, we can be certain that our experiments were performed at a proper range of concentrations.

It should be noted that before performing the experiments on a custom-made CE instrument, which has a remarkable LOQ of 100 pM, we have also performed the same experiment using a commercial instrument, with a poorer LOQ of 5 nM. Here, in the first set of experiments the value of a was determined to be 30 (Figure S14A). In the second set of NECEEM experiments, we used $[T]_0 = [T]_{stock}/30$ and $[L]_0$ varying from $[L]_{0,first} = 5$ nM to $[L]_{0,last} = 40$ nM, with six points in between. The obtained electropherograms are shown in Figure S15. The experimental points from the second step resulted in the best fit with the theoretical dependence of R versus $[L]_0$ at $K_d = 650$ pM and $[T]_0 = 6.7$ nM, meaning that $[T]_{stock} = 202$ nM (Figure S14B). As the resultant K_d is approximately 13 times smaller than the value of $[L]_{0,first}$, we predicted that significant nonstatistical errors were embedded into these results. This prompted us to repeat the study with a more sensitive instrument (discussed above), which indeed provided more refined values.

To assess the accuracy of our method in the determination of $[T]_0$, we estimated the concentration of the MutS stock solution by the BCA assay. The spectroscopically measured value of $[T]_{stock} = 226 \pm 13$ nM is in good agreement with our NECEEM-measured values of 213 and 202 nM, with the custom-made and commercial instruments, respectively. As for K_d measurements, our previous study of the same interaction, conducted with a single-point NECEEM approach by using a commercial CE instrument with a poorer LOQ, resulted in $K_d = 15$ nM,⁷ which is significantly larger than $K_d = 123$ pM obtained with the method presented in this work. The previous K_d measurement was performed as a part of a screening of a large number of aptamer clones, and thus, the $[L]_0$ and $[T]_0$ concentrations were not optimized as rigorously as in this study. The discrepancy between the values, most likely, comes from the fact that in the previous study nanomolar concentrations of $[L]_0$ and $[T]_0$ were used to screen all of the aptamer clones, resulting in large errors for some of the measurements. Nevertheless, the single-point NECEEM method, used previously, did not provide us with any means of assessing the accuracy of the results. In the current approach, however, an insufficient LOQ of the instrument used was evident from the produced results. This suggests that the current multipoint approach is much more suitable for affinity measurements which require a high degree of accuracy and precision.

Concluding Remarks. A procedure for the determination of K_d from two NECEEM electropherograms obtained with two different known $[L]_0$ but the same unknown $[T]_0$ was recently described.¹⁷ In general, the “two-point” method uses our eq 5 to make a system of two equations with two unknowns. Values R_1 and R_2 are obtained for ligand concentrations $[L]_{0,1}$ and $[L]_{0,2}$. The first equation in the system is obtained from eq 5 by (i) writing it twice (for two sets of $[L]_0$ and R), (ii) equating its right-hand side parts, and (iii) expressing the equality in the following form:

$$K_d = \frac{[L]_{0,2}(1 - R_2) - [L]_{0,1}(1 - R_1)}{1/R_1 - 1/R_2} \quad (9)$$

The authors obtained only this expression.¹⁷ However, the solution for $[T]_0$ can also be found by simply placing eq 9 into eq 5:

$$[T]_0 = \frac{[L]_{0,2}(1 - R_2) - [L]_{0,1}(1 - R_1)}{1/R_1 - 1/R_2} (1/R_1 - 1) + [L]_{0,1}(1 - R_1) \quad (10)$$

The two-point method utilizing eqs 9 and 10 can also reveal both K_d and $[T]_0$. However, it uses a smaller data set than the multipoint approach with the fitting described in this work. It also does not incorporate a way of finding the optimum dilution factor for the stock solution of T . As a result the two-point method is expected to produce less accurate values of K_d and $[T]_0$. It may, however, be very useful for fast and rough estimates of K_d and $[T]_0$.

To summarize, we described a NECEEM-based approach for simultaneous determination of the K_d and target concentration, $[T]_0$. The *in silico* validation showed that the approach (mathematics, algorithms, and software) has above 90% accuracy over a wide range of initial conditions. The accuracy of the experimental implementation of this approach will, thus, be mainly affected by the quality of the NECEEM electropherograms, and accuracy of ligand concentration, $[L]_0$, used in calculations. As in the case of any approach including nonlinear regression, our method requires a reasonable judgment of the experimenter about the quality of the experimental data to ensure the reliability of the obtained values of K_d and $[T]_0$.

■ ASSOCIATED CONTENT

● Supporting Information

Supporting mathematics, software, and figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: skrylov@yorku.ca.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the Natural Sciences and Engineering Research Council of Canada (Discovery grant 238990). M.K. was supported by an Alexander Graham Bell Canada Graduate Scholarship.

■ REFERENCES

- (1) Kastritis, P. L.; Bonvin, A. J. R. *Soc. Interface* **2013**, *10*, doi: 10.1098/rsif.2012.0835.
- (2) Krylov, S. N. *Electrophoresis* **2007**, *28*, 69–88.
- (3) Haes, A. J.; Giordano, B. C.; Collins, G. E. *Anal. Chem.* **2006**, *78*, 3758–3764.
- (4) Krylov, S. N. *J. Biomol. Screening* **2006**, *11*, 115–122.
- (5) Drabovich, A. P.; Okhonin, V.; Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2007**, *129*, 7260–7261.
- (6) Cherney, L. T.; Kanoatov, M.; Krylov, S. N. *Anal. Chem.* **2011**, *83*, 8617–8622.
- (7) Drabovich, A. P.; Berezovski, M.; Okhonin, V.; Krylov, S. N. *Anal. Chem.* **2006**, *78*, 3171–3178.
- (8) Kanoatov, M.; Javaherian, S.; Krylov, S. N. *Anal. Chim. Acta* **2010**, *681*, 92–97.
- (9) Drabovich, A.; Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 11224–11225.

- (10) Krylov, S. N.; Starke, D. A.; Arriaga, E. A.; Zhang, Z. R.; Chan, N. W. C.; Palcic, M. M.; Dovichi, N. J. *Anal. Chem.* **2000**, *72*, 872–877.
- (11) Wu, S. L.; Dovichi, N. J. *J. Chromatogr.* **1989**, *480*, 141–155.
- (12) Krylova, S. M.; Wegman, D. W.; Krylov, S. N. *Anal. Chem.* **2010**, *82*, 4428–4433.
- (13) Okhonin, V.; Krylova, S. M.; Krylov, S. N. *Anal. Chem.* **2004**, *76*, 1507–1512.
- (14) Krylova, S. M.; Dove, P. M.; Kanoatov, M.; Krylov, S. N. *Anal. Chem.* **2011**, *83*, 7582–7585.
- (15) Cochran, W. G. *Biometrics* **1954**, *10*, 417–451.
- (16) van de Geer, S. A. In *Encyclopedia of Statistics in Behavioral Science*; Everitt, B. S., Howell, D. C., Eds.; John Wiley & Sons, Ltd.: Chichester, 2005; Vol. 2, pp 1041–1045.
- (17) Zhang, Y. W.; Yan, H. Y.; Fu, P.; Jiang, F.; Zhang, Y.; Wu, W. X.; Li, J. X. *Anal. Bioanal. Chem.* **2013**, *405*, 5549–5555.

SUPPORTING INFORMATION

Using Non-Equilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM) for Simultaneous Determination of Concentration and Equilibrium Constant

Mirzo Kanoatov,¹ Victor A. Galievsky,¹ Svetlana M. Krylova,¹ Leonid T. Cherney,¹ Hanna K. Jankowski,² and Sergey N. Krylov^{1*}

¹Department of Chemistry and Centre for Research on Biomolecular Interactions and

²Department of Mathematics and Statistics
York University, Toronto, Ontario M3J 1P3, Canada

TABLE OF CONTENTS

1. Relationship between $[L]_0$ and error in measuring of K_d	S2
2. Proof for Optimum Relation between K_d and $[T]_0$	S3
3. Program for Finding Areas in NECEEM Electropherograms	S5
3.1. Description of the NECEEM Area Analysis Program	S5
3.2. The general algorithm of the NAAP	S5
3.2.1. Setting of initial parameters	S5
3.2.2. Data import	S6
3.2.3. Determination of corrected signal $S_{\text{corr}}(t)$	S6
3.2.4. Determination of $t_{\text{TL,start}}$ and $t_{\text{L,end}}$ positions	S8
3.2.5. Determinations of t_{TL} and t_{L}	S8
3.2.6. Calculation of areas	S9
3.2.7. Calculation of kinetic constants	S9
4. Program for Fitting Eq. (7) into Experimental Dependence of R on $[L]_0$	S10
4.1. Description of the Fitting Program	S10
4.2. Asymptotic variance of parameters for data with uniform random error	S11
4.3. Asymptotic variance of parameters for data with non-uniform random error	S14
5. NECEEM-electropherogram simulation with COMSOL Multiphysics	S15
6. Supporting Figures	S17
7. Supporting References	S23

1. Relationship between $[L]_0$ and error in measuring of K_d

$$\Delta K_d \approx \frac{dK_d}{dR} \Delta R \quad (S1)$$

Calculations of derivative dK_d/dR using Eq. (S5) give:

$$\frac{dK_d}{dR} = \frac{[L]_0}{\frac{1}{R}-1} + \frac{[T]_0 - [L]_0(1-R)}{\left(\frac{1}{R}-1\right)^2} \frac{1}{R^2} \quad (S2)$$

This expression can be simplified and transferred as follows:

$$\begin{aligned} \frac{dK_d}{dR} &= \frac{[L]_0 R}{1-R} + \frac{[T]_0 - [L]_0(1-R)}{(1-R)^2} = \frac{[L]_0 R}{1-R} + \frac{[T]_0}{(1-R)^2} - \frac{[L]_0}{1-R} = \\ &= \frac{[L]_0 R - [L]_0}{(1-R)^2} + \frac{[T]_0}{(1-R)^2} = -[L]_0 + \frac{[T]_0}{(1-R)^2} \end{aligned} \quad (S3)$$

From Eqs. (S1) and (S3) we have the following relation for small deviations (or errors) denoted by Δ :

$$\Delta K_d \approx \left(-[L]_0 + \frac{[T]_0}{(1-R)^2} \right) \Delta R \quad (S4)$$

Using Eq. (S4), dependence of relative error, $\Delta K_d/K_d$, on dimensionless ligand concentration $[L]_0/K_d$ at constant R and ΔR , can be calculated and plotted (**Fig. S1.A**). Similarly, dependence of $\Delta K_d/K_d$ on value of R can be calculated and plotted at constant $[T]_0$ and ΔR (**Fig. S1.B**).

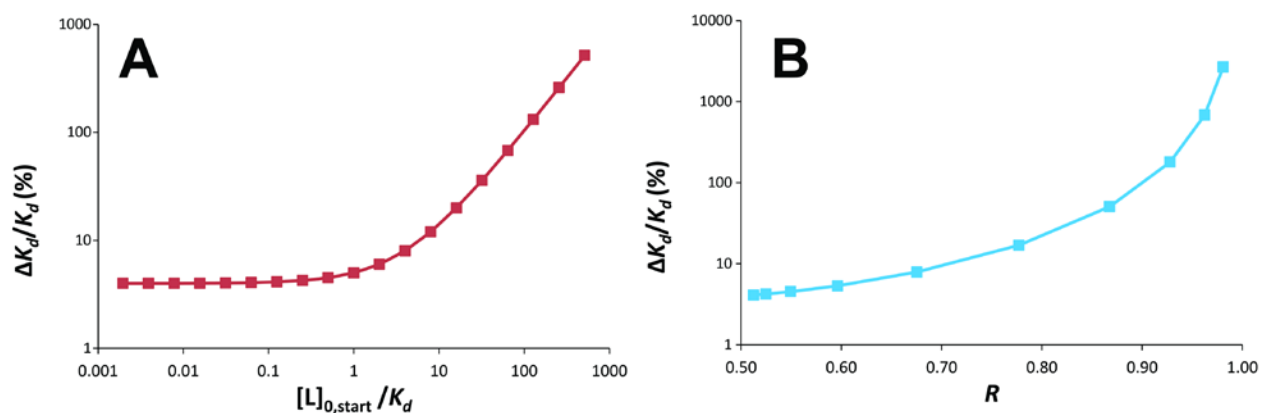


Figure S1: Relative error in determination of K_d , depending on the ratio of $[L]_0$ to K_d (panel A) and the value of R (panel B). **Panel A:** ΔK_d was calculated using Eq. (S4) by varying $[L]_0$ and $[T]_0$ to yield $R = 0.5$. K_d was set to 1, and ΔR equaled 1% of R . **Panel B:** ΔK_d was calculated using Eq. (S4) by varying $[L]_0$ to yield R values in the range of 0.5 and 0.98. K_d was set to 1,

$[T]_0 = K_d$, and ΔR equaled 1% of R . Please note that both y-axes are presented in logarithmic scale.

2. Proof for Optimum Relation between K_d and $[T]_0$

Expression for R has a form (see Eq. (7) in the manuscript):

$$R = -\frac{K_d + [T]_0 - [L]_0}{2[L]_0} + \sqrt{\left(\frac{K_d + [T]_0 - [L]_0}{2[L]_0}\right)^2 + \frac{K_d}{[L]_0}} \quad (S5)$$

At a very small value $[L]_0$ (with respect to $[T]_0$ and K_d) we can approximately transform Eq. (S5) by expanding the square root as follows:

$$\begin{aligned} R_0 = R_{[L]_0 \rightarrow 0} &= -\frac{K_d + [T]_0 - [L]_0}{2[L]_0} + \frac{K_d + [T]_0 - [L]_0}{2[L]_0} \sqrt{1 + \frac{\frac{K_d}{[L]_0}}{\left(\frac{K_d + [T]_0 - [L]_0}{2[L]_0}\right)^2}} \approx \\ &\approx -\frac{K_d + [T]_0 - [L]_0}{2[L]_0} + \frac{K_d + [T]_0 - [L]_0}{2[L]_0} \left(1 + \frac{1}{2} \frac{\frac{K_d}{[L]_0}}{\left(\frac{K_d + [T]_0 - [L]_0}{2[L]_0}\right)^2}\right) = \\ &= \frac{\frac{K_d}{[L]_0}}{\frac{K_d + [T]_0 - [L]_0}{[L]_0}} = \frac{K_d}{K_d + [T]_0 - [L]_0} \approx \frac{K_d}{K_d + [T]_0} \end{aligned} \quad (S6)$$

Let us calculate a derivative of R_0 with respect to K_d that shows how fast R changes with a change in K_d at small $[L]_0$:

$$\frac{\partial R_0}{\partial K_d} = \frac{1}{K_d + [T]_0} - \frac{K_d}{(K_d + [T]_0)^2} = \frac{[T]_0}{(K_d + [T]_0)^2} \quad (S7)$$

If we consider a dependence of $\partial R_0 / \partial K_d$ on $[T]_0$ then we can conclude that $\partial R_0 / \partial K_d$ has a maximum at some value of $[T]_0 = [T]^*$. Indeed, according to Eq. (S7) we have $\partial R_0 / \partial K_d = 0$ at $[T]_0 = 0$, $\partial R_0 / \partial K_d = 0$ at $[T]_0 = \infty$, and $\partial R_0 / \partial K_d > 0$ in the interval $0 < [T]_0 < \infty$. In other words, $\partial R_0 / \partial K_d$ is positive inside an interval $(0, \infty)$ and approaches 0 at its ends. Therefore, Eq. (S7) reaches a maximum in some point $[T]^*$ inside this interval. To find $[T]^*$ we should calculate the second derivative $\partial(\partial R_0 / \partial K_d) / \partial [T]_0$:

$$\frac{\partial}{\partial [T]_0} \left(\frac{\partial R_0}{\partial K_d} \right) = \frac{1}{(K_d + [T]_0)^2} - \frac{2[T]_0}{(K_d + [T]_0)^3} \quad (S8)$$

This derivative should be equal to zero in the point $[T]_0 = [T]^*$ where $\partial R_0 / \partial K_d$ reaches its maximum. As a result we obtain equation:

$$\frac{1}{\left(K_d + [T]^*\right)^2} - \frac{2[T]^*}{\left(K_d + [T]^*\right)^3} = 0 \quad (S9)$$

which can be transformed to a simpler one:

$$1 - \frac{2[T]^*}{K_d + [T]^*} = 0 \quad (S10)$$

Solving Eq. (S10) and taking into account Eq. (S6) we finally obtain:

$$[T]^* = K_d, \quad (R_0)_{[T]_0=[T]^*} = \frac{1}{2} \quad (S11)$$

The second relation in Eq. (11) can be used to approximately satisfy relation $[T]_0 = [T]^*$ in experiments even if K_d and, therefore, $[T]^*$ are unknown. To do that, one should dilute the target so that a value of R becomes approximately 0.5 at the lowest feasible concentration of L . In this case, R is the most affected by changes in K_d , which, in turn, is the most favorable scenario for fitting theoretical expression (S5) into experimental dependence of R on $[L]_0$. After that, NECEEM experiments should be run at higher $[L]_0$ which would result in R in the range $0.5 < R < 1$. Finally, fitting dependence (S5) into such experimental dependence of R on $[L]_0$ should produce accurate values of K_d and $[T]_0$.

3. Program for Finding Areas in NECEEM Electropherograms

3.1. Description of the NECEEM Area Analysis Program

NECEEM Area Analysis Program (NAAP) has been written using Visual Basic, and requires Microsoft Excel version 2007 or later. The program analyzes NECEEM electropherograms, calculates areas of the TL_{int} , TL_{diss} , and L components, and extracts K_d , k_{on} and k_{off} values in a batch regime. It employs the peak deconvolution method developed by Cherney *et al.*¹ The program is designed to minimize the human participation and accelerate the data processing. The most recent version of NAAP can be downloaded from: <http://www.yorku.ca/skrylov/Research/>

The current version is developed for analysis of data obtained by a P/ACE MDQ series of Capillary Electrophoresis instruments (Beckman-Coulter), and exported to ASCII-type by the 32 Karat software, version 7. A template of a compatible ASCII-file type is presented in the “Example_Data” worksheet of the NAAP, and can be used for analysis of data obtained by other CE instruments. Detailed user guide is in the “Instructions” worksheet.

3.2. The general algorithm of the NAAP

3.2.1. Setting of initial parameters

Before the code execution the user has to set initial parameters divided into 4 groups (**Fig. S2**):

- 1) path to data folder;
- 2) parameters which describe the studied system:
 - i) initial concentrations of Ligand and Target,
 - ii) ratio of fluorescence quantum yields of free ligand and complex;
- 3) temporal parameters that define signal area for analysis (**Fig. S3**):
 - i) complete treated range including the TL_{int} , TL_{diss} , L bands and flat zero-signal zones before and after the bands for background calculation,
 - ii) boundaries of these background zones,
 - iii) sample plug propagation time, before the measurements start, if the plug propagation through uncooled zone procedure was used;²
- 4) parameters exploited in the NAAP treated algorithm:
 - i) signal-to-noise ratio coefficient used as criterion for band boundary determination,
 - ii) fractions of peak widths to define data ranges of the TL_{int} - and L-bands that are used for calculation of the maxima of these peaks (t_{TL} and t_L , respectively).

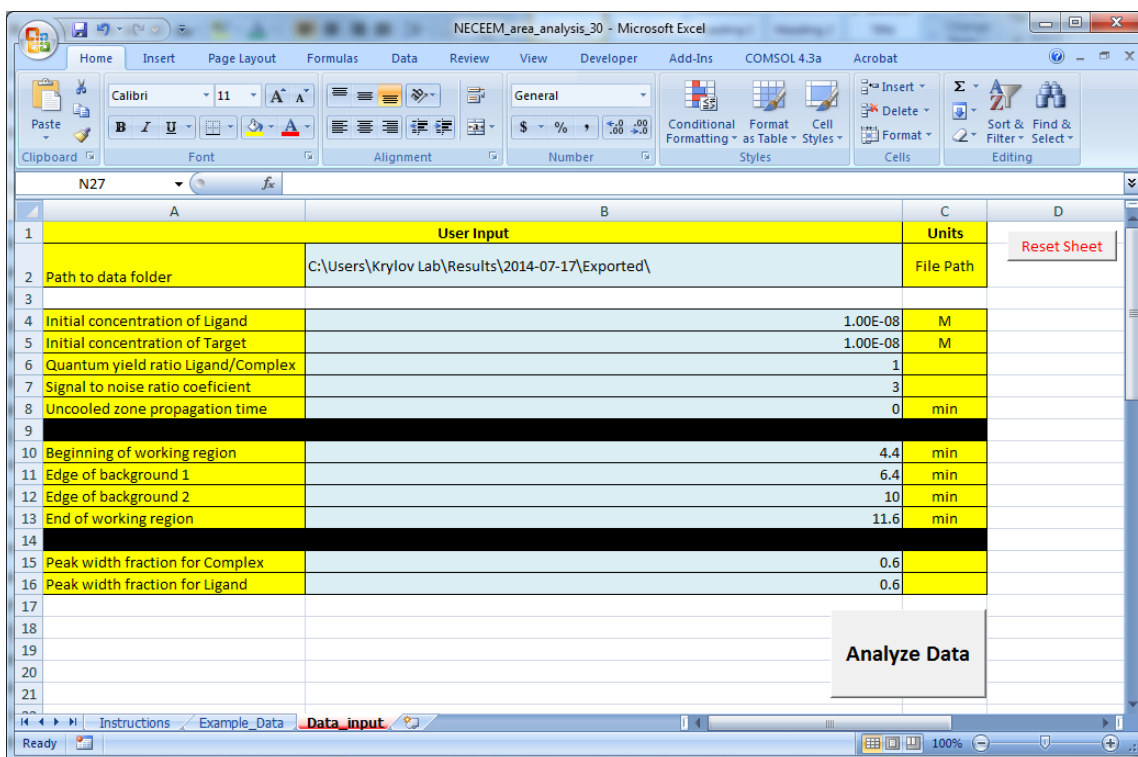


Figure S2. “Data Input” worksheet of the NAAP.

3.2.2. Data import

A batch subroutine reads all ASCII-files within the user-defined folder and imports them into the workbook. For each data file a separate worksheet is created which contains data, parameters, calculation results, and graphs with the electropherogram at different algorithm stages. In this way the user can observe and control both intermediate and final results of the calculations and finetune of the treatment parameters for any data file.

3.2.3. Determination of corrected signal $S_{\text{corr}}(t)$

The program calculates baseline and standard deviation (SD) of a background noise using the predefined flat regions before and after NECEEM peaks (**Fig. S3**). We suppose the background doesn't change substantially during a measurement and so it can be linearized within the user-defined working region of the signal. Baseline is calculated using the average values of background in predefined regions before and after NECEEM signal. The corrected signal $S_{\text{corr}}(t)$ is obtained by subtraction of the calculated baseline from rare data and then is used for all of the following computations (**Fig. S4**).

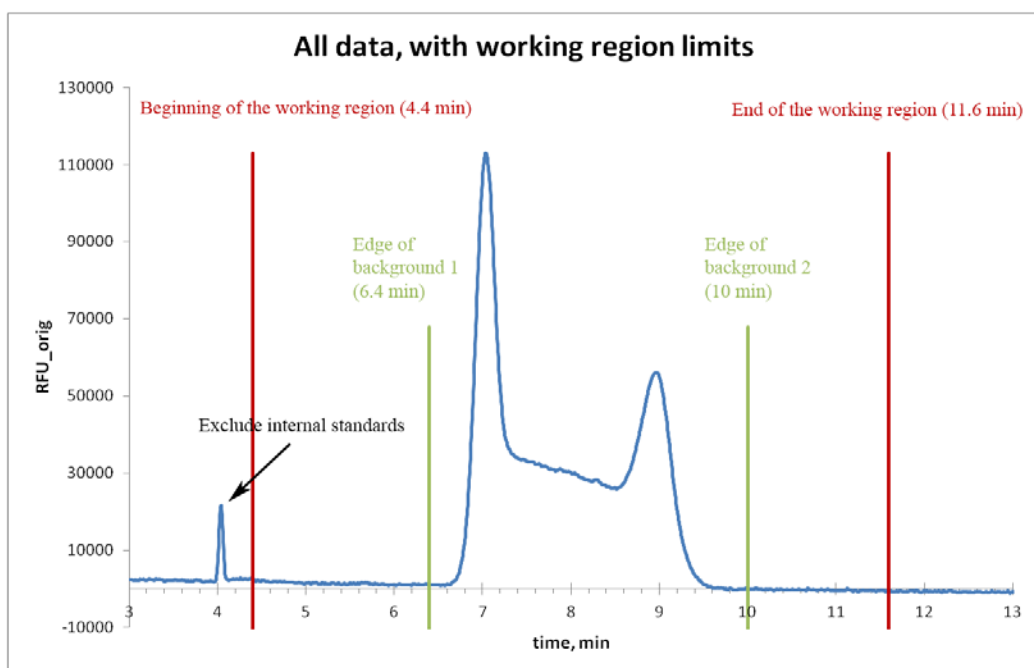


Figure S3. Typical NECEEM electropherogram with defined timing limits.

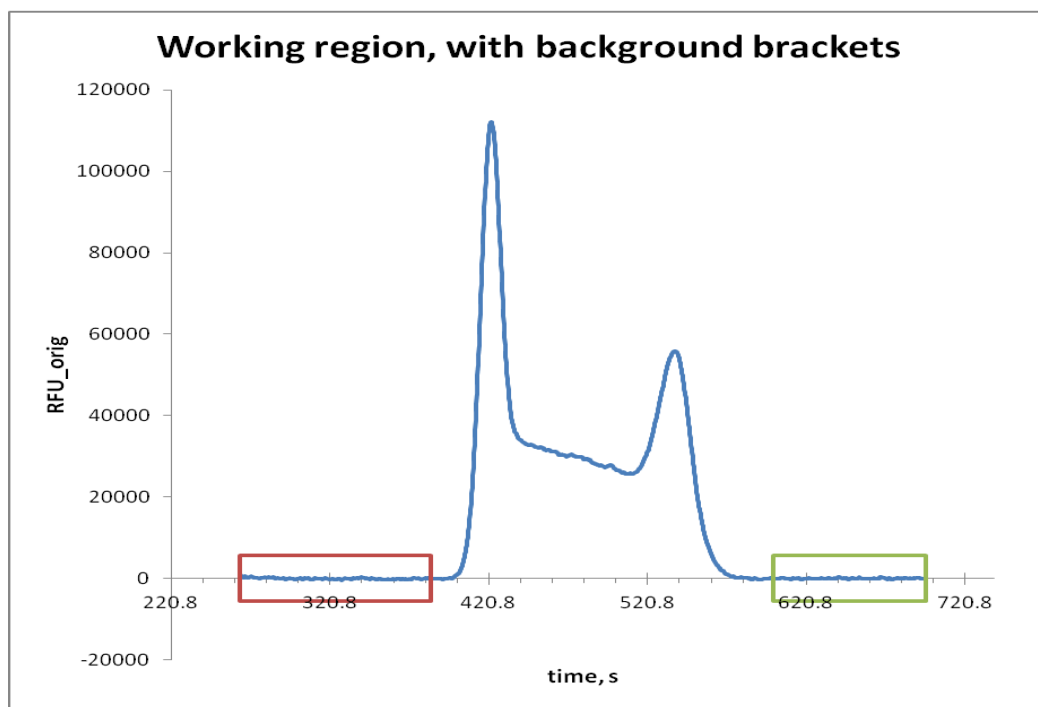


Figure S4. Corrected NECEEM signal $S_{\text{corr}}(t)$ after baseline subtraction. Standard deviation (SD) of a background noise is calculated using signal in selected areas.

3.2.4. Determination of $t_{\text{TL,start}}$ and $t_{\text{L,end}}$ positions

User-defined background regions and a signal-to-noise coefficient n are used to identify non-convoluted edges of the complex and free-ligand peaks. We determine the beginning of the TL_{int} peak ($t_{\text{TL,start}}$) and the end of the L peak ($t_{\text{L,end}}$) at time points where NECEEM signal exceeds the background at n values of the SD. For example at $n = 3$ a confidence level is more than 99% for a normal distribution of the noise. Noise overshoots can produce erroneous responses of the boundary search algorithm, so before executing the boundary identification procedure we apply a smoothing filter to the signal $S_{\text{corr}}(t)$ to decrease the noise level. A Savitzky-Golay algorithm allows smoothing nonlinear data with minimal losses of fine structural details such as peak height and width.³ We use a 29-point Savitzky-Golay smoothing filter which produces a weighted averaging of the 29-neighbor points with the central data point given the most significance. We found experimentally that this is an optimal filter for significant reduction of noise without electropherogram distortion.

3.2.5. Determinations of t_{TL} and t_{L}

Positions of the band maxima are determined in two-step procedure.

A first derivative of the data is used to approximate maxima of the TL_{int} and L bands (**Fig. S5**). The derivative is obtained by applying a second pass of the Savitzky-Golay digital filter over the smoothed data, but with another set of coefficients, which simultaneously realizes both the smoothing and numerical differentiation.⁴ Double passes of the average filter improve algorithm robustness because it eliminates false maxima coming from noise waves near the band base.

For noisy and wide peaks the program may erroneously determine a local maximum which differs from the actual peak maximum. To address this, we introduced a second step to accurately identify the global maximum for each peak. The peak maxima positions t_{TL} and t_{L} are further refined by fitting to a second-order polynomial equation (**Fig. S6**). A width of the extrapolated area is a program parameter and can be tuned by the user as appropriate for each peak.

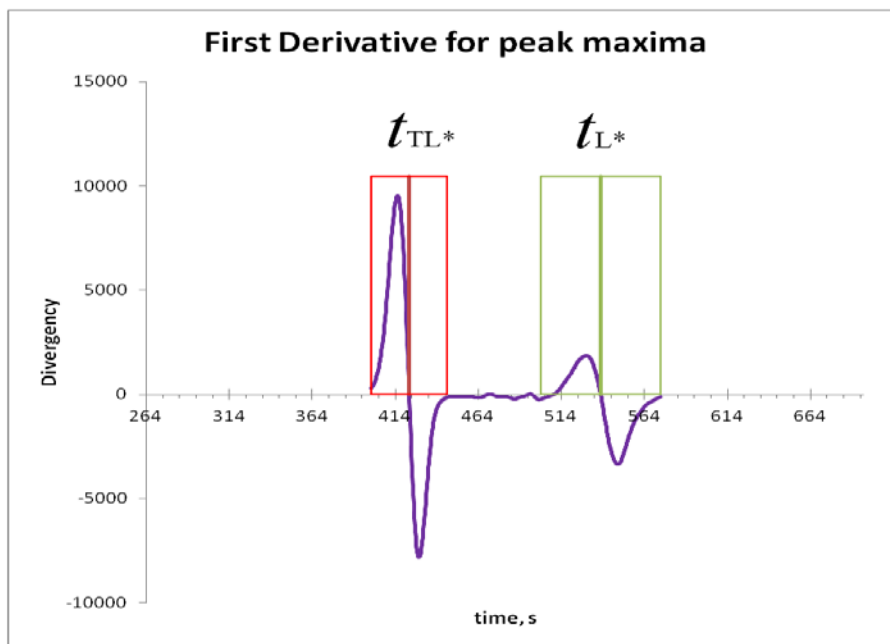


Figure S5. Approximate maxima $t_{\text{TL}*}$ and $t_{\text{L}*}$ of the TL_{int} and L bands, respectively, were defined at time positions where the first derivative of the signal changes his sign.

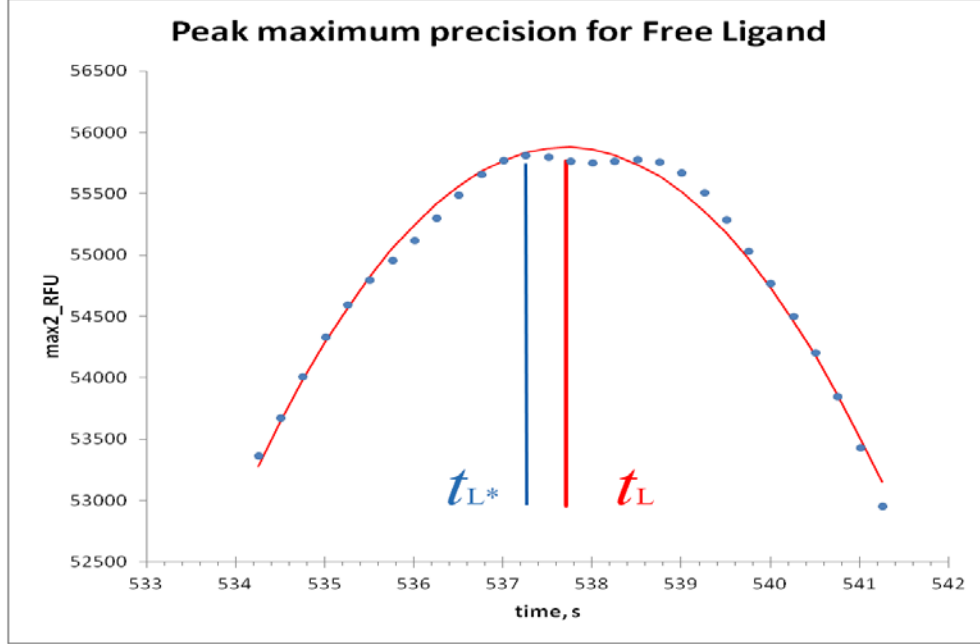


Figure S6. A global maximum t_L is calculated by signal fitting to a second-order polynomial equation near the t_{L*} -peak position.

3.2.6. Calculation of areas

Assuming peak symmetry, convoluted edges of the two peaks are resolved (**Fig. S7**) and areas under the three regions of the NECEEM electropherograms A_L , $A_{TL \text{ int}}$ and $A_{TL \text{ diss}}$ are calculated numerically according to equations S12-S14 (**Fig. S8**):

$$A_L = \sum_{t=2t_L-t_{L,\text{end}}}^{t_{L,\text{end}}} S_{\text{corr}}(t) \times t_{\text{step}} - \frac{2}{3} S_{\text{corr}}(t) \Big|_{t=2t_L-t_{L,\text{end}}} \times (t_{L,\text{end}} - t_L) \quad (\text{S12})$$

$$A_{TL \text{ int}} = \sum_{t=t_{TL,\text{start}}}^{2t_{TL}-t_{TL,\text{start}}} S_{\text{corr}}(t) \times t_{\text{step}} - \frac{2}{3} S_{\text{corr}}(t) \Big|_{t=2t_{TL}-t_{TL,\text{start}}} \times (t_{TL} - t_{TL,\text{start}}) \quad (\text{S13})$$

$$A_{TL \text{ diss}} = \sum_{t=2t_{TL}-t_{TL,\text{start}}}^{2t_L-t_{L,\text{end}}} S_{\text{corr}}(t) \times t_{\text{step}} + \frac{2}{3} S_{\text{corr}}(t) \Big|_{t=2t_L-t_{L,\text{end}}} \times (t_{L,\text{end}} - t_L) + \frac{2}{3} S_{\text{corr}}(t) \Big|_{t=2t_{TL}-t_{TL,\text{start}}} \times (t_{TL} - t_{TL,\text{start}}) \quad (\text{S14})$$

where t_{step} is an acquisition step during the electrophoretic measurement.

3.2.7. Calculation of kinetic constants

The signal areas together with $[L]_0$ and $[T]_0$ (if the latter is known) are used to calculate K_d , k_{on} and k_{off} .¹

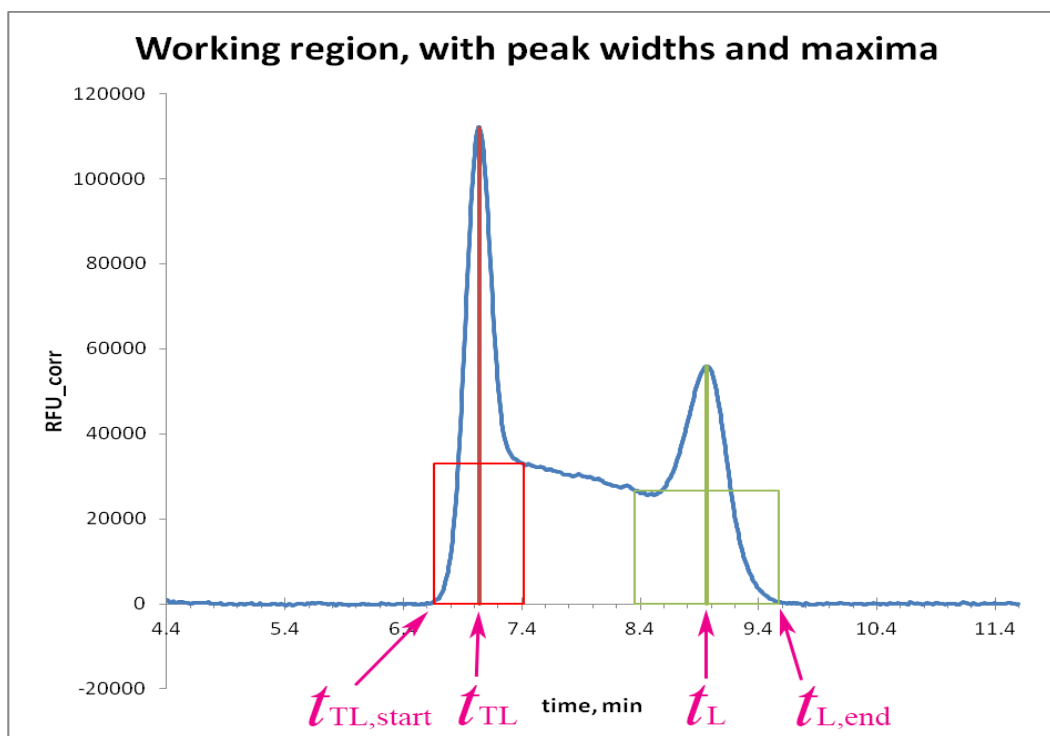


Figure S7. The NECEEM electropherogram with defined time positions: the beginning of the $T_{L_{int}}$ peak ($t_{TL,start}$), the end of the L peak ($t_{L,end}$), and the maxima of these peaks (t_{TL} and t_L , respectively).

Results									
#	File Name	Init. Conc. of Ligand (M)	Init. Conc. of Target (M)	K_d (M)	k_{on} ($M^{-1} s^{-1}$)	k_{off} (s^{-1})	Proportion of Free Ligand	Proportion of Complex	Proportion of Decay
1	002 150 nM AikB, 50 nM AikB aptamer,	1.00E-08	1.00E-08	5.80E-10	3.33E+06	1.93E-03	21%	35%	44%

Figure S8. Result table of the NAAP. “Proportion of Free Ligand”, “Proportion of Complex”, and “Proportion of Decay” are relative areas (i.e. total sum of areas is normalized to 1) of A_L , $A_{TL_{int}}$ and $A_{TL_{diss}}$, respectively.

4. Program for Fitting Eq. (7) into Experimental Dependence of R on $[L]_0$

4.1. Description of the Fitting Program

Experimental data of R (“Proportion of Free Ligand” in **Fig. S8**) measured at different initial concentrations of ligand $[L]_0$ were fitted by Eq. (7) using $[T]_0$ and K_d as variable parameters. For

this procedure Microsoft Excel Solver add-in is used. The Solver is a tool for finding solutions of "what-if" scenarios. The Solver is not available when Excel is installed by default. If the Solver command does not appear on the Data menu (at right side of Data ribbon in **Fig. S9**), you should select the Solver Add-In check box in the Tools menu or install Solver by running the Microsoft Office Setup program again and selecting Solver from the list of available Excel add-ins. Refer to Microsoft Office documentation for a help.

Example of Excel program that used Solver add-in for R-fitting can be downloaded from: <http://www.yorku.ca/skrylov/Research/>

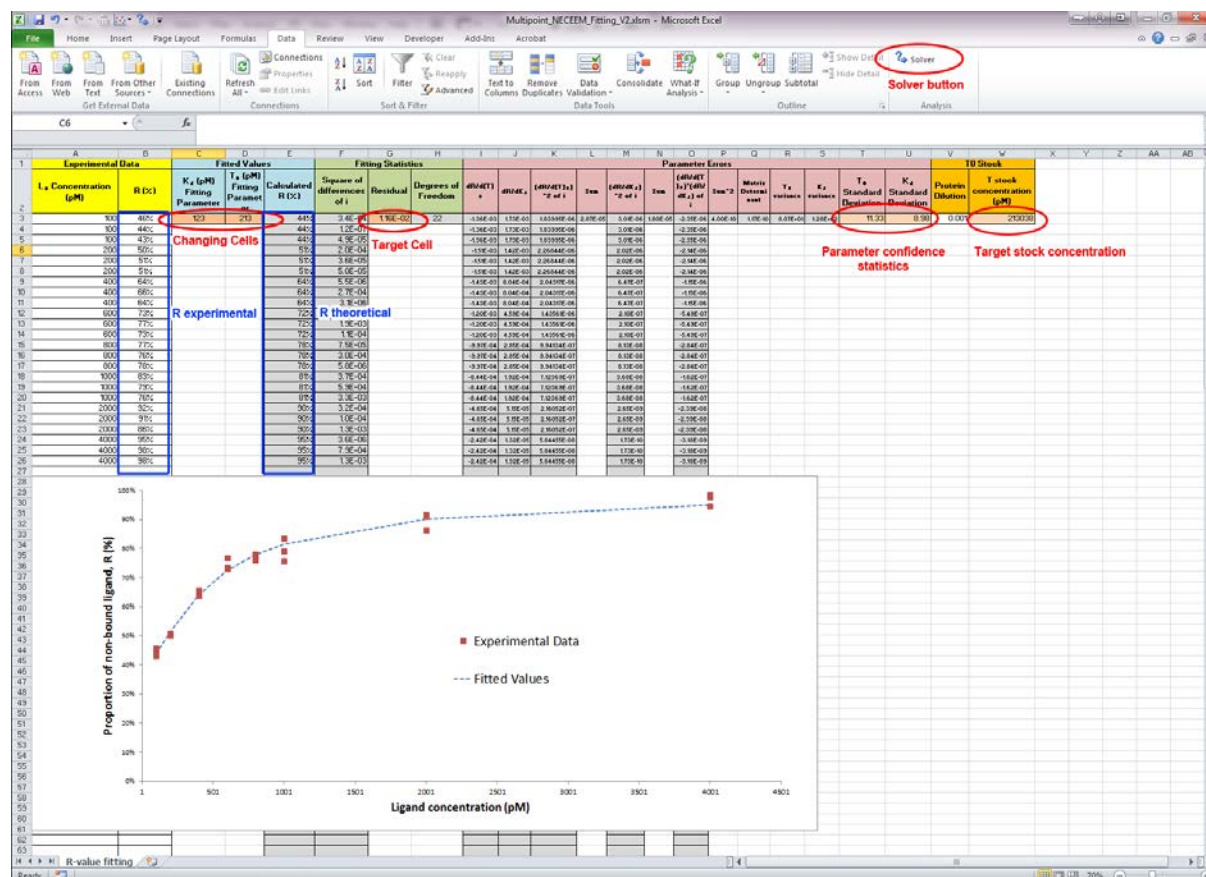


Figure S9. Example of Excel worksheet with fitting of experimental R-data using Eq. (7) and Solver add-in.

4.2. Asymptotic variance of parameters for data with uniform random error

Example of Excel worksheet contains data table that was created by NAAP is demonstrated in **Fig. S9**. Ligand concentrations and corresponding average R-values (Proportion of Free Ligand) are used for the fitting procedure. Theoretical R-values are calculated using Eq. (7) with initial values of $[T]_0$ and K_d that written in special cells ("changing cells" in **Fig. S9**). Solver adjusts the values in these "changing cells" to produce minimal result in the "target cell". Target cell contains the residual of the fitting process, S , which is the sum of squared differences between experimental and theoretical values of $[L]_0$, and can be expressed as:

$$S = \sum_{i=1}^n \left(R_{\text{exp},i} - R\left([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d\right) \right)^2 \quad (\text{S15})$$

where the values of parameter estimators \widehat{K}_d and $[\widehat{T}]_0$ are determined by iterative substitution to produce a minimized S .

Experimental variability of R translates into uncertainties in the parameters obtained by fitting. If it can be assumed that the random error in R follows Gaussian distribution and that it is uniform across all $[L]_0$, the variances of parameters estimated by the method of non-linear least squares can be calculated by assembling a covariance matrix, X_{cv} . To achieve this, the sensitivity of the dependent variable to minute changes in each of the parameters should be stacked into a Jacobian matrix X . This matrix has a size of $n \times p$, where n is the total number of experimental points, and p is the number of fitted parameters. In its general form, the matrix X can be expressed as:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix} \quad (\text{S16})$$

where i is the row index that spans between 1 and n , and j is the column index that spans between 1 and p . In X , the elements of i^{th} row and j^{th} column are calculated by the following expression:⁵

$$x_{ij} = \frac{\partial f(x_i)}{\partial \beta_j} \quad (\text{S17})$$

where $f(x)$ is the function which is fitted to the experimental results, and β_j is the parameter estimated by fitting. The covariance matrix for the set of fitted parameters is defined as:⁵

$$X_{\text{cv}} = \sigma_{\text{res}}^2 \left([X^T X]^{-1} \right) \quad (\text{S18})$$

where σ_{res}^2 is the true variance of the least squares residual, and X^T is transposed matrix X . From this, the estimators of variance for each parameter β_j can be calculated as follows:

$$\widehat{\text{var}}(\beta_j) = \sigma_{\text{res}}^2 \left([X^T X]^{-1} \right)_{jj} \quad (\text{S19})$$

Since the true value of residual variance cannot be determined in the context of regression analysis, it can be approximated by the minimized value of the sum of squared differences S , from Eq. S15, normalized by the number of statistical degrees of freedom.⁵ Thus, Eq. S19 can be written as:

$$\widehat{\text{var}}(\beta_j) \approx \frac{S}{n-p} \left([X^T X]^{-1} \right)_{jj} \quad (\text{S20})$$

For large matrices, manipulations such as multiplication and inversion can be difficult to perform manually, and generally require the use of computational approaches. However, in our case the matrix X contains only 2 columns (due to $p = 2$), which allows for all of the required matrix manipulations to be reduced to a closed form expression:

$$[X^T X]^{-1} = \frac{1}{\Delta} \begin{pmatrix} \sum_{i=1}^n x_{i2}^2 & -\sum_{i=1}^n x_{i1}x_{i2} \\ -\sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i1}^2 \end{pmatrix} \quad (\text{S21})$$

where the determinant Δ is defined by:

$$\Delta = \left(\sum_{i=1}^n x_{i1}^2 \right) \left(\sum_{i=1}^n x_{i2}^2 \right) - \left(\sum_{i=1}^n x_{i1}x_{i2} \right)^2 \quad (\text{S22})$$

Thus, substitution of Eq. S21 into Eq. S20, in the case of a 2 parameter problem, allows the variances of the estimated parameters to be calculated by:

$$\widehat{\text{var}}(\beta_1) \approx \frac{S}{(n-2)} \frac{1}{\Delta} \sum_{i=1}^n \left(\frac{\partial f(x_i)}{\partial \beta_2} \right)^2 \quad (\text{S23})$$

$$\widehat{\text{var}}(\beta_2) \approx \frac{S}{(n-2)} \frac{1}{\Delta} \sum_{i=1}^n \left(\frac{\partial f(x_i)}{\partial \beta_1} \right)^2 \quad (\text{S24})$$

In our specific case, the general equation Eq. S17 takes on the following form:

$$x_{ij} = \frac{\partial R([L]_{0i})}{\partial \beta_j} \quad (\text{S25})$$

which essentially represents partial derivatives of equation S5 with respect to each of the fitted parameters. We can define β_1 and β_2 as $[T]_0$ and K_d , respectively, and calculate the derivatives of Eq. S5 with respect to them:

$$x_{i1} = \frac{\partial R([L]_{0i})}{\partial [T]_0} = -\frac{1}{2[L]_0} + \frac{1}{4} \frac{\frac{\widehat{K}_d + [T]_0 - [L]_0}{[L]_0^2}}{\sqrt{\left(\frac{\widehat{K}_d + [T]_0 - [L]_0}{2[L]_0} \right)^2 + \frac{\widehat{K}_d}{[L]_0}}} \quad (\text{S26})$$

$$x_{i2} = \frac{\partial R([L]_{0i})}{\partial K_d} = -\frac{1}{2[L]_0} + \frac{1}{2} \frac{\frac{\widehat{K}_d + [\widehat{T}]_0 - [L]_0}{2[L]_0^2} + \frac{1}{[L]_0}}{\sqrt{\left(\frac{\widehat{K}_d + [\widehat{T}]_0 - [L]_0}{2[L]_0}\right)^2 + \frac{\widehat{K}_d}{[L]_0}}} \quad (S27)$$

where the values of $[L]_0$ are determined by the experiment, and values of parameter estimators \widehat{K}_d and $[\widehat{T}]_0$ are determined by the results of least squares fitting. Substitution of Eqs. S26 and S27 into Eqs. S23 and S24, respectively, allows variances to be calculated for each of the parameter estimators:

$$\widehat{\text{var}}([\widehat{T}]_0) \approx \frac{S}{(n-2)} \frac{1}{\Delta} \sum_{i=1}^n \left(\frac{\partial R}{\partial K_d}([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d) \right)^2 \quad (S28)$$

$$\widehat{\text{var}}(\widehat{K}_d) \approx \frac{S}{(n-2)} \frac{1}{\Delta} \sum_{i=1}^n \left(\frac{\partial R}{\partial [T]_0}([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d) \right)^2 \quad (S29)$$

where the determinant Δ takes on a specific form:

$$\Delta = \left(\sum_{i=1}^n \left(\frac{\partial R}{\partial [T]_0}([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d) \right)^2 \right) \left(\sum_{i=1}^n \left(\frac{\partial R}{\partial K_d}([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d) \right)^2 \right) - \left(\sum_{i=1}^n \left(\frac{\partial R}{\partial [T]_0}([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d) \right) \left(\frac{\partial R}{\partial K_d}([L]_{0i}, [\widehat{T}]_0, \widehat{K}_d) \right) \right)^2 \quad (S30)$$

and S , which denotes the sum of squared differences, is defined by Eq. 15. Standard deviation associated with each of the parameters can be calculated by taking the square root of its variance. Using this approach should be sufficient for most experiments.

4.3. Asymptotic variance of parameters for data with non-uniform random error

In cases when the random error in R follows Gaussian distribution, but it cannot be assumed to be uniform across all $[L]_0$, a modified approach for calculating variance of parameter estimators can be used. Here, the asymptotic variance of the $[T]_0$, K_d estimators is given by H^{-1} where

$$H = \sum_{i=1}^n \sum_{q=1}^{m_i} \frac{1}{\widehat{\sigma}_i^2} \begin{bmatrix} x_{iq1}^2 & x_{iq1}x_{iq2} \\ x_{iq1}x_{iq2} & x_{iq2}^2 \end{bmatrix} \quad (S31)$$

Here x_{iq1} , x_{iq2} are defined as in S26 and S27, respectively, with m is the number of replicate experiments at each experimental location i , q is the index of replicates, and $\widehat{\sigma}_i^2$ is the estimated variance at each experimental location i . The asymptotic variance of $[\widehat{T}]_0$ is, thus, the (1,1) entry of H_n^{-1} , while the asymptotic variance of \widehat{K}_d is the (2,2) entry. Similarly as in S28 and S29, the variances of the parameter estimators are defined as:

$$\widehat{\text{var}}(\widehat{[T]_0}) = \frac{1}{\Delta_H} \left(\sum_{i=1}^n \sum_{q=1}^{m_i} \frac{1}{\hat{\sigma}_i^2} \left(\frac{\partial R}{\partial K_d} ([L]_{0iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \right) \quad (\text{S32})$$

$$\widehat{\text{var}}(\widehat{K_d}) = \frac{1}{\Delta_H} \left(\sum_{i=1}^n \sum_{q=1}^{m_i} \frac{1}{\hat{\sigma}_i^2} \left(\frac{\partial R}{\partial [T]_0} ([L]_{0iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \right) \quad (\text{S33})$$

where the determinant Δ_H can be written as:

$$\begin{aligned} \Delta_H = & \left(\sum_{i=1}^n \sum_{q=1}^{m_i} \frac{1}{\hat{\sigma}_i^2} \left(\frac{\partial R}{\partial [T]_0} ([L]_{0iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \right) \left(\sum_{i=1}^n \sum_{q=1}^{m_i} \frac{1}{\hat{\sigma}_i^2} \left(\frac{\partial R}{\partial [T]_0} ([L]_{0iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \right) - \\ & - \left(\sum_{i=1}^n \sum_{q=1}^{m_i} \frac{1}{\hat{\sigma}_i^2} \frac{\partial R}{\partial [T]_0} ([L]_{0iq}, \widehat{[T]_0}, \widehat{K_d}) \times \frac{\partial R}{\partial [T]_0} ([L]_{0iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \end{aligned} \quad (\text{S34})$$

Estimates of $\hat{\sigma}_i^2$ based on experimental replication are generally not sufficiently accurate when a typical number of replicates for analytical experiments (3-5 replicates) are produced. Thus, $\hat{\sigma}_i^2$ needs to be estimated in an iterative minimization approach, while fitting the parameter estimators $\widehat{K_d}$ and $\widehat{[T]_0}$. In such minimization, equation S15 is replaced by:

$$S_H = \sum_{i=1}^n \sum_{q=1}^m \frac{1}{\sigma_i^2} \left(R_{iq} - R(L_{iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \quad (\text{S35})$$

Finding the estimator can be obtained by first setting an initial best-guess value for each $\hat{\sigma}_i^0$ and minimizing S_H by iteration of $\widehat{[T]_0}$ and $\widehat{K_d}$. The best-guess value of $\hat{\sigma}_i^0$ is then refined by the following equation:

$$\hat{\sigma}_i^2 = \frac{1}{m_i} \sum_{q=1}^{m_i} \left(R_{iq} - R(L_{iq}, \widehat{[T]_0}, \widehat{K_d}) \right)^2 \quad (\text{S36})$$

The refined value of $\hat{\sigma}_i^0$ is then substituted back into Eq. S35, and new estimates of $\widehat{[T]_0}$ and $\widehat{K_d}$ are calculated. These iterations are continued until some reasonable convergence criterion is met (such as, the estimates for all consecutive iterations change by no more than 10^{-5}).

5. NECEEM-electropherogram simulation with COMSOL Multiphysics

To simulate NECEEM electropherograms we used COMSOL Multiphysics environment, version 4.4, with the “Transport of Diluted Species” module which incorporates mass balance equations, chemical reactions, convective and electrophoretic terms, and Fick’s law of diffusion. Concentrations of defined species were calculated within 1D geometry that modeled a capillary with optical detector zone. Length of the capillary up to detector and extent of the detection window were 20 cm and 0.6 mm, respectively. Injection plug had a Gaussian shape with a half-

width of 10 mm. Injected mixture consisted of ligand, target, and complex with equilibrium concentrations $[L]_{eq}$, $[T]_{eq}$, and $[C]_{eq}$, respectively:

$$[L]_{eq} = [L]_0 - [C]_{eq} \quad (S37)$$

$$[T]_{eq} = \frac{-([L]_0 - [T]_0 + K_d) + \sqrt{([L]_0 - [T]_0 + K_d)^2 + 4 [T]_0 K_d }}{2} \quad (S38)$$

$$C_{0_eq} = [T]_0 - [T]_{eq} \quad (S39)$$

where $[L]_0$ and $[T]_0$ are initial concentrations of ligand and target, respectively.

We used diffusion coefficients of $1.52 \times 10^{-10} \text{ m}^2/\text{s}$ for ligand molecules and $5.9 \times 10^{-11} \text{ m}^2/\text{s}$ for target and complex. Electrophoretic mobilities were set $1.17 \times 10^{-8} \text{ m}^2/(\text{V s})$ for ligand, $4.16 \times 10^{-8} \text{ m}^2/(\text{V s})$ for target, and $3.5 \times 10^{-8} \text{ m}^2/(\text{V s})$ for complex. Applied electric field was 333 V/cm.

Transformations of reactive species inside capillary were described with reaction constants R_L , R_T , and R_C :

$$R_L = -k_{on}[L][T] + k_{off}[C] \quad (S40)$$

$$R_T = -k_{on}[L][T] + k_{off}[C] \quad (S41)$$

$$R_C = k_{on}[L][T] - k_{off}[C] \quad (S42)$$

where $k_{on} = k_{off}/K_d$, and $k_{off} = 0.01 \text{ s}^{-1}$.

Direct PARDISO solver was used to approximate concentrations of reactive species in a time-dependent manner over defined geometry. Output signal was defined as a sum of average concentrations of ligand and complex inside of detection zone. An adjustable COMSOL model file used in this work can be downloaded from: <http://www.yorku.ca/skrylov/Research/>

6. Supporting Figures

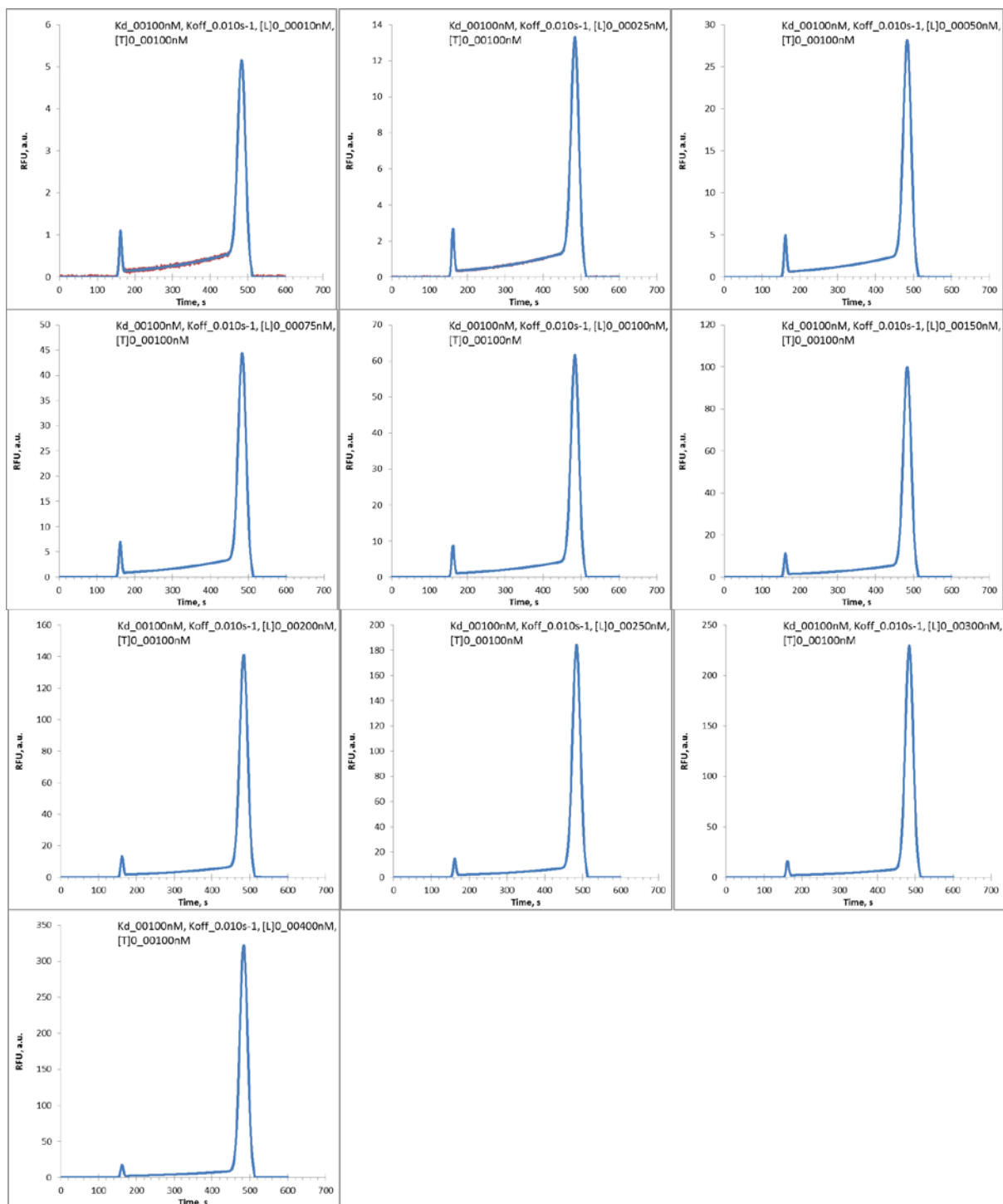


Figure S10. Set of simulated NECEEM electropherograms to find "unknown" values of K_d and $[T]_0$. In this set $[T]_0$, K_d , and k_{off} are fixed at 100 nM, 100 nM, and 0.01 s^{-1} , respectively, whereas $[L]_0$ varies between 10 and 400 nM.

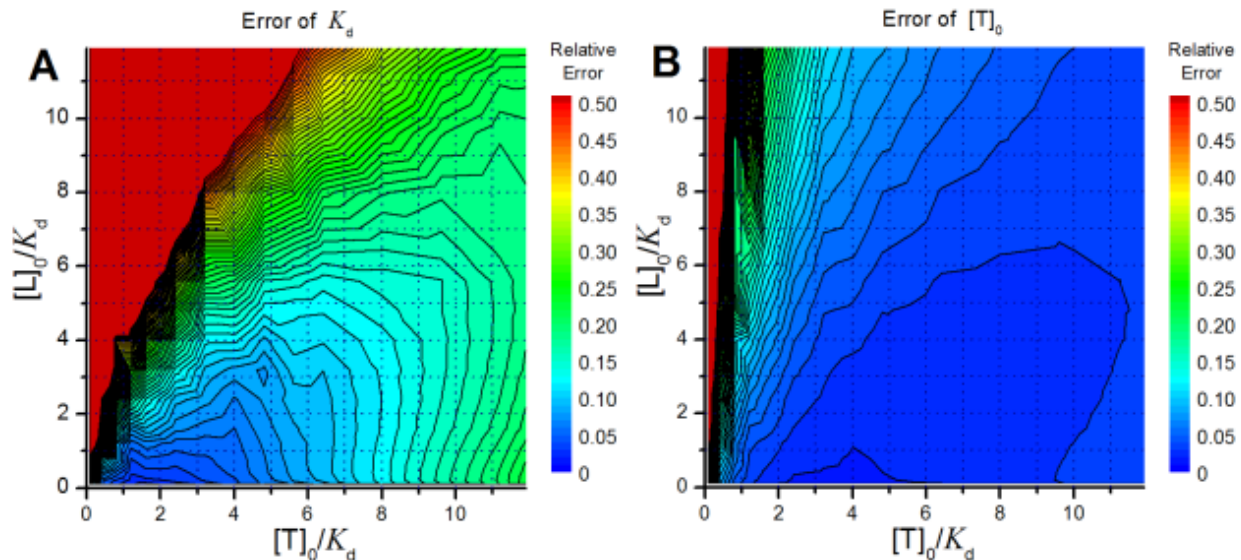


Figure S11. Error relative to K_d in determination of K_d (panel A) and $[T]_0$ (panel B) depending on ratio of concentrations $[L]_0$ and $[T]_0$ to K_d . R values were calculated from simulated electropherograms using NAAP. K_d was fixed to 100 nM. Figure 4 in the main text represents a 2D slices of these contour plots at $[L]_0/K_d = 0.2$.

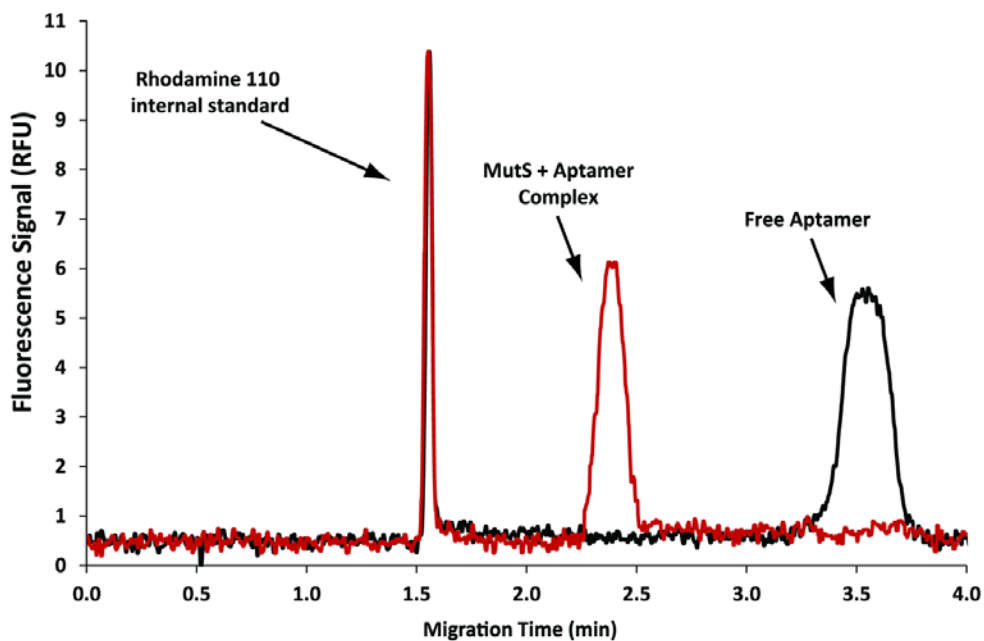


Figure S12. Determination of ϕ . Areas of peaks TL and L, at full complex saturation (red) and at absence of target protein (black) yielded a ratio of ≈ 1 . To obtain these measurements $[L]_0 = 5$ nM in both cases, and $[T]_0 = 0$ in black trace, and $[T]_0 = 20$ nM in red trace.

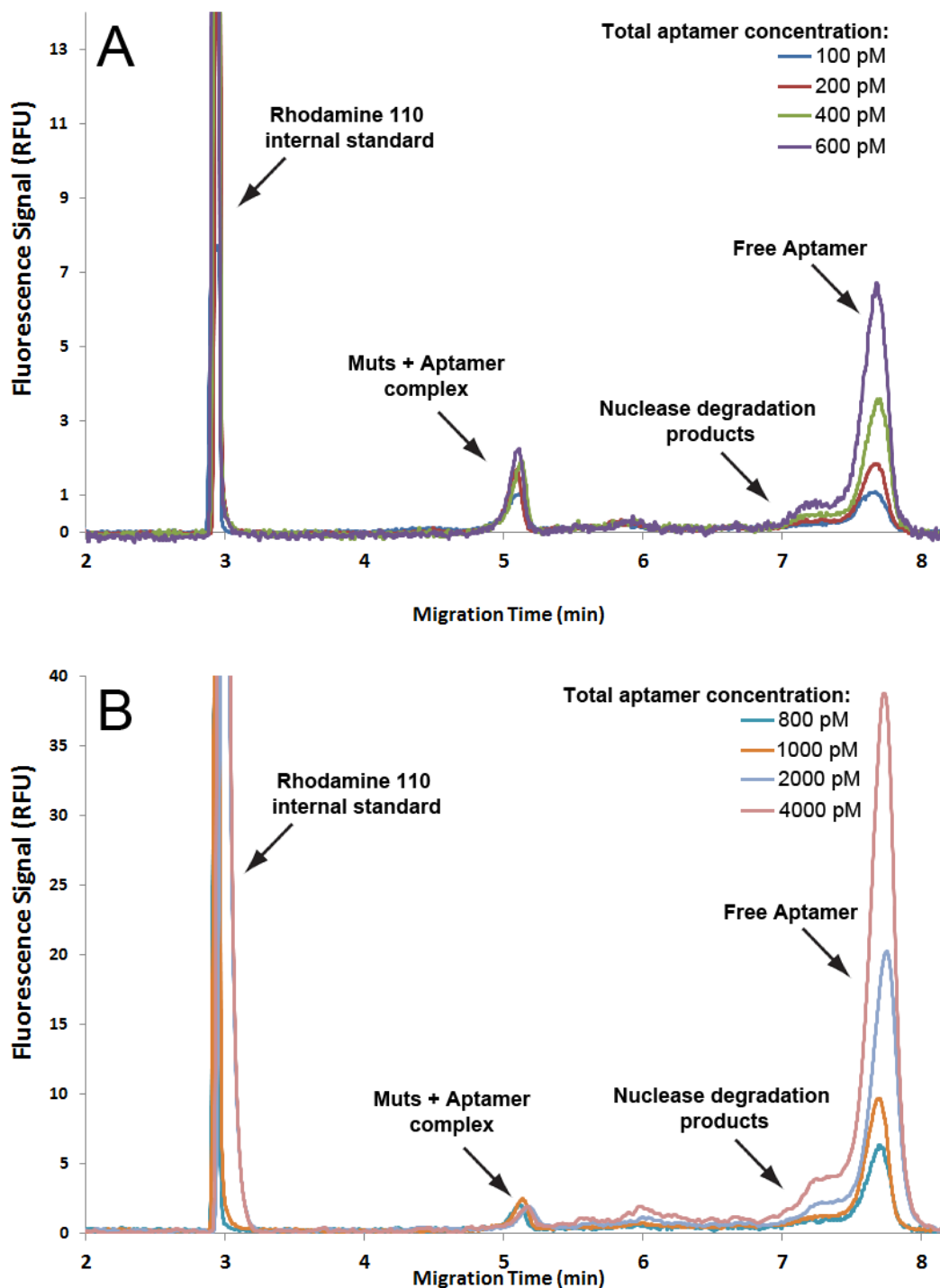


Figure S13. Representative electropherograms obtained with the custom-made instrument. In all experiments $[T]_0$ was maintained at $[T]_{\text{stock}}/1000$ dilution. $[L]_0$ varied between 100 and 600 pM in panel A, and 800 pM to 4 nM in panel B. The results were segregated in two panels to facilitate viewing peaks of significantly different magnitudes. In experiments performed on the custom-made instrument, peak “shoulders” were observed between the TL and L peaks, which were not observed in experiments performed on the commercial instrument (**Fig S12**). As

incubation times for experiments on custom-made instrument were significantly longer, we hypothesized that the shoulders were the result of nuclease-facilitated detachment of the fluorophore from the DNA aptamer. The fact that nuclease activity has been observed in MutS homologues is consistent with this hypothesis.⁶ Furthermore, the size of the “shoulder” regions significantly increased at the expense of the L peak, if the sample mixtures were analyzed 24 h after mixing (results not shown). The migration times of the shoulders are also consistent with migration pattern of free fluorescein. As fluorophore-missing aptamer sequences would still participate in binding, but both their L and TL species would be invisible to detection, their presence would be inconsequential to accuracy of the NECEEM measurements. For this reason, the inter-peak shoulders were eliminated when calculating R for experiments performed on the custom-made instrument.

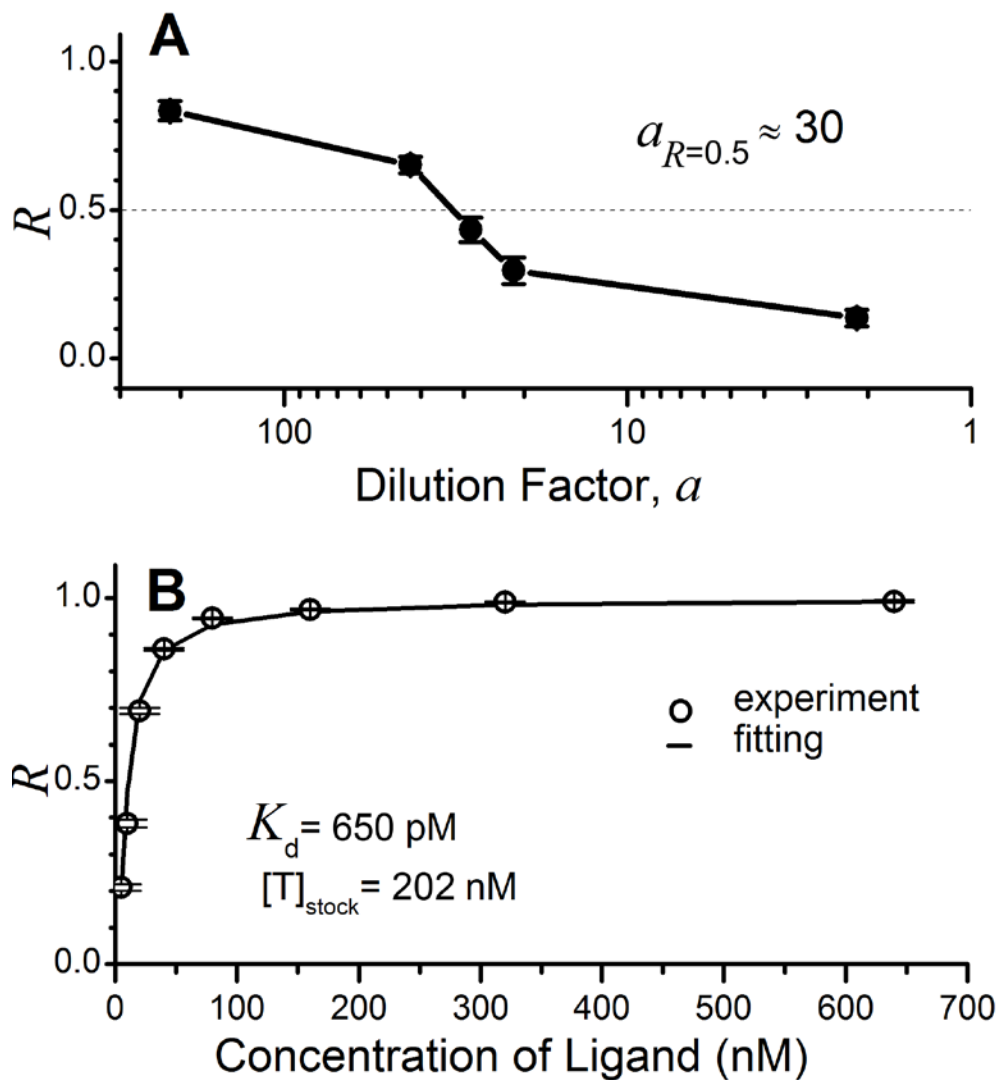


Figure S14. Determination of K_d and $[T]_0$ from experimental NECEEM data performed on a commercial instrument: **(A)** Estimation of optimal dilution factor a from the dependence of R (fraction of unbound L) on a . **(B)** Dependences of R on concentration $[L]_0$, at constant concentration $[T]_0$ obtained from the stock solution by diluting with factors $a = 30$. Open circles represents calculated R values, solid curve is a fitting with Eq. (7).

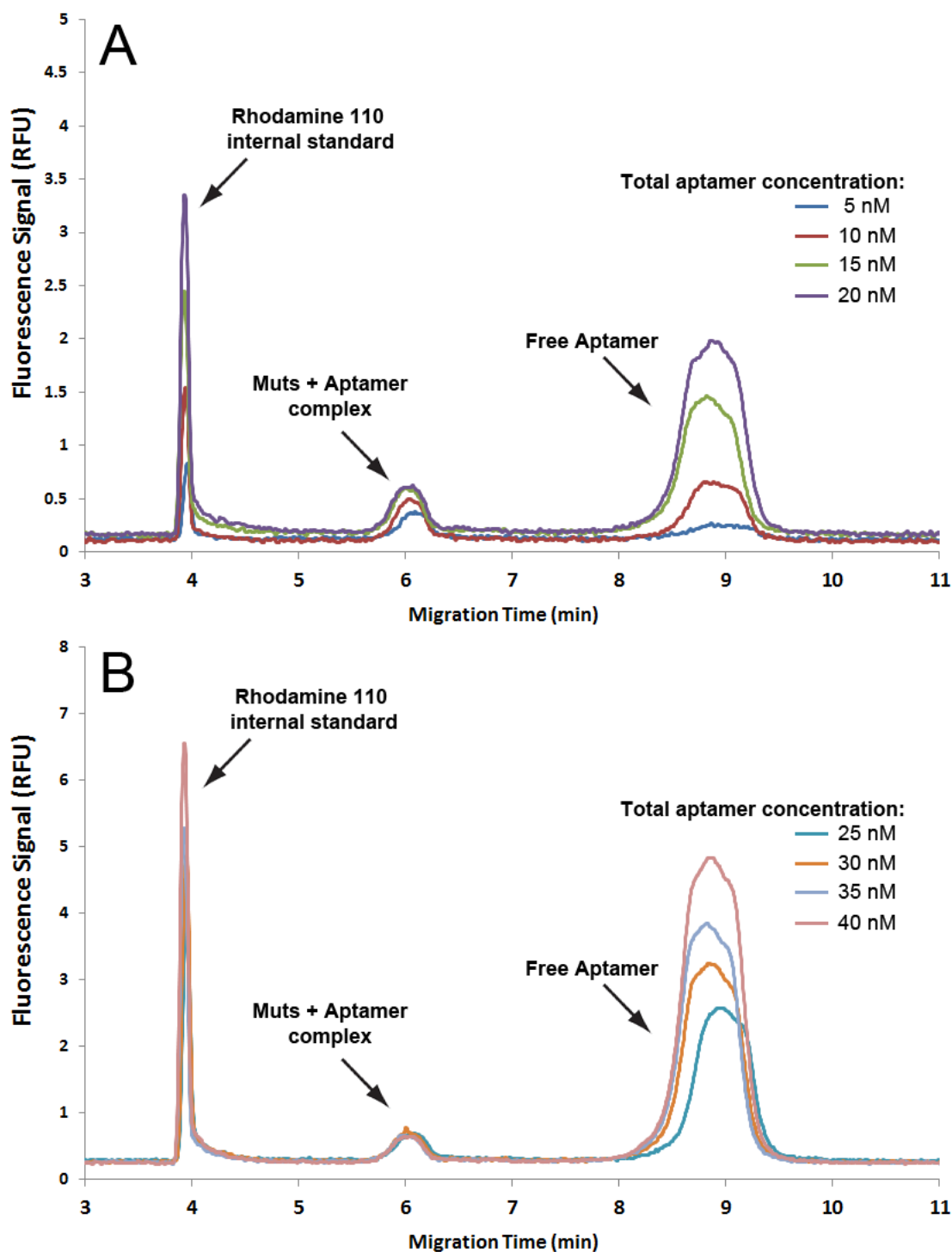


Figure S15. Representative electropherograms obtained with the commercial instrument. In all experiments $[T]_0$ was maintained at $[T]_{\text{stock}}/1000$ dilution. $[L]_0$ varied from 5 to 20 nM in panel A, and from 25 to 40 nM in panel B. The results were segregated in two panels to facilitate viewing peaks of significantly different magnitudes.

Supporting References

- (1) Cherney, L. T.; Kanoatov, M.; Krylov, S. N. *Anal. Chem.* **2011**, 83, 8617-8622
- (2) Musheev, M. U.; Filiptsev, Y.; Krylov, S. N. *Anal. Chem.* **2010**, 82, 8692-8695.
- (3) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, 36, 1627-1639.
- (4) Madden, H. H. *Anal. Chem.* **1978**, 50, 1383-1386.
- (5) van de Geer, S. A. In *Encyclopedia of Statistics in Behavioral Science*; Everitt, B. S., Howell, D. C., Eds.; John Wiley & Sons, Ltd.: Chichester, 2005; Vol. 2, pp 1041-1045.
- (6) Fukui, K.; Kosaka, H.; Kuramitsu, S.; Masui, R. *Nucleic Acids Res.* **2007**, 35, 850-860.