

**ADVANCING KINETIC CAPILLARY ELECTROPHORESIS
FOR HIGH-EFFICIENCY SCREENING OF
OLIGONUCLEOTIDE LIBRARIES IN DRUG DISCOVERY**

AN T. H. LE

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

November 2023

© AN T. H. LE, 2023

ABSTRACT

Identifying protein binders is the first step in drug discovery. The combinatorial approach, in which a library of compounds is subjected to affinity screening against a target protein, is a major way for identifying protein binders. Oligonucleotide libraries constitute the largest source of material for such affinity screening. Selecting protein binders from such libraries requires a highly efficient method for separation of protein–oligonucleotide complexes from the excess of unbound oligonucleotides. Kinetic Capillary Electrophoresis (KCE) is a rapidly advancing technique in affinity applications. It reportedly has superior partitioning efficiency, but screening oligonucleotide libraries by KCE has many challenges which must be addressed before KCE can compete with conventional surface-based screening. The goal of my research is to transform KCE into a versatile technology for screening protein binders from oligonucleotide libraries. To overcome the nonbinder background issue in KCE-based partitioning, I introduce Ideal-Filter Capillary Electrophoresis (IFCE), where binders and nonbinders travel in opposite directions. While successfully implementing IFCE conditions to be compatible with physiological buffers, a remarkable partitioning efficiency of 10^9 is achieved, the highest recorded so far. Further, I develop the first quantitative characterization of all KCE-based partitioning modes for a diverse range of protein target sizes. This systematic analysis provides guidance for CE users on selecting appropriate KCE-based partitioning conditions for a given protein target. Next, I conduct the first experimental investigation into the influence of target concentration on binder selection, aiding researchers in identifying an appropriate range of target concentrations to prevent selection failures. Finally, I gather insights from all these works to demonstrate the first highly efficient KCE-based selection of protein binders from DNA-encoded library of small molecules (DEL). This pioneering achievement showcases the capabilities of KCE-based

partitioning within the context of DEL-based drug discovery, marking a significant advancement in the field.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my profound gratitude to my supervisor, Dr. Sergey Krylov, whose exceptional qualities as a mentor have been a cornerstone of my doctoral journey. Dr. Krylov is the most brilliant and diligent scientist I have had the privilege to know. His unwavering guidance has been helping me achieve far more than I ever thought possible during my PhD. I am indebted to him for his priceless insights and ideas that enriched my research. Thank you for the countless hours dedicated to refining my manuscripts. Thank you for letting me freely explore my interests in research. Thank you for investing substantial time and effort securing grants, ensuring our lab has access to critical materials and instruments which only a few labs can have. Thank you for pushing me beyond my comfort zone, turning me into a scientist and an individual I had not imagined becoming. I am profoundly thankful for your invaluable life advice, genuine care, and the enjoyable conversations we shared. As you often say, you are our father in research, and I am committed to carrying forward your wisdom with care and pride.

I would like to extend my gratitude to my thesis committee members, Dr. Philip Johnson and Dr. Gerald Audette. They have been with me from the outset, generously offering their time, kindness, and invaluable advice throughout my time at York University. I am fortunate and appreciative to have had them as my committee members over the years. I would also like to thank my internal examiner, Dr. Mark Bayfield as well as my external examiner, Dr. Karen Waldron from University of Montreal, for dedicating their time evaluating my dissertation.

I am also deeply appreciative of my collaborators at Nurix Therapeutics, Dr. Justin Hall and his colleagues, who proved to be outstanding partners during the final phase of my doctoral journey. Their support, flexibility, and responsiveness were pivotal in the timely completion of

the last chapter of my thesis. Without them, this last chapter would have taken another 2 years to finish.

My time at York University provided me with the opportunity to connect with exceptional individuals who became mentors I hold in the highest regard: Dr. Lana Krylova, Dr. Mirzo Kanoatov, and Dr. Jiayin Bao. Their patience, expertise, and dedication in training me in the fundamentals of experiment planning, data analysis, and laboratory work have laid a strong foundation of skills and work ethics that I will carry with me throughout my career.

A very special place in my heart goes to my fellow colleagues who accompanied and supported me on this lengthy journey: Stas, Liang, Nikita, JeanLuc, Tong, Eden, Shakiba, and Giammarco. We experienced our life moments together: learning, shedding tears, growing, and achieving success. All these memories, I will forever cherish.

I must also express my sincere thanks to the senior members of the Krylov lab of the past, including Dr. Sven Kochmann, Dr. Victor Galievsky, Dr. Vasilij Koshkin, and Dr. Alexander Stasheuski. Their kind assistance with experiments, instruments, and administrative matters has been invaluable.

Finally, my unwavering courage throughout my PhD journey can only be attributed to the unconditional love and support I received from my family: my husband, Slava; my parents, Ha and Nga; and my parents-in-law, Anatoly and Elena. During moments of doubt and weariness, they stood by me, offering unwavering support, reassurance and encouragement to pursue my dreams. I know that, regardless of the circumstances, they will always be there to take care of everything for me. I would not be where I am today without them. This PhD is dedicated to them.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	vi
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xvi
LIST OF SYMBOLS	xviii
LIST OF EQUATIONS	xxii

CHAPTER 1. INTRODUCTION TO HIGHLY EFFICIENT METHODS OF SCREENING OLIGONUCLEOTIDE LIBRARIES

1.1. OVERVIEW OF SCREENING METHODS FOR OLIONUCLEOTIDE LIBRARIES.....	1
1.2. KCE INSTRUMENTATION AND SELECTION METHODOLOGY.....	6
1.2.1. Concept of KCE	6
1.2.2. Major modes of KCE-based separation used in binder selections	7
1.2.3. Theoretical considerations in adapting KCE-based separation techniques to the selection of binders from oligonucleotide libraries.....	11
1.3. THE PREMIER WORKS ON KCE-BASED BINDER SELECTION.....	14
1.3.1. Work by the Bowser group	14
1.3.2. Work by the Krylov group	15
1.4. EFFORTS OF SCIENTIFIC COMMUNITY TO ADVANCE KCE SELECTION OF BINDERS ...	17
1.4.1. Increasing the input binder quantity (Strategy 1)	17
1.4.2. Increasing the partitioning efficiency (Strategy 2)	19
1.4.3. Improving the post-partitioning process (Strategy 3).....	23
1.4.4. Additional tools to assist KCE selections of binders.....	25

1.4.5. Challenges in expanding KCE adoption in the scientific community and my project goal	27
---	----

**CHAPTER 2. DEVELOPMENT OF IDEAL-FILTER CAPILLARY ELECTROPHORESIS: A
HIGHLY EFFICIENT PARTITIONING METHOD FOR SELECTION OF PROTEIN
BINDERS FROM OLIGONUCLEOTIDE LIBRARIES**

2.1. THE PROOF OF PRINCIPLE WORK.....	33
2.1.1. The significance of single-round selection of binders from oligonucleotide libraries	33
2.1.2. Materials and methods.....	36
2.1.2.1. Materials and solutions	36
2.1.2.2. CE instrumentation and fraction collection conditions	37
2.1.2.3. Quantitative PCR	38
2.1.2.3. Specific of single-round aptamer selection	38
2.1.3. Results and discussion.....	40
2.1.3.1. Efficiency of partitioning required for one-step selection of binders	40
2.1.3.2. Hypothesis for reducing nonbinder background in KCE-based partitioning	42
2.1.3.3. Electric field to guarantee no overheating	44
2.1.3.4. Achieving $v_{P-DNA} > 0 > v_{DNA}$	46
2.1.3.5. Lowering nonbinder background <i>via</i> increasing I_{RB}	47
2.1.3.6. Efficiency of binder collection	49
2.1.3.7. One-step aptamer selection	49
2.1.4. Conclusions	51
2.2. DETERMINATION OF THE EQUILIBRIUM CONSTANT AND RATE CONSTANT OF PROTEIN-OLIGONUCLEOTIDE COMPLEX DISSOCIATION UNDER THE CONDITIONS OF IDEAL-FILTER CAPILLARY ELECTROPHORESIS	53
2.2.1. Requirement of a KCE method for determining affinity and kinetics of binder selected under physiological or IFCE conditions.....	53

2.2.2. Materials and methods.....	56
2.2.2.1. Materials and solutions	56
2.2.2.2. CE instrumentation and default conditions	57
2.2.3. Results and discussion.....	58
2.2.3.1. General procedures for finding K_d and k_{off}	58
2.2.3.2. Theory of the double-passage approach.....	61
2.2.3.3. Conditions for the first passage.....	64
2.2.3.4. Recovery of peak areas	65
2.2.3.5. Recovery of R	67
2.1.3.6. Experimental determination of K_d and k_{off} of protein–aptamer complex	68
2.2.4. Conclusions	71

CHAPTER 3. QUANTITATIVE CHARACTERIZATION OF PARTITIONING IN SELECTION OF BINDERS FOR PROTEIN TARGETS FROM OLIGONUCLEOTIDE LIBRARY BY CAPILLARY ELECTROPHORESIS

3.1. IMPORTANCE OF DETERMINING THE APPROPRIATE KCE-BASED PARTITIONING METHOD IN BINDER SELECTION	72
3.2. MATERIALS AND METHODS	75
3.2.1. Materials and solutions.....	75
3.2.2. CE instrumentations and conditions	76
3.2.3. Quantitative PCR	77
3.3. THEORETICAL AND EXPERIMENTAL CONSIDERATIONS.....	77
3.3.1. Dependence of number of partitioning rounds on k_N & k_B	77
3.3.2. Key differences within the binder-collection windows in different KCE-based partitioning methods	80
3.3.3. Parameters influencing k_N and k_B in CE-Based Partitioning	82

3.3.4. Rationale for experimental design	84
3.4. RESULTS AND DISCUSSION.....	85
3.4.1. Migration profiles of DNA in different modes of CE-based partitioning	85
3.4.2. Influence of MW_P on k_N	88
3.4.3. Influence of MW_P on k_B	90
3.4.4. Influence of MW_P on the number of partitioning rounds	92
3.5. CONCLUSIONS.....	97

CHAPTER 4. INFLUENCE OF TARGET CONCENTRATION ON BINDER SELECTION:
AN EXPERIMENTAL STUDY

4.1. CRITICAL ROLE OF TARGET CONCENTRATION IN SELEX.....	101
4.2. MATERIALS AND METHODS	101
4.2.1. Materials and solutions	103
4.2.2. CE instrumentations and default conditions	104
4.2.3. Specifics of CE-based fraction collection	105
4.2.4. PCR procedures and generation of binder-enriched library	105
4.2.5. Specifics of bulk affinity test	107
4.3. THEORETICAL AND EXPERIMENTAL CONSIDERATIONS IN THE STUDY OF EFFECTS OF TARGET CONCENTRATION IN SELEX	108
4.4. RESULTS AND DISCUSSION.....	111
4.5. CONCLUSIONS.....	115

CHAPTER 5. CAPILLARY ELECTROPHORESIS-BASED SELECTION FROM DNA-
ENCODED LIBRARIES OF SMALL MOLECULES: THE PROOF OF PRINCIPLE STUDY

5.1. AN INTRODUCTION TO DNA-ENCODED LIBRARIES OF SMALL MOLECULES AND THE PROSPECTIVE ROLE OF KCE-BASED PARTITIONING IN DEL PLATFORM	117
5.2. MATERIALS AND METHODS	123

5.2.1. Materials and solutions	123
5.2.2. CE instrumentations and default conditions	125
5.2.3. Quantitative PCR	126
5.2.3. Specifics of KCE-based selection of SM ligands for sEH from a pre-screened DEL library .	127
5.3. THEORETICAL AND EXPERIMENTAL CONSIDERATIONS IN DESIGNING AND PROVING HIGHLY EFFICIENT KCE-BASED SELECTION OF SMALL MOLECULE LIGANDS FROM DELS	128
5.4. RESULTS AND DISCUSSION	130
5.4.1. Identification of suitable KCE conditions to separate target–DEL complexes from unbound DEL	130
5.4.2. Selection of small-molecule binders from hypothetical DEL libraries: a comparative evaluation of KCE-based and bead-based partitioning methods	135
5.4.3. Proposed KCE-based selection of DEL binders for a therapeutic protein target from a real DEL library	140
5.5. CONCLUSIONS	144
LIMITATIONS AND FUTURE DIRECTIONS	145
CONCLUDING REMARKS	151
LIST OF PUBLICATIONS	153
REFERENCES	154
APPENDICES	166
APPENDIX A. SUPPORTING INFORMATION FOR IDEAL-FILTER CAPILLARY ELECTROPHORESIS (PROOF-OF-PRINCIPLE WORK).....	166
A1. Determination of elution windows of the MutS–aptamer complex	166
A2. Details on the determination of efficiency of binder collection.....	167
A3. Evaluation of binding affinity of aptamer pool selected for MutS	170
APPENDIX B. SUPPORTING INFORMATION FOR DOUBLE-PASSAGE APPROACH	171

B1. Calculating R from the velocity-corrected peak areas via double-passage approach	171
B2. Calculating k_{off} by varying E during the second passage	171
B3. Recovery of peak areas	171
B4. Recovery of R	173
B5. Experimental determination of K_d and k_{off} of protein–aptamer complex	176
B6. Determination of v_L	177
APPENDIX C. SUPPORTING INFORMATION FOR QUANTITATIVE CHARACTERIZATION OF KCE-BASED PARTITIONING	178
C1. Dependence of number of partitioning rounds on k_N & k_B	178
C2. Empirical mathematical model to predict the electrophoretic mobility of protein–DNA complexes	180
C3. Determination of elution time of protein–DNA complex, the binder-collection window, and k_N for a given value of MW_p	182
C4. The predicted elution time of protein–DNA complex with MW_p ranging from 15 to 150 kDa in KCE-based partitioning	184
C5. The dependence of m_{min} and m_{max} on MW_p ranging from 15 to 150 kDa in IFCE for DNA library with moderate to high binder abundance and bulk $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$	185
APPENDIX D. SUPPORTING INFORMATION FOR PROTEIN CONCENTRATION STUDY	186
D1. Determination of the binder-elution window	186
D2. Summary of nonbinder background and relative yield of DNA values obtained in MutS and thrombin selections	188
D3. Data analysis for bulk affinity assays.....	191
APPENDIX E. SUPPORTING INFORMATION FOR KCE-BASED SELECTION FROM DEL	196
E1. Quality assessment of q-PCR determined data used in mock selection for SA and sEH	196
E2. Initial analysis of binder-enriched pool obtained from bead-based and KCE-based selection of SM ligands for sEH	199

LIST OF TABLES

Table 5.1. Comparison of KCE-based and bead-based partitioning performance in the selection of SM ligands from DEL.	139
Table B1. Recovery of peak areas for double-passage experiment with 1 μ M GFP	172
Table B2. Recovery of R for double-passage experiment of GFP–DNA mixtures	175
Table B3. Experimental determination of K_d of MutS–aptamer complex	177
Table D1. Relative standard deviation of q	188
Table D2. Summary of k_N and q values obtained in MutS selections.....	189
Table D3. Summary of k_N and q values obtained in thrombin selections	190
Table D4. Summary of R values for the starting library (Round 0) and binder-enriched libraries (Round 1–3) obtained in MutS selections at 4 different target concentration schemes: 1, 10, 100 and 500 nM MutS	193
Table D5. Summary of R values for the starting library (Round 0) and binder-enriched libraries (Round 1–3) obtained in thrombin selections at four different target concentrations: 1, 10, 100 and 500 nM thrombin	195
Table E1. Quantification of DEL molecules at the input and output of partitioning for sEH mock selection using q-PCR analysis	198

LIST OF FIGURES

Figure 1.1. Schematic representation of KCE separation when L is labeled for detection while P is undetectable	8
Figure 1.2. Illustration of different modes of KCE-based separation	9
Figure 1.3. Schematic representation of selection of smart binders with predefined k_{off} by NECEEM and predefined K_d by ECEEM	16
Figure 2.1. Schematic depiction of 1-step selection of binders from an oligonucleotide library .	41
Figure 2.2. Schematics of KCE-based partitioning of protein target–binder complexes (P–DNA) from nonbinders (unbound DNA).....	43
Figure 2.3. Determination of E that guarantees no overheating	45
Figure 2.4. The effect of NaCl added to RB on migration pattern of protein-DNA complexes and unbound DNA in CE with “+” and “–” at the inlet, respectively	47
Figure 2.5. The effect of NaCl added to RB on the migration pattern of DNA in CE with “+” at the inlet.....	48
Figure 2.6. IFCE-based partitioning of DNA binders of MutS protein from the unbound library	50
Figure 2.7. Schematic of the double-passage approach for finding K_d and k_{off} under the IFCE conditions	55
Figure 2.8. Determination of velocity-corrected peak areas via double-passage experiment.	66
Figure 2.9. Recovery of R for double-passage experiments of GFP–DNA mixtures with [DNA]/[GFP] ranging from 0.08 to 4	68
Figure 2.10. Double-passage experiment with a mixture of MutS protein and its aptamer under different electric field strengths	69
Figure 3.1. Schematic representation of the binder collected window in different modes of KCE-based partitioning	81
Figure 3.2. DNA background profiles under conditions of NECEEM and IFCE	86
Figure 3.3. The effect of the molecular weight of protein on the transmittance of KCE-based partitioning for nonbinders under the conditions of NECEEM and IFCE.....	89
Figure 3.4. The dependency of the transmittance of KCE-based partitioning for binders on the molecular weight of target protein under the conditions of NECEEM and IFCE	91

Figure 3.5. The dependence of required and allowed numbers of partitioning rounds on the molecular weight of target protein in non-SELEX selection of aptamers using different modes of KCE-based partitioning	93
Figure 3.6. The dependence of required and allowed numbers of partitioning rounds on the molecular weight of target protein in non-SELEX selection of aptamers using different modes of KCE-based partitioning	98
Figure 4.1. Schematic representation of efficient partitioning of binders from nonbinders in a mixture of oligonucleotide library and protein target	109
Figure 4.2. Comparison of q values and bulk affinities represented by R values to evaluate the selection outcomes for MutS and thrombin under four different target concentrations	113
Figure 5.1. Schematic representation of two general assembly pathways for DELs	119
Figure 5.2. Conceptual depiction of migration patterns of DEL, protein and protein-DEL complex in different KCE-based partitioning modes	122
Figure 5.3. Structure details of the DNA tags used in the DEL ligands for SA and sEH	124
Figure 5.4. Complex-first NECEEM-based binding analysis of SA and its DEL ligands, binder and nonbinder	132
Figure 5.5. Complex-last NECEEM-based binding analysis of sEH and its DEL ligands, binder and nonbinder	134
Figure 5.6. Evaluation of partitioning performance of complex-first NECEEM and complex-last NECEEM in mock selections of SM binder from hypothetical DELs for SA and sEH respectively	137
Figure 5.7. Evaluation of partitioning quality of complex-last NECEEM and magnetic bead-based selection of SM ligands from a real DEL library for sEH	142
Figure A1. Determination of time windows for MutS-aptamer complex collection for RBs based on 50 mM Tris-HCl pH 7.0 with concentrations of NaCl varying from 0 to 100 mM	166
Figure A2. Determination of B_{in} and B_{out} for IFCE-based partitioning of MutS-bound aptamer and the unbound aptamer.	169
Figure A3. Pressure-aided IFCE separation of the components of equilibrium mixture containing 20 nM fluorescently labeled enriched DNA library and 100 nM MutS	170
Figure B1. Double-passage experiment performed with 1 μ M GFP in IFCE running buffer	172
Figure B2. Determination of φ for GFP-DNA mixture	173

Figure B3. Double-passage experiments of GFP–DNA mixtures with [DNA]/[GFP] ratios ranging from 0.08 to 4	174
Figure B4. Determination of φ for MutS–aptamer mixture.....	176
Figure B5. Double-passage experiment with mixture of MutS protein and its aptamer under different electric field strengths	176
Figure B6. Determination of v_L	177
Figure C1. Line of the best fit for the electrophoretic mobility of protein–ssDNA complex as a function of X : $\mu_{P-DNA} = A + BX$, where $X = \mu_{P-DNA} L_{DNA}^{0.68} MW_{P-DNA}^{-1/3}$	181
Figure C2. An example of binder-collection time window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM.....	183
Figure C3. The predicted elution time for the protein–DNA complex as a function of MW_P under conditions of NECEEM and IFCE.....	184
Figure C4. The dependence of m_{min} and m_{max} on MW_P in IFCE for DNA library with moderate and high binder abundance	185
Figure D1. Determination of binder-elution window for NECEEM-based selection for MutS and thrombin	187
Figure D2. Bulk affinity tests of the starting library and the binder-enriched libraries obtained in MutS selections at four different target concentrations using the bulk affinity workflow	192
Figure D3. Bulk affinity tests of the starting library and binder-enriched libraries obtained from thrombin selections at four different target concentrations using the bulk affinity workflow	194
Figure E1. Calibration curves for binders and nonbinders using binder-specific and nonbinder-specific primer sets, respectively.	197
Figure E2. Correlation between a set of controls detected in bead-based and KCE-based selections.	201
Figure E3. Preference of selected DEL ligands for bead-based or KCE-based partitioning.	202

LIST OF ABBREVIATIONS

Bodipy	4,4-difluoro-4-bora-3a,4a-diaza-s-indacene
CE	capillary electrophoresis
ctITP	capillary transient isotachophoresis
DEL	DNA-encoded library of small molecules
DNA	deoxyribonucleic acid
ECEEM	equilibrium capillary electrophoresis of equilibrium mixtures
EM	equilibrium mixture
EOF	electroosmotic flow
HTS	high-throughput sequencing
GFP	green fluorescent protein
GLCBS-l-leucine	Gly-(l)Leu-4-carboxybenzene sulfonamide
IFCE	ideal filter capillary electrophoresis
KCE	kinetic capillary electrophoresis
LIF	laser-induced fluorescence
LOD	limit of detection
LOQ	limit of quantification
MACE	microbead-assisted capillary electrophoresis
NECEEM	non-equilibrium capillary electrophoresis of equilibrium mixtures
nonSELEX	selection of ligands without PCR amplification between rounds of partitioning
PCR	polymerase chain reaction
PVA	poly(vinyl alcohol)
PectI	polymer-enhanced capillary transient isotachophoresis

qPCR	quantitative PCR
RB	running buffer
SA	streptavidin
sEH	soluble epoxide hydrolase
SELEX	systematic evolution of ligands by exponential enrichment
SM	small molecule
SUMET	simplified universal method for predicting the electrolyte temperatures
XNAs	xeno nucleic acids

LIST OF SYMBOLS

a	coefficient of proportionality between the peak area and the concentration of the fluorescent oligonucleotide
A_1	area of the first pressure-driven peak
A_2	area of the second electrophoretic peak
B_{in}	quantity of binders at the input of partitioning
B_{out}	quantity of binders at the output of partitioning
d	distance from the capillary inlet to the center of the detection window
D	diffusion coefficient of the largest-size analyte
d_1	distance between the center of the stopped plug and the center of the detection window
\mathbf{E}	vector of the electric field
E	magnitude of \mathbf{E}
I_{RB}	ionic strength of running buffer
φ	relative quantum yield of a fluorescent label within the protein–oligonucleotide complex with respect to that of the free oligonucleotide
k_{B}	transmittance of partitioning for binders or binder recovery
K_{d}	equilibrium dissociation constant
k_{N}	transmittance of partitioning for nonbinders or nonbinder background
k_{off}	rate constant of dissociation
k_{on}	rate constant of association
L	ligand molecule
l	length of sample plug

L_{DNA}	length of DNA
l_{sep}	separation distance
L_{T}	output quantity of the library in the presence of the target
$[\text{L}]_0$	initial concentration of ligand
$[\text{L}]_{\text{eq}}$	equilibrium concentration of ligand
m	number of partitioning rounds
MW_{P}	molecular weight of protein target
$\text{MW}_{\text{P-DNA}}$	molecular weight of complex of protein target and DNA
μ_{DNA}	electrophoretic mobility of DNA
$\boldsymbol{\mu}_{\text{EOF}}$	electrophoretic mobility vector of electroosmotic flow
μ_{EOF}	electrophoretic mobility of electroosmotic flow (EOF)
$\boldsymbol{\mu}_{\text{L}}$	electrophoretic mobility vector of ligand
μ_{L}	electrophoretic mobility of ligand
$\mu_{\text{P-DNA}}$	electrophoretic mobility of protein–DNA complex
$\boldsymbol{\mu}_{\text{PL}}$	electrophoretic mobility vector of protein–ligand complex
μ_{PL}	electrophoretic mobility of protein–ligand complex
N_{in}	quantity of nonbinders at the input of partitioning
N_{out}	quantity of nonbinders at the output of partitioning
N_{PCR}	level of PCR noise
P	protein target molecule
$[\text{P}]_0$	initial concentration of protein
$[\text{P}]_{\text{eq}}$	equilibrium concentration of protein
P-DNA	protein–DNA complex

PL	protein–ligand complex
$[PL]_{eq}$	equilibrium concentration of protein–ligand complex
q	relative yield of DNA (ratio between the quantity of DNA after partitioning in the presence of the target and that in the absence of the target)
Q_1	factor by which the output quantity of binder must reliably exceed the level of PCR noise
Q_2	desired binder purity
r	capillary radius
R	fraction of unbound ligand at equilibrium
R_s	analytical resolution
t	time of separation (or elution time)
t_1	the time span from the beginning of EM plug propagation to the time when the plug’s center reaches the center of the detection window
t_2	time span from the plug center’s passing the center of detection window to the end of plug propagation
t_3	time span from the beginning of electrophoresis to the moment when the center of the plug of PL reaches the center of the detection window
t_3'	varied value of t_3 by changing the electric field strength
T_{cap}	temperature inside the capillary
$T_{coolant}$	temperature of coolant
T_{cooled}	temperature of the cooled region inside the capillary
$T_{non-cooled}$	temperature of the uncooled region inside the capillary
t_{rebind}	characteristic rebinding time

t_{sep}	characteristic separation time of complex from unbound ligand and protein
v_{av}	pressure-driven velocity during the first passage
\mathbf{v}_{DNA}	electrophoretic velocity vector of DNA
v_{DNA}	electrophoretic velocity of DNA
\mathbf{v}_{L}	electrophoretic velocity vector of ligand
v_{L}	electrophoretic velocity of ligand
\mathbf{v}_{P}	electrophoretic velocity vector of protein
$\mathbf{v}_{\text{P-DNA}}$	electrophoretic velocity vector of protein–DNA complex
$v_{\text{P-DNA}}$	electrophoretic velocity of protein–DNA complex
\mathbf{v}_{PL}	electrophoretic velocity vector of protein–ligand complex
v_{PL}	electrophoretic velocity of protein–ligand complex
Z_{B}	base of the exponent describing PCR amplification of binders
Z_{N}	base of the exponent describing PCR amplification of nonbinders

LIST OF EQUATIONS

1. $P + L \xrightleftharpoons[k_{\text{off}}]{k_{\text{on}}} PL$ 6
2. $K_d = \frac{[P]_{\text{eq}}[L]_{\text{eq}}}{[PL]_{\text{eq}}} = \frac{k_{\text{off}}}{k_{\text{on}}}$ 7
3. $k_B = B_{\text{out}}/B_{\text{in}}$ 40
4. $k_N = N_{\text{out}}/N_{\text{in}}$ 41
5. $k_B/k_N \geq 100/(B_{\text{in}}/N_{\text{in}})$ 41
6. $R = \frac{[L]_{\text{eq}}}{[L]_0}$ 59
7. $K_d = \frac{[P]_0 - [L]_0(1-R)}{1/R - 1}$ 59
8. $[L]_0/K_d \leq 1$
 $[P]_0/K_d \approx 1$ 59
9. $R = -\frac{K_d + [P]_0 - [L]_0}{2[L]_0} + \sqrt{\left(\frac{K_d + [P]_0 - [L]_0}{2[L]_0}\right)^2 + \frac{K_d}{[L]_0}}$ 59
10. $\frac{1-R(t)}{1-R_{t=0}} = e^{-k_{\text{off}}t}$ 60
11. $t_{\text{sep}} \ll t_{\text{rebind}}$ 60
12. $t_{\text{sep}} = \frac{l}{\max\{|\mathbf{v}_{PL} - \mathbf{v}_L|, |\mathbf{v}_{PL} - \mathbf{v}_P|\}}$ 60
13. $t_{\text{sep}} = \frac{l}{v_{PL} + v_L}$ 60
14. $t_{\text{rebind}} = 1/(k_{\text{on}}[P]_{\text{eq}}) \approx 1/(k_{\text{on}}[P]_0)$ 61
15. $A_1 = a(\phi[PL]_{\text{eq}} + [L]_{\text{eq}})/v_{\text{av}}$ 61
16. $v_{\text{av}} = \frac{d-l/2}{t_1}$ 62
17. $d_1 = v_{\text{av}}t_2$ 62

18.	$A_2 = a\varphi[\text{PL}]_{\text{eq}} / v_{\text{PL}}$	62
19.	$v_{\text{PL}} = \frac{d_1}{t_3}$	63
20.	$R = \frac{[\text{L}]_{\text{eq}}}{[\text{L}]_0} = \frac{A_1 v_{\text{av}} - A_2 v_{\text{PL}}}{A_2 v_{\text{PL}} / \varphi + A_1 v_{\text{av}} - A_2 v_{\text{PL}}}$	63
21.	$v_{\text{PL}} = \mathbf{E} \cdot (\boldsymbol{\mu}_{\text{PL}} + \boldsymbol{\mu}_{\text{EOF}})$	63
22.	$k_{\text{off}} = \frac{\ln\left(\frac{1-R(t_3)}{1-R(t_3')}\right)}{t_3' - t_3}$	63
23.	$k_{\text{on}} = k_{\text{off}} / K_{\text{d}}$	64
24.	$v_{\text{L}} = d / t_4$	64
25.	$v_{\text{PL}} + v_{\text{L}} = E(\mu_{\text{PL}} + \mu_{\text{L}})$	64
26.	$t_1 \ll 1 / k_{\text{off}}$	65
27.	$v_{\text{av}} > \frac{l}{r^2 / D}$	65
28.	$\varphi = A_{\text{GFP}} / A_{\text{DNA}} = 0.11 \pm 0.01$	67
29.	$A_{2,t=0} = A_{2,t_3} e^{k_{\text{off}} t_3}$	70
30.	$\frac{B_{\text{out}}}{N_{\text{out}}} = \frac{B_{\text{in}}}{N_{\text{in}}} \frac{k_{\text{B}}}{k_{\text{N}}}$	73
31.	$B_{\text{out}} > Q_1 N_{\text{PCR}}$	78
32.	$B_{\text{out}} / N_{\text{out}} > Q_2$	78
33.	$m \leq \left\lfloor \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log(k_{\text{B}})} \right\rfloor$	78
34.	$m \geq \left\lceil \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log(k_{\text{B}} / k_{\text{N}})} \right\rceil$	78
35.	$\left\lfloor \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log(k_{\text{B}})} \right\rfloor \geq m \geq \left\lceil \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log(k_{\text{B}} / k_{\text{N}})} \right\rceil$	79

36.	$m \geq \left\lceil \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log\left((k_{\text{B}} / k_{\text{N}})(Z_{\text{B}} / Z_{\text{N}})^n\right)} \right\rceil$	79
37.	$n < \left\lfloor \frac{\log(k_{\text{N}} / k_{\text{B}})}{\log(Z_{\text{B}} / Z_{\text{N}})} \right\rfloor$	79
38.	$k_{\text{N}} = F(R_{\text{S}}) = F(\mu_{\text{P-DNA}}, \mu_{\text{DNA}}, t) = F(\text{MW}_{\text{P}}, L_{\text{DNA}}, t, \text{pH}, I_{\text{RB}})$	83
39.	$k_{\text{B,NECEEM}} = \text{const} \approx 1$ $k_{\text{B,IFCE}} = F(k_{\text{off}}, \text{MW}_{\text{P}}, L_{\text{DNA}}, \text{pH}, I_{\text{RB}}, t)$	83
40.	$\mu_{\text{P-DNA}} = A + B\mu_{\text{DNA}} L_{\text{DNA}}^{0.68} \text{MW}_{\text{P-DNA}}^{-1/3}$	84
41.	$k_{\text{B}} = \frac{B_{\text{out}}}{B_{\text{in}}} = \frac{B_{\text{in}} e^{-k_{\text{off}} t}}{B_{\text{in}}} = e^{-k_{\text{off}} t}$	90
42.	$q = L_{\text{T}}/N_{\text{out}} = (B_{\text{out}} + N_{\text{out}})/N_{\text{out}} = 1 + (B_{\text{out}}/N_{\text{out}})$	110

CHAPTER 1. INTRODUCTION TO HIGHLY EFFICIENT METHODS OF SCREENING OLIGONUCLEOTIDE LIBRARIES

The material presented in **Chapter 1** was published previously and reprinted with permission from “**Le, A.T.H.**; Krylova, S.M.; Krylov, S.N. Kinetic capillary electrophoresis in screening oligonucleotide libraries for protein binders. *TrAC – Trends in Analytical Chemistry* **2023**, *162*, 117061”. Copyright 2023 Elsevier. All of the adapted text below was written by me personally.

1.1. OVERVIEW OF SCREENING METHODS FOR OLIGONUCLEOTIDE LIBRARIES

Screening combinatorial libraries for protein binders (i.e., compounds capable of strongly and specifically binding target proteins) is the mainstream approach for identifying hit compounds in drug discovery.¹ Increasing the structural diversity of the library increases the probability of selecting suitable protein binders from the library.² Libraries in which every compound is synthesized, stored, and reacted with the target protein separately have limited diversity of approximately a million compounds.³ On the other hand, libraries synthesized *via* the split-and-mix approach and, thus, being mixtures of compounds, provide virtually unlimited diversity.⁴ They are the most diverse source of hit compounds and, accordingly, are the sole subject of our study.

The screening of a highly diverse combinatorial library for hits can be divided into two stages: binder selection and hit verification.⁵⁻¹⁰ In the selection stage, the initial library is reacted with the target, followed by (*i*) physical separation of binders (library members with affinity to a given molecular target) from nonbinders (library members without any appreciable affinity to the

target) and (ii) identification of the putative hits. The inevitably low number of copies of each unique member in such highly diverse libraries makes the hit identification process challenging.

Oligonucleotide libraries, such as random-sequence DNA and RNA libraries and DNA-encoded libraries of small molecules (DEL), address this challenge of identifying binders in low copies.^{6,9-11} The oligonucleotide moieties of the selected compounds from such libraries can be amplified by PCR, and their sequences can be identified *via* DNA sequencing. For random-sequence DNA and RNA libraries, the sequence information is used to prepare their replicas *via* chemical synthesis of the oligonucleotide sequence.¹² For DEL, decoding the small-molecule structure information through sequencing the DNA “barcode” can reveal the identities of the putative hits (e.g., small molecules capable of binding the target protein).⁶ This information is used to synthesize individual compounds without the DNA barcodes. After the selection stage, the hit compounds are proceeded to the verification stage, where their affinities are assessed by measuring quantitative parameters of binding to the target, such as the equilibrium dissociation constant (K_d) and the rate constants of association (k_{on}) and dissociation (k_{off}).^{5,6,13,14}

The goal of the selection stage is to eliminate the nonbinders while retaining the binders so that the library is reduced to a smaller pool of compounds that can undergo the subsequent identification and verification steps in a practical manner. It is widely accepted that for a given target, only a minute fraction of the library molecules will possess the required affinity and specificity to be identified as binders, with binder-to-nonbinder ratios being possibly as low as one per million or even billion. Thus, methods used for the selection stage must exhibit high efficiency of partitioning, meaning that it must be capable of efficiently eliminating nonbinders, while retaining as many binders as possible. The main source of inefficiencies in the elimination of nonbinders is the non-specific adhesion of library compounds to the surface of the

chromatographic support, filter, magnetic beads, etc.¹⁵ A common source of inefficiencies in the retention of binders is the decreased affinity between the interactants as a result of steric hindrance effects (e.g., caused by surface immobilization of molecular targets), or alteration of their conformation (e.g., caused by chemical derivatization of molecular targets).^{16,17} Thus, to ensure high efficiency of partitioning binders from nonbinders, an ideal selection method must be: (i) resistant to non-specific adhesion of library compounds to surfaces; (ii) immobilization-free; and (iii) free of requirements for chemical derivatization of the molecular target.

Today, most methods used for routine screening of oligonucleotide libraries do not satisfy any of these requirements. The binder-selection stage is usually carried out by surface-based techniques, such as bead-pulldown, affinity chromatography, or filtration.^{12,15,18,19} Bead-pulldown and affinity chromatography are hampered by the requirement of surface immobilization of the molecular target, while filtration and size-exclusion chromatography suffer from poor partitioning efficiencies due to the non-specific adhesion of the library molecules to the surface of the sieving matrices. In such methods, as much as 15% of the entire library can adhere to the surfaces non-specifically, necessitating the selection procedure to be repeated in a number of rounds (typically 10 to 20).¹⁵

The multi-round selection of binders from oligonucleotide libraries has fundamental drawbacks. In multi-round selection of aptamers from a random oligonucleotide library *via* the systematic evolution of ligands by exponential enrichment (SELEX) process (repetition of partitioning and PCR amplification), the number of selection rounds is theoretically unlimited.²⁰⁻²² However, the large number of selection round can lead to selection failure due to the accumulation of sequence biases after repetitive PCR cycles between the selection rounds.²³⁻²⁵

In selection of binders from DEL, due to the inability to PCR-amplify the binders, the consecutive rounds are conducted without PCR amplification between rounds in a process which can be called nonSELEX.^{6,26-29} Since there are always unavoidable losses of binders between selection rounds, the maximum number of rounds in nonSELEX is typically limited to three while the resulting pool is still mostly dominated by nonbinders. In addition to the practical challenges in time and resources, multi-round selection of binders using the surface-based techniques associated with poor separation efficiencies arguably creates a bottleneck in routine screening of oligonucleotide libraries. As a result of this and other limitations, 70% of selection campaigns fail to select binders from non-modified random oligonucleotide (ordinary DNA and RNA) libraries.³⁰ Despite the well appreciated limitations of the surface-based partitioning methods, they still dominate the field of oligonucleotide library screening due to their robustness, simplicity, and accessibility.

Gel-free capillary electrophoresis (CE) is a solution-based separation method and a highly promising alternative to surface-based methods for partitioning binders from nonbinders. The separation in CE is based on difference of electrophoretic mobilities between target–binder complexes and nonbinders as they migrate along a narrow-bore capillary. The appreciation of kinetics in early development of CE-based binder selection significantly enriches the analytical capabilities of the method, coining the term kinetic capillary electrophoresis (KCE) for such a platform.³¹⁻³⁴ KCE does not assume equilibrium between the target and library compounds, simply because separation disturbs equilibrium; thus, uniquely enabling kinetic analysis of the interactions and selection of binders with predefined kinetic parameters. Over the past two decades, KCE has been gradually developed into a reasonably well-understood, yet still moderately used, methodological platform for screening of oligonucleotide libraries.^{31,32,35-44}

Some authors call this approach simply CE; for the sake of uniformity and to emphasize the importance of kinetics in CE-based partitioning, I will use the term of KCE throughout this dissertation.

Being a solution-based separation technique with a superb partitioning efficiency between non-derivatized and non-immobilized target–binder complexes and free oligonucleotide library, KCE satisfies all of the three above-stated performance requirements for the selection stage of oligonucleotide library screening. So far, a variety of KCE methods have been developed and utilized for the selection of binders from oligonucleotide libraries.^{31,32,35-44} Despite the apparent differences between them, all these methods can be classified into 2 major groups: non-equilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) and ideal filter capillary electrophoresis (IFCE). The partitioning efficiency of KCE methods exceeds that of the conventional surface-based methods by 3 to 6 orders of magnitude.^{20,22,31,42} As a result, KCE methods effectively reduce the number of rounds required for completing selection from over 10 (by surface-based methods) to 1-4. Uniquely, with its kinetic capabilities, KCE has been demonstrated to facilitate the selection of “smart” binders with predefined equilibrium and kinetic parameters.³⁴ In the following sections, I will present a detailed description of KCE methodology; discuss the advantages and challenges of using KCE in screening of oligonucleotides libraries; provide a critical analysis of the achievements across the community; and lastly, explain the directions that I have set out for my research project to effectively advance KCE-based screening methodology beyond academic laboratory setting.

1.2. KCE INSTRUMENT AND SELECTION METHODOLOGY

1.2.1. Concept of KCE

As the name implies, CE is the instrumental platform of KCE. Since the introduction of modern CE by Jorgenson and Lukacs in 1981, CE has become a well-established analytical separation technique with essential roles in protein and nucleic acid characterization as well as study of biomolecular interactions.⁴⁵⁻⁴⁷ In contrast to conventional gel electrophoresis, which separate molecules travelling through a slab gel matrix, CE separates molecules as they migrate along the bore of the capillary tube filled with a conductive, liquid buffer (a.k.a., the CE running buffer). CE separation is characterized by an unprecedentedly high resolution due to the use of very thin capillaries with inner diameters ranging typically from 50 to 150 μm . The superior heat-dissipation in such narrow-bore capillaries means that temperature differences across the capillary are greatly minimized making effects of convection and lateral diffusion negligible. As a result, high-magnitude electric field can be used to facilitate fast and high-resolution separation in free solution.

KCE stands for CE-based separation of molecules which interact during separation unless fully separated.³³ KCE involves affinity binding between interacting molecules as the initial stage. In the context of KCE-based screening of oligonucleotide libraries, we assign that a protein target (P) and a binder (a.k.a. ligand, L) from the oligonucleotide library take part in reversible binding event with the formation of their intermolecular complex (PL); this binding can be described by the following reaction equation:



Thermodynamic/kinetic parameters that characterize such reversible intermolecular interactions include the equilibrium dissociation constant K_d , which is used to characterise

complex stability, and rate constants k_{on} and k_{off} , which characterize the kinetic properties of complex formation and dissociation, respectively. The definition of K_d through equilibrium concentrations of P, L, and PL, as well as its relation to k_{on} and k_{off} is:

$$K_d = \frac{[\text{P}]_{\text{eq}}[\text{L}]_{\text{eq}}}{[\text{PL}]_{\text{eq}}} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (2)$$

In KCE, the interacting molecules P, L, and their complex PL are continuously separated from each other based on differences in their electrophoretic mobilities. The re-equilibration of species in reaction 1 during their migration, essentially convolutes the information on K_d , k_{on} , and k_{off} within their migration patterns. Separation of any two species in CE requires that at least one of them bears a significant charge. For a random oligonucleotide library, due to the high density of the negative charge on the oligonucleotide moiety, CE is capable of separating it from other molecules, such as proteins and oligonucleotide–protein complexes, with remarkable efficiency.^{31,35} Obviously, the separation of PL from L improves with increasing molecular size of P. As the electrophoretic properties of DEL are mainly defined by the oligonucleotide moiety of molecules, high partitioning efficiencies can be also achieved between protein–DEL complexes and free DEL.^{48,49} In addition, because the information on affinity and kinetics of an interaction is embedded in the migration patterns of molecules, the use of KCE allows the selection of binders with predefined values of k_{on} , k_{off} , and K_d (smart binders).³⁴ To put it simply, collecting fractions of binders at pre-set migration times can produce smart binders with desired K_d and/or k_{off} depending on the mode of KCE.

1.2.2. Major modes of KCE-based separation used in binder selection

The basic procedure for KCE-based separation involves the following four steps: (i) P and L are mixed together and incubated to approach equilibration of reaction 1, (ii) a short plug of this

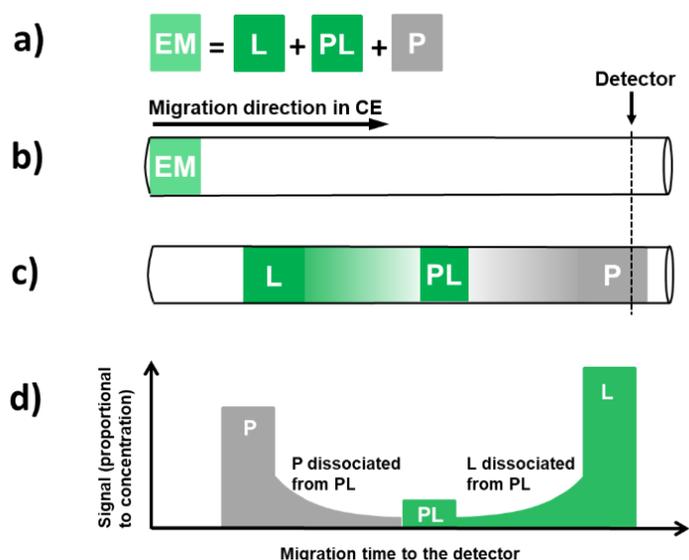


Figure 1.1. Schematic representation of KCE separation when L is labeled for detection (as depicted in bright green color) while P is undetectable (as depicted in faded gray color): **a)** preparation of the equilibrium mixture (EM); **b)** injection of the equilibrium mixture into the capillary; **c)** partitioning of PL from free L; and **d)** detection of separated PL and L as they migrate through the capillary into three distinct zones: free L zone, complex dissociation zone, and intact PL zone. Depending on the mode of KCE separation, not all three features may be present in the resulting electropherogram. See text for details.

equilibrium mixture is injected into the capillary, (iii) the free and protein-bound L are separated from each other based on the difference in electrophoretic velocities of L and PL, and (iv) the separated molecules are detected *via* a suitable detection module (e.g., absorbance or fluorescence detection) as they migrate through the capillary (**Figure 1.1**).³¹⁻³³ In most instances, L is fluorescently labelled for detection, while P is not, resulting in detectable signals from L and PL. During the separation, the equilibrium fraction of free L migrates as a distinct zone, while the PL undergoes continuous dissociation as a result of the disturbed equilibrium. Thus, a typical

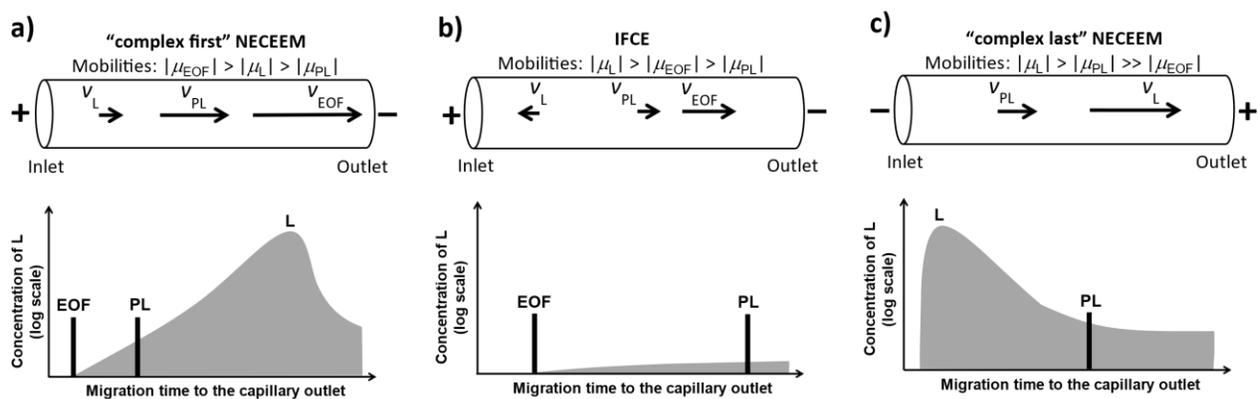


Figure 1.2. Illustration of different modes of KCE-based separation: **a)** “complex-first” NECEEM, **b)** IFCE and **c)** “complex-last” NECEEM. The top parts show electrophoretic mobilities (μ) and observed velocities (v) of the electroosmotic flow (EOF), PL and L. The bottom panels schematically depict the nonbinder background profiles originating from the nonuniform electrophoretic mobility of L in corresponding modes of KCE-based separation.

KCE migration profile contains 3 distinct features: 2 peaks that correspond to the zones of free L and intact PL and a smear-like region of PL dissociation products (referred to as the complex dissociation zone) which merges with both L and PL peaks.

We define two major modes of KCE-based separation using the migration directions of L and PL as criteria. The first mode is NECEEM, in which L and PL move in the same direction. The second mode is IFCE, in which L and PL move in the opposite directions.⁴² As a result, NECEEM electropherograms contain all three features (the two peaks and the complex dissociation zone) while only the peak of intact PL is present in IFCE electropherograms. We further divide the NECEEM platform into two submodes: “complex-first” and “complex-last”.⁵⁰ The conversion between different modes can be achieved by varying the mobility of the electroosmotic flow (EOF) (**Figure 1.2**). In complex-first NECEEM, when an uncoated capillary

is used, and the strong EOF is present, PL moves ahead of L. In complex-last NECEEM, EOF is suppressed *via* a neutral coating and as a result, L moves ahead of PL. In IFCE, the EOF is balanced so that PL moves to the capillary outlet while L moves to the inlet.

If L and PL moved as electrophoretic zone with near-Gaussian concentration distributions, KCE partitioning efficiency would approach infinity and there would be no contamination of the collected PL with the unbound L at the elution end and KCE partitioning efficiency will approach infinity.⁵¹ In reality, there is a small fraction of unbound L that moves to the elution end with the mobility close to that of PL due to a phenomenon of nonuniform migration of DNA in a uniform electric field (**Figure 1.2**, bottom panel).⁵² This contamination of binders by nonbinders mainly defines the nonbinder background, resulting in lower than theoretically expected KCE partitioning efficiency. In IFCE, the nonbinder background is approximately 3 orders of magnitude lower than in NECEEM.⁵⁰ Between the two submodes of NECEEM, the complex-last submode has relatively higher nonbinder background, owing to the additional contamination of faster-moving nonbinders at the capillary end. In the complex-first submode, the complex exits the capillary before the slower-moving nonbinders contaminate the capillary end; thus, the nonbinder background is solely caused by the irregularities in electrophoretic mobility of DNA. However, the use of complex-last NECEEM is still beneficial when the adsorption of the target protein onto the uncoated inner capillary surface is severe; the coating of the walls can suppress such adsorption, typically, at the expense of suppressing the EOF as well. Overall, the partitioning efficiencies of NECEEM and IFCE were estimated to be as high as 10^5 and 10^9 respectively while that of surface-based methods can only reach up to 10^3 , and typically below 10^2 .^{15,20,22,53} Despite having the highest partitioning efficiency on record, IFCE is still immature and limited to large-size protein only due to its very long separation times for small

proteins which leads to dissociation of complexes. Today, the well-established NECEEM still remains the most frequently used KCE methods in CE-based selection of binders from oligonucleotide libraries.

1.2.3. Theoretical considerations in adapting KCE-based separation techniques to the selection of binders from oligonucleotide libraries

A typical round of binder selection procedure involves the following three steps: (i) reacting the library (consisting of binders and nonbinders) with the target to form target–binder complexes, (ii) partitioning the complexes from nonbinders, and (iii) amplifying the collected fraction of oligonucleotides to obtain a binder-enriched library for the next round of selection.^{18,19} SELEX consists of many rounds (up to 20) of the above three steps while nonSELEX is limited to fewer rounds (up to three) of the two initial steps (reacting and partitioning).^{26,27}

The post-partitioning process involves PCR amplification and identification of binders from the binder-enriched library *via* either cloning or high-throughput sequencing (HTS).^{26,27,54} In any binder-selection procedure, the odds of success depend on two critical conditions.⁵⁰ The first condition is that the partitioning efficiency must be high enough to enrich the initial library having a low binder-to-nonbinder ratio (e.g., $<10^{-6}$) to an acceptable level of binder purity at the output (e.g., $>50\%$). A single-round selection can be achieved if partitioning efficiency is higher than the reciprocal fraction of binders in the initial library.⁵⁵ In most cases, a successful selection requires multiple rounds due to the limited partitioning efficiency and the miniscule fraction of high-affinity binders in the initial library. The second condition is that the quantity of binders at the input be sufficient for the next rounds of selection and/or PCR amplification. The quantity of

binders at the output can be simply increased *via* two means: (i) using a library with better affinity to the target (i.e., with a larger fraction of binders) and/or (ii) increasing the quantity of the initial library at the input of selection. In essence, when developing a selection procedure, one needs to aim at increasing the partitioning efficiency and/or the fraction of binders in the starting library.

In the context of KCE, the partitioning efficiency is largely defined by the nonbinder background, which is caused by the heterogeneity of the electrophoretic velocity of DNA in CE.⁵² This nonbinder background increases in orders of magnitude toward the free DNA peak (**Figure 1.2**, bottom panel).⁵⁰ Nonbinder backgrounds will differ for different target sizes (i.e., molecular weights of proteins) and, accordingly, different mobility shifts of DNA upon its binding to the target. Smaller targets (such as peptides and small molecules) experience lesser mobility shifts upon complexation, resulting in elevated levels of nonbinder background as well as inefficient partitioning. Therefore, in KCE, the way to minimize the level of nonbinder background and improve the efficiency of partitioning is to ensure an appreciable mobility shift of the complex from the free DNA regardless of the target sizes. Moreover, while KCE efficiency is almost ideal for binder selection (especially for relatively large target molecules, such as proteins), KCE has one inherent limitation: the low injection volume (e.g., less than a microliter). A typical input quantity of the initial library in KCE is about 10^{11} – 10^{12} molecules, which is 3 to 4 orders of magnitude lower than in surface-based methods.^{18,19,31,35} Despite the uniquely high partitioning efficiency of KCE, the low input quantity of the library can lead to failed selection for more challenging, less “aptagenic” targets, for which the fraction of binders in the initial library is low. Accordingly, efforts should be made to mitigate this limitation.

The application of KCE to selection of binders from oligonucleotide libraries began with the development of the NECEEM mode as a highly efficient separation method for aptamer selection. Between 2004 and 2005, complex-last and complex-first NECEEM modes were introduced to obtain high affinity aptamers for protein targets within only one to four rounds of selection by the Bowser group and Krylov group, respectively.^{31,32,35,36,56} Since then, there have been reports of successful aptamer selection by NECEEM for more than 30 targets of different classes (e.g., small molecules, peptide, proteins, and whole cell/bacteria) by numerous research groups, proving the effectiveness of the KCE approach.⁵⁷ In 2019, IFCE was introduced by the Krylov group as the 2nd mode of KCE in binder selection.⁴² So far, various strategies to improve the use of KCE selection of binders have been reported.^{37-44,58-61} These strategies can be classified as follows: (i) increasing the input quantity and/or binder abundance in the initial library, (ii) increasing the partitioning efficiency, and (iii) optimizing the post-partitioning process (PCR quality and sequencing analysis). It should be emphasized that nothing can compensate for the lack of binders in the initial library and/or insufficient partitioning; thus Strategies 1 and 2 are key to the development of robust methods for KCE-based selection of binders from oligonucleotide libraries. Once binders of sufficient quantity and purity are collected at the output of partitioning, Strategy 3 can help to minimize PCR biases and improve binder identification process; thus, contributing to the overall success of selection. In the next sections, we provide our analysis of the achievements in the field of KCE-based binder selection, from the pioneering works by the Bowser and Krylov groups to the later contributions from other research groups. We critically analyze all the proposed improvements from the original KCE selection procedures based on how their strategies can effectively increase the success rate of KCE-based selection of binders.

1.3. THE PREMIER WORKS ON KCE-BASED BINDER SELECTION

1.3.1. Work by the Bowser group

In 2004, Mendonsa and Bowser successfully introduced the use of KCE as an alternative partitioning method to conventional surface-based separation in SELEX and named this new approach CE-SELEX.³⁵ By using (in essence) the complex-last submode of NECEEM, they selected DNA aptamers with nanomolar affinities against IgE protein from a random DNA library in only four rounds of selection as opposed to 10 rounds in conventional surface-based SELEX for the same target.⁶² They also introduced the use of a “bulk” dissociation constant of the binder-enriched library after each round as a quantitative affinity parameter to monitor the progress of selection. In the following year, the Bowser group successfully obtained aptamers with picomolar affinity for HIV reverse transcriptase in 4 rounds of SELEX using the complex-first NECEEM submode.³⁶

Bowser’s team was the first to demonstrate the potential use of KCE to obtain aptamers for small targets such as peptides and small molecules. Aptamers against neuropeptide Y and small-molecule porphyrin targets with high-nanomolar to low-micromolar affinity were selected after three rounds of CE-SELEX using the complex-first submode of NECEEM.^{56,63} However, they noted the poor resolution between PL and L due to the small mobility shift of the DNA upon binding to small-sized targets and that many more iterative rounds would be required to eventually evolve a reasonably pure pool of high-affinity aptamers. The feasibility of the KCE technology in its application to selection of binders for small targets is still very limited and so far, there have been only a few reports of such binders found by KCE-based selection,^{56,63-66} some of these reports have not presented electropherograms and, therefore, can hardly be critically evaluated.

1.3.2. Work by the Krylov group

In early 2005, the first demonstration of complex-first NECEEM mode in SELEX was presented by the Krylov group.³¹ In this proof-of-principle work, high-affinity aptamers for protein farnesyltransferase with low-nanomolar affinity were selected in a single round for the 1st time. In the same year, this group introduced a new variation of the complex-first mode of NECEEM, namely equilibrium capillary electrophoresis of equilibrium mixtures (ECEEM).³² Conceptually, the only difference between NECEEM and ECEEM is the presence of the target protein in the running buffer for ECEEM which supports the state of dynamic quasi-equilibrium when L, P, and PL migrate through the capillary.

Krylov's group was the first to report the use of KCE to select smart binders with predefined binding parameters. In both complex-first NECEEM and ECEEM, different DNA sequences in the library have similar electrophoretic mobilities and migrate as a single electrophoretic zone while protein–aptamer complexes, which have mobilities different from free DNA, move faster and elute from the capillary earlier. The eluting fractions are collected in specific time windows to facilitate selection of aptamers with mathematically-predefined values of binding parameters (k_{off} and K_d). Since the extents of complex binding and unbinding differ in NECEEM and ECEEM, they have different accuracy of determining different binding parameters.³⁴ In NECEEM, complex unbinding prevails over complex binding events, making it more “sensitive” to k_{off} than k_{on} . In ECEEM, the quasi-equilibrium is maintained during the separation, making it more “sensitive” to K_d . Thus, NECEEM is used to select ligands with predefined k_{off} , while ECEEM allows selection of ligands with predefined K_d . **Figure 1.3** schematically illustrates how the range of k_{off} and K_d values of collected ligands depends on the position and width of the ligand-collection window in complex-first NECEEM and ECEEM respectively. Since the

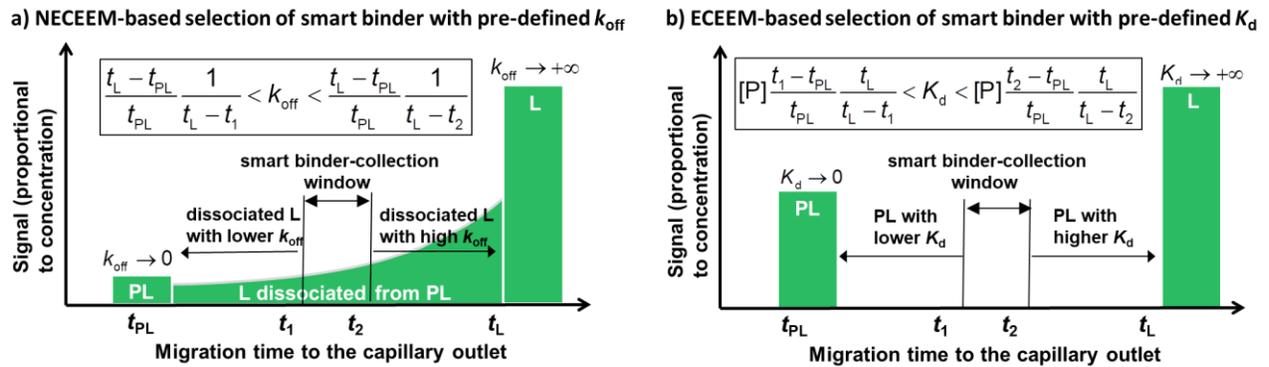


Figure 1.3. Schematic representation of selection of smart binders with **a)** predefined k_{off} by NECEEM and **b)** predefined K_d by ECEEM. Binders collected in the time window t_1-t_2 (smart binder-collection window) will have binding parameters defined by the equation in the figure.

nonbinder background is present in all KCE modes, obtaining smart binders might require more rounds of selection for narrowing the range of parameters around the desirable values. To prove the concept of KCE selection of smart binders, aptamers with pre-defined ranges of k_{off} and K_d were successfully selected for MutS protein after four rounds of SELEX based on complex-first NECEEM and ECEEM, respectively.³⁴

In 2006, the Krylov group introduced nonSELEX as an alternative to SELEX.^{26,27} As mentioned previously, nonSELEX involves repetitive steps of binding and partitioning without PCR amplification steps in between them; thus, removing the accumulated PCR biases during SELEX as well as significantly shortening the selection time. In the proof-of-principle work, DNA aptamers for h-Ras protein were obtained (with a 10^4 -fold improvement in the binding affinity as compared to the initial library) in three rounds of nonSELEX based on complex-first NECEEM. Since then, more DNA aptamers for protein targets have been found by means of KCE-based nonSELEX, proving the feasibility of the procedure.^{60,67,68} Moreover, nonSELEX is the only option for selection of binders from DEL because DEL cannot be amplified by PCR.⁶

Although nonSELEX seems to provide an excellent practical alternative to conventional SELEX, the lack of intermediate PCR amplification step limits the maximum number of rounds to three and requires more abundant population of binders in the initial library or highly-aptagenic targets such as DNA-binding proteins. This requirement is especially crucial for KCE-based nonSELEX due to the trivial loss of binders at every step of KCE partitioning. Fraction collection into a vial containing a CE running buffer leads to binder dilution and the following injection of only a small part of this diluted solution inevitably results to a loss of approximately 99% of binder per round.²⁶ Thus, the adaptation of KCE-based nonSELEX in selection of aptamers or DEL requires major efforts in Strategy 1 to compensate for the significant losses of binders between consecutive rounds of partitioning.

1.4. EFFORTS OF SCIENTIFIC COMMUNITY TO ADVANCE KCE SELECTION OF BINDERS

1.4.1. Increasing the input binder quantity (Strategy 1)

As mentioned above, an obvious limitation of KCE-based selection of binders is the small quantity of the injected initial library, which is about three orders of magnitude smaller than in surface-based selection (10^{12} molecules in KCE *versus* 10^{15} molecules in surface-based methods).^{18,19,31,35} To improve the success rate of KCE-based selection, several research groups have made significant efforts in Strategy 1: increasing the input quantity and/or binder abundance in the initial library. The most-straightforward approaches to Strategy 1 involve simply increasing the concentration of the initial library, the length of injected sample plug, and the diameter of the capillary. However, any obvious increase to the above parameters beyond the current standard setup would result in deterioration of peak shape, poor resolution, or significant

Joule heating.^{36,67,69} There were also efforts to increase the number of the fraction collection experiments for the first round of selection.^{66,70} However, this approach seems to be impractical since tens to hundreds of independent fraction collections would have to be conducted to essentially compensate for the low injection quantity in KCE-based selection.

A promising approach to Strategy 1 is the use of chemically modified oligonucleotide libraries which presumably contain more binders than non-modified DNA and RNA libraries. In 2013, the Kuwahara group succeeded in KCE-based aptamer selection using a xeno nucleic acids (XNAs) library.⁵⁸ In addition to their generally improved chemical and biological stability, XNA decorated with diverse chemical substituents (e.g., hydrophobic groups) can yield improved properties and functionalities such as new structural motifs and enhanced binding capabilities.⁷¹ Using complex-first NECEEM, chemically modified DNA aptamers for human α -thrombin were successfully obtained from an XNA library containing 2'-O, 4'-C-methylene-bridged/linked bicyclic ribonucleotides (2', 4'-BNA/LNA) in the primer region and C5-modified thymidine with N₆-ethyladenine in the randomized region. The selected modified aptamers showed several-fold improvement in binding affinity as well as biostability as compared to thrombin aptamers selected from unmodified library. Interestingly, these modified DNA aptamers did not contain classic G-quadruplex motifs, although the G4 structure is always observed for thrombin-binding aptamers selected from non-modified DNA libraries. Later, BNA/2'-deoxy-2'-fluoro-ribonucleic acid (FNA) chimeric aptamers for thrombin, exhibiting RNA-like conformations, were obtained from a library containing FNA in the randomized region.⁵⁹ Since thrombin is a commonly used model target protein in aptamer selection, it would be interesting to see the application of modified library in KCE selection for “difficult” protein targets which had repeatedly failed SELEX when using unmodified DNA libraries. These works by the Kuwahara group suggest the

feasibility for further development using more novel classes of modified DNA library, which will greatly benefit the area of KCE-based selection of binders.

Another option to increase the sampled quantity of the initial library is to couple KCE at later rounds with surface-based partitioning in the initial rounds. In 2015, Ashley Li and co-authors proposed a two-step method which incorporated partitioning on a nitrocellulose membrane filter followed by additional rounds of KCE-based partitioning.⁶⁰ In 2016, the Liu group presented a similar approach that combined boronate affinity magnetic nanoparticles and KCE-based selection.⁶¹ In this hybrid selection approach, a high input quantity of the initial library (e.g., 10^{15} unique molecules) is subjected to be screened in the first round of surface-based partitioning, thus appreciably reducing nonbinders sequences and diversity of the library before additional rounds of NECEEM-based selection. The Li and Liu groups demonstrated this new hybrid selection approach on the selection of aptamers for cholesterol esterase, ribonuclease and alkaline phosphatase and obtained aptamers with mid-nanomolar affinity. Pre-screening the library with surface-based partitioning is a viable option to produce an improved library with higher binder abundance for KCE-based selection; however, a comparative study is required in the future to evaluate the selection outcome between such hybrid selection approach and the traditional KCE-based selection. In general, the area of screening oligonucleotide libraries will remain more an art than a science until quantitative comparative studies become standard in this challenging area.

1.4.2. Increasing the partitioning efficiency (Strategy 2)

The partitioning efficiency of KCE is limited by the nonbinder background: the result of irregularities in electrophoretic mobility of oligonucleotides.⁵² The smaller the mobility shift of

target–oligonucleotide complex from that of free DNA, the higher the irregular oligonucleotide fraction and, accordingly, the nonbinder background.⁵⁰ Targeting Strategy 2 (i.e., increasing partitioning efficiency) requires a larger mobility shift or, in the other words, improved temporal resolution between target-oligonucleotide complexes and nonbinders, followed by a quantitative assessment of the nonbinder background. Since the nonbinder background is under the limit of detection of any optical system, its measurement requires collection of multiple ~1 min fractions and quantification of oligonucleotide in every fraction by qPCR to build a “oligonucleotide quantity versus migration time” electropherogram.⁵² Unfortunately, most of the reports dedicated to optimizing Strategy 2 were not supported by any measurements of the nonbinder background (with the exception of the 2018 IFCE-based selection, which will be discussed in details in **Chapter 2**). Despite having this limitation, these works still constitute a good starting point for any future efforts in improving partitioning efficiency in KCE.

The first illustration of an “ideal” filter in CE, where the complex and DNA migrate oppositely was conducted by the Qu group in 2014.⁷² Using CE running buffer at pH 2.6 in an uncoated capillary, they proposed low pH CE-SELEX (LpH-CE-SELEX), in which the EOF is suppressed to drive the migration of complex and DNA to the opposite directions. In this proof-of-concept study, the low-pH partitioning was demonstrated for three different mixtures of model proteins (transferrin, bovine serum albumin, and cytochrome c) and a DNA library. Although the idea was novel, the proposed low-pH selection approach is fundamentally problematic, and the validity of the presented data is questionable. The medium at such low pH values drastically affects the stability of both DNA and protein as well as impedes their intermolecular interactions. Moreover, any complex formation will mostly be driven by the nonspecific electrostatic interaction between negatively charged DNA and positively charged

protein at pH 2.6. In the absence of basic understanding of the fundamentals, there is no room for further discussion on the feasibility of LpH-CE-SELEX and/or its contribution to the area of KCE-based selection.

In 2015, a combined effort by the Colyer and Bonin groups led to the proposal of using capillary transient isotachopheresis (ctITP) to enhance the partitioning efficiency of KCE-based selection of aptamers by concentrating both the complex peak and the free DNA peak into narrower zones.³⁷ The authors of this work conducted a mocked selection for human α -thrombin from a random DNA library containing a known amount of thrombin binding aptamer using the proposed procedure of ctITP-NECEEM. After a single round of ctITP-NECEEM, HTS data showed a 40-fold enrichment in the quantity of thrombin-binding aptamer.

In 2016, the Saito group utilized the same approach to develop polymer-enhanced ctITP (PectI) *via* adding polyethyleneoxide (PEO) to both CE running and sample buffer. The procedure was reported to facilitate single-round selection of DNA aptamers for *E. coli*, *S. cerevisiae*, and human lung cancer cell line.^{39,40}

The interpretation of the high efficiency of partitioning, which facilitated single-round selection, assumed that ctITP improved the resolution *via* ITP-based focusing of both the complex zone and the DNA zone. We view this interpretation with scepticism for the following reasons. Firstly, none of the above reports provided any evidence of similar focusing of the zone of target-oligonucleotide complexes and that of unbound DNA. ITP is not a universal focusing technique and it is rather unlikely that species with very different electrophoretic mobilities will focus similarly.⁷³ Secondly, even if both the complex and unbound DNA focused similarly under the chosen ctITP conditions, the nonbinder background, which has the same mobility as the complex, would be focused together with the complex into a narrow zone resulting in no gain in

the efficiency of partitioning. Thirdly, in KCE-based aptamer selection for cells, enhancing resolution by ITP might be irrelevant since any complex between DNA and a cell (a large target) must already experience a significant mobility shift from the free DNA. Lastly, the mock-selection data for one round of ctITP showed only a 40-fold enrichment for the binder sequence while a typical NECEEM round has the partitioning efficiency of approximately 10^5 . Clearly, a 40-fold enrichment cannot support the successful single-round selection claim. The proposed ITP approach neither reduced the nonbinder background nor increased the peak resolution; therefore, no improvement in partitioning efficiency can be expected. In our opinion, the use of ITP or other stacking techniques can only be beneficial if the separation is insufficient and the resolution is poor, e.g., in the case of small molecule targets.

To improve the partitioning efficiency of KCE-based selection for a small-size target, the Yoshimoto group merged bead-based partitioning and NECEEM in a process called microbead-assisted capillary electrophoresis (MACE)-SELEX.⁴¹ During MACE, an equilibrium mixture of target-coupled microbeads and DNA library is directly introduced into a capillary and subjected to NECEEM. The binding of the target-coupled microbead to DNA library results in a large mobility shift of the complex from the free DNA, effectively eluting the complex together with the unbound target-coupled microbeads. Thus, MACE-SELEX is applicable to any type of bead-bound targets whereas KCE-SELEX is limited to only targets that exhibit a significant mobility shift upon binding to DNA. In the proof-of-principle work, thrombin binding aptamer with low nanomolar affinity was successfully obtained in three rounds of MACE-SELEX.⁴¹ To eliminate any nonspecific adhesion of the library to the bead surface, negatively charged beads possessing carboxylic acid groups were used and the PCR product of the non-specifically adsorbed amount of DNA on the thrombin-free beads was virtually undetectable. In principle, introducing a

microbead as a drag tag can improve the efficiency of partitioning for KCE selection of aptamers for small-size targets. However, it appears to be a contradictory approach to the immobilization-free feature of KCE. In MACE-SELEX, surface immobilization of such small-size target might result in more pronounced steric hindrance effects, decreasing the affinity between the target molecule and the library and ultimately lowering the success rate of selection. As a result, there is a trade-off between immobilization-free characteristic of KCE-SELEX and the bead-induced mobility shift enhancement of MACE-SELEX. To demonstrate the advantages of MACE-SELEX over KCE-SELEX more convincingly, one needs to apply MACE-SELEX to a small molecule target which is unapproachable by NECEEM-SELEX owing to the lack of mobility shift upon oligonucleotide complexation to such a target. Conveniently, Yoshimoto and co-authors had initiated this task by conducting 3 rounds of MACE-SELEX for the small molecule drug methotrexate (454 Da) and obtaining DNA aptamers with the best K_d value of 570 nM.⁷⁴ Despite having some inherent limitations for small molecule targets, MACE-SELEX still contributes to facilitate the use of KCE technique to a much broader range of target sizes. I hope that MACE-SELEX will continue to demonstrate its applicability to more small-molecule targets, which could make MACE-SELEX a highly used technique.

1.4.3. Improving the post-partitioning process (Strategy 3)

The optimization of post-partitioning processes (e.g., the analysis of PCR quality and sequencing) also plays an important role in improving the performance of KCE-based screening of oligonucleotide libraries. The classic PCR amplification protocol (used for amplification of homogeneous DNA samples) was found to rapidly generate byproducts and preferentially amplify nonbinder sequences when applied to highly-heterogeneous DNA samples.^{54,75} To

address this issue, in 2016, Yufa *et al.* proposed the use of emulsion PCR (ePCR), which is known to reduce byproduct formation as well amplification biases, as a PCR mode for coupling with KCE-based SELEX.⁷⁶ The effectiveness of ePCR was demonstrated by successful aptamer selection for the DNA repair enzyme ABH2, for which aptamers could not be obtained by NECEEM-based SELEX using conventional PCR. The results of this work emphasize the importance of having an optimized PCR procedure in binder selection and the necessity of monitoring the quality of PCR products throughout the entire selection process.

Sequencing of the selected nucleotide binders is an important stage in the screening process. Originally, the enriched binder pools obtained by KCE-based selection were cloned into a plasmid, and only a few (typically tens) individual clones were sequenced and tested for affinity to the target in the binder-verification stage.^{31,32,35,36,56} Later, the applicability of the HTS technology in the identification of binders from the binder-enriched libraries after KCE-based selection has been evaluated by some research groups. In 2013, Jing and Bowser explored the use of HTS to analyze the evolution process of KCE-based selection.⁷⁰ The study revealed the unique characteristics of the binder-enriched libraries obtained in CE-based selection: (i) the library remains highly heterogenous after four rounds of selection, (ii) there is no prevailing motifs, and (iii) there is no statistically-significant difference in affinity between randomly chosen sequences from the binder-enriched libraries. This work proves that KCE-based selection produces uniquely heterogeneous pools of high-affinity aptamers, providing more options for the secondary screening stage beyond the several binding motifs typically obtained in surface-based selection.

Since the enriched sequences or highly abundant nucleotide motifs are rarely observed in KCE-SELEX, identifying binders from the massive HTS sequences data become very

challenging. In 2015, the Whelan group performed NECEEM-SELEX/HTS approach to select aptamers for the ovarian cancer biomarker HE4.⁷⁷ After five rounds of selection, the HTS sequencing data was analyzed using a bioinformatics pipeline. The aptamer sequences with quite high K_d values ranging from 300 to 780 nM were discovered by clustering the top 1,000 most enriched sequences using the CD-HIT-EST program. In 2016, the Gmeiner group employed HTS to identify sequences in a binder-enriched library obtained after three rounds of ctITP-NECEEM-based aptamer selection for vitronectin protein.⁷⁸ Seven top-populous sequences were first chosen out of 143,845 unique sequences and used an 8-base reading frames to determine all of the aptamers that contained any of these reading frames from the data. These sequences were clustered into families using the Clustal software. A phylogenetic tree was then used to differentiate the families, compared top ten most abundant sequences from these groups, and finally selected an aptamer with $K_d = 405$ nM. Recently, more computer-assisted tools to analyze HTS data from selection, such as FASTAptamer, AptaCluster, and AptaTrace became available.⁷⁹⁻⁸¹ Together, they provide tools to potentially establish a more universal algorithm and metrics for selecting binders from large HTS datasets.

1.4.4. Additional tools to assist KCE selection of binders

There are a few reports where the authors aimed to enhance the automaticity and convenience of KCE-based aptamer-selection procedures. Although these reports do not contribute directly to the success rate of the selection, they can potentially be useful additional tools to assist the users if utilized properly.

In 2013, the Dovichi group reported a flexible and low-cost automated fraction collection system for KCE-based selection.⁸² This system was applied to the selection of DNA aptamers

against thrombin by collecting fractions into wells on a 96-well microtiter plate at a 4-s time interval. To simplify subsequent analysis, qPCR reagents were employed as the sheath liquid of the fraction collector. After fraction collection, the 96-well plate was directly analysed by qPCR to search for the presence of any complex formation in the collected fractions. The pooled contents of the chosen wells were then sequenced to identify the binder sequences. In 2015, the Zhang group undertook a very similar approach, called FCE-SELEX, in which the expected zone of complex was collected in 12 fractions containing oil-sealed qPCR mixture at 30-s time interval.⁴³ The authors claimed that FCE-SELEX can facilitate selection of aptamer in a single round. This is an unjustified statement since multiple-fraction collection approach can only assist in identifying fractions with higher output quantity of DNA and, thus, will not compensate for the lack of binders or insufficient efficiency of partitioning.

With the goal of improving the procedure of KCE-based aptamer selection, the Qu group proposed an online reaction-based, single-step CE-SELEX, called ssCE-SELEX using human α -thrombin as a model protein target.⁴⁴ In ssCE-SELEX, the entire procedures of KCE-based partitioning (sample mixing, incubation, reacting, separation, detection, and collection) are performed in a single CE run by on-capillary mixing of library and protein targets. By using ssCE-SELEX, aptamers for human α -thrombin and bovine lactoferrin were obtained within only two rounds with the best K_d values of 56 and 20 nM, respectively; leading to the conclusion that the partitioning efficiency of ssCE-SELEX was higher than that of regular KCE-SELEX. Clearly, there is no ground for this conclusion while there are justified benefits of using an online mode of KCE-based selection: (i) lower sample consumption, (ii) easier handling, and, (iii) decrease in experimental time.

The Krylov group introduced mobility predictor tools to facilitate the rational design of KCE-based binder selection; such a design requires the knowledge of the time at which the complex exits the capillary.^{49,83} In many cases, due to the low binder abundance in the initial library, finding this time experimentally is very difficult because the concentration of complex might be below the limit of detection, even for highly sensitive fluorescence detection. To resolve this dilemma, mathematical models have been developed for predicting the mobility of protein–aptamer complexes and protein–DEL complexes with less than 5% and 11% deviations from the experimentally determined values, respectively. The *a priori* knowledge of complex mobility will allow accurate blind setting of the time window for complex collection. Although having a mobility predictor does not contribute directly to improving the selection performance, it will serve as an indispensable tool in rational selection of the fractions to be blindly collected in KCE-based selection.

1.4.5. Challenges in expanding KCE adoption within the scientific community and my project goal

Between the premier work of KCE-based selection and the commencement of my research project, the community had made significant contributions aimed at all the aforementioned strategies to further advance the practicality of KCE methods. Despite these efforts, KCE-based selection has still remained as more of an art form rather than a robust and practical tool in the field of ligand selection. It is evident that KCE-based selections have solely been conducted within specialized academic laboratories, with no successful integration observed in clinical or industrial settings. The reluctance of scientific community toward greater use of KCE-based selection can be attributed to the following practical challenges: (i) the limited partitioning

efficiency of NECEEM platform, (ii) the restricted use of highly conductive physiological buffers and (iii) the absence of an instructive formalism to assist users in determining optimal KCE selection conditions tailored to different protein targets.

At the start of my research project, the consensus held that the partitioning efficiency of KCE was limited to approximately 10^5 , primarily due to the exclusive utilization of NECEEM as the mode of KCE in binder selection during that period. In comparison to the conventional bead-based approach, the development of NECEEM resulted in a notable improvement of partitioning efficiency by 2 to 3 orders of magnitude and reduced the required selection rounds from over 10 to just 4.³¹ However, despite being significant, this improvement is worse than anticipated: theoretical prediction suggested that NECEEM-based partitioning efficiency would approach infinity and successful single-round SELEX could be achieved as long as there is sufficient quantity of binders at the input of partitioning.⁵¹ As mentioned earlier, a study conducted by the Krylov lab in 2013 revealed that the higher-than-expected nonbinder background in NECEEM was due to the heterogeneity in electrophoretic mobility of oligonucleotides. This phenomenon led to an underwhelming partitioning efficiency value in NECEEM.⁵² Thus, our anticipation is that the elimination of this nonbinder background in KCE could transform it into a truly invaluable technology, increasing its potential to replace the conventional bead-based method. This is particularly relevant in the context of DEL-based drug discovery platforms, which rely on a nonSELEX approach.⁶ In such a platform, achieving an exceptionally high partitioning efficiency with minimal to no background is crucial to attain a pool primarily consisting of binders, all within the constraint of a maximum of 3 rounds.

Another long-standing problem that hinders the adoption of KCE-based selection is its limited compatibility with highly conductive physiological solutions.^{31,32,35-44,56,58-61} Typical

NECEEM experiments are constrained to a narrow set of low-ionic strength running buffers due to their relatively small degree of Joule heating and slow time of ion depletion. The restricted variety of compatible running buffers in NECEEM-based selection is a major drawback, as this could yield results that inaccurately mirror the affinity interactions taking place in physiological fluid characterized by high ionic strength. Thus, by enabling the use of physiological buffer in KCE-based selection, the practical relevance of its outcomes would be notably enhanced, rendering the selected binders far more pertinent for potential *in vivo* applications.

Last but not least, the slow rate of adoption of KCE in the selection community has to do with the lack of a framework for proper usage and optimization of KCE for different protein targets. Each mode of KCE is characterized by a set of experimental parameters, which include, but are not limited to, ionic strength and pH of the running buffer, strength of the applied electric field, length and diameter of the capillary, temperature inside the capillary and coating of the inner capillary wall. For a given set of experimental parameters in a particular KCE mode, the partitioning efficiencies over a wide range of protein target sizes will greatly vary. It means that the chosen conditions in previously successful KCE selection for one protein target might be detrimental for another protein of different size. On the other hand, optimizing KCE conditions for a given protein target is not an easy task, due to the complex interrelation between various experimental factors and their counteracting effects on the produced results (e.g., higher ionic strength of the running buffer and longer capillary leads to better temporal resolution but lengthy separation time and more complex dissociation). To ensure that KCE-based selection is effectively utilized in a broader community, the influence of different experimental conditions on the outcome of KCE selection must be examined for a wide range of protein target sizes. Such study is pivotal to the rational planning and optimization of KCE-based selection.

The purpose of my research project is to resolve the aforementioned practical challenges, with the ultimate goal of transforming KCE into a truly disruptive technology for screening protein binders from oligonucleotide libraries, particularly in the context of DEL-based screening platforms, where KCE methods have not been previously utilized. To prepare KCE for successful implementation in DEL platform, I outline the following key objectives: firstly, mitigating the restricted partitioning efficiency of KCE-based selection by introducing a novel KCE-based partitioning mode devoid of nonbinder background; secondly, enhancing the practical relevance of KCE-selected binders by making KCE-based selection compatible with physiological buffers; and thirdly, improving the overall applicability of KCE-based selection by developing a quantitative characterization of all KCE partitioning modes for a diverse range of protein target sizes. The progress made towards accomplishing all three objectives is described in **Chapter 2** and **3**. In **Chapter 2**, I introduce a new mode of KCE partitioning, in which target–binder complexes and nonbinders move in the opposite direction. This unique mode of KCE, thus, would function as a real filter, giving it a name of IFCE.⁴² By reducing the nonbinder background to under the limit of detection in qPCR, the efficiency of partitioning in IFCE is improved by three orders of magnitude as compared to that of NECEEM, reaching 10^9 – the highest value reported so far. Advantageously, the condition of IFCE is achieved when the running buffer becomes more physiological, providing greater biological relevance of selected protein binders. To address the Joule heating issue associated with the use of physiological running buffer in IFCE, I utilize a simplified universal method for predicting the electrolyte temperatures (SUMET) to rationally lower the electric field so that the in-capillary temperature is kept under a reasonable value during the separation.⁸⁴ I validate the newly developed IFCE platform by conducting a successful one-step selection of a high-affinity aptamer pool to a target

protein. Moreover, I also propose a double-passage KCE approach for finding K_d and k_{off} of protein binders selected under the physiological conditions of IFCE, and validate the approach using an interacting pair of a protein and its previously selected aptamer.⁸⁵ Moving on to **Chapter 3**, I describe a quantitative study investigating the impact of specific critical experimental conditions (e.g., composition of running buffer and coating of inner capillary wall) on the efficacy of different modes of KCE partitioning.⁵⁰ The study spans a diverse range of protein target sizes. A comprehensive guideline is formulated based on this study, enabling the identification of appropriate KCE-based partitioning conditions for a given protein target. This guideline takes into consideration the complex interrelation between numerous experimental parameters, allowing researchers to approach the process in a systematic and rational manner. The combined findings from **Chapters 2** and **3** successfully fulfill my threefold objectives of establishing KCE-based selection as universally applicable and suitable for integration into DEL-based screening platforms. Prior to transitioning into the utilization of KCE in the DEL platform, I explore a pivotal variable known to exert a significant influence on the process of binder selection: the total concentration of the target protein. In **Chapter 4**, I introduce the first experimental investigation into the correlation between target concentration and the selection of protein binders from oligonucleotide libraries. To facilitate the selection process, I suggest employing a straightforward experimental parameter that can be readily determined. This parameter serves as a simple tool for researchers to identify an appropriate range of target concentration to avoid selection failure. Finally, in **Chapter 5**, I integrate the insights garnered from all our previous works to facilitate the application of KCE in selection of protein binders from DEL. By utilizing the KCE and target concentration optimization framework, I present the first experimental demonstration of high-efficiency KCE-based selection of small molecule

binders for protein target from a hypothetical DEL library — this being the ultimate goal of my project. This pioneering achievement showcases the capabilities of KCE-based partitioning within the context of DEL-based drug discovery, marking a significant advancement in the field.

**CHAPTER 2. DEVELOPMENT OF IDEAL-FILTER CAPILLARY
ELECTROPHORESIS (IFCE): A HIGHLY EFFICIENT PARTITIONING METHOD
FOR SELECTION OF PROTEIN BINDERS FROM OLIGONUCLEOTIDE LIBRARIES**

2.1. THE PROOF OF PRINCIPLE WORK

The presented material was published previously and reprinted with permission from “**Le, A.T.H.**; Krylova, S.M.; Kanoatov, M.; Desai, S.; Krylov, S.N. Ideal-filter capillary electrophoresis (IFCE) facilitates the one-step selection of aptamers. *Angewandte Chemie International Edition* **2019**, 58, 14610 – 14615”. Copyright 2019 John Wiley and Sons. My contribution to the article was: (i) planning all experiments, (ii) performing all experiments, (iii) interpreting results, (iv) preparing figures and (v) writing the manuscript.

2.1.1. The significance of single-round selection of binders from oligonucleotide libraries

As previously discussed in **Section 1.1.**, efficient physical partitioning of binders from nonbinders constitutes a major challenge for the field of binder selection from oligonucleotide libraries. The abundance of binders in a naive (unbiased for a specific target) oligonucleotide library may be as low as one suitable binder per million or billion library molecules. Separating the few binders from the overwhelming majority of nonbinders in a single partitioning step is limited by a nonbinder background.^{15,20-22} Consequently, multiple rounds of partitioning are typically used for in vitro selection of binders from oligonucleotide libraries.

Multiround selection of binders from oligonucleotide libraries has fundamental drawbacks. Oligonucleotide aptamers are typically obtained by the SELEX process, which involves iterative cycles of binding/partitioning and enrichment by PCR amplification and theoretically enables an

unlimited number of rounds.²⁰⁻²² Remarkably, SELEX fails to select binders in 70% of attempts.³⁰ This multi-round procedure is prone to failure inherently because PCR preferentially amplifies nonbinders, which are less structured oligonucleotides than binders and are, hence, more easily accessible to polymerases.^{54,75} As a result, a large number of rounds may not be able to compensate for inefficient partitioning if the sequence bias of PCR toward the amplification of nonbinders is greater than the preference of partitioning for the isolation of binders.

The limited efficiency of partitioning is even more detrimental in the isolation of binders from DELs. Binders isolated from DELs cannot be amplified by PCR, while every round of partitioning results in binder losses.⁶ Thus, a typical binder selection from a highly diverse DEL is limited to three rounds and produces a pool greatly dominated by nonbinders. Secondary screening or the hit verification stage, which involves deep DNA sequencing of the enriched pool followed by *in silico* removal of known nonbinders (e.g., binders to the matrix), synthesis of selected DNA-free compounds, and testing them in binding assays, is used as a means of compensating for the inefficiency of partitioning.^{6,28,29,86} Sequencing large amounts of DNA is expensive, and the fact that a significant part of the material processed in DNA sequencing is nonbinders makes selection from highly diverse DELs cost inefficient for industrial R&D and largely cost prohibitive for academic laboratories.

An ultimate solution for problems originating from multi-round selection would be a partition method that could enrich binders to the required level (e.g., 99% of binders in the binder-enriched library) in one step of partitioning. There have been several reports of one-step selection of oligonucleotide aptamer; however, they have not been independently confirmed raising doubts in method practicality, transferability, and reliability.^{31,87-92} We think there are two major reasons for slow progress in creating a practical way of single-step selection of binders

from oligonucleotide libraries. The first reason is technological: it is extremely difficult to achieve high efficiencies of partitioning, e.g., due to adsorption of nonbinders to surfaces in solid-phase selection methods,^{15,20-22} and non-uniform migration of the nonbinders in homogeneous capillary electrophoresis-based methods.⁵² The second reason is methodological: while high efficiencies of partitioning are the goal, the efficiency of partitioning was not used to guide developments or substantiate claims of one-step selection; moreover, it was rarely measured.⁸⁶⁻⁹¹

In this work, we address both the technological and methodological issues. First, we estimate theoretically the efficiency of partitioning required for one-step selection to be roughly 10^9 , which is at least a million times higher than efficiencies of typical solid-phase methods of partitioning, and ten thousand times higher than the efficiency of classical KCE-based selection methods. Second, we develop a new partitioning method, Ideal-Filter Capillary Electrophoresis (IFCE), in which target–binder complexes and nonbinders move in the opposite directions inside the capillary. Third, the efficiency of partitioning was adjusted by changing the ionic strength of the running buffer (I_{RB}) to reach the required value of 10^9 . Finally, we demonstrate one-step selection of a high-affinity aptamer pool to a protein target (MutS protein) to confirm that the estimated efficiency of 10^9 required for one-step selection is correct. Our work anticipates that IFCE partitioning holds great promise as an essential tool for the rapid and robust selection of binders from oligonucleotide libraries.

2.1.2. Materials and methods

2.1.2.1. Materials and solutions

All chemicals were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries with inner and outer diameters of 75 and 360 μm , respectively, were purchased from Molex Polymicro (Phoenix, AZ, USA). Recombinant *Thermus aquaticus* MutS protein (MW \approx 90 kDa, pI 6.0) was expressed and purified as described previously.⁹³ All DNA molecules were custom synthesized by Integrated DNA Technologies (Coralville, IA, USA). Bodipy (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) was purchased from Life Technologies Inc. (Burlington, ON, Canada). The running buffer (RB) was 50 mM Tris-HCl with NaCl ranging from 0 to 150 mM at pH 7.0. The sample buffer was always identical to RB to prevent adverse effects of buffer mismatch. Accordingly, all dilutions of sample components used in CE experiments were done by adding the corresponding RB.

The DNA aptamer with affinity toward MutS protein was previously selected in our laboratory (clone 2-06), and its fluorescein amidite (FAM)-labeled version was used here: 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'.³⁴ For aptamer-selection study, a synthetic FAM-labeled DNA library (N40) with a 40-nt random region was used: 5'-FAM-AGC CTA ACG CAG AAC AAT GG-(N40)-CGA TGC CAG GTT AAA GCA CT-3'. The following primers were used for PCR amplification of the MutS aptamer: forward primer (MutS_uF), 5'-CTT CTG CCC GCC TCC TTC C-3'; reverse primer (MutS_uR), 5'-AGT GTC CGC CTA TCT CGT CTC C-3'. Two sets of primers were used to amplify binders selected from the naïve library. The first set of primers was an unlabelled forward primer (N40_uF), 5'-AGC CTA ACG CAG AAC AAT GG-3', and an unlabeled reverse primer (N40_uR), 5'-AGT GCT TTA ACC

TGG CAT CG-3'. The second set contained a FAM-labeled forward primer (N40-famF), 5'-FAM-AGC CTA ACG CAG AAC AAT GG-3', and a biotin-labeled reverse primer (N40-biotinR), 5'-Biotin-TEG-AGT GCT TTA ACC TGG CAT CG-3'.

2.1.2.2. CE instrumentation and fraction collection conditions

All CE experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detection system. Fluorescence was excited with a blue line (488 nm) of a solid-state laser and detected at 520 nm using a spectrally-optimized emission filter system. Uncoated fused-silica capillaries, with a total length of 50 cm and a 10.2-cm distance from one of the ends to the detection zone were used. The two capillary ends were used as inlets interchangeably in experiments requiring different separation distances. The separation distance was defined as the distance in the capillary from the sample position at the time of electric field application to the detection point. Prior to every run, the capillary was rinsed successively with 0.1 M HCl, 0.1 M NaOH, deionized H₂O, and a run buffer at 20 psi (138 kPa) for 3 min each. The sample contained 10 μM annealed oligonucleotides (melted at 90 °C for 2 min and gradually cooled down to 20 °C at a rate of 0.5 °C/sec) and 150 nM Bodipy (as EOF marker). When specified, the sample also contained 100 nM MutS protein. The sample mixture was incubated for 30 min at a room temperature (22–24 °C) and then injected with a pressure pulse of 0.5 psi (3.5 kPa) × 10 s to yield a 10 mm long sample plug. The injected sample plug was propagated through the uncooled part of the capillary at the inlet by injecting a 5.7 cm long plug of the RB with a pressure pulse of 0.3 (2.2 kPa) psi × 90 s. CE was carried out at an electric field of 200 V/cm (10 kV over 50 cm). The duration of electrophoretic runs without fraction collection was 50 min. The duration of electrophoretic runs with fraction

collection was 64 min. Collection vials contained 20 μ L of the RB each and were switched every 2 min of voltage applied; 32 fractions were collected for each run.

2.1.2.3. Quantitative PCR

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

2.1.2.4. Specifics of single-round aptamer selection

Fraction collection and qPCR detection were similar to the procedures described in the previous two sections with a few modifications specified below. The equilibrium mixture contained: 10 μ M N40 library, 100 nM MutS protein, and 150 nM Bodipy. For qPCR, 1 \times SYBR Green Supermix and the unlabeled primers for N40 library (100 nM of each N40_uF and N40_uR) constituted the qPCR reagent mixture. The fraction which eluted at minute 29 and contained the highest amount of MutS–DNA complexes was subjected to preparative PCR. The procedure of preparative PCR involved two rounds of amplification. In the first round, the fraction was amplified by qPCR in quintuplicates as previously described. An S-shaped

amplification curve was plotted, and the PCR product was removed two cycles into the exponential phase of the curve. After qPCR, 100 μL of the five combined PCR reactions was purified using MinElute[®] PCR purification kit (QIAGEN, Mississauga, ON, Canada) as per manufacturer's instructions. DNA was then eluted using 20 μL of 50 mM TrisHCl, pH 7.0. Once product's purity was verified by native PAGE, it was subjected to asymmetric PCR. Five μL of DNA was added to 45 μL of asymmetric PCR reagent mixture from New England Biolabs Inc. (MA, USA). Final concentrations of PCR reagents in the reaction mixture were: 1 \times Standard *Taq* Reaction Buffer, 1 μM N40-famF, 50 nM N40-biotinR, 2.5 units/ μL *Taq* DNA Polymerase, and 200 μM dNTPs mix. The reaction was performed in duplicates with the following temperature protocol: 94 $^{\circ}\text{C}$ for 30 s (initial denaturation, performed once), 94 $^{\circ}\text{C}$ for 10 s (denaturation), 56 $^{\circ}\text{C}$ for 10 s (annealing), and 72 $^{\circ}\text{C}$ for 10 s (extension). Seventeen cycles of asymmetric PCR were run. Ten μL of MagnaBind[™] streptavidin beads suspension (Thermo Scientific, IL, USA) was washed three times and resuspended in bead washing/binding buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA pH 8.0). Once amplified, the duplicate PCR reactions were combined and incubated with streptavidin magnetic beads for 30 min at room temperature. The beads were magnetized, discarded, and the PCR product was then purified using the MinElute[®] PCR purification kit. The final product was eluted using 20 μL of 50 mM TrisHCl, pH 7.0, and 2 μL of 1 M NaCl was added to bring NaCl concentration to 100 mM.

To determine DNA concentration in the enriched library pool, serial dilutions of N40-famF (2 μM , 1 μM , 500 nM, 250 nM, 125 nM, 62.5 nM, and 31.25 nM) were prepared to build a standard curve by measuring fluorescence intensity at 519 nM with NanoDrop 3300 Fluorospectrometer (Thermo Scientific, IL, USA). DNA concentration in the enriched library pool was found to be 1.2 μM .

For a pressure-aided IFCE-based binding test of the enriched pool (used to determine $K_{d,app}$), a 47 nL plug of the equilibrium mixture containing 20 nM enriched library and 100 nM MutS in 50 mM Tris-HCl, 100 mM NaCl, pH 7.0, (default running buffer for IFCE) was injected into 50-cm-long capillary by a 0.5 psi (3.5 kPa) \times 10 s pressure pulse. The sample mixture was propagated through the non-cooled portion of the capillary by injecting a 5.7-cm-long plug of RB with a pressure pulse of 0.3 psi (2.2 kPa) \times 90 s. CE was carried out at an electric field of +200 V/cm at the inlet. In addition to applying voltage, a pressure of 0.20 psi (1.4 kPa) was applied to the capillary inlet to supplement the electric field and ensure that the nonbinders reach the detector. The pressure-aided IFCE allowed detection of target–binder complexes and nonbinders, which is required for determination of $K_{d,app}$.⁹⁴

2.1.3. Results and discussion

2.1.3.1. Efficiency of partitioning required for one-step selection of binders

If the library contains molecules capable of binding the target protein under the chosen selection conditions, then the equilibrium mixture of the library and the target contains target-bound library molecules (binders) and target-unbound library molecules (nonbinders). The selection scheme is depicted in **Figure 2.1**. The purpose of partitioning in binder selection is to separate binders from nonbinders physically. Thus, the partitioning process is conceptually filtration that should let binders through and cut nonbinders. To describe partitioning quantitatively, we adopt a term of “transmittance” which is used for optical filters. Transmittance of partitioning for binders, k_B , is the ratio between quantities of binder molecules at the output, B_{out} , and input, B_{in} , of partitioning, respectively:

$$k_B = B_{out} / B_{in} \quad (3)$$

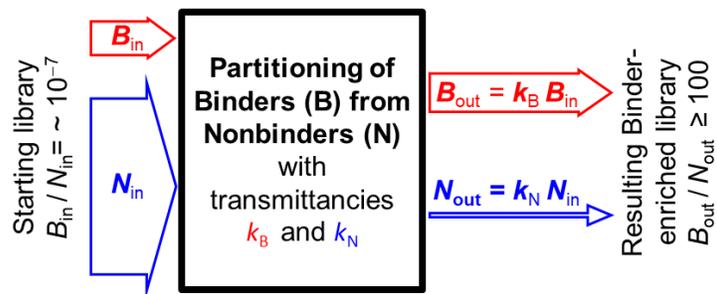


Figure 2.1. Schematic depiction of one-step partitioning of binders (B) from nonbinders (N) in a mixture of oligonucleotide library and protein target. See text for details.

Transmittance of partitioning for nonbinders, k_N , is the ratio between the quantities of nonbinder molecules at the output, N_{out} , and input, N_{in} , of partitioning, respectively:

$$k_N = N_{out} / N_{in} \quad (4)$$

Note that k_N is a fraction of nonbinders that penetrates through the step of partitioning; this fraction contaminates binders and creates nonbinder background in the selection process. Thus, k_N is a quantitative measure of nonbinder background.

If we chose $B_{out} / N_{out} \geq 100$ as a criterion of completed selection (a criterion of $B_{out} / N_{out} \geq 1$ is typically considered acceptable), then one step of partitioning is sufficient for completing selection when the efficiency of partitioning (defined as k_B / k_N) relates to the starting binder abundance (B_{in} / N_{in}) as:

$$k_B / k_N \geq 100 / (B_{in} / N_{in}) \quad (5)$$

Values of B_{in} / N_{in} are hard to estimate in SELEX. Our estimate *via* binder selection from a random-sequence DNA library in three consecutive rounds of partitioning without PCR amplification between them gave $B_{in} / N_{in} \sim 10^{-7}$.²⁶ According to **Eq. 5**, this estimate suggests $k_B / k_N \sim 10^9$ as an efficiency of partitioning which should suffice binder selection in one step. Assuming that $k_B \approx 1$ (which is typically satisfied), we can conclude that reaching $k_N \sim 10^{-9}$ is

sufficient for one-step selection. Further, we use this value as a guide in our development of a novel KCE-based partitioning method suitable for one-step selection of binders from oligonucleotide libraries.

2.1.3.2. Hypothesis for reducing nonbinder background in KCE-based partitioning

Partitioning by KCE is based on free-flow separation of protein target–binder complexes from nonbinders in an electric field. The main reason for nonbinder background in KCE-based partitioning is nonuniform migration of oligonucleotides: there is always a small part of nonbinders that tails towards binders and creates nonbinder background in the binder-collection time window.⁵² Two known practical modes of KCE-based partitioning differ by voltage polarity (namely, complex-first and complex-last NECEEM), but in both of them, nonbinders (denoted as unbound DNA) and protein target–binder complexes (denoted as P–DNA) move in the same direction (**Figure 2.2a, b**).^{31,35} These modes do not operate as a physical filter which is supposed to let binders through but reject nonbinders. We hypothesized that the k_N in KCE-based partitioning could be decreased if the target–binder complexes and nonbinders moved in the opposite directions (**Figure 2.2c**) making KCE function as a physical filter (but without the issue of non-specific adsorption inherent to real filters) and giving this mode of KCE a name of Ideal-Filter CE (IFCE).

IFCE is equivalent to the following relation between the velocity of nonbinders, v_{DNA} , and that of binders, v_{P-DNA} : $v_{DNA} < 0 < v_{P-DNA}$. This relation can be achieved if electrophoretic mobilities of the target–binder complex, μ_{P-DNA} , and nonbinders, μ_{DNA} , relate to the mobility of electroosmotic flow (EOF), μ_{EOF} , as: $|\mu_{DNA}| > |\mu_{EOF}| > |\mu_{P-DNA}|$. The latter relation can be achieved, in turn, by decreasing $|\mu_{EOF}|$ via increasing the ionic strength of the running buffer, I_{RB} .

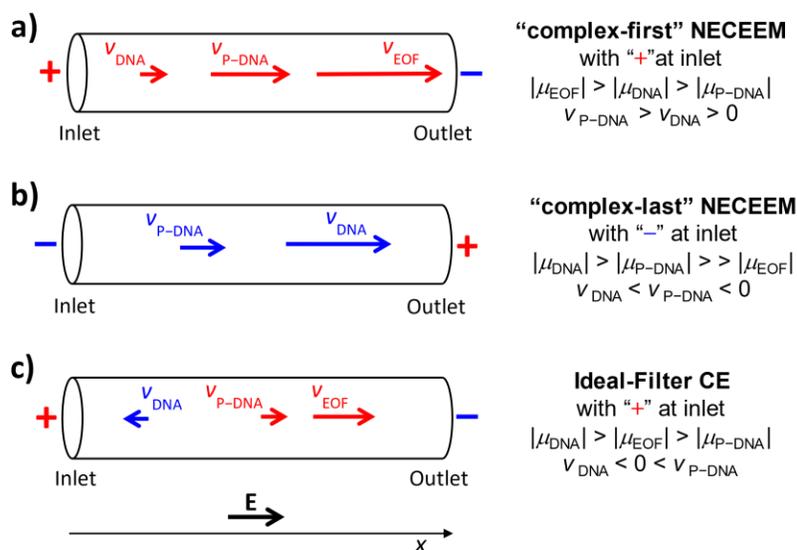


Figure 2.2. Schematics of KCE-based partitioning of protein target–binder complexes (P–DNA) from nonbinders (unbound DNA). Panels (a) and (b) show standard KCE-based partition with velocity vectors of complexes and nonbinders directed towards the capillary outlet biased at “–” and “+”, respectively. Panel (c) shows IFCE-based partitioning with the velocity vector of complexes directed towards the capillary outlet at “–” and with a counter-directed velocity vector of nonbinders.

However, this approach has a limitation: increasing I_{RB} leads to greater Joule heat generation and increased temperature inside the capillary, T_{cap} , that can lead to dissociation of target–binder complexes. Luckily, this limitation can be overcome with a recently introduced simplified universal method for predicting electrolyte temperatures (SUMET) which allows adjusting T_{cap} to a desirable value by rationally lowering the electric field strength, E .⁸⁴ Advantageously, increasing I_{RB} makes the running buffer more physiological and can allow selection of binders intended for use *in vivo* (e.g., as detection probes, drugs, or drug-delivery vehicles). Therefore, we decided to explore varying I_{RB} to test our main hypothesis that achieving $v_{DNA} < 0 < v_{P-DNA}$ could reduce the nonbinder background in KCE-based partitioning.

In order to solve the nonbinder background problem in the context of our hypothesis, we formulated five specific goals. Goal 1 was to find E that could guarantee a suitable range of T_{cap} for I_{RB} ranging from low values typically used in CE to high values approaching physiological. Goal 2 was to test whether a condition of IFCE ($v_{\text{DNA}} < 0 < v_{\text{P-DNA}}$) could be achieved in CE by increasing I_{RB} to near-physiological values while keeping elution time for the target–binder complexes under 1 h. Goal 3 was to test if satisfying this condition could drastically reduce the nonbinder background (reduce k_{N} to the desired value of $\sim 10^{-9}$). Goal 4 was to confirm that $k_{\text{B}} \approx 1$. Goal 5 was to examine if the obtained $k_{\text{B}}/k_{\text{N}} \approx 1/k_{\text{N}}$ was sufficient to facilitate selection of binders from a naïve oligonucleotide library in one step of IFCE-based partitioning. The following sections describes how the five goals were achieved.

2.1.3.3. Electric field to guarantee no overheating

The value of I_{RB} was changed by introducing NaCl in concentration ranging from 25 to 150 mM to the running buffer (RB). The value of I_{RB} for 50 mM Tris-HCl is 46 mM and every additional 25 mM NaCl adds 25 mM to the I_{RB} value. Accordingly, the values of I_{RB} in our experiments ranged from 46 to 196 mM. To place these values in the context of biological relevance, note that I_{RB} of phosphate buffered saline is 162 mM. A SCIEX CE instrument utilized in our study allowed capillary thermo-stabilization *via* its contact with a liquid coolant stabilized at T_{coolant} . The central longest part of the capillary was thermostated by the coolant and had temperature $T_{\text{cooled}} > T_{\text{coolant}}$; while the two short flanking parts of the capillary were not in contact with the coolant and their temperature $T_{\text{non-cooled}}$ was higher than T_{cooled} . The goal of this part of the study was to find E which guaranteed $T_{\text{cooled}} \leq 20$ °C and $T_{\text{non-cooled}} \leq 42$ °C at $T_{\text{coolant}} = 15$ °C (20 and 42 °C flank the range of vital temperatures of the human body).

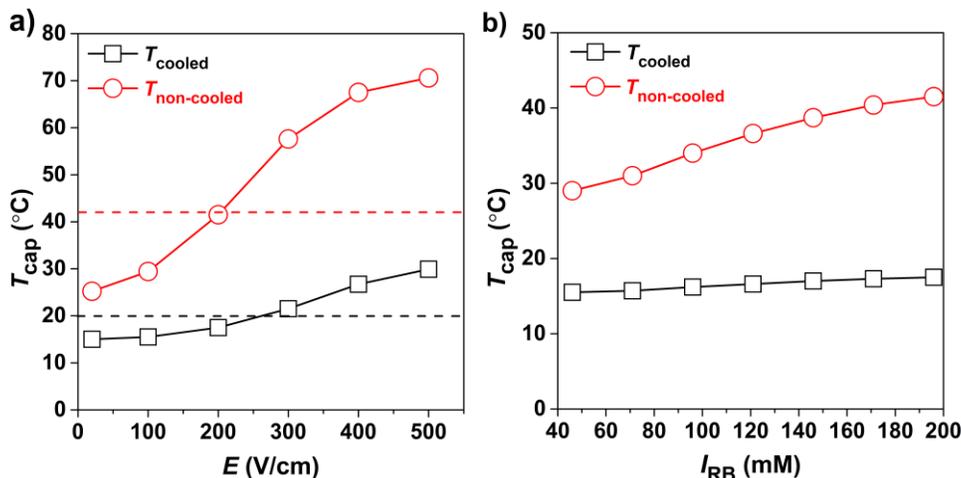


Figure 2.3. Determination of E that guarantees no overheating. Panel **a** shows the dependency of T_{cooled} and $T_{\text{non-cooled}}$ on E for $I_{\text{RB}} = 196$ mM. The value of E satisfying both $T_{\text{cooled}} \leq 20$ °C and $T_{\text{non-cooled}} \leq 42$ °C was 200 V/cm. Panel **b** shows the values of T_{cooled} and $T_{\text{non-cooled}}$ for different I_{RB} at $E = 200$ V/cm.

Our SUMET method was utilized for determination of T_{cooled} and $T_{\text{non-cooled}}$ for different values of E .⁸⁴ Since the highest I_{RB} corresponds to the greatest Joule heat, it was sufficient for us to determine E which satisfied these conditions for RB containing 150 mM NaCl ($I_{\text{RB}} = 196$ mM); RBs with lower I_{RB} would then automatically satisfy the two conditions. Accordingly, the electric current was measured for running buffers containing 150 mM NaCl during 1 min for each of six values of applied voltage and for six corresponding values of E : 20, 100, 200, 300, 400 and 500 V/cm. The collected current-voltage data and the SUMET program were used to determine the dependence of T_{cooled} and $T_{\text{non-cooled}}$ on E (**Figure 2.3a**). The highest E satisfying both $T_{\text{cooled}} \leq 20$ °C and $T_{\text{non-cooled}} \leq 42$ °C was 200 V/cm; the sought temperatures were $T_{\text{cooled}} = 17.5$ °C and $T_{\text{non-cooled}} = 41.5$ °C.

We then conducted current-voltage measurements at $E = 200$ V/cm for RBs with lower ionic strengths and determined T_{cooled} and $T_{\text{non-cooled}}$ as functions of I_{RB} (**Figure 2.3b**). As expected, all

values of T_{cooled} were in a range of 15.0–17.5 °C and all values of $T_{\text{non-cooled}}$ were higher than 17.5 °C and lower than 41.5 °C. Thus, an electric field of 200 V/cm guaranteed $T_{\text{cooled}} \leq 20$ °C and $T_{\text{non-cooled}} \leq 42$ °C and was used in the rest of this study.

To avoid sample exposure to higher temperatures in the inlet non-cooled part of the capillary, the sample was propagated through this part by pressure at $E = 0$, and the voltage was applied only when the sample was 2 cm inside the cooled part of the capillary. The intact complexes were still exposed to the elevated $T_{\text{non-cooled}} \leq 41.5$ °C while passing through the non-cooled part at the capillary outlet.

2.1.3.4. Achieving $v_{\text{P-DNA}} > 0 > v_{\text{DNA}}$

In a set of experiments testing whether $v_{\text{P-DNA}} > 0 > v_{\text{DNA}}$ could be achieved in KCE by increasing I_{RB} , the sample was an equilibrium mixture of MutS (MW \approx 90 kDa, pI \approx 5.2) with its previously selected and characterized fluorescently-labeled DNA aptamer.³⁴ The equilibrium mixture contained the MutS–aptamer (target–binder) complex and an unbound aptamer (nonbinder). To be able to detect nonbinders moving with very small velocities, we conducted these experiments with a short separation distance of $l_{\text{sep}} = 4.5$ cm and with a long run time of $t = 50$ min. Every equilibrium mixture was run with two CE polarities: “+” at the inlet (**Figure 2.4a**) and “–” at the inlet (**Figure 2.4b**); the second was required to detect nonbinders when $v_{\text{DNA}} < 0$. The magnitude of the minimum “detectable” velocity was $|v_{\text{min}}| = l_{\text{sep}}/t = 0.9$ mm/min. The peak of the complex was detected only with “+” at the inlet and for all concentrations of NaCl suggesting that $v_{\text{P-DNA}} > 0$ for all I_{RB} values tested. The peak of nonbinders was detected with “+” at the inlet for $[\text{NaCl}] \leq 50$ mM, and with “–” at the inlet for $[\text{NaCl}] \geq 100$ mM, suggesting that $v_{\text{DNA}} > |v_{\text{min}}|$ for $[\text{NaCl}] \leq 50$ mM and $v_{\text{DNA}} < -|v_{\text{min}}|$ for $[\text{NaCl}] \geq 100$ mM. The

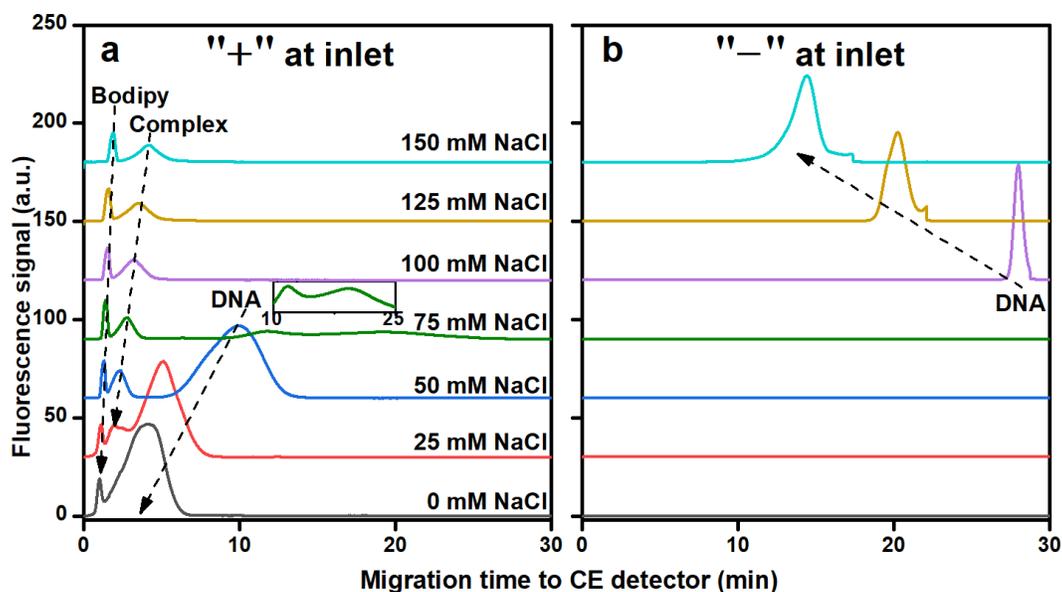


Figure 2.4. The effect of NaCl added to RB on migration pattern of protein–DNA complexes and unbound DNA in CE with “+” (a) and “–” (b) at the inlet, respectively. The equilibrium mixture contained 100 nM MutS protein, 100 nM fluorescently-labeled DNA aptamer of MutS, and 150 nM Bodipy (EOF marker) and was incubated for 30 min at room temperature. No peaks were observed after 30 min; therefore, only the first 30 min of 50-min runs are shown.

peak of nonbinders was not detected with either polarity for $[\text{NaCl}] = 75 \text{ mM}$ suggesting $|v_{\text{DNA}}| < |v_{\text{min}}|$. These results prove that $v_{\text{P-DNA}} > 0 > v_{\text{DNA}}$ can be satisfied by increasing I_{RB} (to $I_{\text{RB}} \geq 150 \text{ mM}$ in our case, which corresponds to $[\text{NaCl}] \geq 100 \text{ mM}$).

2.1.3.5. Lowering nonbinder background *via* increasing I_{RB}

We then studied how increasing I_{RB} affected the nonbinder background. The sample was DNA without any target. Two-minute fractions were collected and DNA quantities in them were determined *via* quantitative PCR (qPCR). Such experiments were carried out with different I_{RB} , and the results were presented as “DNA quantity in a corresponding fraction *vs* migration time of

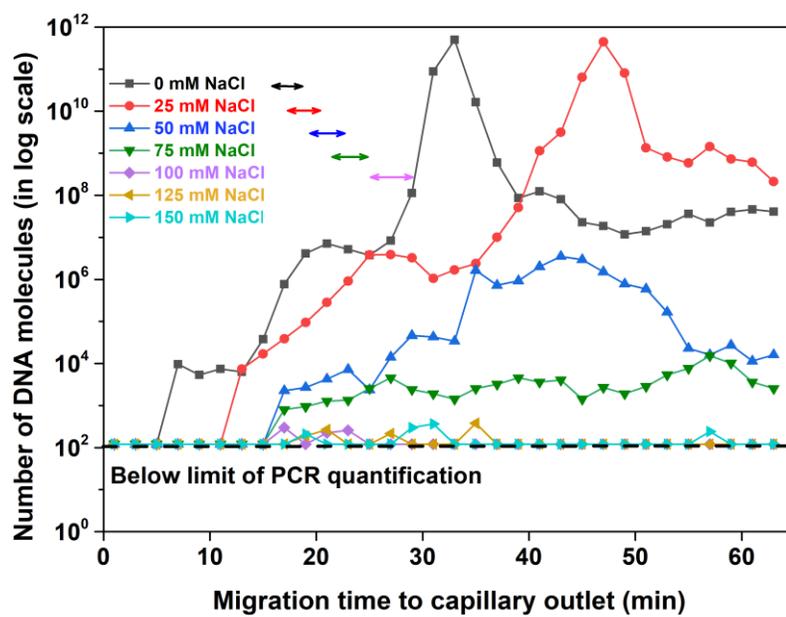


Figure 2.5. The effect of NaCl added to RB on the migration pattern of DNA in CE with “+” at the inlet. The sample contained 10 μM DNA and 150 nM Bodipy (EOF marker and a reference for correct start time of collecting the first fraction). Separation distance was 34 cm. Fractions were collected every 2 min and concentrations of DNA in them were determined with qPCR and used to calculate DNA quantities in these fractions. These quantities are shown on the y -axis in the graph. The double-headed arrows indicate estimated elution windows of the aptamer (see **Appendix A1**).

this fraction to the capillary outlet” (**Figure 2.5**). In agreement with our hypothesis, increasing I_{RB} led to decreasing the DNA (nonbinder) background. Remarkably, the background decreased down to and below the limit of quantitation (LOQ) of qPCR at $[\text{NaCl}] \geq 100$ mM. The values of k_{N} were calculated as integrals under DNA curves within the binder-collection time windows (see **Appendix A1**) divided by the total quantity of DNA sampled into the capillary. The latter was calculated as an integral under the DNA curve for $[\text{NaCl}] = 0$ within a 0 to 50 min time window. For different concentrations of NaCl in RB we obtained the following values of k_{N} :

8×10^{-6} (0 mM NaCl), 6×10^{-7} (25 mM NaCl), 2×10^{-8} (50 mM NaCl), 9×10^{-9} (75 mM NaCl), and 6×10^{-10} (100 mM NaCl). Adding 100 mM NaCl to RB resulted in 1.3×10^4 times decrease in k_N in comparison to no NaCl in RB. We did not calculate k_N for $[\text{NaCl}] \geq 125$ mM as the quantity of DNA was well below the LOQ of qPCR. These experiments prove that IFCE (i.e., $v_{\text{P-DNA}} > 0 > v_{\text{DNA}}$) can drastically decrease the nonbinder background and reach $k_N \sim 10^{-9}$. In the rest of the study, IFCE was conducted with RB containing 100 mM NaCl ($I_{\text{RB}} \approx 150$ mM).

2.1.3.6. Efficiency of binder collection

So far we assumed that $k_B \approx 1$, and, thus, k_B/k_N was anticipated to be predominantly defined by k_N . In principle, k_B can be much lower than unity due to binder loss in partitioning. In solid-phase partitioning, the “best” aptamers can be lost due to the inability to soft-dissociate them from the surface-immobilized target. In KCE-based partitioning, aptamers can be lost due to an incorrectly-determined binder-collection time window. Here, we proved our assumption of $k_B \approx 1$ by (i) determining the quantity of MutS–aptamer (target–binder) complex sampled, $B_{\text{in}} = (8.9 \pm 0.9) \times 10^8$, (ii) determining the quantity of aptamers (binders) collected in the binder-collection time window corresponding to the elution time window of the MutS–aptamer complex, $B_{\text{out}} = (7.3 \pm 0.5) \times 10^8$, and (iii) calculating $k_B = B_{\text{out}}/B_{\text{in}} = 0.8 \pm 0.3 \approx 1$. See **Appendix A2** for details.

2.1.3.7. One-step aptamer selection

To test if the achieved IFCE efficiency of partitioning of $k_B/k_N > 1.7 \times 10^9$ could facilitate one-step selection of aptamers we conducted partitioning of MutS binders from a random-sequence DNA library. To exclude the effect of potential contamination of solutions with traces

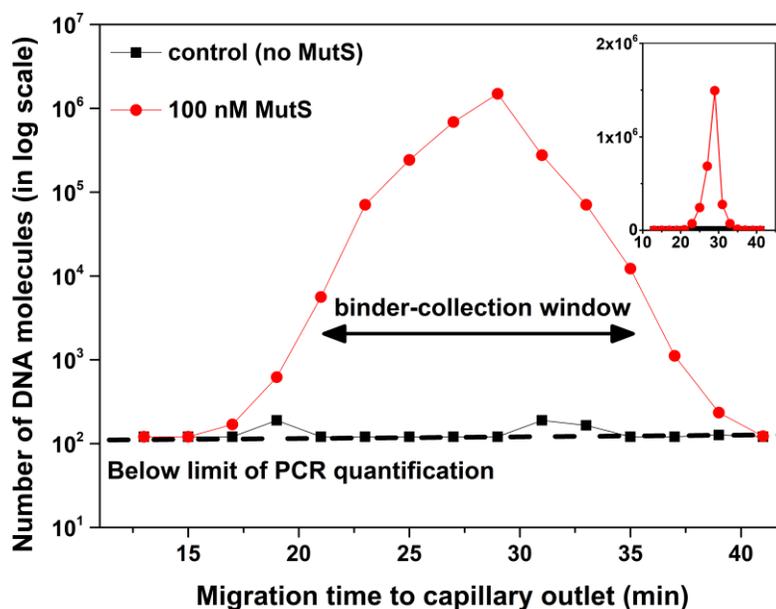


Figure 2.6. IFCE-based partitioning of DNA binders of MutS protein from the unbound library.

The sampled equilibrium mixture was 47 nL in volume and contained 10 μ M random-sequence DNA library and 100 nM MutS protein. Separation distance was 34 cm. The inset shows the same data but with a linear ordinate.

of the aptamer used to measure k_N and k_B , the library was designed with PCR-priming regions different from those of the aptamer. A sample of the equilibrium mixture containing the library ($B_{in} + N_{in} \approx 2.8 \times 10^{11}$) and MutS was subjected to IFCE. Two-minute fractions were collected and analyzed by qPCR to build a “DNA quantity vs migration time to the capillary outlet” electropherogram; the control experiment was similar, but MutS in the equilibrium mixture was replaced with RB (**Figure 2.6**). B_{out} and N_{out} were calculated as integrals under the curves within the target–binder complex collection time window, 23 to 35 min, in IFCE and control experiment, respectively. They were found to be $B_{out} \approx 2.9 \times 10^6$ and $N_{out} \approx 1.1 \times 10^3$, and, accordingly, $B_{out}/N_{out} \approx 2.6 \times 10^3$. Thus, IFCE could support $B_{out}/N_{out} \gg 100$, which confirmed completed selection using the chosen very strong criterion of selection completion. This experiment independently confirmed that $k_N = N_{out}/N_{in} \sim 10^{-9}$ can be reached in a real selection

(from a random-sequence DNA library and in the presence of a protein target). The knowledge of the quantity of the sampled nonbinders $N_{\text{in}} \approx B_{\text{in}} + N_{\text{in}}$ and the quantity of sampled binders $B_{\text{in}} \approx B_{\text{out}}$ (as $k_{\text{B}} \approx 1$) uniquely allowed estimation of the initial binder abundance: $B_{\text{in}}/N_{\text{in}} \approx 1.0 \times 10^{-5}$. In other words, approximately 0.001% of the random-sequence library were bound to MutS in the equilibrium mixture containing 100 nM MutS and stayed bound for the duration of the 1-h long IFCE partitioning run. The initial abundance is obviously not an invariant; it depends on the natures of target and library, their concentrations, incubation time, *etc.* IFCE can uniquely facilitate studies which are needed to understand how $B_{\text{in}}/N_{\text{in}}$ depends on these parameters.

As a final step, we amplified a fraction containing the highest quantity of complexes by PCR using a fluorescently-labeled primer. After amplifying DNA in this fraction by PCR, we performed a pressure-aided IFCE-based binding test with fluorescence detection (see **Appendix A3**). This test revealed an apparent equilibrium dissociation constant of the enriched library of $K_{\text{d, app}} \approx 40$ nM and confirmed successful selection of a high-affinity aptamer pool in one step of IFCE partitioning. For comparison, selecting a pool with similar $K_{\text{d, app}}$ by classical KCE-based partitioning (complex-first NECEEM mode) required three rounds of SELEX.³⁴ This successful one-step selection, in turn, confirmed correctness of our estimate that $k_{\text{N}} \sim 10^{-9}$ was sufficient for one-step selection. Cumulatively, this study proves that our approach of rationally developing a partitioning method for one-step selection of binders from oligonucleotide libraries does work.

2.1.4. Conclusions

To summarize, the condition of IFCE, i.e., the migration of target–binder complexes and nonbinders in the opposite directions, was achieved by raising I_{RB} to a physiological value at

physiological pH. We demonstrated that IFCE allows reaching uniquely low nonbinder background with $k_N \sim 10^{-9}$. This value is $\sim 10^7$ lower than k_N values of practical solid-phase partitioning methods. Importantly, such extremely low k_N was reached without sacrificing k_B which was near unity. The resulting k_B/k_N was sufficient for selection of a potent aptamer pool for MutS protein in one step of partitioning. An additional advantage of IFCE is that the very high efficiency of partitioning can be achieved by simply increasing the ionic strength of the running buffer to a physiological level of 146 mM. This feature makes IFCE suitable for the selection of binding molecules intended for *in vivo* applications. The successful utilization of physiological conditions in IFCE has expanded the scope of KCE-based selection to encompass the selection of binding molecules under physiological conditions. Previously, only low ionic strength running buffers were used for such purposes.

Note that in IFCE-based binder selection, oligonucleotide amplification by PCR or another process is needed only for binder identification; therefore, high-fidelity amplifying enzymes (e.g., polymerases) must be used. While the proof of IFCE was done here with a random-sequence DNA library, we foresee that IFCE will be directly applicable to selection of binders from other anionic libraries with a uniform charge. For example, IFCE should greatly benefit selection of binders from DNA-encoded libraries to which SELEX is not applicable due to inability to PCR-amplify such libraries.

2.2. DETERMINATION OF THE EQUILIBRIUM CONSTANT AND RATE CONSTANT OF PROTEIN–OLIGONUCLEOTIDE COMPLEX DISSOCIATION UNDER THE CONDITIONS OF IDEAL-FILTER CAPILLARY ELECTROPHORESIS

The presented material was published previously and reprinted with permission from “**Le, A.T.H.**; Krylova, S.M.; Krylov, S.N. Determination of the equilibrium constant and rate constant of protein–oligonucleotide complex dissociation under the conditions of ideal-filter capillary electrophoresis. *Analytical Chemistry* **2019**, *91*, 12680–12687.” Copyright 2019 American Chemical Society. My contributions to the article were: (i) performing all presented experiments, (ii) preparing all figures, (iii) interpreting the results and (iv) writing the manuscript.

2.2.1. Requirement of a KCE method for determining affinity and kinetics of binders selected under physiological or IFCE conditions

In the previous section, we have introduced ideal-filter capillary electrophoresis (IFCE), in which protein-bound and unbound oligonucleotides move in the opposite directions inside the capillary. The condition of IFCE is achieved when the magnitude of the mobility of electroosmotic flow (EOF) is larger than that of protein–oligonucleotide complexes and smaller than that of free oligonucleotides while the signs of the latter two mobilities are negative. The efficiency of IFCE-based partitioning of binders from nonbinders reaches 10^9 , which is 10^7 times greater than those of solid-phase partitioning methods and 10^4 times greater than that of classical KCE modes (i.e., NECEEM), in which protein-bound and unbound oligonucleotides move in the same direction. The classical KCE-based partitioning is carried out at higher-than-physiological pH and lower-than-physiological ionic strength of the background electrolyte.^{31,32,35-44,56,58-61}

Advantageously, the conditions of IFCE are achieved at near-physiological values of pH and ionic strength providing greater biological relevance of selected protein binders and suppressed non-specific binding of the protein with the oligonucleotides and with the capillary surface.⁴² Partitioning by IFCE was shown to facilitate the one-step selection of DNA aptamers, and it promises to be applicable to selection of protein binders from DELs.

The movement of the unbound oligonucleotides away from the detection end of the capillary makes impossible their detection in IFCE. This creates a problem for measuring the equilibrium constant (K_d) and rate constant (k_{off}) of protein–oligonucleotide complex dissociation by IFCE. (It is worth mentioning that this problem does not exist in non-equilibrium capillary electrophoresis of equilibrium mixtures, in which protein-bound and unbound oligonucleotides move in the same direction.) Here we describe a solution for this problem *via* double passage of the sample through the detection window: the first passage is driven by pressure, and the second passage is driven by an electric field.

The essence of the proposed double-passage approach is the following (**Figure 2.7**). The oligonucleotide is fluorescently labeled allowing for fluorescence detection of both free oligonucleotide and protein–oligonucleotide complex. A plug of the protein–oligonucleotide mixture at equilibrium is injected into the capillary from its close-to-the-detector end. The plug is moved slowly by pressure to and through the detection window and stopped after passing it. Since the protein-bound oligonucleotide is not separated from the unbound oligonucleotide in a slow pressure-driven flow, both contribute to the area of the first recorded peak. The voltage is then applied with a negative polarity being at the injection end of the capillary, and the protein-bound oligonucleotide moves back to the detector while the unbound oligonucleotide moves further away from the detector. Only the protein-bound oligonucleotide passes through the

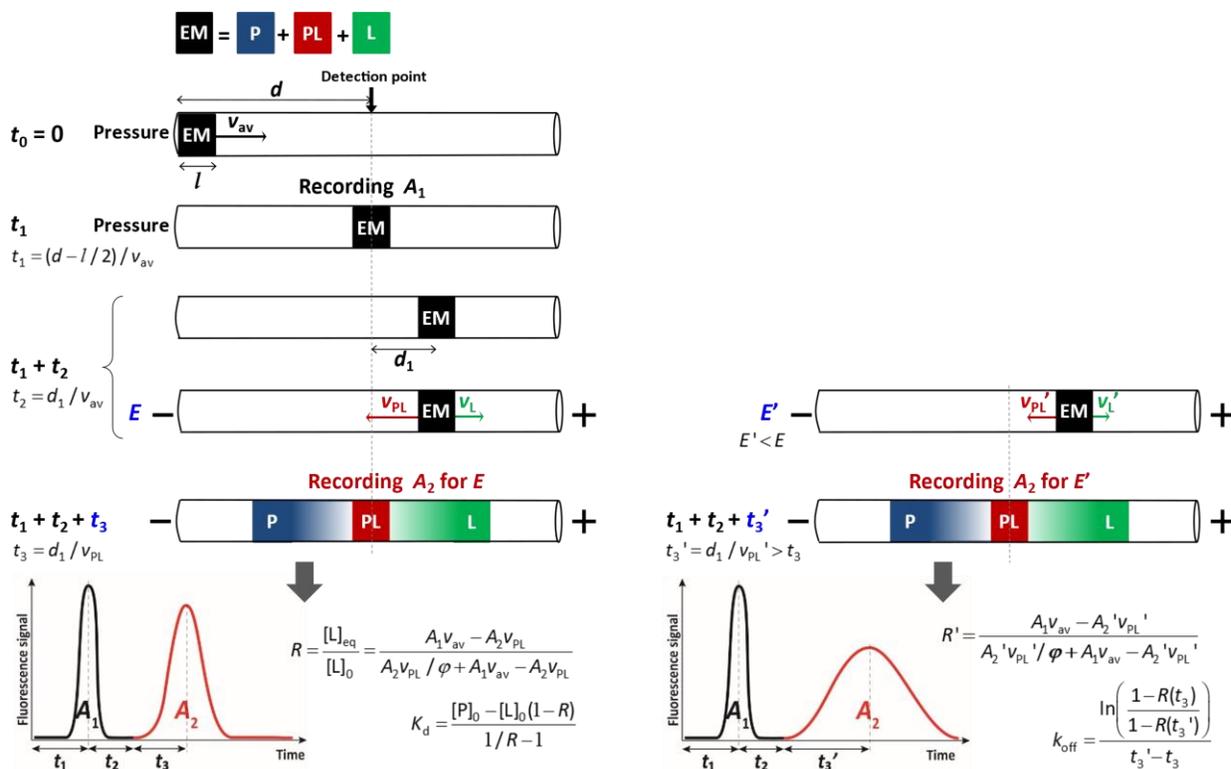


Figure 2.7. Schematic of the double-passage approach for finding K_d (left) and k_{off} (right) under the IFCE conditions. The pressure-driven step is identical for the left and right parts. See text for details.

detector the second time and contributes to the area of the second recorded peak. The two peak areas are corrected for different velocities of pressure-driven migration and electric-field-driven migration. The relative quantum yield of the fluorescent label within the protein-bound oligonucleotide with respect to that of unbound oligonucleotide can be determined as a ratio of first-passage peak areas for pure oligonucleotide and near-pure complex that exists in the equilibrium mixture with protein concentration much greater than the value of K_d . Finally, the areas corrected for different velocities, different quantum yields, and also label photobleaching (if significant) are used to calculate accurate values of K_d and k_{off} .

In this work, general procedures for finding K_d and k_{off} are summarized, and the theory of the double-passage approach in the context of these general procedures is presented. Selection of suitable conditions for the pressure-driven stage is discussed. Good recoveries of peak areas and a fraction of unbound oligonucleotide were demonstrated using a model system of non-interacting green fluorescent protein (GFP) and a fluorescently-labeled oligonucleotide mixed at different ratios. Finally, the double-passage approach was utilized to find K_d and k_{off} values for an affinity complex between MutS protein and its DNA aptamer. The double-passage approach provides a solution for accurate determination of K_d and k_{off} of protein–oligonucleotide complexes under the near physiological conditions of IFCE. This kind of measurements is impossible in regular KCE mode, such as NECEEM, which is carried out at higher-than-physiological pH and lower-than-physiological ionic strength. The double-passage approach for finding K_d and k_{off} of protein–oligonucleotide complexes under IFCE conditions is a perfect complement for IFCE-based one-step partitioning of protein binders from oligonucleotide libraries.

2.2.2. Materials and methods

2.2.2.1. Materials and solutions

All chemicals and buffer components were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries with inner and outer diameters of 75 and 360 μm , respectively, were purchased from Molex Polymicro (Phoenix, AZ, USA). Recombinant *A. victoria* GFP protein (MW \approx 27 kDa, pI 5.7) was obtained from Abcam (Cambridge, UK). Recombinant *T. aquaticus* MutS protein (MW \approx 90 kDa, pI 6.0) was expressed and purified as described previously.⁹³ All DNA molecules were custom synthesized by Integrated DNA

Technologies (Coralville, IA, USA). To study the recovery of *R*, a synthetic FAM-labeled DNA library (N40) with a 40-nt random region was used: 5'-FAM-CT ACG GTA AAT CGG CAG TCA -N40- AT CTG AAG CAT AGT CCA GGC. To determine K_d and k_{off} of a protein–aptamer pair, the DNA aptamer with affinity toward MutS protein was previously selected in our laboratory (clone 2-06),³⁴ and its Alexa Fluor 488-labeled version was used here: 5'-AlexaFluor488-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 determine concentrations of protein and DNA stock solutions by measuring light absorbance at 280 and 260 nm, respectively, and dividing the absorbance by the corresponding molar extinction coefficients. A solution of 5% bovine serum albumin was used to coat all the sample vials to minimize absorption of DNA and protein onto the reservoir walls. The running buffer was 50 mM Tris-HCl pH 7.0 supplemented with 100 mM NaCl. The sample buffer was always identical to the running buffer to prevent adverse effects of buffer mismatch. Accordingly, all dilutions of sample components used in CE experiments were done by adding the running buffer.

2.2.2.2. CE instrumentation and default conditions

Our default experimental parameters were: $d = 10.2$ cm, $r = 37.5$ μm , $l = 1.09$ cm, pressure for carrying out the first passage = 0.30 psi, total capillary length = 50 cm, and the electric field for carrying out the second passage $E = 200$ V/cm. These default experimental parameters were used for all experiments in this work except for measurements of k_{off} in which we also used a lower electric field strength of $E' = 120$ V/cm. Under our experimental conditions, magnitudes of electrophoretic mobilities of EOF, protein–oligonucleotide complex, and oligonucleotide were: $\mu_{EOF} = 20$, $\mu_{PL} = 13$, and $\mu_L = 23$ $\text{mm}^2/(\text{kVs})$.

All experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) capable of applying both pressure and electric field and equipped with a laser-induced fluorescence detection system. Fluorescence was excited with a blue line (488 nm) of a solid-state laser and detected at 520 nm using a spectrally-optimized emission filter system. An uncoated fused-silica capillary, with a total length of 50 cm and a 10.2-cm distance from one of the ends to the detection window was used. The temperature of the capillary coolant was set to 15°C. Prior to every run, the capillary was rinsed successively with 0.1 M HCl, 0.1 M NaOH, deionized H₂O, and a running buffer for 3 min each under a pressure of 20 psi.

Mixtures of various concentration combinations were prepared, as described in the results and discussion section. Prior to the mixing, the DNA stock solution was incubated at 90 °C for 2 min and gradually cooled down to 20 °C at a rate of 0.5 °C/s. For double-passage experiments with GFP–DNA mixtures, the concentration of GFP was 250 nM while the concentration of DNA ranged from 20 to 1000 nM (20, 50, 100, 200, 400, 600, and 1000). For experiments with MutS and its aptamers, the mixture was incubated at room temperature for 1 h. The prepared samples were injected into the capillary by a 0.3 psi × 17 s pressure pulse to yield a 1.09-cm-long sample plug inside the capillary. The injected sample plug was moved by pressure at 0.3 psi for 3.5 min. The pressure was then stopped, and either 10 or 6 kV was applied to the capillary ends (cathode at the injection end) resulting in the electric field strengths of 200 or 120 V/cm, respectively. Peak areas and migration times were obtained by analyzing electropherograms with 32 Karat Software.

2.2.3. Results and discussion

2.2.3.1. General procedures for finding K_d and k_{off}

At its basic level, the study of an interacting molecular system, such as the one represented in **Eq. 1**, involves the measurement of its equilibrium dissociation constant (K_d), which characterizes the binding affinity the two molecules (i.e., P and L), and the rate constants of the association (k_{on}) and dissociation (k_{off}), respectively. The definition of K_d , through the equilibrium concentrations of P ($[P]_{eq}$), L ($[L]_{eq}$) and PL ($[PL]_{eq}$), as well as its relationship with k_{on} and k_{off} , is illustrated in **Eq. 2**. In general, finding K_d requires the determination of a fraction R of unbound L in a mixture of L and P in the state of equilibrium with the initial concentrations $[L]_0$ and $[P]_0$:

$$R = \frac{[L]_{eq}}{[L]_0} \quad (6)$$

K_d can then be calculated using the following expression:

$$K_d = \frac{[P]_0 - [L]_0(1 - R)}{1/R - 1} \quad (7)$$

Finding K_d is associated with a systematic error that is minimal when the following relationships are satisfied⁹⁵:

$$\begin{aligned} [L]_0 / K_d &\leq 1 \\ [P]_0 / K_d &\approx 1 \end{aligned} \quad (8)$$

Since K_d is not known a priori, accurate determination of K_d may require measurement of R at different $[P]_0$ and fitting the R vs $[P]_0$ binding isotherm with their theoretical relationship while using K_d as a variable:

$$R = -\frac{K_d + [P]_0 - [L]_0}{2[L]_0} + \sqrt{\left(\frac{K_d + [P]_0 - [L]_0}{2[L]_0}\right)^2 + \frac{K_d}{[L]_0}} \quad (9)$$

Further, if the determined K_d does not satisfy the first inequality in **Eq. 8**, measurements should be repeated at a lower $[L]_0$.⁹⁵

The determination k_{off} requires inducing complex dissociation by physical separation of PL from L and P and measuring R as a function of time. The experimental dependence of $(1 - R(t))/(1 - R_{t=0})$ on time should then be fitted with the theoretical dependence:

$$\frac{1 - R(t)}{1 - R_{t=0}} = e^{-k_{\text{off}}t} \quad (10)$$

while using k_{off} as a fitting parameter. The complex-dissociation experiments should be conducted under conditions that minimize the rebinding reaction during the separation process. The rebinding can be neglected if a characteristic separation time of PL from the unbound L and P is much shorter than a characteristic rebinding time:

$$t_{\text{sep}} \ll t_{\text{rebind}} \quad (11)$$

When separation is carried out by KCE, the characteristic separation time is defined *via* a plug length l , and velocity vectors \mathbf{v}_{PL} , \mathbf{v}_{L} , and \mathbf{v}_{P} for PL, L, and P, respectively:

$$t_{\text{sep}} = \frac{l}{\max\{|\mathbf{v}_{\text{PL}} - \mathbf{v}_{\text{L}}|, |\mathbf{v}_{\text{PL}} - \mathbf{v}_{\text{P}}|\}} \quad (12)$$

For conditions of IFCE, $|\mathbf{v}_{\text{PL}} - \mathbf{v}_{\text{L}}| > |\mathbf{v}_{\text{PL}} - \mathbf{v}_{\text{P}}|$ and $|\mathbf{v}_{\text{PL}} - \mathbf{v}_{\text{L}}| = v_{\text{PL}} + v_{\text{L}}$, where v_{PL} and v_{L} are magnitudes (positive values) of the corresponding velocities. Thus, for the conditions of IFCE, **Eq. 12** can be further simplified:

$$t_{\text{sep}} = \frac{l}{v_{\text{PL}} + v_{\text{L}}} \quad (13)$$

The rate at which the rebinding reaction between P and L occurs during the separation process is directly proportional to the binding rate constant k_{on} and the concentration of unbound protein at equilibrium $[\text{P}]_{\text{eq}}$; the latter can be assumed to be approximately equal to $[\text{P}]_0$ for a typical condition of $[\text{L}]_0 \ll [\text{P}]_0$ used when finding k_{off} . The characteristic time of rebinding, t_{rebind} , thus, can be defined as:

$$t_{\text{rebind}} = 1/(k_{\text{on}}[\text{P}]_{\text{eq}}) \approx 1/(k_{\text{on}}[\text{P}]_0) \quad (14)$$

Thus, minimizing rebinding of L formed from complex dissociation calls for minimizing $[\text{P}]_0$. On the other hand, a condition of $[\text{P}]_0 \gg [\text{L}]_0$ should still be satisfied, and the dynamic range of R should be sufficient for accurately measuring the kinetics of R increase upon complex dissociation. As a rule of thumb, $[\text{P}]_0 = K_d$ can be used for the determination of k_{off} . In such a case, $R_{t=0} = 1/2$, and the dynamic range of R is from $1/2$ to 1. In general, it is practical to determine K_d first, which can require varying $[\text{P}]_0$, and, then, determine k_{off} at a single concentration of $[\text{P}]_0 = K_d$.

2.2.3.2. Theory of the double-passage approach

The double-passage approach for finding K_d and k_{off} of protein–oligonucleotide complexes under the ICFE conditions is schematically illustrated in **Figure 2.7**. The equilibrium mixture (EM) of the protein (P) and the oligonucleotide that it binds (L) is prepared outside the capillary in the IFCE running buffer. A plug of EM of length l is injected into the capillary by pressure at a low flow velocity v_{av} to minimize plug elongation due to Taylor dispersion. The running buffer is then injected by pressure into the capillary at the same v_{av} to move the plug of EM through the detection window. Since the PL and L are not separated from each other in a slow pressure-driven flow, the area of the first recorded peak will represent a sum of equilibrium concentrations of PL and L:

$$A_1 = a(\phi[\text{PL}]_{\text{eq}} + [\text{L}]_{\text{eq}}) / v_{\text{av}} \quad (15)$$

where a is a coefficient of proportionality between the peak area and the concentration of the fluorescent oligonucleotide, and ϕ is a relative quantum yield of a fluorescent label within the protein–oligonucleotide complex with respect to that of the free oligonucleotide. The value of a

is constant for the fixed geometry of the detection system, constant intensity of fluorescence-excitation light, and constant sensitivity of the photodetector. The value of φ is not expected to be much less than unity as the fluorophore is not involved in the protein–oligonucleotide binding reaction. The value of φ can be determined as a ratio between A_1 found for pure oligonucleotide ($[P]_0 = 0$) and A_1 for near-pure complex (the same $[L]_0$, but $[P]_0 \gg K_d$). The value of v_{av} can be calculated as:

$$v_{av} = \frac{d-l/2}{t_1} \quad (16)$$

where d is the distance from the capillary inlet to the center of the detection window, and t_1 is time span from the beginning of EM plug propagation to the time when the plug's center reaches the center of the detection window.

The pressure-driven flow is stopped when the plug passes the detection window. The time span from the plug center's passing the center of detection window to the end of plug propagation is t_2 , and the distance between the center of the stopped plug and the center of the detection window is:

$$d_1 = v_{av}t_2 \quad (17)$$

The voltage is then applied with a negative polarity being at the injection end of the capillary. Under the conditions of IFCE, the protein-bound oligonucleotide moves back to the detector while the unbound oligonucleotide moves in the opposite direction, i.e., further away from the detector. Only PL passes through the detector the second time and contributes to the area of the second recorded peak. In the assumption of negligible complex dissociation during the second passage the area of the second peak is:

$$A_2 = a\varphi[PL]_{eq} / v_{PL} \quad (18)$$

where v_{PL} is the electrophoretic velocity of PL which can be determined as:

$$v_{\text{PL}} = \frac{d_1}{t_3} \quad (19)$$

Here t_3 is time from the beginning of electrophoresis to the moment when the center of the plug of PL reaches the center of the detection window.

If the two peak areas are measured and v_{av} , v_{PL} , and φ are determined as described above, then the value of R can be calculated as (see **Appendix B1** for derivation):

$$R = \frac{[\text{L}]_{\text{eq}}}{[\text{L}]_0} = \frac{A_1 v_{\text{av}} - A_2 v_{\text{PL}}}{A_2 v_{\text{PL}} / \varphi + A_1 v_{\text{av}} - A_2 v_{\text{PL}}} \quad (20)$$

and used to determine K_{d} with **Eq. 7**. If the found value of K_{d} does not satisfy the first inequality in **Eq. 8**, the experiments should be repeated at a lower value of $[\text{L}]_0$.

The value of k_{off} can be determined in a series of experiments with $[\text{P}]_0$ and $[\text{L}]_0$ satisfying both inequalities in **Eq. 8** and for varying $t_3 = v_{\text{PL}} d_1$ via changing either v_{PL} or d_1 . The value of v_{PL} can be changed by changing the magnitude of vector \mathbf{E} of the electric field. The value of v_{PL} is a scalar product of this vector and a sum of electrophoretic mobility vectors of PL, $\boldsymbol{\mu}_{\text{PL}}$, and EOF, $\boldsymbol{\mu}_{\text{EOF}}$, respectively:

$$v_{\text{PL}} = \mathbf{E} \cdot (\boldsymbol{\mu}_{\text{PL}} + \boldsymbol{\mu}_{\text{EOF}}) \quad (21)$$

The value of d_1 can be changed by propagating the plug of EM further from the detector during the first passage. A two-point approach (with two different times of electrophoretic propagation: $t_3 < t_3'$) can be used for a simple assessment of k_{off} (see **Appendix B2** for derivation):

$$k_{\text{off}} = \frac{\ln\left(\frac{1-R(t_3)}{1-R(t_3')}\right)}{t_3' - t_3} \quad (22)$$

The found value of k_{off} should be used to assess if the rebinding could be neglected, i.e., if the inequality in **Eq. 11** is satisfied. First, the value of k_{on} should be calculated with the found values of k_{off} and K_d by re-arranging **Eq. 2**:

$$k_{\text{on}} = k_{\text{off}} / K_d \quad (23)$$

This value of k_{on} should be used to calculate the characteristic rebinding time using **Eq. 14**. The value of v_L should be determined in a classical way by injecting a plug of pure L and propagating it to the detector by an electric field with a cathode at the injection end of the capillary and the magnitude of the electric field identical to that used in experiments utilized for finding K_d . If the migration time of the center of the plug of L to the center of the detector is t_4 then:

$$v_L = d / t_4 \quad (24)$$

If the found values of t_{sep} and t_{rebind} do not satisfy the inequality in **Eq. 11**, the experiments should be repeated with faster separation (smaller t_{sep}) which can be achieved by shortening the length l of the plug of EM and/or by increasing the differential velocity ($v_{\text{PL}} + v_L$) *via* increasing the electric field strength E , taking into account the proportionality of the differential velocity to E :

$$v_{\text{PL}} + v_L = E(\mu_{\text{PL}} + \mu_L) \quad (25)$$

where μ_{PL} and μ_L are magnitudes (positive values) of electrophoretic mobilities of PL and L, respectively.

2.2.3.3. Conditions for the first passage.

During the first, pressure-driven, propagation of the plug of EM to and through the detection window, equilibrium in the reversible binding reaction shown in **Eq. 1** should not be perturbed significantly. In the other words, the amounts of PL, L, and P should not change significantly

during the pressure-driven propagation. The equilibrium will not be perturbed if the propagation time, which is defined in **Eq. 16**, is much shorter than the characteristic time of complex dissociation:

$$t_1 \ll 1/k_{\text{off}} \quad (26)$$

Since we do not know k_{off} a priori, it appears natural to use as small t_1 as can be achieved; however, small t_1 requires high v_{av} which inevitably leads to sample dilution due to Taylor dispersion and, thus, perturbed equilibrium. The dilution effect of Taylor dispersion becomes significant when:

$$v_{\text{av}} > \frac{l}{r^2/D} \quad (27)$$

where r is capillary radius and D is the diffusion coefficient of the largest-size analyte, which is PL in our case (r^2/D is the characteristic time of transverse diffusion of this analyte across the capillary). Too high a value of v_{av} may even lead to transient incomplete separation (TIS) of L from PL, which produces a two-peak profile.^{96,97} In addition to complicating data interpretation, TIS can potentially affect accuracy of quantitation due to radial non-uniformity of signal readout. Thus, while no simple quantitative instruction can be given for an a priori choice of v_{av} , we suggest that TIS be avoided and satisfying **Eq. 26** be tested when k_{off} is determined.⁹⁶

2.2.3.4. Recovery of peak areas

The double-passage approach measures velocity-corrected areas $A_1 v_{\text{av}}$ and $A_2 v_{\text{PL}}$ (see **Eq. 20**), where A_1 and A_2 are peak areas determined for “signal vs time” coordinates. The goal of this part of our study was to confirm that $A_1 v_{\text{av}}$ and $A_2 v_{\text{PL}}$ could be accurately determined using the approach for determination of v_{av} and v_{PL} described in **Section 2.2.3.2**. We used 1 μM GFP as an analyte mimicking PL in terms of electrophoretic mobility and the following default

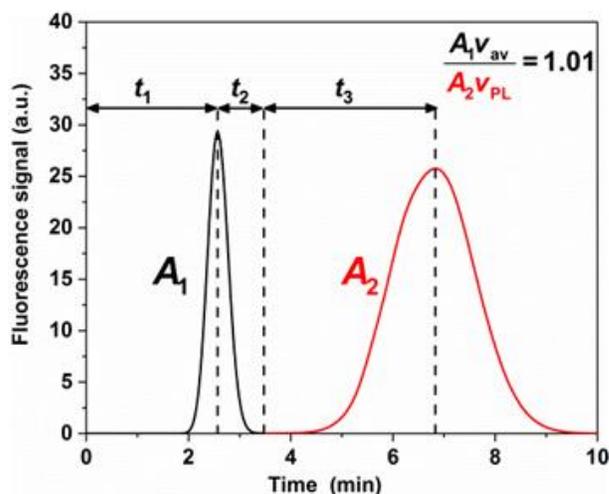


Figure 2.8. Experimental determination of velocity-corrected peak areas *via* double-passage experiment. The sample mixtures contained 1 μM GFP in 50 mM Tris-HCl pH 7.0, 100 mM NaCl.

experimental parameters: $d = 10.2$ cm, $r = 37.5$ μm , $l = 1.09$ cm, pressure equal to 0.30 psi, total capillary length equal to 50 cm, and the electric field for carrying out the second passage being $E = 200$ V/cm. These default experimental parameters were used for all experiments in this work except for determination of k_{off} in which $E' = 120$ V/cm (see **Figure 2.7**, right) was used along with $E = 200$ V/cm.

The results of the peak-area recovery experiment are shown in **Figure 2.8**. The following five parameters were directly found from this time dependence: $t_1 = 171$ s, $t_2 = 56.0$ s, $t_3 = 201$ s, $A_1 = 8.69 \times 10^7$ a.u., and $A_2 = 3.11 \times 10^8$ a.u. Then, the value of $v_{\text{av}} = 0.565$ mm/s was found with **Eq. 16**; the value of $d_1 = 3.16$ cm was found with **Eq. 17**; the value of $v_{\text{PL}} = 0.0157$ mm/s was found with **Eq. 19**. The ratio between the velocity-corrected areas was found to be $A_1 v_{\text{av}} / (A_2 v_{\text{PL}}) = 1.01$ for the results shown in **Figure 2.8**. The experiment was repeated 10 times and the average value was found to be $A_1 v_{\text{av}} / (A_2 v_{\text{PL}}) = 1.00 \pm 0.01$ (see **Appendix B3**) suggesting perfect recovery of the velocity-corrected peak areas. It is important to emphasize that this ratio

may be greater than unity in case of significant photobleaching of the fluorophore during the first passage. GFP is photostable and its photobleaching under our illumination conditions was negligible.

2.2.3.5. Recovery of R

To prove the recovery of R experimentally, we modelled equilibrium mixtures of a protein–oligonucleotide binding pair by mixing GFP and a fluorescently-labeled DNA at known values of $[\text{DNA}]/[\text{GFP}]$ ranging from 0.08 to 4. The quantum yield of GFP with respect to that of DNA was determined by identically injecting and pressure-propagating identical-length plugs of equimolar solutions of GFP and DNA, measuring peak areas A_{GFP} and A_{DNA} , and finding their ratio in 10 repetitions (**Appendix B4**):

$$\varphi = A_{\text{GFP}} / A_{\text{DNA}} = 0.11 \pm 0.01 \quad (28)$$

A double-passage experiment with GFP–DNA mixtures was carried out using the default experimental parameters. Time dependencies of the fluorescence signal are shown in **Figure 2.9a**. The values of t_1 , t_2 , t_3 , A_1 , and A_2 were found directly from these time dependencies (**Appendix B4**). The values of secondary parameters v_{av} , d_1 , and v_{PL} were calculated with **Eqs. 16, 17 and 19**, respectively (**Appendix B4**). The found values of φ , A_1 , A_2 , v_{av} , and v_{PL} were used in **Eq. 20** to calculate the values of R which were plotted against the actual values of R (**Figure 2.9b**). The recovery of R proved to be satisfactory for the whole its range. The lower recovery at low concentrations of DNA were due to DNA adsorption on the walls of sample tubes.⁹⁸

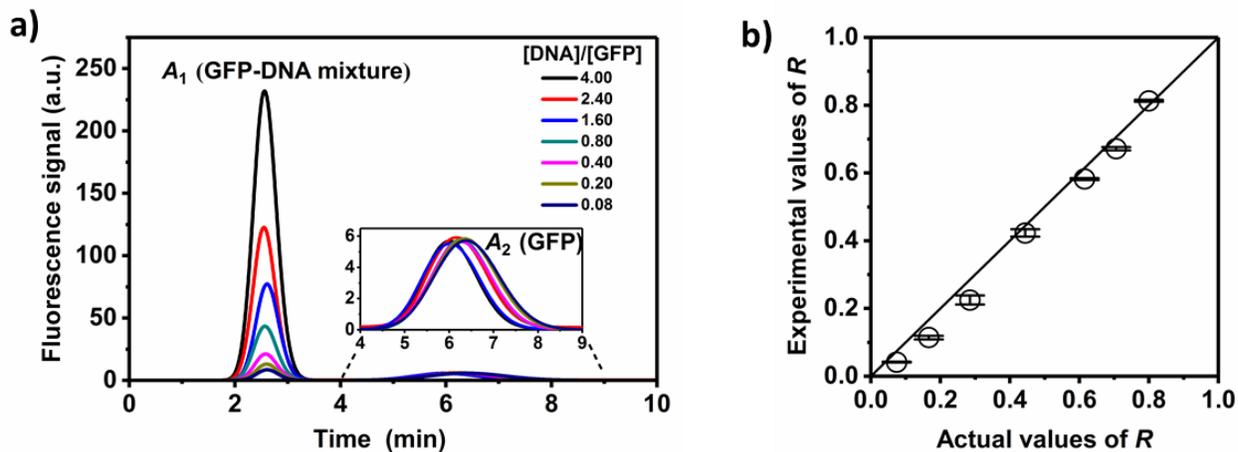


Figure 2.9. Recovery of R for double-passage experiments of GFP–DNA mixtures with $[\text{DNA}]/[\text{GFP}]$ ranging from 0.08 to 4. Panel **a** shows time dependencies of fluorescence signal for different $[\text{DNA}]/[\text{GFP}]$. Panel **b** shows the comparison between the experimental and the actual values of R .

2.2.3.6. Experimental determination of K_d and k_{off} of protein–aptamer complex

In the final demonstration of analytical utility of the double-passage approach, we used an interacting pair of a protein (MutS) and its previously selected and characterized DNA aptamer.^{34,95} An equilibrium mixture of MutS and the aptamer was prepared by mixing $[\text{MutS}]_0 = 0.50$ nM with $[\text{aptamer}]_0 = 0.20$ nM and incubating the mixture for 1 h. The mixture was sampled for the double-passage experiment using the above-defined default experimental parameters. A typical time dependence of the fluorescence signal is shown in **Figure 2.10**, right. The experiment was repeated 3 times and the values of t_1 , t_2 , t_3 , A_1 , and A_2 were found directly from these time dependences (**Appendix B5**). The values of v_{av} , d_1 , and v_{PL} were then calculated with **Eqs. 16, 17** and **19**, respectively (**Appendix B5**). The found values of A_1 , A_2 , v_{av} , and v_{PL} were used in **Eq. 20** to calculate $R = 0.53 \pm 0.01$ under an assumption of $\varphi = 1$. The value of

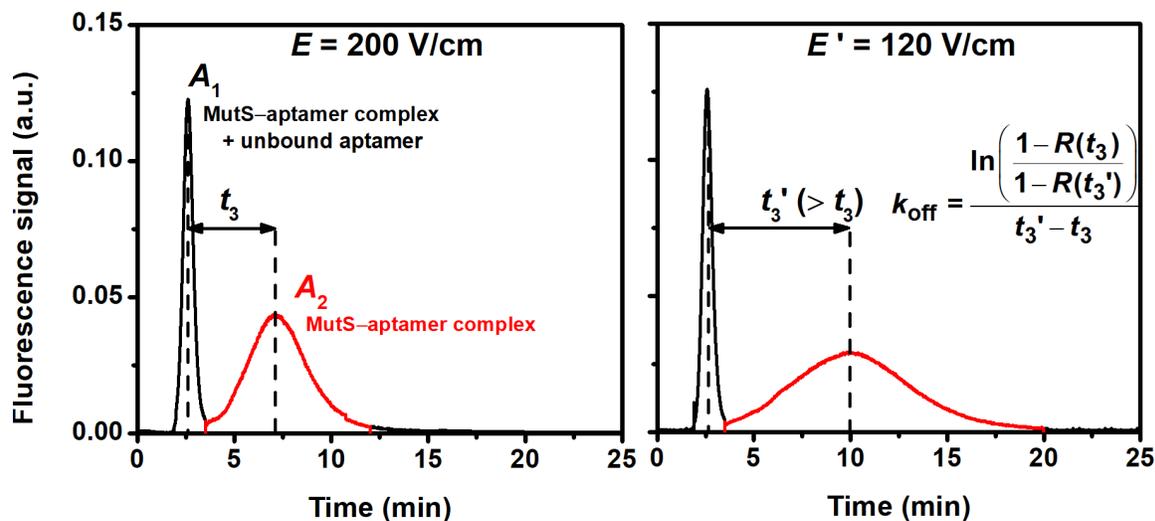


Figure 2.10. Double-passage experiment with a mixture of MutS protein and its aptamer under different electric field strengths: 200 and 120 V/cm. The sample mixture contained 0.50 nM MutS and 0.20 nM aptamer. The experiment was carrying out under using the default experimental parameters (see **Section 2.2.2.2.** CE instrumentation and conditions).

$K_d = 0.47 \pm 0.03$ nM was then calculated with **Eq. 7**. The value of $\varphi = 0.80 \pm 0.01$ was further calculated as a ratio between A_1 determined for near-pure complex and A_1 determined for pure aptamer (**Appendix B5**). A more accurate value of $R = 0.48 \pm 0.01$ was calculated with **Eq. 20** using $\varphi = 0.80$. A φ -corrected value of $K_d = 0.36 \pm 0.02$ nM was calculated by using $R = 0.48$ in **Eq. 7**.

The value of k_{off} was determined by using varying t_3 via changing the electric field strength. Along with the default value of $E = 200$ V/cm we used a lower value of $E' = 120$ V/cm (**Figure 2.10**). Three experiments were carried out to find the values of $t_3 = 225 \pm 13$ s for $E = 200$ V/cm and $t_3' = 375 \pm 17$ s for $E' = 120$ V/cm. The value of $k_{\text{off}} = (1.1 \pm 0.5) \times 10^{-3} \text{ s}^{-1}$ was calculated with **Eq. 22**.

We then carried out a quick test for our experiment's satisfying the inequality of **Eq. 11**. The value of $v_L = 0.059$ mm/s was determined by electrophoretically running a plug of pure L with a cathode at the inlet (**Appendix B6**) and using **Eq. 24**. The value of $t_{\text{sep}} = 54$ s was calculated with **Eq. 13**. The value of $k_{\text{on}} = (3.0 \pm 1.2) \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ was calculated with **Eq. 23** by using the determined values of k_{off} and K_d . Finally, the value of $t_{\text{rebind}} = 670$ s was estimated with **Eq. 14**. Comparing t_{sep} and t_{rebind} suggests that the inequality of **Eq. 11** was satisfied under our experimental conditions ($54 \ll 670$) and, thus, the rebinding did not affect our determination of k_{off} .

Knowing k_{off} allowed us to make the final adjustment of K_d by taking into account the dissociation of PL during its electrophoretic migration to and through the detection window. Peak area $A_{2,t=0}$ corrected for this dissociation is:

$$A_{2,t=0} = A_{2,t_3} e^{k_{\text{off}} t_3} \quad (29)$$

Using the corrected value of A_2 founds from **Eq. 29** in **Eq. 20** for finding R and subsequently using the corrected R in **Eq. 7** we found $K_d = 0.20 \pm 0.02$ nM corrected for complex dissociation during the electrophoretic migration of PL to and through the detection window. The values of $K_d = 0.20$ nM and $k_{\text{off}} = 1.1 \times 10^{-3} \text{ s}^{-1}$ determined by us under the IFCE conditions (ionic strength of 150 mM and pH 7.0) are very close to those determined for the same binding pair but at a much lower ionic strength of 13 mM and higher pH of 8.5: $K_d = 0.1$ nM and $k_{\text{off}} = 0.4 \times 10^{-3} \text{ s}^{-1}$.^{34,95} This similarity suggests relative insensitivity of MutS–aptamer binding to ionic strength and pH.

2.2.4. Conclusions

To conclude, we developed the double-passage approach for determination of K_d and k_{off} of protein–oligonucleotide complex under the IFCE conditions (near-physiological ionic strength and pH). This approach is needed for assessing stability of protein–oligonucleotide complexes selected by IFCE. It will also be useful for assessing stability under the near-physiological conditions of protein–oligonucleotide complexes selected by NECEEM under non-physiological conditions.

CHAPTER 3. QUANTITATIVE CHARACTERIZATION OF PARTITIONING IN SELECTION OF BINDERS FOR PROTEIN TARGETS FROM OLIGONUCLEOTIDE LIBRARY BY CAPILLARY ELECTROPHORESIS

The presented material was published previously and reprinted with permission from “**Le, A.T.H.**; Wang, T.Y.; Krylova, S.M.; Beloborodov, S.S.; Krylov, S.N. Quantitative Characterization of Partitioning in Selection of DNA Aptamers for Protein Targets by Capillary Electrophoresis. *Analytical Chemistry* **2022**, *94*, 2578–2588”. Copyright 2022 American Chemical Society. My contributions to the article were: (i) performing all presented experiments, (ii) preparing all figures, (iii) interpreting the results and (iv) writing the manuscript.

3.1. IMPORTANCE OF DETERMINING THE APPROPRIATE KCE-BASED PARTITIONING METHOD IN BINDER SELECTION

In the preceding chapter, we introduced IFCE into the KCE-based partitioning toolbox alongside the two known NECEEM-based partitioning modes: complex-first NECEEM and complex-last NECEEM. Each of these methods exhibits unique characteristics, defined by two well-established partitioning parameters: the transmittance for protein-bound DNA (binders) and the transmittance for unbound DNA (nonbinders). Understanding the distinctions in these fundamental partitioning parameters among different KCE methods is critical for effective planning and control of selection processes. The lack of such studies is arguably the major reason for KCE-based selection still being more an art than science. Here, by combining a theoretical model and experimental data, we present the first comprehensive study demonstrating the variation of these fundamental partitioning parameters across different KCE methods, regarding

different protein targets. This study aims to provide guidance to KCE users in selecting the most appropriate partitioning methods and conditions for successful binder selection from oligonucleotide libraries.

Partitioning is evidently a key step in binder selection — increasing the efficiency of partitioning allows completion of binder selection in fewer rounds and can help avoid selection failures.⁵⁵ In **Chapter 2**, we quantitatively described partitioning using the term “transmittance”, originally employed in spectrophotometry to characterize spectral filters, but can be generalized to any filter. In general, the transmittance of partitioning for binders (B), k_B , is defined as the ratio between quantities of binders at the output, B_{out} , and input, B_{in} , of partitioning (see **Eq. 3**). Accordingly, transmittance of partitioning for nonbinders (N), k_N , is defined as the ratio between the quantities of nonbinders at the output, N_{out} , and input, N_{in} , of partitioning (see **Eq. 4**).

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

$$\frac{B_{out}}{N_{out}} = \frac{B_{in}}{N_{in}} \frac{k_B}{k_N} \quad (30)$$

When optimizing partitioning, one needs to minimize k_N without proportionally decreasing k_B . Knowing how k_N and k_B depend on multiple experimental parameters for a specific method of partitioning is, therefore, pivotal to planning efficient aptamer selection and to controlling it.

Surprisingly, there is a lack of quantitative studies dedicated to the factors affecting k_N and k_B . This work results from our effort to initiate such studies and make quantitative characterization of partitioning a foundation for technological advancement of aptamer selection. The efficiency and versatility of KCE-based partitioning make it a prime subject for the first comprehensive study on quantitative partitioning characterization in binder selection. We employ a combination of theoretical modeling and experimentation to understand the underlying concept and obtain empirical information required for this study to be instructive for practical users.

In essence, here we study the dependence of k_N and k_B on the molecular weight of protein target for varying pH and ionic strength of the running buffer. For a given protein size, the optimal running buffer conditions should guarantee low k_N and high k_B for a higher chance of successful selection. When the running buffer has lower than physiological ionic strength and/or higher than physiological pH, the target–binder complexes and nonbinders move in the same direction; this mode of partitioning is known as nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM).³¹ We found that in NECEEM, k_N increases by several orders of magnitude while k_B remains close to unity when the molecular weight of the protein target decreases. When the running buffer has both ionic strength and pH near physiological, the target–binder complexes and nonbinders move in the opposite directions. Such conditions facilitate so-called ideal-filter capillary electrophoresis (IFCE).⁴² IFCE is characterized by the lowest k_N values on record (10^{-9}) achieved at the expense of large decrease in k_B . With decreasing molecular weight of the protein target, k_N does not change much while k_B decreases by as much as orders of magnitude in IFCE. The decrease in k_B discourages from using IFCE conditions for small-size protein targets. Our results suggest that IFCE conditions are most suitable for large-size protein targets to obtain high affinity binders in a minimal number of

partitioning rounds. When the non-SELEX approach is used, losses of binders are significant between partitioning rounds, and these losses cannot be compensated as there is no PCR amplification in between the rounds.²⁶ Therefore, in non-SELEX, NECEEM (in which k_B is close to unity) is preferred over IFCE to retain a sufficient quantity of binders for the next rounds of partitioning and the concluding PCR amplification.

3.2. MATERIALS AND METHODS

3.2.1. Materials and solutions

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

The CE running buffers were 50 mM Tris-Acetate at pH 8.2 without NaCl for NECEEM and 50 mM Tris-HCl at pH 7.0 supplemented with 100 mM NaCl for IFCE, resulting in ionic strength of the running buffer (I_{RB}) of 25 and 146 mM, respectively. The sample buffer was always the same as the running buffer to prevent adverse effects of buffer mismatch. Accordingly, all dilutions of sample components used in CE experiments were done by adding the corresponding buffer.

We used a synthetic FAM-labeled DNA library (N40) with a 40-nt random region: 5'-FAM-CT ACG GTA AAT CGG CAG TCA-(N40)-AT CTG AAG CAT AGT CCA GGC-3'. The nucleotide sequence of the forward primer was 5'-CTA CGG TAA ATC GGC AGT CA-3', and

the sequence of the reverse primer was 5'-GCC TGG ACT ATG CTT CAG AT-3'. All solutions were prepared in deionized water filtered through a 0.22- μm Milipore filter membrane (Nepean, ON).

3.2.2. CE instrumentations and conditions

All CE experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detection system. Fluorescence was excited with a blue line (488 nm) of a solid-state laser and detected at 520 nm using a spectrally optimized emission filter system. All capillaries were 50-cm long (40 cm to the detector) and had an inner diameter of 75 μm and an outer diameter of 360 μm . The poly(vinyl alcohol) (PVA)-coated capillary was prepared as described elsewhere.⁹⁹ Prior to every fraction-collection experiment, a new capillary was installed and conditioned successively with MeOH at 20 psi for 10 min, 0.1 M HCl at 20 psi for 3 min, 0.1 M NaOH at 10 psi for 6 min, water at 20 psi for 3 min and a running buffer at 40 psi for 40 min. Prior to every run, the capillary was rinsed successively with 0.1 M HCl, 0.1 M NaOH, deionized H₂O, and a running buffer for 3 min each. Conditioning steps were not required for PVA-coated capillaries; such capillaries were rinsed with the running buffer only at 20 psi for 10 min prior to the fraction-collection experiment.

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

CE was carried out at an electric field of 200 V/cm (10 kV over 50 cm). CE-run duration was 34 min for NECEEM conditions and 128 min for IFCE. For uncoated capillaries, CE was carried out with the positive electrode at the injection end of the capillary; for PVA-coated capillaries, the polarity was reversed. Collection vials contained 20 μ L of the running buffer each and were switched every 4 min for IFCE and every 1 min for NECEEM.

3.2.3. Quantitative PCR

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

3.3. THEORETICAL AND EXPERIMENTAL CONSIDERATIONS

3.3.1. Dependence of number of partitioning rounds on k_N & k_B

This analysis is general and does not depend on the type of partitioning method. Let us consider major conditions to be satisfied to ensure successful selection. PCR amplification of collected DNA is always used as a final step of binder selection before DNA sequencing. If

selection is done in a single round, then the quantity of binders at the output of partitioning, B_{out} , must exceed the level of PCR noise (N_{PCR}), produced during the PCR amplification of the selected binders, by a set number $Q_1 > 1$:

$$B_{\text{out}} > Q_1 N_{\text{PCR}} \quad (31)$$

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

$$B_{\text{out}} / N_{\text{out}} > Q_2 \quad (32)$$

A single round of selection is rarely sufficient to satisfy the inequality in **Eq. 32**. If multiple consecutive rounds of selection are conducted without PCR amplification between them (applicable to non-SELEX), then the number of rounds m should, in turn, satisfy two conditions. Firstly, m should be not too high to prevent the excessive loss of binders, i.e., satisfy inequality in **Eq. 31**:

$$m \leq \left\lfloor \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log(k_B)} \right\rfloor \quad (33)$$

where $\lfloor x \rfloor$ represent a mathematical function that rounds x down to the nearest integer (see **Appendix C1** for derivation of **Eq. 33**). Secondly, m should be high enough to satisfy inequality in **Eq. 32**:

$$m \geq \left\lceil \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log(k_B / k_N)} \right\rceil \quad (34)$$

where $\lceil x \rceil$ represent a mathematical function that rounds x up to the nearest integer (see **Appendix C1** for derivation of **Eq. 34**). The last two inequalities establish the range of acceptable values of m :

$$\left\lfloor \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log(k_{\text{B}})} \right\rfloor \geq m \geq \left\lceil \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log(k_{\text{B}} / k_{\text{N}})} \right\rceil \quad (35)$$

The values of Q_1 and Q_2 may be set taking into account secondary considerations; for Q_2 it may be, for instance, the cost of post-selection screening.⁵⁵ The values of k_{B} and k_{N} depend on the partitioning method of choice and can be estimated or determined experimentally. The values of B_{in} and N_{in} are not known for real selections (unlike mock selections in which known binders are spiked controllably into known nonbinders), but different scenarios can be considered for them conclusively when analyzing any specific selection case.

In the case of classic aptamer selection by SELEX, PCR is used between the consecutive rounds to maintain the quantity of binders. Therefore, m does not have an upper limit, but there is still a lower limit of m :

$$m \geq \left\lceil \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log\left((k_{\text{B}} / k_{\text{N}})(Z_{\text{B}} / Z_{\text{N}})^n\right)} \right\rceil \quad (36)$$

where Z_{B} and Z_{N} are the bases of the exponent describing PCR amplification of binders and nonbinders, respectively. In an unbiased amplification, binders and nonbinders are amplified with the same efficiency, i.e., $Z_{\text{B}} = Z_{\text{N}}$. It is likely, however, that $Z_{\text{B}} < Z_{\text{N}}$ (due to the more folded structure of aptamers), which imposes an upper limit for the number of PCR cycles:²³⁻²⁵

$$n < \left\lfloor \frac{\log(k_{\text{N}} / k_{\text{B}})}{\log(Z_{\text{B}} / Z_{\text{N}})} \right\rfloor \quad (37)$$

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

Accurately assessing the limits for m and n using **Eqs. 35, 36, and 37** *a priori* is impossible due to the uncertainties in B_{in} , N_{in} , Z_{B} , and Z_{N} . However, some quantitative analysis of these

limits can be conducted upon reasonable assumptions for the values of B_{in} , N_{in} , Z_B , and Z_N , and instructive conclusions can be made.

3.3.2. Key differences within the binder-collection windows in different KCE-based partitioning methods

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

In both NECEEM sub-modes, the electropherograms contain all three features (the two peaks and the bridge between them); only the order of the peaks and the direction of the bridge change (**Figure 3.1a, c**). In IFCE, P–DNA moves towards the collection end of the capillary while all unbound DNA (including binders dissociated from P–DNA during partitioning) moves in the opposite direction resulting in an electropherogram which contains only the peak of intact complex (**Figure 3.1b**).

The purpose of partitioning is to collect binders and reject nonbinders. In KCE-based partitioning, this is achieved by collecting a sample fraction at the capillary outlet in a specific binder-collection window. In NECEEM, the binder-collection window can cover both intact P–

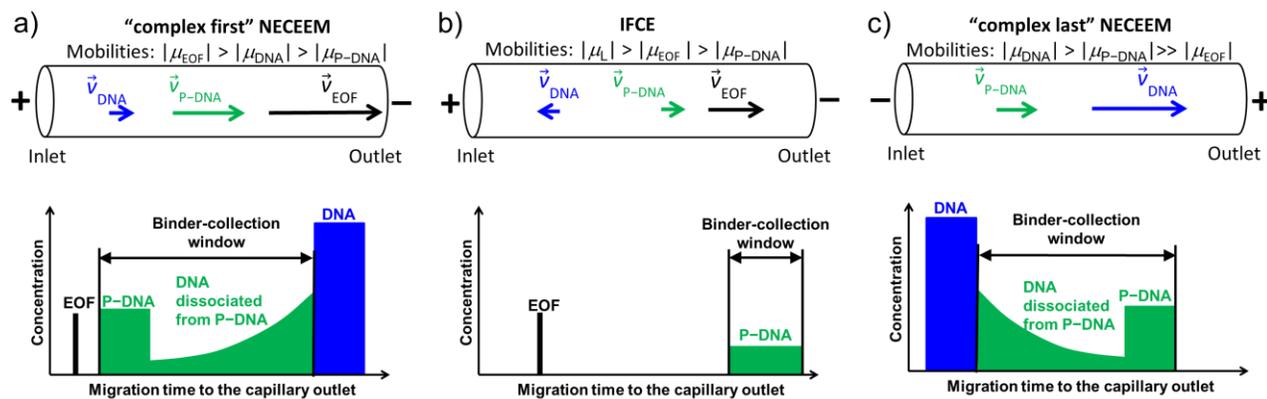


Figure 3.1. Schematic representation of the binder-collection window in different modes of KCE-based partitioning: **a)** “complex first” NECEEM, **b)** IFCE and **c)** “complex last” NECEEM. The EOF bar indicates migration time of an EOF marker (a neutral molecule). See text for details.

DNA and DNA dissociated from P–DNA during CE separation (see **Figure 3.1a, c**). In IFCE, the binder-collection window can only cover the intact P–DNA (see **Figure 3.1b**) as the bridge moves along with the peak of nonbinders towards the injection end of the capillary.

The binder-collection window in NECEEM includes a tail of the unbound DNA peak which constitutes the nonbinder background (DNA background). In contrast, IFCE appears to be free of this effect based on the fundamentals of CE separation. In reality, both NECEEM and IFCE partitioning always have DNA background in the binder-collection window due to a phenomenon of non-uniform migration of DNA in a uniform electric field.⁵² The phenomenon is hypothetically caused by the effect of the electric field on very stable complexes of DNA with counterions. There is no quantitative theory of this effect that could help to predict the level of the DNA background in KCE-based partitioning; therefore, this background should be studied empirically.

Sequential transition from complex-first NECEEM to IFCE and then to complex-last NECEEM is achieved by reducing the mobility of the electroosmotic flow (EOF, μ_{EOF}). The top part of **Figure 3.1** schematically shows directions of velocities of P-DNA and DNA along with relations between $\mu_{\text{P-DNA}}$, μ_{DNA} , and μ_{EOF} for these three cases. The value of μ_{EOF} for a bare fused-silica capillary depends on pH of the running buffer and its ionic strength I_{RB} : lowering pH and/or increasing I_{RB} of the running buffer lead to decreasing μ_{EOF} . Coating the inner wall of the capillary with a non-ionizable layer suppresses EOF and can lead to $\mu_{\text{EOF}} \ll \mu_{\text{P-DNA}}, \mu_{\text{DNA}}$ and, thus, to complex-last NECEEM for a broad spectrum of values of running buffer pH and I_{RB} . The following two sections consider what k_{N} and k_{B} depend on in KCE-based partitioning. The sole purpose of this consideration is to assist in rationally designing our empirical study.

3.3.3. Parameters influencing k_{N} and k_{B} in KCE-based partitioning

The value of k_{N} is a function of analytical resolution R_{s} of the peaks of P-DNA and DNA (k_{N} decreases with increasing R_{s}).⁵¹ The value of R_{s} , in turn, depends on $\mu_{\text{P-DNA}}$, μ_{DNA} , and time of separation (or elution time) t . The value of $\mu_{\text{P-DNA}}$ greatly depends on 1) the size of the protein, which is linked with its molecular weight (MW_{P}), 2) the length of DNA (L_{DNA}), which is the same for binders and nonbinders, and 3) pH and ionic strength I_{RB} of the running buffer. Notably, the dependence of $\mu_{\text{P-DNA}}$ on the charge of the protein is negligible in the first approximation because of the much higher charge density on DNA.⁸³ The value of μ_{DNA} depends on I_{RB} , and slightly depends on L_{DNA} in gel-free CE.¹⁰⁰ As we described in the previous section, the value of k_{N} is greatly influenced by the nonbinder background (DNA background) caused by the nonuniform mobility of DNA in KCE.⁵² The background itself depends (directly or

indirectly) on pH, I_{RB} , MW_P , and t . Cumulatively, the value of k_N is a function of five parameters:

$$k_N = F(R_S) = F(\mu_{P-DNA}, \mu_{DNA}, t) = F(MW_P, L_{DNA}, t, pH, I_{RB}) \quad (38)$$

There may be some cross-influence of the parameters, e.g., the choice of t may depend on MW_P , but such nuances do not affect the essence of **Eq. 38**, and, therefore, are not a subject of this conceptual consideration.

The value of k_B depends on whether the aptamer-collection window covers the entire span of binders in the profile (**Figure 3.1**). NECEEM and IFCE are radically different with regards to k_B . In theory, k_B decreases with time exponentially with a rate constant k_{off} of dissociation of P–DNA complexes. In NECEEM, both intact P–DNA complexes and DNA dissociated from P–DNA during CE migrate in the same direction. Therefore, nearly all binders can be collected, and k_B in NECEEM can be assumed to be close to unity: $B_{out} \approx B_{in}$. In IFCE, P–DNA moves towards the collection end of the capillary, but the unbound DNA migrates in the opposite direction. As a result, only the intact complexes are collected while binders dissociated from the complexes during CE are not. P–DNA dissociates following the monomolecular decay: $B_{out} = B_{in} e^{-k_{off}t}$. The time t during which P–DNA is allowed to dissociate before elution (i.e., elution time) is defined by μ_{P-DNA} and, thus, depends on MW_P and L_{DNA} . A mixture of binders with different k_{off} values is not characterized by a specific k_{off} value; therefore, we use k_{off} here as a loose term. Thus, we can write for k_B in NECEEM and IFCE, respectively:

$$\begin{aligned} k_{B,NECEEM} &= const \approx 1 \\ k_{B,IFCE} &= F(k_{off}, MW_P, L_{DNA}, pH, I_{RB}, t) \end{aligned} \quad (39)$$

As seen from **Eq. 38** and **39**, k_N and k_B in KCE-based partitioning are defined by a total of 6 parameters: MW_P , L_{DNA} , k_{off} , t , pH, and I_{RB} . Rational design of KCE-based aptamer selection,

thus, requires an experimental study that would lead to understanding how these parameters affect k_N and k_B .

3.3.4. Rationale for experimental design

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

Advantageously, the study of how MW_P affects the values of k_N and k_B for varying pH and I_{RB} can be conducted without using proteins as there is a recently published empirical function that links the mobility of the protein–DNA complex with the molecular weight of the complex (MW_{P-DNA}):⁸³

$$\mu_{P-DNA} = A + B\mu_{DNA}L_{DNA}^{0.68}MW_{P-DNA}^{-1/3} \quad (40)$$

where electrophoretic mobilities are expressed in $\text{mm}^2/(\text{kVs})$, L_{DNA} is expressed in the number of nucleotides, MW_{P-DNA} is expressed in kDa, while A and B are empirical constants. For a running

buffer with $I_{\text{RB}} < 50$ mM, these constants are $A = -9.95 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.0929 \text{ kDa}^{1/3}$. For a running buffer with $I_{\text{RB}} = 146$ mM, the constants are $A = 10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.2365 \text{ kDa}^{1/3}$ (see **Appendix C2**). As $MW_{\text{P-DNA}} = MW_{\text{P}} + MW_{\text{DNA}}$, the predicted mobility values obtained with **Eq. 40** are used to estimate the associated velocity and the elution time of protein–DNA complex for given value of MW_{P} (see **Appendix C3**).

3.4. RESULTS AND DISCUSSION

3.4.1. Migration profiles of DNA in different modes of KCE-based partitioning

To evaluate DNA background in the binder-collection window, we experimentally obtained DNA migration profiles in different modes of KCE-based partitioning: both sub-modes of NECEEM and IFCE. The high value of EOF required for complex-first NECEEM can be achieved in the buffer system with low I_{RB} and/or high pH (typical $I_{\text{RB}} < 50$ mM) in a bare fused-silica capillary. In IFCE, the EOF is reduced by using buffer with high I and/or low pH (typical $I_{\text{RB}} > 100$ mM) in a bare fused-silica capillary. For complex-last NECEEM, the EOF is suppressed *via* coating the inner wall of capillary (e.g., with PVA). In this study, we chose two previously published NECEEM and IFCE running buffers to generate qualitatively distinct migration profiles of DNA.^{31,42} The first buffer was 50 mM Tris-Acetate pH 8.2 ($I_{\text{RB}} = 21$ mM), corresponding to NECEEM, in which P–DNA and DNA move in the same direction towards the capillary outlet. For consistency, we used the same buffer for both complex-first NECEEM in a bare-silica capillary and complex-last NECEEM in a PVA-coated capillary. The second buffer was 50 mM Tris-HCl pH 7.0 ($I_{\text{RB}} = 146$ mM), corresponding to IFCE, in which P–DNA moves to the outlet while DNA moves to the inlet in a bare-silica capillary.

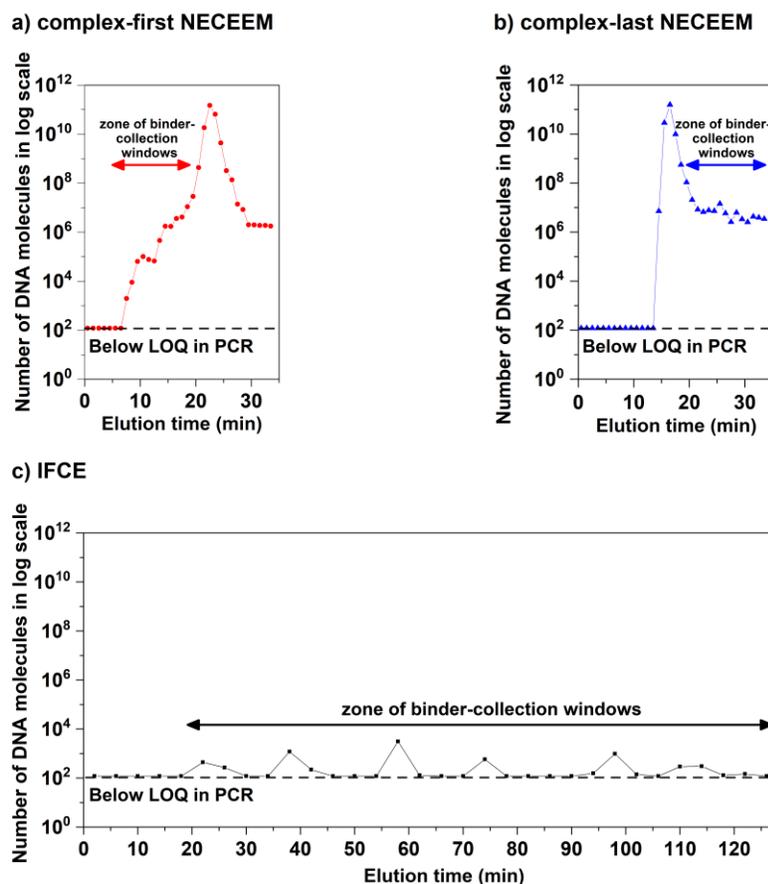


Figure 3.2. DNA background profiles under conditions of NECEEM (a and b) and IFCE (c). A sample of 2.8×10^{11} molecules of 80-nt DNA was subjected to KCE-based partitioning. Fractions were collected every 1 min in NECEEM and every 4 min in IFCE. Concentration of DNA in every fraction was quantitated using qPCR, and these quantities are shown on the y-axis in the graph. The double-arrow indicates the appropriate zone of binder-collection windows in corresponding mode of CE-based partitioning.

The sample of 2.8×10^{11} molecules of 80-nt DNA was subjected to both NECEEM and IFCE. One- and four-minute fractions were collected for NECEEM and IFCE, respectively; all collected fractions were analyzed by qPCR to build a “DNA quantity versus migration time to the capillary outlet” electropherogram. The results are shown in **Figure 3.2**.

In complex-first NECEEM, the main DNA peak eluted in 23 min after the start of CE (**Figure 3.2a**). A small part of DNA sample migrated in front of the main DNA peak and created the nonbinder background of a total of approximately 10^7 molecules in the zone of NECEEM binder-collection windows. As I_{RB} increased to reach the IFCE condition, the migration direction of the main DNA peak switched from the direction towards the outlet end of the capillary to the opposite direction (towards the inlet end) while the nonbinder background became stretched out. In IFCE, the nonbinder background was reduced to below the limit of detection of qPCR for a 2-h binder-collection window (**Figure 3.2c**). In complex-last NECEEM, the main DNA peak moved ahead of the binder-collection window and eluted in 17 min, while the nonbinder background of a total of approximately 10^7 molecules tailed behind the main DNA zone in the binder-collection window (**Figure 3.2b**).

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

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

3.4.2. Influence of MW_P on k_N

Knowing the predicted mobility of protein–DNA complex (**Eq 40**) allows one to calculate the binder-collection window in which P–DNA should elute from the capillary (see **Appendix C3**). Knowing this time window, in turn, allows the determination of the transmittance of partitioning for nonbinders k_N . The value of k_N was calculated based on **Eq. 4** as the total number of background DNA (N_{out}) collected within the binder-collection window divided by the total number of DNA injected into the capillary (N_{in}).

The resulting dependence of k_N on MW_P ranging between 25 to 150 kDa is shown in **Figure 3.3**. In both NECEEM sub-modes, as MW_P decreased from 150 to 25 kDa, k_N increases approximately 3 orders of magnitude. As expected, k_N values for complex-last NECEEM are higher than those for complex-first NECEEM due to the elevated DNA background in complex-last NECEEM (**Figure 3.2**). In IFCE, the background profile is stable with k_N remaining near 10^{-9} throughout the 2-h run; thus, k_N is similar for target proteins with different MW_P . In general, the IFCE running buffer with higher ionic strength and/or lower pH leads to lower k_N values as

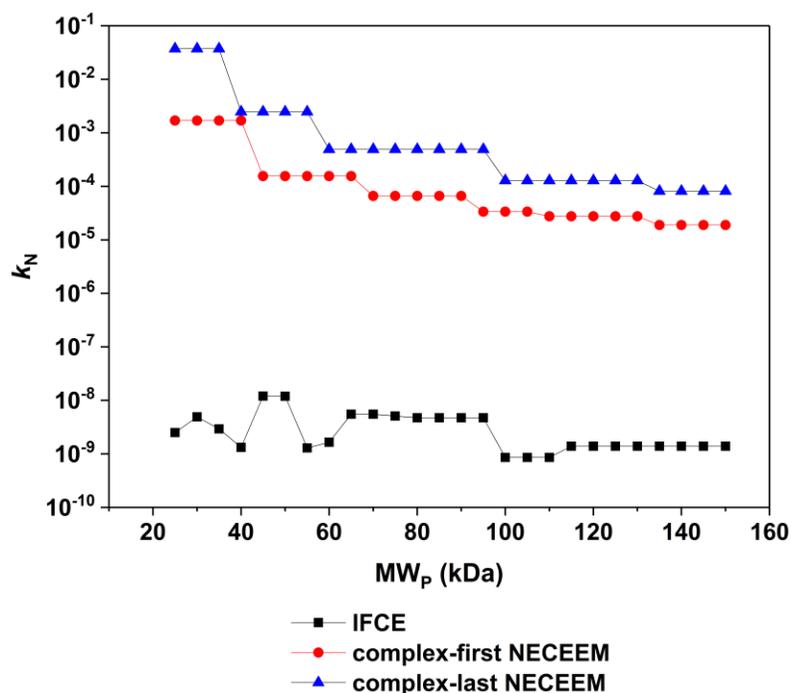


Figure 3.3. The effect of the molecular weight of protein (MW_P) on the transmittance of KCE-based partitioning for nonbinders (k_N) under the conditions of NECEEM (colored lines) and IFCE (black line).

well as weaker dependence of k_N on MW_P . However, the low EOF obtained with such a running buffer increases the predicted complex migration time to over 3 h for smaller proteins with $MW_P < 25$ kDa (see **Appendix C4**). Note that the effect of MW_P on k_N cannot be measured experimentally for target–binder complexes with $MW_P < 25$ kDa under IFCE conditions due to an unreasonably long CE run. Given such a high stringency of partitioning (very long separation), dissociation of protein–DNA complex might reduce the level of intact complex to below the noise of PCR. Therefore, one has to adjust stringency of partitioning to achieve reasonably low DNA background while maintaining a sufficient quantity of intact protein–DNA complexes. The extent of P–DNA dissociation in different KCE-based partitioning modes will be evaluated in the next section.

3.4.3. Influence of MW_P on k_B

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

$$k_B = \frac{B_{\text{out}}}{B_{\text{in}}} = \frac{B_{\text{in}} e^{-k_{\text{off}} t}}{B_{\text{in}}} = e^{-k_{\text{off}} t} \quad (41)$$

The values of t as well as the binder-collection windows for complexes of 80-nt DNA and protein targets of different MW_P were estimated using the predicted mobility values of protein–DNA complexes obtained with **Eq. 40**. The transmittance of partitioning for binders as defined by **Eq. 41** was estimated for two values of k_{off} , 10^{-3} and 10^{-4} s^{-1} , for which the characteristic complex-dissociation times ($\tau = 1/k_{\text{off}}$) are 20 min and ~ 3 h, respectively. The values of k_{off} outside of the 10^{-4} – 10^{-3} s^{-1} range are more likely to be unpreferred for therapeutic targets since the complex either dissociate too quickly ($\tau = 1$ min for $k_{\text{off}} = 10^{-2} \text{ s}^{-1}$) or remain stable for too long ($\tau = 28$ h for $k_{\text{off}} = 10^{-5} \text{ s}^{-1}$).¹⁰¹

The predicted dependence of k_B on MW_P is shown in **Figure 3.4**. For $k_{\text{off}} = 10^{-3} \text{ s}^{-1}$, decreasing MW_P from 150 to 25 kDa leads to decreasing the value of k_B by up to 3 orders of magnitude in IFCE. In contrast, in NECEEM, k_B is not affected by changing MW_P . Notably, in IFCE, k_B is predicted to be lower than 10^{-3} when $MW_P < 30$ kDa, meaning that less than 0.1% of the total quantity of protein–DNA complex in the equilibrium mixture would survive separation until elution.

For $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$, the decrease in k_B is much less pronounced for both IFCE and NECEEM. In both NECEEM sub-modes, k_B remains relatively stable over the specified range of MW_P with more than 80% of protein–DNA complexes reaching the capillary end intact. This finding agrees

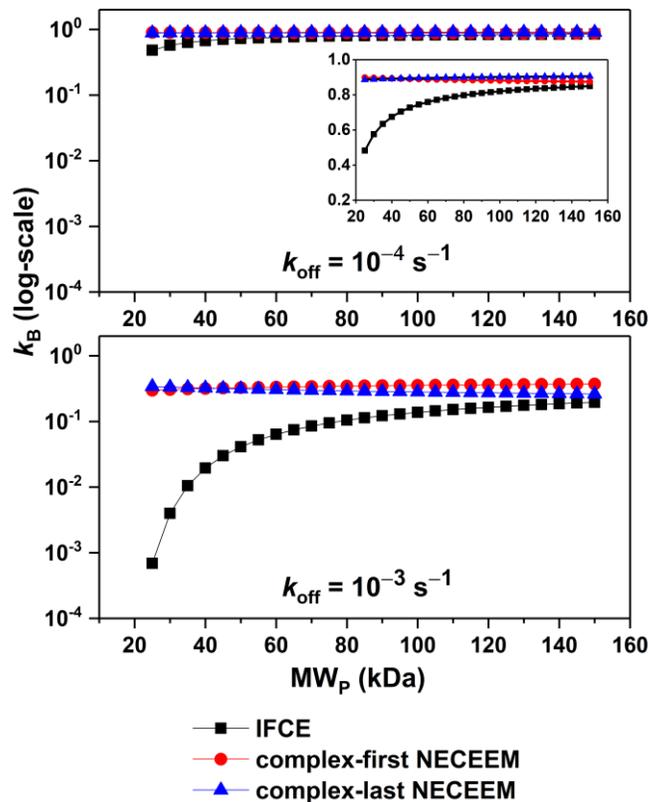


Figure 3.4. The dependency of the transmittance of KCE-based partitioning for binders (k_B) on the molecular weight of target protein (MW_P) under the conditions of NECEEM (colored lines) and IFCE (black lines). The value of k_B was estimated based on two k_{off} values: 10^{-3} s^{-1} and 10^{-4} s^{-1} . In the graph with k_{off} of 10^{-4} s^{-1} , the inset shows the same data but with a linear y-scale for k_B .

with **Eq. 39**, in which the k_B value in NECEEM is assumed to be constant. In IFCE, our prediction shows that up to 60% of protein–DNA complexes would dissociate as the values of MW_P decrease to 25 kDa. Overall, the dissociation of protein–DNA complexes is much less pronounced in NECEEM than in IFCE over the specified range of MW_P . On the other hand, k_N value in IFCE is orders of magnitudes lower than in NECEEM. A balance between k_B and k_N must be achieved in order to obtain a target level of binder purity Q_2 after partitioning.

3.4.4. Influence of MW_P on the number of partitioning rounds

The odds of successful selections depend on two major conditions. The first condition is a high value of efficiency of partitioning (k_B/k_N) to enrich the initial library with a low level of binder abundance (B_{in}/N_{in}) to a desirable level of binder purity at the output ($B_{out}/N_{out} > Q_2$). The second condition is a sufficient quantity of binders at the input (B_{in}) so that the output quantity of binder (B_{out}) can exceed the PCR noise. The efficiency of partitioning for IFCE and NECEEM can be derived from the k_N and k_B values for the two preferred k_{off} values considered above: 10^{-3} and $10^{-4} s^{-1}$. To estimate B_{in}/N_{in} , we hypothesized different scenarios for aptamer selection based on the affinity of random DNA library to the target protein: (i) high abundance of binders ($B_{in}/N_{in} = 10^{-5}$ or approximately 1 binder per 10^5 nonbinders), (ii) moderate abundance of binders ($B_{in}/N_{in} = 10^{-6.5}$ or ~ 1 binder per 3×10^6 nonbinders), and (iii) low abundance of binders ($B_{in}/N_{in} < 10^{-8}$ or less than 1 binder per 10^8 nonbinders).

We then estimated the range of required partitioning rounds (m) to obtain a binder-enriched pool at the output of non-SELEX (without PCR amplification of the collected pools between the rounds) with each of NECEEM sub-modes and IFCE. The upper limit of m (m_{max}) and the lower limit of m (m_{min}) were calculated using **Eqs. 33** and **34**, respectively, for $Q_1 = 100$ (i.e., B_{out} exceeds PCR noise of 120 molecules of DNA by a factor of 100) and $Q_2 = 1$ (i.e., binders constitute 50% of the final DNA pool). In principle, a high value of m_{max} indicates a low level of binder losses throughout the selection process; therefore, many rounds of partitioning can be conducted to further enrich the pool without detrimental losses of binders. On the other hand, a low value of m_{min} is preferable to minimize the number of partitioning rounds required to reach a target level of binder purity in the resulting pool.

cycles.²⁵ Values of m_{\min} for non-SELEX presented in this section are also applicable to SELEX under optimal PCR conditions. Note that when there is a PCR bias, m_{\min} in SELEX will be higher than in non-SELEX.

Figure 3.5 shows the predicted dependence of acceptable numbers of partitioning rounds in IFCE and NECEEM (both sub-modes) on MW_P for 3 different values of binder abundance in the initial library ($B_{\text{in}}/N_{\text{in}}$). Suitability of the partitioning method for selection can be assessed based on the following criteria: (i) a range of m is wide (m_{\max} and m_{\min} points are far apart on the graph), (ii) m_{\max} is greater than m_{\min} (m_{\max} point is not lower than m_{\min} point on the graph) and (iii) m_{\max} is not smaller than 1 (the otherwise case indicates insufficient quantity of binders at the output). Based on **Figure 3.5**, KCE-based selection of aptamers can be successful in the following scenarios: (i) the input library has high binder abundance ($\sim 10^{-5}$) or (ii) the input library has moderate binder abundance ($>10^{-8}$) with a bulk value of $k_{\text{off}} \leq 10^{-4}$. In these scenarios, as MW_P increases from 25 to 150 kDa, the range of m increases in IFCE and complex-first NECEEM. This means that the extent of binder losses is smaller (higher m_{\max}) and fewer partitioning rounds are required (smaller m_{\min}) for larger protein targets in IFCE and complex-first NECEEM. On the other hand, in complex-last NECEEM, both m_{\max} and m_{\min} decrease as MW_P increases from 25 to 150 kDa. In complex-last NECEEM, the elution time of protein–DNA complex increases with increasing MW_P ; thus, the extent of binder losses is higher while fewer partitioning rounds are allowed for larger protein targets. Moreover, as both NECEEM sub-modes have higher range of k_N values (**Figure 3.5**), their range of m_{\min} values is also higher than that in IFCE. This relation means that more partitioning rounds in NECEEM would be required to enrich the pool to a certain level of binder purity as compared to IFCE (particularly, complex-

last NECEEM with the highest range of m_{\min} values requires more partitioning rounds than complex-first NECEEM).

Despite the low value of k_B for small protein targets, the extremely low k_N values in IFCE suggest that IFCE could support the enrichment of binders in a single step of partitioning (50% binders in the resulting pool) over the whole range of MW_P . However, in IFCE, low values of m_{\max} were observed for small-size protein targets, indicating that excessive losses of binders can potentially hinder the success of IFCE-based selection of aptamers for such targets. In some cases, the m_{\max} value in IFCE were unacceptable ($m_{\max} \leq m_{\min}$ and/or $m_{\max} \leq 1$) for protein targets with the following ranges of molecular weights: (i) $MW_P < 60$ kDa for the input library with high binder abundance and $k_{\text{off}} = 10^{-3} \text{ s}^{-1}$, and (ii) $MW_P < 20$ kDa for the input library with higher binder abundance and $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$ and (iii) $MW_P < 25$ kDa for the input library with moderate binder abundance and $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$. The latter two were extrapolated by assuming a constant k_N in the range of 10^{-9} for protein–DNA complex with MW_P ranging from 15 to 25 kDa in IFCE (see **Appendix C5**).

To ensure the collection of a sufficient quantity of binders at the output, NECEEM is the method of choice for selection of binders for small-size protein targets. This statement is especially true in non-SELEX, in which no PCR amplification is used between the rounds of partitioning to compensate for the dilution-induced losses of binders between the rounds. Our results suggest that IFCE is not preferred for small-size protein targets due to the excessive binder losses within every round of partitioning owing to complex dissociation. However, IFCE is the most suitable method for large-size protein targets to obtain high affinity binders in a minimal number of partitioning rounds (as few as a single round as our data suggest). Our previous selection of aptamers for a large-size protein target ($MW_P \approx 90$ kDa) showed that a

high-affinity pool of the enriched library could be obtained after a single round of IFCE or 3 rounds of complex-first NECEEM.^{34,42} Based on **Figure 3.5** and given the value of binder abundance of 10^{-5} (estimated from a single-round IFCE) and the value of $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$, the predicted m values for KCE-based selection of aptamer for this protein are as follows: $m_{\text{min,IFCE}} = 1$ and $m_{\text{max,IFCE}} = 25$ while $m_{\text{min,complex-first NECEEM}} = 2$ and $m_{\text{max,complex-first NECEEM}} = 52$. This prediction means that both IFCE and complex-first NECEEM could be used to select aptamers for a 90-kDa protein from this random-sequence DNA library with binder abundance $\approx 10^{-5}$. The predicted minimum number of partitioning rounds depends on the mode of partitioning. A single round is required in IFCE and two rounds are required in NECEEM, which agrees with the results of experimental selection: a single round in IFCE and 3 rounds in complex-first NECEEM. Another important conclusion from data shown in **Figure 3.5** is that both NECEEM and IFCE fail to retain a sufficient quantity of binders after one round of partitioning when the binder abundance in the initial library is as low as 10^{-8} or 1 binder per one hundred million nonbinders ($m_{\text{max}} < 1$ in all cases in **Figure 3.5c**). When the bulk value of k_{off} is high ($k_{\text{off}} = 10^{-3} \text{ s}^{-1}$), this lower limit of binder abundance in the initial library for successful KCE-based selection increases to $10^{-6.5}$ or $\sim 3 \times 10^{-7}$ ($m_{\text{max}} \leq m_{\text{min}}$ in all cases with $k_{\text{off}} = 10^{-3} \text{ s}^{-1}$ in **Figure 3.5b**).

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

final equilibrium mixture with the sample plug length of 1 cm. Our preliminary data suggest that increasing the sample-plug length by an order of magnitude (from 1 to 10 cm) increases k_N by multiple orders of magnitude, resulting in insufficient separation of DNA nonbinders from P-DNA complexes. Due to such an inherent limitation on the size of the injected sample in KCE-based partitioning, the ultimate solution to improve the success rate of aptamer selection for less aptagenic target proteins is to use DNA libraries with higher binder abundance, such as modified oligonucleotide libraries with functionalized protein-like groups.^{30,102-104} Selection of aptamers from modified DNA libraries have yielded high-affinity aptamers to many difficult-to-select-for proteins that had repeatedly failed SELEX with unmodified DNA libraries.³⁰ Application of KCE-based partitioning to selection of aptamers from modified oligonucleotide libraries is a promising direction of further development of this partitioning approach.

3.5. CONCLUSIONS

This work clearly demonstrates high productivity of our simple formalism based on considering partitioning as a filter with differential transmittance for binders and nonbinders. Not only does this formalism simplify and help to understand the bases of partitioning but also the application of this formalism to a specific mode of partitioning can lead to practical recommendations for the users. Below, we summarize recommendations derived in this work for KCE-based partitioning (**Figure 3.6**). The size of protein target dictates the choice of the mode of KCE-based partitioning in aptamer selection. In NECEEM, as the size of protein target decreases, k_N increases by several orders of magnitude while the k_B is relatively stable and close to unity. On the other hand, IFCE improves k_N values (which are as low as 10^{-9} and do not change much with varying size of protein target) at the expense of sacrificing k_B . The k_B values

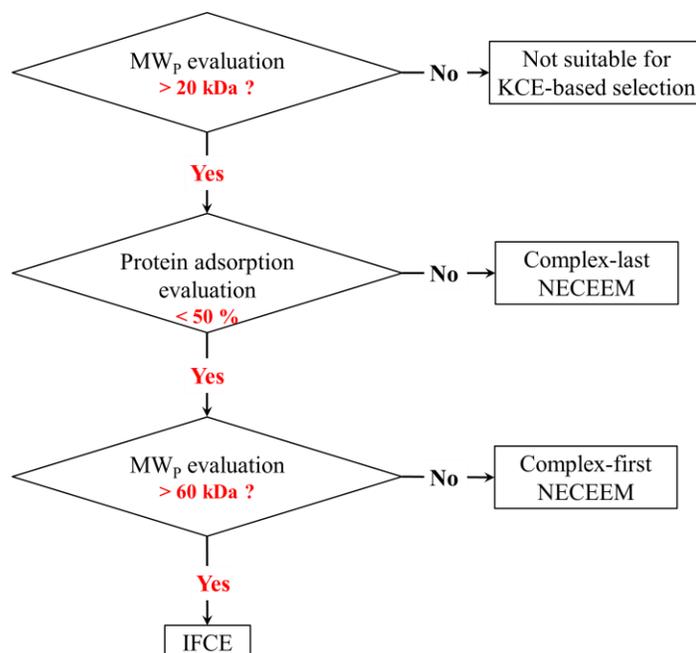


Figure 3.6. Choice of KCE-based partitioning according to the characteristics of protein targets.

in IFCE decrease drastically when the size of protein target decreases, thus, hindering the collection of sufficient quantity of intact complexes for small-size protein targets. To ensure obtaining a pool of binder of sufficient purity and quantity, one must find a balance between k_N and k_B to determine the most suitable mode of KCE-based partitioning and corresponding running buffer. For large-size protein targets, IFCE is preferred in order to obtain high affinity aptamers in fewer rounds of partitioning (single-round selection could be possible as our data suggest). We recommend that NECEEM be used in selection of aptamers for small-size protein targets, especially in non-SELEX, in which there is no PCR amplification of the collected pools between the rounds of partitioning. Between the two sub-modes of NECEEM, the complex-first sub-mode is proven to facilitate selection of aptamer in fewer rounds due to lower k_N values. However, the use of complex-last NECEEM is still beneficial when the adsorption of some protein targets to the uncoated inner capillary surface is severe and detrimental for the selection;

the coating of the walls can suppress such adsorption. For more difficult protein targets (e.g., those that had repeatedly failed SELEX), KCE-based selection fails when the binder abundance in the initial library is as low as 10^{-8} . Under such circumstances, using a better library, such as chemically modified DNA library with a greater abundance of binders, could be a promising direction for future development in the field of highly efficient CE-based aptamer selection.^{30,102-}

104

While this study considers only two characteristic sets of pH and I_{RB} values of the running buffer to illustrate the two modes of KCE-based partitioning (NECEEM and IFCE), these modes can be achieved with three types of conditions. Complex-first NECEEM can be achieved using a running buffer with low I_{RB} and/or high pH values in a bare fused-silica capillary. IFCE can be achieved using a running buffer with high I and/or low pH values in a bare fused-silica capillary. Complex-last NECEEM can be achieved using a running buffer with broad ranges of pH and I_{RB} values in a coated capillary with suppressed EOF. Since the analytical resolution R_s is a function of μ_{P-DNA} , μ_{DNA} , and time of separation t , the value of R_s can be fine-tuned (to achieve desired outcomes of selection) by modulating t or varying the running buffer composition within the acceptable ranges of pH and I_{RB} for each mode of KCE-based partitioning. For example, the stringency of selection can be increased to drive the selection process towards obtaining aptamers with low K_d and/or low k_{off} values. Aptamers with low K_d can be obtained by using a lower target concentration for preparation of the equilibrium mixture, while aptamers with low k_{off} can be obtained by increasing the separation time. The latter can be achieved *via* several practical means, such as (i) decreasing pH and/or increasing I_{RB} (ii) decreasing the applied voltage, (iii) increasing the capillary length and (iv) decreasing the running buffer temperature. The means of (ii), (iii) and (iv) are applicable for all modes of KCE-based partitioning, while (i)

is only applicable for complex-first NECEEM and IFCE. It should be noted that while increasing the selection stringency favours stronger and more stable binders, too high a stringency can be detrimental for the selection. As such, when optimizing the experimental conditions, one must balance k_B and k_N carefully to ensure the collection of binders of sufficient purity and quantity. This balancing can be done in a rational way using our formalism.

CHAPTER 4. INFLUENCE OF TARGET CONCENTRATION ON BINDER

SELECTION: AN EXPERIMENTAL STUDY

The presented material was submitted to *Angewandte Chemie International Edition* **2023** (currently under review): **Le, A.T.H.**; Teclemichael, E.; Krylova, S.M.; Krylov, S.N. Influence of target concentration on aptamer selection: the experimental study. My contributions to this manuscript were: (i) performing the thrombin selection experiments, (ii) preparing all figures, (iii) interpreting the results and (iv) writing the manuscript. My colleague, Eden Teclemichael, conducted the MutS selection portion of the study.

4.1. CRITICAL ROLE OF TARGET CONCENTRATION IN SELEX

The combined findings from **Chapters 2** and **3** successfully achieve the three objectives outlined in Chapter 1: (i) improving the partitioning efficiency of KCE, (ii) adapting the conditions of KCE-based partitioning to be physiologically relevant for potential in vivo applications, and (iii) providing a comprehensive formalism for the proper implementation of all KCE-based partitioning methods. Before transition into the utilization of KCE within the DEL platform, in this chapter, I aim to explore a crucial variable that significantly influences the binder selection process: the concentration of the target protein. This chapter marks the introduction of the first experimental investigation into the relationship between target concentration and the selection of protein binders from oligonucleotide libraries.

Target concentration is a key variable in SELEX for any given target and library, and it is broadly accepted that SELEX performance (e.g., round-to-round progress in binder enrichment) depends on target concentration.^{20,22,105,106} However, the effect of target concentration on

quantitative parameters of selection has never been studied experimentally owing to the complexity of such a study. While being conceptually simple, SELEX is very cumbersome and fails to select binders from non-modified oligonucleotide libraries in approximately 70% of attempts.³⁰ Completing one successful selection, i.e., obtaining aptamers, is an achievement on its own. Completing multiple successful selections for systematically varied target concentrations is objectively very difficult. Doing this for multiple targets would further multiply the difficulty. Performing such a study quantitatively, so that conclusive results could be obtained for the influence of target concentration on the quantitative characteristics of SELEX, is the ultimate leap. Here, we report on performing such a study.

Whereas there are no experimental studies, there are multiple theoretical works on the effect of target concentration on aptamer selection.^{20,22,105,106} However, all theoretical works (including our own investigation conducted before this experimental study) unavoidably utilize hard-to-prove assumptions and fundamentally-unknown parameters thus leading to non-instructive and impossible-to-test conclusions. Our analysis of the literature on experimental SELEX suggests that experimenters always choose target concentration arbitrarily — in this respect, SELEX is still more an art than a science. Here we present the first experimental study of the effect of target concentration on aptamer selection. Aptamers were selected for two protein targets at varying target concentrations for each. The results suggest that selection fails if target concentration is below a certain threshold value and succeeds if it is above this value. The threshold value depends on target nature as well as efficiencies of partitioning and PCR amplification. We propose a single experimental parameter which is easily determinable and indicative of target concentration being above or below the threshold: a quantity of DNA after partitioning in the presence of target relative to that in the absence of target (q). Our results lead

to a simple instructive conclusion: to avoid selection failure, experimentalists must use target concentrations which result in values of q statistically significantly greater than unity.

4.2. MATERIALS AND METHODS

4.2.1. Materials and solutions

All chemicals were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries with inner and outer diameters of 75 and 360 μm , respectively, were purchased from Molex Polymicro (Phoenix, AZ, USA). Recombinant His-tagged MutS protein (MW \approx 90 kDa, pI 6.0) was purchased from Prospec Protein Specialist (Ness Ziona, Israel). Recombinant human alpha-thrombin protein (MW \approx 36.7 kDa, pI 6.4–7.6) was purchased from Fisher Scientific (Mississauga, ON, Canada). All DNA molecules were custom synthesized by Integrated DNA Technologies (Coralville, IA, USA). CE running buffers were 50 mM Tris-HCl pH 8.0 and 50 mM Tris-acetate pH 8.2 for MutS and thrombin selections, respectively. The sample buffer was always identical to the running buffer to avoid the adverse effects of buffer mismatch. Accordingly, all dilutions of sample components in CE experiments were done by adding the corresponding running buffer.

All DNA stock solutions were subjected to annealing by incubation at 90 °C for 2 min before being cooled to 20 °C at a rate of 0.5 °C/s, prior to the dilution and preparation of the equilibrium mixtures. To avoid cross-contamination between the selections for two different protein targets, distinct synthetic fluorescein amidite (FAM)-labeled, 40-nt random DNA libraries (referred to as N40) with unique primer regions were used as follows: (i) for MutS: 5'-FAM-CTC CTC TGA CTG TAA CCA CG-N40-GC ATA GGT AGT CCA GAA GCC-3', and (ii) for thrombin: 5'-FAM-CTA CGG TAA ATC GGC AGT CA-(N40)-AT CTG AAG CAT AGT CCA GGC-3'.

Two sets of primers were used to amplify binders selected from the starting library. The primers in the first set were unlabeled and employed for quantitative PCR (qPCR). These primers had the following sequences: (i) for MutS: 5'-CTC CTC TGA CTG TAA CCA CG-3'(forward) and 5'-GGC TTC TGG ACT ACC TAT GC-3'(reverse), and (ii) for thrombin: 5'-CTA CGG TAA ATC GGC AGT CA-3' (forward) and 5'-GCC TGG ACT ATG CTT CAG AT-3'(reverse). For asymmetric PCR (aPCR), the second set of primers included a fluorescently labeled version of the forward primer and a biotin-labeled version of the reverse primer: (i) for MutS: 5'-Alexa Fluor488-CTC CTC TGA CTG TAA CCA CG-3'(forward) and 5'-Biotin-TEG-GGC TTC TGG ACT ACC TAT GC (reverse), and (ii) for thrombin: 5'-Alexa Fluor488-CTA CGG TAA ATC GGC AGT CA-3' (forward) and 5'-Biotin-TEG-GCC TGG ACT ATG CTT CAG AT-3'(reverse).

4.2.2. CE instrumentations and default conditions

All CE experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detection system. Fluorescence was excited with a blue line (488 nm) of a solid-state laser and detected at 520 nm using a spectrally-optimized emission filter system. The poly(vinyl alcohol) (PVA)-coated capillaries were prepared as described elsewhere.⁹⁹ The total length of the capillary was 80 cm for most of the experiments, except for the bulk affinity tests conducted for MutS selection, where the capillary length was 50 cm. In all cases, the detection window was positioned 10 cm away from the outlet of the capillary. Prior to every run, the PVA-coated capillary was rinsed with the running buffer at 20 psi (138 kPa) for 8 min. The coolant temperature was set at 15 °C.

4.2.3. Specifics of CE-based fraction collection

In Round 1, the equilibrium mixture contained the annealed starting library of 10 μM and the protein target of chosen concentration; 330 nM binder-enriched library was used for Rounds 2 and 3 instead of 10 μM . The target concentration in the equilibrium mixture was kept constant throughout the three rounds of selection. The equilibrium mixtures were incubated at room temperature (21 $^{\circ}\text{C}$) for 1 h to approach chemical equilibrium in the binding reaction. The equilibrium mixture was injected into the capillary by a pressure pulse of 1 psi (6.9 kPa) \times 28 s, resulting in a sample plug of 3.7 cm in length. The sample plug was propagated by a pressure pulse of 0.9 psi (6.2 kPa) \times 45 s (to yield a 5.4 cm-long buffer plug) to pass the uncooled region of the capillary before applying the electric field. Partitioning was carried out using reversed polarity (anode at the outlet) at 25 kV for 26 and 20 min for MutS and thrombin selections, respectively. After KCE-based partitioning, elution of the target–binder complex was facilitated by pressure at 5 psi (34.5 kPa) for 1 min into a fraction-collection vial containing 20 μL of the running buffer.

4.2.4. PCR procedures and generation of binder-enriched library

The eluted binder-enriched library was amplified and quantitated by two rounds of quantitative PCR (qPCR) using CFX Connect instrument (Bio-Rad, ON, Canada). The qPCR reagent mixture was prepared to obtain final concentrations of 1 \times Q5 High-Fidelity 2 \times Master Mix (New England BioLabs, Whitby, ON, Canada), 1 \times SYBR Green (Fisher Scientific, Mississauga, ON, Canada), 500 nM unlabeled forward primer, and 500 nM unlabeled reverse primer. Before thermocycling, the qPCR reaction mixture was prepared by adding a 2 μL aliquot of the eluted fraction to 18 μL of the qPCR reagent mixture. The PCR thermocycling protocol

was as follows: 98 °C for 30 s (initialization, performed once), 98 °C for 10 s (denaturation), 65 °C for 20 s (annealing), and 72 °C for 20 s (extension), followed by a plate read at 72 °C and a return to the denaturation step for a total of 40 cycles. All qPCR reactions were performed in duplicate. In the first round of qPCR, the eluted fraction was quantitated using an eight-point calibration curve. An S-shaped amplification curve was then plotted for the eluted fraction. In the second round of the qPCR, the qPCR product of the eluted fraction was removed when it was two cycles into the exponential phase of the previously plotted amplification curve. After qPCR, 100 µL of the qPCR product was later purified using the MinElute[®] PCR purification kit (QIAGEN, Mississauga, ON, Canada) as per manufacturer's instructions. Once product's purity was verified by native PAGE, it was subjected to asymmetric PCR (aPCR). Five µL of DNA was added to 45 µL of aPCR reagent mixture from New England Biolabs Inc. (Whitby, ON, Canada). Final concentrations of PCR reagents in the reaction mixture were: 1×Q5[®] Reaction Buffer, 1 µM fluorescently labeled forward primer, 50 nM biotin-labeled reverse primer, 0.02 units/µL Q5[®] High-Fidelity DNA Polymerase, and 200 µM dNTPs mix. The reaction was performed in duplicates with the following temperature protocol: 98 °C for 30 s (initial denaturation, performed once), 98 °C for 10 s (denaturation), 65 °C for 20 s (annealing), and 72 °C for 20 s (extension). Eighteen cycles of aPCR were run. Ten µL of MagnaBind[™] streptavidin beads suspension (Fisher Scientific, Mississauga, ON, Canada) was washed three times and resuspended in bead washing/binding buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA pH 8.0). Once amplified, the duplicate PCR reactions were combined and incubated with streptavidin magnetic beads for 30 min at a room temperature (23 ± 1 °C). The beads were magnetized, discarded, and the PCR product was then purified using the MinElute[®] PCR purification kit as per manufacturer's instructions.

To quantitate the DNA concentration in the binder-enriched library, serial dilutions of the fluorescently labelled forward primer (2 μ M, 1 μ M, 500 nM, 250 nM, 125 nM, 62.5 nM, and 31.25 nM) were prepared to build a standard curve by measuring fluorescence intensity at 519 nm with NanoDrop 3300 Fluorospectrometer (Fisher Scientific, Mississauga, ON, Canada). The purified binder-enriched library was then ready for the next round of selection.

4.2.5. Specifics of bulk affinity test

Equilibrium mixtures of either the starting library or the binder-enriched library and varying target concentrations were prepared and incubated at room temperature for 1 h prior to injection into the capillary. Throughout all the bulk affinity tests, the concentrations of the starting library or the binder-enriched library remained constant (i.e., 1 nM for MutS selection and 20 nM for thrombin selection). In case of MutS bulk affinity tests, a 50-cm capillary was used to shorten the separation time while still ensuring the desired resolution between the unbound library and the target–binder complex. As such, the conditions for MutS bulk affinity tests were readjusted as follows: (i) sample injection at 0.5 psi (3.4 kPa) \times 20 s to create a 2.1 cm-long sample plug, (ii) buffer propagation at 0.9 psi (6.2 kPa) \times 30 s to yield a 5.8 cm-long buffer plug and pass the uncooled capillary region and (iii) separation at 25 kV with reversed polarity (anode at the capillary outlet) for a duration of 15 min. Due to the poor resolution in thrombin selections, the bulk affinity tests were continued to be conducted using an 80-cm capillary. The conditions for thrombin bulk affinity tests were the same as conditions used in the binder selection with the total separation time of 25 min.

4.3. THEORETICAL AND EXPERIMENTAL CONSIDERATIONS IN THE STUDY OF EFFECTS OF TARGET CONCENTRATION IN SELEX

To be successful and instructive, an experimental study of the influence of target concentration on aptamer selection by SELEX requires several right choices to be made. First, such a study requires a highly efficient partitioning method so that the chances of failure are minimized, and the timescale of the study is kept reasonable (months rather than years). We chose partitioning by KCE, which reliably supports partitioning efficiency of 10^4 – 10^9 (orders of magnitude higher than that of surface-based partitioning).^{31,35,36,42,107,108} Partitioning by KCE typically facilitates aptamer selection in less than five rounds of SELEX. Second, an objective comparison of the SELEX results for different target concentrations requires a reliable bulk affinity assay. We used our recently-published bulk-affinity workflow relying on optimized measurements of the fraction of unbound library (R).¹⁰⁹ We chose four target concentrations covering two orders of magnitude to work with: 500, 100, 10, and 1 nM. Accordingly, we planned for conducting four selections with constant target concentrations (500, 100, 10, and 1 nM). Finally, such a study requires well understood and confirmed targets for aptamer selection by SELEX. We chose two proteins, His-tagged MutS (93 kDa) and non-tagged thrombin (35 kDa) for which successful aptamer selections were performed several times.^{32,34,110,111}

Our first task was to determine suitable modes of KCE-based partitioning for the two protein targets: His-tagged MutS and non-tagged thrombin. We found that both proteins excessively adhered to the inner fused silica capillary wall, while coating the wall with polyvinyl alcohol (PVA) suppressed this adsorption along with suppressing the electroosmotic flow.^{99,112} The suppressed electroosmotic flow forced us to use the “complex-last” mode of NECEEM-based partitioning in which the unbound DNA (nonbinders) moves inside the capillary faster than the

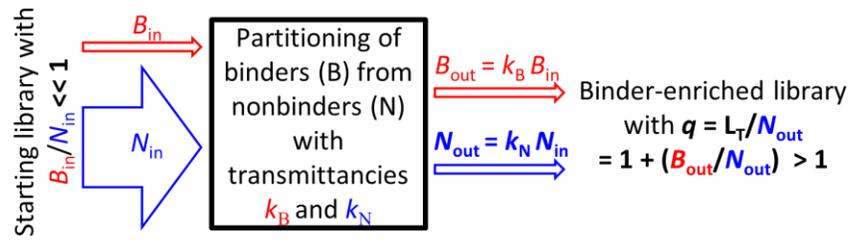


Figure 4.1. Schematic representation of efficient partitioning of binders (B) from nonbinders (N) in a mixture of oligonucleotide library and protein target. At the input of partitioning, the binder-to-nonbinder ratio (B_{in}/N_{in}) is typically $\ll 1$. At the output of partitioning, with transmittance equal to k_B and k_N for binder and nonbinder, respectively, the binder-enriched library is expected to have binder-to-nonbinder ratio (B_{out}/N_{out}) $> B_{in}/N_{in}$, with the relative yield of DNA, $q > 1$. In this study, we utilized two experimental parameters, k_N and q to draw a conclusion about the choice of target concentration for selection.

protein-DNA (target-binder) complexes.⁵⁰ For each target concentration, we conducted 3 rounds of selection, as typical for a NECEEM-based selection routine.

We then conducted the NECEEM-based selections of binders for MutS and thrombin from a DNA library with 40 randomly positioned nucleotides. In Step 1, the target was mixed with the library, and the mixture was incubated for 1 h to allow the formation of target-binder complexes, serving as a positive control. As a negative control, we used a mixture of the library with target matrix void of the target. In Step 2, a small volume of the mixture was injected inside the capillary (the length of the resulting sample plug was approximately 5% of the capillary length), and target-binder complexes were separated from the unbound library by electrophoresis. A fraction was collected in a pre-determined time window, where binders should elute (see **Appendix D1** for the determination of the binder-elution window). In Step 3, the collected

fraction was subjected to a two-stage PCR amplification — quantitative PCR (qPCR) followed by asymmetric PCR (aPCR) — to produce the binder-enriched library for the next round of SELEX. The bulk-affinity assay was performed for the starting library and for each of the binder-enriched libraries, and its results were used to judge the progress of selection. qPCR was used to obtain two quantitative parameters (**Figure 4.1**): (i) a transmittance of partitioning for nonbinders (k_N), often referred to as the nonbinder background ^{42,50}, and (ii) a relative yield of DNA (q). The transmittance is defined as the ratio between the quantities of nonbinders at the output (N_{out}) and input (N_{in}) of partitioning, in the absence of the target (negative control) (see **Eq. 4**).

In the presence of the target and, thus, when target–binder complexes are formed, the output quantity of the library (L_T) is a sum of the output quantity of binders (B_{out}) and nonbinders (N_{out}). We define the relative yield of DNA (q) as a ratio between the output quantity of the library in the presence of the target (L_T) and the output quantity of nonbinders (N_{out}).

$$q = L_T/N_{out} = (B_{out} + N_{out})/N_{out} = 1 + (B_{out}/N_{out}) \quad (42)$$

For the selection to be productive, B_{out} , must not only be larger than zero but also exceed the limit of detection in PCR (typically ranging from 10 to 100 DNA molecules). Therefore, q must be statistically significantly greater than unity for the selection to proceed. To assess the experimental errors associated with calculation of q , multiple sets of positive-control and negative-control experiments for one constant target concentration were conducted to determine L_T and N_{out} , respectively. The mean value of q and its relative standard deviation (RSD) were then estimated (**Appendix D2**).

4.4. RESULTS AND DISCUSSION

We completed three-round selections for both MutS and thrombin with four constant (round-to-round) target concentrations. We followed the established bulk affinity workflow, as previously published,¹⁰⁹ to evaluate the bulk affinity of the binder-enriched libraries to the protein target after each round. For consistency across selections, we used a constant 10- μ M concentration of the 80-nt starting library in Round 1. Subsequently, we employed a 330-nM binder-enriched library for later rounds. To confirm the robustness and the reproducibility of the results of selections, we repeated two of the four selections for thrombin, particularly the selections for 10-nM and 500-nM target concentrations.

After every round of selection, k_N and q values were determined *via* qPCR using an above-outlined procedure; a detailed summary of k_N and q values can be found in **Appendix D2**. Notably, the k_N values were found to be in the range of 10^{-4} to 10^{-3} and 10^{-6} to 10^{-5} for NECEEM-based selection for thrombin and MutS, respectively. While the k_N value for MutS selection was typical of NECEEM-based partitioning, the k_N for thrombin selection was higher than expected.⁵⁰ This variance in k_N values was attributed to different resolutions of the proteins–DNA complexes from the DNA nonbinders between the two protein targets of different sizes (thrombin is smaller than MutS). In essence, thrombin selections were performed with an approximately 100 times higher nonbinder background than MutS selections. Considering that q is inversely proportional to the nonbinder background (**Eqs. 4** and **42**), it is expected that the theoretical range of q values in MutS selection would be approximately two orders of magnitude higher than in thrombin selection, given the same target concentration and input quantity of library molecules. Indeed, the experimental values of q in MutS selections were consistently within 1 to 2 orders of magnitude higher than in thrombin selections for the same target

concentrations (**Figure 4.2a**). In addition to the nonbinder background, variations in experimental q values could arise from the following factors. First, the uncertainty associated with the qPCR-determined L_T and N_{out} values can lead to potential variations of up to ~15% under the same experimental conditions (see calculations of RSD in **Appendix D2**). Second, the nature of the target, specifically its aptagenicity or affinity to oligonucleotide binders, played a pivotal role in dictating the binder abundance in the starting library and setting the upper theoretical limit of L_T . Intuitively, an ideal selection would have a high binder abundance in the starting library (high L_T) and low nonbinder background (low k_N or N_{out}), leading to a high q value much greater than unity.

A consistent trend observed in the q values for both targets was a decrease in q with decreasing target concentration, ultimately reaching $q \approx 1$ (**Figure 4.2a**). This trend was expected; at higher target concentration, more target molecules will be available to bind DNA molecules, increasing L_T and q . On the opposite, when the target is in deficiency, only the most tightly bound binders remain bound to target and are collected at the output of partitioning, resulting in lower L_T and q values. For thrombin selection, the q value decreased to near unity when the target concentration reached (on the way down) 10 nM. Since MutS selection experienced a lower nonbinder background, the q value was close to unity at a lower target concentration of 1 nM. Another important observation from **Figure 4.2a** was that when q value in Round 1 was near unity (as seen in the 1-nM and 10-nM selections for Thrombin, as well as the 1-nM selection for MutS), there was no subsequent increase in the q value between rounds for a given target concentration. However, when q value in Round 1 was significantly greater than 1, q consistently increased from round to round, peaking in Round 3 (e.g., 100-nM and 500-nM thrombin selections; and 10-nM, 100-nM, and 500-nM MutS selections). This increase in q

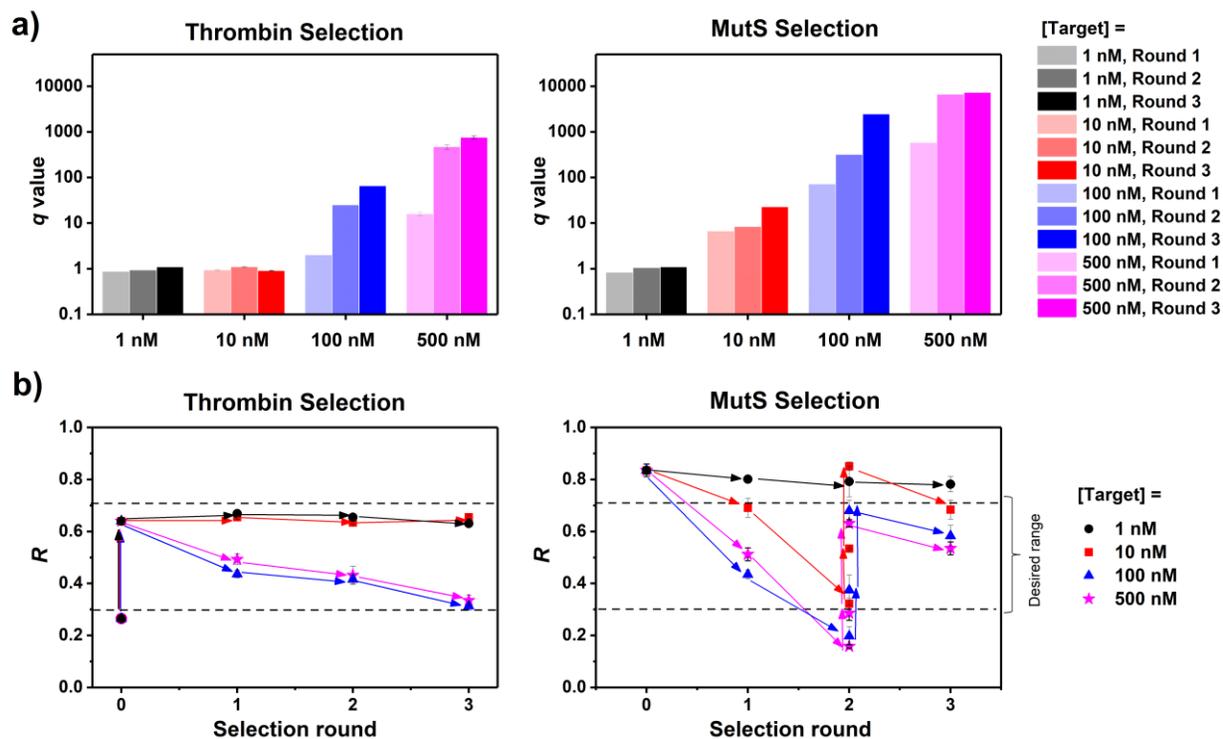


Figure 4.2. Comparison of q values (a) and bulk affinities represented by R values (b) to evaluate the selection outcomes for MutS and thrombin under four different constant (throughout the rounds of selection) target concentrations. In (b), the measurements of R in each selection followed a previously established workflow for assessment of bulk affinity, starting with a protein concentration of 1 μM . To ensure that R remained within the desired range, we systematically adjusted the target concentration in the bulk affinity workflow in a stepwise manner. The vertical arrow connecting points on the graph indicates a 10-fold decrease in target concentration for the same selection round, which was implemented to maintain R within the desired range.

between rounds for such target concentrations means that the yield of DNA binders was enriched throughout the selection process, possibly indicating a positive selection outcome (which will be discussed in detail below).

To accurately evaluate and compare the selection outcomes across different target concentrations, we conducted a previously published bulk affinity workflow for every binder-enriched library obtained in the selections for both targets.¹⁰⁹ The results of bulk affinity assays are summarized in **Figure 4.2b**, where R is plotted against the selection round for every target concentration, with Round 0 being the starting library prior to the selection (see **Appendix D3** for the electropherograms and calculations of R values). R , representing the fraction of unbound library obtained in the bulk affinity test, serves as an indicator of enrichment progress: lower R values indicate improved affinity to the target. It is expected that $R > 0$ is typically observed for Round 0 while enrichment progressively reduces R for subsequent rounds. To mitigate poor accuracy associated with R measurement close to its limits (0–1), we systematically adjusted the protein concentration in the bulk affinity test in a stepwise fashion to maintain R within the desired range of 0.3 to 0.7 (as indicated by the vertical arrow between the points in **Figure 4.2b**).¹⁰⁹

In the case of thrombin selection, the improvement (decrease) of R throughout the selection was observed for 100-nM and 500-nM target concentrations while R value remained unchanged when the target concentration decreased down to 10 nM — corresponding to the point where q reached near unity. A similar trend was seen in MutS selection: the progression of R eventually ceased at a target concentration of 1 nM, where q was close to unity. Overall, successful selections with increasing binder enrichment were achieved for those target concentrations for which q values in Round 1 exceeded unity.

The combined results from qPCR measurements of DNA quantity and bulk affinity assays allow us to draw the following key conclusions. First, selection outcomes are contingent upon a critical “threshold” target concentration; selection fails below this threshold but proceeds when

target concentration exceeds it. Second, as target concentration decreases, q value is getting smaller until it gets closer to unity (i.e., $q = 1 + n\sigma$, where σ is the standard deviation of q at $q = 1$, which was ~ 0.15 in our assay, as detailed in **Appendix D2**) at the “threshold” target concentration. Therefore, we recommend employing target concentrations that yield q values statistically significantly greater than unity (i.e., $q > 1 + n\sigma$) for a successful selection with enrichment of binders. In practical terms, this means setting n at a minimum of 3, indicating that q must be at least 3 standard deviations above unity to achieve a positive selection outcome. Should experimentalists aim to increase the selection stringency by utilizing a lower target concentration to drive the selection towards selecting binders with higher affinity, it is crucial to ensure that the resulting q value remains above unity. In certain cases, it may be necessary to expand the q range. When the range of q is expanded for a given target, the threshold target concentration at which q reaches unity decreases, facilitating binder enrichment at lower target concentrations. In essence, a more reliable selection with higher chance of success, even at low target concentrations, is characterized by high q values. While the range of q may be largely influenced by the nature of protein target, increasing q can be achieved by decreasing the nonbinder background (k_N) of the partitioning method or increasing L_T through the means of a more superior starting library, such as a chemically modified DNA library with greater affinity to the target.^{30,102-104}

4.5. CONCLUSIONS

To conclude, our data further underscores the critical role of target concentration in the performance of SELEX. In particular, we have established that SELEX succeeds when the target concentration surpasses a specific threshold value and halts when the concentration falls below

this threshold. This threshold value varies depending on several factors, including the nature of the target, the nonbinder background, and the quality of PCR; however, it can be readily identified on a case-to-case basis using a simple quantitative parameter: the relative yield of DNA after partitioning (q) which is defined as the ratio between the quantity of collected DNA in the presence the target and the absence of the target. We encourage experimentalists to choose the target concentration that results in q values statistically significantly greater than unity. The use of this straightforward parameter will prove invaluable in rational determination of the target concentration, ultimately enhancing the prospect of success within the SELEX community.

CHAPTER 5. CAPILLARY ELECTROPHORESIS-BASED SELECTION FROM DNA-ENCODED LIBRARIES OF SMALL MOLECULES: THE PROOF OF PRINCIPLE STUDY

My contributions to this study were: (i) performing all the presented experiments (unless otherwise specified), (ii) preparing all figures and (iii) interpreting the results.

5.1. AN INTRODUCTION TO DNA-ENCODED LIBRARIES OF SMALL MOLECULES AND THE PROSPECTIVE ROLE OF KCE-BASED PARTITIONING IN DEL PLATFORM

In **Chapter 3**, I have established a formalism for determining the appropriate CE conditions for selecting oligonucleotide binders for a given protein target. This formalism includes a detailed characterization of KCE-based partitioning methodology as well as guidance on choosing different KCE-based partitioning modes, namely, complex-first NECEEM, complex-last NECEEM and IFCE. Additionally, in **Chapter 4**, I developed practical guidelines to evaluate the choice of target concentration in SELEX to prevent selection failure. In this final chapter of my thesis, I gathered the insights from these two preceding chapters to conduct the first-ever KCE-based selection of small molecule binders from DNA-encoded libraries of small molecules (DELs).

Discovering specific ligands that interact with pharmaceutical-relevant targets is the primary step in the modern drug development.¹¹³⁻¹¹⁵ Traditional approaches to discover such therapeutic ligands have heavily relied on high-throughput screening (HTS) of large compound libraries against specific drug targets.¹¹³ However, the resource-intensive and time-consuming nature of

these approaches, combined with limitations in chemical diversity and target specificity, have led researchers to explore new frontiers in drug discovery. One such frontier that has emerged as a game-changer in the field is the use of DNA-encoded libraries of small molecules (DELs).^{116,117} DELs provide a solution for the main dilemma of selecting ligands from highly diverse combinatorial libraries of molecules. According to probability rules, an increased structural diversity in the library results in a greater number of suitable ligands.² However, greater diversity reduces the number of unique molecule copies, making binder identification more challenging. In DELs, each small molecule's structure is encoded within its DNA tag, enabling identification through tag amplification and sequencing. The efficiency of polymerase chain reaction (PCR) and DNA sequencing is so high that selecting a few copies of each ligand from a DEL is sufficient for structure identification.¹¹⁸ Unlike traditional drug discovery methods that rely solely on physically screening individual compounds, DELs allow for the simultaneous screening of millions to billions of compounds in a single experiment, providing an unprecedented level of diversity and efficiency in drug discovery.¹¹⁹

The concept of using DNA to encode a combinatorial chemistry library was first introduced by Brenner and Lerner in 1992,¹¹⁷ and since then, various approaches to DEL synthesis have been developed and employed.^{6,120-125} These approaches fall into two general categories: “DNA-directed chemistry”, where the oligonucleotide strand both identifies and directs the synthesis of each library member, and “DNA-recorded chemistry”, where the oligonucleotide strand serves as an identification tag for corresponding chemical moieties (**Figure 5.1**).^{126,127} In principle, DNA-directed approaches allow the application of an evolution process to small molecule selection through alternating rounds of selection, PCR amplification, and translation, in a complete analogy to biological display systems.¹²⁰⁻¹²³ To regenerate small molecules between selection

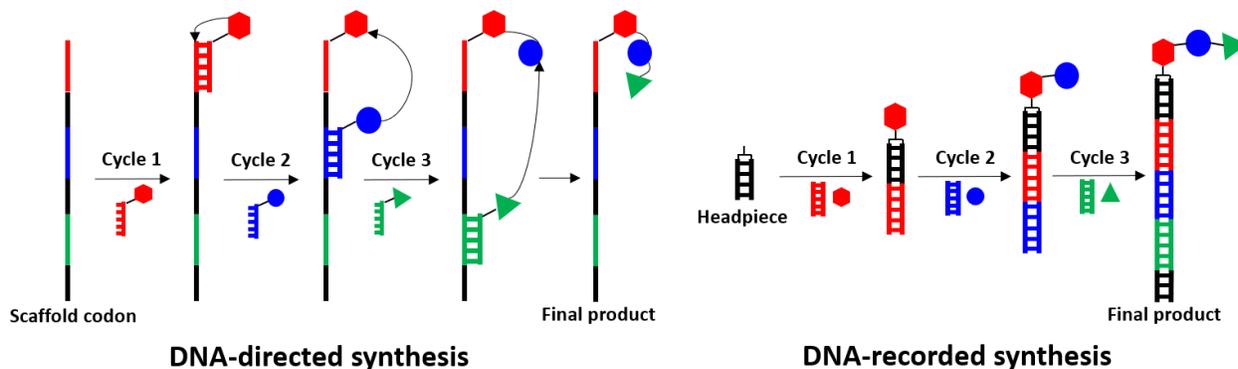


Figure 5.1. Schematic representation of two general assembly pathways for DELs. Building blocks of the small-molecule head and DNA fragments encoding them are shown by the same color. The end product in this illustration is a DEL comprising three building blocks, obtained from three cycles of building block synthesis. Potentially, more cycles can be introduced to create a library with greater number of building blocks.

rounds through a PCR-amplified genetic code (scaffold codon), prior conjugation of a large set of chemical building blocks to DNA oligonucleotides before library assembly is necessary. However, this preparatory step can be time-consuming and labor-intensive process, potentially limiting its application in a fast-paced industrial setting.

The “DNA-recorded” approaches, developed by researchers at Praecis Pharmaceuticals (now owned by GlaxoSmithKline) and the Neri group at the Institute of Pharmaceutical Science in Zurich, Switzerland, have collectively overcome the complexity associated with library generation seen in “DNA-directed” chemistry.^{6,124,125} In this approach, oligonucleotides carrying distinct short sequences serving as identification barcodes for the corresponding chemical moieties are ligated to a DNA scaffold in each cycle of chemical synthesis. The encoded compounds are combined and split into aliquots for the subsequent synthesis stage, involving different sets of chemical building blocks and the ligation of the corresponding oligonucleotide

codes. This process efficiently generates a remarkable level of library diversity over iterative cycles of division, synthesis, and pooling in a “split-and-pool” manner. Unlike DNA-directed libraries, DNA-recorded libraries cannot be regenerated and, thus, they are not subject to iterative rounds of selection or evolution principles. Nonetheless, DNA-recorded approaches offer significant advantages over DNA-directed approaches, including the use of off-the-shelf reagents, ease of library synthesis, and the ability to use any encoding DNA sequences that may not be compatible with DNA-directed library synthesis methods. These features have enabled GlaxoSmithKline to construct DNA-recorded libraries containing up to 800 million individual library members, the largest DNA-encoded small-molecule library reported to date, with just four rounds of DNA ligation and chemical synthesis on a triazine scaffold.

Given their prominence in the industrial setting and the accessibility from our collaborators, our study will focus exclusively on the DEL platform based on DNA-recorded chemistry.

As previously detailed in **Section 1.1**, the selection of ligands from DELs involves three major steps: *(i)* mixing the protein target with DELs, *(ii)* partitioning the target-bound ligands (binders) from the target-unbound ligands (nonbinders), and *(iii)* identifying the binders through sequencing of the DNA tags. The efficiency of the partitioning step is crucial for the successful selection from DELs since binders selected from DELs cannot be amplified or regenerated *via* molecular biology tools, and each round of partitioning results in binder losses.⁶ Consequently, a typical binder selection from a highly diverse DEL is limited to three rounds and necessitates highly efficient partitioning. Unfortunately, surface-based techniques with poor partitioning efficiency, particularly those involving protein-immobilized magnetic beads, are typically used for the partitioning step in DEL.¹²⁸ The limited partitioning efficiency of such techniques primarily stems from the nonspecific adhesion of DEL molecules to the expansive surface area

of the bead matrices. To address this limitation, binding-induced hybridization/ligation-based techniques, such as interaction-dependent PCR, have been actively developed for DEL-based screening platforms and have shown promise.^{129,130} However, these techniques require the attachment of complex DNA constructs to the target molecule, which may interfere with intermolecular interactions. Additionally, practical challenges related to assembling and attaching DNA constructs to the target molecule create bottlenecks in screening a large number of target/library combinations; thus, these methods are still confined to a relatively small number of specialized laboratories. Consequently, despite their limited partitioning efficiency, bead-based techniques continue to dominate DEL screening due to their robustness, simplicity, and accessibility.

In this context, we propose the use of KCE as an innovation, solution-based alternative to the traditional surface-based methods for separating binders from nonbinders in DEL-based screening platforms. KCE is renowned for its exceptional partitioning efficiency and has been utilized successfully in the selection of binders from random oligonucleotide library. As a result, we anticipate that KCE holds great promise as a highly efficient method for selecting ligands from DELs. Our rationale for this proposal is as follows:

In DELs, the DNA tag is significantly larger and carries a greater charge than the small molecule head. Consequently, the physical properties of DELs, including size and charge, are primarily determined by their DNA tags.⁴⁹ Each DEL has a single structure of DNA tags, resulting in their migration in KCE as a single zone. Protein–DEL complexes will also migrate as a single zone with a mobility intermediate between that of the protein and the DELs. **Figure 5.2** illustrates the migration patterns of DELs, proteins, and protein–DEL complexes in different modes of KCE-based partitioning: complex-first NECEEM, complex-last NECEEM, and IFCE.

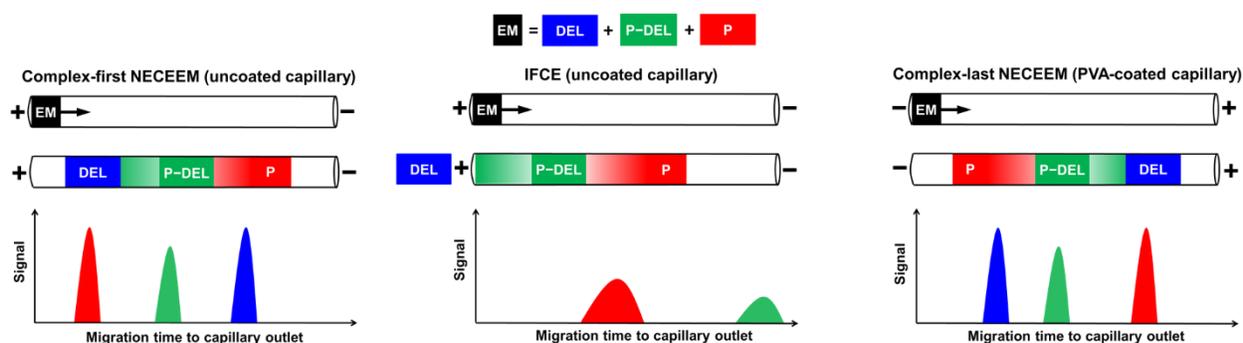


Figure 5.2. Conceptual depiction of migration patterns of DEL, protein (P) and protein-DEL complex (P-DEL) in different KCE-based partitioning modes. The positions of the three corresponding electrophoretic zones are similar to those observed in KCE-based aptamer selection from random oligonucleotide libraries. See text for details.

The positions of these zones are similar to those observed in KCE-based aptamer selection from random oligonucleotide libraries (**Figure 5.2**).

In this proof-of-principle study, we utilized the interaction between several DEL ligands and their respective protein targets to identify suitable KCE modes for partitioning based on our prior knowledge of KCE-based partitioning regarding specific protein target characteristics. To experimentally validate that KCE methodology can indeed achieve highly efficient selection of small-molecule ligands from DELs, we created hypothetical DELs using the known DEL ligands. We then conducted mock selections for DEL binders corresponding to chosen protein targets, employing different KCE-based partitioning modes. When compared to the partitioning performance of bead-based techniques for hypothetical DELs, KCE-based partitioning significantly outperforms in both binder recovery and nonbinder background levels. Following the results of the mock selections, we performed the selection of binders from a real DEL library targeting one of the protein targets chosen in the mock selections. We applied the previously

identified KCE-based partitioning conditions from the mock selection and achieved a successful binder selection outcome, as confirmed through high throughput sequencing (HTS) data analysis of the binder-enriched library. Collectively, these results affirm that KCE can be employed with remarkable efficiency in the selection of small-molecule ligands from DELs, surpassing conventional bead-based techniques. This advancement paves the way for the use of such libraries in identifying drug leads and diagnostic probes, marking a significant step forward in this field.

5.2. MATERIALS AND METHODS

5.2.1. Materials and solutions

All chemicals were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries with inner and outer diameters of 75 and 360 μm , respectively, were obtained from Molex Polymicro (Phoenix, AZ, USA). Bodipy (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) was purchased from Life Technologies Inc. (Burlington, ON, Canada). Recombinant streptavidin (SA) protein (MW = 52 kDa, *pI* 6.8) was purchased from Sigma-Aldrich (Oakville, ON, Canada). Recombinant soluble epoxide hydrolase (sEH) protein (MW = 63 kDa, *pI* 5.9) was produced and purified by Nurix Therapeutics (San Francisco, ON, Canada). All DNA primers were custom synthesized by Integrated DNA Technologies (Coralville, IA, USA). The CE running buffers were 50 mM Tris-HCl pH 8.0, 40 mM NaCl for complex-first NECEEM experiments and 50 mM Tris-HCl pH 7.5, 100 mM NaCl with addition of 0.1% Tween 20 for complex-last NECEEM experiments. The sample buffer was the same as the running buffer to prevent adverse effects of buffer mismatch.

Biotin DNA tag (SA binder)

5'-TGACTCCCAAATCGATGTGTTCCGCAAGAAGCCTGGTAAGCGGAGAAAGGTCGTTTTACTGCCCGGTCTAC (N12) CTGATGGCGCGAGGGAGGCAACAGCTATGACCATA/3AlexaF488N/
3'-ACCGCGCTCCCTCCG-5' (Forward primer)
5'-TGACTCCCAAATCGA-3' (Reverse primer)
3'-ACTGAGGGTTAGCTACACAAGGCGTTCTTCGGACCATTCGCCTCTTTCCAGCAAATGACGGGCCAGATG

GLCBS-L-Leu DNA tag (SA nonbinder)

5'-TGACTCCCCAGGAAACAGCTATGACAAGAAGCCTGGTAAGCGGAGAAAGGTCGTTATTGCTCCCGGTCTAC (N12) CTGACACACGCCTACCGCCCATTTGCGAACAGCTA/3AlexaF488N/
3'-CGGATGGCGGGTAAA-5' (Forward primer)
5'-CAGGAAACAGCTATGAC-3' (Reverse primer)
3'-ACTGAGGGGTCCTTTGTCGATACTGTTCCTTCGGACCATTCGCCTCTTTCCAGCAATAACGAGGGCCAGATG

Figure 5.3. Structure details of the DNA tags used in the DEL ligands for SA. The annealing sites of qPCR primers were also included.

In this study, DEL ligands for SA were supplied by GlaxoSmithKline, while DEL ligands for sEH were provided by Nurix Therapeutics. The DEL ligands for SA consisted of biotin and Gly-(1)Leu-4-carboxybenzene sulfonamide (GLCBS-1-leucine) as the SM heads of the binder and nonbinder, respectively.¹³¹ In case of sEH, (1R,3S)-N-(4-cyano-2-(trifluoromethyl)benzyl)-3-((4-methyl-6-(methylamino)-1,3,5-triazin-2-yl)amino)cyclohexane-1-carboxamide (a.k.a, GSK-2256294) and benzoic acid served as SM heads of the binder and nonbinder, respectively.^{53,132}

Detailed synthetic procedures for the SA ligands were previously described, with a modification of the closing primer ligation method.⁶ In this modified approach, Klenow polymerization was omitted, and the longer oligo strand was positioned at the top, leaving a 35-nucleotide 3' overhang to provide a noncompetitive priming site for more efficient PCR amplification. The detailed structures of the DNA tags of SA ligands can be found in **Figure 5.3**. The structures and sequence information of the DNA tags of sEH ligands were kept confidential by Nurix Therapeutics and are not disclosed here. It is worth noting that all DNA tags for the DEL ligands were fluorescently labeled for detection.

For DEL ligands specific to SA, two sets of specific primers were used to quantitate the binder and nonbinder separately: (i) for binder: 5'-GCC TCC CTC GCG CCA-3' (forward) and

5'-TGA CTC CCA AAT CGA-3' (reverse), and (ii) for nonbinder: 5'-AAA TGG GCG GTA GGC-3' (forward) and 5'-CAG GAA ACA GCT ATG AC-3' (reverse). For DEL ligands for sEH, both binder and nonbinder shared a universal primer set: 5'-TGA CTC CCG ACC GAA GG-3' (forward) and 5'-GGT TAG GCA GGG ATG AGA GA-3' (reverse).

5.2.2. CE instrumentation and default conditions

All CE experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detection system. The fluorescence was excited with a blue line (488 nm) of a solid-state laser and detected at 520 nm using a spectrally optimized emission filter system. All capillaries were 50-cm long (40 cm to the detector) and had an inner diameter of 75 μm and an outer diameter of 360 μm . The detection window was positioned 10 cm away from the outlet of the capillary. The poly(vinyl alcohol) (PVA)-coated capillary was prepared as described elsewhere.⁹⁹ New capillary was conditioned successively with MeOH at 20 psi (138 kPa) for 10 min, 0.1 M HCl at 20 psi for 3 min, 0.1 M NaOH at 10 psi (7 kPa) for 6 min, water at 20 psi for 3 min and a running buffer at 40 psi (276 kPa) for 40 min. Prior to every run, the capillary was rinsed successively with 0.1 M HCl, 0.1 M NaOH, deionized H₂O, and a running buffer at 20 psi for 3 min each. Conditioning steps were not required for PVA-coated capillaries; such capillaries were rinsed with the running buffer only at 20 psi for 10 min prior to every run. In all cases, the coolant temperature was set at 15 °C.

Equilibrium mixtures of protein targets and DEL ligands were incubated at room temperature (21 °C) for 1 h to approach chemical equilibrium. When specified, 500 nM of Bodipy (a neutral molecule) was added to the equilibrium mixture to monitor the electroosmotic flow (EOF). In

case of negative control experiment, the protein target was omitted from the equilibrium mixture. In the binding tests for the protein targets and their DEL ligands, the equilibrium mixtures were consisted of 100 nM DEL ligand (either binders or nonbinders) and 1 μ M protein target. In the mock selection, the equilibrium mixtures contained 1 nM binders, 1 μ M nonbinders and 1 μ M of the corresponding protein target. In all instances, the equilibrium mixture was injected into the capillary by a pressure pulse of 0.5 psi (3.5 kPa) \times 10 s, resulting in a sample plug of 1 cm in length. The sample plug was propagated by a pressure pulse of 0.9 psi (6.2 kPa) \times 45 s (to yield a 5.4 cm-long buffer plug) to pass the uncooled region of the capillary before applying the electric field. In this study, two modes of NECEEM-based partitioning were conducted. For complex-first NECEEM (used for SA-DEL ligand pair), CE was carried out at an electric field of 400 V/cm (20 kV over 50 cm), with the positive electrode located at the injection end of the capillary. For complex-last NECEEM (used for the sEH-DEL ligand pair), CE was performed at an electric field of 240 V/cm (12 kV over 50 cm), with the negative electrode placed at the injection end of the capillary. When specifying fraction collection, we collected fractions within predefined time windows, where the protein target–DEL complex was expected to elute. For SA-DEL ligand pair, the fraction was collected between 15 and 25 minutes of partitioning using complex-first NECEEM. For sEH-DEL ligand pair, the fraction was collected by applying pressure at 4 psi (28 kPa) for 1 min after 26 minutes of partitioning using complex-last NECEEM. Collection vials contained 20 μ L of the corresponding running buffer.

5.2.3. Quantitative PCR

DNA tags of DEL molecules in the collected fractions were amplified and quantitated by qPCR using a CFX Connect instrument from Bio-Rad (Mississauga, ON, Canada). The qPCR

reagent mixture was prepared to obtain final concentrations of 1×Q5 High-Fidelity 2×Master Mix (New England BioLabs, Whitby, ON, Canada), 1×SYBR Green (Fisher Scientific, Mississauga, ON, Canada), 500 nM forward primer, and 500 nM reverse primer.

In the case of SA mock selection, we prepared different qPCR reagent mixtures with primer sets specific for binders and nonbinders to quantitate them separately. For sEH, a universal set of primers was used to quantitate both binders and nonbinders. Before thermocycling, the qPCR reaction mixture was prepared by adding a 2 μ L aliquot of the collected fraction to 18 μ L of the qPCR reagent mixture. The PCR thermocycling protocol was as follows: 98 °C for 30 s (initialization, performed once), 98 °C for 10 s (denaturation), 65 °C for 20 s (annealing), and 72 °C for 20 s (extension), followed by a plate read at 72 °C and a return to the denaturation step for a total of 40 cycles. All qPCR reactions were performed in duplicate.

5.2.4. Specifics of KCE-based selection of SM ligands for sEH from a pre-screened DEL library

A DEL library with a diversity of 4.74 billion compounds was produced and subjected to 2 rounds of pre-screening, using target-immobilized magnetic bead-based partitioning by our collaborator, Nurix Therapeutics. This pre-screened library was then subjected to a single round of complex-last NECEEM-based selection of SM ligands for sEH. The conditions for fraction collection and qPCR detection were similar to the procedures described in the previous two sections, with a few modifications specified below.

The equilibrium mixture consisted of 10 nM pre-screened library, 1 μ M sEH and 1 nM sEH nonbinders. Since this DEL library was unlabeled, the inclusion of 1 nM sEH nonbinders (a fluorescently-labelled DEL ligand with a known DNA tag sequence) served to precisely monitor

and ensure that both unbound DEL and the pre-determined binder elution window were in their expected positions during CE. To maximize the screening and collection efficiency of the DEL library molecules, a total of seven repetitions of complex-last NECEEM experiments were conducted over a 10-hour period, using the same equilibrium mixture. The collected fractions were subsequently subjected to qPCR, pooled, and sent to Nurix Therapeutics for next-generation sequencing (NGS) and further analysis.

5.3. THEORETICAL AND EXPERIMENTAL CONSIDERATIONS IN DESIGNING AND PROVING HIGHLY EFFICIENT KCE-BASED SELECTION OF SMALL MOLECULE LIGANDS FROM DELS

In order to demonstrate the efficiency of KCE in selecting small molecule (SM) ligands from DELs, it is crucial to design a study that allows for a quantitative and conclusive comparison with traditional surface-based techniques. To achieve this, we have adopted a universal framework, based on published guidelines for the development and validation of high-efficiency ligand selection from oligonucleotide libraries, to structure our comparative study.⁵⁵

To set the stage for a meaningful comparison, the first step is to choose a target-DEL pair. This pair must include: a binder, which is a DEL ligand containing a SM head with high affinity for the target, and a nonbinder, which is a DEL ligand containing a SM head with minimum to no affinity for the target. These SM heads should either be well-documented in the literature to exhibit the desired affinity to the target or must have been previously selected against the target from actual DELs.

Once the appropriate target, binder, and nonbinder have been identified, the next step is to determine the optimal KCE conditions to effectively separate the target-DEL complexes from

unbound DELs. This is based on prior research regarding suitable KCE conditions for selecting binders from oligonucleotide libraries (as detailed in **Chapter 3**).⁵⁰ Under these chosen experimental conditions, our objective is to ensure that the complex of the binder with the target can be readily formed and detected in CE, while the nonbinder should not exhibit any affinity to the target. After establishing the optimal KCE-based separation conditions, the partitioning performance of KCE in DEL selection can be quantitatively assessed. This involves the following steps: (i) assembling a hypothetical DEL with known input quantities of binders (B_{in}) and nonbinders (N_{in}), (ii) conducting KCE-based partitioning using the hypothetical DEL and (iii) determining the quantities of binders (B_{out}) and nonbinders (N_{out}) at the output of partitioning by quantitative PCR (qPCR).

To ensure an accurate characterization of partitioning by qPCR, it is essential that the amplification of binders and nonbinders is highly specific and free from any cross-amplification biases between them. The qPCR-determined quantities of binders and nonbinders at both the input and output of KCE-based partitioning allow us to characterize the quality of partitioning and compare it to the results obtained using surface-based techniques. Specifically, this data enables the calculation of the transmittance of partitioning of binders (k_B), often referred to as the binder recovery, by determining the ratio of B_{out} to B_{in} (see **Eq. 3**). Similarly, the transmittance of partitioning of nonbinders (k_N), commonly known as the binder recovery, is calculated by establishing the ratio of N_{out} to N_{in} (see **Eq. 4**).

Furthermore, adhering to the previously described guidelines for evaluating the choice of target concentration to avoid any selection failure (refer to **Chapter 4**), we will assess the choice of target concentrations utilized in our selections. This assessment involves the determination of the relative yield of library molecules, denoted as the q value. The q value is derived by

calculating the ratio of total library molecules in the presence of the target (L_T) to that in the absence of the target (N_{out}) (see **Eq. 42**). A q value significantly greater than unity indicates a positive selection outcome with an enrichment of binders.

In the following sections, we will experimentally demonstrate KCE-based selection of SM ligands from DELs, using the established quantitative criteria, while carefully addressing the experimental considerations outlined throughout the process.

5.4. RESULTS AND DISCUSSION

5.4.1. Identification of suitable KCE conditions to separate target–DEL complexes from unbound DEL

In this proof-of-principle study, we employed two model systems, each consisting of a protein target and its DEL ligands featuring DNA tag structures identical to those found in a typical DEL. The selected protein targets were untagged streptavidin (SA) with a molecular weight (MW) of 52 kDa and isoelectric point (pI) of 6.8, as well as His-tagged soluble epoxide hydrolase (sEH) with a MW of 63 kDa and a pI of 5.9. For each protein target, we employed a set of DEL ligands, including a binder and a nonbinder, featuring similar DNA tag structures but distinct small molecule heads. Notably, all DNA tags were fluorescently labelled, enabling the detection of DEL ligands through laser-induced fluorescence (LIF) module at 520 nm.

For SA, the SM head of the DEL binder was biotin, a well-known high-affinity binding partner for streptavidin. In contrast, the SM head of the DEL nonbinder was Gly-(1)Leu-4-carboxybenzene sulfonamide (GLCBS-1-leucine), a compound recognized as a SM binder for carbonic anhydrase II and not expected to exhibit affinity for SA.¹³¹ The second protein target, sEH, is a member of the α,β -hydrolase-fold family of epoxide hydrolases and represents a

therapeutic target for chronic obstructive pulmonary disease (COPD). For sEH, the SM head of the DEL binder was GSK2256294, a SM drug candidate developed by GlaxoSmithKline against sEH.^{53,132} On the other hand, the nonbinder for sEH featured only a benzoic acid functional group as its SM head. All the DEL ligands used in this study were synthesized and provided by our collaborators: GlaxoSmithKline supplied the DEL ligands designed for SA, while Nurix Therapeutics provided the DEL ligands intended for sEH.

Our initial task in determining the appropriate mode of KCE-based partitioning for non-tagged SA and His-tagged sEH was to assess the adsorption of these target proteins to the inner surface of the fused silica capillary. We found that non-tagged SA exhibited minimal adsorption to the silica capillary wall, whereas His-tagged sEH displayed a significant degree of adsorption. Consequently, for the selection to yield productive results, mitigating the high level of sEH adsorption to the silica capillary wall required the use of a capillary with a neutrally coated inner surface.^{99,112} In contrast, the selection for SA could be performed in a regular, uncoated capillary.

For SA, the use of an uncoated capillary allowed us to employ either the “complex-first” mode of NECEEM-based partitioning or IFCE-based partitioning for the selection process. Given the size of SA, which is 52 kDa, the most suitable partitioning mode was the “complex-first” NECEEM since IFCE primarily benefits very large protein targets in the range of approximately 100 kDa and above (refer to **Chapter 3** for quantitative characterization of KCE-based selection of binders from an oligonucleotide library).⁵⁰ Consequently, we opted for the “complex-first” NECEEM mode for SA, wherein the protein target–DEL complex migrates faster than the unbound DEL. To facilitate “complex-first” NECEEM mode, we used a sample/running buffer consisting of 50 mM TrisHCl at pH 8.0, with an additional 40 mM NaCl to reduce the electroosmotic flow (EOF) and enhance the resolution between SA–DEL complex

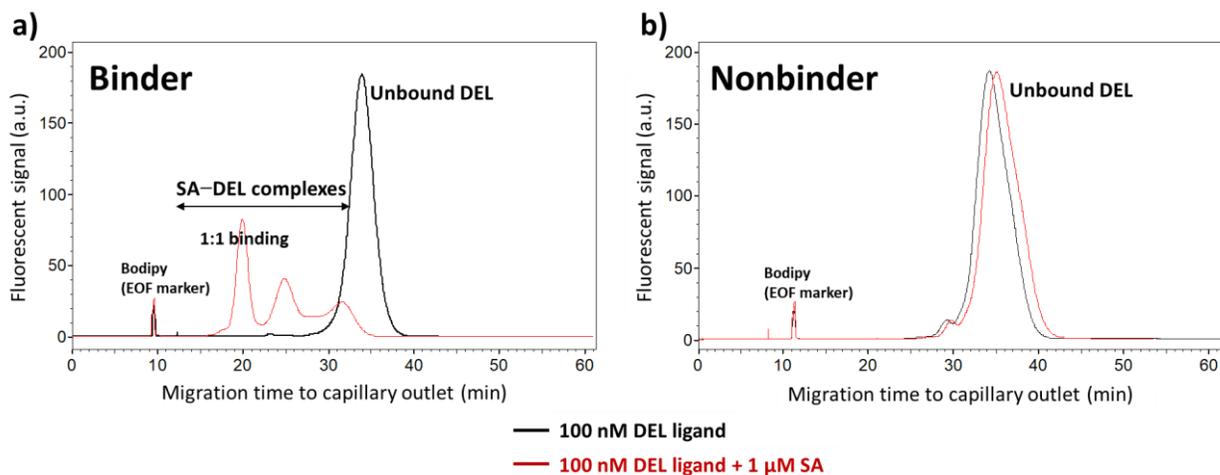


Figure 5.4. Complex-first NECEEM-based binding analysis of SA and its DEL ligands, (a) binder and (b) nonbinder. Multiple peaks of SA–DEL complexes were formed between SA and its biotin-containing ligand (binder). The 1:1 stoichiometry complex peak was indicated in (a) as “1:1 binding”.

and the unbound DEL peak without compromising significantly on the total separation time. SA, being a homotetramer capable of binding up to four biotin-containing DEL ligands, exhibited multiple SA–DEL complex peaks with different binding stoichiometries when introduced at a relatively high concentration in the micromolar range (**Figure 5.4a**, red trace). In our case, the primary focus was the 1:1 stoichiometry complex, which eluted at around 20 minutes. Notably, this 1:1 complex peak was effectively separated from the unbound DEL peak, which eluted at approximately 33 minutes (**Figure 5.4a**, black trace). In contrast, when the nonbinder for SA was subjected to complex-first NECEEM under similar conditions, no complex formation was observed (**Figure 5.4b**). Thus, under the chosen complex-first NECEEM-based partitioning conditions, we identified a suitable pair of binder and nonbinder for SA, forming the basis to generate a hypothetical library targeting SA in the mock selection.

For sEH and its ligands, a more challenging target-DEL ligand pair than SA and biotin-containing DEL ligand, we introduced a neutral coating, specifically polyvinyl alcohol (PVA), to minimize protein adhesion to the capillary wall. This neutral coating resulted in the suppression of EOF, necessitating the use of the “complex-last” mode of NECEEM-based partitioning. In this mode, the unbound DEL moves faster than the protein–DEL complex and elutes from the capillary prior to the complex. Since the condition of complex-last NECEEM is compatible with a wide range of running buffer pH and ionic strength (I_{RB}),⁵⁰ we opted for a sample/running buffer similar to the selection buffer used in the original DEL screening to discover GSK2256294, the SM head of the binder for sEH in our study. The chosen buffer consisted of 50 mM TrisHCl at pH 7.5, 100 mM NaCl,⁵³ with the addition of 0.1% Tween 20 to further minimize adsorption of sEH to the capillary wall and the inner surface of the sample-containing vial. This buffer had a higher ionic strength than conventional CE buffers, resulting in increased Joule heat generation and temperature inside the capillary. To address this, we employed SUMET to predict and adjust electrolyte temperatures by rationally reducing the electric field strength.⁸⁴

With the suitable KCE conditions identified, we conducted complex-last NECEEM to evaluate the binding affinity of DEL ligands to sEH. The DEL binder exhibited sub-micromolar range affinity, with over 40% of the unbound DEL peak remaining in the presence of 1 μ M sEH. However, due to the substantial adsorption of His-tagged sEH, the complex peak extended from approximately 25 to 45 minutes, with a dissociation bridge to the unbound DEL peak (**Figure 5.5a**, upper panel). In the case of the nonbinder, some binding was also observed, as the unbound DEL peak decreased by about 30% when 1 μ M of sEH was added (**Figure 5.5b**, upper panel). This scenario, where there is not a significant discrepancy between the affinities of the binder

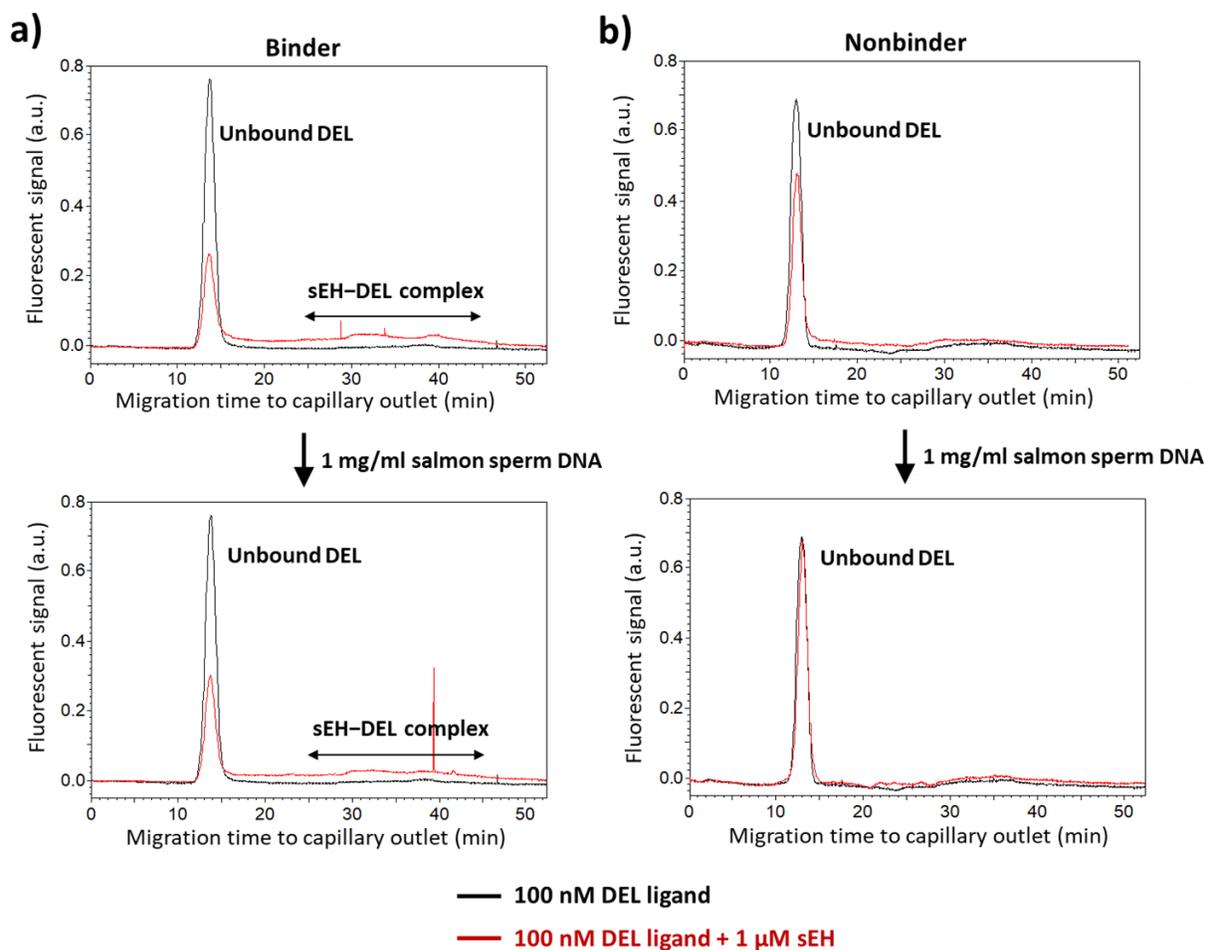


Figure 5.5. Complex-last NECEEM-based binding analysis of sEH and its DEL ligands, (a) binder and (b) nonbinder. In (a) and (b), the top panel represents the data without any blocking DNA in the sample, while the bottom panel corresponds to the data when 1 mg/ml of salmon sperm DNA was introduced.

and nonbinder for the protein target, is not ideal for the mock selection, as it would adversely impact the selection performance parameters. In such a case, a substantial quantity of nonbinders would be eluted alongside the binders, as both would form complexes with the protein. Later, we discovered that sEH exhibited affinity to the oligonucleotide strands of DEL, as evidenced by the

same level of nonspecific binding of sEH to a standard, random oligonucleotide library (without any SM moiety) (data not shown).

To address this issue, we introduced a high concentration of blocking DNA (specifically, 1 mg/ml salmon sperm DNA) to minimize the nonspecific binding of sEH to the DNA strands of DEL. The addition of blocking DNA not only had no adverse impact on the binding of the binders to sEH (**Figure 5.5a**, lower panel) but also significantly reduced the unwanted binding of the nonbinder to sEH (**Figure 5.5b**, lower panel). With the inclusion of blocking DNA in the sample, the binder and nonbinder for sEH, under the chosen complex-last NECEEM partitioning conditions, were now considered suitable for the creation of a hypothetical library for the mock selection.

The utilization of distinct KCE-based partitioning modes for SA (complex-first NECEEM) and sEH (complex-last NECEEM) is expected to yield different partitioning performance parameters for these two protein targets, which will be discussed further in **Section 5.4.2**.

5.4.2. Selection of small-molecule binders from hypothetical DEL libraries: a comparative evaluation of KCE-based and bead-based partitioning methods

In this section, we performed the mock selection for SA and sEH using the established KCE-based partitioning conditions outlined in the previous section. To create a hypothetical library, we spiked a known quantity of binders into nonbinders at a B_{in}/N_{in} ratio of 1 to 1000. This hypothetical library was then mixed with 1 μ M of the target protein, and the mixture was subjected to the chosen KCE-based partitioning method for the respective protein target: complex-first NECEEM for SA and complex-last NECEEM for sEH.

We collected a fraction within a pre-determined time window, where binders should elute. For SA, we collected the fraction between 15 and 25 minutes of partitioning using complex-first NECEEM. For sEH, we eluted the target–binder complex by pressure after 26 minutes of partitioning with complex-last NECEEM. To evaluate the partitioning performance, both the input library and the collected (output) fraction were subjected to qPCR to estimate the input and output quantities of binders (B_{in} and B_{out}) and nonbinders (N_{in} and N_{out}). For SA, we employed two sets of primers specific to binder and nonbinder to quantify them separately. We tested the quality of qPCR by building different calibration curves with serial dilutions of binders, nonbinders, and their combinations, ensuring proper efficiency and no cross-amplification. In the case of sEH, the DNA tags of both binder and nonbinder were originally designed to be amplified by the same primer set. Therefore, we performed complex-last NECEEM separately for binder and nonbinder using the same concentrations in the hypothetical library to estimate the quantities of binders and nonbinders for calculation of k_B and k_N . We then compared the quantities of binders and nonbinders collected in individual experiments with the quantity of DEL molecules collected using the hypothetical library to ensure consistency. All the data used to assess the quality of PCR procedures in mock selections were included in **Appendix E1**.

Figure 5.6 illustrated the mock selection results for SA and sEH, with qPCR-determined quantities of binders and nonbinders in the input library and the collected (output) fraction, both in the presence and absence of the protein target. For SA and sEH mock selections, the relative yield of library molecules after partitioning, represented as the q value (indicating the suitability of protein concentration to prevent selection failure), significantly exceeded unity. Notably, q values were 34 and 4 for SA and sEH mock selections respectively. This suggests that the chosen

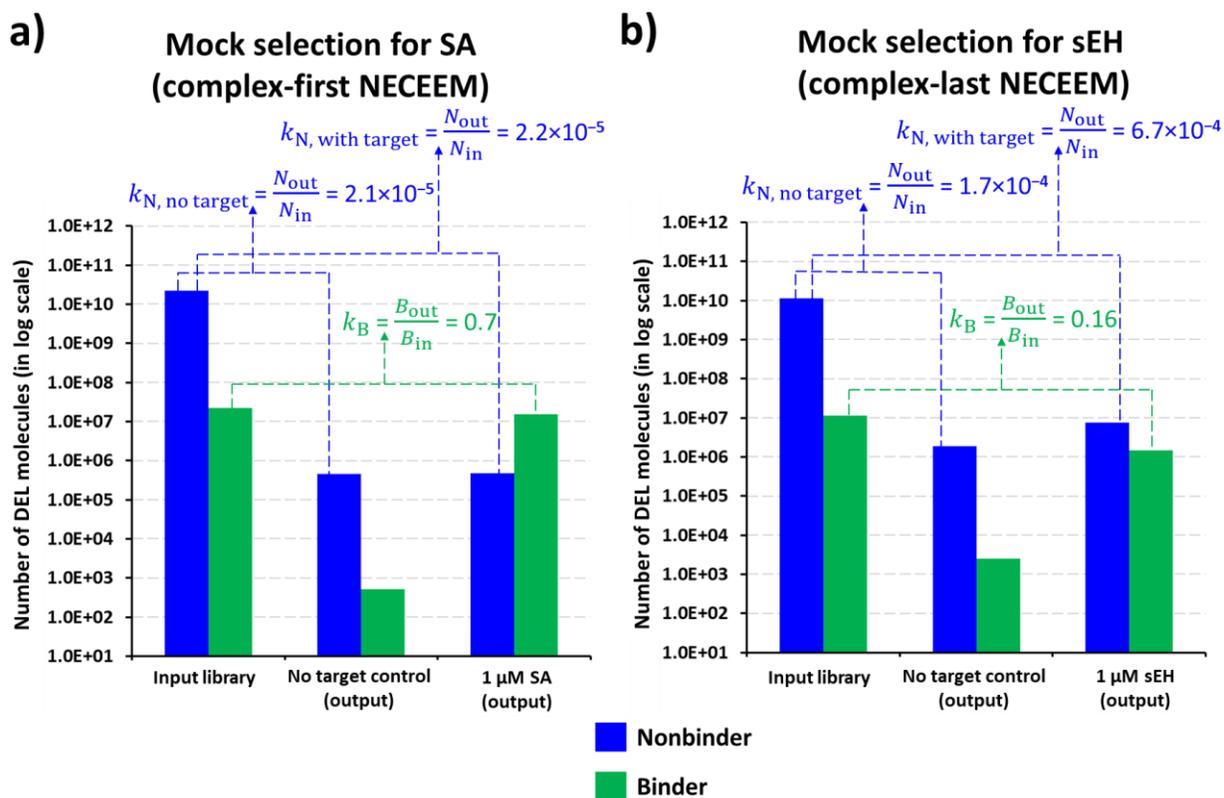


Figure 5.6. Evaluation of partitioning performance of (a) complex-first NECEEM and (b) complex-last NECEEM in mock selections of SM binder from hypothetical DELs for SA and sEH respectively. See text for more details.

target concentrations for these mock selections were appropriate and would likely lead to a positive selection outcome with an enrichment of DEL binders for both targets.

As we observed in **Figure 5.6a**, for SA, after a single round of complex-first NECEEM, the quantity of nonbinders decreased by approximately 5 orders of magnitude ($k_{N, SA} = 2.1 \times 10^{-5}$), while most of the binders were recovered ($k_{B, SA} = 0.7$). For sEH, after a single round of complex-last NECEEM, the quantity of nonbinders decreased by about 4 orders of magnitude ($k_{N, sEH} = 1.7 \times 10^{-4}$), and approximately 15% of binders were recovered after partitioning ($k_{B, sEH} = 0.16$) (**Figure 5.6b**).

The levels of nonbinder background observed in SA and sEH mock selections using different KCE-based partitioning modes were typical and aligned with our previously published quantitative characterization of partitioning in KCE (**Chapter 3**).⁵⁰ Typically, nonbinder background values range in the range of 10^{-5} to 10^{-3} for complex-last NECEEM and 10^{-6} to 10^{-4} for complex-first NECEEM. In complex-last mode, the nonbinder background is relatively higher, primarily due to contamination by faster-moving nonbinders at the capillary end before the complex elutes from the capillary end. As such, the sEH mock selection using complex-last NECEEM experienced approximately 10 times higher nonbinder background than the SA mock selection when complex-first NECEEM was used. Another important observation regarding the nonbinder background in the sEH mock selection is that significantly more nonbinder was collected in the presence of the protein target compared to its absence ($k_{N, sEH} = 1.7 \times 10^{-4}$ and 6.7×10^{-4} with and without protein, respectively). This indicates that, even with the addition of a high concentration of blocking DNA, some protein binding to the nonbinder still occurred. This binding was not observable under LIF detection in CE and required qPCR for detection, leading to a higher nonbinder background when the protein was present.

Regarding the binder recovery after partitioning, sEH mock selection showed lower binder recovery compared to SA selection ($k_{B, sEH} = 0.16$ versus $k_{B, SA} = 0.7$). This difference might be attributed to the loss of complexes involving the sticky His-tagged sEH, despite the use of a PVA-coated capillary and the addition of Tween 20 to the buffer to minimize adsorption. Given the low quantity of binder in the hypothetical library, this phenomenon was challenging to completely eliminate.

Finally, we quantitatively compared the partitioning performance of KCE-based selection to bead-based selection. **Table 5.1** summarizes the partitioning parameter values obtained for

Table 5.1. Comparison of KCE-based and bead-based partitioning performance in the selection of SM ligands from DEL.

Method of partitioning	Nonbinder background (k_N)	Binder recovery (k_B)
Complex-first NECEEM	2.1×10^{-5}	0.71
Complex-last NECEEM	1.7×10^{-4}	0.16
Magnetic bead*	1.0×10^{-3}	0.14

*Values for magnetic bead-based partitioning were obtained from Ref [53]

selection of binders from DEL using bead-based partitioning and KCE-based partitioning, particularly complex-first NECEEM and complex-last NECEEM modes of KCE. The data for bead-based selection was obtained from a previously published mock selection for sEH by GlaxoSmithKline researchers, where magnetic beads were used for partitioning binders from a hypothetical library with a similar binder-to-nonbinder ratio of 1:1000.⁵³ Compared to bead-based selection, KCE-based selection demonstrated significantly lower nonbinder background, ranging from one to several orders of magnitude lower, depending on the chosen CE partitioning mode. As for binder recovery, KCE-based selection also outperformed bead-based selection by up to a few folds. This value can vary greatly in KCE and depend on the k_{off} value of the selected binders as well as the nature of the protein, particularly its degree of adsorption to the inner capillary wall. In general, KCE-based selection for DEL consistently excelled over bead-based selection in terms of nonbinder background and binder recovery with the most notable improvement seen in the level of background binding.

5.4.3. Proposed KCE-based selection of DEL binders for a therapeutic protein target from a real DEL library

Following our quantitative assessment of KCE-based partitioning for selecting binders from a hypothetical DEL library, we moved on to the final phase of our study, where we applied KCE-based partitioning to select binders from a real DEL library. Our chosen protein target was sEH, a therapeutic target for chronic obstructive pulmonary disease (COPD) and one of the protein targets we used in our previous mock selection study to characterize KCE-based partitioning for DEL selection. To ensure consistency, we employed the exact KCE-based partitioning conditions (i.e., complex-last NECEEM) used in the mock selection for this real selection.

One of the major practical challenges of KCE-based selection, as outlined in the **Chapter 1** and published guidelines for KCE-based partitioning (**Chapter 3**) is its limitation in injection volumes, which restricts the quantity of the input library that can be injected into the capillary. Despite the high efficiency in KCE-based partitioning, this limitation can lead to a pool of binders with insufficient quantity for downstream processes like PCR amplification and sequencing. To address this issue, we decided to enhance the abundance of binders in the input library by utilizing a pre-screened DEL library, obtained after two rounds of magnetic bead-based partitioning, as the input library for our KCE-based partitioning. The DEL library for this selection was synthesized and provided by our collaborator, Nurix Therapeutics. Initially, approximately 5.1×10^{15} molecules from the library were used as input for pre-screening by 2 rounds of magnetic bead-based partitioning. Following the pre-screening process, about 5.8×10^{10} molecules were obtained, which were then divided into two equivalent samples, each containing 2.9×10^{10} molecules. One of these samples was used for an additional round of bead-based partitioning, while the other was subjected to one round of complex-last NECEEM-based

partitioning. The two binder-enriched pools, one obtained exclusively from 3 rounds of bead-based partitioning and the other from 2 pre-screened rounds of bead-based and 1 final round of complex-last NECEEM-based partitioning, were subjected to qPCR to quantitate the number of collected DEL molecules and later underwent high-throughput sequencing (HTS) to decode the DNA tags of the collected DEL molecules. As part of our collaborative efforts, the bead-based partitioning was conducted by Nurix Therapeutics, while we performed the KCE-based partitioning procedures.

Figure 5.7 provides a summary of the qPCR-determined quantities of DEL obtained before (input) and after the final round of partitioning (output) for both bead-based and complex-last NECEEM-based partitioning. It can be observed that input library quantity in the final round of bead-based partitioning was about an order of magnitude higher than that of complex-last NECEEM-based partitioning. As previously explained, CE-based partitioning suffers from a limited injection amount, allowing only a very small fraction of the pre-screened library to be used as the input library in this selection. To mitigate this limitation, we conducted multiple repetitions of complex-last NECEEM-based partitioning from the same input library and pooled all the output fractions from these repetitions to maximize the quantity of DEL molecules that could be screened and collected. In this study, a total of seven repetitions of complex-last NECEEM experiments were conducted for the same input sample. This resulted in a total input quantity of 2.3×10^9 library molecules being screened in one round of KCE-based partitioning, which is about 8% of the total DEL molecules in the pre-screened input library.

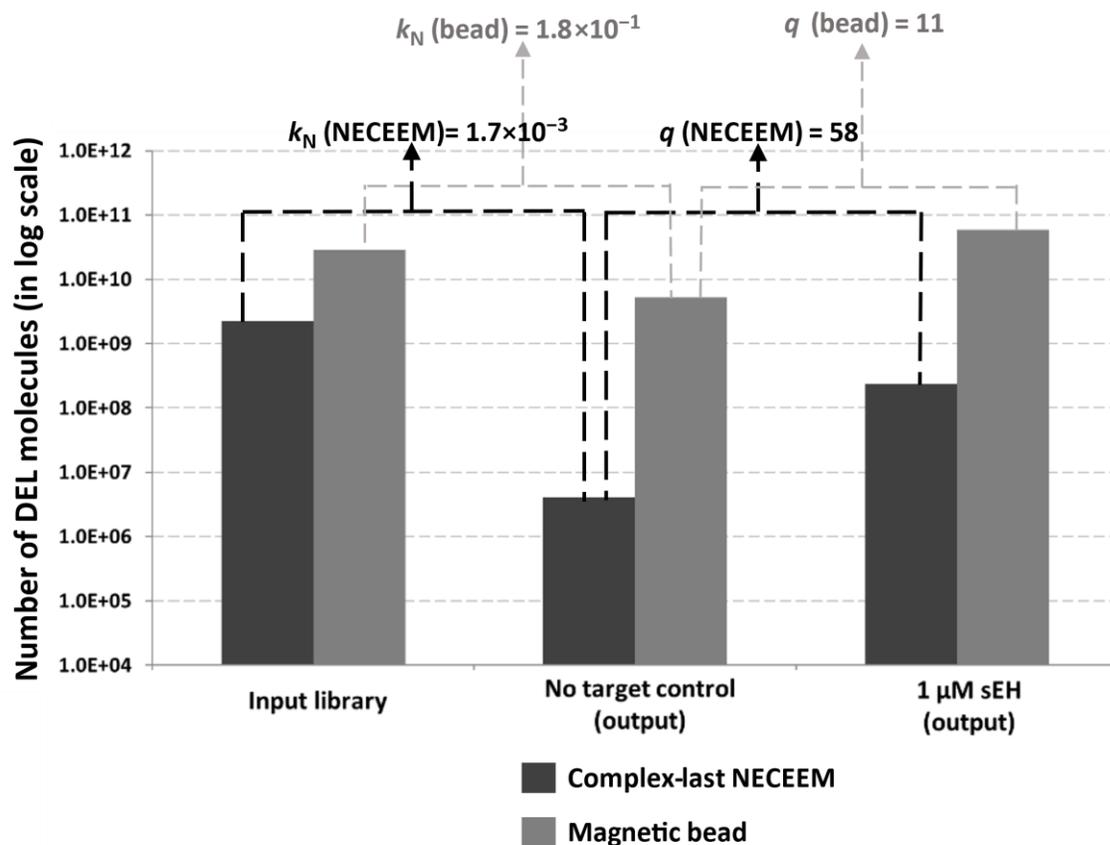


Figure 5.7. Evaluation of partitioning quality of complex-last NECEEM and magnetic bead-based selection of SM ligands from a real DEL library for sEH. See text for more details.

To assess whether the amount of protein used in both selections (i.e., 1 μM sEH) was sufficient, we followed the previously developed guidelines for target concentration evaluation. We obtained q values significantly larger than unity for both selections, with approximately 58 for complex-last NECEEM and 11 for bead. These q values indicate a positive selection outcome with the enrichment of DEL binders for both selections. Since q represents the yield of library molecules in the presence of target relative to that in the absence of target, it could be interpreted that complex-last NECEEM-based partitioning exhibited a higher (~5-fold) protein-dependent filtration compared to bead-based partitioning. Moreover, when evaluating the quality of

partitioning in the final round of selection for bead-based and KCE-based partitioning, we observed that KCE-based partitioning was more stringent. This stringency was evident from the measured nonbinder background, which was about two orders of magnitude lower for complex-last NECEEM-based partitioning ($k_{N, \text{complex-last NECEEM}} = 1.7 \times 10^{-3}$) compared to bead-based partitioning ($k_{N, \text{bead}} = 1.8 \times 10^{-1}$).

Based on Nurix's sequencing analysis of the output pools obtained in the final round of selection, positive selection outcomes were indeed observed for both bead-based and complex-last NECEEM-based partitioning (**Appendix E2**). In the input library used in both selections, a group of DEL controls with variable affinity to the protein target was integrated prior to partitioning. By monitoring the counts of these controls in binder-enriched pool obtained from bead-based and KCE-based selections, a strong correlation was found in the two HTS datasets. Specifically, the affinity of DEL controls highly correlated with their quantity in the final round of both selections, indicating that tighter ligands were enriched over weaker ones in both selections. Overall, qPCR data combined with HTS analysis suggests that both bead-based and KCE-based selection of SM ligands for sEH from a real DEL library were successful in enriching SM ligands for sEH from DEL, with KCE-based partitioning exhibiting more stringent partitioning performance (i.e., significantly lower nonbinder background). In the future, it is necessary to conduct a comprehensive analysis of HTS data to compare the binder-enriched pools obtained through KCE-based and bead-based selections. Additionally, a thorough evaluation of the affinity and kinetic characteristics of the selected DEL ligands is required to gain a deeper understanding of the potential advantages of DEL molecules selected via KCE in comparison to bead-based methods.

5.5. CONCLUSIONS

In conclusion, for the first time, we have demonstrated a step-by-step experimental development of KCE-based selection of SM ligands from a DEL library. This involved determining the appropriate KCE mode for partitioning for different protein targets, optimizing experimental conditions to separate target–DEL complexes from unbound DEL, and quantitatively evaluating KCE-based partitioning performance using a well-established mock selection protocol. We have successfully demonstrated that, in comparison to the traditional bead-based techniques, KCE-based partitioning excels in every aspect of selection performance, achieving significantly lower nonbinder background and higher binder recovery. Furthermore, we have successfully conducted the selection of SM ligands for sEH, a therapeutic target for COPD, from a real DEL library. It is also worth to note that the HTS data from the selected pool has revealed different classes of DEL ligands, some favoring bead-based partitioning and others favoring KCE-based partitioning. Further study is needed to compare the HTS data sets from both KCE-based and bead-based selections, which will provide insights into the potential practical applications of DEL compounds selected by KCE-based partitioning *versus* bead-based partitioning. By establishing a novel and highly efficient KCE-based partitioning platform for selecting SM ligands from DEL, this accomplishment highlights the potential of KCE-based partitioning methodology and is expected to promote its adoption in the field of DEL-based drug discovery.

LIMITATIONS AND FUTURE DIRECTIONS

Despite the significant advances in the development of KCE, the following challenges still remain in using the KCE technology in routine screening of oligonucleotide libraries: firstly, the low screening capacity due to the inherent limitation on the sample injection volume; secondly, the poor efficiency of KCE in separating molecules with a small difference in size-to-charge ratio (this issue is particularly noticeable when working with small target proteins.); and thirdly, restricted use of physiological buffers. (While there has been great progress in the compatibility of physiological buffers in KCE-based separations, these adaptations come at the cost of long separation time and increased complex dissociation.)

Conventional KCE-based selection suffers greatly from low injection volume, which decreases the total number of screened library molecules by at least 3 orders of magnitude as compared to that of surface-based selection. To address this issue, some strategies have been proposed. One approach involves using modified oligonucleotide libraries with improved target binding affinity.^{58,59} This approach aims to maximize the fraction of binders in the sampled library and ease the inherent issue of the greatly limited quantity of input library. Another approach is the pre-screening of a large quantity of the starting library using surface-based partitioning.^{60,61} This step generates a binder-enriched library prior to KCE-based selection, an approach that was employed in our selection of SM ligands from a real DEL for sEH (see **Section 5.4.3**). While this hybrid selection approach yielded positive outcomes, it also revealed certain limitations. One significant drawback is that, in the case of binder selection from DEL, the pre-screened library obtained by surface-based partitioning cannot be amplified by PCR. After pre-screening procedures, its quantity is significantly reduced to several orders of magnitude less than that of a regular, un-screened (naïve) library. Even with efforts to

concentrate this pre-screened library, its concentration remains much lower than the typical concentration of a naïve library used in KCE. Consequently, the injected quantity of the pre-screened library is vastly lower than the typical input quantity of a naïve library in KCE. This, in turn, can diminish the advantage of having a binder-enriched library, as it becomes challenging to collect a sufficient quantity of DEL binders at the output of KCE-based partitioning. In our specific case for sEH selection, we sought to maximize the input quantity of pre-screened libraries by conducting multiple repetitions of KCE-based partitioning using the same sample mixture (i.e., 7 repetitions over a 10-hour period). However, this approach may not be sustainable in other cases, especially considering that not all complexes can survive for such extended durations. Even with this approach, only 7% of the pre-screened library was sampled in KCE, in contrast to 100% coverage achieved by bead-based methods. For more challenging targets with lower affinity to the DEL library than sEH, this limitation becomes more pronounced, requiring a significantly larger input quantity to isolate binders. Consequently, while the hybrid selection approach holds promise for regular KCE-based SELEX from random oligonucleotide libraries, the non-SELEX feature of DEL-based platform necessitates a different approach to overcome the limitation of poor input quantity in KCE-based selection.

In this context, IFCE platform can potentially offer a great solution to increase the sampled quantity of the library in KCE-based partitioning without compromising peak resolution for selections of binders from both random oligonucleotide library and DEL. In IFCE, the complexes and nonbinders move in the opposite directions and only the complexes will be eluted at the output of partitioning; thus, decreasing interferences due to the poor peak shape of nonbinders. Increasing the sampled quantity in IFCE could be done *via* each of the following means or a combination of them: (i) using higher concentration of the starting library,

(ii) injecting a longer sample plug (the maximum sample plug length must be smaller than the length of the capillary) and (iii) applying a continuous electrokinetic injection approach. The above means can be utilized in any round of IFCE-based partitioning to maximize the quantity of collected binders and help to reduce the level of binder loss associated with the long separation time in IFCE. The last approach may provide a theoretically unlimited quantity of binders collected at the output as long as the complexes with suitable kinetic parameters are continuously fed into the capillary *via* electrokinetic injection while the nonbinders never enter the capillary. It is important to keep in mind that the odds of success in IFCE becomes slim for small-size protein target due to hours-long separation time.⁵⁰ Since the occurrence of highly stable complexes that survive such long separation is extremely low in most cases, it will be very interesting to see if the application of electrokinetic injection or any of the above means to increase the sampled quantity can improve the performance of IFCE for smaller protein targets.

Regardless of the modes, the success of KCE-based selection depends greatly on the size of target molecules: the larger targets have a greater chance to support successful selection of binders (refer to **Chapter 3** for a comprehensive characterization of KCE-based partitioning for different target sizes). The mobility of target–binder complexes for small-size targets differs little from the mobility of nonbinders. Accordingly, the efficiency of partitioning in NECEEM for such targets will be low while IFCE will require exceptionally long time of such complexes to elute.⁵⁰ Indeed, we found that the condition of IFCE leads to a drastic decrease in the output quantity of binders when the size of the protein target decreases. Thus, IFCE-based selection of binders for small-size protein target will certainly fail due to the excessive binder loss and the inability to collect a sufficient quantity of intact complex. On the other hand, under the IFCE conditions, target–oligonucleotide complexes for large-size target proteins elute faster and are

subject to lesser extent of dissociation, and, in turn, lesser loss of binders. It is recommended that IFCE be used only in selection of binders for large-size targets in order to obtain high-affinity binders in a minimal number of partitioning rounds. In the proof-of-principle work (**Chapter 2**), the feasibility of IFCE was proven by a successful one-step selection of a high-affinity aptamer pool to a large target protein, MutS (90 kDa). For comparison, selecting a pool with similar affinity to MutS by NECEEM-based partitioning required three rounds of SELEX. Recently, Martinez Roque *et al.* successfully obtained DNA aptamers for SARS-CoV-2 spike glycoprotein (130 kDa) with $K_d = 90$ nM after only two rounds of IFCE-SELEX.¹⁰⁷ Being the newest separation mode in the field of KCE-based screening of oligonucleotide libraries, IFCE is still to find considerable confirmation of its effectiveness. However, with its ability to facilitate highly efficient KCE selection of binders, more adaptations and uses are certainly anticipated for IFCE.

To resolve the limitation of KCE-based selection when dealing with small sized targets, valuable effort in the community have been directed towards increasing the effective size of these targets by coupling them to beads, leading to a drastic increase in the mobility shift upon complexation with bead-immobilized targets.⁴¹ However, the immobilization of target on a large scaffold (e.g., magnetic beads) undoubtedly introduces steric hindrances for binding of library molecules to the target. The use of smaller drag tag, such as fusing another protein of relatively moderate size to the target protein, can be very alluring to induce an appreciable mobility shift while eliminating any pronounced steric effects associated with the use of magnetic beads. The protein tag must be inert, meaning that it does not interfere with the binding of the library molecules to the protein target of interest. In such fusion-protein system, the linker region should be long enough to minimize the steric hindrance between the tag and the protein of interest. One good candidate to act as a protein drag tag in selection of binders from random DNA libraries is

green fluorescent protein (GFP, 27 kDa), which has been demonstrated to be inert in binding to DNA due to its acidic isoelectric point and smooth surface topography.¹³³ Moreover, the use of fluorescent protein tag like GFP also bring in an additional benefit of providing a means of fluorescently detecting the complexes without the need to fluorescently-label the DNA library. The major challenge in this approach is that it requires a different set of expertise and experimental efforts to generate a well-designed fusion protein system for every protein target of interest.

In addition to the above methodology-related challenges, the main practical consideration that remain partially unresolved in the adoption of KCE-based selection is the restricted variety of KCE compatible buffers. Complex-first NECEEM experiments are limited to a narrow set of low ionic strength and low ionic mobility running buffers, with Tris and tetraborate solutions of 20–50 mM ionic strength being the most common. The choice of these buffers can be justified by their resulting in low Joule heating and slow time of ion depletion under the condition of strong electric field in complex-first NECEEM. In the course of developing IFCE (as discussed in **Chapter 2**), we effectively addressed the issue of Joule heating associated with the use of physiological running buffers. We achieved this by employing SUMET, an in-capillary temperature determination method, to determine an appropriate electric field strength that would yield the desired in-capillary temperature during the separation process.⁸⁴ This innovative technique allowed for the utilization of high ionic strength buffers in complex-last NECEEM (see **Section 5.4.1**). However, it comes with a certain compromise. Under relatively low electric field strengths, both IFCE and complex-last NECEEM suffers from lengthy separation time and, thus, extensive complex dissociation. To expedite the separation process and maintain the integrity of complexes, a constant pressure-driven hydrodynamic flow can be employed during

electrophoresis.⁹⁴ However, this approach introduces an undesirable increase in peak broadening, particularly for the separation of nonbinders from target–binder complexes in the case of small-sized targets. While the Joule heating problem can be solved conveniently by systematic adjustment of the applied electric field using SUMET, the extended separation time and high degree of binder loss still pose a challenge to the use of physiological running buffer in KCE and limit the choice of targets to large proteins.

CONCLUDING STATEMENTS

Two decades since its inception, KCE has emerged as a feasible partitioning platform for screening oligonucleotide libraries to identify protein binders. Compared to the conventional surface-based techniques, KCE allows the interaction and separation to take place in free solution; thus, eliminating all the sources of inefficiencies associated with surface immobilization, chemical labelling, and non-specific surface adhesion of the interacting molecules. Despite its remarkable efficiency, screening oligonucleotide libraries using KCE was initially regarded as more art than science, presenting numerous challenges that needed resolution before KCE could rival conventional surface-based methods.

Over the course of my research journey, I have transformed KCE into a well-understood and quantitative methodological platform for screening of oligonucleotide libraries. My primary achievements can be summarized as follows:

- (i) **Minimized nonbinder background:** I successfully addressed the issue of nonbinder background and further improved the partitioning efficiency in KCE by introducing IFCE, a novel KCE-based partitioning method conducted using physiological buffers. With IFCE, partitioning efficiency can reach up to 10^9 , a thousand times higher than conventional KCE-based methods and a million times higher than typical surface-based techniques, enabling single-round selection of high-affinity binders. To complement IFCE, I also introduced double-passage KCE approach, uniquely allowing the measurement of affinity and kinetics of target–binder complexes selected under physiological or IFCE conditions.
- (ii) **Enhanced compatibility with physiological buffers:** I resolved the problem of KCE-based selection's compatibility with physiological buffers by utilizing SUMET, an in-capillary

temperature determination method, to rationalize the choice of the electric field strength during separation. This breakthrough allows for the seamless performance of KCE-based partitioning in high ionic strength buffers, as demonstrated in IFCE and certain cases of complex-last NECEEM (see **Section 5.4.1** for optimizing the KCE conditions to study binding of sEH and its DEL ligands).

- (iii) Quantitative characterization: I provided a comprehensive and quantitative characterization of all the KCE-based partitioning methods, guiding CE users on how to identify and optimize appropriate KCE-based partitioning conditions for a given protein target.
- (iv) Optimized target concentration: To reduce the uncertainty in the general procedures of binder selection, I introduced a quantitative experimental parameter that can be readily determined to identify an appropriate range of target concentration, avoiding selection failures.

While some challenges remain (e.g., low input capacity, limited mobility shift for complexes of oligonucleotide binders and small-size protein targets, and compromised separation time and binder loss when using physiological buffers), the observed results from proposed improvements in KCE-based selections still outweigh the limitations in many proven cases.^{41,42,58-60,107} With further advances in the screening capacity, KCE has the potential to become more universal and robust. While the past 20 years of KCE-based selection has been mainly devoted to generating aptamers from random-sequence oligonucleotide libraries, our successful demonstration of highly efficient KCE-based selection of protein binders from DNA-encoded library of small molecules marks a significant entry for KCE-based partitioning methodology into the field of pharmaceutical drug discovery and development platform.

LIST OF PUBLICATIONS

10. Wang, T.Y.; Ji, W.; Everton, D.; Le, A.T.H.; Krylova, S.M.; Fournier, R.; Krylov, S.N. Fundamental determinants of the accuracy of equilibrium constants for affinity complexes. *Analytical Chemistry* **2023**, *95*, 15826–15832.
9. Le, A.T.H.; Krylova, S.M.; Krylov, S.N. Kinetic capillary electrophoresis in screening oligonucleotide libraries for protein binders. *TrAC – Trends in Analytical Chemistry* **2023**, *162*, 117061.
8. Teclmichael, E.; Le, A.T.H.; Krylova, S.M.; Wang, T.Y.; Krylov, S.N. Bulk affinity assays in aptamer selection: challenges, theory, and workflow. *Analytical Chemistry* **2022**, *94*, 15183–15188.
7. Wang, T.Y.; Rukundo, J.-L.; Le, A.T.H.; Ivanov, N.A.; Le Blanc, J.C.Y.; Gorin, B.I. Krylov, S.N. Transient incomplete separation of species with close diffusivity to study stability of affinity complexes. *Analytical Chemistry* **2022**, *94*, 15415–15422.
6. Le, A.T.H.; Wang, T.Y.; Krylova, S.M.; Beloborodov, S.S.; Krylov, S.N. Quantitative characterization of partitioning in selection of DNA aptamers for protein targets by capillary electrophoresis. *Analytical Chemistry* **2022**, *94*, 2578–2588.
5. Le, A.T.H.; Krylova, S.M.; Beloborodov, S.S.; Wang, T.Y.; Hili, R.; Johnson, P.E.; Li, F.; Veedu, R.N.; Belyanskaya, S.; Krylov, S.N. How to develop and prove high-efficiency selection of ligands from oligonucleotide libraries: a universal framework for aptamers and DNA-encoded small-molecule ligands. *Analytical Chemistry* **2021**, *93*, 5343–5354.
4. Le, A.T.H.; Krylova, S.M.; Krylov, S.N. Determination of the equilibrium constant and rate constant of protein–oligonucleotide complex dissociation under the conditions of ideal-filter capillary electrophoresis. *Analytical Chemistry* **2019**, *91*, 8532–8539.
3. Le, A.T.H.; Krylova, S.M.; Krylov, S.N. Ideal-filter capillary electrophoresis (IFCE): a highly efficient partitioning method for selection of protein binders from oligonucleotide libraries. *Electrophoresis* **2019**, *40*, 2553–2564.
2. Le, A.T.H.; Krylova, S.M.; Kanoatov, M.; Desai, S.; Krylov, S.N. Ideal-filter capillary electrophoresis (IFCE) facilitates the one-step selection of aptamers. *Angewandte Chemie International Edition* **2019**, *58*, 2739–2743.
1. Kochmann, S.; Le, A.T.H.; Hili, R.; Krylov, S.N. Predicting efficiency of NECEEM-based partitioning of protein binders from non-binders in DNA-encoded libraries. *Electrophoresis* **2018**, *39*, 2991–2996.

REFERENCES

1. Liu, R.; Li, X.; Lam, K. S. Combinatorial chemistry in drug discovery. *Current Opinion in Chemical Biology* **2017**, *38*, 117-126.
2. Li, J.; Murray, C. W.; Waszkowycz, B.; Young, S. C. Targeted molecular diversity in drug discovery: Integration of structure-based design and combinatorial chemistry. *Drug Discovery Today* **1998**, *3*, 105-112.
3. Plunkett, M. J.; Ellman, J. A. Combinatorial Chemistry and New Drugs. *Scientific American* **1997**, *276*, 68-73.
4. Furka, Á. Forty years of combinatorial technology. *Drug Discovery Today* **2022**, *27*, 103308.
5. Quartararo, A. J.; Gates, Z. P.; Somsen, B. A.; Hartrampf, N.; Ye, X.; Shimada, A.; Kajihara, Y.; Ottmann, C.; Pentelute, B. L. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. *Nature Communications* **2020**, *11*, 3183.
6. 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., et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature Chemical Biology* **2009**, *5*, 647-654.
7. (Kaur, S.; McGuire, L.; Tang, D.; Dollinger, G.; Huebner, V. Affinity Selection and Mass Spectrometry-Based Strategies to Identify Lead Compounds in Combinatorial Libraries. *Journal of Protein Chemistry* **1997**, *16*, 505-511.
8. Evans, D. M.; Williams, K. P.; McGuinness, B.; Tarr, G.; Regnier, F.; Afeyan, N.; Jindal, S. Affinity-based screening of combinatorial libraries using automated, serial-column chromatography. *Nature Biotechnology* **1996**, *14*, 504-507.
9. Vant-Hull, B.; Gold, L.; Zichi, D. A. Theoretical Principles of In Vitro Selection Using Combinatorial Nucleic Acid Libraries. *Current Protocols in Nucleic Acid Chemistry* **2000**, *00*, 9.1.1-9.1.16.
10. Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. In Vitro Selection of a DNA-Templated Small-Molecule Library Reveals a Class of Macrocyclic Kinase Inhibitors. *Journal of the American Chemical Society* **2010**, *132*, 11779-11791.
11. Ecker, D. J.; Vickers, T. A.; Hanecak, R.; Driver, V.; Anderson, K. Rational screening of oligonucleotide combinatorial libraries for drug discovery. *Nucleic Acids Research* **1993**, *21*, 1853-1856.
12. Blind, M.; Blank, M. Aptamer Selection Technology and Recent Advances. *Molecular Therapy Nucleic Acids* **2015**, *4*, e223.

13. Franzini, R. M.; Ekblad, T.; Zhong, N.; Wichert, M.; Decurtins, W.; Nauer, A.; Zimmermann, M.; Samain, F.; Scheuermann, J.; Brown, P. J.; Hall, J.; Gräslund, S.; Schüler, H.; Neri, D. Identification of Structure–Activity Relationships from Screening a Structurally Compact DNA-Encoded Chemical Library. *Angewandte Chemie International Edition* **2015**, *54*, 3927-3931.
14. Zimmermann, G.; Li, Y.; Rieder, U.; Mattarella, M.; Neri, D.; Scheuermann, J. Hit-Validation Methodologies for Ligands Isolated from DNA-Encoded Chemical Libraries. *ChemBioChem* **2017**, *18*, 853-857.
15. Papoulas, O. Rapid Separation of Protein-Bound DNA from Free DNA Using Nitrocellulose Filters. *Current Protocols in Molecular Biology* **1996**, *36*, 12.18.11-12.18.19.
16. Vijayendran, R. A.; Leckband, D. E. A Quantitative Assessment of Heterogeneity for Surface-Immobilized Proteins. *Analytical Chemistry* **2001**, *73*, 471-480.
17. Nieba, L.; Krebber, A.; Plückthun, A. Competition BIAcore for Measuring True Affinities: Large Differences from Values Determined from Binding Kinetics. *Analytical Biochemistry* **1996**, *234*, 155-165.
18. Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818-822.
19. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505-510.
20. Irvine, D.; Tuerk, C.; Gold, L. Selexion: Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *Journal of Molecular Biology* **1991**, *222*, 739-761.
21. Ciesiolka, J.; Illangasekare, M.; Majerfeld, I.; Nickles, T.; Welch, M.; Yarus, M.; Zinnen, S. *Methods in Enzymology* **1996**, *267*, 315-335.
22. Wang, J.; Rudzinski, J. F.; Gong, Q.; Soh, H. T.; Atzberger, P. J. Influence of Target Concentration and Background Binding on In Vitro Selection of Affinity Reagents. *PLOS One* **2012**, *7*, e43940.
23. Multer, G. L.; Boynton, K. A. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. *Nucleic Acids Research* **1995**, *23*, 1411-1418.
24. Warnecke, P. M.; Stirzaker, C.; Melki, J. R.; Millar, D. S.; Paul, C. L.; Clark, S. J. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Research* **1997**, *25*, 4422-4426.

25. Takahashi, M.; Wu, X.; Ho, M.; Chomchan, P.; Rossi, J. J.; Burnett, J. C.; Zhou, J. High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Scientific Reports* **2016**, *6*, 33697.
26. Berezovski, M.; Musheev, M.; Drabovich, A.; Krylov, S. N. Non-SELEX Selection of Aptamers. *Journal of the American Chemical Society* **2006**, *128*, 1410-1411.
27. Berezovski, M. V.; Musheev, M. U.; Drabovich, A. P.; Jitkova, J. V.; Krylov, S. N. Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides. *Nature Protocols* **2006**, *1*, 1359-1369.
28. Goodnow, R. A.; Dumelin, C. E.; Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery* **2017**, *16*, 131-147.
29. Chan, A. I.; McGregor, L. M.; Liu, D. R. Novel selection methods for DNA-encoded chemical libraries. *Current Opinion in Chemical Biology* **2015**, *26*, 55-61.
30. Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T.; Flather, D.; Forbes, A.; Foreman, T.; Fowler, C.; Gawande, B.; Goss, M.; Gunn, M.; Gupta, S.; Halladay, D.; Heil, J., et al. Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLOS One* **2010**, *5*, e15004.
31. Berezovski, M.; Drabovich, A.; Krylova, S. M.; Musheev, M.; Okhonin, V.; Petrov, A.; Krylov, S. N. Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures: A Universal Tool for Development of Aptamers. *Journal of the American Chemical Society* **2005**, *127*, 3165-3171.
32. Drabovich, A.; Berezovski, M.; Krylov, S. N. Selection of Smart Aptamers by Equilibrium Capillary Electrophoresis of Equilibrium Mixtures (ECEEM). *Journal of the American Chemical Society* **2005**, *127*, 11224-11225.
33. Petrov, A.; Okhonin, V.; Berezovski, M.; Krylov, S. N. Kinetic Capillary Electrophoresis (KCE): A Conceptual Platform for Kinetic Homogeneous Affinity Methods. *Journal of the American Chemical Society* **2005**, *127*, 17104-17110.
34. Drabovich, A. P.; Berezovski, M.; Okhonin, V.; Krylov, S. N. Selection of Smart Aptamers by Methods of Kinetic Capillary Electrophoresis. *Analytical Chemistry* **2006**, *78*, 3171-3178.
35. Mendonsa, S. D.; Bowser, M. T. In Vitro Evolution of Functional DNA Using Capillary Electrophoresis. *Journal of the American Chemical Society* **2004**, *126*, 20-21.
36. Mosing, R. K.; Mendonsa, S. D.; Bowser, M. T. Capillary Electrophoresis-SELEX Selection of Aptamers with Affinity for HIV-1 Reverse Transcriptase. *Analytical Chemistry* **2005**, *77*, 6107-6112.

37. Riley, K. R.; Gagliano, J.; Xiao, J.; Libby, K.; Saito, S.; Yu, G.; Cubicciotti, R.; Macosko, J.; Colyer, C. L.; Guthold, M.; Bonin, K. Combining capillary electrophoresis and next-generation sequencing for aptamer selection. *Analytical and Bioanalytical Chemistry* **2015**, *407*, 1527-1532.
38. Riley, K. R.; Saito, S.; Gagliano, J.; Colyer, C. L. Facilitating aptamer selection and collection by capillary transient isotachopheresis with laser-induced fluorescence detection. *Journal of Chromatography A* **2014**, *1368*, 183-189.
39. Saito, S.; Hirose, K.; Tsuchida, M.; Wakui, K.; Yoshimoto, K.; Nishiyama, Y.; Shibukawa, M. Rapid acquisition of high-affinity DNA aptamer motifs recognizing microbial cell surfaces using polymer-enhanced capillary transient isotachopheresis. *Chemical Communications* **2016**, *52*, 461-464.
40. Hirose, K.; Tsuchida, M.; Asakura, H.; Wakui, K.; Yoshimoto, K.; Iida, K.; Sato, M.; Shibukawa, M.; Suganuma, M.; Saito, S. A single-round selection of selective DNA aptamers for mammalian cells by polymer-enhanced capillary transient isotachopheresis. *Analyst* **2017**, *142*, 4030-4038.
41. Wakui, K.; Yoshitomi, T.; Yamaguchi, A.; Tsuchida, M.; Saito, S.; Shibukawa, M.; Furusho, H.; Yoshimoto, K. Rapidly Neutralizable and Highly Anticoagulant Thrombin-Binding DNA Aptamer Discovered by MACE SELEX. *Molecular Therapy - Nucleic Acids* **2019**, *16*, 348-359.
42. Le, A. T. H.; Krylova, S. M.; Kanoatov, M.; Desai, S.; Krylov, S. N. Ideal-Filter Capillary Electrophoresis (IFCE) Facilitates the One-Step Selection of Aptamers. *Angewandte Chemie International Edition* **2019**, *58*, 2739-2743.
43. Luo, Z.; Zhou, H.; Jiang, H.; Ou, H.; Li, X.; Zhang, L. Development of a fraction collection approach in capillary electrophoresis SELEX for aptamer selection. *Analyst* **2015**, *140*, 2664-2670.
44. Zhu, C.; Li, L.; Yang, G.; Fang, S.; Liu, M.; Ghulam, M.; Hao, C.; Chen, Y.; Qu, F. Online reaction based single-step capillary electrophoresis-systematic evolution of ligands by exponential enrichment for ssDNA aptamers selection. *Analytica Chimica Acta* **2019**, *1070*, 112-122.
45. Voeten, R. L. C.; Ventouri, I. K.; Haselberg, R.; Somsen, G. W. Capillary Electrophoresis: Trends and Recent Advances. *Analytical Chemistry* **2018**, *90*, 1464-1481.
46. Galievsky, V. A.; Stasheuski, A. S.; Krylov, S. N. Capillary Electrophoresis for Quantitative Studies of Biomolecular Interactions. *Analytical Chemistry* **2015**, *87*, 157-171.
47. Jorgenson, J. W.; Lukacs, K. D. Zone electrophoresis in open-tubular glass capillaries. *Analytical Chemistry* **1981**, *53*, 1298-1302.

48. Drabovich, A. P.; Berezovski, M. V.; Musheev, M. U.; Krylov, S. N. Selection of Smart Small-Molecule Ligands: The Proof of Principle. *Analytical Chemistry* **2009**, *81*, 490-494.
49. Bao, J.; Krylova, S. M.; Cherney, L. T.; Hale, R. L.; Belyanskaya, S. L.; Chiu, C. H.; Shaginian, A.; Arico-Muendel, C. C.; Krylov, S. N. Predicting Electrophoretic Mobility of Protein–Ligand Complexes for Ligands from DNA-Encoded Libraries of Small Molecules. *Analytical Chemistry* **2016**, *88*, 5498-5506.
50. Le, A. T. H.; Wang, T. Y.; Krylova, S. M.; Beloborodov, S. S.; Krylov, S. N. Quantitative Characterization of Partitioning in Selection of DNA Aptamers for Protein Targets by Capillary Electrophoresis. *Analytical Chemistry* **2022**, *94*, 2578-2588.
51. Kochmann, S.; Le, A. T. H.; Hili, R.; Krylov, S. N. Predicting efficiency of NECEEM-based partitioning of protein binders from nonbinders in DNA-encoded libraries. *Electrophoresis* **2018**, *39*, 2991-2996.
52. Musheev, M. U.; Kanoatov, M.; Krylov, S. N. Non-uniform Velocity of Homogeneous DNA in a Uniform Electric Field: Consequence of Electric-Field-Induced Slow Dissociation of Highly Stable DNA–Counterion Complexes. *Journal of the American Chemical Society* **2013**, *135*, 8041-8046.
53. Belyanskaya, S. L.; Ding, Y.; Callahan, J. F.; Lazaar, A. L.; Israel, D. I. Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *ChemBioChem* **2017**, *18*, 837-842.
54. Schütze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Mörl, M.; Erdmann, V. A.; Lehrach, H.; Konthur, Z.; Menger, M.; Arndt, P. F.; Glökler, J. Probing the SELEX Process with Next-Generation Sequencing. *PLOS One* **2011**, *6*, e29604.
55. Le, A. T. H.; Krylova, S. M.; Beloborodov, S. S.; Wang, T. Y.; Hili, R.; Johnson, P. E.; Li, F.; Veedu, R. N.; Belyanskaya, S.; Krylov, S. N. How to Develop and Prove High-Efficiency Selection of Ligands from Oligonucleotide Libraries: A Universal Framework for Aptamers and DNA-Encoded Small-Molecule Ligands. *Analytical Chemistry* **2021**, *93*, 5343-5354.
56. Mendonsa, S. D.; Bowser, M. T. In Vitro Selection of Aptamers with Affinity for Neuropeptide Y Using Capillary Electrophoresis. *Journal of the American Chemical Society* **2005**, *127*, 9382-9383.
57. Zhu, C.; Yang, G.; Ghulam, M.; Li, L.; Qu, F. Evolution of multi-functional capillary electrophoresis for high-efficiency selection of aptamers. *Biotechnology Advances* **2019**, *37*, 107432.
58. Kasahara, Y.; Irisawa, Y.; Ozaki, H.; Obika, S.; Kuwahara, M. 2',4'-BNA/LNA aptamers: CE-SELEX using a DNA-based library of full-length 2'-O,4'-C-methylene-bridged/linked bicyclic ribonucleotides. *Bioorganic & Medicinal Chemistry Letters* **2013**, *23*, 1288-1292.

59. Hagiwara, K.; Fujita, H.; Kasahara, Y.; Irisawa, Y.; Obika, S.; Kuwahara, M. In vitro selection of DNA-based aptamers that exhibit RNA-like conformations using a chimeric oligonucleotide library that contains two different xeno-nucleic acids. *Molecular BioSystems* **2015**, *11*, 71-76.
60. Ashley, J.; Ji, K.; Li, S. F. Y. Selection of cholesterol esterase aptamers using a dual-partitioning approach. *Electrophoresis* **2015**, *36*, 2616-2621.
61. Li, X.; He, Y.; Ma, Y.; Bie, Z.; Liu, B.; Liu, Z. Hybrid Approach Combining Boronate Affinity Magnetic Nanoparticles and Capillary Electrophoresis for Efficient Selection of Glycoprotein-Binding Aptamers. *Analytical Chemistry* **2016**, *88*, 9805-9812.
62. Wiegand, T. W.; Williams, P. B.; Dreskin, S. C.; Jouvin, M. H.; Kinet, J. P.; Tasset, D. High-affinity oligonucleotide ligands to human IgE inhibit binding to Fc epsilon receptor I. *The Journal of Immunology* **1996**, *157*, 221-230.
63. Yang, J.; Bowser, M. T. Capillary Electrophoresis-SELEX Selection of Catalytic DNA Aptamers for a Small-Molecule Porphyrin Target. *Analytical Chemistry* **2013**, *85*, 1525-1530.
64. Lin, L.; Hom, D.; Lindsay, S. M.; Chaput, J. C. In Vitro Selection of Histone H4 Aptamers for Recognition Imaging Microscopy. *Journal of the American Chemical Society* **2007**, *129*, 14568-14569.
65. Williams, B. A. R.; Lin, L.; Lindsay, S. M.; Chaput, J. C. Evolution of a Histone H4-K16 Acetyl-Specific DNA Aptamer. *Journal of the American Chemical Society* **2009**, *131*, 6330-6331.
66. Rose, C. M.; Hayes, M. J.; Stettler, G. R.; Hickey, S. F.; Axelrod, T. M.; Giustini, N. P.; Suljak, S. W. Capillary electrophoretic development of aptamers for a glycosylated VEGF peptide fragment. *Analyst* **2010**, *135*, 2945-2951.
67. Ashley, J.; Ji, K.; Li, S. F. Y. Selection of bovine catalase aptamers using non-SELEX. *Electrophoresis* **2012**, *33*, 2783-2789.
68. Tok, J.; Lai, J.; Leung, T.; Li, S. F. Y. Selection of aptamers for signal transduction proteins by capillary electrophoresis. *Electrophoresis* **2010**, *31*, 2055-2062.
69. Mendonsa, S. D.; Bowser, M. T. In Vitro Selection of High-Affinity DNA Ligands for Human IgE Using Capillary Electrophoresis. *Analytical Chemistry* **2004**, *76*, 5387-5392.
70. Jing, M.; Bowser, M. T. Tracking the Emergence of High Affinity Aptamers for rhVEGF165 During Capillary Electrophoresis-Systematic Evolution of Ligands by Exponential Enrichment Using High Throughput Sequencing. *Analytical Chemistry* **2013**, *85*, 10761-10770.
71. Duffy, K.; Arangundy-Franklin, S.; Holliger, P. Modified nucleic acids: replication, evolution, and next-generation therapeutics. *BMC Biology* **2020**, *18*, 112.

72. Li, Q.; Zhao, X.; Liu, H.; Qu, F. Low pH capillary electrophoresis application to improve capillary electrophoresis-systematic evolution of ligands by exponential enrichment. *Journal of Chromatography A* **2014**, *1364*, 289-294.
73. Bahga, S. S.; Santiago, J. G. Coupling isotachopheresis and capillary electrophoresis: a review and comparison of methods. *Analyst* **2013**, *138*, 735-754.
74. Nagano, M.; Toda, T.; Makino, K.; Miki, H.; Sugizaki, Y.; Tomizawa, H.; Isobayashi, A.; Yoshimoto, K. Discovery of a Highly Specific Anti-methotrexate (MTX) DNA Aptamer for Antibody-Independent MTX Detection. *Analytical Chemistry* **2022**, *94*, 17255-17262.
75. Musheev, M. U.; Krylov, S. N. Selection of aptamers by systematic evolution of ligands by exponential enrichment: Addressing the polymerase chain reaction issue. *Analytica Chimica Acta* **2006**, *564*, 91-96.
76. Yufa, R.; Krylova, S. M.; Bruce, C.; Bagg, E. A.; Schofield, C. J.; Krylov, S. N. Emulsion PCR Significantly Improves Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures-Based Aptamer Selection: Allowing for Efficient and Rapid Selection of Aptamer to Unmodified ABH2 Protein. *Analytical Chemistry* **2015**, *87*, 1411-1419.
77. Eaton, R. M.; Shallcross, J. A.; Mael, L. E.; Mears, K. S.; Minkoff, L.; Scoville, D. J.; Whelan, R. J. Selection of DNA aptamers for ovarian cancer biomarker HE4 using CE-SELEX and high-throughput sequencing. *Analytical and Bioanalytical Chemistry* **2015**, *407*, 6965-6973.
78. Stuart, C. H.; Riley, K. R.; Boyacioglu, O.; Herpai, D. M.; Debinski, W.; Qasem, S.; Marini, F. C.; Colyer, C. L.; Gmeiner, W. H. Selection of a Novel Aptamer Against Vitronectin Using Capillary Electrophoresis and Next Generation Sequencing. *Molecular Therapy - Nucleic Acids* **2016**, *5*, e386.
79. Alam, K. K.; Chang, J. L.; Burke, D. H. FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Molecular Therapy - Nucleic Acids* **2015**, *4*, e230.
80. Hoinka, J.; Berezhnoy, A.; Sauna, Z. E.; Gilboa, E.; Przytycka, T. M. AptaCluster - A Method to Cluster HT-SELEX Aptamer Pools and Lessons from its Application. *Research in Computational Molecular Biology* **2014**, *8394*, 115-128.
81. Dao, P.; Hoinka, J.; Takahashi, M.; Zhou, J.; Ho, M.; Wang, Y.; Costa, F.; Rossi, J. J.; Backofen, R.; Burnett, J.; Przytycka, T. M. AptaTRACE Elucidates RNA Sequence-Structure Motifs from Selection Trends in HT-SELEX Experiments. *Cell Systems* **2016**, *3*, 62-70.
82. Hüge, B. J.; Flaherty, R. J.; Dada, O. O.; Dovichi, N. J. Capillary electrophoresis coupled with automated fraction collection. *Talanta* **2014**, *130*, 288-293.

83. Beloborodov, S. S.; Krylova, S. M.; Krylov, S. N. Spherical-Shape Assumption for Protein–Aptamer Complexes Facilitates Prediction of Their Electrophoretic Mobility. *Analytical Chemistry* **2019**, *91*, 12680-12687.
84. Patel, K. H.; Evenhuis, C. J.; Cherney, L. T.; Krylov, S. N. Simplified universal method for determining electrolyte temperatures in a capillary electrophoresis instrument with forced-air cooling. *Electrophoresis* **2012**, *33*, 1079-1085.
85. Le, A. T. H.; Krylova, S. M. Determination of the Equilibrium Constant and Rate Constant of Protein-Oligonucleotide Complex Dissociation under the Conditions of Ideal-Filter Capillary Electrophoresis. *Analytical Chemistry* **2019**, *91*, 8532-8539.
86. Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Bühlmann, P.; Supuran, C. T.; Neri, D. Selection of Carbonic Anhydrase IX Inhibitors from One Million DNA-Encoded Compounds. *ACS Chemical Biology* **2011**, *6*, 336-344.
87. Liu, Y.; Wang, C.; Li, F.; Shen, S.; Tyrrell, D. L.; Le, X. C.; Li, X. F. DNase-mediated single-cycle selection of aptamers for proteins blotted on a membrane. *Analytical Chemistry* **2012**, *84*, 7603-7606.
88. Imashimizu, M.; Takahashi, M.; Amano, R.; Nakamura, Y. Single-round isolation of diverse RNA aptamers from a random sequence pool. *Biology Methods and Protocols* **2018**, *3*.
89. Lauridsen, L. H.; Shamaileh, H. A.; Edwards, S. L.; Taran, E.; Veedu, R. N. Rapid one-step selection method for generating nucleic acid aptamers: development of a DNA aptamer against α -bungarotoxin. *PLOS One* **2012**, *7*, e41702.
90. Peng, L.; Stephens, B. J.; Bonin, K.; Cubicciotti, R.; Guthold, M. A combined atomic force/fluorescence microscopy technique to select aptamers in a single cycle from a small pool of random oligonucleotides. *Microscopy Research and Technique* **2007**, *70*, 372-381.
91. Nitsche, A.; Kurth, A.; Dunkhorst, A.; Pänke, O.; Sielaff, H.; Junge, W.; Muth, D.; Scheller, F.; Stöcklein, W.; Dahmen, C.; Pauli, G.; Kage, A. One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. *BMC Biotechnol* **2007**, *7*, 48.
92. Lou, X.; Qian, J.; Xiao, Y.; Viel, L.; Gerdon, A. E.; Lagally, E. T.; Atzberger, P.; Tarasow, T. M.; Heeger, A. J.; Soh, H. T. Micromagnetic selection of aptamers in microfluidic channels. *Proceedings of the National Academy of Sciences USA* **2009**, *106*, 2989-2994.
93. Beloborodov, S. S.; Bao, J.; Krylova, S. M.; Shala-Lawrence, A.; Johnson, P. E.; Krylov, S. N. Aptamer facilitated purification of functional proteins. *Journal of Chromatography B* **2018**, *1073*, 201-206.
94. Kanoatov, M.; Krylov, S. N. Analysis of DNA in Phosphate Buffered Saline Using Kinetic Capillary Electrophoresis. *Analytical Chemistry* **2016**, *88*, 7421-7428.

95. Kanoatov, M.; Galievsky, V. A.; Krylova, S. M.; Cherney, L. T.; Jankowski, H. K.; Krylov, S. N. Using Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM) for Simultaneous Determination of Concentration and Equilibrium Constant. *Analytical Chemistry* **2015**, *87*, 3099-3106.
96. Sisavath, N.; Rukundo, J.-L.; Le Blanc, J. C. Y.; Galievsky, V. A.; Bao, J.; Kochmann, S.; Stasheuski, A. S.; Krylov, S. N. Transient Incomplete Separation Facilitates Finding Accurate Equilibrium Dissociation Constant of Protein–Small Molecule Complex. *Angewandte Chemie International Edition* **2019**, *58*, 6635-6639.
97. Chamieh, J.; Leclercq, L.; Martin, M.; Slaoui, S.; Jensen, H.; Østergaard, J.; Cottet, H. Limits in Size of Taylor Dispersion Analysis: Representation of the Different Hydrodynamic Regimes and Application to the Size-Characterization of Cubosomes. *Analytical Chemistry* **2017**, *89*, 13487-13493.
98. Kanoatov, M.; Krylov, S. N. DNA Adsorption to the Reservoir Walls Causing Irreproducibility in Studies of Protein–DNA Interactions by Methods of Kinetic Capillary Electrophoresis. *Analytical Chemistry* **2011**, *83*, 8041-8045.
99. de Jong, S.; Krylov, S. N. Pressure-based approach for the analysis of protein adsorption in capillary electrophoresis. *Analytical Chemistry* **2012**, *84*, 453-458.
100. Desruisseaux, C.; Long, D.; Drouin, G.; Slater, G. W. Electrophoresis of Composite Molecular Objects. 1. Relation between Friction, Charge, and Ionic Strength in Free Solution. *Macromolecules* **2001**, *34*, 44-52.
101. Dahl, G.; Akerud, T. Pharmacokinetics and the drug–target residence time concept. *Drug Discovery Today* **2013**, *18*, 697-707.
102. Eaton, B. E. The joys of in vitro selection: chemically dressing oligonucleotides to satiate protein targets. *Current Opinion in Chemical Biology* **1997**, *1*, 10-16.
103. Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E. Expanding the Chemistry of DNA for in Vitro Selection. *Journal of the American Chemical Society* **2010**, *132*, 4141-4151.
104. Kong, D.; Yeung, W.; Hili, R. In Vitro Selection of Diversely Functionalized Aptamers. *Journal of the American Chemical Society* **2017**, *139*, 13977-13980.
105. Levine, H. A.; Nilsen-Hamilton, M. A mathematical analysis of SELEX. *Computational Biology and Chemistry* **2007**, *31*, 11-35.
106. Spill, F.; Weinstein, Z. B.; Irani Shemirani, A.; Ho, N.; Desai, D.; Zaman, M. H. Controlling uncertainty in aptamer selection. *Proceedings of the National Academy of Sciences USA* **2016**, *113*, 12076-12081.
107. Martínez-Roque, M. A.; Franco-Urquijo, P. A.; García-Velásquez, V. M.; Choukeife, M.; Mayer, G.; Molina-Ramírez, S. R.; Figueroa-Miranda, G.; Mayer, D.; Alvarez-Salas, L.

- M. DNA aptamer selection for SARS-CoV-2 spike glycoprotein detection. *Analytical Biochemistry* **2022**, *645*, 114633.
108. Yang, G.; Li, Z.; Mohammed, I.; Zhao, L.; Wei, W.; Xiao, H.; Guo, W.; Zhao, Y.; Qu, F.; Huang, Y. Identification of SARS-CoV-2-against aptamer with high neutralization activity by blocking the RBD domain of spike protein 1. *Signal Transduction and Targeted Therapy* **2021**, *6*, 227.
109. Teclemichael, E.; Le, A. T. H.; Krylova, S. M.; Wang, T. Y.; Krylov, S. N. Bulk Affinity Assays in Aptamer Selection: Challenges, Theory, and Workflow. *Analytical Chemistry* **2022**, *94*, 15183-15188.
110. Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992**, *355*, 564-566.
111. Tasset, D. M.; Kubik, M. F.; Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *Journal of Molecular Biology* **1997**, *272*, 688-698.
112. Liyanage, R.; Krylova, S. M.; Krylov, S. N. Minimizing adsorption of histidine-tagged proteins for the study of protein-deoxyribonucleic acid interactions by kinetic capillary electrophoresis. *Journal of Chromatography A* **2013**, *1322*, 90-96.
113. Hughes, J.; Rees, S.; Kalindjian, S.; Philpott, K. Principles of early drug discovery. *British Journal of Pharmacology* **2011**, *162*, 1239-1249.
114. Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Accounts of Chemical Research* **2009**, *42*, 724-733.
115. Keserü, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discovery Today* **2006**, *11*, 741-748.
116. Gironda-Martínez, A.; Donckele, E. J.; Samain, F.; Neri, D. DNA-Encoded Chemical Libraries: A Comprehensive Review with Successful Stories and Future Challenges. *ACS Pharmacology & Translational Science* **2021**, *4*, 1265-1279.
117. Brenner, S.; Lerner, R. A. Encoded combinatorial chemistry. *Proceedings of the National Academy of Sciences USA* **1992**, *89*, 5381-5383.
118. Pfaffl, M. W.; Hageleit, M. Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnology Letters* **2001**, *23*, 275-282.
119. Franzini, R. M.; Neri, D.; Scheuermann, J. DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries. *Accounts of Chemical Research* **2014**, *47*, 1247-1255.

120. Tse, B. N.; Snyder, T. M.; Shen, Y.; Liu, D. R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection. *Journal of American Chemistry Society* **2008**, *130*, 15611-15626.
121. Hansen, M. H.; Blakskjaer, P.; Petersen, L. K.; Hansen, T. H.; Højfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. A yoctoliter-scale DNA reactor for small-molecule evolution. *Journal of American Chemistry Society* **2009**, *131*, 1322-1327.
122. Halpin, D. R.; Lee, J. A.; Wrenn, S. J.; Harbury, P. B. DNA Display III. Solid-Phase Organic Synthesis on Unprotected DNA. *PLOS Biology* **2004**, *2*, e175.
123. Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded self-assembling chemical libraries. *Nature Biotechnology* **2004**, *22*, 568-574.
124. Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions. *Bioorganic and Medicinal Chemistry Letters* **2008**, *18*, 5926-5931.
125. Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proceedings of the National Academy of Sciences USA* **2008**, *105*, 17670-17675.
126. Gironde-Martínez, A.; Donckele, E. J. DNA-Encoded Chemical Libraries: A Comprehensive Review with Successful Stories and Future Challenges. *ACS Pharmacology & Translational Science* **2021**, *4*, 1265-1279.
127. Kleiner, R. E.; Dumelin, C. E.; Liu, D. R. Small-molecule discovery from DNA-encoded chemical libraries. *Chemical Society Reviews* **2011**, *40*, 5707-5717.
128. Doyon, J. B.; Snyder, T. M.; Liu, D. R. Highly sensitive in vitro selections for DNA-linked synthetic small molecules with protein binding affinity and specificity. *Journal of American Chemistry Society* **2003**, *125*, 12372-12373.
129. Gorin, D. J.; Kamlet, A. S.; Liu, D. R. Reactivity-dependent PCR: direct, solution-phase in vitro selection for bond formation. *Journal of American Chemistry Society* **2009**, *131*, 9189-9191.
130. McGregor, L. M.; Gorin, D. J.; Dumelin, C. E.; Liu, D. R. Interaction-dependent PCR: identification of ligand-target pairs from libraries of ligands and libraries of targets in a single solution-phase experiment. *Journal of American Chemistry Society* **2010**, *132*, 15522-15524.
131. Supuran, C. T.; Scozzafava, A. Carbonic anhydrase inhibitors: aromatic sulfonamides and disulfonamides act as efficient tumor growth inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2000**, *15*, 597-610.

132. Podolin, P. L.; Bolognese, B. J.; Foley, J. F.; Long, E., 3rd; Peck, B.; Umbrecht, S.; Zhang, X.; Zhu, P.; Schwartz, B.; Xie, W.; Quinn, C.; Qi, H.; Sweitzer, S.; Chen, S.; Galop, M.; Ding, Y.; Belyanskaya, S. L.; Israel, D. I.; Morgan, B. A.; Behm, D. J., et al. In vitro and in vivo characterization of a novel soluble epoxide hydrolase inhibitor. *Prostaglandins & Other Lipid Mediators* **2013**, *104-105*, 25-31.
133. Stanlis, K. K. H.; McIntosh, J. R. Single-strand DNA Aptamers as Probes for Protein Localization in Cells. *Journal of Histochemistry & Cytochemistry* **2003**, *51*, 797-808.

APPENDICES

APPENDIX A. SUPPORTING INFORMATION FOR IDEAL-FILTER CAPILLARY ELECTROPHORESIS (PROOF-OF-PRINCIPLE WORK)

A1. Determination of elution windows of the MutS–aptamer complex

Aptamer-collection time windows were determined as widths of the bases of peaks of MutS–aptamer complexes in CE experiments performed with NaCl-free RB and fluorescence detection (**Figure A1**).

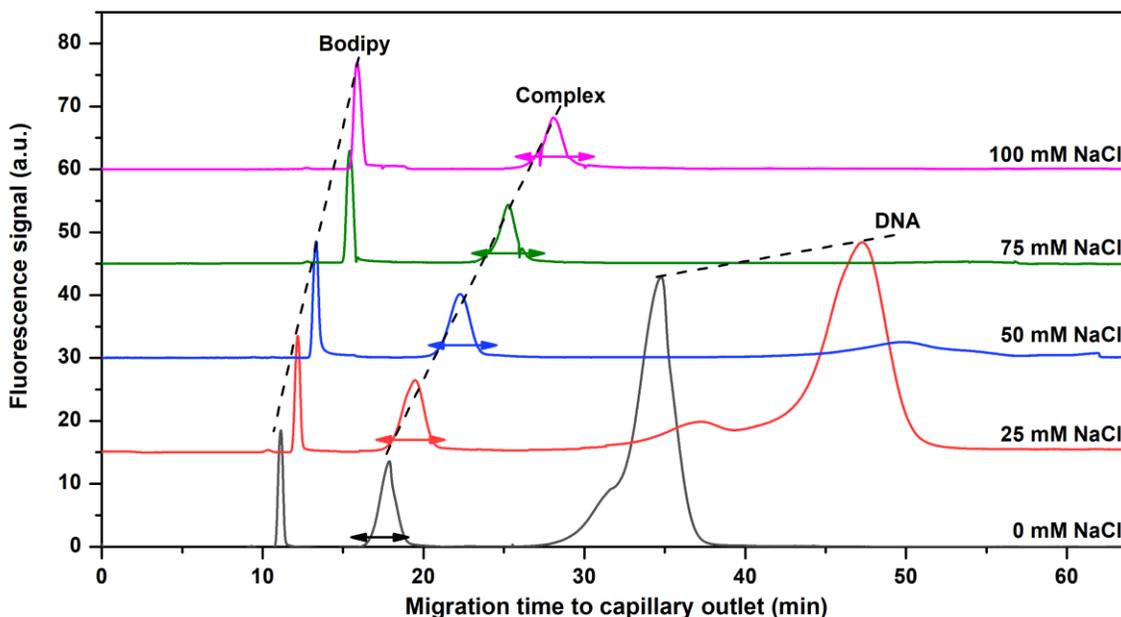


Figure A1. Determination of time windows for MutS–aptamer complex collection for RBs based on 50 mM Tris-HCl pH 7.0 and containing concentrations of NaCl varying from 0 to 100 mM. The equilibrium mixture contained 100 nM MutS, 100 nM fluorescently-labeled aptamer, and 150 nM Bodipy (EOF marker). Electrophoresis was carried out with an electric field of 200 V/cm, “+” at the capillary inlet, and a separation distance of 34 cm. The double-headed arrows indicate estimated elution windows of the MutS–aptamer complex.

A2. Details on the determination of efficiency of binder collection

We found k_B experimentally as $k_B = B_{\text{out}}/B_{\text{in}}$, where B_{in} is the number of target–binder complexes sampled and B_{out} is the number of binders collected in the binder-collection time window corresponding to the elution time window of the target–binder complexes. We first used NaCl-free RB in which the target–binder complexes and nonbinders migrate in the same direction (complex-first NECEEM) allowing for very accurate determination of B_{in} by using fluorescence detection. A known volume of the equilibrium mixture containing MutS and its DNA aptamer was sampled for a CE run, in which both fluorescence and qPCR detections were used leading to two electropherograms for each CE run (**Figure A2a, b**).

The value of $B_{\text{in}} = (8.9 \pm 0.9) \times 10^8$ was found from an electropherogram with fluorescence detection shown in **Figure A2a** as the multiplication product of (i) the total concentration of aptamer in the equilibrium mixture, (ii) the volume of the sampled equilibrium mixture, and (iii) a relative amount of aptamer that was bound to MutS in the equilibrium mixture (relative to the total sampled amount of aptamer). We made no correction for complex dissociation during CE as less than 1% of the MutS–aptamer complex dissociated during complex migration to the capillary exit. The degree of complex dissociation during KCE-based partitioning of MutS–aptamer complex from the unbound aptamer was assessed *via* determination of the value of the rate constant of complex dissociation (k_{off}):

$$k_{\text{off}} = \frac{\ln((A_C + A_D)/A_C)}{t_{\text{detection}}} = 0.003 \text{ min}^{-1}$$

where A_D and A_C are areas under the electropherograms segments (**Figure A2a**) corresponding to complex decay and the intact complex at the time of its detection, $t_{\text{detection}}$, respectively. The quantity of intact complex at the capillary outlet (B_{out}) was calculated as:

$$B_{\text{out}} = B_{\text{detected}} \exp(-k_{\text{off}}(t_{\text{out}} - t_{\text{detection}})) = 8.8 \times 10^8$$

Here, B_{detected} is the amount of complex that was intact at the time of its passing the detector ($B_{\text{detected}} = 8.9 \times 10^8$) and t_{out} is the time when the complex reached the capillary outlet. The value of $(t_{\text{out}} - t_{\text{detection}})$ was found to be 4 min. Thus, less than 1% of the complex dissociated during complex migration to the outlet from the detector.

The value of $B_{\text{out}} = (9.2 \pm 0.4) \times 10^8$ was then determined from the electropherogram with qPCR detection shown in **Figure A2b** by calculating an integral under the DNA curve within the binder-collection time window of 15 to 23 min. Knowing B_{out} and B_{in} allowed us to calculate $k_{\text{B}} \equiv B_{\text{out}} / B_{\text{in}} = 1.0 \pm 0.1$.

We then conducted an IFCE experiment (with RB containing 100 mM NaCl) and found $B_{\text{out}} = (7.3 \pm 0.5) \times 10^8$, which is slightly lower than the value of $(9.2 \pm 0.4) \times 10^8$ obtained for NaCl-free RB. The difference was presumably due to additional ions' in IFCE affecting the strength of ionic bonds in protein–DNA complexes. These experiments demonstrated that by choosing a proper binder-collection time window we are able to satisfy our assumption of $k_{\text{B}} \approx 1$.

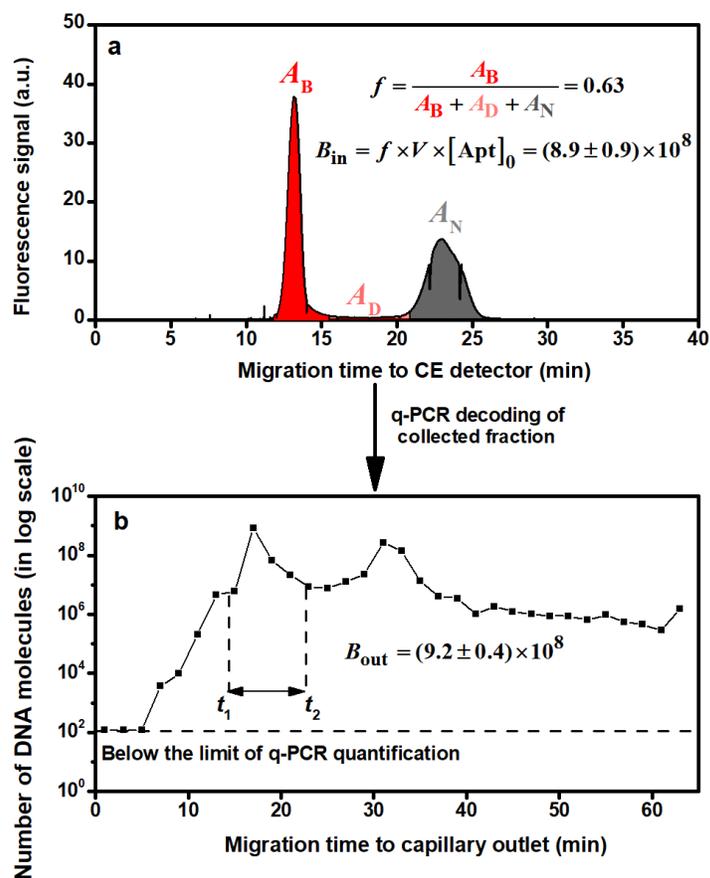


Figure A2. Determination of B_{in} (**a**) and B_{out} (**b**) for IFCE-based partitioning of MutS-bound aptamer and the unbound aptamer. A known volume, $V = 47$ nL, of the equilibrium mixture containing 200 nM MutS and 50 nM fluorescently-labeled aptamer ($[Apt]_0 = 50$ nM) was injected and CE was carried out with an electric field of 200 V/cm with a separation distance of 34 cm. Fluorescence detection was used to obtain an ordinary electropherogram shown in Panel **a**. Two-min fractions were collected and the number of aptamer copies was determined in each fraction by qPCR to produce an electropherogram in Panel **b**. A relative amount of MutS-bound aptamer (f) was found from the shaded areas in Panel **a**, and B_{in} was determined with formulas shown within the panel. B_{out} was calculated as an integral under the curve in Panel **b** within the binder-collection time window $t_1 - t_2$.

A3. Evaluation of binding affinity of aptamer pool selected for MutS

The equilibrium mixture of the enriched DNA library with 100 nM MutS was subjected to pressure-aided IFCE with fluorescent detection. A distinct peak corresponding to MutS–DNA complexes was observed in a time window of 7 to 9 min followed by a smaller peak of unbound DNA with a maximum at 11 min. The apparent equilibrium dissociation constant of $K_{d,app} \approx 40$ nM for the interaction of the enriched library with MutS protein was found from

Figure A3.

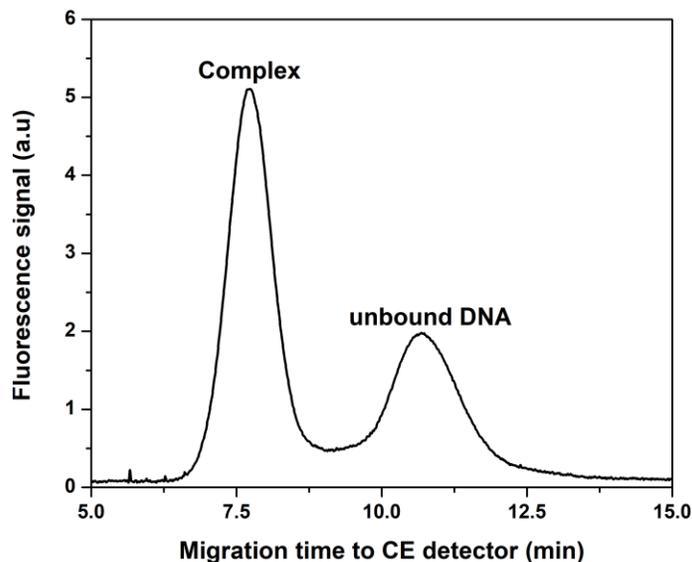


Figure A3. Pressure-aided IFCE separation of the components of equilibrium mixture containing 20 nM fluorescently labeled enriched DNA library and 100 nM MutS. The running buffer was 50 mM TrisHCl pH 7.0, 100 mM NaCl. CE was carried out with an electric field of 200 V/cm (“+” at the inlet) with a pressure supplement of 0.2 psi. The separation distance was 34 cm.

APPENDIX B. SUPPORTING INFORMATION FOR DOUBLE-PASSAGE APPROACH

B1. Calculating R from the velocity-corrected peak areas via double-passage approach

$$R = \frac{[L]_{\text{eq}}}{[L]_0} = \frac{[L]_{\text{eq}}}{[PL]_{\text{eq}} + [L]_{\text{eq}}} = \frac{(\phi[PL]_{\text{eq}} + [L]_{\text{eq}}) - \phi[PL]_{\text{eq}}}{\{\phi[PL]_{\text{eq}} / \phi\} + \{(\phi[PL]_{\text{eq}} + [L]_{\text{eq}}) - \phi[PL]_{\text{eq}}\}} = \frac{A_1 v_{\text{av}} / a - A_2 v_{\text{PL}} / a}{\{(A_2 v_{\text{PL}} / a) / \phi\} + \{A_1 v_{\text{av}} / a - A_2 v_{\text{PL}} / a\}} = \frac{A_1 v_{\text{av}} - A_2 v_{\text{PL}}}{A_2 v_{\text{PL}} / \phi + A_1 v_{\text{av}} - A_2 v_{\text{PL}}}$$

B2. Calculating k_{off} by varying E during the second passage

The value of k_{off} can be determined by using varying t_3 via changing the electric field strength ($E' < E$). A two-point approach (with two different values of t_3 : $t_3 < t_3'$) can be used for a simple assessment of k_{off} :

$$1 - R(t_3) = (1 - R_{t=0})e^{-k_{\text{off}}t_3}, \quad 1 - R(t_3') = (1 - R_{t=0})e^{-k_{\text{off}}t_3'}$$
$$\frac{1 - R(t_3')}{1 - R(t_3)} = e^{-k_{\text{off}}(t_3' - t_3)}, \quad \ln\left(\frac{1 - R(t_3')}{1 - R(t_3)}\right) = -k_{\text{off}}(t_3' - t_3)$$
$$k_{\text{off}} = \frac{\ln\left(\frac{1 - R(t_3)}{1 - R(t_3')}\right)}{t_3' - t_3}$$

B3. Recovery of peak areas

In order to evaluate the recovery of velocity-corrected peak areas via the double-passage approach, we used 1 μM GFP as an analyte mimicking PL in terms of electrophoretic mobility. The experiment was repeated 10 times using the default conditions: $d = 10.2$ cm, $r = 37.5$ μm , $l = 1.09$ cm, pressure equal to 0.30 psi, total capillary length = 50 cm, $E = 200$ V/cm (**Figure B1**). Detailed calculation for recovery of peaks areas can be found from **Table B1**.

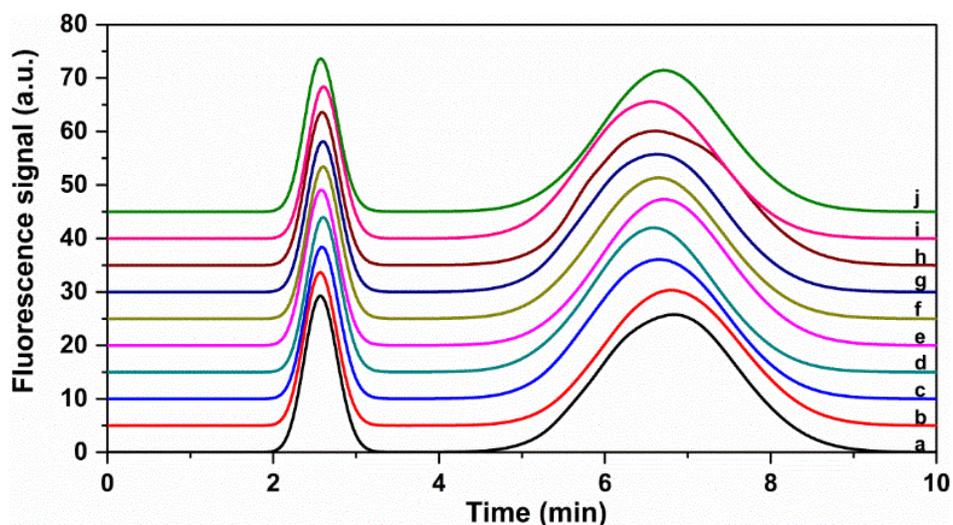


Figure B1. Double-passage experiment performed with 1 μM GFP in IFCE running buffer (50 mM Tris-HCl pH 7.0, 100 mM NaCl). Ten repetitions were done to determine the average velocity-corrected areas and to assess the error.

Table B1. Recovery of peak areas for double-passage experiment with 1 μM GFP.

Figure	A_1	A_2	t_1 min	v_{av} cm/min	t_2 min	d_1 cm	t_3 min	v_{PL} cm/min	$(A_1 v_{av}) / (A_2 v_{PL})$
B1, trace a	8.70×10^7	3.11×10^8	2.85	3.39	0.933	3.16	3.36	0.941	1.01
B1, trace b	8.43×10^7	2.91×10^8	2.85	3.39	0.933	3.16	3.28	0.963	1.02
B1, trace c	8.44×10^7	2.92×10^8	2.87	3.36	0.913	3.07	3.16	0.971	1.00
B1, trace d	8.60×10^7	2.92×10^8	2.88	3.35	0.900	3.01	3.11	0.969	1.02
B1, trace e	8.65×10^7	2.98×10^8	2.86	3.37	0.921	3.11	3.20	0.972	1.01
B1, trace f	8.41×10^7	2.92×10^8	2.88	3.35	0.900	3.01	3.14	0.959	1.01
B1, trace g	8.38×10^7	2.95×10^8	2.88	3.35	0.900	3.01	3.12	0.967	0.98
B1, trace h	8.48×10^7	2.89×10^8	2.88	3.36	0.908	3.05	3.11	0.981	1.00
B1, trace i	8.41×10^7	2.84×10^8	2.89	3.34	0.896	3.00	3.05	0.984	1.01
B1, trace j	8.40×10^7	2.92×10^8	2.85	3.38	0.929	3.14	3.21	0.980	0.99

B4. Recovery of R

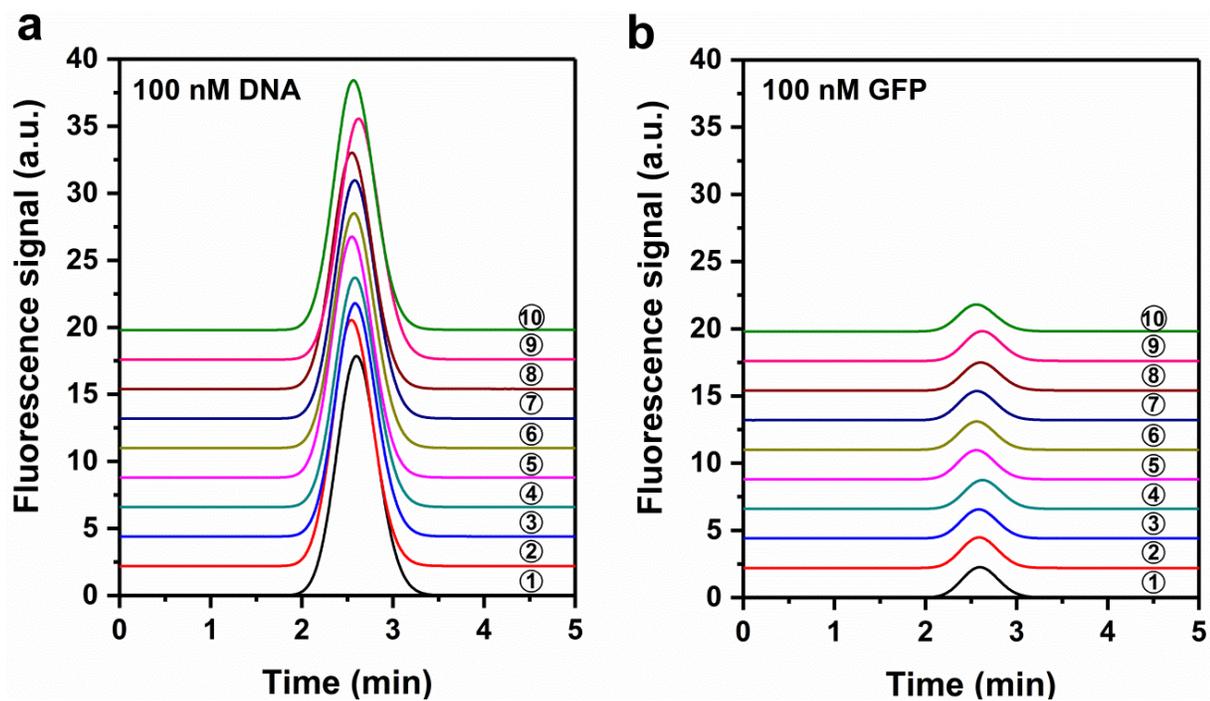


Figure B2. Determination of ϕ for GFP–DNA mixture. Areas of peaks of GFP and DNA yield a ratio of ≈ 0.11 . Each set of experiments was performed 10 times.

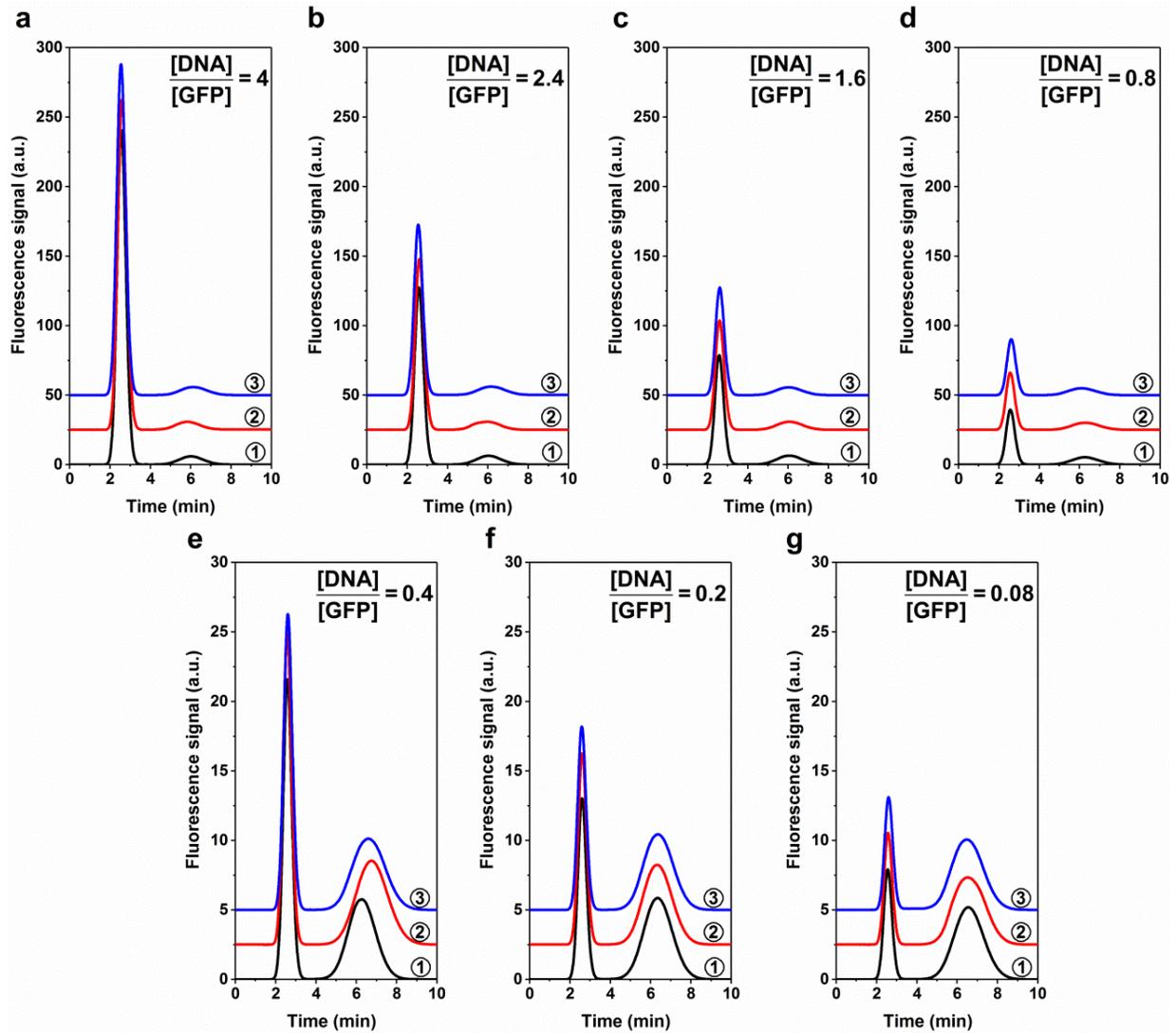


Figure B3. Double-passage experiments of GFP-DNA mixtures with $\frac{[DNA]}{[GFP]}$ ratios ranging from 0.08 to 4. Each set of experiments was performed in triplicates.

Table B2. Recovery of R for double-passage experiment of GFP–DNA mixtures.

$\frac{[\text{DNA}]}{[\text{GFP}]}$	Figure	A_1	A_2	t_1 min	v_{av} cm/min	t_2 min	d_1 cm	t_3 min	v_{PL} cm/min	R	Mean $R \pm \text{SD}$
4	B3-a, 1	7.88×10^8	5.18×10^7	2.85	3.39	0.937	3.18	2.52	1.26	0.814	
	B3-a, 2	7.84×10^8	4.93×10^7	2.88	3.36	0.908	3.05	2.34	1.30	0.815	0.813 ± 0.003
	B3-a, 3	7.79×10^8	5.44×10^7	2.83	3.41	0.954	3.26	2.64	1.23	0.810	
2.4	B3-b, 1	4.15×10^8	5.70×10^7	2.86	3.37	0.921	3.11	2.54	1.22	0.677	
	B3-b, 2	4.00×10^8	5.65×10^7	2.89	3.34	0.896	3.00	2.43	1.23	0.667	0.672 ± 0.005
	B3-b, 3	3.93×10^8	5.59×10^7	2.83	3.41	0.950	3.24	2.64	1.23	0.671	
1.6	B3-c, 1	2.56×10^8	5.15×10^7	2.87	3.37	0.917	3.09	2.55	1.21	0.585	
	B3-c, 2	2.56×10^8	5.40×10^7	2.88	3.35	0.900	3.01	2.58	1.17	0.580	0.582 ± 0.003
	B3-c, 3	2.51×10^8	5.23×10^7	2.90	3.34	0.888	2.96	2.52	1.18	0.581	
0.8	B3-d, 1	1.27×10^8	5.13×10^7	2.85	3.39	0.933	3.16	2.76	1.15	0.411	
	B3-d, 2	1.33×10^8	5.13×10^7	2.85	3.39	0.937	3.18	2.79	1.14	0.424	0.423 ± 0.011
	B3-d, 3	1.30×10^8	4.84×10^7	2.90	3.34	0.888	2.96	2.62	1.13	0.433	
0.4	B3-e, 1	6.90×10^7	5.80×10^7	2.86	3.37	0.921	3.11	2.78	1.12	0.222	
	B3-e, 2	7.00×10^7	7.03×10^7	2.85	3.38	0.929	3.14	3.23	0.97	0.213	0.225 ± 0.013
	B3-e, 3	6.71×10^7	5.97×10^7	2.89	3.34	0.896	3.00	3.08	0.97	0.240	
0.2	B3-f, 1	4.06×10^7	6.12×10^7	2.88	3.35	0.900	3.01	2.85	1.06	0.108	
	B3-f, 2	4.10×10^7	5.69×10^7	2.86	3.37	0.919	3.10	2.83	1.10	0.118	0.114 ± 0.005
	B3-f, 3	4.08×10^7	5.85×10^7	2.87	3.36	0.913	3.07	2.87	1.07	0.116	
0.08	B3-g, 1	2.41×10^7	5.64×10^7	2.84	3.40	0.942	3.20	3.05	1.05	0.041	
	B3-g, 2	2.43×10^7	5.73×10^7	2.86	3.38	0.925	3.12	3.03	1.03	0.041	0.042 ± 0.001
	B3-g, 3	2.48×10^7	5.85×10^7	2.88	3.35	0.900	3.01	2.99	1.01	0.043	

B5. Experimental determination of K_d and k_{off} of protein–aptamer complex

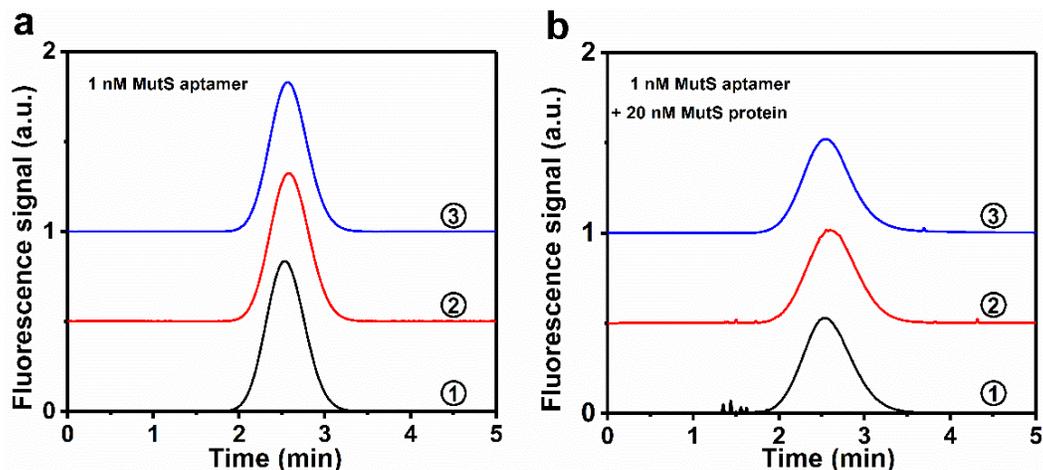


Figure B4. Determination of φ for MutS–aptamer mixture. Areas of peaks PL and L, at full complex saturation and at absence of target protein yield a ratio of ≈ 0.80 . Each set of experiments was performed in triplicates.

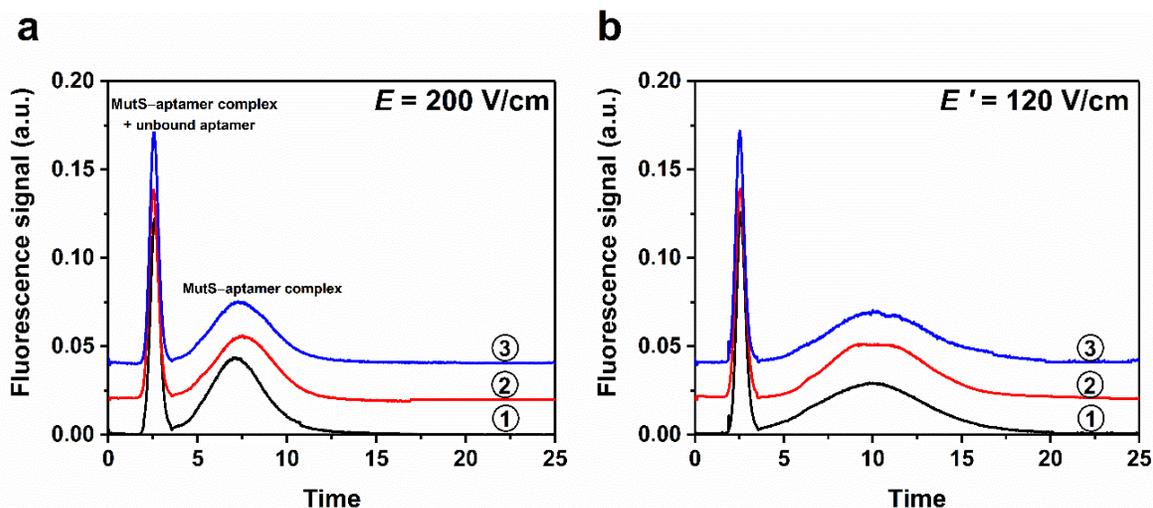


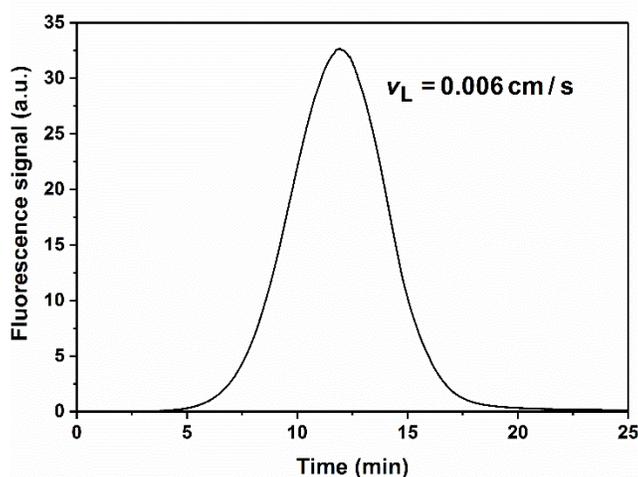
Figure B5. Double-pass experiment with mixture of MutS protein and its aptamer under different electric field strengths: (a) 200 V/cm, and (b) 120 V/cm. Each set of experiments was performed in triplicates. The sample mixture contained 0.50 nM MutS and 0.20 nM aptamer.

Table B3. Experimental determination of K_d of MutS–aptamer complex.

Figure	E V/cm	A_1	A_2	t_1 min	v_{av} cm/min	t_2 min	d_1 cm	t_3 min	v_{PL} cm/min	$R\phi = 1$	$R\phi = 0.80$	$K_d\phi = 1$ nM	$K_d\phi = 0.80$ nM
B5-a, 1		4.44×10^5	8.17×10^5	2.89	3.34	0.892	2.98	3.54	0.842	0.536	0.480	0.470	0.366
B5-a, 2	200	4.57×10^5	9.07×10^5	2.82	3.43	0.967	3.32	3.99	0.831	0.519	0.464	0.437	0.340
B5-a, 3		4.79×10^5	8.78×10^5	2.85	3.38	0.929	3.14	3.74	0.840	0.545	0.489	0.489	0.381
B5-b, 1		4.71×10^5	1.21×10^6	2.87	3.37	0.917	3.09	6.31	0.489	0.626	0.573	0.713	0.556
B5-b, 2	120	4.82×10^5	1.28×10^6	2.83	3.41	0.950	3.24	5.93	0.546	0.574	0.519	0.559	0.435
B5-b, 3		4.64×10^5	1.15×10^6	2.81	3.43	0.971	3.33	6.50	0.513	0.630	0.577	0.726	0.566

B6. Determination of v_L

The value of v_L was determined by electrophoretically running a plug of pure L (MutS aptamer) with a cathode at the inlet using the following conditions: $d = 4.4$ cm, $r = 37.5$ μm , $l = 1.09$ cm, pressure equal to 0.30 psi, total capillary length = 50 cm, $E = 200$ V/cm (**Figure B6**). To achieve faster detection time of pure L under standard IFCE condition, the sample plug was propagated by injecting a 5.8 cm long plug of the run buffer with a pressure pulse of 0.3 psi \times 90 s, resulting in a shorter d (4.4 cm). The value of t_4 was found directly from the electropherogram in **Figure B6**: $t_4 = 714$ s. The, the value of $v_L = 0.006$ cm/s was found with **Eq. 24** in the main text.

**Figure B6.** Determination of v_L . Concentration of L (MutS aptamer) was 50 nM.

**APPENDIX C. SUPPORTING INFORMATION FOR QUANTITATIVE
CHARACTERIZATION OF KCE-BASED PARTITIONING**

C1. Dependence of number of partitioning rounds on k_N & k_B

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

$$B_{\text{out}} = B_{\text{in}} k_B^m$$

$$B_{\text{out}} > Q_1 N_{\text{PCR}}$$

$$B_{\text{in}} k_B^m > Q_1 N_{\text{PCR}}$$

$$k_B^m > Q_1 N_{\text{PCR}} / B_{\text{in}}$$

$$m \log(k_B) > \log(Q_1 N_{\text{PCR}} / B_{\text{in}})$$

$$\because k_B < 1 \text{ and } (Q_1 N_{\text{PCR}} / B_{\text{in}}) < 1; \log(k_B) < 0 \text{ and } \log(Q_1 N_{\text{PCR}} / B_{\text{in}}) < 0$$

$$m < \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log(k_B)}$$

$$B_{\text{out}} / N_{\text{out}} > Q_2$$

$$\because B_{\text{out}} = B_{\text{in}} k_B^m \text{ and } N_{\text{out}} = N_{\text{in}} k_N^m$$

$$B_{\text{in}} k_B^m / (N_{\text{in}} k_N^m) > Q_2$$

$$(k_B / k_N)^m > Q_2 N_{\text{in}} / B_{\text{in}}$$

$$m \log(k_B / k_N) > \log(Q_2 N_{\text{in}} / B_{\text{in}})$$

$$\because k_B / k_N > 1 \text{ and } Q_2 N_{\text{in}} / B_{\text{in}} < 1; \log(k_B / k_N) > 0 \text{ and } \log(Q_2 N_{\text{in}} / B_{\text{in}}) < 0$$

$$m > \frac{\log(Q_2 N_{\text{in}} / B_{\text{in}})}{\log(k_B / k_N)}$$

Multi-round selection with PCR amplification (SELEX):

$$B_{\text{out}} = B_{\text{in}} \left(k_{\text{B}} (Z_{\text{B}})^n \right)^m$$

$$B_{\text{out}} > Q_1 N_{\text{PCR}}$$

$$B_{\text{in}} \left(k_{\text{B}} (Z_{\text{B}})^n \right)^m > Q_1 N_{\text{PCR}}$$

$$\left(k_{\text{B}} (Z_{\text{B}})^n \right)^m > Q_1 N_{\text{PCR}} / B_{\text{in}}$$

$$m \log \left(k_{\text{B}} (Z_{\text{B}})^n \right) > \log(Q_1 N_{\text{PCR}} / B_{\text{in}}),$$

$$\because k_{\text{B}} (Z_{\text{B}})^n > 1 \text{ and } (Q_1 N_{\text{PCR}} / B_{\text{in}}) < 1; \log \left(k_{\text{B}} (Z_{\text{B}})^n \right) > 0 \text{ and } \log(Q_1 N_{\text{PCR}} / B_{\text{in}}) < 0$$

$$m > \frac{\log(Q_1 N_{\text{PCR}} / B_{\text{in}})}{\log \left(k_{\text{B}} (Z_{\text{B}})^n \right)}$$

$$(k_{\text{B}} / k_{\text{N}}) (Z_{\text{B}} / Z_{\text{N}})^n > 1$$

$$(Z_{\text{B}} / Z_{\text{N}})^n > k_{\text{N}} / k_{\text{B}}$$

$$\log \left((Z_{\text{B}} / Z_{\text{N}})^n \right) > \log(k_{\text{N}} / k_{\text{B}})$$

$$n \log(Z_{\text{B}} / Z_{\text{N}}) > \log(k_{\text{N}} / k_{\text{B}}),$$

$$\because Z_{\text{B}} / Z_{\text{N}} < 1 \text{ and } k_{\text{N}} / k_{\text{B}} < 1; \log(Z_{\text{B}} / Z_{\text{N}}) < 0 \text{ and } \log(k_{\text{N}} / k_{\text{B}}) < 0$$

$$n < \left\lfloor \frac{\log(k_{\text{N}} / k_{\text{B}})}{\log(Z_{\text{B}} / Z_{\text{N}})} \right\rfloor$$

C2. Empirical mathematical model to predict the electrophoretic mobility of protein–DNA complexes

The mobility of the protein–DNA complex is linked with the molecular weight of the complex (MW_{P-DNA}) based on the **Eq. 42** (Beloborodov, S. S.; Krylova, S. M.; Krylov, S. N. Spherical-Shape Assumption for Protein–Aptamer Complexes Facilitates Prediction of Their Electrophoretic Mobility. *Analytical Chemistry* 2019, *91*, 12680–12687) in the main-text:

$$\mu_{P-DNA} = A + B\mu_{DNA}L_{DNA}^{0.68}MW_{P-DNA}^{-1/3} \quad (42)$$

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

Since μ_{DNA} is dependent on I_{RB} , the new empirical constants A and B were re-established for running buffer with $I_{RB} = 146$ mM. Linear fitting of experimental mobility data for six protein–DNA complexes with **Eq. C1** resulted in $A = 10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $B = 0.2365 \text{ kDa}^{1/3}$ with a correlation coefficient (R^2) of 0.946 (**Figure C1**).

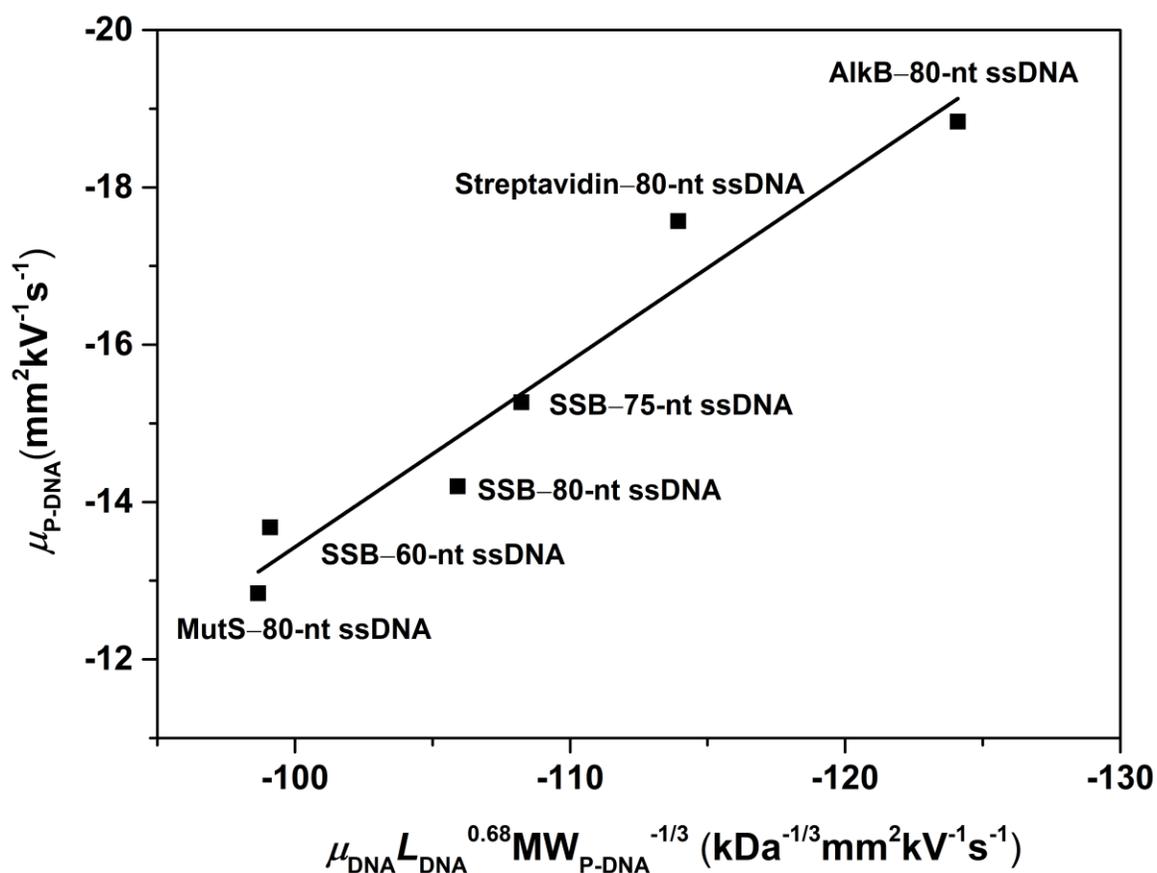


Figure C1. Line of the best fit for the electrophoretic mobility of protein-ssDNA complex as a function of X : $\mu_{\text{P-DNA}} = A + BX$, where $X = \mu_{\text{P-DNA}} L_{\text{DNA}}^{0.68} \text{MW}_{\text{P-DNA}}^{-1/3}$. Calculated values for A and B were $10.225 \text{ mm}^2 \text{ kV}^{-1} \text{ s}^{-1}$ and 0.2365 , respectively. The correlation coefficient was $R^2 = 0.946$.

C3. Determination of elution time of protein–DNA complex, the binder-collection window, and k_N for a given value of MW_P

The time of separation (or elution time) of protein–DNA complex (t) for a given value of MW_P was estimated using the mobilities value obtained from **Eq. 42** (main-text). For both NECEEM sub-modes ($I_{RB} < 50$ mM), the A and B constants are $-9.95 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $0.0929 \text{ kDa}^{1/3}$ respectively. For IFCE ($I_{RB} = 146$ mM), the A and B constants are $10.225 \text{ mm}^2\text{kV}^{-1}\text{s}^{-1}$ and $0.2365 \text{ kDa}^{1/3}$, respectively. The binder-collection time window for a given value of MW_P was calculated as the elution time $\pm 5\%$. An example of determining the binder-collection time window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM is given below (**Figure C2**).

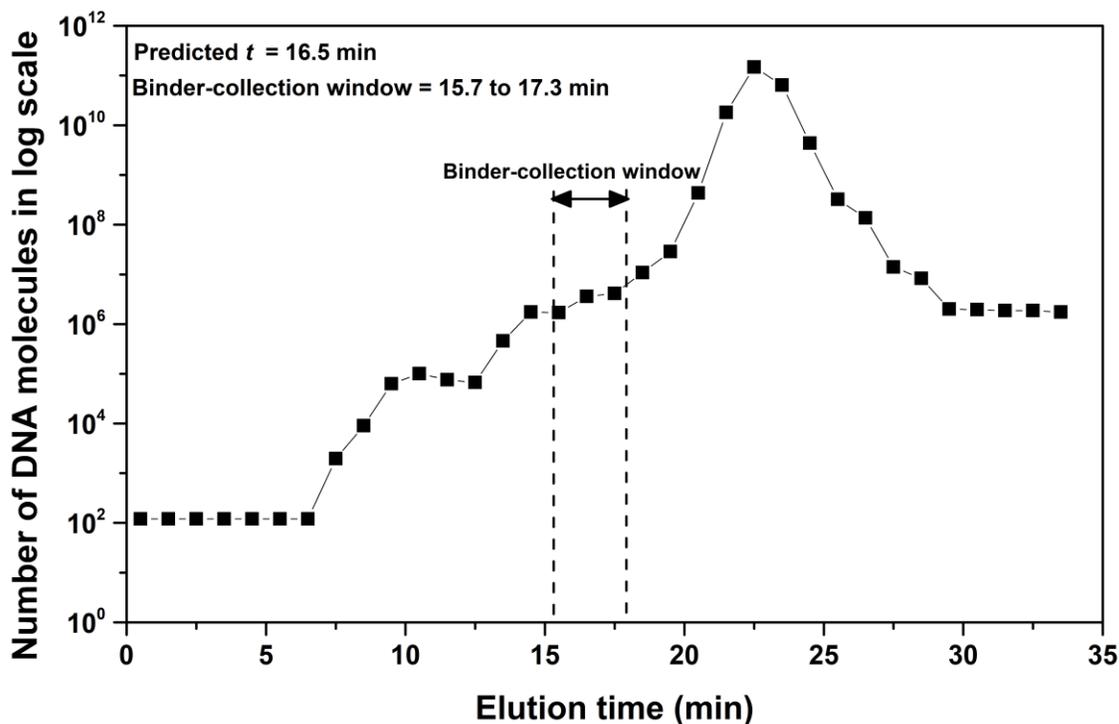


Figure C2. An example of binder-collection time window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM. The black trace indicates the DNA background profile (DNA quantity versus elution time under qPCR detection) in complex-first NECEEM. The predicted elution time (predicted t) for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM was estimated to be 16.5 min. The binder-collection window was defined as $(16.5 \pm 5\%)$ min or 15.7 to 17.3 min. The double-headed arrow indicates the defined binder-collection window for protein–DNA complex with $MW_P = 150$ kDa in complex-first NECEEM. The associated k_N for this protein–DNA complex was calculated as the integral under the DNA-background-profile curve within the binder-collection time window divided by the total quantity of DNA sampled into the capillary.

C4. The predicted elution time of protein–DNA complex with MW_P ranging from 15 to 150 kDa in KCE-based partitioning

The predicted elution times of protein–DNA complexes with MW_P ranging from 15 to 150 kDa are shown in **Figure C3**. For both NECEEM sub-modes, the predicted elution times of protein–DNA complexes are within 25 min over the whole specified range of MW_P . For IFCE, the elution times of the complexes are highly sensitive to the variation in MW_P . When $MW_P < 25$ kDa, the predicted elution time of protein–DNA complexes is beyond 3 h in IFCE.

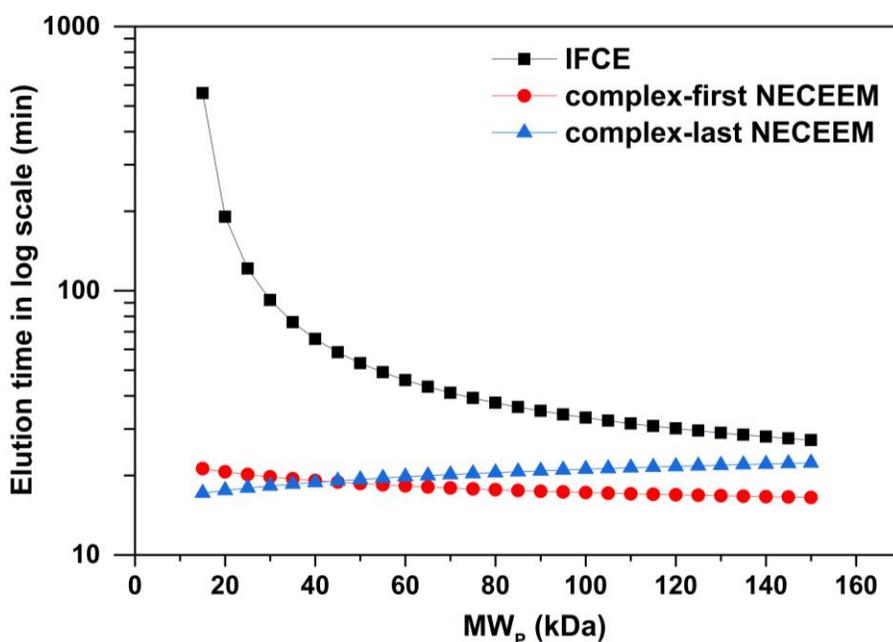


Figure C3. The predicted elution time for the protein–DNA complex as a function of MW_P under conditions of NECEEM and IFCE.

Note C5. The dependence of m_{\min} and m_{\max} on MW_P ranging from 15 to 150 kDa in IFCE for DNA library with moderate to high binder abundance and bulk $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$

The values of m_{\min} and m_{\max} were calculated using Eqs. 33 and 34 in the main-text, respectively, for $Q_1 = 100$ (*i.e.*, B_{out} exceeds PCR noise of 120 molecules of DNA by a factor of 100) and $Q_2 = 1$. As k_N cannot be measured experimentally for target–binder complexes with $MW_P < 25$ kDa under IFCE conditions due to an unreasonably long CE run, the values of k_N for $15 \text{ kDa} < MW_P < 25 \text{ kDa}$ were assumed to be constant (in the range of 10^{-9}) and equal to the experimental k_N value obtained for protein–DNA complex with $MW_P = 25$ kDa.

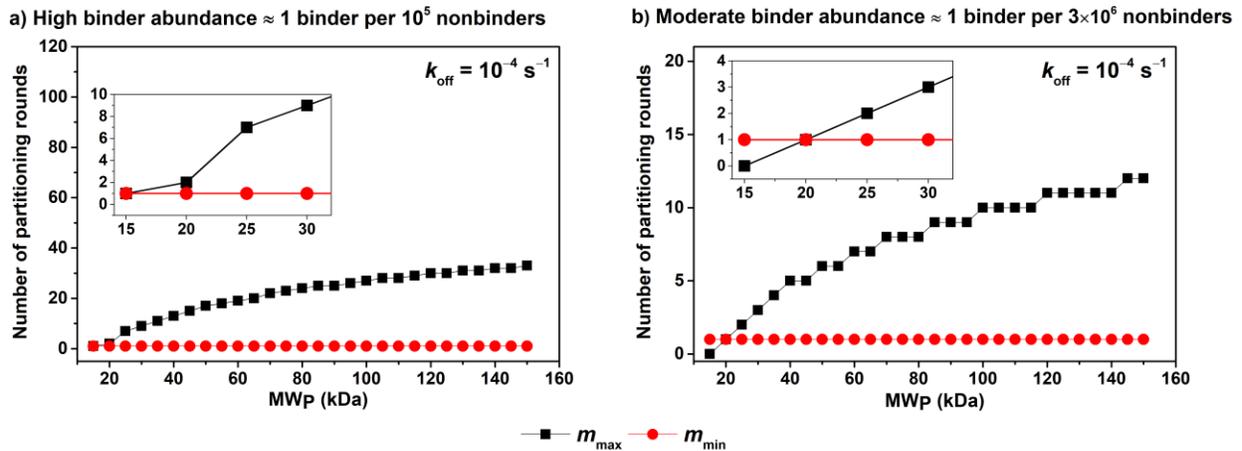


Figure C4. The dependence of m_{\min} and m_{\max} on MW_P in IFCE for DNA library with moderate (a) and high binder abundance (b). The values of k_B were estimated based on a bulk k_{off} of 10^{-4} s^{-1} .

APPENDIX D. SUPPORTING INFORMATION FOR PROTEIN CONCENTRATION STUDY

D1. Determination of the binder-elution window

To determine the binder-elution window, NECEEM-based partitioning was conducted using a mixture of the starting DNA library (100 nM) and a relatively high concentration of the target (1 μ M). Peaks of protein–DNA complexes were detected for both protein targets at such high concentrations, allowing us to identify binder-elution windows to be used in aptamer selection. It is noted that in the case of thrombin, the resolution between target–binder complexes and DNA nonbinders was poorer than in the MutS case, primarily due to the smaller size of thrombin. As such, the partitioning in thrombin selections experienced a much higher nonbinder background (10^{-4} – 10^{-3} for thrombin *versus* 10^{-6} – 10^{-5} for MutS).

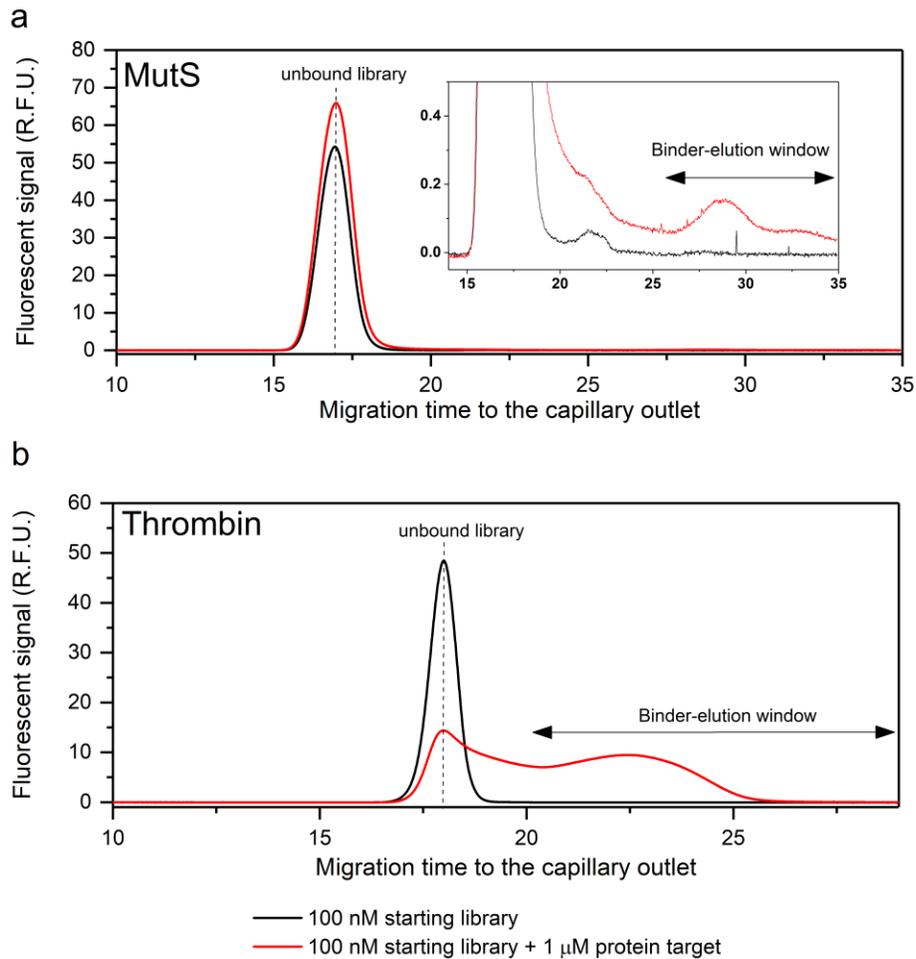


Figure D1. Determination of binder-elution window for NECEEM-based selection for MutS (a) and thrombin (b). Based on the migration profile of DNA nonbinders and target–binder complexes, elution of target–binder complexes was conducted using pressure after NECEEM-based partitioning for 26 min and 20 min for MutS and thrombin selections, respectively. In this complex-last NECEEM mode, the first peak (from the left) corresponds to the unbound library while the second peak corresponds to the target–binder complex.

D2. Summary of nonbinder background (k_N) and relative yield of DNA (q) values obtained in MutS and thrombin selections

D2.1. Estimation of relative standard deviation of q

We conducted 5 sets of positive control (presence of target) and negative control (absence of target) using 10 μ M DNA library and 500 nM thrombin; 5 sets of L_T and N_{out} values were obtained respectively.

The mean value of q (\bar{q}) was calculated from mean values of L_T ($\overline{L_T}$) and N_{out} ($\overline{N_{out}}$) ($n = 5$):

$$\bar{q} = \overline{L_T} / \overline{N_{out}}$$

Subsequently, the standard deviation (σ) of q was estimated by applying the error-propagation rule for noncorrelated standard deviations of L_T and N_{out} :

$$\frac{\sigma_q}{\bar{q}} = \sqrt{\left(\frac{\sigma_{L_T}}{\overline{L_T}}\right)^2 + \left(\frac{\sigma_{N_{out}}}{\overline{N_{out}}}\right)^2}$$

The relative standard deviation (RSD) of q could then be calculated:

$$\text{RSD}_q = \frac{\sigma_q}{\bar{q}}$$

The results of calculations re shown in **Table D1** below.

Table D1. Relative standard deviation of q . Five sets of positive and negative controls were conducted using 10 μ M DNA library and 500 nM thrombin to find L_T and N_{out} , respectively.

Parameters	Quantity (number of DNA molecules)					Mean	σ	\bar{q}	RSD_q
L_T	7.1×10^9	5.9×10^9	7.5×10^9	6.9×10^9	6.9×10^9	6.9×10^9	5.9×10^8	15	0.14 or 14%
N_{out}	4.3×10^8	4.0×10^8	5.4×10^8	4.6×10^8	4.7×10^8	4.6×10^8	5.3×10^7		

D2.2. Summary of k_N and q values obtained in MutS selections

We conducted three-round SELEX for His-tagged MutS with four constant round-to-round target concentrations: 1, 10, 100 and 500 nM. Each round involved a set of positive controls (in the presence of target) and negative controls (in the absence of target) to determine L_T and N_{out} , respectively. After every round, DNA was quantitated with qPCR and k_N and q values were estimated. The results are shown in **Table D2** below.

Table D2. Summary of k_N and q values obtained in MutS selections. The values in the Table were presented as (Round 1, Round 2, Round 3).

Target concentration	N_{in}	N_{out}	L_T	k_N ($=N_{out}/N_{in}$)	q ($=L_T/N_{out}$)
1 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(6.90×10^7 , 2.67×10^6 , 1.29×10^6)	(5.73×10^7 , 2.80×10^6 , 1.42×10^6)	(6.94×10^{-5} , 8.14×10^{-5} , 3.94×10^{-5})	(0.83, 1.05, 1.1)
10 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(5.61×10^7 , 7.04×10^5 , 2.77×10^6)	(3.76×10^8 , 5.91×10^6 , 6.37×10^7)	(5.65×10^{-5} , 2.15×10^{-5} , 8.44×10^{-5})	(6.7, 8.4, 23)
100 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(6.90×10^6 , 8.68×10^5 , 1.43×10^5)	(4.94×10^8 , 2.77×10^8 , 3.55×10^8)	(6.94×10^{-6} , 2.65×10^{-5} , 4.36×10^{-6})	(71.6, 319, 2480)
500 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(5.23×10^6 , 7.69×10^5 , 1.86×10^6)	(3.05×10^9 , 5.08×10^9 , 1.36×10^{10})	(5.26×10^{-6} , 2.35×10^{-5} , 5.67×10^{-5})	(583, 6610, 7300)

D2.3. Summary of k_N and q values obtained in thrombin selections

Similar to MutS selection procedures, we completed three-round SELEX for thrombin with four constant round-to-round target concentrations and estimated k_N and q values for every round after DNA quantitation with qPCR. We repeated two of four selections for thrombin to ensure the reproducibility of the results (10 nM and 500 nM selections). The data are shown in **Table D3** below.

Table D3. Summary of k_N and q values obtained in thrombin selections. The values are presented as (Round 1, Round 2, Round 3).

Target concentration	N_{in}	N_{out}	L_T	k_N ($=N_{out}/N_{in}$)	q ($=L_T/N_{out}$)
1 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(5.17×10^8 , 2.12×10^7 , 2.24×10^7)	(4.48×10^8 , 2.00×10^7 , 2.36×10^7)	(5.20×10^{-4} , 6.45×10^{-4} , 6.83×10^{-4})	(0.87, 0.94, 1.1)
10 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(4.86×10^8 , 9.61×10^6 , 1.36×10^7)	(4.69×10^8 , 1.01×10^7 , 1.27×10^7)	(4.89×10^{-4} , 2.93×10^{-4} , 4.14×10^{-4})	(0.97, 1.1, 0.93)
10 nM (repetition)	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(4.61×10^8 , 1.22×10^7 , 2.11×10^7)	(4.23×10^8 , 1.35×10^7 , 1.89×10^7)	(4.64×10^{-4} , 3.73×10^{-4} , 6.44×10^{-4})	(0.92, 1.1, 0.90)
100 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(5.02×10^8 , 1.69×10^7 , 1.13×10^7)	(1.00×10^9 , 4.21×10^8 , 7.39×10^9)	(5.05×10^{-4} , 5.15×10^{-4} , 3.44×10^{-4})	(2, 25, 66)
500 nM	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(4.25×10^8 , 1.57×10^7 , 2.17×10^7)	(7.11×10^9 , 7.99×10^9 , 1.56×10^{10})	(4.27×10^{-4} , 4.78×10^{-4} , 6.61×10^{-4})	(17, 510, 720)
500 nM (repetition)	(9.94×10^{11} , 3.28×10^{10} , 3.28×10^{10})	(3.96×10^8 , 1.60×10^7 , 1.85×10^7)	(5.90×10^9 , 6.88×10^9 , 1.48×10^{10})	(3.98×10^{-4} , 4.87×10^{-4} , 5.63×10^{-4})	(15, 430, 800)

D3. Data analysis for bulk affinity assays

D3.1. Electropherograms and calculation of R for MutS selection

We used a previously published bulk affinity workflow to evaluate the progress of selection for MutS selection at four different target concentrations (Teclmichael, E.; Le, A. T. H.; Krylova, S. M.; Wang, T. Y.; Krylov, S. N. Bulk Affinity Assays in Aptamer Selection: Challenges, Theory, and Workflow. *Analytical Chemistry* 2022, 94, 15183-15188). The bulk affinity assay was conducted using a constant DNA concentration of 1 nM and a starting target concentration of 1 μ M (**Figure D2**). The target concentration in the bulk affinity assay was subsequently decreased in a stepwise fashion (i.e., 1 μ M \rightarrow 100 nM \rightarrow 10 nM) to ensure that R value (fraction of unbound library) stays within its desired range of 0.3–0.7.

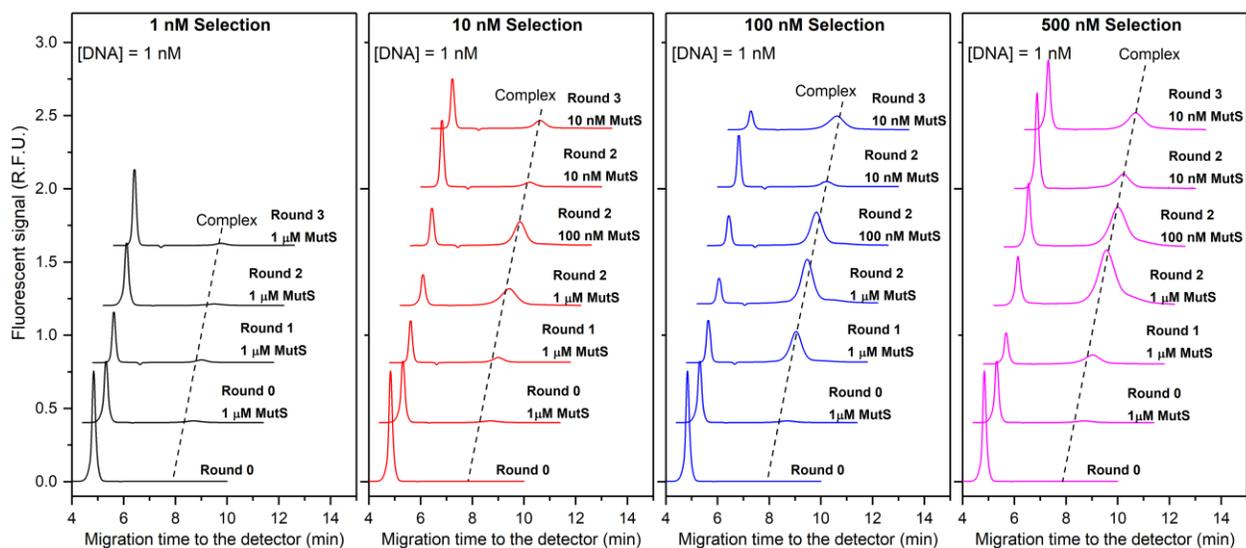


Figure D2. Bulk affinity tests of the starting library and the binder-enriched libraries obtained in MutS selections at four different target concentrations using the published bulk affinity workflow. Black, red, blue, and magenta traces represent selections using 1, 10, 100, and 500 nM MutS, respectively. The experiments were conducted in triplicates and only the representative electropherograms are shown here. The dashed line indicates the position of the target–binder complex in each electropherogram while the leftmost peak corresponds to the unbound DNA library.

The R value in the bulk affinity test was then estimated from the ratio of the peak area of unbound DNA library to the total peak area of unbound library and target–binder complex in the corresponding electropherograms using the NAAP program (Kanoatov, M.; Galievsky, V. A.; Krylova, S. M.; Cherney, L. T.; Jankowski, H. K.; Krylov, S. N. Using Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM) for Simultaneous Determination of Concentration and Equilibrium Constant. *Analytical Chemistry* 2015, 87, 3099-3106). The results are shown in **Table D4** below.

Table D4. Summary of R values for the starting library (Round 0) and binder-enriched libraries (Round 1–3) obtained in MutS selections at 4 different target concentration schemes: 1, 10, 100 and 500 nM MutS. Note that R values for Round 0 were the same for all selections.

Selection	Round	Target concentration used in the affinity test	R values for triplicate runs	Mean $R \pm$ standard deviation
	0	1 μ M	0.802, 0.845, 0.857	0.835 \pm 0.024
1 nM MutS	1	1 μ M	0.822, 0.791, 0.791	0.801 \pm 0.015
	2	1 μ M	0.790, 0.865, 0.721	0.792 \pm 0.059
	3	1 μ M	0.763, 0.759, 0.823	0.782 \pm 0.029
10 nM MutS	1	1 μ M	0.643, 0.700, 0.731	0.691 \pm 0.036
	2	1 μ M	0.324, 0.296, 0.346	0.322 \pm 0.021
	2	100 nM	0.531, 0.537, 0.535	0.534 \pm 0.003
	2	10 nM	0.850, 0.871, 0.833	0.851 \pm 0.016
	3	10 nM	0.634, 0.700, 0.719	0.684 \pm 0.037
100 nM MutS	1	1 μ M	0.411, 0.443, 0.451	0.435 \pm 0.018
	2	1 μ M	0.214, 0.231, 0.151	0.198 \pm 0.035
	2	100 nM	0.376, 0.445, 0.304	0.375 \pm 0.058
	2	10 nM	0.625, 0.709, 0.707	0.680 \pm 0.039
	3	10 nM	0.637, 0.570, 0.542	0.583 \pm 0.400
500 nM MutS	1	1 μ M	0.511, 0.482, 0.544	0.512 \pm 0.025
	2	1 μ M	0.164, 0.159, 0.151	0.158 \pm 0.006
	2	100 nM	0.325, 0.260, 0.274	0.286 \pm 0.028
	2	10 nM	0.646, 0.633, 0.613	0.631 \pm 0.014
	3	10 nM	0.508, 0.569, 0.528	0.535 \pm 0.025

D3.2. Electropherograms and calculation of R for thrombin selections

Similar to our treatment of MutS selection results, we applied the bulk affinity workflow to assess the progress of selection for thrombin across four different target concentrations. The workflow maintained a constant DNA concentration of 20 nM and began with a target concentration of 1 μ M (**Figure D3**). At 1 μ M target concentration, significant binding of the starting library to thrombin was observed as the corresponding R value was below 0.3. Consequently, the target concentration was reduced by 10 folds from 1 μ M to 100 nM. This adjustment was made to elevate the R value of the starting library (Round 0) to a level within the desired range (0.3–0.7); this target concentration (100 nM) remained fixed for later rounds.

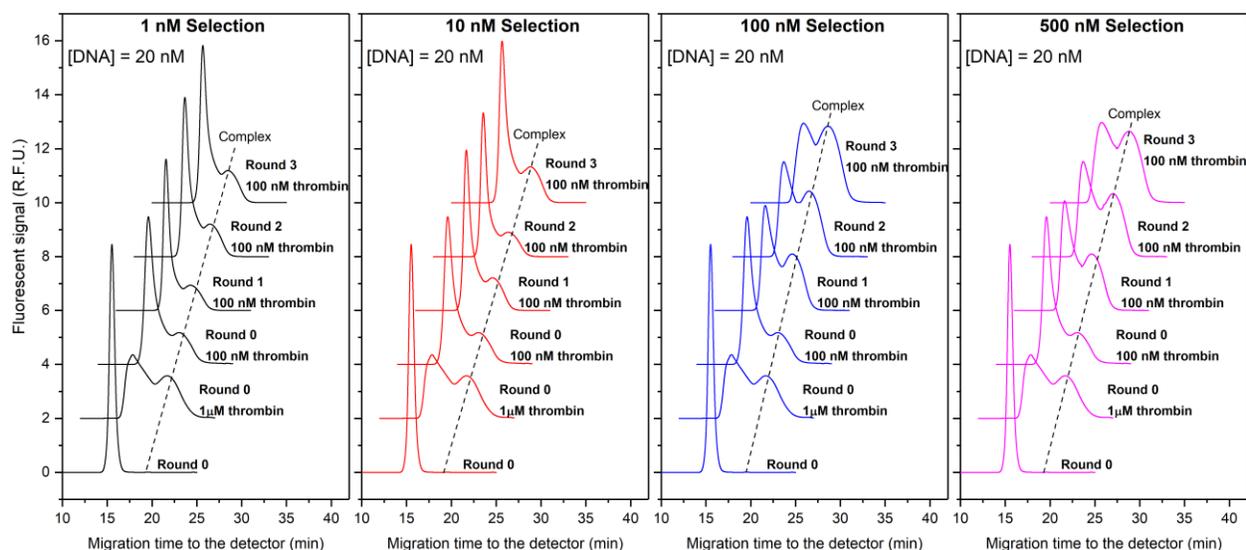


Figure D3. Bulk affinity tests of the starting library and binder-enriched libraries obtained from thrombin selections at four different target concentrations using the published bulk affinity workflow. Black, red, blue, and magenta traces represent selections using 1, 10, 100, and 500 nM thrombin, respectively. The affinity test for every round was conducted in triplicates and only the representative electropherograms are shown here. The dashed lines indicate positions of the target–binder complexes while the leftmost peak corresponds to the unbound DNA library.

In thrombin selections, the resolution between the unbound library and the target–binder complex was relatively poor (**Figure D3**), leading to challenges in calculating *R* based on distinct peak areas of the unbound library and target–binder complex. Therefore, for the thrombin case, we determined *R* value by utilizing the peak height ratio of unbound library in the presence of target to that in its absence. The peak heights and migration times were obtained with 32 Karat Software. The results can be found in **Table D5** below.

Table D5. Summary of *R* values for the starting library (Round 0) and binder-enriched libraries (Round 1–3) obtained in thrombin selections at four different target concentrations: 1, 10, 100 and 500 nM thrombin. Note, *R* values for Round 0 were the same for all selections.

Selection	Round	Target concentration used in the affinity test	<i>R</i> values for triplicate runs	Mean <i>R</i> ± standard deviation
–	0	1 μM	0.261, 0.266, 0.268	0.265 ± 0.004
	0	100 nM	0.640, 0.636, 0.644	0.640 ± 0.004
1 nM thrombin	1	100 nM	0.660, 0.663, 0.682	0.669 ± 0.012
	2	100 nM	0.644, 0.656, 0.665	0.655 ± 0.011
	3	100 nM	0.635, 0.627, 0.630	0.631 ± 0.004
10 nM thrombin	1	100 nM	0.649, 0.654, 0.662	0.655 ± 0.007
	2	100 nM	0.627, 0.635, 0.640	0.634 ± 0.007
	3	100 nM	0.653, 0.660, 0.651	0.655 ± 0.005
10 nM thrombin (repetition)	1	100 nM	0.683, 0.685, 0.682	0.684 ± 0.002
	2	100 nM	0.671, 0.671, 0.674	0.672 ± 0.002
	3	100 nM	0.640, 0.651, 0.653	0.648 ± 0.007
100 nM thrombin	1	100 nM	0.453, 0.429, 0.427	0.436 ± 0.015
	2	100 nM	0.403, 0.411, 0.435	0.416 ± 0.016
	3	100 nM	0.322, 0.320, 0.296	0.313 ± 0.015
500 nM thrombin	1	100 nM	0.474, 0.490, 0.515	0.493 ± 0.021
	2	100 nM	0.403, 0.470, 0.420	0.431 ± 0.035
	3	100 nM	0.341, 0.350, 0.313	0.335 ± 0.020
500 nM thrombin (repetition)	1	100 nM	0.489, 0.502, 0.501	0.497 ± 0.008
	2	100 nM	0.402, 0.427, 0.443	0.424 ± 0.020
	3	100 nM	0.349, 0.337, 0.338	0.341 ± 0.007

APPENDIX E. SUPPORTING INFORMATION FOR KCE-BASED SELECTION FROM DEL

E1. Quality assessment of q-PCR determined data used in mock selection for SA and sEH

For the SA mock selection, we employed two distinct sets of primers, one for binders and another for nonbinders, allowing us to quantitate them separately. To ensure the efficiency and specificity of these qPCR primer sets, we conducted thorough testing by constructing various calibration curves. These curves covered a range of serial dilution standards, from 10^{-17} to 10^{-10} M, consisting of binders, nonbinders, and their combination. This rigorous testing ensured proper amplification efficiency and confirmed the absence of cross-amplification.

Specifically, we used the binder-specific primer set to establish calibration curves for binders and their combinations, as illustrated in **Figure E1a** below. Similarly, the nonbinder-specific primer set was used to construct calibration curves for nonbinders and their combinations, as displayed in **Figure E1b** below. In both figures, the calibration curves for either binders or nonbinders alone and for the combination of both were nearly identical. These results indicate that the presence of either binders or nonbinders did not affect the efficiency of amplification of the other. Furthermore, the amplification of binders and nonbinders using their respective primer sets demonstrated high specificity, with no observed amplification of either binders or nonbinders using the primer set that was specific for the other.

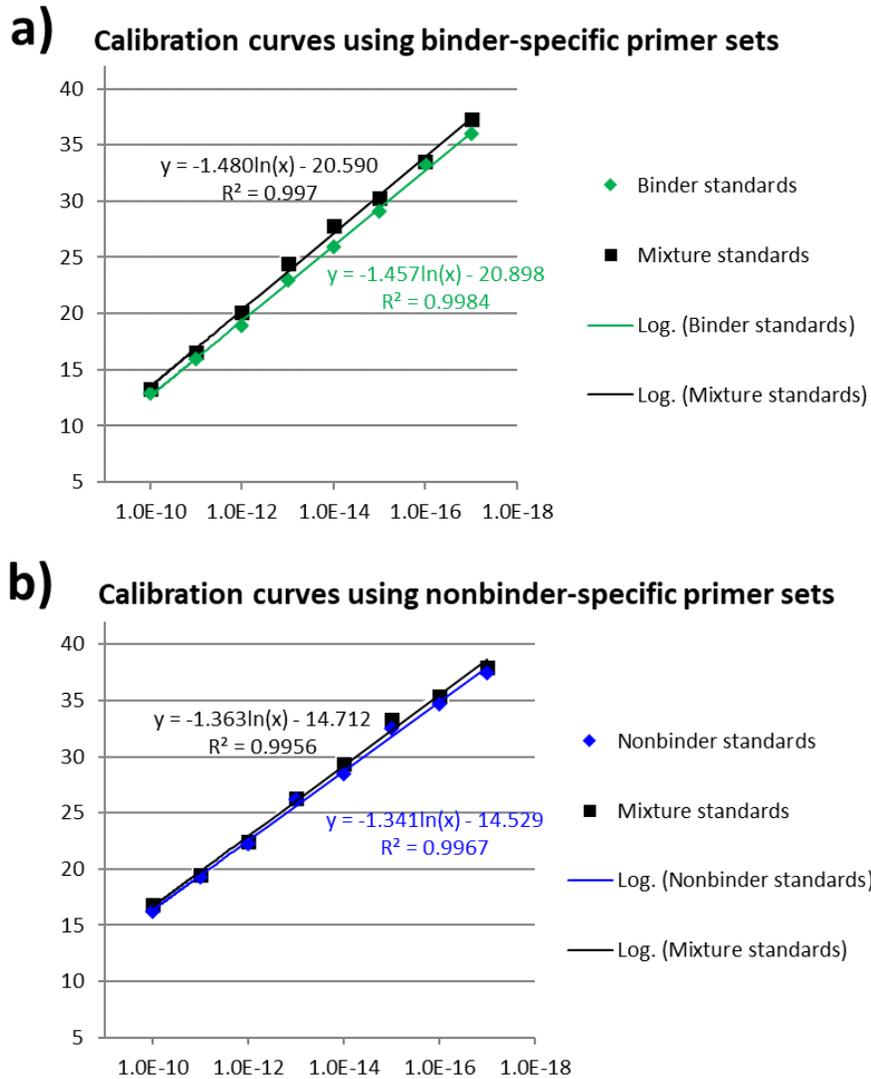


Figure E1. Calibration curves for binders and nonbinders using binder-specific (a) and nonbinder-specific (b) primer sets, respectively. Mixture standards, combining both binder and nonbinder standards, were also incorporated to assess the potential for cross-amplification.

For the sEH mock selection, the DNA tags for both binders and nonbinders were originally designed to be amplified using the same primer set. To estimate the quantities of binders and nonbinders when mixture of binders and nonbinders was used, as in the hypothetical library, we conducted complex-last NECEEM experiments for binders and nonbinders separately, using concentrations similar to those in the mixture. Subsequently, we compared the quantities of binders and nonbinders obtained in these individual experiments with the quantity of DEL molecules obtained from experiment with the mixture (the hypothetical library) to ensure that the numbers matched. The summarized results are presented in **Table E1** below, demonstrating that the quantities of DEL molecules obtained when using the mixture of binders and nonbinders were in agreement with the combined quantities obtained from individual experiments using either binders or nonbinders (less than 5% differences).

Table E1. Quantification of DEL molecules at the input and output of partitioning for sEH mock selection using q-PCR analysis.

	Nonbinder only	Binder only	Sum from individual experiments	Mixture of binder and nonbinder
Input	1.2×10^{10}	1.1×10^7	1.2×10^{10}	1.3×10^{10}
No target control (output)	1.9×10^6	2.5×10^3	1.9×10^6	2.1×10^6
1 μM sEH (output)	7.6×10^6	1.8×10^6	9.4×10^6	9.2×10^6

E2. Initial analysis of binder-enriched pool obtained from bead-based and KCE-based selection of SM ligands for sEH

The output pools from the final round of bead-based and KCE-based selection of SM ligands for sEH were PCR amplified, sequenced, translated into chemical structures and analyzed using TIBCO Spotfire software as described previously (Clark, M. A. et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature Chemical Biology* 2009, 5, 647-654). A group of 23 DEL ligand controls with known literature affinities to sEH were spiked into the initial input library (control molecules/library molecules = 2×10^{-9}) prior to the selections. Subsequently, we monitored and quantified the presence of these controls in the output pool after the selections via HTS. In **Figure E2a** below, the counts per control in the bead-based selection were plotted against the counts for the same control in the KCE-based selection. It's worth noting that the count numbers in the KCE-based selection were lower (given that less than 10% of the input library was screened using KCE compared to bead-based selection). However, a strong correlation emerged in terms of the control abundance between the two datasets. For instance, the most abundant control in the bead-based selection was also the most abundant control in the KCE-based selection.

To further assess the data, we normalized the count numbers of all the controls relative to the count numbers of the most abundant control. We then plotted the normalized counts per control against their literature IC50 values for both the bead-based and KCE-based selections. This analysis revealed a strong correlation between the affinity of controls and their quantities in both selections, suggesting that both methods effectively enriched tighter ligands over weaker ones (**Figure E2b**).

Lastly, when we plotted the counts of all the selected molecules in the bead-based selection against their counts in the KCE-based selection, distinct patterns emerged. Some classes of DEL ligands favored bead-based partitioning, while others exhibited a preference for KCE-based partitioning (**Figure E3**).

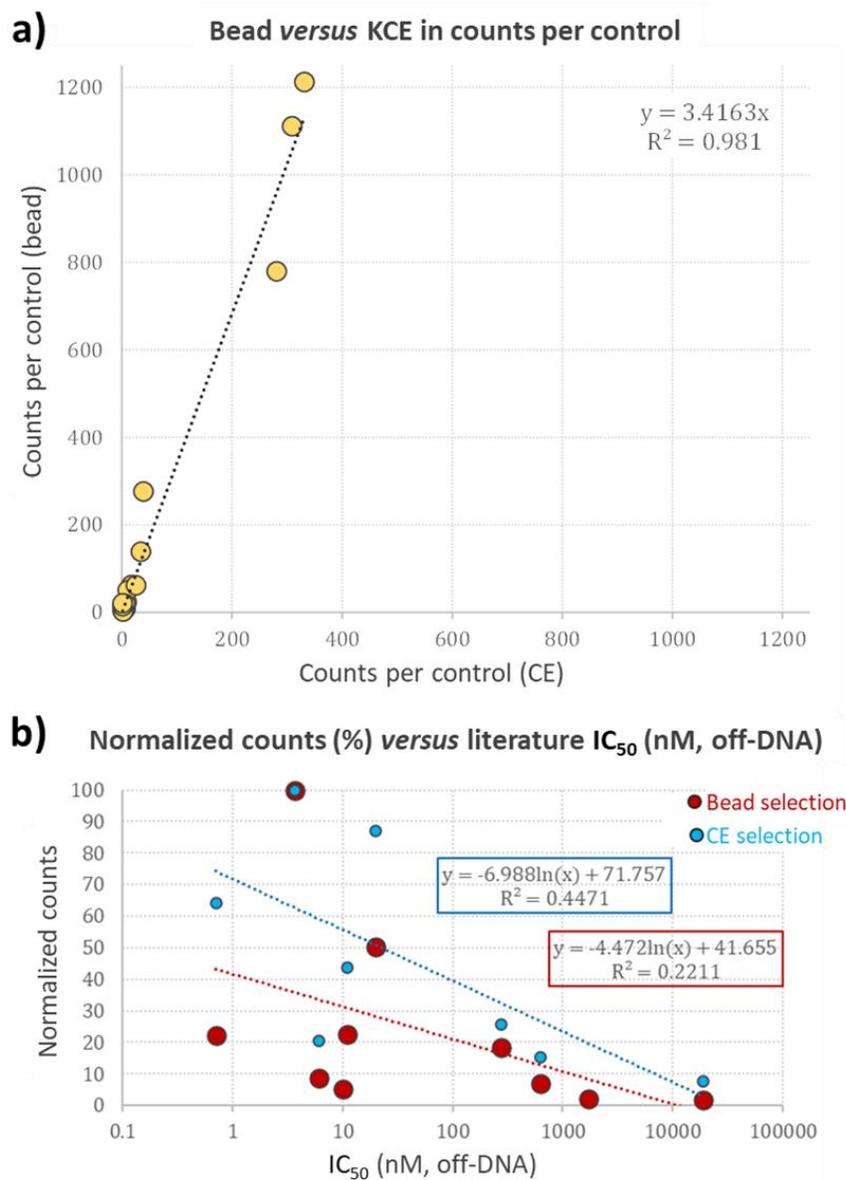


Figure E2. Correlation between a set of controls detected in bead-based and KCE-based selections. In (a), the counts per control from the bead dataset were plotted against the counts for the same control from the CE dataset. In (b), the counts per control (normalized to the most abundant control) were plotted against their respective literature IC₅₀ values for both bead-based and KCE-based selections.

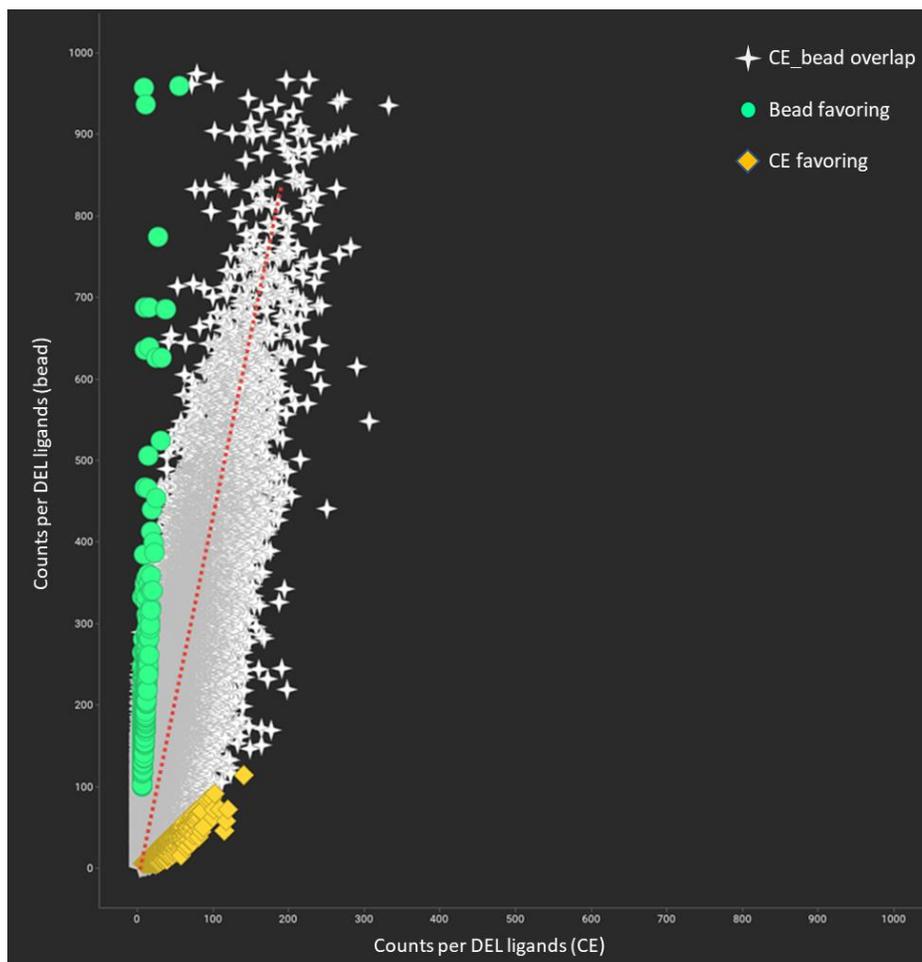


Figure E3. Preference of selected DEL ligands for bead-based or KCE-based partitioning. The green or yellow data are $\geq 4\sigma$ favoring bead-based and KCE-based selection respectively.