Improving the Accuracy, Versatility, and Partitioning Efficiency of Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures

Mirzo Kanoatov

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 BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

November 2016

ABSTRACT

Studying the affinity and the kinetics of biomolecular interactions is an essential task in biology and pharmacology. Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) is a promising technique that allows the measurement of the equilibrium dissociation constant (K_d), and the rate constant of dissociation (k_{off}), and can also be used as the partitioning method in selection of aptamers. There are three critical issues that have hindered the wide adoption of NECEEM by the analytical community: **(1)** poor resilience of the method against systematic errors, **(2)** limited versatility in terms of emulating physiological conditions, and **(3)** restricted partitioning efficiency in DNA aptamer selection. The goals of my research project were **(i)** to develop strategies for minimizing the systematic errors in NECEEM; **(ii)** to make NECEEM compatible with the use of physiological buffers; and **(iii)** to eliminate the restricted partitioning efficiency of NECEEM in aptamer selection. To improve the accuracy of NECEEM, I have developed an approach for simultaneous determination of both the K_d and the concentration of one of the interacting molecules, which eliminates errors caused by the presence of improperly folded molecules in samples. Further, I have developed an "algorithmic" approach for optimization of NECEEM, which takes the interrelation between its experimental parameters into account and minimizes the systematic error in an objective manner. To make NECEEM compatible with physiological buffers, I have uncovered the cause behind the poor detectability of DNA in phosphate buffered saline (PBS), and have created a pressure-assisted modification of NECEEM that overcomes this problem. Lastly, I have determined that the irregular electrophoretic migration of DNA in strong electric fields is caused by the ability of DNA to form unusually stable complexes with its counterions. Based on these findings, I have developed two independent approaches that increase the efficiency of NECEEM in selection of DNA aptamers.

to Mom and Dad,

for always placing the needs and aspirations

of their children first

and to Lise,

for being my source of encouragement and motivation,

without whom I would probably take another 10 years to finish

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my academic supervisor, Professor Sergey N. Krylov, without whose guidance and support this work would be impossible. Dr. Krylov is a brilliant scientist, whose passion and dedication to this art will always be my model for imitation. I was extremely lucky to have such an exceptional teacher, who is genuinely dedicated to the success of his students. Through his tireless efforts, he is able to provide us with opportunities that few other students have access to. Thank you for all of the time, energy and resources that you have dedicated to my development. Most importantly, thank you for giving me your trust, freedom to pursue my interests, and your genuine care. You have shaped me as a scientist and as a person, and I carry your gifts with pride.

I would like to express my sincere gratitude to the members of my supervisory committee, Professors Katalin A. Hudak and Samuel Benchimol, whose wisdom and insight has steered my ever-changing research project into the right direction. Thank you for your caring, patience and the invaluable advice that you have given to me over the years. I would also like to thank the members of my examining committee, Professors Mark Bayfield and Derek J. Wilson, as well as my external examiner, Professor Andrei Drabovich from University of Toronto, for dedicating their invaluable time to the evaluation of my work.

In my time spent at the Krylov Lab, I have had the privilege of rubbing shoulders with some of the brightest and most exceptional people in the history of the Universe. Many have become my mentors: Sahar Javaherian, Maxim Berezovski, Svetlana Krylova, Michael Musheev, Alexander Petrov, Victor Galievsky, and Leonid Cherney, all have dedicated countless hours of their time to share their knowledge and expertise with me. A very special place in my heart will always be reserved for my fellow grad students: David Wegman, Jiayin Bao, Fletcher Agostino,

who have been with me from day one, and soon joined by Natalia Obrecht, Roman Yufa, and Ruchi Liyanage. We grew together, learned from each other, shared in our successes and failures, saw each other getting married and becoming parents, and downed many liters of beer while at it. A special mention goes out to Roseanne Evenhuis, whose loving work has made the lab feel like a second home to us all. I was also extremely fortunate to collaborate with some exceptional undergraduate students: Coraline Retif, Asela Gamage, Sina Mehrabanfar and An Le Thi Hoai. The projects that we have worked on together would not have succeeded without your brilliant ideas, hard work and dedication. All of the members of the Krylov Lab, past and present, thank you for your friendship and your support.

All of my successes in life I owe to my family. My wife, Lise Geffray, is my rock without whom I would collapse. Thank you for the love and happiness that you have given me, it has kept me going through these trying times. Thank you to my parents, Diloro Tursun-Zadeh and Daler Kanoatov, for your unconditional love, your hard work raising us, and for making me believe that I can achieve anything. Special recognition goes out to my grandfather, Masud Tursun-Zadeh, and my uncles, Nodir and Irage Tursun-Zadeh, who have instilled in me the love of science and technology from a young age. All of the members of my large family, thank you for your love, kindness and support. I would not be where I am today without you (quite literally).

I would also like to express my sincere appreciation to Biology Graduate Assistants, Adrienne Dome and Cristalina Del Biondo, whose tireless efforts made the lives of countless graduate students much easier. My endless gratitude goes out to York University, for being my second home for over a decade. I would also like to thank Natural Sciences and Engineering Research Council of Canada and the Government of Ontario for financial support.

TABLE OF CONTENTS

ABSTRACT ii
DEDICATION iii
ACKNOWLEDGEMENTS iv
TABLE OF CONTENTS vi
LIST OF TABLES x
LIST OF FIGURES xi
LIST OF ABBREVIATIONS xiii
LIST OF SYMBOLS xv
LIST OF EQUATIONS xix

Chapter 1. Introduction

1.1. STUDYING BIOMOLECULAR INTERACTIONS 1
1.1.1. Characteristics of biomolecular interactions 1
1.1.2. Analytical performance parameters of methods for studying BMIs 2
1.1.3. Importance of measuring kinetics of BMIs 4
1.2. METHODS FOR STUDYING THE KINETICS OF BMIS 6
1.2.1. Surface plasmon resonance spectroscopy and bio-layer interferometry 6
1.2.2. Stopped-flow techniques 7
1.2.3. Fluorescence correlation spectroscopy 8
1.2.4. Kinetic capillary electrophoresis 9
1.2.4.1. Basic principles of CE 9
1.2.4.2. Concept of KCE 11
1.3. NECEEM – THE PRINCIPLE KCE METHOD 13
1.3.1. Concept of NECEEM 13
1.3.2. History of NECEEM development and application 17
1.3.3. Technological limitations of NECEEM and project goals 21

Chapter 2. Using NECEEM for Simultaneous Determination of Concentration and Equilibrium Constant

2.1. ERRORS DUE TO UNCERTAINTY IN TARGET CONCENTRATION 25
2.2. EXPERIMENTAL SECTION 27
2.2.1. Materials 27
2.2.2. Capillary electrophoresis 27

2.2.3. NECEEM 29
2.3. RESULTS AND DISCUSSION 30
2.3.1. Mathematical solution for simultaneous determination of K_d and $[T]_0$ 30
2.3.2. Considerations for avoiding systematic errors 30
2.3.3. *In silico* modeling and validation. 34
2.3.4. Experimental determination of K_d and $[T]_0$ 40
2.4. CONCLUSIONS 43

Chapter 3. Systematic Approach to Optimization of NECEEM Experimental Conditions

3.1. INTRODUCTION TO NECEEM OPTIMIZATION 44
3.2. EXPERIMENTAL SECTION 46
3.2.1. Materials 46
3.2.2. NECEEM 46
3.3. RESULTS AND DISCUSSION 48
3.3.1. Defining the Parameters and the Variables 48
3.3.2. Phase 1: Experiment validation 50
3.3.3. Phase 2: Recommendations for adjustment of experimental conditions 56
3.3.4. Modifications for preparative applications 62
3.3.5. Recommended conditions for an initial experiment 63
3.3.6. Software implementation 64
3.3.7. Practical application 66
3.4. CONCLUSIONS 68

Chapter 4. NECEEM Analysis of DNA in Physiological Conditions

4.1. IMPORTANCE OF MEASUREMENTS IN PHYSIOLOGICAL CONDITIONS 69
4.2. EXPERIMENTAL SECTION 71
4.2.1. Materials 71
4.2.2. Capillary electrophoresis 71
4.2.3. NECEEM 72
4.2.4. Mobility measurements 74
4.2.5. Peak broadening measurements 75
4.2.6. Characterization of pressure-assisted NECEEM 75

4.3. RESULTS AND DISCUSSION 75
4.3.1. Issue with DNA detectability in PBS-NECEEM 75
4.3.2. Fluorescence intensity 79
4.3.3. Analyte velocity 81
4.3.4. Peak broadening 86
4.3.5. Pressure-facilitated NECEEM 90
4.4. CONCLUSIONS 94

Chapter 5. Non-Uniform Velocity of Homogeneous DNA in a Uniform Electric Field

5.1. REDUCED PARTITIONING EFFICIENCY OF NECEEM FOR DNA MOLECULES 96
5.2. EXPERIMENTAL SECTION 98
5.2.1. Materials 98
5.2.2. Capillary electrophoresis 100
5.2.3. Quantitative polymerase chain reaction 101
5.2.4. DNA dialysis 101
5.2.5. Solubility of DNA aggregates. 102
5.2.6. DNA integrity 103
5.3. RESULTS AND DISCUSSION 104
5.3.1. Non-uniform velocity of DNA 104
5.3.2. Possible causes for non-uniform velocity of DNA 106
5.3.3. Interactions of DNA with its counterions 108
5.3.4. DNA precipitation by electro-dialysis 115
5.3.5. Improving partitioning efficiency of NECEEM without desalting 120
5.4. CONCLUSIONS 123

Chapter 6. Concluding Statements

6.1. OUTCOMES AND IMMEDIATE APPLICATIONS 125
6.2. LIMITATIONS AND FUTURE DIRECTIONS 127

LIST OF PUBLICATIONS 131
REFERENCES 132

APPENDIX A. MATHEMATICAL DERIVATIONS AND PROOFS 143
A.1. Relationship between $[L]_0$ and error in measuring of K_d 143
A.2. Proof for Optimum Relation between K_d and $[T]_0$ 143

APPENDIX B. EXTRACTING KINETICS FROM AFFINITY CE DATA 146
B.1. Summary 146
B.2. Introduction 146
B.3. Results and discussion 149
B.4. Concluding remarks 169
B.5. Appendix references 169

LIST OF TABLES

Table 1.1. Comparison of methods for studying BMI kinetics..... 5
Table 3.1. Required information about the limitations of the CE instrument..... 57
Table 3.2. Recommendations for experimental conditions for a study-initiating experiment..... 64

LIST OF FIGURES

Figure 1.1. Schematic diagram of a CE instrument. 10
Figure 1.2. Methods comprising the KCE toolbox. 12
Figure 1.3. Schematic diagram of NECEEM initial conditions and the resulting data. 15
Figure 2.1. Dependence of relative error in the determination of K_d on initial concentrations. ... 32
Figure 2.2. Validation of the procedure for determination of optimal dilution factor a. 36
Figure 2.3. Validation of the procedure for determination of unknown $[T]_0$. 37
Figure 2.4. Validation of the fitting procedure over a wide range of initial concentrations. 38
Figure 2.5. General trends in the accuracy of determination of K_d and $[T]_0$. 39
Figure 2.6. Determination of K_d and $[T]_0$ from data obtained with a commercial instrument. 40
Figure 2.7. Determination of K_d and $[T]_0$ from data obtained with a custom-made instrument... 42
Figure 3.1. Schematic diagram of the optimization algorithm. 49
Figure 3.2. Minimum value of S required for accurate deconvolution of peak areas. 53
Figure 3.3. Range of values of τ for accurate deconvolution of peak areas. 55
Figure 3.4. Minimum signal-to-noise ratio required for accurate measurement of peak areas. ... 58
Figure 3.5. Application of the algorithm to computer simulated experiments. 65
Figure 3.6. Application of the algorithm within an experimental study. 67
Figure 4.1. NECEEM analysis performed in CB. 77
Figure 4.2. Lack of DNA detectability in PBS-NECEEM. 78
Figure 4.3. Signal repeatability of fluorescently labeled DNA in CB and PBS buffers. 80
Figure 4.4. Mobility of DNA in PBS-NECEEM. 82
Figure 4.5. Effects of pressure propagation step on precision of μ_{eph} measurements in CB. 84
Figure 4.6. EOF stability test. 86
Figure 4.7. Analysis of DNA adsorption onto capillary walls. 88
Figure 4.8. Dynamic peak broadening in PBS. 89
Figure 4.9. Effects of pressure-assisted electrophoresis on DNA peak shape. 91
Figure 4.10. Pressure-assisted NECEEM analysis performed in PBS. 93
Figure 5.1. Heterogeneous migration of homogeneous DNA in a uniform electric field. 105
Figure 5.2. Migration profile of dsDNA in a uniform electric field. 108
Figure 5.3. Effect of Na^+ counterions on velocity non-uniformity of DNA. 111
Figure 5.4. The effect of counterion depletion and reintroduction on DNA migration. 113

Figure 5.5. Removal of DNA condensed counterions by electro-dialysis. 114
Figure 5.6. Ion-dependent solubility of electro-precipitated DNA. 116
Figure 5.7. Agarose gel electrophoresis analysis of DNA integrity after electro-precipitation. 117
Figure 5.8. Improving the partitioning efficiency of NECEEM by modulation of μ_{EOF}. 122
Figure B.1. Comparison of ACE (panel A) and MASKE (panel B) methods. 150
Figure B.2. Effects of concentration of B on deviation between ACE and MASKE results..... 155
Figure B.3. Accuracy of ACE kinetic measurements at various B_0 and A_0. 163
Figure B.4. Fitting of ACE experimental data for interaction between CA and ASL. 166
Figure B.5. ACE results for interaction between boronic acids and fructose. 168

LIST OF ABBREVIATIONS

AptaPIC	aptamer-facilitated protein isolation from cells
BCA	bicinchoninic acid assay
BGE	background electrolyte
BLI	bio-layer interferometry
BMI	biomolecular interaction
BODIPY	boron-dipyrromethene
BSA	bovine serum albumin
CB	conventional CE buffer
CCD	contactless conductivity detection
CE	capillary electrophoresis
CGE	capillary gel electrophoresis
CIEF	capillary isoelectric focusing
CZE	capillary zone electrophoresis
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
ECEEM	equilibrium CE of equilibrium mixtures
EOF	electroosmotic flow
FCS	fluorescence correlation spectroscopy
ITP	capillary isotachophoresis
KCE	kinetic capillary electrophoresis
LIF	laser-induced fluorescence
LOD	limit of detection
LOQ	limit of quantitation
MASKE	macroscopic approach to studying kinetics at equilibrium

MEKC	micellar electrokinetic chromatography
MS	mass spectrometry
NAAP	NECEEM area analysis program
NECEEM	nonequilibrium CE of equilibrium mixtures
PAF	peak asymmetry factor
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDE	partial differential equation
PDGF	platelet-derived growth factor
pH	potential of hydrogen
pI	isoelectric point
ppKCE	plug-plug KCE
qPCR	quantitative PCR
RNA	ribonucleic acid
SELEX	systematic evolution of ligands by exponential enrichment
SPR	surface plasmon resonance
SSB	single-strand DNA-binding protein
ssDNA	single-stranded DNA
SUMET	simplified UMET
SweepCE	sweeping CE
UMET	universal method for determining electrolyte temperatures
UV	ultraviolet

LIST OF SYMBOLS

a	optimal dilution factor of $[T]_{stock}$
b, c, g, k, n	empirical constants for determination of in-capillary temperature
A_C	area of C peak on an electropherogram
A_D	area of the decay bridge on an electropherogram
A_L	area of free-L peak on an electropherogram
A_{front}	area of the leading half of a peak
A_{tail}	area of the trailing half of a peak
C	intermolecular complex between L and T
[C]	local (point) concentration of C
$[C]_{eq}$	equilibrium concentration of C
E	strength of applied electric field
EM	equilibrium mixture with labelled L
EM^0	equilibrium mixture with unlabelled L
E_{max}	largest acceptable magnitude of applied electric field strength
$E_{max.I}$	largest magnitude of applied electric field strength, limited by $I_{instr.max}$
$E_{max.Tef}$	largest magnitude of applied electric field strength, limited by T_{ef}
$E_{max.V}$	largest magnitude of applied electric field strength, limited by $V_{instr.max}$
$E_{max.\tau}$	largest magnitude of applied electric field strength, limited by τ
E_{min}	smallest acceptable magnitude of applied electric field strength
$E_{min.dis}$	smallest magnitude of applied electric field strength, limited by t_{dis}
$E_{min.trun}$	smallest magnitude of applied electric field strength, limited by $t_{run.max}$
$E_{min.V}$	smallest magnitude of applied electric field strength, limited by $V_{instr.min}$
$E_{min.\tau}$	smallest magnitude of applied electric field strength, limited by τ
E_{range}	range of the interval between E_{max} and E_{min}

f	minimum desired fraction of C to reach the detector intact
γ_C	signal response factor for C
γ_L	signal response factor for L
h_C	height of C peak
h_L	height of L peak
I	electric current
I_{av}	average electric current during electrophoresis
$I_{instr.max}$	maximum current supplied by the instrument
K_d	equilibrium dissociation constant
K_d^*	apparent (non-validated) value of K_d
ΔK_d	random error in calculating K_d
k_{off}	rate constant of C dissociation
k_{off}^*	apparent (non-validated) value of k_{off}
k_{on}	rate constant of C formation
L	ligand molecule
[L]	local (point) concentration of L
$[L]_0$	initial (pre-equilibrium) concentration of L
$[L]_{0,first}$	smallest initial concentration of L used in a series of experiments
$[L]_{0,last}$	highest initial concentration of L used in a series of experiments
$[L]_{eq}$	equilibrium concentration of L
l_{det}	length between capillary inlet and detector
$l_{det.instr.max}$	maximum capillary length to detector
$l_{det.instr.min}$	minimum capillary length to detector
$l_{det.min}$	shortest acceptable length between capillary inlet and detector
$l_{det.new}$	value of l_{det} used in a given iteration

$l_{det.previos}$	value of l_{det} used in the previous iteration
l_{out}	length between the detector and capillary outlet
l_{prop}	length of initial pressure propagation zone
l_{sep}	capillary length available for separation
$l_{sep.dis}$	capillary length available for separation which results in t_{run} equal to t_{dis}
$l_{sep.max}$	capillary length available for separation which results in highest acceptable value of S
$l_{sep.min}$	shortest acceptable capillary length available for separation
l_{tot}	total capillary length
m_2, m_1, m_0	coefficients determined by fitting the relationship between E and T_{ef}
μ_C	total mobility of C
μ_{EOF}	mobility of the EOF
μ_{eph}	electrophoretic mobility of DNA
μ_L	total mobility of L
μ_{tot}	total mobility of DNA
$\Delta\mu$	difference in absolute mobilities of L and C
R	fraction of non-bound L
ΔR	random error in measuring R
r	inner radius of capillary
ρ	capillary resistivity
S	number of zone separations events that occur during an experiment
σ	average amplitude of noise at baseline
T	target molecule
[T]	local (point) concentration of T
$[T]_0$	initial (pre-equilibrium) concentration of T
$[T]_{eq}$	equilibrium concentration of T

$[T]_{stock}$	stock concentration of T
T_{amb}	ambient room temperature
T_{boil}	boiling temperature of the BGE
t_C	time of migration of the C peak to the detector
T_{cool}	coolant temperature used during the experiment
$T_{cool.min}$	minimum coolant temperature
t_{dis}	time that it takes for the intact C peak height to reduce to LOQ
T_{ef}	temperature in cooled portion of capillary
$T_{ef.goal}$	desired temperature in cooled portion of the capillary
t_{eq}	characteristic time of equilibration
T_{inef}	temperature in uncooled portion of capillary
t_L	elution time of L
t_{run}	total analysis time
$t_{run.max}$	longest possible analysis time
t_{sep}	characteristic time of separation
τ	ratio of characteristic times t_{sep} and t_{eq}
$V_{instr.max}$	maximum electrophoresis voltage supplied by the instrument
$V_{instr.min}$	minimum electrophoresis voltage supplied by the instrument
v_C	velocity of C
v_{EOF}	velocity of EOF
v_{ep}	electrophoretic velocity of DNA
v_L	velocity of L
v_T	velocity of T
v_{tot}	total velocity of DNA
w	width of sample injection zone

LIST OF EQUATIONS

1. $\mathrm{T} + \mathrm{L} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \mathrm{C}$ 1

2. $K_{\mathrm{d}} = \dfrac{k_{\mathrm{off}}}{k_{\mathrm{on}}} = \dfrac{[\mathrm{T}]_{\mathrm{eq}}[\mathrm{L}]_{\mathrm{eq}}}{[\mathrm{C}]_{\mathrm{eq}}}$ 2

3. $\dfrac{d[\mathrm{C}]}{dt} = -k_{\mathrm{off}}[\mathrm{C}] + k_{\mathrm{on}}[\mathrm{T}][\mathrm{L}]$ 2

4. $l_{\mathrm{sep}} = l_{\mathrm{tot}} - l_{\mathrm{out}} - l_{\mathrm{prop}} = l_{\mathrm{det}} - l_{\mathrm{prop}}$ 14

5. $\mu_{\mathrm{L}} = \dfrac{v_{\mathrm{L}}}{E}$ 14

6. $\mu_{\mathrm{C}} = \dfrac{v_{\mathrm{C}}}{E}$ 14

7. $\Delta\mu = \left|\mu_{\mathrm{L}} - \mu_{\mathrm{C}}\right|$ 14

8. $\gamma_{\mathrm{L}} = \dfrac{h_{\mathrm{L}}}{[\mathrm{L}]_{\mathrm{eq}}}$ 16

9. $\gamma_{\mathrm{C}} = \dfrac{h_{\mathrm{C}}}{[\mathrm{C}]_{\mathrm{intact}}}$ 16

10. $[\mathrm{C}]_{\mathrm{intact}} = [\mathrm{C}]_{\mathrm{eq}} e^{-k_{\mathrm{off}} t_{\mathrm{C}}}$ 16

11. $R = \dfrac{A_{\mathrm{L}}}{A_{\mathrm{L}} + (\gamma_{\mathrm{L}} / \gamma_{\mathrm{C}}) A_{\mathrm{C}} + A_{\mathrm{D}}}$ 16

12. $K_d = \dfrac{[\mathrm{T}]_0 - [\mathrm{L}]_0 (1 - R)}{(1/R) - 1}$ 16

13. $k_{\mathrm{off}} = \dfrac{\ln\left(\dfrac{A_{\mathrm{C}} + A_{\mathrm{D}}}{A_{\mathrm{C}}}\right)}{t_{\mathrm{C}}}$ 16

14. $[\mathrm{T}]_0 = K_{\mathrm{d}}((1/R) - 1) + [\mathrm{L}]_0 (1 - R)$ 17

15. $R^2 + \dfrac{K_{\mathrm{d}} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{[\mathrm{L}]_0} R - \dfrac{K_{\mathrm{d}}}{[\mathrm{L}]_0} = 0$ 30

16. $R = -\dfrac{K_{\mathrm{d}} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0} + \sqrt{\left(\dfrac{K_{\mathrm{d}} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0}\right)^2 + \dfrac{K_{\mathrm{d}}}{[\mathrm{L}]_0}}$ 30

17. $\Delta K_d \approx \left(-[\mathrm{L}]_0 + \frac{[\mathrm{T}]_0}{(1-R)^2} \Delta R \right)$ 31

18. $[\mathrm{L}]_0 \le 2K_\mathrm{d}^*$ 51

19. $(K_\mathrm{d}^* + 0.5[\mathrm{L}]_0) \le [\mathrm{T}]_0 \le 2K_\mathrm{d}^*$ 51

20. $t_{\mathrm{sep}} = \frac{w}{\Delta \mu E}$ 51

21. $t_{\mathrm{run}} = \frac{l_{\mathrm{sep}}}{\min(\mu_\mathrm{L}, \mu_\mathrm{C}) E}$ 51

22. $S = \frac{t_{\mathrm{run}}}{t_{\mathrm{sep}}} = \frac{\Delta \mu l_{\mathrm{sep}}}{\min(\mu_\mathrm{L}, \mu_\mathrm{C}) w}$ 51

23. $t_{\mathrm{dis}} = \frac{\ln\left(\frac{[\mathrm{C}]_{\mathrm{eq}} \gamma_\mathrm{C}}{10\sigma} \right)}{k_{\mathrm{off}}}$ 52

24. $t_{\mathrm{eq}} = \frac{1}{(k_{\mathrm{on}}[\mathrm{T}] + k_{\mathrm{off}})}$ 53

25. $\tau = \frac{t_{\mathrm{sep}}}{t_{\mathrm{eq}}} = \frac{w(k_{\mathrm{on}}[\mathrm{T}]_{\mathrm{eq}} + k_{\mathrm{off}})}{E \Delta \mu}$ 53

26. $S \ge 5$ 54

27. $t_{\mathrm{run}} \le t_{\mathrm{dis}}$ 54

28. $0.15 \le \tau \le 0.6$ 54

29. $\tau \le 0.6$ 54

30. $T_{\mathrm{ef}} = T_{\mathrm{cool}} + \frac{c(EI_{\mathrm{av}})^{n+1}}{g + (EI_{\mathrm{av}})^n}$ 55

31. $T_{\mathrm{inef}} = T_{\mathrm{amb}} + kb(EI_{\mathrm{av}})^2$ 56

32. $T_{\mathrm{ef}} = T_{\mathrm{ef.goal}}$ 56

33. $T_{\mathrm{inef}} < T_{boil}$ 56

34. $[\mathrm{L}]_0 = \begin{cases} 0.2K_\mathrm{d}^* & \text{if } \frac{100\sigma}{\gamma_\mathrm{L}} \le 0.2K_\mathrm{d}^* \\ \frac{100\sigma}{\gamma_\mathrm{L}} & \text{if } 0.2K_\mathrm{d}^* \le \frac{100\sigma}{\gamma_\mathrm{L}} \le 2K_\mathrm{d}^* \end{cases}$ 59

35. $[\mathrm{T}]_0 = K_\mathrm{d}^* + 0.5[\mathrm{L}]_0$ 59

36. $l_{\mathrm{sep.min}} = \dfrac{5\min(\mu_\mathrm{L}, \mu_\mathrm{C})w}{\Delta\mu}$ 59

37. $l_{\mathrm{det.min}} = l_{\mathrm{sep.min}} + l_{\mathrm{prop}}$ 59

38. $l_{\mathrm{det}} = \max\left(l_{\mathrm{det.min}}, l_{\mathrm{det.instr.min}}\right)$ 59

39. $E_{\mathrm{max.V}} = \dfrac{V_{\mathrm{instr.max}}}{l_{\mathrm{det}} + l_{\mathrm{out}}}$ 60

40. $E_{\mathrm{max.I}} = \dfrac{\rho I_{\mathrm{instr.max}}}{\pi r^2}$ 60

41. $E_{\mathrm{max.Tef}} = \dfrac{-m_1 + \sqrt{m_1^{\,2} - 4m_2(m_0 - T_{\mathrm{ef.goal}} + T_{\mathrm{cool.min}})}}{2m_2}$ 60

42. $E_{\mathrm{max.}\tau} = \dfrac{w(k_{\mathrm{on}}[\mathrm{T}]_{\mathrm{eq}} + k_{\mathrm{off}})}{0.15\Delta\mu}$ 60

43. $\rho = \dfrac{E\pi r^2}{I_{\mathrm{av}}}$ 60

44. $E_{\mathrm{min.V}} = \dfrac{V_{\mathrm{instr.min}}}{l_{\mathrm{det}} + l_{\mathrm{out}}}$ 60

45. $E_{\mathrm{min.trun}} = \dfrac{l_{\mathrm{sep}}}{\min(\mu_\mathrm{C}, \mu_\mathrm{L})t_{\mathrm{run.max}}}$ 60

46. $E_{\mathrm{min.dis}} = \dfrac{l_{\mathrm{sep}}}{\mu_\mathrm{C} t_{\mathrm{dis}}}$ 60

47. $E_{\mathrm{min.}\tau} = \dfrac{w(k_{\mathrm{on}}[\mathrm{T}]_{\mathrm{eq}} + k_{\mathrm{off}})}{0.6\Delta\mu}$ 60

48. $E_{\mathrm{max}} = \min(E_{\mathrm{max.V}}, E_{\mathrm{max.I}}, E_{\mathrm{max.Tef}}, E_{\mathrm{max.}\tau})$ 60

49. $E_{\mathrm{min}} = \max(E_{\mathrm{min.V}}, E_{\mathrm{min.trun}}, E_{\mathrm{min.dis}}, E_{\mathrm{min.}\tau})$ 61

50. $E_{\mathrm{range}} = E_{\mathrm{max}} - E_{\mathrm{min}}$ 61

51. $l_{\mathrm{sep.dis}} = t_{\mathrm{dis}} \min(\mu_\mathrm{L}, \mu_\mathrm{C})E_{\mathrm{max}}$ 61

52. $l_{\mathrm{prop}} = l_{\mathrm{det.instr.min}} - l_{\mathrm{sep.dis}}$ 61

53. $l_{\mathrm{sep.max}} = \min(l_{\mathrm{det.instr.max}}, l_{\mathrm{sep.dis}})$ 62

54. $l_{\text{det.new}} = l_{\text{det.previous}} + 0.5(l_{\text{det.previous}} - (l_{\text{sep.dis}} + l_{\text{prop}}))$ 62

55. $t_{\text{dis}} = \dfrac{-\ln(f)}{k_{\text{off}}}$ 62

56. $\text{PAF} = \dfrac{A_{\text{front}} - A_{\text{tail}}}{A_{\text{front}} + A_{\text{tail}}}$ 75

57. $\Delta K_{\text{d}} \approx \dfrac{dK_{\text{d}}}{dR}\Delta R$ 143

58. $\dfrac{dK_{\text{d}}}{dR} = \dfrac{[\text{L}]_0}{\dfrac{1}{R}-1} + \dfrac{[\text{T}]_0 - [\text{L}]_0(1-R)}{\left(\dfrac{1}{R}-1\right)^2}\dfrac{1}{R^2}$ 143

59. $\dfrac{dK_{\text{d}}}{dR} = \dfrac{[\text{L}]_0 R}{1-R} + \dfrac{[\text{T}]_0 - [\text{L}]_0(1-R)}{(1-R)^2} = \dfrac{[\text{L}]_0 R}{1-R} + \dfrac{[\text{T}]_0}{(1-R)^2} - \dfrac{[\text{L}]_0}{1-R} =$
$= \dfrac{[\text{L}]_0 R - [\text{L}]_0}{1-R} + \dfrac{[\text{T}]_0}{(1-R)^2} = -[\text{L}]_0 + \dfrac{[\text{T}]_0}{(1-R)^2}$ 143

$R_0 = R_{[\text{L}]_0 \to 0} = -\dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{2[\text{L}]_0} + \dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{2[\text{L}]_0}\sqrt{1 + \dfrac{\dfrac{K_{\text{d}}}{[\text{L}]_0}}{\left(\dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{2[\text{L}]_0}\right)^2}} \approx$

60. $\approx -\dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{2[\text{L}]_0} + \dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{2[\text{L}]_0}\left(1 + \dfrac{1}{2}\dfrac{\dfrac{K_{\text{d}}}{[\text{L}]_0}}{\left(\dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{2[\text{L}]_0}\right)^2}\right) =$ 144

$= \dfrac{\dfrac{K_{\text{d}}}{[\text{L}]_0}}{\dfrac{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0}{[\text{L}]_0}} = \dfrac{K_{\text{d}}}{K_{\text{d}} + [\text{T}]_0 - [\text{L}]_0} \approx \dfrac{K_{\text{d}}}{K_{\text{d}} + [\text{T}]_0}$

61. $\dfrac{\partial R_0}{\partial K_{\text{d}}} = \dfrac{1}{K_{\text{d}} + [T]_0} - \dfrac{K_{\text{d}}}{(K_{\text{d}} + [T]_0)^2} = \dfrac{[T]_0}{(K_{\text{d}} + [T]_0)^2}$ 144

62. $\dfrac{\partial}{\partial [\text{T}]_0}\left(\dfrac{\partial R_0}{\partial K_{\text{d}}}\right) = \dfrac{1}{(K_{\text{d}} + [\text{T}]_0)^2} - \dfrac{2[\text{T}]_0}{(K_{\text{d}} + [\text{T}]_0)^3}$ 144

63. $\frac{1}{\left(K_{\mathrm{d}}+[T]^{*}\right)^{2}}-\frac{2[T]^{*}}{\left(K_{\mathrm{d}}+[T]^{*}\right)^{3}}=0$ 145

64. $1-\frac{2[\mathrm{T}]^{*}}{K_{\mathrm{d}}+[\mathrm{T}]^{*}}=0$ 145

65. $[\mathrm{T}]^{*}=K_{\mathrm{d}}, \quad \left(R_{0}\right)_{[\mathrm{T}]_{0}=[\mathrm{T}]^{*}}=\frac{1}{2}$ 145

“Do. Or do not. There is no try.”

-Yoda

Chapter 1. Introduction

Some of the material presented in **Section 1.3.1** was published previously and reprinted with permission from "Kanoatov, M.; Mehrabanfar, S.; Krylov, S.N. Systematic approach to optimization of experimental conditions in nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry*, 2016, 88, 9300–9308". Copyright 2016 American Chemical Society. All of the adapted text was written by me personally.

1.1. STUDYING BIOMOLECULAR INTERACTIONS

1.1.1. Characteristics of biomolecular interactions

Non-covalent biomolecular interactions (BMIs) are involved in regulation of every cellular process. Elucidation of mechanisms that govern these interactions is of tremendous importance to our understanding of normal cell function, disease, and drug action. As such, research of BMIs is an imperative task in the fields of molecular biology, medicine, and pharmacology **[1-3].**

Reversible binding between a pair of biomolecules, designated as the target (T) and the ligand (L), with the formation of their intermolecular complex (C), can be described by the following reaction equation:

$$T + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C \qquad (1)$$

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 between the two molecules, and the rate constants of association (k_{on}) and dissociation (k_{off}), which characterize the temporal properties of binding and

unbinding, respectively. The value of K_d, as well as its relationship with k_{on} and k_{off}, is defined as follows:

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[T]_{eq}[L]_{eq}}{[C]_{eq}} \quad (2)$$

where $[T]_{eq}$, $[L]_{eq}$, and $[C]_{eq}$ are the concentrations of the three components of an interacting system after it has reached dynamic equilibrium **[4]**. The constants k_{on} and k_{off} define the rates of association and dissociation of C in the following manner:

$$\frac{d[C]}{dt} = -k_{off}[C] + k_{on}[T][L] \quad (3)$$

where [C], [T], and [L] are the concentrations of the three interacting components at a given time after their mixing, *t*. Analytical techniques that aim at measuring only the value of K_d (equilibrium methods) typically require the quantitation of the relative proportion of $[C]_{eq}$ to one of the non-bound (free) components. Kinetic techniques, which measure k_{on} and k_{off} (and by extension K_d), also rely on measuring the concentrations of the bound and free components, but these measurements must be performed in a time-resolved manner, either before the equilibrium is established (*e.g.,* right after mixing), or after the equilibrium has been disturbed (*e.g.,* after one of the components is separated from the mixture).

1.1.2. Analytical performance parameters of methods for studying BMIs

The molecular interactions that underlie biological processes are very diverse, and so are the values of the constants that characterize them. The K_d values of relevant BMIs are known to range between 10^{-3} and 10^{-15} M **[5, 6]**; k_{on} values between 10^3 and 10^9 $M^{-1}s^{-1}$; and k_{off} values between 10^{-6} and 1 s^{-1} **[7]**. For convenience, we can divide the ranges of possible values of K_d, k_{on}, and k_{off} into arbitrary groups: K_d values can be classified as high (10^{-3} to 10^{-6} M),

intermediate (10^{-6} to 10^{-9} M), and low (below 10^{-9} M); characteristic times of association and dissociation, or the reciprocal values of rates of association and dissociation, respectively, as defined in **Eq. 3**, can be classified as slow (on the scale of hours), intermediate (minutes to seconds), and fast (sub-second). There is not a single analytical technique that can be applied to study the entire diversity of BMIs and the full range of values of their constants; instead, multiple methods are used to perform measurements in different subranges in a complimentary manner. Prior to discussing existing methods for studying BMIs, it is useful to define some of the *analytical performance parameters* that help in their comparison **[8]**. Every method that aims at measuring analyte concentrations has a limited *dynamic range*, within the bounds of which it performs reliably. The lower end of the dynamic range is defined by the *limit of quantitation* (LOQ), the concentration of a given analyte that results in a signal exceeding the noise of the detection system by a factor of 10. Similarly, *limit of detection* (LOD) is the concentration of a given analyte that results in a signal exceeding the noise of the detection system by a factor of 3. Methods that are capable of attaining a low LOQ (*i.e.*, approaching 10^{-15} M) are desirable, as they enable the study of high-affinity interactions characterized by low K_d values (this point will be discussed in more detail in **Section 2.3.2** on **page 30**). *Time resolution*, or the fastest rate at which consecutive measurements are possible, is an important consideration for kinetic measurements, with the fastest of interactions requiring time resolution on the ms-scale. Fast-paced interactions are difficult to characterize due to technological limitations of attaining small time resolution. *Accuracy* and *precision* of an analytical method describe its resilience against systematic and random errors, respectively. *Versatility* of a technique describes its capacity to be adapted for use under a variety of experimental conditions, as well as its applicability to a wide range of analytes and their interactions. *Robustness* describes the insensitivity of a method to

small, but deliberate, variations in the controlled experimental factors, while *ruggedness* describes the insensitivity of a method to changes in uncontrolled factors (*e.g.,* room temperature, ambient light), which translates to inter-laboratory reproducibility of measurements.

1.1.3. Importance of measuring kinetics of BMIs

Currently, a wide variety of equilibrium methods is available, which includes such approaches as membrane- and filter-binding **[9]**, ultracentrifugation, dialysis **[10]**, electrophoresis mobility shift assay **[11]**, isothermal titration calorimetry **[12]**, nuclear magnetic resonance **[13]**, and microscale thermophoresis **[14]** to name a few. This methodological variety accommodates the study of a wide diversity of biomolecules and has made K_d the most commonly used quantitative characteristic of intermolecular interactions, with many conclusions and decisions (*e.g.,* ranking of candidate drug compounds) based largely on its value **[15, 16]**. However, it is becoming evident that knowledge of K_d is not sufficient for a comprehensive characterization of an interacting system, and that the interaction kinetics may play a far more important role **[17]**. This stems from the fact that very few processes in living organisms occur at equilibrium (in fact, equilibrium is more characteristic of dead organisms). All cellular processes are precisely controlled in a time-dependent manner; as a result, interactions characterized by identical K_d values, but with different (but still proportional) k_{on} and k_{off} values will lead to drastically different biological outcomes **[18]**. Thus, information on the kinetics of BMIs enables us to make more biologically-relevant predictions.

Conceptually, any equilibrium technique listed in the beginning of the previous paragraph can be adapted for the study of interaction kinetics; in practice, however, most of these methods lack sufficient time resolution capabilities to facilitate the study of the majority of relevant BMIs **[19]**. Due to this challenge, the available variety of kinetic methods for studying BMIs is much more

limited than that of equilibrium methods, with only surface plasmon resonance (SPR) spectroscopy **[20]**, bio-layer interferometry (BLI) **[21]**, stopped-flow **[22]**, and fluorescence correlation spectroscopy (FCS) finding widespread practical use **[23]**. The versatility of this group of methods is restricted by a number of inherent limitations, summarized in **Table 1.1** and discussed in detail in the next section, which necessitates further development and diversification of kinetic techniques in order to promote a more extensive use of kinetic information in biology. Over the past decade, our laboratory has been developing a new set of methods called kinetic capillary electrophoresis (KCE) **[24]**, with the goal of complementing some of the restrictions of the established methods. The next section summarizes and compares the advantages and limitations of each of these techniques.

Table 1.1. Comparison of methods for studying BMI kinetics

Method	**Measurements in free solution?**	**Label-free analysis?**	**Poor compatibility with:**
SPR and BLI	No, surface immobilization required	Yes	Small molecules (under 1 kDa)
Stopped-Flow techniques	Yes	No, labeling of both components required	Molecules that cannot be labeled
FSC	Yes	No, labeling of at least one of the components required	small target molecules, molecules that cannot be labeled
KCE	Yes	Yes, optional labeling of one of the components is sufficient	Pairs of neutral molecules

1.2. METHODS FOR STUDYING THE KINETICS OF BMIS

1.2.1. Surface plasmon resonance spectroscopy and bio-layer interferometry

From the set of existing kinetic methods, SPR is the most commonly used (with citations in more than 14,000 publications over the past 33 years), while the recently commercialized BLI is quickly gaining in popularity (more than 200 publications over the past 11 years). The two technologies are similar to each other, and, thus, share a comparable set of advantages and limitations **[20, 21]**. Both methods rely on immobilization of T on the surface of a sensor, which upon interrogation by light produces a signal dependent on the size of the immobilized entity. If the binding of L can sufficiently increase the size of the surface-immobilized construct, these detection techniques can be used to monitor complexation in real-time, *i.e.*, only limited by the rates of optical data-acquisition electronics, which can operate on the sub-μs time scale. As such, both SPR and BLI offer outstanding time resolution. In SPR, the binding experiments are carried out by delivering a steady stream of buffer-dissolved L to the sensor-immobilized T by a fluidic system to monitor their association. Upon reaching binding saturation, the L-containing buffer is substituted with the L-devoid buffer, and the dissociation of C is observed. In contrast, the current commercial implementations of BLI employ a well-plate design, in which the sensor is dipped into a succession of L-containing and L-devoid buffers, and orbital shaking is applied to ensure homogeneous distribution of L in the solutions. For both methods, the generated experimental curves of binding and dissociation are fitted to theoretical reaction equations of an appropriate order, and k_{on} and k_{off} are extracted as the best-fit parameters.

Besides the exceptional time resolution, other advantages of SPR and BLI include the freedom from sample-labeling requirements, resilience to crude samples, as well as good robustness and ruggedness. The major disadvantage of both SPR and BLI has to do with the fact

that they are surface-based methods: immobilization of T can influence its ability to interact with L by affecting its conformation or by introducing steric hindrance **[25, 26]**, reducing the accuracy of the techniques. In addition, the mass transfer of analytes to and from the surface of the sensor by diffusion may often be rate-limiting and, thus, prevent the measurements of fast interactions. The methods also have poor sensitivity for compounds with molecular weight under 1 kDa, making them not suitable for analysis of interactions of biomolecules with small organic compounds (which are often used as pharmaceuticals), thus, limiting the versatility of the methods **[27, 28]**.

1.2.2. Stopped-flow techniques

The stopped-flow techniques employ specialized microfluidic systems that can efficiently mix L and T together and quickly (within a few ms of dead time) deliver them to an observation cell. Binding is monitored by one of the compatible optical methods, such as ultraviolet (UV) absorbance, fluorescence, or circular dichroism spectroscopies **[29]**. The major advantage of stopped-flow approaches is that the measurements are performed in free-solution, which helps to avoid issues due to surface-immobilization of analytes, inherent to SPR and BLI. While the observation of binding and unbinding does not occur in real-time, the dead times in modern stopped-flow instrumentation are sufficiently short to allow characterization of even the fastest BMIs. The major limitation of stopped-flow techniques stems from the employed detection schemes, which all necessitate a change in the optical properties of the intermolecular system upon binding **[30]**. In most cases, this requires that the interacting molecules are derivatized to contain optically-detectable labels. This can make the measurements prone to error, as optically active chemical groups are often large, charged or contain hydrophobic structures, and can, thus, significantly influence the properties of the derivatized molecules and their interactions. Further,

the required modifications of analyte molecules for use with high-sensitivity detection techniques, such as Förster resonance energy transfer or fluorophore–quencher pairing, require derivatization of both T and L, which further aggravates the issue.

1.2.3. Fluorescence correlation spectroscopy

FCS relies on high-resolution microscopy to optically isolate and observe a very small volume (in fL-scale) in a solution that contains the interacting molecules. Brownian motion of the molecules in and out of the observation volume is influenced by their size; thus, complexation can be detected if the size of C is significantly larger than those of the free molecules **[31]**. Molecular size and concentration information can be extracted from FCS signal by subjecting it to autocorrelation function analysis, which reveals the characteristic times of recurring events. Due to the fact that FCS data is based on the fluctuations of a small number of molecules (between 1 and 100), rather than the bulk of the sample, it can be used to extract information on interaction kinetics even at conditions of equilibrium (*i.e.*, the system does not need to be disturbed at the beginning of measurements). This opens up unique prospects of performing measurements *in situ, e.g.,* in biofilms or in live cells **[32, 33]**. FCS can also be used to perform direct measurements of interaction stoichiometry, which is especially useful for characterization of multi-component interactions. Real-time nature of the observations and the remarkable LOQs inherent to FCS enable its use in measurements of very fast and high-affinity interactions. The major disadvantage of FCS is similar to that of the stopped-flow techniques: studied analytes must be fluorescently-labelled in order to ensure sufficient LOQ for the single-molecule level of analysis. The T molecule studied by FCS (*i.e,* the unlabelled molecule) must be comparable or larger than L, to ensure that their combined diffusion coefficient increases by at least a factor of

1.6 **[31]**, a requirement that restricts the method in the analysis of interactions between large biological polymers with small molecules.

1.2.4. Kinetic capillary electrophoresis

KCE is defined as capillary electrophoresis (CE) of species which interact during the separation **[24]**. In the following subsections, I will describe the basic principles of CE and the concepts of its application in measurement of kinetics.

1.2.4.1. Basic principles of CE

CE is a separation technique that relies on superior heat-dissipation properties of narrow-bore capillaries (with 10-200-μm inner diameters) to perform electrophoresis in a high-magnitude applied electric field (100-1000 V cm^{-1}) **[34]**. The use of a strong electric field in electrophoresis increases the migration rates of electrically charged species, making it possible to achieve substantial spatial separation between different molecules before the detrimental effects of sample diffusion become considerable. Multiple modes of CE exist, which allow the separation of species on the basis of different physical properties: Capillary Zone Electrophoresis (CZE) separates on the basis of size-to-charge ratio, Capillary Gel Electrophoresis (CGE) on the basis of size, Capillary Isoelectric Focusing (CIEF) on the basis of pI, Capillary Isotachophoresis (ITP) on the basis of ionic mobility, and Micellar Electrokinetic Chromatography (MEKC), on the basis of hydrophobicity. CZE is the most commonly used mode of CE, and the two terms are often used interchangeably (including in this Dissertation). In CZE the separation is accomplished in free solution, without the use of sieving matrixes, surface immobilization or pseudo-stationary phases. In its most basic setup, CZE consists of a capillary suspended between two background electrolyte (BGE)-filled reservoirs, with an electrode suspended in each (**Figure 1.1**). Analytes are introduced into the capillary as zones by pressure or electrokinetic injection,

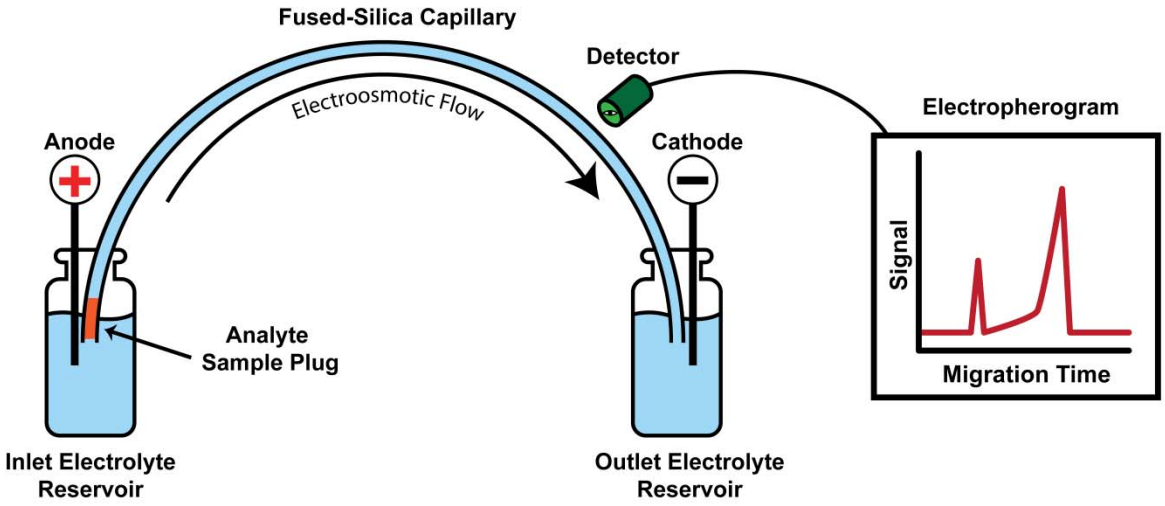

Figure 1.1. Schematic diagram of a CE instrument. A narrow-bore capillary is suspended between two electrolyte-filled reservoirs, with electrodes in each. In the depicted example, a mixture of analytes is injected as a single zone. When a fused-silica capillary is used, a strong EOF is directed from the anode to the cathode. Under the depicted arrangement, positive, negative and neutral molecules have a net velocity directed toward the detector. An electropherogram is generated as the separated species pass through the point of detection.

and electrophoresis is initiated by establishing a uniform electric field across the capillary. When the surface of the capillary bears a charge, such as in the case of fused-silica capillaries, the electrophoretic movement of analytes is supplemented by an electroosmotic flow (EOF), which is driven by the movement of a layer of counterions at the surface–solution interface. A strong EOF is beneficial for CE analysis, as it allows for both negatively and positively charged species to move in the same direction toward the detector. CE is compatible with a large number of on-line detection techniques, including UV absorption spectroscopy, laser-induced fluorescence (LIF) detection, contactless conductivity detection (CCD), and mass spectrometry (MS) **[35, 36]**. Ultrasensitive fluorescence-based detection techniques are available, with LOQs as low as a few hundreds of molecules of fluorescein **[37]**; however, most commercial instruments have an LOQ on the scale of 10^{-10} M for the same standard **[38]**. Upon reaching the detector, the analytes

generate a signal that is presented as a function of migration time to the detector, referred to as an electropherogram plot.

1.2.4.2. Concept of KCE

KCE is a collection of tools that rely on the superb separation efficiency of CZE to elucidate the kinetics of BMIs. Interactions between molecules during electrophoresis will affect their combined migration patterns, essentially convoluting the kinetic and affinity information within specific features of the resulting electropherogram. The initial and boundary conditions of a CE experiment (*i.e.*, the conditions established before commencing electrophoresis) can be arranged in a variety of ways (**Figure 1.2**, leftmost column), such that different features of the electropherograms, defined by either k_{on}, k_{off} or K_d, become more prominent. This allows us to define a set of KCE methods, on the basis of boundary and initial conditions, with each method preferentially suited to study a particular subset of BMI types. The current "toolbox" of KCE methods that are capable of measuring kinetics includes Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM), plug-plug KCE (ppKCE), Sweeping CE (sweepCE), Equilibrium CE of Equilibrium Mixtures (ECEEM), and Macroscopic Approach to Studying Kinetics at Equilibrium (MASKE).

The major advantage of KCE methods is that they allow the study of BMIs in free solution, thus, avoiding the biases that are characteristic of sensor-based methods like SPR and BLI. Furthermore, when coupled with UV absorption-, CCD-, or MS-based detection, KCE can be used to study non-derivatized analytes, thus, avoiding biases characteristic of label-based methods, such as stopped-flow and FCS. When LIF detection is used, and labeling is required, derivatization of only one of the two interacting species is always sufficient for a KCE

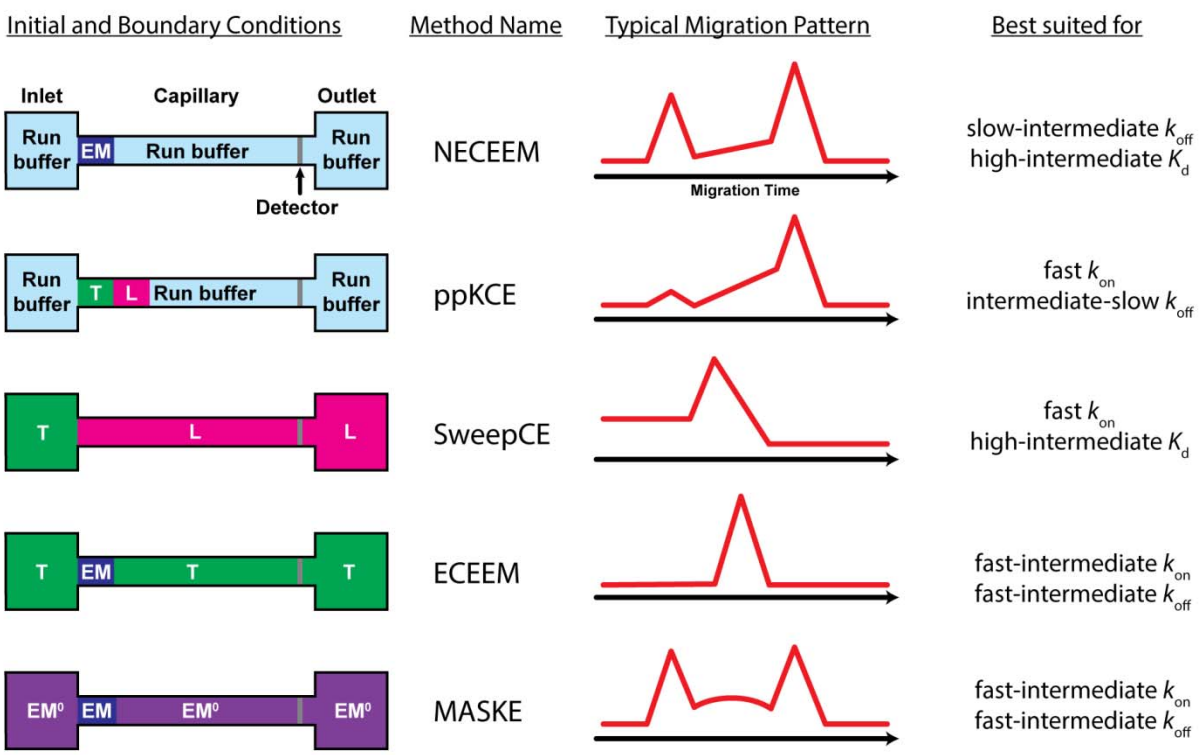

Figure 1.2. Methods comprising the KCE toolbox. The leftmost column depicts the boundary and initial conditions of the system prior to commencing electrophoresis. Large squares represent the electrolyte reservoirs, which define the boundary conditions of the system. The connecting segment represents the capillary, with the initial distributions of analyte zones and the content of the BGE. EM stands for "equilibrium mixture". In the case of MASKE, EM contains L which is labeled for detection, while EM^0 contains non-detectable L. The third column from the left depicts the migration patterns of the detectable species (usually L and C) resulting from an experiment. The rightmost column describes the types of BMIs the method is best suited to study. The figure was adapted with permission from reference **[39]**. Copyright 2015 American Chemical Society.

experiment. As such, out of all of the discussed kinetic methods, KCE methods are least affected by sources of bias and inaccuracy. As kinetic information is embedded into a continuously-recorded electropherogram, KCE time resolution capabilities are on par with the real-time detection methods like SPR and BLI. Combined with this property, the wide range of k_{on} and k_{off}

values that different KCE methods can collectively analyze, as shown in the rightmost column of **Figure 1.2**, makes the versatility of the toolbox difficult to match.

The major inherent limitation of the KCE methods, being an electrophoresis-based technology, is that they can only be applied to BMI systems that can be separated from each other on the basis of size-to-charge ratio. The best results are typically achieved when one of the species is highly charged, such as nucleic acids. In fact, the intermolecular interactions of deoxyribonucleic acid (DNA) are the most common type of BMIs studied by KCE, due to the remarkable charge density of this polymer **[40, 41]**.

1.3. NECEEM – THE PRINCIPLE KCE METHOD

The goal of my research project is the continued development of NECEEM, the most popular of the KCE methods. In the following subsections, I present a detailed conceptual description of NECEEM, summarize the history of its development and application, and discuss the technological limitations of this technique that I have set out to resolve.

1.3.1. Concept of NECEEM

In a NECEEM experiment, T and L are mixed together at initial concentrations $[T]_0$ and $[L]_0$, and incubated to achieve equilibration of the binding reaction **[24, 42]**. NECEEM analysis relies on CE to separate and quantitate the components of the equilibrium mixture. This is done by injecting a short zone of the mixture (with a width w) into a narrow-bore capillary (with an inner radius r) prefilled with the BGE, and subjecting the capillary to a uniform electric field (E). The buffer that is used to prepare the sample mixture is matched with the BGE as closely as possible in order to avoid electrodispersive phenomena. The electric current (I) passing through the capillary causes generation of Joule heat, which often necessitates active cooling to maintain a desired in-capillary temperature. Most modern CE instruments are equipped with liquid or

forced-air cooling systems, which are unable to cool the entire capillary efficiently: often stretches of the capillary at the inlet, outlet, and point of contact with the detector cannot be enclosed within the cooling system **[43]**. Temperatures in the efficiently cooled (T_{ef}) and inefficiently cooled (T_{inef}) portions of the capillary can differ dramatically **[44]**. To avoid the detrimental effects of heating on the studied molecules and CE separation, the sample zone is pressure propagated through the inefficiently cooled portion at the inlet prior to the application of the electric potential. Sample propagation through the inefficiently cooled portion of the capillary reduces the length of the capillary available for separation (l_{sep}), defined as:

$$l_{sep} = l_{tot} - l_{out} - l_{prop} = l_{det} - l_{prop} \quad (4)$$

where l_{tot} is the total length of the capillary; l_{det} is the distance between the capillary inlet and the detector; l_{out} is the distance between the detector and the capillary outlet (for instruments with on-column detection); l_{prop} is the length of the initial pressure propagation zone (**Figure 1.3A**). When the electric field is established, molecules with different size-to-charge ratios migrate through the capillary with different constant velocities (v_T, $v_{L,}$ and v_C) and separate from each other into distinct spatial zones. Mobilities of relevant analytes are defined as follows:

$$\mu_L = \frac{v_L}{E} \quad (5)$$

$$\mu_C = \frac{v_C}{E} \quad (6)$$

$$\Delta\mu = \left|\mu_L - \mu_C\right| \quad (7)$$

It is sufficient that only one of the components (L by convention) is detectable in both its free and bound states. In such a case, it is sufficient to achieve separation between the zones of L and

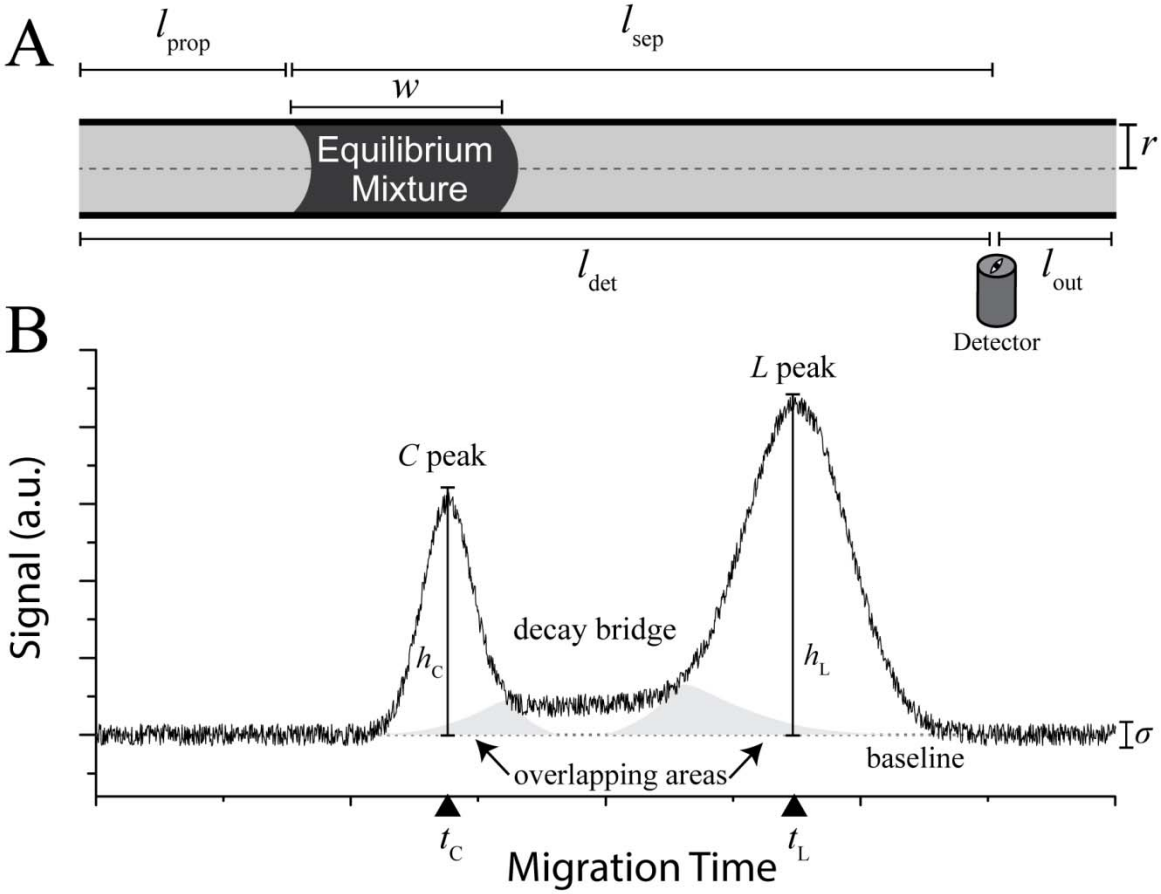

Figure 1.3. Schematic diagram of NECEEM initial conditions and the resulting data. Panel A: Initial conditions in the capillary prior to application of the electric field. A plug of equilibrium mixture is injected into the capillary and propagated through the inefficiently cooled portion. **Panel B:** Example of a NECEEM electropherogram, which consists of 3 features: peaks of free L and intact C, and the decay bridge. Overlapping areas between the peaks and the bridge necessitate the use of a deconvolution procedure for accurate measurement of areas.

C only. During CE separation, the equilibrium fractions of free T and L migrate as distinct zones, while C undergoes continuous dissociation as a result of disturbed equilibrium. NECEEM electropherograms typically contain 3 distinct features: 2 peaks, which correspond to the zones of free L and intact C, and a smear-like region of C dissociation products (decay bridge), which merges with both L and C peaks (**Figure 1.3B**). The amplitude of the generated signal is defined

by the local concentrations of the detected analytes and their response factors (γ_L and γ_C, respectively):

$$\gamma_L = \frac{h_L}{[L]_{eq}} \tag{8}$$

$$\gamma_C = \frac{h_C}{[C]_{intact}} \tag{9}$$

where h_L and h_C are the heights of the free ligand and intact complex peaks, respectively; and $[C]_{intact}$ is the concentration of intact complex at the time of its detection, t_C, calculated as:

$$[C]_{intact} = [C]_{eq} e^{-k_{off} t_C} \tag{10}$$

Once an experimental signal is recorded, the areas under the three features of a NECEEM electropherogram (A_L, A_C, and A_D for peaks of L and C, and the decay bridge, respectively) are integrated, with a deconvolution procedure applied to the overlapping areas between the features (**Figure 1.3B**) **[45]**, and used to calculate the fraction of unbound ligand, R:

$$R = \frac{A_L}{A_L + (\gamma_L / \gamma_C) A_C + A_D} \tag{11}$$

K_d and k_{off} are then calculated as follows:

$$K_d = \frac{[T]_0 - [L]_0 (1 - R)}{(1/R) - 1} \tag{12}$$

$$k_{off} = \frac{\ln\left(\frac{A_C + A_D}{A_C}\right)}{t_C} \tag{13}$$

The value of k_{on} is calculated using **Eq. 2**.

In addition to studying the properties of BMIs, NECEEM can also be used as an affinity probe-based analyte quantitation assay: when the K_d is known, but the initial concentration of T is not, **Eq. 12** can be rearranged to find the unknown concentration:

$$[\mathrm{T}]_0 = K_\mathrm{d}((1/R)-1)+[\mathrm{L}]_0\left(1-R\right) \qquad (14)$$

Lastly, besides being useful as an analytical technique, NECEEM can also be used for preparative applications, with the selection of DNA aptamers being the most prominent one. Aptamers are *in vitro*-selected DNA- or RNA-based affinity ligands that display strong binding affinities and specificities toward their molecular targets **[46, 47]**. DNA aptamers are obtained from diverse combinatorial libraries of oligonucleotides through a procedure of Systematic Evolution of Ligands by Exponential Enrichment (SELEX) **[46]**. Briefly, SELEX is performed in four steps: 1) a target molecule is incubated with a library of single-stranded (ss)DNA molecules of random nucleotide sequences; 2) target-bound DNA ligands are separated from sequences that do not have affinity for the target; 3) collected DNA molecules are amplified by polymerase chain reaction (PCR), and the "sense" strands are purified from the complementary strands; steps 1-3 are then repeated in a number of rounds, until the resultant pool of DNA molecules is sufficiently enriched with aptamers; 4) lastly, individual aptamer sequences are extracted from the DNA pool, and their affinities and specificities for the target are characterized. The superb separation efficiency of NECEEM **[48]**, as well as its capability of generating aptamers with pre-defined affinity and kinetic properties (smart aptamers) **[49]**, makes it a powerful partitioning technique for aptamer selection.

1.3.2. History of NECEEM development and application

NECEEM was invented in 2002 to become the inceptive method in the KCE toolbox **[50]**. In that proof of principle work, **Eqs. 12** and **13** were derived and the utility of NECEEM was

demonstrated using a model interaction between a DNA oligonucleotide and Single-strand DNA-binding protein (SSB) from *E. coli*. In 2005, it was demonstrated that NECEEM can be used to study the thermodynamic properties of BMIs, which was exemplified by model systems of DNA and two proteins: *Taq* polymerase and SSB **[51]**. By measuring k_{on} and k_{off} at different temperatures, controlled by a sheath-liquid cooling system, and subjecting the data to Van't Hoff analysis, the changes in enthalpy (ΔH) and entropy (ΔS) of binding were determined. In that work it was assumed that the temperature inside the capillary was equal to the temperature of the coolant; this assumption was challenged when a new method for measuring in-capillary temperatures was developed, based on diffusion of a species with a known diffusion coefficient **[52]**. It became apparent, that not only is the in-capillary temperature significantly different from that of the coolant liquid, but also that the temperatures in the cooled and non-cooled portions of the capillary differ dramatically, necessitating the propagation of the sample zones through the non-cooled portion **[43, 44]**. The diffusion-based temperature measurements were prohibitively slow for routine monitoring of in-capillary conditions, with each experiment requiring as long as an hour. In response, the Universal Method for Determining Electrolyte Temperatures (UMET) in CE was developed, which allows the estimation in-capillary temperatures on the basis of electrolyte conductivity and electrophoresis power consumption, and allows a series of measurements to be performed within a few minutes **[53]**. A simplified version of UMET (SUMET) reduced the total time of temperature measurement to under one minute, without significant loss of accuracy **[54]**.

By 2011, it had become apparent that the existing data extraction approach from NECEEM electropherograms had a major flaw, in that it relied on subjective placement of boundaries between the peaks and the decay bridge (**Figure 1.3B**), leading to a very large user-to-user

variation in the analysis of the same data. To eliminate this issue, an objective method for deconvolution of the overlapping areas between electropherogram features was developed, which is based on the assumption of axial symmetry of the peaks **[45]**. When it was determined that NECEEM peaks are often asymmetric, due to the initial shape of the sample injection zone, a pressure propagation-based method for correcting this asymmetry was introduced **[55]**.

As evidenced by the previous examples, analysis of DNA–protein interactions is the primary application of NECEEM; however, such interactions often do not occur at 1:1 stoichiometry, which is assumed in most of NECEEM's theory, but instead involve large multi-protein complexes. To enable the study of such interactions, new theory and a mode of data analysis have been developed **[56, 57]**. The new approach was tested using a DNA oligonucleotide which was designed to bind up to 4 SSB proteins. It was concluded, based on experimental and *in silica* studies, that the NECEEM-based approach for studying multi-protein–DNA complexes is much more accurate than the non-separation approaches, such as SPR. A similar approach was later applied to study DNA-binding of multiple molecules of the methyl-CpG binding domain **[58]**. Besides the analysis of DNA, NECEEM has been applied to study the interactions of other types of molecules, including antibody–peptide **[59]**, vitamin–protein **[60]**, organic polymers and a dye **[61]**, and protein-small molecule compounds **[62]**. Advances in interfacing CE with MS-detection are particularly promising, as they enable studies of interactions between such pairs of molecules as organic polymers and metal ions **[63]**, and proteins with DNA-barcoded small-molecule drugs **[64, 65]**.

As mentioned previously, in addition to having great analytical utility, NECEEM also serves as an efficient partitioning tool in aptamer selection. This functionality was first noted within a year of development of NECEEM, when the method was applied to the analysis of an interaction

between thrombin and its DNA aptamer, with the authors suggesting that the method can be used not only for analysis of aptamers but also for their selection **[66]**. Between 2004 and 2005, two groups have shown the superb efficacy of NECEEM as a partitioning method in aptamer selection within a short period of time **[48, 67]**. Bowser's group generated an aptamer with K_d = 30 nM to human Immunoglobulin E within 4 rounds of selection, while our group generated an aptamer with K_d = 0.5 nM to protein farnesyltransferase from *S. cerevisiae* within a single round (although this turned out to be an exceptional case, with most subsequent NECEEM-based selections requiring between 4 and 8 rounds). As aptamer selection performed with conventional partitioning methods, such as filtration or chromatography, takes an average of 20 rounds to complete, both examples have demonstrated the outstanding improvement in the partitioning efficiency that NECEEM had provided for this application. In fact, the separation efficiency of NECEEM has proven to be so high that it has enabled the development of a completely new approach to aptamer selection, non-SELEX, which omits the PCR amplification of the collected pools between the selection rounds, significantly simplifying the process **[68]**. The kinetic information from NECEEM was utilized in aptamer selection for the first time in 2006 when a number of aptamers with pre-defined k_{off} and K_d values were selected for MutS protein from *T. aquaticus* **[49]**. The usefulness of such an approach was demonstrated when three of these aptamers were later used to enable detection of the target protein with an ultra-wide dynamic range **[69]**. Another feat in aptamer selection that was enabled by NECEEM's remarkable separation efficiency was the development of the aptamer-facilitated protein isolation from cells (AptaPIC) technology, which allows the selection of aptamers for an overexpressed protein within the context of a cell lysate, and subsequently enables affinity purification of the target protein using the developed aptamers. AptaPIC was successfully applied to MutS and human

platelet-derived growth factor (PDGF) proteins **[70, 71]**. Since its inception, NECEEM had been used to select aptamers for such targets as cell division control protein 42, myotonic dystrophy kinase-related Cdc42-binding kinase, p21-activated kinase 1 **[72]**, AlkB **[73]**, catalase **[74]**, Ara h1 **[75]**, lipopolysaccharide endotoxin **[76]**, vascular endothelial growth factor **[77]**, leptin **[78]**, and thrombin **[79]**.

1.3.3. Technological limitations of NECEEM and project goals

Between the proof-of-principle work and the time of commencement of my research project, significant advances in development of NECEEM have been made; however, a number of limitations precluded it from becoming a truly practical technique. This is evidenced by the relatively slow rates of adoption of NECEEM, with alternative approaches still largely predominating in academic laboratories, and even to a larger degree in clinical and industrial settings. The reluctance of the analytical community toward the greater use of NECEEM can be attributed to three critical technological issues: **(i)** poor resilience of the method against systematic errors, **(ii)** limited versatility in terms of emulating physiological conditions in experiment, and **(iii)** restricted partitioning efficiency in DNA aptamer selection applications.

It has been observed, by us and by others, that NECEEM often produces results that differ significantly from those produced by alternative methods, largely due to the incorporation of considerable systematic errors into the measurements **[80]**. In particular, the ruggedness and robustness of the method have proven to be unsatisfactory. Our internal testing had revealed that measurements performed with the same interacting molecular pair at different points in time often result in progressively changing K_d values. These findings contradict the conceptual basis of NECEEM, which was developed on the premise of improved accuracy. Examination of the experiments that produced imprecise and inaccurate data had revealed two sources of such

issues. The first issue has to do with the fact that the accuracy of NECEEM measurements strongly depends on the certainty of our knowledge of the concentration of T, especially if it is a protein. Most protein concentration assays cannot distinguish between active and inactive forms of a protein, and if a given ligand binds only to active forms, using the total protein concentration can lead to the introduction of significant errors. The second issue has to do with the optimization of NECEEM experiments. Even when the concentrations of interacting species are known precisely, any equilibrium and kinetic measurement (performed with NECEEM or any other method) may be susceptible to large systematic errors if the experimental conditions are not chosen correctly. We have discovered that, unlike most other techniques, KCE methods are quite difficult to optimize, due to the complex interrelation between various experimental factors and their counteracting effects on the produced results. Another long-standing limitation of KCE methods to applications in the study of BMIs was their poor compatibility with highly conductive physiological solutions. All of the studies discussed in **Section 1.3.2** were performed using a narrow set of low-ionic strength run buffers, due to their relatively small tendency to generate Joule heat and slow time of ion depletion. The restricted variety of compatible run buffers is a major drawback of NECEEM, as it may yield results that are not representative of actual behaviors that occur in cells. Thus, KCE methods will benefit greatly if they can be performed with physiological solutions as the BGE. The final issue has to do with the partitioning efficiency of NECEEM in aptamer selection. As mentioned previously, the use of NECEEM in aptamer selection has reduced the number of required rounds from 20 to fewer than 8. While significant, this improvement is worse than expected: theoretical predictions suggest that CE partitioning can reduce aptamer selection process down to a single round of SELEX. Understanding and eliminating the cause of this underwhelming efficiency of NECEEM

partitioning can make the method a truly disruptive technology in the field of evolutionary selection of ligands.

The purpose of my research project is to resolve the above-described technological limitations of NECEEM in order to make it a more practical tool. The milestones of my project include the following goals: to improve the accuracy of NECEEM by **(i)** developing an approach that allows simultaneous determination of K_d and the active concentration of T, and by **(ii)** developing an objective and systematic approach for optimization of NECEEM experimental conditions; **(iii)** to improve the versatility of NECEEM by making it compatible with the use of physiological buffers; and **(iv)** to elucidate the cause and to eliminate the restricted partitioning efficiency of NECEEM in aptamer selection. My progress toward these goals is described in the following Chapters. In Chapter 2, I describe an approach for simultaneous determination of both K_d and concentration of T, which is based on a new set of mathematical tools that extract information from multiple NECEEM experiments by means of fitting. I describe the validation of the new approach using an *in silico* study, and its subsequent application in experiment. In Chapter 3, I present an "algorithmic" approach to NECEEM optimization, which takes the complex interrelation between its numerous experimental parameters into account and allows researchers to approach it in a systematic manner. I validate the approach using comprehensive *in silico* analysis and show its utility within an experiment. In Chapter 4, I test the feasibility of using phosphate buffered saline (PBS) as a NECEEM run buffer and show that its usage renders DNA samples undetectable. I uncover the causes of this previously unreported detrimental effect and come up with a modification of NECEEM that overcomes the issue. Lastly, in Chapter 5, I address the preparative aspects of NECEEM and report on the surprising findings that the cause of its decreased partitioning efficiency lies in the ability of DNA to form unusually stable

complexes with its counterions, and that the electric field-induced dissociation of such complexes results in irregular electrophoretic migration of DNA. Another unusual effect of the electric field-induced dissociation of DNA–counterion complexes is the precipitation of DNA out of the solution. Finally, I suggest two approaches to minimizing the detrimental effects of this unusual phenomenon in its application to aptamer selection.

Chapter 2. Using NECEEM for Simultaneous Determination of Concentration and Equilibrium Constant

The presented material was published previously and reprinted with permission from "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". Copyright 2015 American Chemical Society.

My contribution to the article was: (i) developing the approach for determination of optimal conditions without *a priori* knowledge of K_d, (ii) developing the *in silico* model for NECEEM, (iii) developing the software for fitting of the experimental data (iv) performing a part of the *in silico* experiments and all of the *in vitro* experiments, (v) writing a major part of the manuscript.

2.1. ERRORS DUE TO UNCERTAINTY IN TARGET CONCENTRATION

In NECEEM, K_d and k_{off} are measured directly, while k_{on} is calculated using **Eq.** 2. As a result, any inaccuracies in the measurement of K_d will also be incorporated into the value of k_{on}, thus, affecting the accuracy of the kinetic measurements. A common issue which affects the accuracy of K_d measurements is the uncertainty in the concentration of T, especially if it is a protein. Most protein concentration assays, like Bradford assay or Bicinchoninic Acid Assay (BCA), rely on quantitation of specific chemical bonds (*e.g.* peptide bond), and cannot distinguish between properly folded and denatured or proteolysed forms of a protein **[81]**. If a given ligand binds only the properly folded forms of the protein (which is a common and desirable property of affinity probes), using an assay-measured protein concentration (with active and inactive forms) in the calculation of the equilibrium constant can lead to significant errors. This problem is especially severe in cases of unstable proteins, or proteins which are difficult to

purify, as their assay-measured concentrations will often be confounded by the presence of impurities.

The previously used mathematical treatment for NECEEM allowed us to determine either an unknown K_d if the concentrations of both T and L are known or the unknown concentration of T if K_d and the concentration of L are known **[69, 82]**; simultaneous determination of both K_d and T was not possible. Here I describe an approach that allows such measurements to be performed. This approach requires NECEEM experiments to be performed at varying concentrations of L but at a constant unknown concentration of T. The experimental results are presented as a fraction R against concentration of L and fitted with a theoretical dependence of these variables, in which K_d and the concentration of T are used as fitting parameters. These parameters are determined as those resulting in the best fit. To test the suggested approach for finding K_d and target concentration, as well as to validate the software serving it, COMSOL multiphysics software was utilized to accurately model the NECEEM processes and simulate electropherograms at various sets of initial conditions *in silico*. The simulated electropherograms were processed as experimental ones and the returned values of K_d and concentration of T were confirmed to be within 10% of the values programmed into the simulations over a wide range of initial conditions, demonstrating the remarkable stability of the developed approach. Lastly, the *in silico*-validated method was used to determine the protein concentration of in-house expressed and purified MutS protein and K_d of its interaction with a DNA aptamer. The results suggest, that the previously measured K_d value for this interaction was significantly overestimated by a single-point NECEEM approach **[49]**. The determined concentration of the protein was in perfect agreement with the BCA assay measurements.

2.2. EXPERIMENTAL SECTION

2.2.1. Materials

All chemicals and buffer components were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries were purchased from Molex (Phoenix, AZ). *T. aquaticus* MutS recombinant protein was previously expressed in-house, and purified by an aptamer-facilitated affinity pulldown method, similar to recently described **[71]**. The stock concentration of the purified MutS protein was measured using a BCA assay kit (Thermo Scientific, Wilmington, DE), according to manufacturer's instructions. A solution of Bovine Serum Albumin (BSA) of known concentration was used to make calibration curve standards. As BCA assay detects peptide bonds, the final concentration values were normalized according to the ratio of peptide bonds between MutS (816 bonds) and BSA (582 bonds). DNA aptamer (clone ID: 2-06) with affinity toward MutS protein was selected previously **[49, 83]**, and was custom synthesized by Integrated DNA Technologies (Coralville, IA). The nucleotide sequence of the fluorescently labeled, ssDNA MutS aptamer was: 5'- Fluorescein - 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'. NanoDrop-1000 spectrometer (Thermo Scientific, Wilmington, DE) was used to verify the concentration of DNA stock by measuring light absorbance at 260 nm, and dividing the absorbance by a manufacturer-provided extinction coefficient.

2.2.2. Capillary electrophoresis

Analyses were carried out using both a commercial and a custom-made CE instruments. The commercial instrument, P/ACE MDQ (Beckman Coulter, Mississauga, ON, Canada), was equipped with a standard fluorescence detector and a 488 nm line of continuous wave solid-state

laser (JDSU, Santa Rosa, CA) for fluorescence excitation. Runs were performed in a 30 cm-long uncoated capillary, with on-column detection 20 cm away from the capillary inlet. The fused-silica capillary had an inner radius of 25 μm and an outer radius of 180 μm. The temperature of the capillary coolant was set to 25 °C.

The custom instrument was described in detail previously **[84, 85]**. The instrument was equipped with a highly sensitive off-column fluorescence detection scheme, with a sheath-flow hydrodynamic focusing cell. A few modifications were incorporated into the existing design to facilitate fluorescent measurements in the required pM range of aptamer concentrations. A diode-pumped solid-state laser (AixiZ, Houston, TX) was applied for excitation of fluorescence. The laser beam had a diameter of about 2 mm and a power of 60 mW (attenuated to 40 mW) at 473 nm. An optical bandpass filter of 525 ± 25 nm (Semrock, Rochester, NY) was used to select for the fluorescent signal. A 20 Hz low-pass electronic filter was introduced between the photomultiplier tube R1477 (Hamamatsu) and the analog-to-digital signal converter PCI-6036E (National Instruments), to reduce electronic noise. To eliminate random spikes in the signal due to laser light scattering by gas bubbles and dust particles, a Millex liquid filter unit (Millipore, Billerica, MA) with 0.22 μm pore size was incorporated into sheath-flow fluidic line. To further attenuate random signal spikes, the digitized signal was treated by a 3-point-median digital filter. Electrophoresis high-voltage electronics were segregated from the detection electronics to reduce effects of electromagnetic interference on the recorded signal. Runs were performed in a 40 cm-long uncoated capillary, with off-column detection at the end of the capillary. The fused-silica capillary inner radius was 25 μm and outer radius 75 μm. The measurements were carried out at room temperature of 23°C.

For both instruments, CE experiments were conducted in the following manner. Both the inlet and the outlet reservoirs contained the electrophoresis run buffer (50 mM Tris-HCl at pH 8.5). Prior to every run, the capillary was rinsed with the run buffer solution at 30 psi (206.8 kPa) pressure for 1 min. At the end of each run, the capillary was rinsed with 100 mM HCl, 100 mM NaOH, and deionized water, with the same pressure for 1 min each. The samples were injected into the capillary, pre-filled with the run buffer, by a 1 psi (6.9 kPa) pressure pulse of 3.7 and 5.0 s for commercial and custom-made instruments, respectively. The length of the sample plugs in both cases was calculated to be 6.8 mm. Electrophoresis was carried out with the anode at the injection end of the capillary, with electric field strengths of 600 V/cm and 400 V/cm for commercial and custom-made instruments, respectively.

2.2.3. NECEEM

NECEEM experiments were performed to study the interaction between MutS protein and a DNA aptamer. All dilutions were prepared with electrophoresis run buffer, to minimize issues from electrolyte mismatch between sample plug and the run buffer. For all experiments, protein stock solution concentration was treated as unknown, and samples were expressed as a dilution relative-to-stock. DNA stock solution was prepared at a concentration of 1 μM, and had Rhodamine 110 mixed in at a 7 μM concentration, to act as an internal standard. Mixtures of various concentration combinations were prepared, as described in the following sections. For experiments with the commercial instrument, performed at nanomolar concentrations, mixtures were incubated at room temperature for 10 min to achieve equilibration. For experiments with the custom-made instrument, performed at pM concentrations, mixtures were incubated at room temperature for 60 min, as equilibration proceeded more slowly. After incubation, the equilibrium mixtures were subjected to CE as described above. Each experimental point was

measured in triplicates, with a fresh equilibrium mixture prepared for each replicate. R coefficients were extracted from resultant electropherograms by NAAP program in batch **[45]**. Values of K_d and $[T]_0$ parameters were obtained by fitting the dependence of R on $[L]_0$ into **Eq. 16**, using Microsoft Excel 2010 Solver Add-in.

2.3. RESULTS AND DISCUSSION

2.3.1. Mathematical solution for simultaneous determination of K_d and $[T]_0$

The mathematical solution for simultaneous determination of K_d and $[T]_0$ is simple. By multiplying **Eq. 12** by its denominator we can transform this equation into a quadratic equation with respect to R:

$$R^2 + \frac{K_d + [T]_0 - [L]_0}{[L]_0} R - \frac{K_d}{[L]_0} = 0 \qquad (15)$$

A physically meaningful positive solution of **Eq. 15** has a form:

$$R = -\frac{K_d + [T]_0 - [L]_0}{2[L]_0} + \sqrt{\left(\frac{K_d + [T]_0 - [L]_0}{2[L]_0}\right)^2 + \frac{K_d}{[L]_0}} \qquad (16)$$

Eq. 16 explicitly relates R with K_d, $[T]_0$ and $[L]_0$. It can be used to find unknown K_d and $[T]_0$ if it is applied to treat data from multiple NECEEM experiments performed with varying known $[L]_0$ but a fixed unknown $[T]_0$ (*i.e.* by titration with L). Such a series of experiments will produce a set of R values that can be plotted as a function of $[L]_0$, and this experimental dataset can be fitted with **Eq. 16** in which R and $[L]_0$ are dependent and independent variables, respectively, and K_d and $[T]_0$ are the fitting parameters.

2.3.2. Considerations for avoiding systematic errors

The described mathematical procedure is simple, but utilizing it in practice requires some additional considerations. There are three questions that will concern an experimenter: (i) in

which range should the $[L]_0$ be varied during its titration (*i.e.* what should be the values of $[L]_{0,first}$ and $[L]_{0,last}$)?; (ii) how many points should be measured between $[L]_{0,first}$ and $[L]_{0,last}$?; and (iii) what relative-to-stock-solution concentration of T (*i.e.* dilution of $[T]_{stock}$) should be used during the titration with L? Answers to these questions are important, as an improper choice of $[L]_0$ and dilution of $[T]_{stock}$ can significantly diminish the accuracy of measurements.

When determining K_d, the experimental error in measuring R can become amplified if the experiments are performed at $[L]_0$ significantly larger than K_d. This can be made apparent by examining the derivative of **Eq. 17** with respect to R (see **Appendix A.1** on **page 143** for differentiation), which relates small variations in experimental measurements (ΔR) to systematic error in calculated final result (ΔK_d):

$$\Delta K_d \approx \left(-[L]_0 + \frac{[T]_0}{(1-R)^2} \Delta R \right) \qquad (17)$$

Using **Eq. 17**, the dependence of relative error, $\Delta K_d/K_d$, on dimensionless ligand concentration $[L]_0/K_d$ at constant R and ΔR, can be calculated and plotted (**Figure 2.1A**). Similarly, the dependence of $\Delta K_d/K_d$ on the value of R can be calculated and plotted at constant $[T]_0$ and ΔR (**Figure 2.1B**). From **Eq. 17**, increasing the value of $[L]_0$ will, at a certain point, result in ΔK_d value which is larger than K_d itself. If the sign of ΔR is positive, this can lead to 0 or negative values of K_d, which are physically meaningless for reversible interactions. As apparent from **Figure 2.1A**, the values of ΔK_d are consistently small as long as $[L]_0$ is smaller or equal to

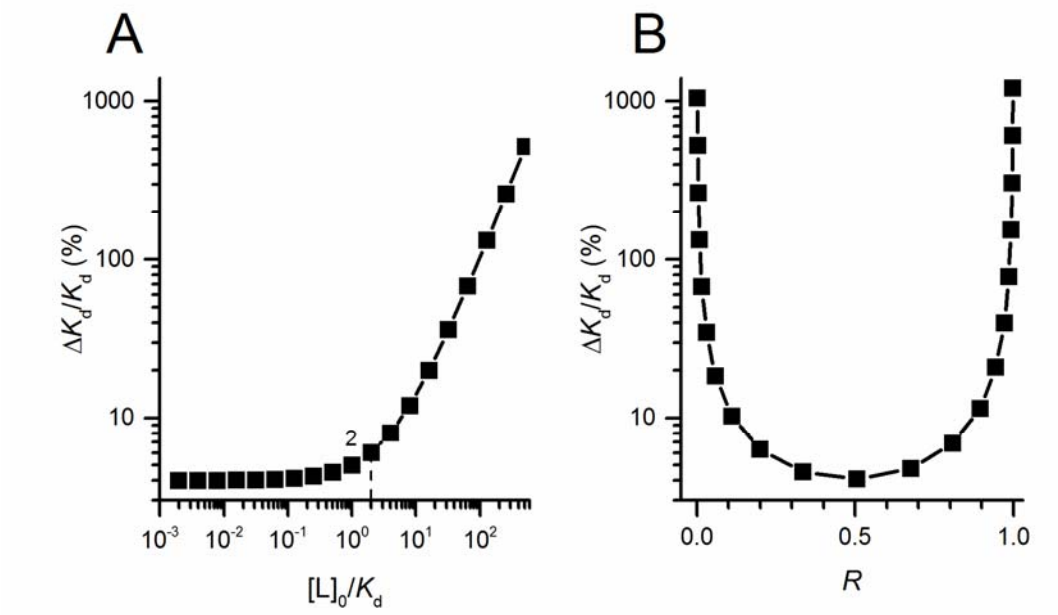

Figure 2.1. Dependence of relative error in the determination of K_d on initial concentrations. Figure shows the change in relative error in the determination of K_d, depending on the ratio of $[L]_0$ to K_d (panel A) and the value of R (panel B). **Panel A:** ΔK_d was calculated using **Eq. 17** by varying $[L]_0$ and $[T]_0$ to yield $R = 0.5$. K_d was set to 1, and ΔR equaled 1% of R. **Panel B:** ΔK_d was calculated using **Eq. 17** by varying $[L]_0$ to yield R values in the range of 0.01 and 0.99. K_d was set to 1, $[T]_0 = K_d$, and ΔR equaled 1% of R. Note that both y-axes are presented in logarithmic scale.

$2K_d$, which represents the point at which concentrations of equilibrium components are perfectly balanced (*i.e.* $[L]_{eq} = [T]_{eq} = [C]_{eq} = K_d$). As $[L]_0$ increases over $2K_d$, however, the values of ΔK_d begin to increase exponentially. Thus, when the K_d value is not known *a priori*, it is practical to keep the value of $[L]_{0,first}$ as the lowest concentration of L which produces reliably measurable areas of all peaks in a NECEEM electropherogram. The exact value is determined by the LOQ of the NECEEM method, which depends on both the sensitivity of the instrument used for the measurements and the accuracy of the data-treatment approach. In this work the LOQ of NECEEM is defined solely by the sensitivity of fluorescence detection to facilitate this consideration. In CE, the LOQ is defined as the concentration of an analyte which results in a

signal with a peak height that exceeds the background noise by a certain threshold value. To ensure that all 3 areas characteristic of NECEEM electropherograms (described in **Figure 1.3**) are accurately quantified in our experiments, I defined the threshold value to equal 100 (this choice is further justified in **Section 3.3.2** on **page 50**). It should be noted, that if an experiment performed at $[L]_{0,first}$ = LOQ results in an apparent K_d value which is smaller than $2[L]_{0,first}$, then the produced results may be inaccurate.

Once the value of $[L]_{0,first}$ is determined, the concentration of $[L]_0$ should be increased until some maximum value $[L]_{0,last}$ is reached. At a constant $[T]_{0,}$ increasing $[L]_0$ will result in R asymptotically approaching the value of 1. According to **Eq. 17**, small variations in R measured close to the asymptote, will yield large errors in K_d. The dependence of ΔK_d on increasing values of R, at constant $[T]_0$ and $\Delta R_,$ is plotted in **Figure 2.1B**. According to this plot, with ΔR set to 1% of K_d, a relative error typical for NECEEM experiments, the value of ΔK_d surpasses the value of K_d at $R \approx 0.9$. This means that points above $R = 0.9$ carry little information on the value of K_d. While the precise value at which ΔK_d overcomes K_d is strongly dependent on the precision of experimental measurements, it is suggested, as a general rule, that the value of $[L]_{0,last}$ should not result in R values above 0.9.

As a fitting-based method, the accuracy of the results produced by the proposed approach relies on the statistical significance of the available data. The least squares technique, used in the fitting of experimental data, requires that at least one degree of statistical freedom be available to the non-linear regression algorithm **[86]**. One degree of freedom is lost when experimental replicate values are averaged and their standard deviation is calculated. During the fitting, the adjustment of two parameters, K_d and $[T]_{0,}$ results in the loss of another two degrees of freedom, for a total of 3 degrees of freedom lost. Thus, a minimum of 4 statistically significant

experimental points (*i.e.* points with non-overlapping standard deviation ranges), including $[L]_{0,first}$ and $[L]_{0,last}$, are required for non-linear regression analysis. Increasing the number of points will further strengthen the confidence levels associated with the resulting values.

Lastly, accurate determination of K_d requires that the value of R is sensitive to the changes in K_d. By differentiating **Eq. 16** with respect to K_d under an assumption that $[L]_0 << [T]_0$, a condition which should be satisfied if $[L]_0 << K_d$, we obtain that R is most sensitive to K_d at $[T]_0 \approx K_d$ (for proof, see **Appendix A.2** on **page 143**). Although neither K_d nor the concentration of T is known, satisfying this conditions is possible by performing one set of preliminary experiments. If $[L]_0 << [T]_0$, a consideration that has been discussed above, then $[T]_0 \approx K_d$ when $R \approx 0.5$. Thus, a series of NECEEM experiments must be performed at $[L]_{0,first}$ held constant, and varying unknown concentrations of $[T]_0$, until the optimal dilution factor, a, is that yields $R \approx 0.5$ is determined. This dilution factor can then be used in NECEEM experiments with varying $[L]_0$ described above.

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

2.3.3. *In silico* modeling and validation.

The best way to validate the proposed approach, as well as to test its robustness and stability, is to subject it to an *in silico* study. Essentially, we can define a specific set of values for K_d and $[T]_0$ and use them to simulate accurate and realistic NECEEM electropherograms, which can then be treated with the proposed method. The accuracy of the proposed approach can be

objectively evaluated by comparing the procedure-extracted values of K_d and $[T]_0$ to the ones pre-defined in the simulations. I have previously developed an *in silica* model for a different KCE method, MASKE, in COMSOL Multiphysics environment (see **Appendix B**), which I have since adapted to NECEEM. Briefly, the capillary was defined as a one-dimensional interval consisting of a discrete number of mesh points, and the chemical species of L, T and C were distributed along this geometry to match the initial conditions used in NECEEM. The direct PARDISO solver was used to calculate the concentrations of L, T, and C at each mesh point in this geometry in a time-dependent manner. The simulation incorporated the mass balance equations, reactions of binding and unbinding, convective and electrophoretic terms, and Fick's law of diffusion. For most adequate modeling, I used realistic values for all experimental parameters including capillary length, position and size of detection window, shape of injected plug of the equilibrium mixture, diffusion coefficients and electrophoretic mobilities of the species, and an electric field. Simulated electropherograms were plotted as a time dependence of the sum of concentrations of L and C averaged over the detection volume with a constant level of noise. The noise was generated as random signal with a standard deviation of the amplitude corresponding to 1% of the lowest simulated signal.

The first set of NECEEM experiments was simulated to find the dilution factor, $a = [T]_{stock}/[T]_0$, by which the stock solution of T of concentration $[T]_{stock}$ needs to be diluted to satisfy $[T]_0 \approx K_d$. $[L]_0 = 10$ nM was chosen as the typical concentration of a fluorescently labeled molecule that can generate signal suitable for NECEEM measurements. We chose $K_d = 100$ nM as a typical value for affinity binding of biological molecules. The "unknown" concentration of target stock solution, $[T]_{stock}$, was chosen to be 10 μM. NECEEM electropherograms were simulated for the stock concentration of T and a number of 2-time dilutions (**Figure 2.2A**). The

three areas required for calculation of R in **Eq. 11** were found in an unbiased way by using the "NECEEM Area Analysis Program" (NAAP), which I have helped to develop **[45]**. R values were plotted as a function of the dilution factor a (**Figure 2.2B**). The value of a that resulted in $R \approx 0.5$ was estimated from the graph ($a = 100$ in this example). At this point, we could calculate $[T]_0$ that corresponded to $R \approx 0.5$ and compare it with K_d defined as 100 nM: $[T]_0 = [T]_{stock}/a = 10\ \mu M/100 = 100$ nM. The equality between $[T]_0$ and K_d confirms that this part of the procedure works correctly.

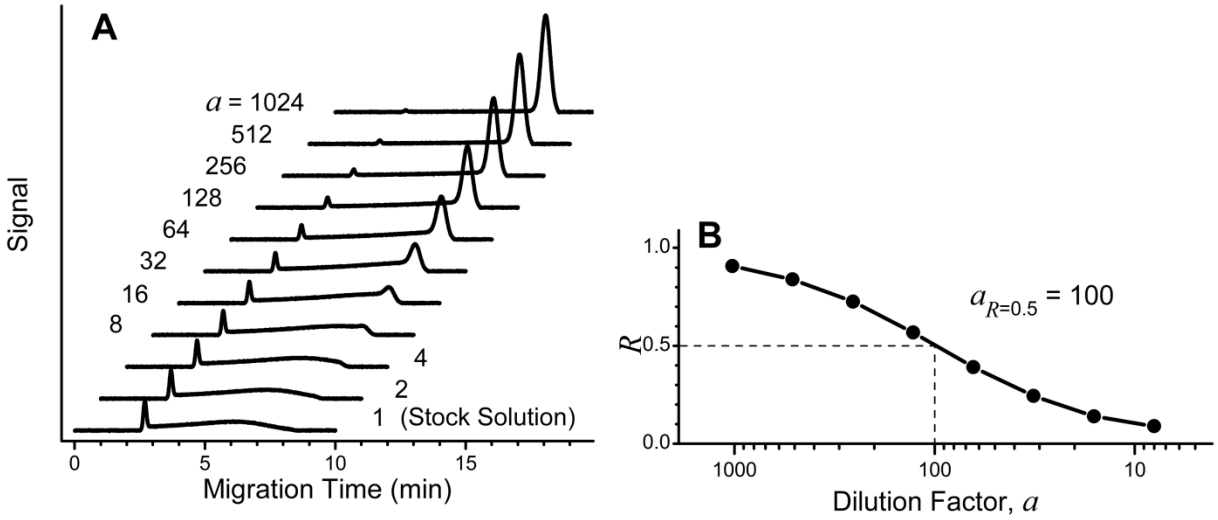

Figure 2.2. Validation of the procedure for determination of optimal dilution factor *a*. **Panel A**: A set of simulated NECEEM electropherograms reveals the effect of serial dilutions of $[T]_{stock}$, whereas concentration $[L]_0$ is fixed to 10 nM. A dilution factor a is changed from 1 to 1024. **Panel B**: Dependence of R (fraction of unbound L) on a. Values of R are calculated from the simulated electropherograms. The optimal factor $a = 100$ corresponds to a dilution at which $R = 0.5$.

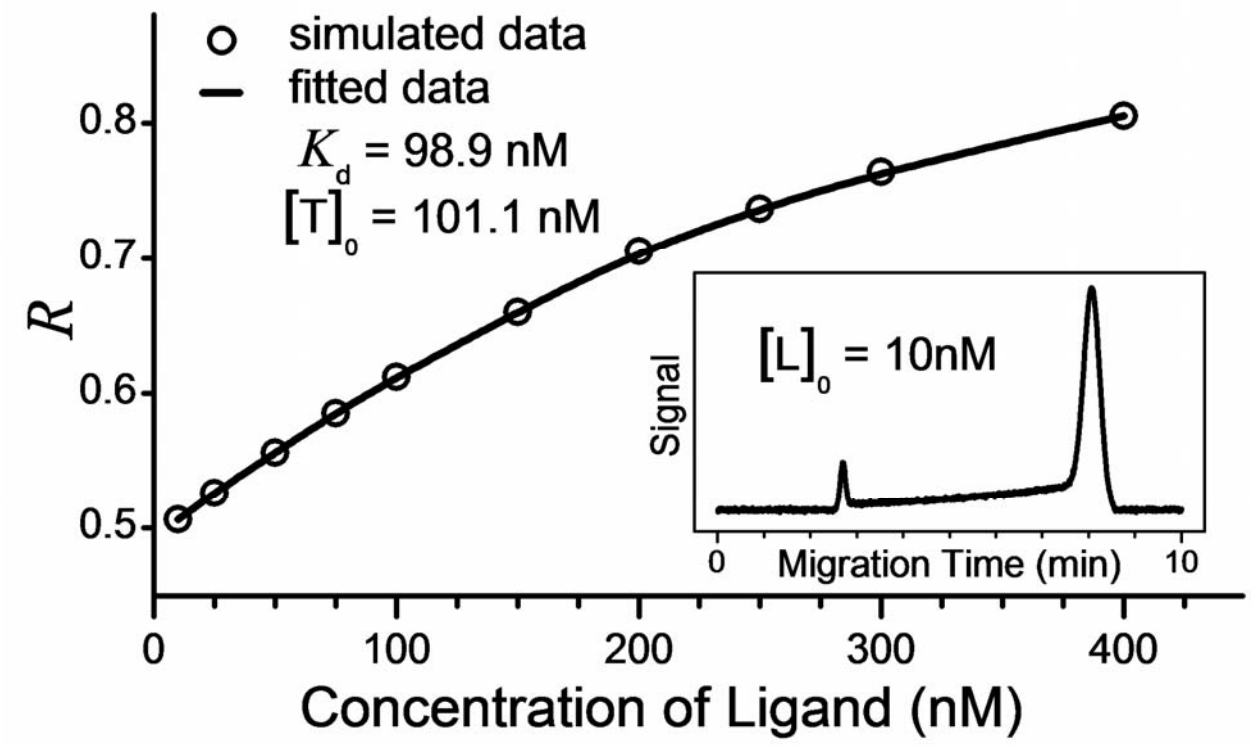

Figure 2.3. Validation of the procedure for determination of unknown $[T]_0$. Dependence of *R* (fraction of unbound L) on concentration $[L]_0$ at constant concentration $[T]_0$ obtained from the stock solution by diluting with factor $a = 100$. Open circles represent calculated *R*-values, solid curve is a fitting with **Eq. 16.** The insert demonstrates typical simulated NECEEM electropherogram used for the *R* calculation.

The second set of NECEEM experiments was simulated to find "unknown" values of K_d and $[T]_0$. In this set, I used $[T]_0 = [T]_{stock}/a = 100$ nM and varied $[L]_0$ between 10 and 400 nM. An example of the resulting simulated NECEEM electropherograms is shown in the insert of **Figure 2.3**. *R* values were calculated from these electropherograms and plotted as a function of $[L]_0$ (open circles in **Figure 2.3**). The solid curve in **Figure 2.3** is the fitting of the *R* values with the theoretical dependence of *R* versus $[L]_0$ in **Eq. 16**. The best fit was obtained for $K_d = 98.9$ nM and $[T]_0 = 101.1$ nM. These values are equal (with 99% accuracy) to the values used in

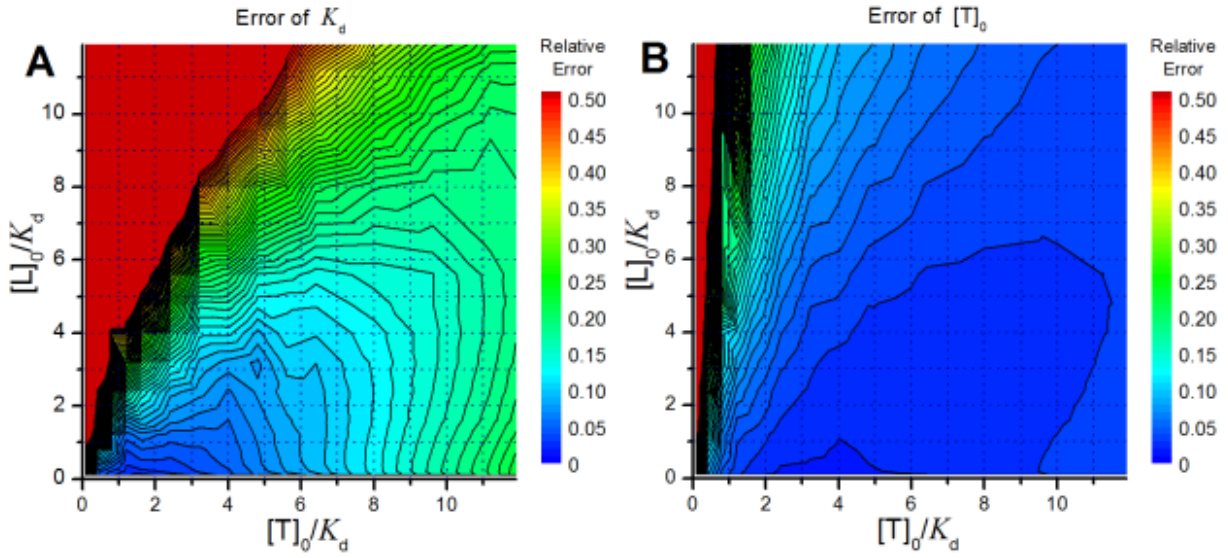

Figure 2.4. Validation of the fitting procedure over a wide range of initial concentrations. Error relative to K_d in determination of K_d (**Panel A**) and $[T]_0$ (**Panel B**) depending on the ratio of concentrations $[L]_0$ and $[T]_0$ to K_d. R values were calculated from simulated electropherograms. K_d was fixed to 100 nM. Black contour lines represent steep changes in values of the errors.

COMSOL simulations, which confirms that the entire procedure (the mathematics, the algorithms, and the software) works correctly.

To assess the robustness of the approach used in the second set of NECEEM experiments, we simulated this set of NECEEM electropherograms at a wide range of $[L]_{0,first}$ and $[T]_0$. In essence, we calculated R values at increasing $[L]_0$ for a series of $[T]_0$, both greater and lesser than K_d. For every graph, we found K_d and $[T]_0$ and compared them with the correct ones (pre-set in COMSOL calculations to simulate the electropherograms). The error was calculated as the absolute difference between the determined and correct values relative to the correct value (**Figure 2.4**).

Slices of these contour plots at $[L]_0/K_d = 0.2$ appear in **Figure 2.5**, and illustrate the general trends in the dependence of the errors on $[T]_0/K_d$. In general, a 3-time overstepping of $[T]_0$ above K_d did not cause the error to surpass more than 5%, which suggests that an overestimate of $[T]_0$ from the first set of NECEEM experiments (**Figure 2.2**) can be suitable for the second set of NECEEM experiments (**Figure 2.3**). As expected, using $[L]_{0,first}$ values which were significantly above $2K_d$ led to unacceptably high errors. Interestingly, these results also suggest that the accuracy of determination of $[T]_0$ is not affected by the improper choice of $[L]_{0,first}$ and $[T]_0$, as severely as the accuracy of K_d determination. These encouraging results have prompted me to apply the proposed approach to an *in vitro* experiment.

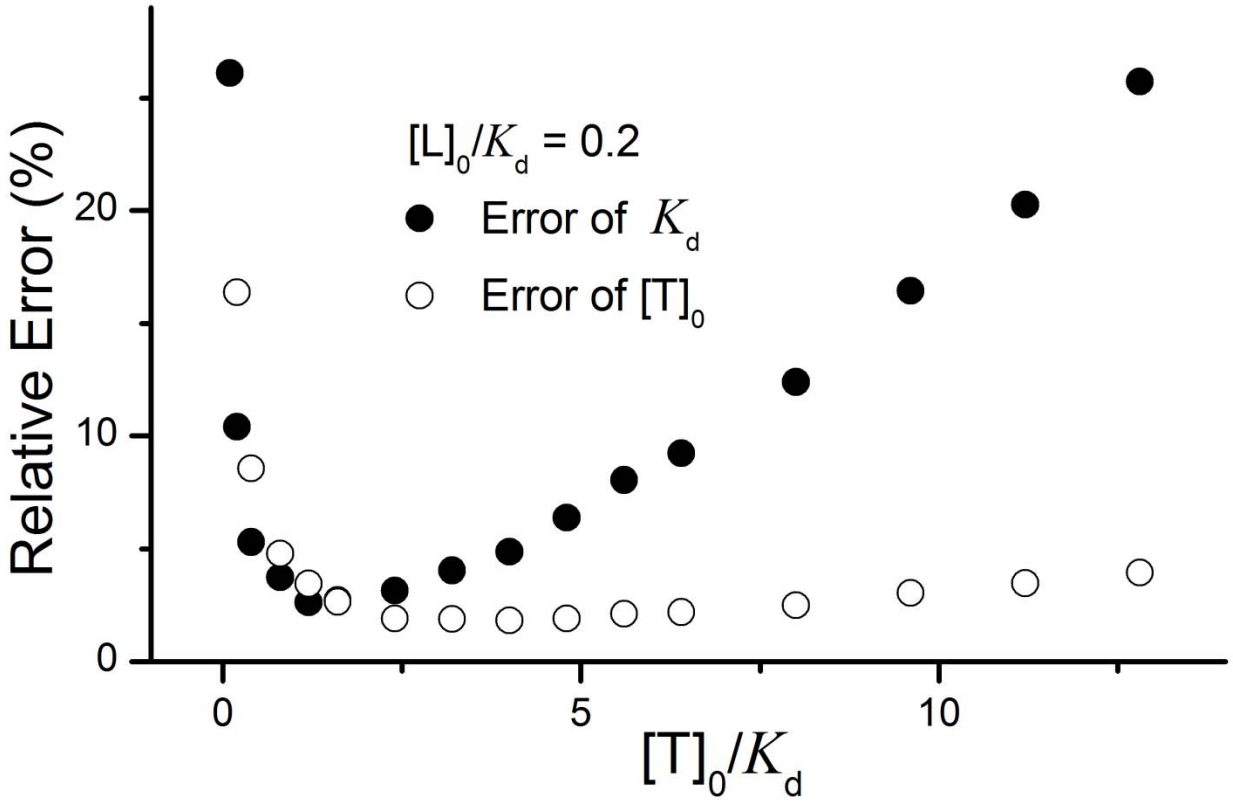

Figure 2.5. General trends in the accuracy of determination of K_d and $[T]_0$. Accuracy of determination of K_d and $[T]_0$ estimated *in silico* at dimensionless ligand concentration $[L]_0/K_d = 0.2$. Closed and open circles represent relative errors of K_d and $[T]_0$, respectively.

2.3.4. Experimental determination of K_d and $[T]_0$

As an experimental model, I used a pair of binding molecules: MutS protein from *T. aquaticus* and a previously selected DNA aptamer **[49]**. I assumed that the concentration of the stock solution of MutS was unknown. MutS was, thus, the target and the DNA aptamer was the ligand in our experiments.

I have first performed the experiment using a commercial instrument, with an LOQ of 5 nM of the aptamer. In the first set of experiments, the value of a was determined to be 30 (**Figure 2.6A**). In the second set of NECEEM experiments, I used $[T]_0 = [T]_{stock}/30$ and $[L]_0$ varying from $[L]_{0,first} = 5$ nM to $[L]_{0,last} = 40$ nM, with 6 points in between. The experimental points from the second step resulted in the best fit with the theoretical dependence of R versus $[L]_0$ at $K_d = 650$ pM and $[T]_0 = 6.7$ nM, meaning that $[T]_{stock} = 202$ nM (**Figure 2.6B**). As the resultant K_d is approximately 13 times smaller than the value of $[L]_{0,first}$, I concluded that a significant

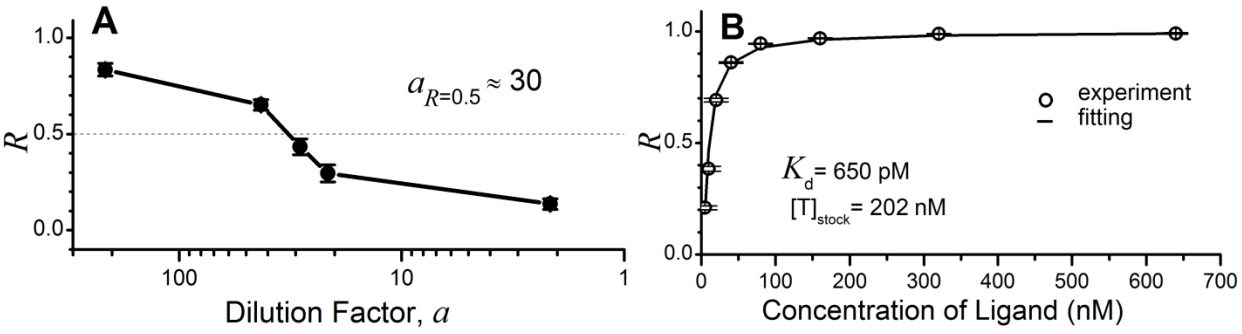

Figure 2.6. Determination of K_d and $[T]_0$ from data obtained with a commercial instrument.Panel A: Estimation of optimal dilution factor a from the dependence of R (fraction of unbound L) on a. **Panel B:** Dependences of R on concentration $[L]_0$, at constant concentration $[T]_0$ obtained from the stock solution by diluting with factors $a = 30$. Open circles represent calculated R values, solid curve is a fitting with **Eq. 16**.

systematic error may be embedded into these results. This prompted me to repeat the study with a more sensitive instrument.

The second attempt at the measurements was performed with a more sensitive custom-made CE instrument, which has an LOQ of 100 pM of the aptamer. In the first set of experiments, I determined dilution of the MutS stock which yielded in $R \approx 0.5$. $[L]_{0,first}$ was set at the LOQ of the instrument (signal-to-noise ratio of 100). Various dilutions of the protein were used to obtain a set of NECEEM electropherograms, and the values of R were plotted as a function of dilution factors used to prepare the protein solutions from the stock solution (**Figure 2.7A**). The $a = 1000$ was chosen as being the closest to $R \approx 0.5$ and was used in the second set of experiments.

In the second set of NECEEM experiments, I used $[T]_0 = [T]_{stock}/1000$ and $[L]_0$ varying from $[L]_{0,first} = 100$ pM to $[L]_{0,last} = 4$ nM, with 6 points in between. The values of R were calculated and plotted as a function of $[L]_0$ (**Figure 2.7B**). The experimental points were fitted with the theoretical dependence of R versus $[L]_0$ presented by **Eq. 16**. The best fit was obtained at $K_d = 123$ pM and $[T]_0 = 213$ pM, meaning that $[T]_{stock} = 213\ \text{pM} \times 1000 = 213$ nM. The experimental variability in measuring R translates into errors in the parameters obtained by fitting, and can be estimated by constructing a covariance matrix based on each of the parameters **[87]**. I determined the relative errors to be 9% and 4% for K_d and $[T]_0$, respectively. As the resultant K_d is larger than the value of $[L]_{0,first}$ it can be concluded that the experiments were performed at a proper range of concentrations

To assess the accuracy of the method in determination of $[T]_0$, I estimated the concentration of the MutS stock solution by the BCA assay. The spectroscopically measured value of $[T]_{stock} = 226 \pm 13$ nM was in good agreement with the NECEEM-measured values of 213 and

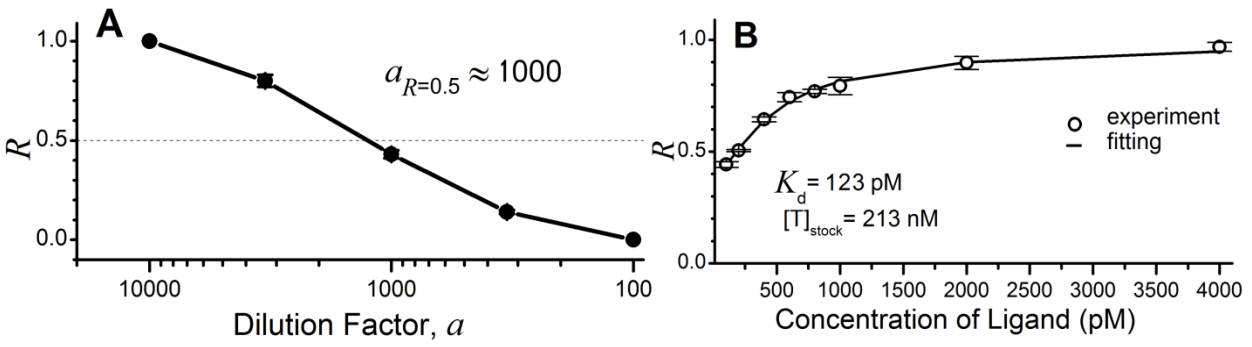

Figure 2.7. Determination of K_d and $[T]_0$ from data obtained with a custom-made instrument.Panel A: Estimation of optimal dilution factor a from the dependence of R (fraction of unbound L) on a. **Panel B:** Dependences of R on concentration $[L]_0$, at constant concentration $[T]_0$ obtained from the stock solution by diluting with factors $a = 1000$. Open circles represent calculated R-values, solid curve is a fitting with **Eq. 16**.

202 nM, with the custom-made and commercial instruments, respectively. As for K_d measurements, a previous study of the same interaction, conducted with a single-point NECEEM approach by using a commercial CE instrument with poorer LOQ, resulted in $K_d = 15$ nM **[49]**, which is significantly larger than $K_d = 123$ pM obtained with the method presented in this work. The previous K_d measurement was performed as a part of a screening of a large number of aptamer clones, and thus, the $[L]_0$ and $[T]_0$ concentrations were not optimized as rigorously as in this study. The discrepancy between the values, most likely, comes from the fact that in the previous study nanomolar concentrations of $[L]_0$ and $[T]_0$ were used to screen all of the aptamer clones, resulting in large errors for some of the measurements. Nevertheless, the single-point NECEEM method, used previously, did not provide us with any means of assessing the accuracy of the results. In the current approach, however, an insufficient LOQ of the instrument used was evident from the produced results. This suggests that the current multi-point approach is much more suitable for affinity measurements which require a high degree of accuracy and precision.

2.4. CONCLUSIONS

I have described a NECEEM-based approach for simultaneous determination of K_d and target concentration, $[T]_0$, which improves the accuracy of NECEEM experiments by eliminating the systematic errors due to uncertainties in $[T]_{stock}$. The *in silico* validation showed that the approach (mathematics, algorithms, and software) has above 90% accuracy over a wide range of initial conditions. The accuracy of the experimental implementation of this approach will, thus, be mainly affected by the quality of NECEEM electropherograms, and accuracy of ligand concentration, $[L]_0$, used in calculations. As any approach including non-linear regression, our method requires a reasonable judgment of the experimenter about the quality of experimental data to ensure the reliability of the obtained values of K_d and $[T]_0$.

Chapter 3. Systematic Approach to Optimization of NECEEM Experimental Conditions

The presented material was published previously and reprinted with permission from "Kanoatov, M.; Mehrabanfar, S.; Krylov, S.N. Systematic approach to optimization of experimental conditions in nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry*, 2016, 88, 9300–9308". Copyright 2016 American Chemical Society.

My contribution to the article was: (i) formulating the original idea and plan of development, (ii) developing a major part of the algorithm, (iii) developing the software implementation (iv) performing all of the *in silico* and *in vitro* experiments, (v) writing the manuscript.

3.1. INTRODUCTION TO NECEEM OPTIMIZATION

As it was illustrated in the previous Chapter, the choice of initial concentrations of the interacting components can have a great influence on the accuracy of the produced results. It is widely appreciated that proper design and optimization of all experimental conditions of kinetic measurements is necessary to ensure accuracy and precision **[88]**. Experimental conditions (*e.g.*, analyte concentrations, time scale of measurement) must be chosen with comprehension and care in order to avoid introducing large systematic errors into the measurements. The need for optimization stems from the fact that molecular interactions underlying biological processes are highly diverse in their properties, a point which was discussed in **Section 1.1.2**. Experimental conditions must be selected such that the sensitive range of a given method matches the magnitudes of the measured parameters. Over the years, a variety of strategies and guidelines have been developed for design and optimization of SPR **[28]**, BLI **[89]**, and stopped-flow **[90]** experiments. For each of these methods, the optimization usually occurs in a one-factor-at-a-time fashion, due to the independent nature of different experimental parameters. Optimization of

NECEEM, however, cannot proceed in such a linear fashion, due to the complex interrelation between its experimental parameters, adjustment of which often produces counteracting effects. As an example, to perform NECEEM measurements in high-conductivity electrolytes (*e.g.,* physiological buffers) the users are required to decrease the applied electric field strength to avoid excessive Joule heating; decreasing the electric field strength, in turn, results in longer analysis times; prolonged analysis may invalidate kinetic NECEEM measurements by rendering parts of electropherograms (*e.g.* the intermolecular complex peak) undetectable. Performing such adjustments properly often requires an effort of a well-trained specialist, with a comprehensive understanding of the underlying physical and chemical phenomena. Without proper guidance, the optimization of NECEEM can become a prolonged task, significantly undermining the processivity of the method. What further aggravates the issue is that improperly optimized NECEEM experiments are often difficult to recognize, which degrades the resilience of the method against errors.

To alleviate this issue, a systematic approach for optimization of NECEEM experimental conditions is required. In this Chapter, I describe such an approach, which takes the complex interrelation between NECEEM experimental conditions into account, and deals with them in a step-wise manner. Our optimization "algorithm" allows the obtained results to be validated and provides objective recommendations on improving experimental conditions when the validation fails. I have implemented the developed algorithm in a form of user-friendly software that is able to automatically process experimental data. I demonstrate the practical utility of this "expert system" by applying it to an *in vitro* NECEEM experiment. The developed optimization approach allows for the application of a wider range of experimental conditions in NECEEM,

making this method more reliable, robust and amenable for practical use by researchers with varying experience in CE.

3.2. EXPERIMENTAL SECTION

3.2.1. Materials

All chemicals and buffer components were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Fused-silica capillaries were purchased from Molex (Phoenix, AZ). *T. aquaticus* MutS recombinant protein was expressed and purified as described in **Section 2.2.1 on page 27**. The truncated version of the MutS aptamer (variant 5) was designed by Prof. Philip E. Johnson (York University), based on a DNA aptamer previously selected (clone 2-06) **[49]**. The aptamer variant was custom synthesized by Integrated DNA Technologies (Coralville, IA). The nucleotide sequence of the fluorescein labeled, ssDNA aptamer was: 5'- Fluorescein - GCC CGC CTC CTT CCT GGT AAA GTC ATT AAT AGG TGT GGG GTT TCG GAG ACG AGA TAG GCGG - 3'.

3.2.2. NECEEM

All NECEEM experiments were carried out using CESI-8000 instrument (Sciex, Concord, ON), equipped with a standard LIF detection system (488 nm excitation, 520 nm emission). Runs were performed in an uncoated fused-silica capillary, with an inner radius of 10 μm and an outer radius of 180 μm. The total length of the capillary was 30 cm, with the detection window placed 20 cm from the inlet.

Dilutions of all sample components were prepared with the electrophoresis run buffer: 50 mM Tris-HCl pH 7.4, 20 mM NaCl. Sample mixtures were incubated at room temperature for 20 min prior to analysis in order to achieve equilibration in the binding reaction. Prior to every run, the capillary was rinsed with the run buffer at 60.0 psi (414 kPa) for 3 min (to pump 10 capillary

volumes). At the end of each run, the capillary was rinsed with a succession of 100 mM HCl, 100 mM NaOH, and deionized water, at the same pressure/time settings. The samples were injected into the capillary, prefilled with the run buffer, by a 1.0 psi (6.9 kPa) pressure for 16 s to yield a 5 mm-long sample plug. Prior to applying voltage, the sample mixture was propagated 4 cm through the uncooled portion of the capillary by 2.0 psi (13.8 kPa) pressure for 1 min.

A total of 4 types of experiments were performed: 3 preliminary and 1 optimized. Each preliminary experiment was performed only once, while the optimized experiment was repeated 3 times. For all of the preliminary experiments the sample mixture was injected from the inlet end of the capillary (20 cm distance to detector) and electrophoresis was carried out with the anode at the inlet, at a constant voltage of 25.0 kV. The coolant temperature was set to 15 °C. The first preliminary experiment aimed at estimating the mobility (μ_L) and the response factor (γ_L) of the ligand, and was performed without the addition of the target. The concentration of the aptamer used in this experiment was 10 nM. This experiment also allowed me to estimate the resistivity of the run buffer (ρ) and the average amplitude of the background noise (σ) in the obtained signal. The second preliminary experiment aimed at obtaining the mobility (μ_C) and the response factor (γ_C) of the complex, and was performed at the highest possible concentration of the target to result in binding saturation. The concentration of the aptamer was 10 nM, and the concentration of the protein was 4.5 μM. The third preliminary experiment aimed at estimating the K_d and the k_{off} values, and was performed according to the recommendations from **Section 3.3.5 on page 63**. The concentration of the aptamer was 300 nM, and the concentration of the protein was 1.5 μM. The last of the preliminary experiments was subjected to validation and optimization using the further-described algorithm, and the optimized set of experiments was performed using the algorithm-provided conditions. The sample mixture was injected from the

outlet end of the capillary (10 cm distance to detector) and electrophoresis was carried out with the anode at the outlet, at a constant voltage of 2.0 kV. The coolant temperature was set to 24 °C. In these experiments, the concentration of the aptamer was 60 nM, while the concentration of the protein was 300 nM.

3.3. RESULTS AND DISCUSSION

3.3.1. Defining the Parameters and the Variables

NECEEM can be used for both analytical (*e.g.,* determination of equilibrium and kinetic rate constants) and preparative purposes (*e.g.,* affinity purification, selection of aptamers) **[91-93]**. To adapt NECEEM for the intended application (analytical or preparative), the users are free to decide on the values of four experimental *parameters*: **(1)** composition of the background electrolyte (BGE), **(2)** intended temperature in the efficiently cooled portion of the capillary ($T_{ef.goal}$), **(3)** internal radius of the capillary (r), and **(4)** width of the sample injection zone (w). The choice of BGE composition, which includes the electrolyte's ionic strength, concentration of the buffering components, pH, and presence of cofactors and stabilizing agents, should imitate the environment in which the intermolecular interaction usually takes place. Similarly, the chosen in-capillary temperature should reflect either the natural conditions of the given interaction (*e.g.*, core temperature of the host organism), or the conditions of intended use (*e.g.*, room temperature for affinity probe-based detection kits). The choice of a capillary radius is dictated by the aims of the application, with smaller radii being preferential in analytical studies to reduce sample consumption, while larger radii being preferential in preparative applications to increase processivity. Larger capillary radii may also be chosen to improve the system's LOQ, as

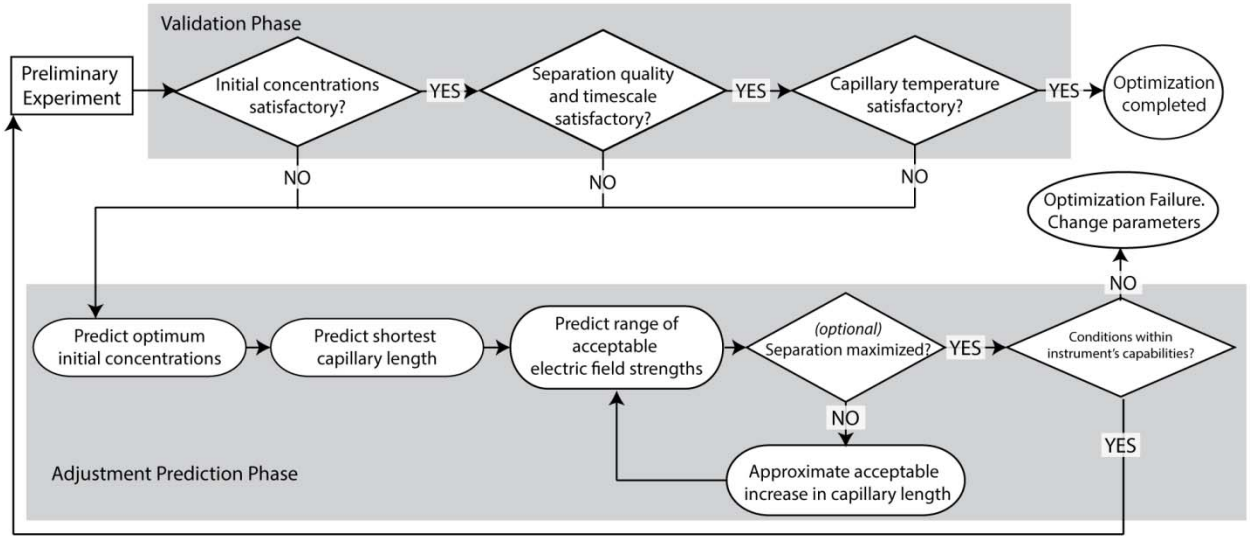

Figure 3.1. Schematic diagram of the optimization algorithm. Information from a preliminary experiment is first subjected to a validation test. If the validation fails, necessary adjustments to experimental variables are predicted based on values from the preliminary experiment. If the predicted values are within capabilities of the CE instrument, another experiment is performed using the newly defined conditions. This refinement of preliminary data continues until the criteria of the experiment validation test are satisfied.

wider cross-sections enable higher analyte flux, and thus, yield a stronger signal. On the other hand, capillaries with smaller inner radii are better at dissipating heat, and are, thus, more suitable for experiments that use high-conductivity run buffers. The width of the sample injection zone has an impact on the separation efficiency of CE, with narrower injection zones used when the resolution between the species is a priority. As the length of the injection zone defines the volume of the sample (along with the capillary radius), wider sample injection plugs are also used to improve method's processivity.

The user-selected values of the four experimental *parameters* will have an effect on the optimal values of three experimental *variables*: **(1)** initial analyte concentrations, $[L]_0$ and $[T]_0$, **(2)** length of capillary to detector, l_{det}, and **(3)** strength of the applied electric field, E. The

purpose of the optimization procedure is to determine the combination of values of the three experimental variables which ensure accurate and bias-free NECEEM analysis.

The optimization is performed in a succession of refining rounds, where the apparent values of k_{off} and K_d (denoted by k_{off}^* and K_d^*, respectively) are measured and subjected to an experiment validation test. If the experiment validation fails, the apparent values are rejected, and the experiment is performed again, with the experimental conditions rationally adjusted to specifically address the shortcomings revealed by the validation test. Once the experiment validation passes, the apparent values are accepted as accurate. The optimization algorithm, thus, consists of two phases: validation of the experiment, followed by generation of recommendations for adjustment of conditions. A schematic overview of the algorithm is presented in **Figure 3.1**.

3.3.2. Phase 1: Experiment validation

Upon deciding on the values of the experimental parameters, and performing a preliminary experiment at a non-optimized set of conditions (the best way to choose conditions for an initial experiment is described in **Section 3.3.5 on page 63**), the users must validate the experiment based on the following criteria: (i) suitability of $[L]_0$ and $[T]_0$, (ii) suitability of separation quality and timescale, and (iii) suitability of in-capillary temperatures.

As described **in Section 2.3.2 on page 30**, a small experimental error in the measurement of K_d can become greatly amplified if $[T]_0$ and $[L]_0$ are improperly chosen. Since the variability in experimental measurement (ΔR) may differ widely between instruments and experimental systems, it is difficult to define an error threshold which will be universally applicable. Thus, optimization criteria should be based on general properties of curves in **Figure 2.1**, such as their inflection points and minima. As a result, the most accurate determination of K_d requires the measurements to be performed at $[L]_0$ that is smaller or equal to $2K_d$ (**Figure 2.1A**), and that $[T]_0$

is chosen so that $R = 0.5$ (**Figure 2.1B**), a concentration that can be calculated by rearranging **Eq. 12**. Thus, values of $[\mathrm{L}]_0$ and $[\mathrm{T}]_0$ can be considered as suitable if they satisfy the following conditions:

$$[\mathrm{L}]_0 \le 2K_\mathrm{d}^* \tag{18}$$

$$(K_\mathrm{d}^* + 0.5[\mathrm{L}]_0) \le [\mathrm{T}]_0 \le 2K_\mathrm{d}^* \tag{19}$$

The extent and the rate of separation between L and C play an important role in our ability to accurately extract information from NECEEM electropherograms. A useful measure of separation efficiency in CE is the characteristic time of separation, t_{sep}, which is the time required for the initially superimposed zones of L and C to completely separate from each other:

$$t_{\mathrm{sep}} = \frac{w}{\Delta\mu E} \tag{20}$$

By dividing the total analysis time, t_{run} (the time required for the slowest detectable component to reach the detector), by t_{sep} we can calculate the number of separation events, S, which took place during a CE experiment:

$$t_{\mathrm{run}} = \frac{l_{\mathrm{sep}}}{\min(\mu_\mathrm{L}, \mu_\mathrm{C})E} \tag{21}$$

$$S = \frac{t_{\mathrm{run}}}{t_{\mathrm{sep}}} = \frac{\Delta\mu l_{\mathrm{sep}}}{\min(\mu_\mathrm{L}, \mu_\mathrm{C})w} \tag{22}$$

S is a parameter similar to chromatographic resolution, with the exception that it neglects zone broadening. Since zone broadening is difficult to predict in a generic way, S is a more amenable measure for a predictive optimization algorithm. For S to be an accurate predictor of separation quality, however, zone broadening effects must be minimized by matching the sample and run buffers (to avoid electrodispersive phenomena) and by avoiding excessively long experiments, in

which the effects of longitudinal diffusion become non-negligible (usually longer than 5 h for macromolecules). Insufficient separation between L and C peaks will obscure the decay bridge, thus, a required minimum number of separation events must occur during an experiment to facilitate accurate extraction of A_L, A_C, and A_D. This minimum is defined by the capabilities of the area deconvolution procedure used in data analysis, with NAAP being the most efficient tool for processing NECEEM electropherograms at this time **[45]**. The best way to study the capabilities of the area deconvolution procedure is to apply it to a set of computer simulated NECEEM electropherograms, extract the K_d and k_{off}, and compare them to the original values predefined in the simulations. For this purpose, I have used the same *in silico* NECEEM modeling tool which I have described in **Section 2.3.3 on page 34**. To find the minimum S required for accurate area analysis, I have simulated a set of NECEEM experiments with a range of S values (by varying l_{sep}), but with otherwise constant and optimized conditions, and applied the area deconvolution approach to them. **Figure 3.2** shows that accurate determination of both K_d and k_{off} requires a minimum value of $S = 5$.

On the other hand, unnecessarily long separation, characterized by a large t_{run}, may lead to a scenario where the complex peak becomes too small for accurate quantitation, *i.e.,* when its height becomes smaller than the LOQ of the instrument, usually defined as 10-times the amplitude of the background noise, σ **[38, 94]**. From **Eq.10**, the time that it takes for the C peak to dissociate to this level, t_{dis}, can be calculated:

$$t_{dis} = \frac{\ln\left(\frac{[C]_{eq}\gamma_C}{10\sigma}\right)}{k_{off}} \qquad (23)$$

Thus, to ensure proper quantitation of intact C, t_{run} may not exceed t_{dis}

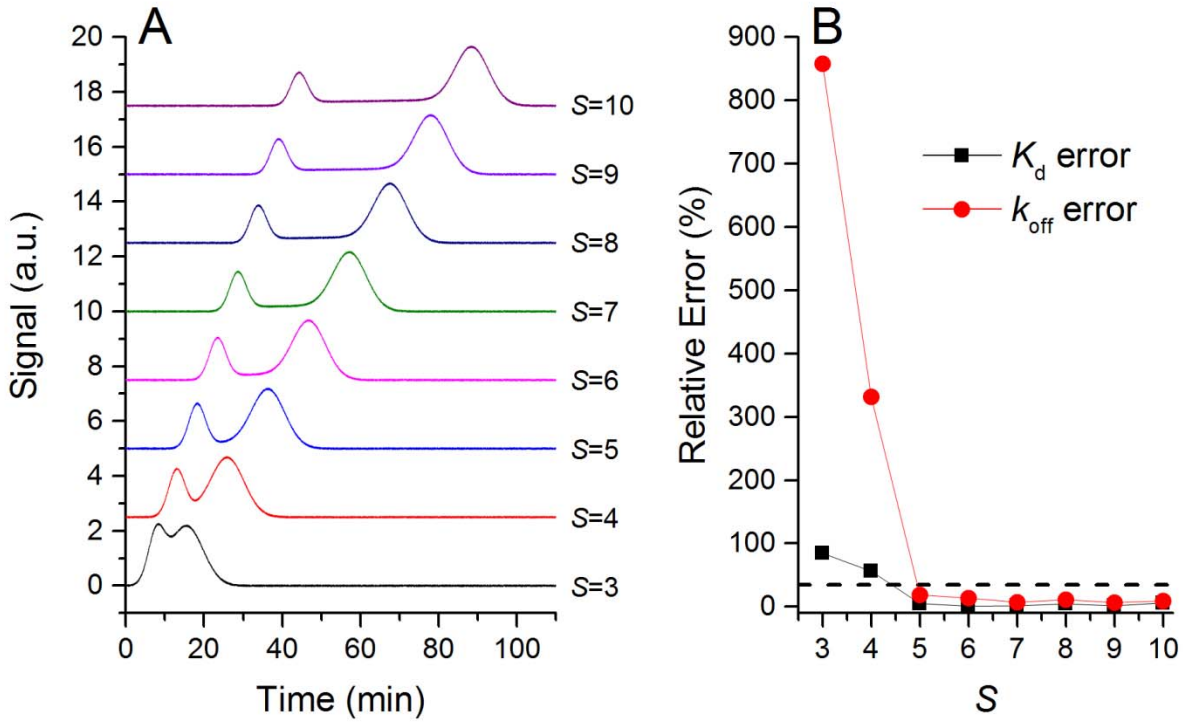

Figure 3.2. Minimum value of *S* required for accurate deconvolution of peak areas. Panel A: a set of simulated NECEEM electropherograms was generated at increasing values of *S*. For the simulations, all of the conditions were fixed, except for an increasing value of l_{det}. **Panel B:** Relative error in extracting K_d and k_{off} depending on the value of *S*. By setting a maximum error threshold at 20%, we can establish the minimum value $S = 5$.

Lastly, the prominence of the decay bridge relative to the other features of an electropherogram influences the accuracy with which the overlapping areas can be deconvoluted. The relative size of the decay bridge is defined by τ, the ratio of two characteristic times: t_{sep}, and the characteristic time of equilibration, t_{eq}:

$$t_{eq} = \frac{1}{(k_{on}[\mathrm{T}] + k_{off})} \tag{24}$$

$$\tau = \frac{t_{sep}}{t_{eq}} = \frac{w(k_{on}[\mathrm{T}]_{eq} + k_{off})}{E\Delta\mu} \tag{25}$$

where [T] is the local concentration of T at a given point in the capillary, which is usually substituted by its maximum value of $[T]_{eq}$. Small values of τ will lead to an undetectable decay bridge, which will prevent accurate determination of k_{off}, while large values will cause the decay bridge to obscure the peaks, which prevents accurate determination of K_d. Thus, a valid experiment must be characterized by an intermediate range of τ values, the boundaries of which can be determined by an *in silico* study. I have simulated a set of NECEEM electropherograms with a range of τ values (by varying k_{off} and k_{on}), but with otherwise optimized conditions, and used them to test the accuracy of both k_{off} and K_d measurements (**Figure 3.3**). By defining a relative systematic error threshold at 20%, it can be established that the optimum range of τ values to lies between 0.15 and 0.6. It should be noted that all of the error thresholds described in this Chapter may be re-defined to satisfy specific requirements of a given study. The recommended error thresholds only serve the purpose of making the algorithm applicable to a general case.

To summarize, for NECEEM separation conditions to be considered as suitable, they must satisfy the following conditions:

$$S \geq 5 \tag{26}$$

$$t_{run} \leq t_{dis} \tag{27}$$

$$0.15 \leq \tau \leq 0.6 \tag{28}$$

It should be noted, that for studies which aim at measuring only the value of K_d, and not k_{off}, a smaller τ is preferable, so that the overlaps between the decay bridge and the peaks are minimal. For such studies, condition 28 may be substituted by the following:

$$\tau \leq 0.6 \tag{29}$$

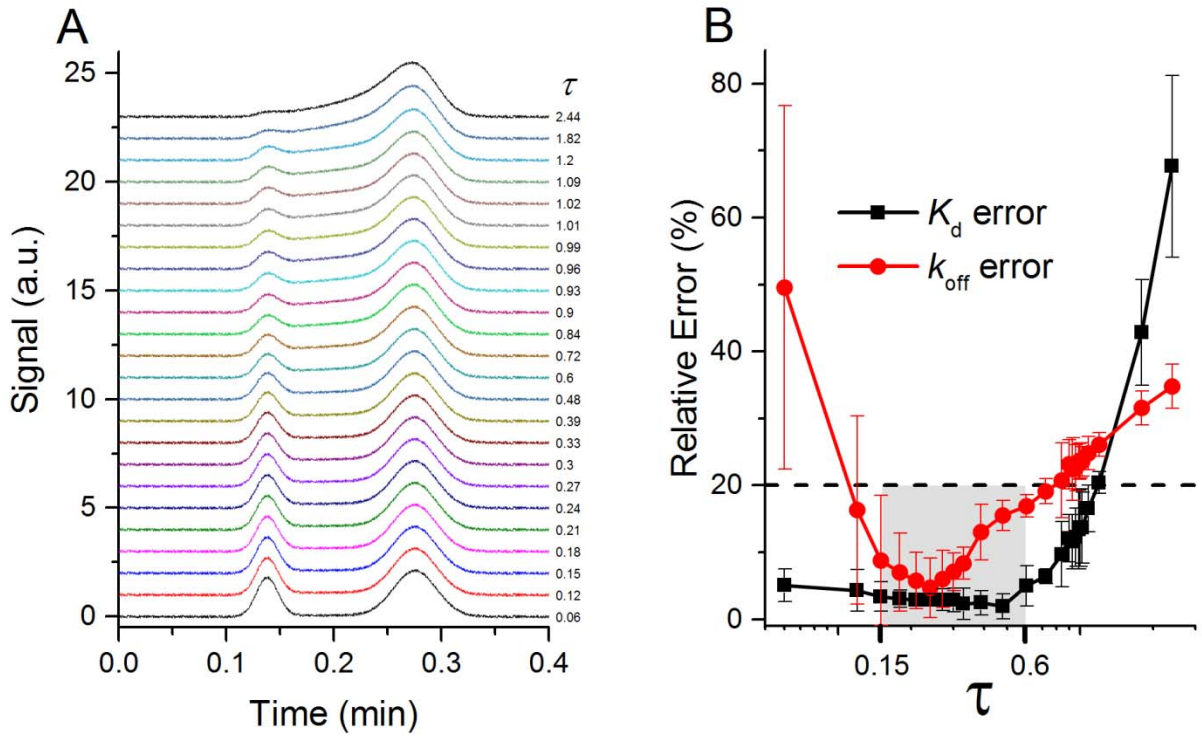

Figure 3.3. Range of values of τ for accurate deconvolution of peak areas. Panel A: a set of simulated NECEEM electropherograms at increasing values of τ. For the simulations, all of the conditions were fixed, except the values of k_{off} and k_{on}. **Panel B:** relative error in extracting K_d and k_{off} dependent on the value of τ. Due to the fact that random noise played a significant role in the magnitude of the error, each electropherogram was regenerated 5 times, with new values of the noise. The markers on the graph show the average value of the error at a given τ value, while the error bars show one standard deviation of the error based on 5 measurements. By setting a maximum error threshold at 20%, we can establish the optimum range of τ values to lie between 0.15 and 0.6.

Temperatures in both efficiently and inefficiently cooled portions of the capillary can be calculated using the simplified universal method for determining electrolyte temperatures (SUMET) **[54]**.

$$T_{ef} = T_{cool} + \frac{c(EI_{av})^{n+1}}{g + (EI_{av})^{n}} \qquad (30)$$

$$T_{\text{inef}} = T_{\text{amb}} + kb(EI_{\text{av}})^2 \quad (31)$$

where T_{cool} is the temperature of the coolant, T_{amb} is the ambient temperature surrounding the inefficiently cooled portions of the capillary, I_{av} is the average electric current during electrophoresis, and b, c, g, k, and n are empirically determined constants defined by r **[53]**.

For the T_{ef} to be considered as valid it must be equal to the chosen value of $T_{\text{ef.goal}}$. As long as the sample mixture is propagated through the inefficiently cooled portion of the capillary, the value of T_{inef} does not have a significant effect on the accuracy of the experiment and is only limited by the boiling temperature of the electrolyte, T_{boil}. However, depending on the specifics of the study, the users may define a more conservative limit of T_{inef} based on some other considerations (*e.g.*, maximum allowable change in BGE viscosity). Thus, in-capillary temperatures can be considered as suitable if they satisfy the following conditions:

$$T_{\text{ef}} = T_{\text{ef.goal}} \quad (32)$$

$$T_{\text{inef}} < T_{boil} \quad (33)$$

If an experiment satisfies conditions described in **Eqs. 18, 19, 26-28, 32** and **33** it can be considered as valid, and the values of K_{d}^{*} and k_{off}^{*} as accurate; otherwise, more suitable experimental conditions should be found, using the steps described below.

3.3.3. Phase 2: Recommendations for adjustment of experimental conditions

To generate practicable recommendations for adjusting experimental conditions, the limitations of the employed CE instrument must be taken into account. Table 3.1 lists the information regarding the experimental system, which needs to be ascertained by the user.

Table 3.1. Required information about the limitations of the CE instrument

Parameter	Description
$V_{instr.min}$	Minimum electrophoresis voltage supplied by the instrument
$V_{instr.max}$	Maximum electrophoresis voltage supplied by the instrument
$I_{instr.max}$	Maximum electrophoresis current supplied by the instrument
$l_{det.instr.min}$	Minimum capillary length to detector
$l_{det.instr.max}$	Maximum capillary length to detector
$T_{cool.min}$	Minimum coolant temperature
$t_{run.max}$	Maximum analysis time[1,2]
σ	Average amplitude of noise at baseline

[1] – usually limited by data-acquisition time, or buffer depletion time
[2] – the minimum time of analysis is subsumed by *S* and, thus, does not need to be considered separately

In some cases, it may be impossible to generate recommendations which fall within the instrument's limitations. In these cases, the optimization algorithm aborts with a "failure" output and provides a message on how the overall design of the experiment must be modified to facilitate the measurement. Recommendations for adjusting experimental variables are generated in 4 steps: (1) prediction of optimum $[L]_0$ and $[T]_0$, (2) prediction of minimum l_{det}; (3) prediction of optimum range of values of *E*; and (4) iterative maximization of *S*.

Step 1. The choice of $[L]_0$ and $[T]_0$ should aim to satisfy conditions 18 and 19. The recommended value of $[L]_0$ must be smaller or equal to $2K_d^*$, but this difference should not exceed 10-fold, as this would lead to an unnecessarily small $[C]_{eq}$. Often, the choice of $[L]_0$ is limited by the LOQ of the instrument. Using an *in silico* study, I have determined that $[L]_0$ must be larger than the concentration of free L that results in a peak with a height 100 times larger than the amplitude of background noise, σ (**Figure 3.4**).

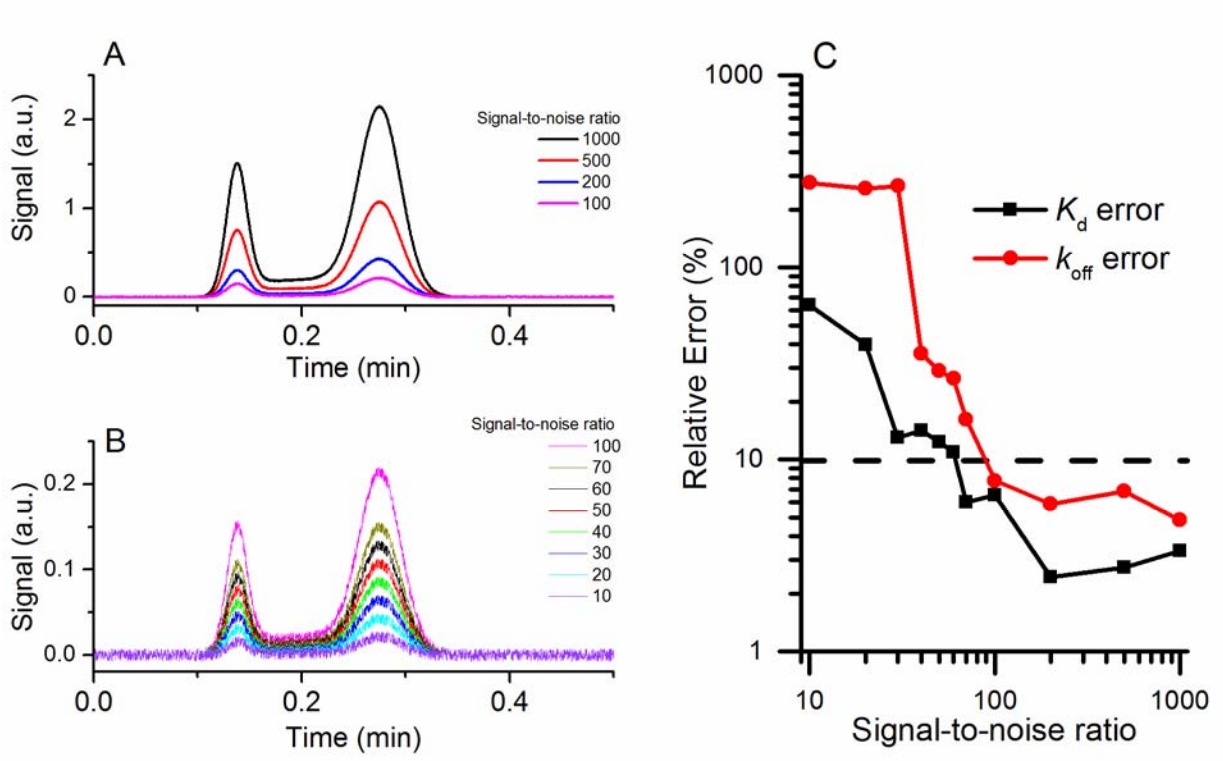

Figure 3.4. Minimum signal-to-noise ratio required for accurate measurement of peak areas. Panel A: a set of simulated NECEEM electropherograms at decreasing signal-to-noise ratios ranging between 1000 and 100. **Panel B:** a set of simulated NECEEM electropherograms at decreasing signal-to-noise ratios ranging between 100 and 10. For the simulations, all of the conditions were fixed, except the values of γ_C and γ_L (which were always equal to each other). **Panel C:** Relative error in extracting K_d and k_{off} dependent on the signal-to-noise ratio. By setting a maximum error threshold at 10%, we can establish the minimum value of signal-to-noise ratio at 100. Signal-to-noise ratios are based on the height of the peak produced by a given concentration of ligand in absence of the target.

To put all of the above conditions for $[L]_0$ together:

$$[L]_0 = \begin{cases} 0.2K_d^* & \text{if } \dfrac{100\sigma}{\gamma_L} \le 0.2K_d^* \\ \dfrac{100\sigma}{\gamma_L} & \text{if } 0.2K_d^* \le \dfrac{100\sigma}{\gamma_L} \le 2K_d^* \end{cases} \tag{34}$$

Optimization of $[L]_0$ fails if $(100\sigma/\gamma_L) > 2K_d^*$, providing a suggestion to improve the LOQ of the system (*e.g.,* by increasing r). Once the value of $[L]_0$ is determined, the optimum value of $[T]_0$ can be calculated:

$$[T]_0 = K_d^* + 0.5[L]_0 \tag{35}$$

Step 2. The choice of l_{det} should aim to satisfy **condition 26**. The shortest l_{sep} that satisfies $S \ge 5$, $l_{sep.min}$, can be calculated by rearranging **Eq. 22**, which can then be substituted into **Eq. 4** to calculate the corresponding distance to the detector, $l_{det.min}$:

$$l_{sep.min} = \frac{5\min(\mu_L, \mu_C)w}{\Delta\mu} \tag{36}$$

$$l_{det.min} = l_{sep.min} + l_{prop} \tag{37}$$

Between $l_{det.min}$ and $l_{det.instr.min}$, the larger value is used as the recommendation:

$$l_{det} = \max\left(l_{det.min}, l_{det.instr.min}\right) \tag{38}$$

Optimization of l_{det} fails if $l_{det.min} > l_{det.instr.max}$, with a suggestion to improve $\Delta\mu$ by modifying separation conditions (*e.g.,* by choosing a BGE that results in a slower EOF).

Step 3. Once l_{det} is defined, the range of optimum values of E can be determined. The choice of E should satisfy all of the remaining conditions, presented in **Eqs. 27, 28, 30,** and **31**. The maximum acceptable E is limited by 4 parameters: $V_{instr.max}$, $I_{instr.max}$, $T_{cool.min,}$ and τ:

$$E_{\max.\mathrm{V}} = \frac{V_{\mathrm{instr.max}}}{l_{\mathrm{det}} + l_{\mathrm{out}}} \tag{39}$$

$$E_{\max.\mathrm{I}} = \frac{\rho I_{\mathrm{instr.max}}}{\pi r^2} \tag{40}$$

$$E_{\max.\mathrm{Tef}} = \frac{-m_1 + \sqrt{m_1^{\,2} - 4m_2(m_0 - T_{\mathrm{ef.goal}} + T_{\mathrm{cool.min}})}}{2m_2} \tag{41}$$

$$E_{\max.\tau} = \frac{w(k_{\mathrm{on}}[\mathrm{T}]_{\mathrm{eq}} + k_{\mathrm{off}})}{0.15\Delta\mu} \tag{42}$$

where m_2, m_1, m_0 are coefficients determined by approximating the relationship between E and T_{ef} by a second-order polynomial; and ρ is the experimentally measured resistivity of the BGE, calculated as:

$$\rho = \frac{E\pi r^2}{I_{\mathrm{av}}} \tag{43}$$

The minimum acceptable E is limited by 4 parameters: $V_{\mathrm{instr.min}}$, $t_{\mathrm{run.max}}$, t_{dis}, and τ:

$$E_{\min.\mathrm{V}} = \frac{V_{\mathrm{instr.min}}}{l_{\mathrm{det}} + l_{\mathrm{out}}} \tag{44}$$

$$E_{\mathrm{min.trun}} = \frac{l_{\mathrm{sep}}}{\min(\mu_{\mathrm{C}}, \mu_{\mathrm{L}})t_{\mathrm{run.max}}} \tag{45}$$

$$E_{\mathrm{min.dis}} = \frac{l_{\mathrm{sep}}}{\mu_{\mathrm{C}} t_{\mathrm{dis}}} \tag{46}$$

$$E_{\min.\tau} = \frac{w(k_{\mathrm{on}}[\mathrm{T}]_{\mathrm{eq}} + k_{\mathrm{off}})}{0.6\Delta\mu} \tag{47}$$

The most conservative range of acceptable values of E is defined by the smallest of the maximum values and the largest of the minimum values.

$$E_{\max} = \min(E_{\max.\mathrm{V}}, E_{\max.\mathrm{I}}, E_{\max.\mathrm{Tef}}, E_{\max.\tau}) \tag{48}$$

$$E_{\min} = \max(E_{\min.V}, E_{\min.trun}, E_{\min.dis}, E_{\min.\tau}) \quad (49)$$

$$E_{range} = E_{max} - E_{min} \quad (50)$$

The interval between E_{min} and E_{max} represents the set of all acceptable values of E. Since choosing a value from this interval smaller than E_{max} will lead to an unnecessarily prolonged experiment, E_{max} is used as the recommendation. Optimization of E fails if E_{range} is smaller than zero. There is one special case: when E_{min} is defined by $E_{min.dis}$ (*i.e.*, if C becomes undetectable before it can reach the detector even at the highest E), it may be possible to find an acceptable set of conditions by shortening l_{sep}. This can be achieved by increasing l_{prop} beyond the value required to propagate the sample through the inefficiently cooled portion of the capillary. The longest l_{sep} which allows for the complex to be detected, $l_{sep.dis}$, as well as the adjusted l_{prop} can be calculated as follows:

$$l_{sep.dis} = t_{dis} \min(\mu_L, \mu_C) E_{max} \quad (51)$$

$$l_{prop} = l_{det.instr.min} - l_{sep.dis} \quad (52)$$

The resulting separation quality must still satisfy **condition 26**, thus, the algorithm fails if the calculated $l_{sep.dis} < l_{sep.min}$. Optimization of E also fails if the determined E_{max} leads to $T_{inef} \geq T_{boil}$, with the suggestion to decrease the generation of Joule heating (*e.g.,* by increasing the resistivity of the BGE), or by improving heat dissipation (*e.g.,* by decreasing r).

Step 4. If the user wishes to perform analysis in the shortest possible time (a common goal in analytical applications), the algorithm completes at the end of Step 3. However, if the user wishes to achieve the best possible separation, *i.e.,* to maximize S (a common goal in preparative applications), then the algorithm proceeds with iterative maximization of S, by increasing l_{sep}. Maximum l_{sep} is limited by the smaller of the values between $l_{det.instr.max}$ and $l_{sep.dis}$.

$$l_{\text{sep.max}} = \min(l_{\text{det.instr.max}}, l_{\text{sep.dis}}) \quad (53)$$

Due to a circular dependence between $l_{\text{sep.dis}}$, E_{max}, and l_{det}, the value of $l_{\text{sep.max}}$ cannot be calculated directly, but must be approached iteratively with "half-step" increments. Every iteration begins by estimating an updated value of l_{det} ($l_{\text{det.new}}$) which will yield increased separation while being limited by $l_{\text{sep.dis}}$. An estimate value of $l_{\text{sep.dis}}$ is calculated by **Eq. 51**, using the current value of E_{max}, but to avoid overshooting the goal, $l_{\text{det.new}}$ is increased by half of the difference between l_{det} value from the previous iteration ($l_{\text{det.previous}}$) and the latest estimate of $l_{\text{sep.dis}}$.

$$l_{\text{det.new}} = l_{\text{det.previous}} + 0.5(l_{\text{det.previous}} - (l_{\text{sep.dis}} + l_{\text{prop}})) \quad (54)$$

Once $l_{\text{det.new}}$ is defined, the algorithm proceeds to step 3, where a new set of values of E_{max} and E_{min} are calculated. The closer is the estimated $l_{\text{sep.max}}$ to its true value, the smaller is the interval between E_{max} and E_{min}. The iterations are repeated until E_{range} collapses to a value of 0, or until $l_{\text{det.new}} = l_{\text{det.instr.max}}$. At the end of the iterative procedure, the value of l_{det} will yield the best quality of separation (maximum S), within the limits of **condition 27**.

3.3.4. Modifications for preparative applications

For preparative applications (such as affinity purification or aptamer selection) complex detectability may not serve as a useful optimization criterion. Instead, it may be more advantageous to define the smallest allowable fraction of intact complex, f, and apply it in the calculation of t_{dis} using **Eq. 10**:

$$t_{\text{dis}} = \frac{-\ln(f)}{k_{\text{off}}} \quad (55)$$

For example, to improve the yield of affinity purification the users might decide that no more than 20% of the initial C may dissociate during an experiment, thus defining $f = 0.8$. For aptamer

selection, on the other hand, it may be useful to define f as a small fraction, and substitute the value of k_{off} in **Eq. 55** with its maximum desired value, thus defining the stringency of the selection process. It may also be advantageous to modify **conditions 18** and **19** to increase the yield of preparative applications by maximizing complex formation: $[L]_0$ may be set at the highest available concentration, while $[T]_0$ may be calculated so that $R > 0.9$ using **Eq. 12**.

3.3.5. Recommended conditions for an initial experiment

Novel studies of biomolecular interactions begin without any *a priori* knowledge of their properties. The conditions for a study-initiating experiment cannot be tailored for accurate determination of K_d and k_{off}, but instead should be aimed at minimizing the odds of producing uninformative electropherograms with indistinguishable or missing features. To facilitate subsequent optimization, the initial experiment should at the least yield estimates of analyte mobilities (μ_L and μ_C) and response factors (γ_L and γ_C), thus, it must produce an electropherogram with discernible peaks of intact C and free L. The best way to achieve this is to perform a short experiment (small t_{run}, to prevent complete dissociation of C) at the fastest possible rate of separation (small t_{sep}, to prevent excessively large decay bridge or insufficient peak separation). Thus, the initial experiment should be performed using the conditions outlined **in** Table 3.2.

Table 3.2. Recommendations for experimental conditions for a study-initiating experiment

Parameter or variable	Value
r	smallest available
$[L]_0$	$100\sigma/\gamma_L$
$[T]_0$	$5[L]_0$
l_{det}	shortest possible (usually 20-30 cm)
l_{prop}	to the edge of the cooled portion
w	shortest possible
E	maximum (voltage- or current-limited)
T_{cool}	smallest possible

If the preliminary experiment does not yield a detectable *C* peak and a decay bridge, then the $[L]_0$ and $[T]_0$ should be increased 10-fold to promote complexation. If the preliminary experiment does not yield a detectable C peak but features a prominent decay bridge, then l_{det} should be shortened to decrease the degree of complex dissociation.

3.3.6. Software implementation

All parts of the described algorithm, including the iterative step, can be easily implemented in any of the available spreadsheet software, such as Microsoft Excel, or OpenOffice Calc. My Excel implementation may be downloaded at http://www.yorku.ca/skrylov/resources.html. Besides the described algorithm, my application also includes a NAAP-based electropherogram processing subroutine **[95]**, which further reduces user involvement and makes the method more convenient and less prone to bias.

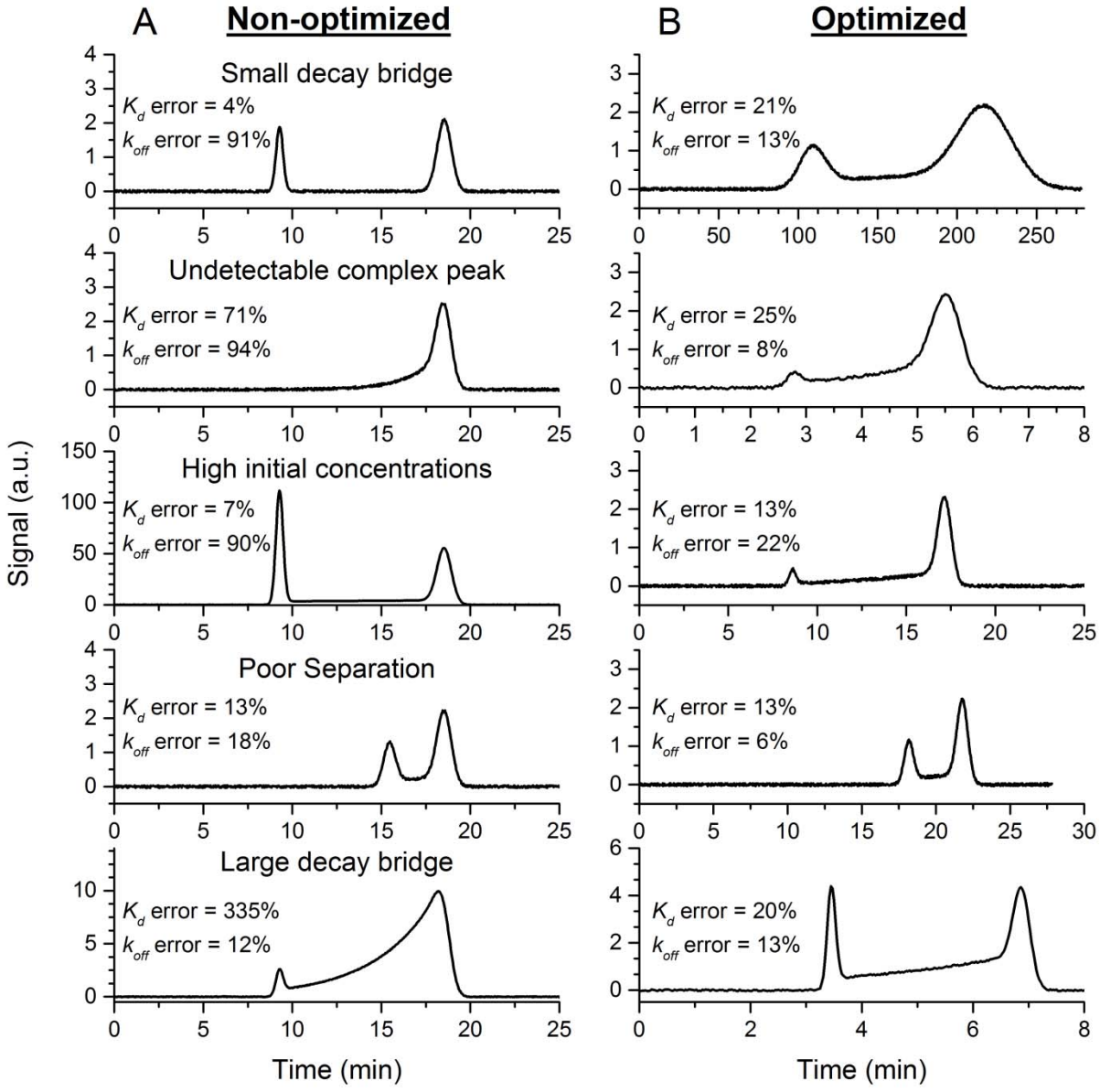

Figure 3.5. Application of the algorithm to computer simulated experiments. Electropherograms were simulated at various non-optimized conditions, representing some of the more commonly occurring problematic cases, and were subjected to optimization by the developed algorithm. **Panel A** shows electropherograms before the optimization, while **Panel B** shows the results after algorithm-guided adjustment of experimental variables.

To test the performance of our software package, I have generated a number of NECEEM electropherograms with non-optimized conditions, subjected them to automatic optimization, and then regenerated a new set of electropherograms following the recommendations from the software. Electropherograms before and after optimization are presented in **Figure 3.5**, along with the measure of accuracy in k_{off} and K_d extraction. The optimization, essentially, "standardizes" the shape of the electropherogram, ensuring that useful information can be extracted with the smallest possible error. This is confirmed by the fact that the optimized set of electropherograms consistently yields a more accurate combination of values than the non-optimized ones.

3.3.7. Practical application

To test the developed optimization algorithm within a real experimental study, I used a model system of MutS protein (target) and a truncated version of its DNA aptamer (ligand). The goal of the analytical study was to measure both K_d and k_{off} accurately. First, I defined the study parameters: (i) I decided to use the protein storage buffer, 50 mM Tris-HCl pH 7.5 with added 20 mM NaCl, as the BGE for the measurements in order to avoid any buffer mismatch phenomena; (ii) I was interested in studying the interaction at 25 °C; (iii) to minimize sample consumption, I chose to perform the measurement in a capillary with an internal radius of 10 μm; (iv) as processivity was not a priority, I used the shortest reproducible width of injection plug for the employed CE instrument, which is 5 mm. Prior to analysis of the equilibrium mixture, I have performed a CE run with only the aptamer at 10 nM, which allowed me to measure $\sigma = 4.6 \times 10^{-9}$ arbitrary units (a.u.), $\gamma_L = 1.1 \times 10^7$ a.u. M^{-1}, and $\rho = 4.1$ Ωm. I then performed another preliminary experiment, where I analyzed an equilibrium mixture prepared with

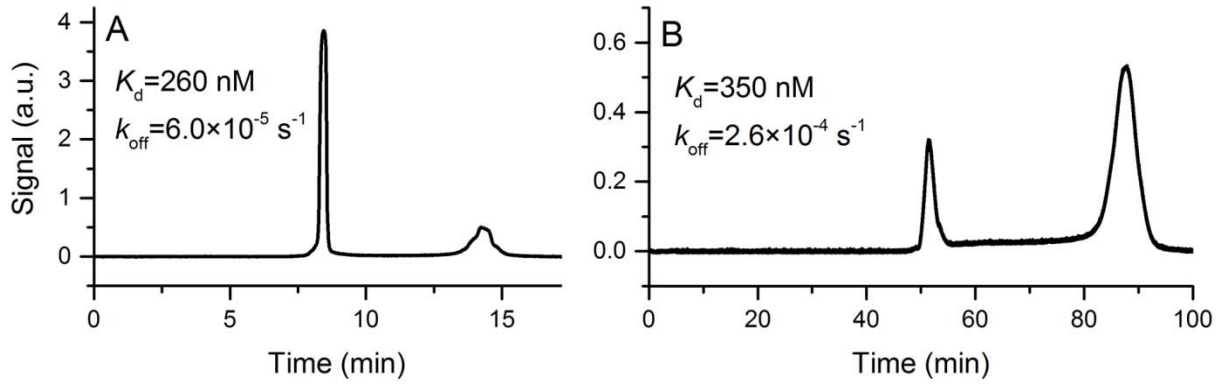

Figure 3.6. Application of the algorithm within an experimental study. NECEEM was used to study the interaction between MutS protein and a truncated aptamer (with decreased affinity). **Panel A:** electropherogram resulting from a non-optimized experiment. This electropherogram was subjected to the software implementation of the algorithm and failed the validation test. **Panel B:** recommendations from the software were applied to a new experiment, which resulted in the presented electropherogram. The electropherogram in panel B passed the validation test.

$[L]_0 \approx 100\sigma/\gamma_L$ (10 nM) and the highest possible $[T]_0$ (4.5 μM), which saturated the binding between T and L and allowed me to measure $\gamma_C = 0.9\gamma_L$. The initial analytical experiments were performed by following the recommendations described in **Section 3.3.5 on page 63**: the first experiment was performed at $[L]_0 \approx 100\sigma/\gamma_L$, but no complex peak was detected; thus, $[L]_0$ was increased 5-10 fold until a prominent complex peak was observed. A resulting electropherogram, presented in **Figure 3.6A**, contained both peaks but did not contain a discernible decay bridge. The electropherogram was subjected to validation and optimization using the algorithm and was rejected based on the unsatisfactory combination of initial concentrations and a small value for τ. The software provided the following recommendation: decrease initial concentrations to $[L]_0 = 60$ nM and $[T]_0 = 300$ nM; decrease l_{det} to 10 cm, and decrease E to 66 V cm^{-1}. These recommendations were applied to a second analytical experiment, along with the desired values for the experimental parameters. The resulting electropherogram (**Figure 3.6B**), contained all of

the expected NECEEM features at correct proportions, and when subjected to the algorithm, passed the validation test. As a result, the obtained K_d = 350 nM ± 14% and k_{off} = 2.6×10^{-4} s^{-1} ± 8% were accepted as accurate (the presented error is variability in experimental measurements based on 3 trials). This demonstrates, that the developed algorithm allows an experiment to be optimized and the results validated within only a few experimental runs.

3.4. CONCLUSIONS

KCE methods, such as NECEEM, possess a number of unique capabilities that address the needs of a wide niche of analytical applications; however, their use is still predominantly limited to academic settings, where the opportunities of an educational environment allow the operators to become highly skilled CE specialists. To promote the use of KCE in industrial and clinical settings, it must be made more accessible to less-experienced users, such that investments in operator training can be converted into productive work within the shortest possible time. I believe that the presented optimization algorithm can be an efficient tool in the hands of novice NECEEM experimenters, which will allow them to produce reliable data while their experience with the methodology grows. Gaining familiarity with the inner-workings of the optimization algorithm can by itself serve as a valuable educational tool and a basis for developing best-practices in CE experimentation. I also believe that this optimization algorithm may be an important step toward greater automation of NECEEM analysis for applications in the industry. Lastly, the presented algorithm may serve as a useful tool during the establishment of the analytical *design space* **[96]**, as part of the implementation of the *Quality by Design* principles in pharmaceutical assay development **[97]**. Other KCE methods stand to benefit from the development of similar optimization approaches.

Chapter 4. NECEEM Analysis of DNA in Physiological Conditions

The presented material was published previously and reprinted with permission from "Kanoatov, M.; Krylov, S.N. Analysis of DNA in phosphate buffered saline using kinetic capillary electrophoresis. *Analytical Chemistry*, 2016, 88, 7421–7428". Copyright 2016 American Chemical Society.

My contribution to the article was: (i) making the initial observations about the nature of the artifact, (ii) formulating the research plan, (iii) performing all of the experiments, (iv) writing the manuscript.

4.1. IMPORTANCE OF MEASUREMENTS IN PHYSIOLOGICAL CONDITIONS

Due to the remarkable charge density and the easily predictable size-to-charge ratio of its chains, electrophoresis-based methods have been especially useful in analysis and manipulation of DNA **[98]**. In particular, CE methods have been instrumental in such applications as DNA sequencing **[99]**, hybridization assays **[100, 101]**, and post-synthesis and amplification purity control **[102]**. Furthermore, in the past two decades KCE methods, NECEEM in particular, have become an attractive means to study the kinetics and thermodynamics of interactions between DNA and their binding partners (ligands) **[39, 51]**, and to select DNA aptamers with high efficiency **[93]**.

Like all electrophoresis-based methods, NECEEM is limited in the type of run buffers it is compatible with. A review of recent literature has revealed that the vast majority of KCE experiments, either aimed at analysis of DNA or other types of molecules, are performed using a narrow set of low ionic strength and low ionic mobility run buffers, with Tris and tetraborate solutions of 20-50 mM ionic strength being the most common **[74, 78, 79, 91, 95, 103-108]**.

Preference for these buffers is often justified by their relatively small degree of Joule heating and slow time of ion depletion. The restricted variety of compatible run buffers, however, is a major drawback of NECEEM, as in some cases it may weaken the relevance of produced analytical data to the corresponding processes in cells. Affinity interactions are sensitive to their environment, and may be altered by such factors as the ionic strength of the solution, its pH, temperature, and the presence of co-factors and stabilizing agents. Thus, for NECEEM measurements to be meaningful, affinity experiments must be performed in environments that enable the studied molecules to assume their natural, or intended, conformation. As DNA aptamers are often created for use as *in vivo* imaging probes, drug delivery vectors, and even drugs themselves **[109]**, selection and analysis of such DNA aptamers must be performed under physiological conditions, *e.g.* in PBS, the most commonly used physiological buffer in biology and medicine. Our own experience shows (as it will be exemplified in this Chapter) that results of NECEEM measurements performed in non-physiological buffers often diverge from values obtained by alternative methods in PBS. Thus, NECEEM would benefit greatly if they could be performed with PBS as the run buffer. Despite the obvious advantages of such applications, PBS has been used as a CE run buffer only a few times **[110-113]**. More surprisingly, I did not find any reports where PBS was used as a CE run buffer for analysis of DNA samples.

In this Chapter, I test the suitability of PBS as a NECEEM run buffer for analysis of DNA-ligand binding and show that its use under standard NECEEM conditions renders DNA undetectable. I show that this previously unreported detrimental effect is caused by a combination of phenomena which include: (i) rapid buffer depletion, (ii) unstable EOF, (iii) severe peak broadening, and (iv) extremely low mobility of DNA. After unraveling the causes for this compounded effect, I suggest an approach to overcome the DNA detection problem in

PBS by employing a pressure-assisted modification of NECEEM. I demonstrate the feasibility of this approach by applying it to a study of an experimental model of a PDGF protein and its DNA aptamer. I show that the results obtained in PBS run buffer (PBS-NECEEM) are much closer to previously reported values than the results obtained with a conventional CE buffer (CB)-NECEEM. This approach significantly improves the analytical value of NECEEM by making the results more relevant to *in vivo* applications.

4.2. EXPERIMENTAL SECTION

4.2.1. Materials

All chemicals and buffer components were purchased from Sigma-Aldrich (Oakville, ON, Canada) unless otherwise stated. Human PDGF recombinant protein was purchased from R&D Systems (MN, USA). Fused-silica capillaries were purchased from Molex (AZ, USA). DNA aptamer (36t) with affinity toward PDGF protein was selected previously by others **[114]**, and was custom synthesized by Integrated DNA Technologies (Coralville, IA). The nucleotide sequence of the Alexa Fluor 488-labelled, ssDNA aptamer was: 5'- Alexa 488 - CAC AGG CTA CGG CAC GTA GAG CAT CAC CAT GAT CCT GTG - 3'. NanoDrop-1000 spectrometer (Thermo Scientific, Wilmington, DE) was used to verify DNA concentration in the stock solution by measuring light absorbance at 260 nm, and dividing the absorbance by a manufacturer-provided extinction coefficient. The concentration of the PDGF protein was measured using a BCA assay kit from ThermoFisher Scientific (Waltham, MA) according to manufacturer's instructions and using the same spectrophotometer.

4.2.2. Capillary electrophoresis

All CE experiments were carried out using P/ACE MDQ instrument (Beckman Coulter, Mississauga, ON), equipped with a standard fluorescence detector and a 488 nm line of

continuous Wave Solid-State laser (JDSU, Santa Rosa, CA) for fluorescence excitation. Runs were performed in uncoated fused-silica capillaries with inner radii of 37.5 μm and outer radii of 180 μm. Capillaries of total lengths of 30 or 50 cm were used as denoted. In both cases, the detection window was 10.1 cm away from the outlet end of the capillary. Times of pressure application which yielded desired volumes of hydrodynamic injection were calculated using CE Expert software, version 2.2 from Sciex (Concord, ON), for each given total length of the capillary.

Two electrophoresis run buffers were compared in these experiments: CB: 50 mM Tris-HCl pH 7.4; and PBS: 8.1 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4. Both the inlet and the outlet reservoirs always contained the electrophoresis run buffer of choice. Prior to every run, the capillary was rinsed with the run buffer at 20.0 psi (138 kPa) for a time sufficient to pump 10 capillary volumes. At the end of each run, the capillary was rinsed with a succession of 100 mM HCl, 100 mM NaOH, and deionized water, at the same pressure/time settings. The samples were injected into the capillary, which was pre-filled with the run buffer, by a 0.50 psi (3.4 kPa) pressure to yield a 5 mm-long sample plug.

To test if the DNA sample interacts with the walls of the capillary, a sample of 50 nM DNA was driven through the capillary by 0.5 psi (3.4 kPa) pressure, with subsequent washes with 100 mM NaOH and 100 mM HCl, at the same pressure. The length of the capillary to the detector was 10.1 cm (total length of the capillary was 30 cm).

4.2.3. NECEEM

NECEEM experiments were performed to study the interaction between PDGF protein and its DNA aptamer. Total capillary length in NECEEM experiments was 50 cm. Plugs of equilibrium sample mixtures were injected from the inlet end of the capillary and electrophoresis was carried

out with the anode at the injection end (inlet end). The sample mixture was propagated through the uncooled portion of the capillary by injecting a 4 cm-long plug of running buffer with a 0.50 psi (3.4 kPa) pressure at zero voltage **[44]**. Dilutions of all sample components were prepared with the electrophoresis run buffer, to minimize issues from electrolyte mismatch between the sample plug and the run buffer. Sample mixtures were incubated at room temperature for 15 min prior to injection to achieve equilibration in the binding reaction. To ensure the best accuracy of the measured K_d values, the concentrations of equilibrium mixture components were adjusted based on K_d values obtained from preliminary experiments according to the criteria described in **Section 3.3.2** on **page 50**. The presented NECEEM experiments were performed with protein concentrations similar to preliminary K_d, while the concentration of the aptamer was lower than the preliminary K_d. For the CB experiments the final concentrations of equilibrium mixture components were: 100 nM PDGF, 40 nM aptamer and 100 nM boron-dipyrromethene (BODIPY), used as an internal standard and EOF marker. Electrophoresis was carried at a 500 V/cm electric field. The capillary coolant temperature was set to 21 °C. The internal capillary temperature of 25 °C (in the cooled region) was calculated using SUMET **[53, 54]**. For PBS experiments, the final concentrations of equilibrium mixture components were: 5 nM PDGF, 1 nM aptamer, and 2.5 nM BODIPY. In addition, 1 mg/mL of BSA protein was added to the equilibrium mixture to prevent adsorption of DNA onto the plastic sample vial. Electrophoresis was carried at a 200 V/cm electric field. The capillary coolant temperature was set to 22 °C, with internal capillary temperature calculated to be 25 °C. Further, for PBS-NECEEM experiments, a pressure of 0.30 psi (2.1 kPa) was applied from the inlet end of the capillary to supplement the electric field. Each experimental point was measured in triplicates,

with a fresh equilibrium mixture prepared for each replicate. Electropherograms were analyzed using NAAP area deconvolution software **[95]**.

4.2.4. Mobility measurements

To estimate the mobility of DNA with the standard NECEEM procedure, a sample mixture of 1 μM BODIPY (mobility marker), 100 nM fluorescein (mobility marker), 100 nM DNA aptamer and 1 mg/mL BSA were used. Electrophoresis was performed in both buffers using the same conditions as in the NECEEM experiments, except that, in the PBS experiments, no pressure was co-applied with the electric field.

To estimate the mobility of DNA in a capillary with a shortened effective length, the post-injection pressure propagation step was modified to increase the pressure propagation distance to 9 cm. A 30-cm (total length) capillary was used for these experiments, with a sample being injected from the outlet end of the capillary. This reduced the distance between the initial sample position and the detection window to 1.1 cm. The electric field was applied with the positive electrode at the injection end (outlet). Due to the poor separation in the very short effective length of the capillary, fluorescein was excluded from the sample mixture to avoid its convolution with the BODIPY peak.

To estimate the stability of EOF over time, a 30-cm (total length) capillary was used, with sample injected from the outlet end of the capillary, and the electric field applied with the positive electrode at the same end. Plugs of 100 nM BODIPY were injected into the capillary with intervals of 260 s, and electrophoresis was carried out without run buffer replenishment between the injections. No pressure propagation of the sample was applied after each sample injection. Signal was only recorded during electrophoresis, and not during vial change-over and sample injection. After all of the electrophoresis steps, a small volume of the buffer from

electrophoresis reservoirs was deposited onto Alkacid Test Ribbon (ThermoFisher Scientific, Waltham, MA) to approximate the pH of the solution.

4.2.5. Peak broadening measurements

For all experiments aimed at characterizing the peak broadening phenomena, electrophoresis was performed as described in **Section 4.2.4**, with the exception that no sample propagation was applied in the CB experiments. A 10.1-cm to detector capillary was used. For all PBS experiments, the post-injection pressure propagation step was 9 cm, making the distance between the initial sample position and the detection window 1.1 cm.

4.2.6. Characterization of pressure-assisted NECEEM

To characterize the effects of pressure-assisted NECEEM on the shape of the DNA peak, as well as the reproducibility of experiments, electrophoresis was performed as described in **Section 4.2.4**. The length of the capillary to the detector was 39.9 cm (total length of the capillary was 50 cm). Peak asymmetry factor (PAF) was calculated as follows:

$$\mathrm{PAF}=\frac{A_{\mathrm{front}}-A_{\mathrm{tail}}}{A_{\mathrm{front}}+A_{\mathrm{tail}}} \tag{56}$$

where A_{front} and A_{tail} are front and tail areas of the peak, respectively, separated at the peak maximum. Values of PAF range between −1 and 1, with the value of 0 corresponding to a perfectly symmetric peak.

4.3. RESULTS AND DISCUSSION

4.3.1. Issue with DNA detectability in PBS-NECEEM

To illustrate how the choice of a run buffer can influence NECEEM results, I performed experiments with a binding pair of PDGF protein and its DNA aptamer. The aptamer for PDGF protein has been selected in PBS. As a point of reference, the K_{d} of the interaction measured in

PBS using nitrocellulose binding assay was reported as 96 pM **[114]**. I first performed NECEEM with a conventional CE run buffer (CB), 50 mM Tris-HCl, which has the same pH of 7.4 as PBS, but a 3-time lower ionic strength. The in-capillary temperature was calculated to be 25 °C. The K_d value, which resulted from these CB-NECEEM experiments, was 56 nM (**Figure 4.1**), almost three orders of magnitude higher than the value reported in the original publication on the PDGF-aptamer interaction.

To test if this buffer difference could be responsible for such a drastic discrepancy in measured K_d, or if it arises from more inherent differences in the methodologies, I set out to perform NECEEM experiments with PBS as the run buffer. Based on the available recommendations from the literature, I expected that the main difficulty in using PBS will lie in excessive Joule heating due to its high ionic strength and high mobility of its ions, compared to conventional CE buffers. This issue was relatively simple to handle, as convenient in-capillary temperature determination methods, *i.e.* SUMET **[53, 54]**, allowed me to select the strength of the applied electric field which results in the desired temperature. Thus, for the PBS-NECEEM experiments, I decreased the applied electric field to yield the same 25 °C in-capillary temperature as in CB-NECEEM. The results of the experiment were quite unexpected. As seen in **Figure 4.2**, the two migration markers, which were included in the sample mixture, had

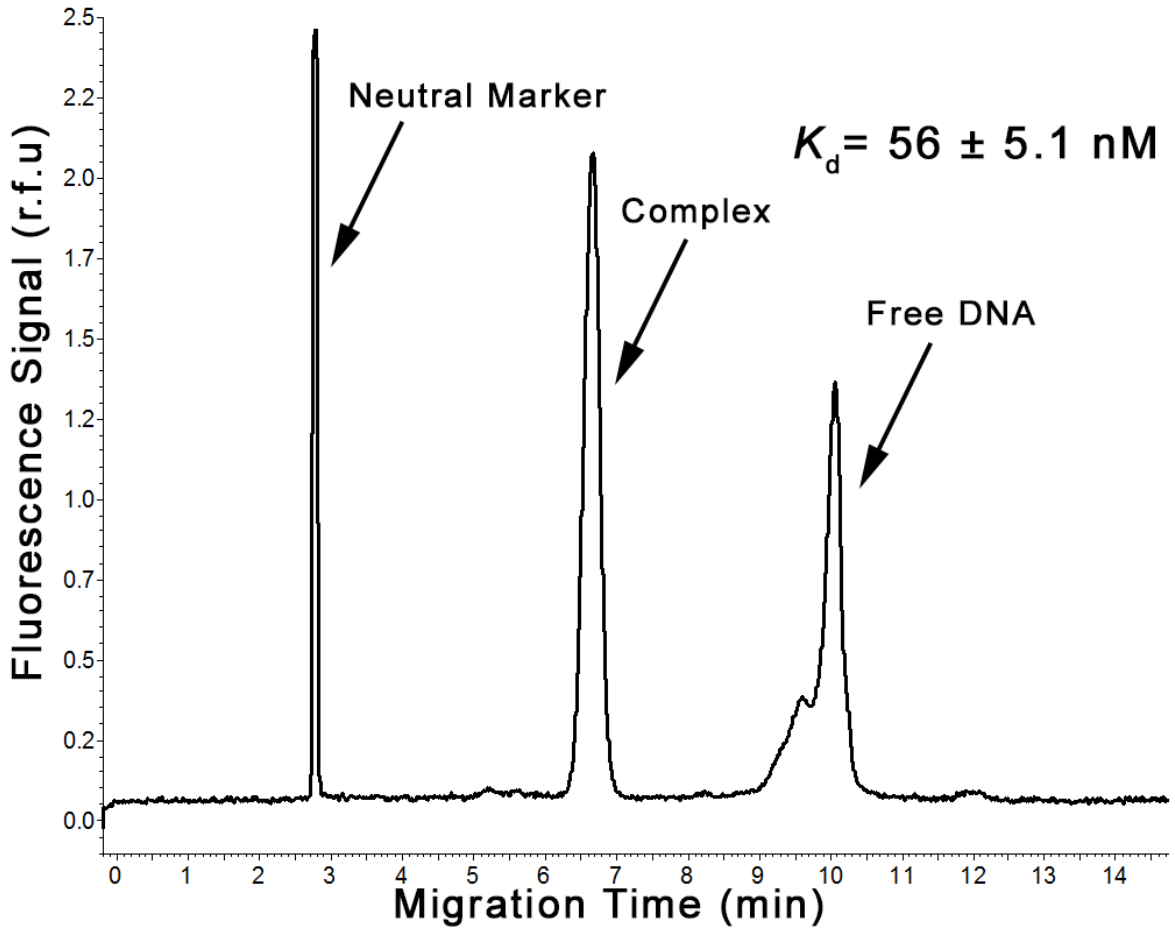

Figure 4.1. NECEEM analysis performed in CB. Sample contained 100 nM BODIPY, 100 nM PDGF, and 40 nM aptamer. The resultant K_d is significantly higher than the value obtained using nitrocellulose binding assay performed in PBS. Asymmetry in the free DNA peak is due to the presence of short DNA sequence impurities.

traveled to the detector with velocities only twice slower than in CB-NECEEM; the signal for the fluorescently labeled DNA, however, did not appear even after 3 h of electrophoresis. The same was true in both the presence and absence of the protein target and for experiments performed with electrode polarities in both positive-to-negative and negative-to-positive arrangements. These results reveal a potential reason as to why the use of PBS as a CE run buffer

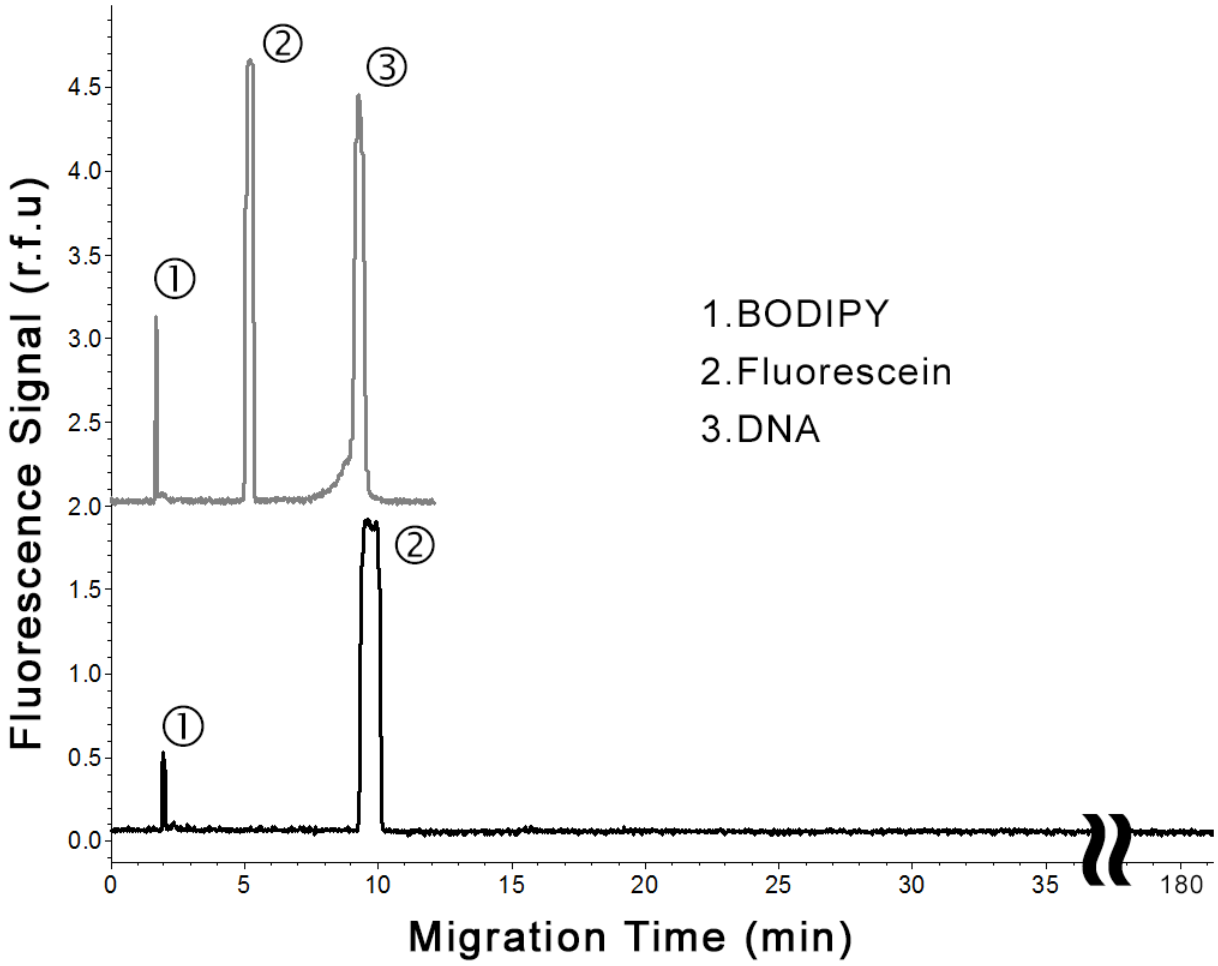

Figure 4.2. Lack of DNA detectability in PBS-NECEEM. A sample mixture of 1 µM BODIPY, 100 nM fluorescein and 100 nM of fluorescently labeled DNA aptamer was analyzed by CE with two different running buffers: CB (top trace) and PBS (bottom trace). In PBS, while both migration markers, BODIPY (neutral) and fluorescein (negative) are detected, DNA is not detected even after 3 hours of analysis. The PDGF target was not included in these experiments. Migration times of BODIPY reflect velocity of EOF.

for analysis of DNA had not been previously documented, despite the attractiveness of such an application. It had become apparent that to facilitate PBS-KCE experiments a better understanding of phenomena underlying the DNA detectability problem is required.

Possible reasons for analyte-specific lack of detectability in CE-LIF can be reduced to three basic causes: (i) reduced detectability of the fluorophore label in a given environment;

(ii) insufficient duration of the experiment for the analyte to reach the detector; or (iii) severe smearing of an analyte over the length of the capillary. To eliminate the observed artifact we need to understand which of these possible causes contribute to the lack of DNA detectability.

4.3.2. Fluorescence intensity

The most discernible reason for the lack of sample detectability in LIF is the poor intensity of fluorescence. The intensity of a fluorophore strongly depends on its environment **[115]**, and the fluorescent label on the employed DNA aptamer may have a diminished capacity to absorb or emit photons in the PBS buffer. To decouple the effects associated with fluorescence intensity from the effects associated with electrophoretic migration, I tested if the fluorescently labeled DNA can be detected with pressure-driven propagation through the capillary. As the presence of PDGF did not influence the outcome of the previous runs, I simplified the sample mixture by excluding the protein from all of the subsequent experiments, unless otherwise stated. As shown in **Figure 4.3A**, fluorescence from labeled DNA was observed with pressure-driven propagation, albeit with poor signal repeatability. I had previously reported on this repeatability issue in DNA analysis which is caused by the interaction of DNA with the walls of the vial used for sample mixture preparation **[116]**. Interestingly, this behavior seems to be more prominent in PBS than in CB, suggesting that the additional ions in PBS facilitate the interaction of DNA with the plastic walls of sample vials. To minimize this behavior, I have added 1 mg/mL of BSA to the sample mixture to passivate the surface of the vial, which has dramatically improved the repeatability of signal intensity (**Figure 4.3B**). It should be noted, that using BSA as a surface passivating agent may not be suitable for all applications, and users are encouraged to consider some of the many available alternatives depending on the specifics of the experiment **[116]**. These experiments show that the fluorescence of Alexa Fluor 488-labelled DNA is not

suppressed in PBS. This is not a surprising finding, as Alexa Fluor 488 is known to be resilient to environmental changes, such as pH and temperature, and is routinely used in PBS **[117, 118]**. To test if the DNA-vial interaction was responsible for the lack of DNA detectability in PBS-NECEEM, I have performed electrophoresis with the BSA-containing sample, but still failed to observe the DNA peak. This suggests that electrophoresis plays a central role in the DNA detectability problem. To avoid the signal repeatability issue due to adsorption of DNA to the vial walls, all subsequent experiments were carried out with 1 mg/mL BSA added to the sample mixture.

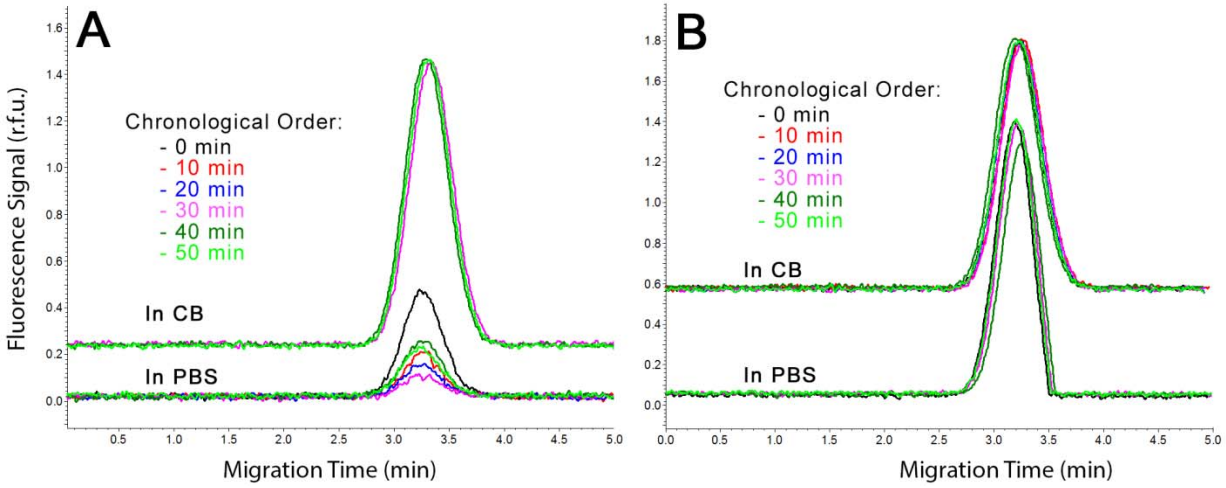

Figure 4.3. Signal repeatability of fluorescently labeled DNA in CB and PBS buffers. A plug of 10 nM DNA was propagated through 10.1 cm of capillary length to detector (total length of capillary was 30 cm) by 0.30 psi (2.1 kPa) pressure. Six repetitions of each experiment were conducted in the chronological order signified by trace colors and presented in the legend. In each panel, top set of traces were obtained with CB as the sample and the run buffers, while bottom set was obtained with PBS. **Panel A:** signal repeatability with no additives in the sample; **Panel B:** signal repeatability with 1 mg/mL of BSA added to the sample for vial surface passivation. Addition of BSA significantly improves signal reproducibility. Fluorescence efficiency of labeled DNA in PBS is similar to one in CB.

4.3.3. Analyte velocity

In CE, an analyte may not be detected if an insufficient amount of time had been allotted for it to reach the detector. Even though my previous PBS-NECEEM experiments had lasted for 3 h, an excessive time for a typical CE run, it might not be sufficient if the analyte velocity is less than 1 mm min^{-1}. To directly measure a low velocity in a manageable timeframe I needed to perform electrophoresis in a shorter length of capillary. While the employed commercial CE instrument allows a minimum of 10-cm capillary length to the detector, the effective distance of migration can be shortened further by pressure-propagating the sample closer to the detector before the application of the electric field. By employing this technique, I performed electrophoresis with the sample plug starting position being 1.1 cm away from the detector. Shortening the electrophoresis distance had allowed me to observe the DNA peak in PBS-KCE (**Figure 4.4**), which migrated with low velocities in a range of 0.2-0.6 mm min^{-1}. At this range of velocities, it would take over 20 h for DNA to traverse the standard capillary length to the detector of 40 cm. To understand what causes the DNA to slow down to a near-stationary migration in PBS-NECEEM, we need to take into account the different mobility vectors that define it. The velocity of an analyte in CE is determined by its total mobility (μ_{tot}) and the strength of the applied electric field. The above measurement shows that μ_{tot} of DNA is ~0.003 $mm^2\ V^{-1}\ min^{-1}$. In bare-silica capillaries, μ_{tot} of DNA is a sum of two opposing vectors: mobility of the EOF (μ_{EOF}), directed toward the cathode, and the electrophoretic mobility of DNA (μ_{eph}), directed toward the anode. EOF is a bulk flow, meaning that its mobility vector affects all analytes equally. The magnitude of μ_{EOF} is inversely proportional to the ionic strength

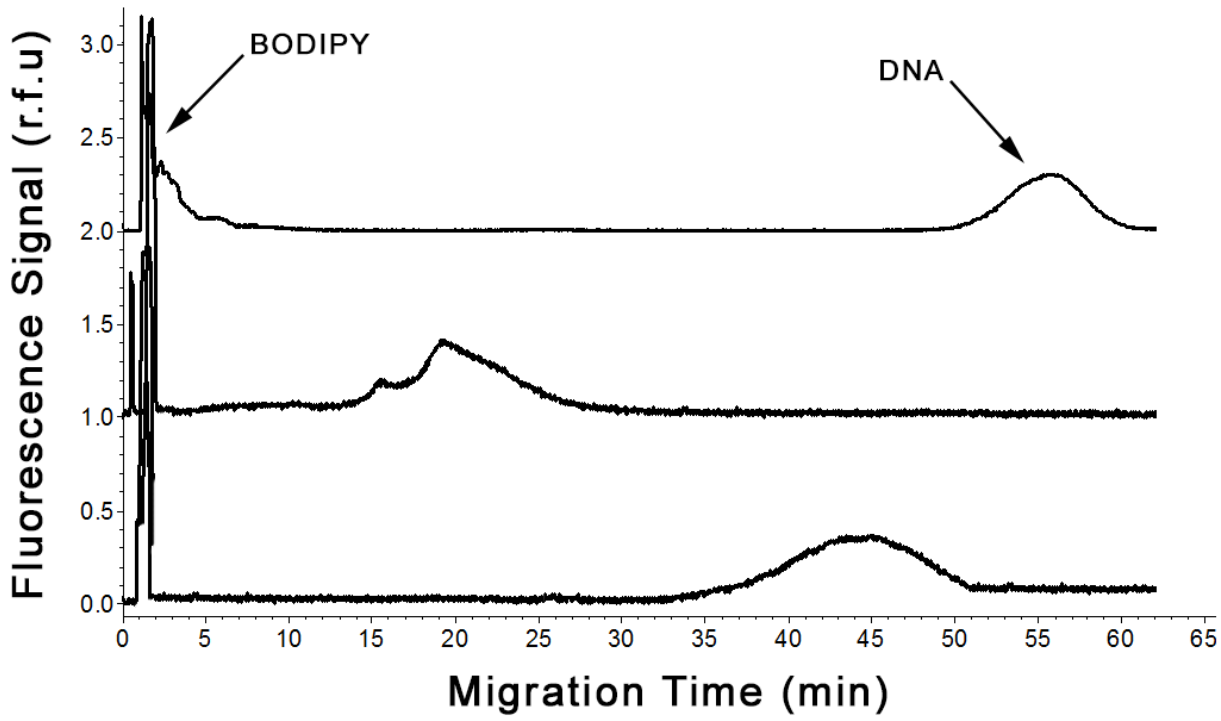

Figure 4.4. Mobility of DNA in PBS-NECEEM. The experiments were performed in 11 mm effective capillary length. A plug of 1 μM BODIPY and 100 nM fluorescently labeled DNA aptamer was injected into a 10.1 cm to detector capillary and was propagated by pressure 9 cm. The three stacked traces show trials of the same experiment in chronological order (first experiment at the bottom, last at the top). DNA migration velocity has very poor repeatability.

of the solution, and since the ionic strength of PBS is greater, its μ_{EOF} is lower than in CB, 2.25 $mm^2\,V^{-1}\,min^{-1}$ and 3.85 $mm^2\,V^{-1}\,min^{-1}$, respectively (**Figure 4.2**). The near-stationary μ_{tot} of DNA, thus, can arise if the magnitude of μ_{eph} is similar to the magnitude of $\mu_{EOF.}$ Interestingly, the μ_{eph} of DNA in CB-NECEEM is −2.20 $mm^2\,V^{-1}\,min^{-1}$ (negative sign signifies that it opposes the flow of the electrical current), suggesting that it does not change significantly upon transition from CB to PBS.

Besides revealing the small magnitude of DNA $\mu_{tot,}$ results in **Figure 4.4** reveal two additional complications: first, the migration of DNA in PBS-NECEEM has very poor repeatability, with

elution times ranging between 20 to 60 min; second, the peaks manifested significant broadening and loss of intensity. It is possible that the poor repeatability of DNA migration time, observed in **Figure 4.4**, was caused by imprecisions in the pressure propagation step, introduced to shorten the effective capillary length; however, a similar experiment performed with CB demonstrates that sample propagation by itself does not cause migration repeatability issues (**Figure 4.5**), establishing the fact that these effects are tied to electrophoresis in PBS. Many (if not most) migration no-repeatability issues in CE are caused by inconsistent velocities of the EOF. Unless the analyte undergoes a time-dependent change in its molecular size-to-charge ratio, or the applied electric field strength is imprecisely controlled (both issues are rare), the μ_{eph} of an analyte remains stable. In contrast, even under optimized conditions the μ_{EOF} fluctuates in response to changes in the ionic strength of the buffer and the conditioning of the capillary surface and slight variations in pH and temperature **[119]**. In conventional CE experiments, when analyte mobilities are relatively high, small variations in EOF result in insignificant changes to μ_{tot}; however, when μ_{tot} is small, the same variations in EOF will have a pronounced impact on the repeatability of analyte migration. Moreover, it is likely that the stability of EOF in PBS is worse than in CB. Electric current passing through an electrolyte solution causes ion depletion and formation of pH gradients, as consequences of electrophoresis and electrolysis, respectively. In CB-KCE these changes occur on a much larger time scale than the time of analysis, ensuring that constant velocities are maintained throughout the experiment. Compared to CB, PBS consists of smaller, more mobile ions, and has a lower buffering capacity, making it likely that considerable depletion of ions and establishment of pH gradients occur before analysis is complete. All of these instabilities will have a pronounced effect on μ_{EOF}.

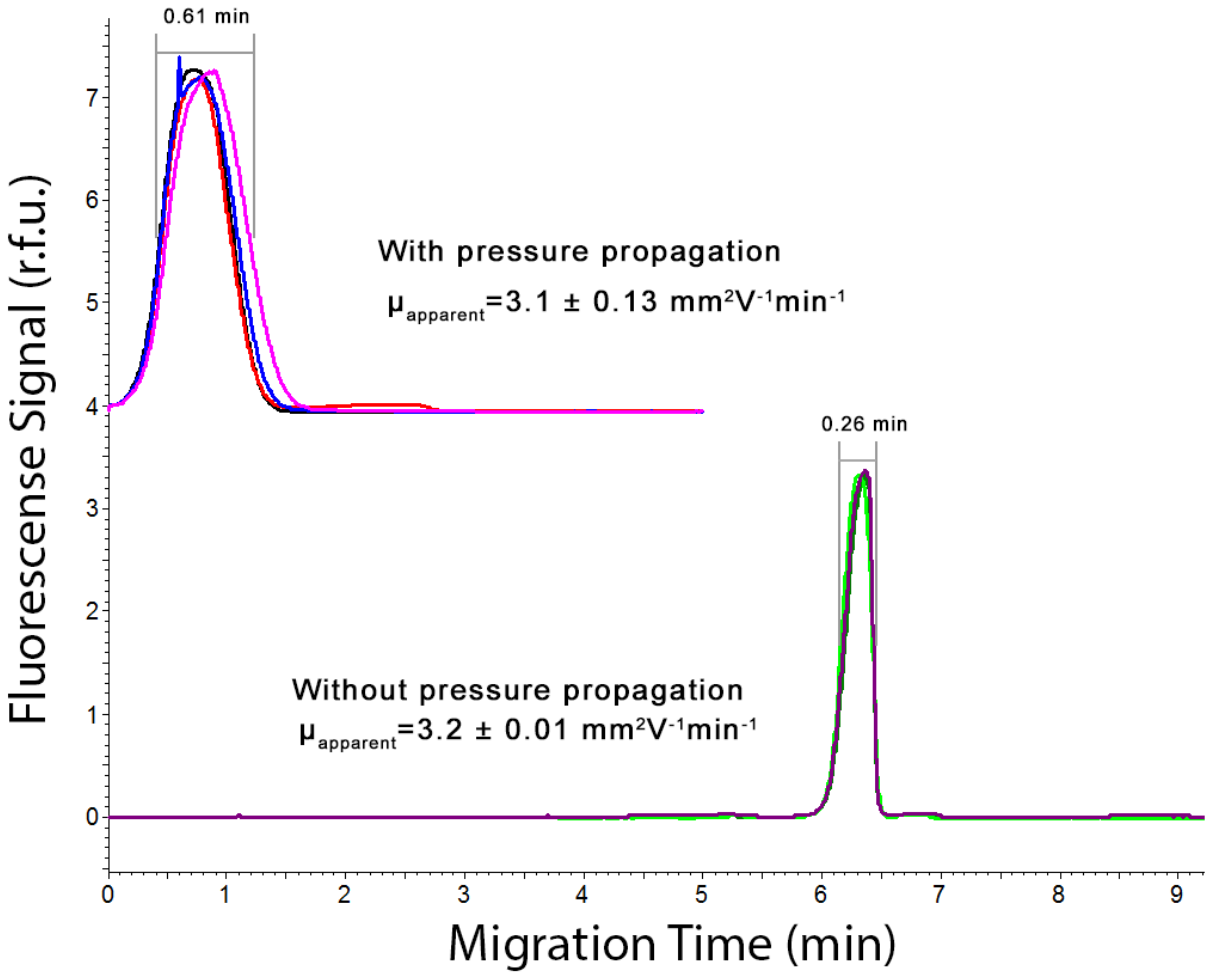

Figure 4.5. Effects of pressure propagation step on precision of μ_{eph} measurements in CB. All analyses were performed with the sample mixture of 50 nM DNA aptamer and 1 mg/mL BSA. The bottom set of traces shows migration of DNA in CB electrophoresis experiment without pressure propagation, and serves as a reference point for reproducibility, velocity, and peak widths. The top set of traces shows migration of the same DNA sample after it was pressure-propagated 9 cm closer to the detector. The pressure propagation step was performed the same way as in experiments described in **Figure 4.4**. The electrophoresis step was performed at a reduced strength of applied electric field (50 V/cm) to account for higher velocity of EOF in CB. Based on the velocity value from the bottom trace, the accuracy of the pressure propagation step is estimated to be within 0.03 mm, with a precision of 0.05 mm. Pressure propagation step increases the peak width 2.3-times. Thus, the introduction of the pressure propagation step cannot account for the poor repeatability of migration times and peak broadening observed under similar conditions in PBS.

To assess and compare the stability of EOF in CB and PBS, I conducted an interval injection experiment, in which several injections of neutral marker were made at equal time intervals, and electrophoresis was performed without buffer replenishment between the injections (**Figure 4.6**). Any changes in the elution time of the neutral marker with consecutive injections indicate a change in μ_{EOF} and signify the depletion of the buffer in the reservoirs. As expected, for CB the elution velocity of the neutral marker remained stable over the 20 min duration of the experiment. For PBS, however, within the first 14 min of electrophoresis the velocity of EOF had decreased sufficiently to prevent marker detection. This confirmed my speculations about the susceptibility of PBS to fast ion depletion. Interestingly, no significant pH or conductivity changes were observed in the PBS buffer reservoirs after the 14 min of electrophoresis, suggesting that the decrease in μ_{EOF} was predominantly caused by the change in non-proton ion content of the buffer. Given the complicated and transient nature of surface charge distribution in response to a changing ionic background, it is the most likely explanation for the poor repeatability of migration times in PBS-KCE.

The poor stability of EOF in PBS-NECEEM suggests that the above-measured estimate of DNA's $\mu_{tot} = 0.003\ mm^2\ V^{-1}\ min^{-1}$ represents a time-averaged value; *i.e.* its magnitude is larger in the beginning of the experiment and becomes smaller as buffer depletion continues. It may even be possible, if allowed enough time, that μ_{EOF} decreases to a point that μ_{tot} of DNA becomes negative, causing the DNA molecules to start migrating back toward the anode, never reaching the detector. As such, it is important to perform PBS-KCE experiments in less than 10 min, before the effects of buffer depletion set in.

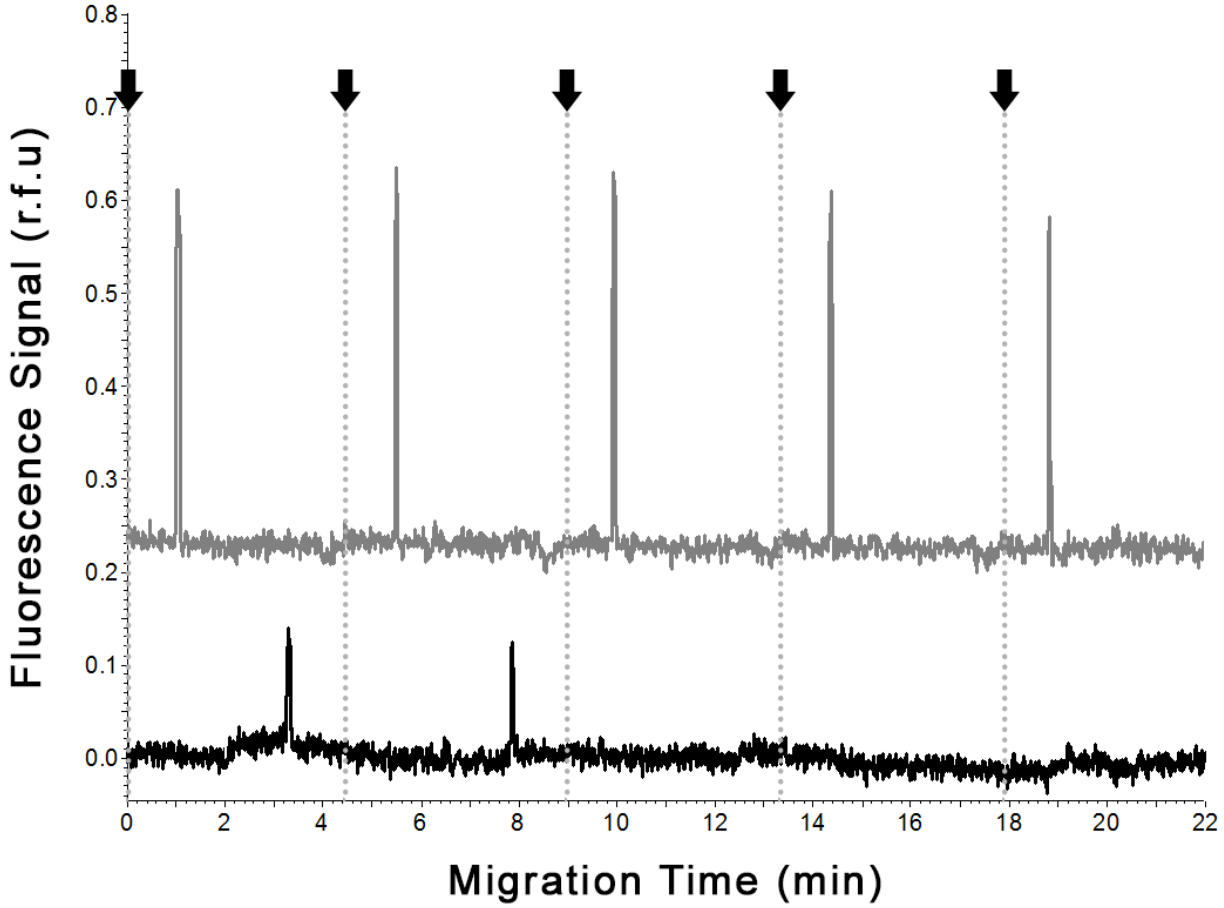

Figure 4.6. EOF stability test. A plug of 100 nM BODIPY was injected every 260 s, and electrophoresis was carried out between the injections from the same set of buffer reservoirs. Elution time of the neutral marker peak from the time of each injection (marked by the arrow and dotted vertical line) is indicative of the EOF velocity over the increasing time of electrophoresis. Experiment was performed with CB **(top trace)** and PBS **(bottom trace)** as run buffers. No signal was recorded during vial switching between sample injection and electrophoresis steps. PBS buffer depletion occurs just under 14 minutes.

4.3.4. Peak broadening

Lastly, as observed in **Figure 4.4**, electrophoresis in PBS causes the width of the DNA peak to increase by approximately 10-times compared to peak widths in CB (after normalizing by residence time in the detector). Some of this peak broadening can be ascribed to the pressure-propagation step, which was introduced to decrease the effective length of the capillary to

1.1 cm, as it increases peak widths by approximately 3-times (**Figure 4.5**). However, all of my NECEEM experiments incorporate a pressure-propagation step to avoid the uncooled portion of the capillary, meaning that this level of broadening will be present in all my experiments. Additional peak broadening, however, can aggravate the issue of poor DNA detectability, if the local concentration of the analyte decreases below the limit of detection in significant parts of the sample zone.

In CE, peak smearing can occur either as a result of electro-dispersive phenomena (*e.g.* anti-stacking), sample interaction with the walls of the capillary or due to the interaction of the analyte with other species in the solution. Electro-dispersive peak smearing (anti-stacking) occurs when the ionic strength of the sample solution is significantly higher than that of the run buffer **[34]**. In my experiments, the sample buffer was always matched to the run buffer, eliminating the possibility of anti-stacking. Moreover, anti-stacking affects all analytes in the sample, while in my previous tests no signs of dispersal for peaks of BODIPY or fluorescein were observed (**Figure 4.2**). The surface of a bare-silica capillary has a negative charge, same as the charge of DNA, which makes the attractive interactions between the DNA molecules and the walls of the capillary unlikely. Nevertheless, to exclude the possibility of ion-facilitated DNA-silica interactions, I have conducted a pressure propagation test for analyte adsorption onto capillary walls, as described previously **[120]**, where a sample of DNA is driven through the capillary by pressure, with subsequent washes with NaOH and HCl. The harsh pH conditions disrupt electrostatic interactions between the sample and the walls and allow the adsorbed molecules to elute and be detected. Expectedly, during both the NaOH and HCl washes no additional elution of DNA was observed, suggesting that no adsorption of DNA on the capillary walls occur (**Figure 4.7**).

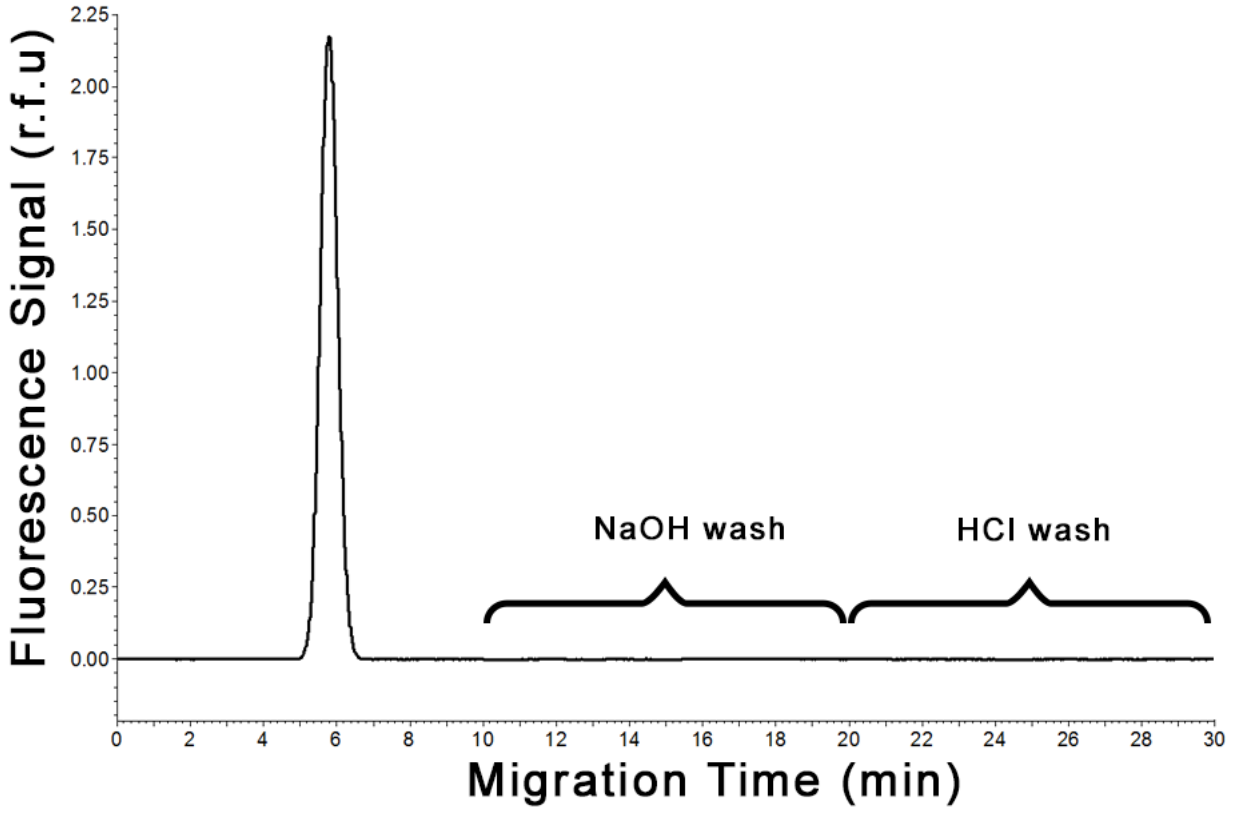

Figure 4.7. Analysis of DNA adsorption onto capillary walls. Lack of additional elution of DNA during the base and acid washes suggests that no adsorption of DNA on capillary walls occurs.

In PBS-KCE experiments, DNA is always surrounded by a high concentration of Na^+ and K^+, which can interact with its backbone and partially neutralize its charge. If these counterion-DNA complexes are stable over a comparable amount of time that it takes to spatially separate DNA molecules of various degree of neutralization (*i.e.* on the order of seconds or minutes), then dynamic peak broadening will ensue. In fact, the KCE technique of ECEEM relies on such peak broadening to measure the kinetics of molecular interactions **[121]**. As it will be further discussed in Chapter 5, I have shown that DNA can form stable complexes with metal counterions, including Na^+ and K^+, which dissociate on a minute-timescale.

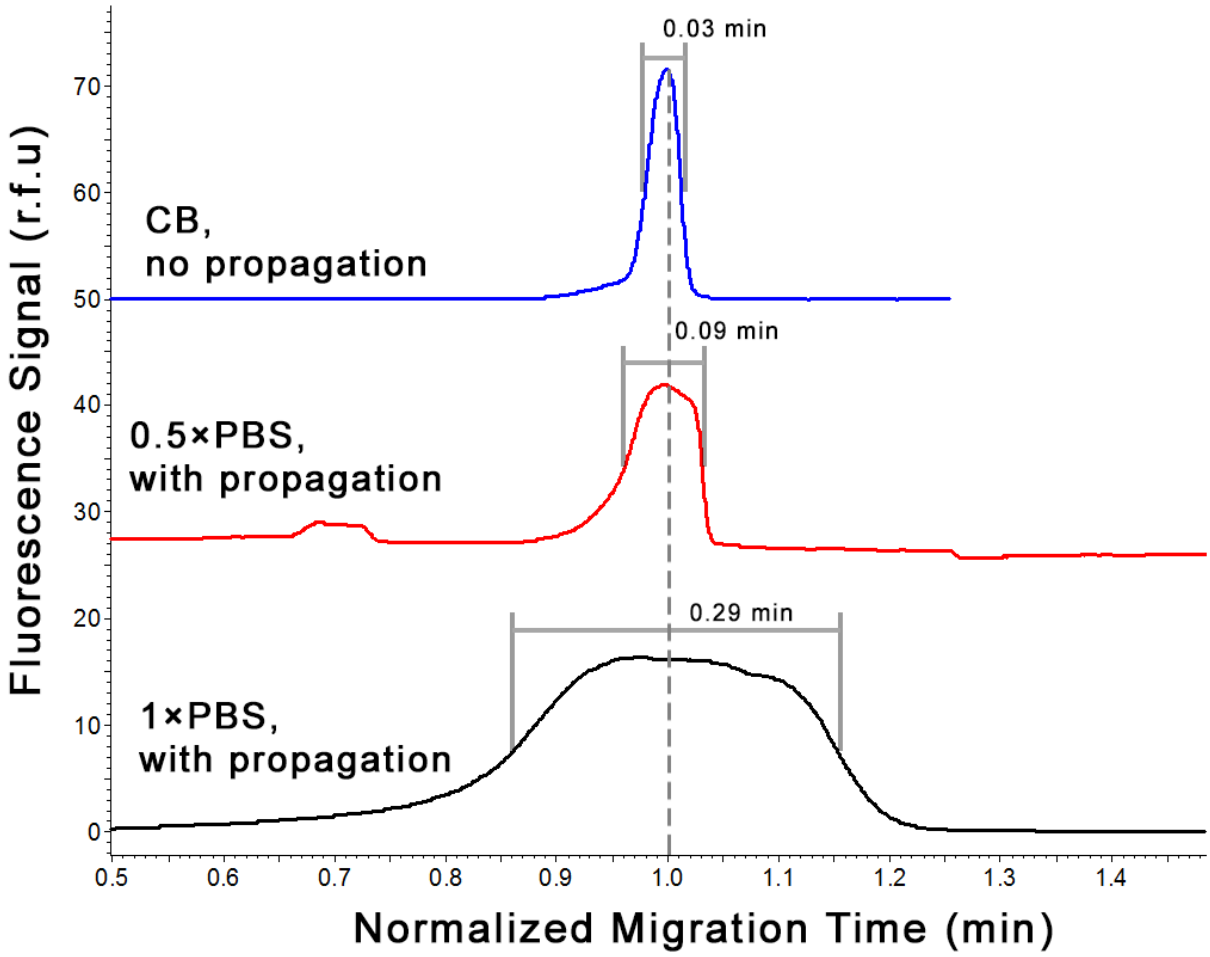

Figure 4.8. Dynamic peak broadening in PBS. Comparison of widths of DNA peak in CB (blue trace), 2-times diluted PBS (0.5×PBS, red trace), and non-diluted PBS (1×PBS, black trace). Data shown is after normalization by residence time in the detector; the normalization was done by applying a multiplier to the *x*-axis equal to the reciprocal of the migration time of the peak. After normalization, the DNA peak width in non-diluted PBS is 3.2 times wider than in 2-times diluted PBS. This suggests that the ionic strength of the buffer plays a role in peak broadening. The 3-times increased peak width between CB and 2-times diluted PBS is not due to the difference in ionic content (as the concentration of the major ionic species is similar), but due to the pressure propagation step.

Such interactions, thus, are a feasible explanation for the observed peak broadening. To verify this hypothesis, I have performed electrophoresis in a 2-time dilution of PBS and observed

approximately 3-time reduction in peak width when normalized by migration velocity (**Figure 4.8**). These observations strongly support the notion that DNA peak broadening in PBS-KCE occurs as a result of its interaction with the increased background of counterions. In this scenario, the longer the DNA sample is subjected to electrophoresis in the high ionic strength environment, the broader its peak becomes; thus, decreasing the duration of PBS-KCE experiments should help mediate this issue as well.

4.3.5. Pressure-facilitated NECEEM

The phenomena described in the previous sections combine in a synergistic manner. The near-stationary migration of DNA causes its retention time to be excessively long, while buffer depletion and peak broadening are progressively aggravated with time. Thus, to alleviate the DNA detection issue in PBS-KCE the analysis time must be significantly shortened. The most direct approach to achieve this is to reduce the physical length of the capillary; however, given the estimates of DNA mobility and ion-depletion time reported in previous sections, the capillary would have to be shortened to less than 1 cm, a feature difficult to implement. This can be somewhat offset by increasing the volume of the buffer reservoirs, to allow for slower buffer depletion. However, both of these changes require a high degree of customization to the CE instrument. Given the prevalence of commercial CE instrumentation, a convenient solution to the DNA detectability problem in PBS-KCE should rely on commonly available features and capabilities of such tools. An alternative approach to reducing CE analysis time, which is easily realized in most CE machines, involves supplementation of electrophoresis with a pressure-driven hydrodynamic flow. This, essentially, adds an additional velocity vector to all analytes, which offsets the drop in μ_{EOF} in PBS, and brings μ_{tot} of DNA to a more practical magnitude.

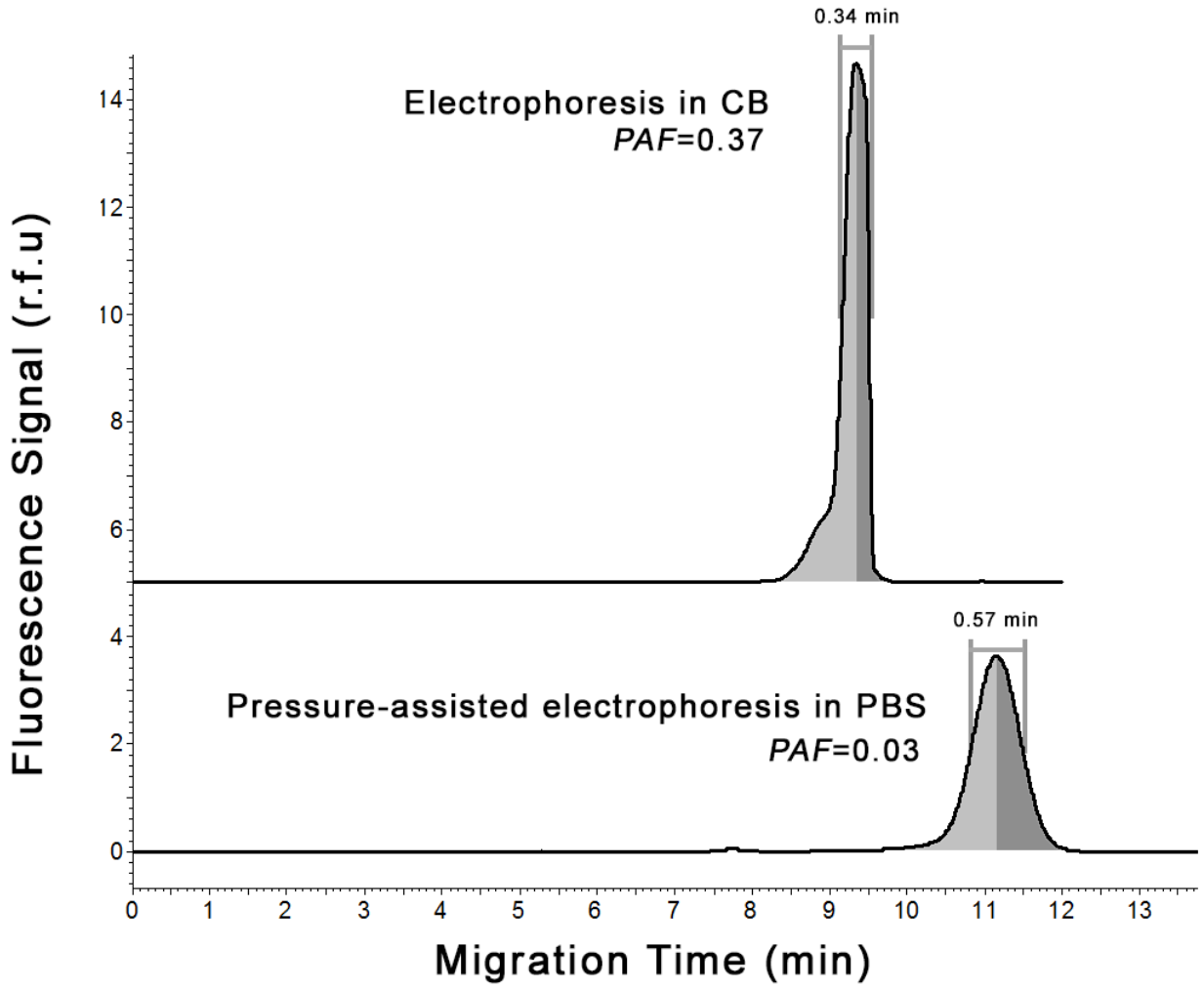

Figure 4.9. Effects of pressure-assisted electrophoresis on DNA peak shape. DNA peak shape and width as a result of electrophoresis in CB (top trace) and pressure-assisted electrophoresis in PBS (bottom trace). Both analyses were performed with a sample mixture of 50 nM DNA aptamer and 1 mg/mL of BSA. Pressure-assisted electrophoresis increased peak width by 1.7 times, and decreased the peak asymmetry factor (*PAF*) by 10 times. The increased peak width is not expected to be detrimental to KCE data analysis, while the improvement in peak symmetry is expected to be beneficial, as described previously **[55]**.

Since pressure-driven velocity can be finely controlled, the total analysis time in PBS-NECEEM can be matched to that in CB-NECEEM, an added advantage for my comparative study.

When applying a hydrodynamic flow, we must keep in mind the shape of its profile, which, unlike the flat-fronted EOF, is parabolic. A significant flow velocity difference between the center and the boundaries of the capillary lumen cause deformation and asymmetry in resulting peaks, which makes signal area deconvolution more difficult in NECEEM **[55]**. This behavior, however, can be alleviated by using slow hydrodynamic flows, or by using capillaries of a smaller inner radius, in which the parabolicity of the flow profile is smoothed out by transverse diffusion. Before performing PBS-NECEEM analysis, I have confirmed that pressure-assisted NECEEM in PBS, performed in the same capillary setup as in the previously discussed CB-NECEEM experiment, does not cause significant peak aberrations (**Figure 4.9**). These experiments were also used to estimate analytical performance parameters of pressure-assisted KCE, namely the precision of analyte migration time to the detector, which had a relative error of 0.6%, and the reproducibility of peak area measurements for a given analyte, which had a relative error of less than 2.4%.

Using the pressure-assisted NECEEM has, at last, allowed me to perform the analysis in PBS which has initiated this troubleshooting effort (**Figure 4.10**). The K_d value measured with PBS-NECEEM was 6.2 ± 1.2 nM, which is smaller than the CB-NECEEM results by nearly an order of magnitude. This suggests that the conformations of the aptamer and the protein are more favorable to form and maintain the complex in the high-ionic environment of PBS, in which the aptamer was selected. The K_d value obtained in PBS-NECEEM value is also much closer to the results of the nitrocellulose binding assay reported upon the original characterization of the aptamer **[114]**. While the original publication reports the K_d as 0.096 nM, closer inspection of the data suggests that this value might have been underestimated. The value was obtained by fitting

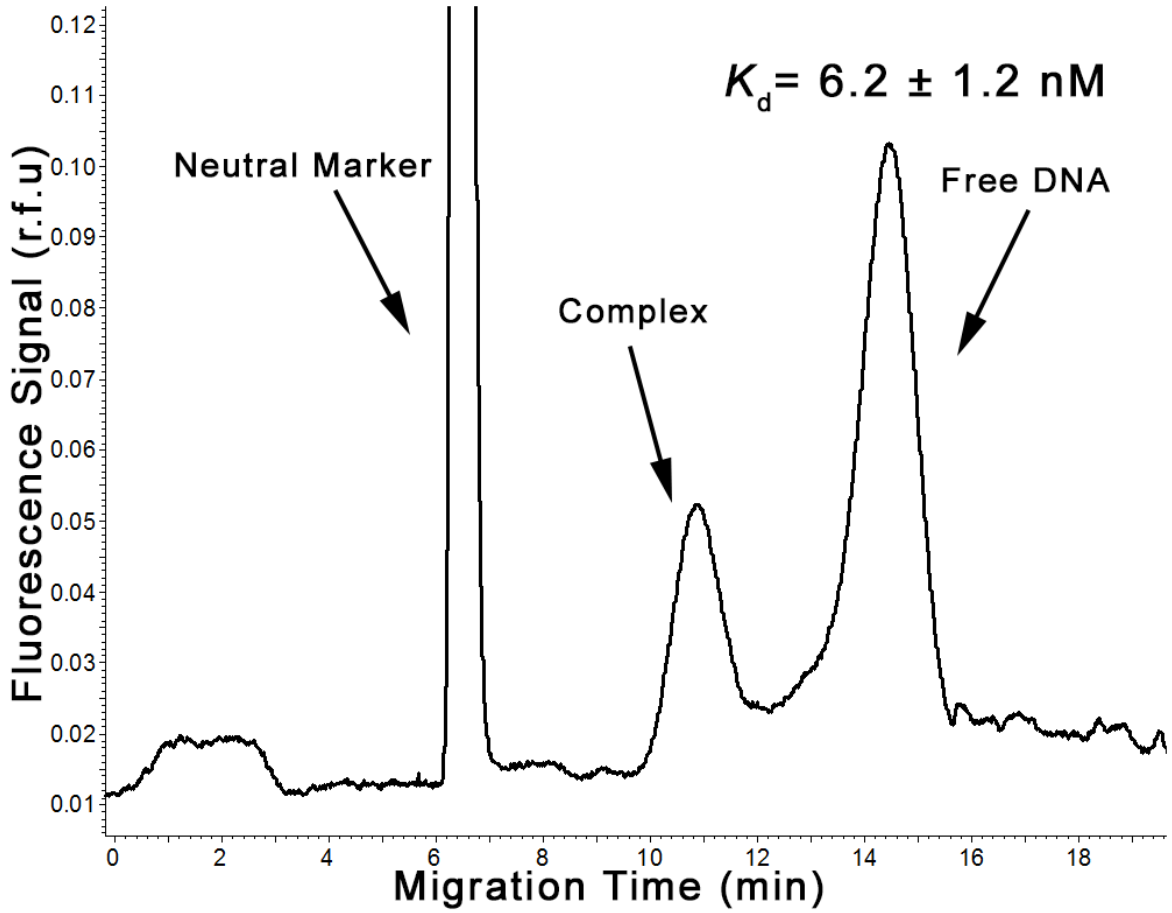

Figure 4.10. Pressure-assisted NECEEM analysis performed in PBS. The sample contained 2.5 nM BODIPY, 5 nM PDGF, and 1 nM aptamer. The experiment was performed with an added pressure of 0.30 psi (2.1 kPa). The K_d value measured in PBS is significantly closer to the value measured by nitrocellulose membrane-binding assay (1 nM), than the value measured in CB (56 nM).

experimental data to a theoretical equation of a binding curve; however, the fairly large variance in replicates and abnormalities in the shape of the binding curve, which can be observed in the original figure (**Figure 5.C** in reference **[114]**), may have caused the fitting algorithm to return a value with a significant systematic error. Visual analysis of the published binding curve suggests that the half-binding point (which corresponds to K_d for 1:1 interactions) is reached at PDGF

concentration of approximately 1 nM, which is a more conservative estimate of the K_d of this interaction. With this in mind, the value produced by our PBS-NECEEM (6.2 nM) is in line with the re-evaluated characterization data of the aptamer (1 nM).

4.4. CONCLUSIONS

The main cause of poor DNA detectability in PBS-NECEEM is the close match between time-averaged magnitudes of μ_{EOF} and μ_{eph} vectors of opposing signs, which combine into a near-zero value of DNA μ_{tot}. When dealing with phenomena that are based on narrowly defined thresholds, it is important to question whether they result from a very specific set of employed conditions, or whether they will manifest in a broad range of scenarios. In other words, it is important to understand whether other researchers are likely to experience similar problems and whether the described method of troubleshooting will be of a wide interest. The use of PBS as a NECEEM running buffer is warranted when emulating physiological conditions, a common provision in the analysis of biomolecules. Depending on the model organism in question, such physiological conditions are defined quite narrowly in terms of temperature, pH, and ionic content. PBS, in particular, is universally formulated, with occasional supplementation with sub-mM concentrations of $MgCl_2$ and $CaCl_2$, which will have a negligible effect on μ_{EOF}. Theoretically, μ_{EOF} depends on the zeta potential of capillary wall (dependent on pH), the dielectric constant of the background electrolyte (dependent on ionic strength), and the viscosity (dependent on temperature), all factors that will hold constant for all PBS-NECEEM experiments that use bare-silica capillaries. As for the μ_{eph} of DNA, it remains unchanged across a wide range of molecular sizes **[122-124]**, as a result of the nearly constant mass-to-charge ratio of this polymer. The use of capillary coatings, especially those that stabilize EOF, might be an alternative method to overcome the described issues of poor DNA detectability; however, the use

of the negatively charged bare-silica capillaries is preferential in the analysis of DNA due to its simplicity and absence of DNA-capillary wall interactions. The use of capillary coatings which suppress EOF is not desired, as it will cause issues in the analysis of positively charged DNA binders. Under suppressed EOF the negative DNA and any positively charged analytes will migrate in opposite directions, making it difficult to quantify both in the same run. Suppressed EOF will also lead to longer analysis times (due to slower analyte velocities), which will exacerbate the issue of PBS depletion. The use of coatings that reverse the EOF will decrease the separation efficiency between DNA and most proteins, large molecules which tend to co-migrate with the EOF, making this option non-robust. Due to better heat dissipation properties, experiments performed in capillaries of smaller inner-radii may employ higher electric field strengths to maintain the required in-capillary temperatures; however, such increases in the applied electric field strength would not exceed 2-5 times, and, thus, would not be sufficient to decrease the DNA migration time to the detector to a practical level. It should be noted, that applying a pressure flow (to decrease the analysis time), while at the same time decreasing the strength of the applied electric field (to account for the increased Joule heating) can significantly reduce the separation efficiency of CE. In our case, the separation between DNA and the DNA-protein complex was still sufficient to accurately measure analyte peak areas; however, depending on the studied molecules, the experimenters must ensure that a sufficiently long capillary is used to allow for a better-than-baseline separation between the DNA and its complex. Taking a stock of all these considerations, it is likely that the majority of KCE experiments that aim at analysis of DNA-ligand interactions in PBS will experience the described issues. It is plausible that the insofar absence of PBS-KCE results for DNA is explained by the fact that some researchers have already encountered this problem in the past.

Chapter 5. Non-Uniform Velocity of Homogeneous DNA in a Uniform Electric Field

The presented material was published previously and reprinted with permission from "Musheev, M.U.; Kanoatov, M.*; Krylov, S.N. Non-uniform velocity of DNA in a uniform electric field: the consequence of electric-field-induced slow dissociation of highly stable DNA-counterion complexes. *Journal of the American Chemical Society* 2013, 135, 8041-8046" and "Musheev, M.U.; Kanoatov, M.*; Retif, C.; Krylov, S.N. Stable DNA aggregation by removal of counterions. *Analytical Chemistry* 2013, 85, 10004-10007". Copyright 2013 American Chemical Society".

(* - equal contribution with the first author)

My contribution to the articles was: (i) formulating the counterion condensation hypothesis, (i) formulating the plan to improve NECEEM separation efficiency without desalting, (iii) performing the majority of the experiments, and (iv) writing major parts of the two manuscripts.

5.1. REDUCED PARTITIONING EFFICIENCY OF NECEEM FOR DNA MOLECULES

The efficiency of the separation (partitioning) step in SELEX is crucial to the success of a given aptamer selection project. Conventionally, this partitioning is achieved either through filtration or liquid chromatography. These heterogeneous-phase separation methods have poor efficiencies, as they are prone to non-specific DNA adsorption. As a result, conventional aptamer selection projects require multiple repetitions of SELEX, often demanding as many as 20 rounds **[93]**. More importantly, poor efficiency of partitioning often leads to complete loss of aptamer sequences, and as a result, to failed SELEX projects. Our lab has pioneered the use of NECEEM

for SELEX partitioning **[48]**. Due to the homogeneous nature of CE separation, it provides a significantly improved partitioning efficiency over conventional methods. The use of CE-SELEX has reduced the average number of required selection rounds from 20 to fewer than 8 **[48]**. While significant, this improvement is worse than expected: our theoretical predictions suggest that CE partitioning can reduce aptamer selection process down to a single round of SELEX.

All electrophoretic techniques operate under the assumption that in a uniform electric field, within a uniform electrolyte, DNA migrates with a velocity that is defined by its polymer length and conformation **[125]**. By extension, it is also assumed that identical DNA molecules migrate with a uniform velocity, when diffusion is accounted for. This is confirmed on a daily basis by thousands of researchers that observe uniform zones of DNA in properly performed electrophoresis experiments. Indeed, when accompanied by prevailing optical or radioactivity-based detection approaches, electrophoresis produces apparently uniform velocity profiles for homogeneous DNA samples. The assumption about the uniform velocity of DNA migration in electrophoresis is, thus, widely accepted and not challenged.

In aptamer selection, however, the products of DNA electrophoresis separation are subjected to PCR amplification, a procedure that is much more sensitive to the presence of minute quantities of DNA **[126]**. Through PCR, it has been observed that a considerable amount of DNA appears in fractions that, theoretically, should be devoid of it **[48]**. These unusual migration patterns of DNA prevent the partitioning efficiency of NECEEM to approach its theoretical maximum. In the past, it has been proposed that this phenomenon is caused by the high diversity of structures in the random-sequence DNA libraries used in aptamer selection, but this hypothesis was never conclusively proven.

This work was partially motivated by our realization that typical electrophoresis experiments do not reveal detailed velocity profiles of DNA. They utilize optical or radioactivity-based detection approaches which are characterized by relatively poor limits of detection and dynamic ranges. Irregularities in DNA velocity cannot be noticed if the amount of DNA that moves with irregular velocities is below the LOD. We, hence, decided to study DNA velocity profiles in a larger dynamic range of DNA concentrations by using quantitative polymerase chain reaction (qPCR). This detection approach has a much lower LOD, when compared to the more commonly used alternatives, and can quantitate DNA in a range of concentrations of 10 orders of magnitude **[127]**. In this Chapter, I report on our findings that while the major fraction of DNA molecules does migrate with predictable velocities, there exists a fraction of DNA molecules that migrate with a wide range of irregular velocities. We have determined that the cause for this velocity heterogeneity, within a homogeneous DNA sample, lies in the ability of DNA to form unusually-stable complexes with its counterions. Irregular DNA migration profiles arise due to electric field-induced dissociation of such complexes. Furthermore, I describe another unusual effect of the electric field-induced dissociation of DNA-condensed counterions, which causes DNA to precipitate out of the solution.

5.2. EXPERIMENTAL SECTION

5.2.1. Materials

All chemicals were purchased from Sigma-Aldrich (Oakville, ON) unless otherwise stated. Fused-silica capillaries were purchased from Molex Polymicro (Phoenix, AZ). The deionized water was freshly produced by a Millipore Milli-Q UV Plus instrument and had an electrical resistance of ~18 MΩ. DNA sequences were custom synthesized by Integrated DNA Technologies (IDT) (Coralville, IA). 80-nt DNA samples had an additional HPLC purification

cycle to remove truncated oligomers. All synthetic DNA was subjected to a standard desalting by IDT, a procedure which is claimed by the company to remove almost completely all of the non-reacted protective groups and diffusely bound counterions from the DNA samples. All DNA was received as lyophilized pellet and resuspended in the sample/run buffer (50 mM Tris-acetate at pH 8.3) unless stated otherwise. All electrophoresis and NECEEM experiments were performed using 10 μM DNA samples. NanoDrop-1000 spectrometer (Thermo Scientific, Wilmington, DE) was used to verify DNA concentrations by measuring absorbance at 260 nm. Values were calculated based on manufacturer-provided extinction coefficients. For electrophoretic migration studies, a fluorescein labeled ssDNA molecule was employed, with a sequence: 5'- Fluorescein - 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'. For dsDNA migration studies, an unlabelled complimentary strand was separately synthesized, mixed in at 1:1 ratio, heated to 90°C and then slowly cooled down to 4°C. For the experiments with a heterogeneous library a synthetic ssDNA library of the following sequence was used: 5'- Fluorescein -CTT CTG CCC GCC TCC TTC CT -(N40)- AGA CGA GAT AGG CGG ACA CT-3', where N represents a random nucleotide.

In addition to the above-described DNA molecules, electro-dialysis experiments were also performed with double-stranded (ds)DNA extract from herring sperm (Sigma-Aldrich), circular plasmid DNA purified from bacteria by QIAGEN Midiprep Kit, an AlexaFluor 488-labelled ssDNA molecule (5'- Alexa - 488-CTC CTC TGA CTG TAA CCA CGT GCC TAG CGT TTC ATT GTC CCT TCT TAT TAG GTG ATA ATA GCA TAG GTA GTC CAG AAG CC-3'), a non-labeled ssDNA molecule (5'-GGT GGT GGT GGT GGT GGT GTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GGT GGG TGG GTG GGT GG-3'), and a

synthetic ssDNA library (5'- Fluorescein -CTT CTG CCC GCC TCC TTC CT -(N40)- AGA CGA GAT AGG CGG ACA CT-3').

5.2.2. Capillary electrophoresis

All capillary electrophoresis (CE) procedures were performed using a P/ACE MDQ instrument (Beckman Coulter, Mississauga, ON,) equipped with a standard LIF detection system. All capillaries were 80 cm-long (70 cm to the detector) and had an inner radius of 37.5 μm and an outer radius of 180 μm. The Poly(vinyl alcohol)-coated capillary was prepared as described previously **[120]**. Samples were injected into the capillary, pre-filled with the run buffer (which was matched with the sample buffer, 50 mM Tris-Acetate at pH 8.3), by a pressure pulse of 0.5 psi (3.44 kPa) for 11 s, resulting in a plug with a length of 9 mm. Where stated, the run buffer was also supplemented with 120 mM of NaCl. Prior to every run, the coated capillaries were rinsed with a sequence of deionised water and the sample/run buffer. Uncoated capillaries were rinsed with a sequence of 6,000 ppm NaClO (bleach) solution, 100 mM HCl, 100 mM NaOH, deionized water, and run buffer prior to every experiment. During electrophoresis, both inlet and outlet reservoirs contained the sample/run buffer solution. Separations were carried out by an electric field of 375 V/cm. For coated capillaries, electrophoresis was carried out with the cathode at the injection end of the capillary. For uncoated capillaries, electrophoresis was carried out with the anode at the injection end of the capillary; the direction of EOF was from the inlet to the outlet reservoir. The temperature in the cooled region of the capillary was maintained at 15ºC during separations. For experiments with fraction collections, uninterrupted electrophoresis was performed for the first 5 min of the run, after which the collection vial was switched every minute. Eluent was collected into vials containing 10 μL of run buffer. A total of 34 fractions were collected for each experiment. To

prevent fraction cross-contamination, the outlet end of the capillary was momentarily dipped into a reservoir with a large volume of sample/run buffer in between every fraction collection step. Collected fractions were immediately analyzed through qPCR. All NECEEM experiments were conducted in uncoated capillaries, with the same procedures as described above, except for the fact that no fractions were collected, and electrophoresis was performed uninterrupted for 40 min. In NECEEM experiments, 250 nM ssDNA samples contained either 0, 1, or 2 M NaCl.

5.2.3. Quantitative polymerase chain reaction

qPCR reaction mixture was prepared by combining IQ SYBR Green Supermix from Bio-Rad (Mississauga, ON) with unlabelled DNA primers. The nucleotide sequence of the sense primer was 5'-CTT CTG CCC GCC TCC TTCC-3' and the sequence of the anti-sense primer was 5'-AGT GTC CGC CTA TCT CGT CTC C-3'. Two μL of each fraction were mixed with 18 μL of qPCR reaction mixture immediately before thermocycling. Besides the collected fractions, each qPCR experiment also included negative controls (no template control, fraction-collection buffer control and run buffer control), a set of concentration standards (containing from 10^3 to 10^7 template molecules) and a set of 100-times dilutions of the seven fractions with the highest expected DNA concentration. Each qPCR reaction was performed in duplicate. Thermocycling and real-time fluorescence signal collection were performed with iCycler IQ system from Bio-Rad (Mississauga, ON). Raw fluorescence signal data was background-subtracted and amplitude normalized, as described elsewhere **[128]**.

5.2.4. DNA dialysis

For all dialysis procedures, 50 to 200 μL of each DNA sample was prepared at 100 μM concentration, with deionized water as the solvent. Roughly 3 cm-long portions of Spectra/Por 6 dialysis membrane bags (Spectrum Laboratories Inc., Rancho Dominguez, CA) with molecular

weight cut-off value of 25 kDa, were used for all experiments. Prior to each procedure, dialysis membrane bags were soaked in deionized water for 30 min and thoroughly rinsed. The DNA sample was then transferred into the dialysis bag, and clamped off at each end, ensuring that no air bubbles were trapped inside the membrane bags. For passive dialysis, the membrane bags were placed into 500 mL of de-ionized water and incubated for 8 h. The diluent water was exchanged every hour, for a total of 8 times. The sample solution was then transferred from the membrane bag into a test tube. For electro-dialysis, the membrane bags with DNA samples were placed into a Minive Blotter chamber (Amersham-GE Healthcare, Baie d'Urfe, QC) containing 300 mL of de-ionized water. An electric field of 600 V/cm was then applied across the blotter chamber for one minute. At this point, the diluent water was exchanged and the procedure repeated. This procedure was repeated up to 20 times, with visible DNA precipitates usually appearing after 5-7 pulses. The precipitates were then picked up from the membrane bag using a micropipette tip and transferred into a test tube that contained 100 μL of either deionized water or 50 mM Tris-Acetate buffer at pH 8.3. Control electro-dialysis experiments were performed with deionized water and 1 mg/mL of BSA solution instead of the DNA solution. The pH of solutions was determined by depositing small drops of the sample onto Alkacid Test litmus paper ribbon (Fisher Scientific, Pittsburgh, PA, USA).

5.2.5. Solubility of DNA aggregates.

One hundred μL of 50 μM fluorescein-labelled ssDNA solution was prepared and split equally between two dialysis bags to test the influence of counterions on the solubility of DNA precipitate. The samples were concurrently subjected to electro-dialysis in the same electro-blotting chamber. Each of the formed precipitates was transferred into a separate vial containing 50 μL of either 50 mM Tris-Acetate buffer solution at pH 8.3 or deionized water and thoroughly

vortexed. The first set of photographs of the test tubes was taken 1 min after the transfer. The samples were further incubated for additional 24 h at room temperature and the second set of photographs was taken. Finally, 1 μL of 50 mM NaCl solution was added to each sample, to a final concentration of 1 mM of NaCl. Samples were thoroughly mixed by pipetting and photographed for the third time. The fluorescently labeled DNA precipitate was resuspended in 10 mM Tris-HCl buffer pH 8.0 and the fluorescence was measured by Nanodrop-3300 fluorometer (Thermo Scientific, Wilmington, DE) to calculate the efficiency of aggregate formation.

5.2.6. DNA integrity

DNA integrity experiments were performed with two identical 100 μL aliquots of 50 μM DNA. Electro-dialysis was concurrently performed with both DNA samples for five 1-min pulses of a 600 V/cm electric field. At that point, one of the samples was removed from the electro-dialysis chamber, and 15 additional 1-min cycles of electro-dialysis were performed with the remaining sample. The precipitates from both samples were transferred into new vials, both containing 100 μL of 50 mM Tris-acetate buffer at pH 8.3. The supernatants that remained after electro-dialysis were also collected. Samples of the original (non-dialysed) DNA solution, both resuspended precipitates and their supernatants were diluted 100 times and loaded onto a 2.2% agarose gel. Electrophoresis was performed for 30 min at 100 V. DNA molecular weight standards were visualized by Ethidium Bromide staining, while the ssDNA was visualized through fluorescein labeling.

5.3. RESULTS AND DISCUSSION

5.3.1. Non-uniform velocity of DNA

To study the electrophoretic migration patterns of DNA in detail, I performed CE experiments with concurrent detection by LIF and off-line qPCR. For qPCR detection, fractions were collected from the capillary output at 1-min intervals and subjected to amplification all at once. In CE, the net velocity of DNA is a vector sum of the electrophoretic velocity (v_{ep}) and velocity of the EOF (v_{EOF}) (**Figure 5.1, top row**). To differentiate between the two migration velocity components, experiments were carried out in both a coated capillary (with suppressed EOF, **Figure 5.1A**), and in a bare-silica capillary (with a strong EOF directed from anode to cathode, **Figure 5.1B**). Since the total velocity vector (v_{tot}) of DNA has opposite directions in the two types of capillaries, opposing polarities must be applied to ensure proper direction of movement of DNA. A short plug of an 80-nt synthetic fluorescein-labeled ssDNA was injected into a capillary, a high voltage was applied, and migration time of DNA to a detection point was recorded both by both detection methods. As I expected, fluorescence detection resulted in a typical Gaussian-shape peak (**Figure 5.1, middle row**), which suggests uniform velocity of DNA normally dispersed due to peak-broadening phenomena, such as diffusion **[129]**. The dynamic range of fluorescence detection was approximately 3 orders of magnitude. Presenting data in a log-scale did not reveal any additional information. The results were different with qPCR detection, which allowed us to determine the quantity of DNA with a range of 10 orders of magnitude. Presenting the results of qPCR measurements in a log-scale revealed a non-uniform velocity profile (**Figure 5.1, bottom row**). The predominant portion of DNA molecules migrated with a single velocity, normally dispersed to yield a Gaussian distribution. However, there was also a small fraction of DNA molecules that migrated with velocities different from that of the

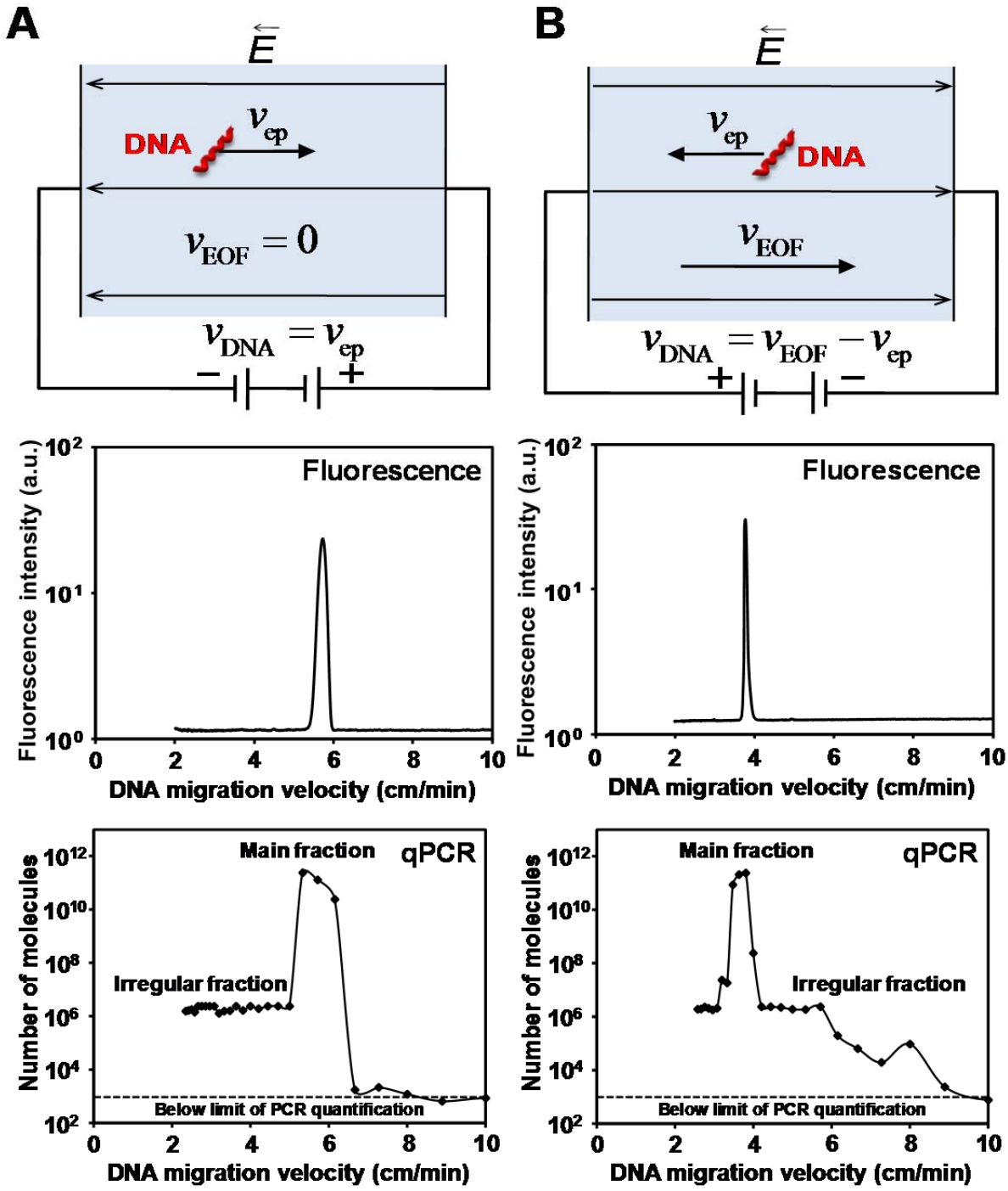

Figure 5.1. Heterogeneous migration of homogeneous DNA in a uniform electric field. Migration of DNA in the absence (**Panel A**) and presence (**Panel B**) of EOF. Top row shows the schematic representations of the DNA mobility vectors. The middle and bottom graphs show electrophoretic migration profiles of ssDNA detected by LIF and qPCR, respectively.

main DNA zone, varying within a several-fold range. This "irregular" fraction of DNA was below the LOD of fluorescence. The length of the DNA molecule did not significantly affect the observed velocity heterogeneity, as similar velocity profiles were observed for a 71-nt and a 120-nt DNA samples (data not shown).

5.3.2. Possible causes for non-uniform velocity of DNA

Strong distortion (fronting or tailing) of DNA sample zones in CE may be caused by: (i) sample diffusion; (ii) the anti-stacking effect; (iii) interaction of DNA molecules with the walls of the capillary; (iv) heterogeneity in DNA polymer lengths; (v) heterogeneity in DNA tertiary structure; and (vi) interaction of DNA with other molecules in the solution. Some of these effects can be refuted immediately as possible explanations for the observed phenomenon. DNA molecules are characterized by slow diffusion coefficients (on the order of 1×10^{-8} cm^2/s). In my experiments, the front of the irregular fraction and the peak of the main fraction were separated by approximately 40 cm. Even with the greatest estimate for DNA diffusion coefficient, calculations show that it would take approximately 1500 days to establish the observed DNA distribution pattern by diffusion alone, while the front of the irregular fraction eluted within 10 min of electrophoresis. Thus, the irregular velocity of DNA cannot be explained by sample diffusion. Neither can it be explained by the anti-stacking effect, as in our experiments the sample electrolyte and the electrophoresis electrolyte were matched and the DNA concentration was negligibly low when compared to the concentration of the buffer. Boundary phenomena, such as slowing down of DNA molecules near the walls of the capillary due to "friction" also cannot explain the observed results: if DNA molecules were slowed down by interacting with the capillary surface, the irregular fraction of DNA would migrate slower than the major fraction regardless of the presence of EOF. However, in the presence of EOF (and

reversed direction of the electric field), the irregular fraction migrated faster than the major fraction (**Figure 5.1A, bottom row**). Lastly, the observed irregular fraction cannot be explained by heterogeneity in DNA polymer lengths (*e.g.* as a result of sample degradation). Matrix-free electrophoresis methods, such as CE, separate molecules based on size-to-charge ratios. DNA molecules of various lengths are poorly resolved in matrix-free CE, as the size-to-charge ratio of DNA is approximately constant across a wide range of polymer lengths **[122-124]**. Generally, the CE mobilities between 120-nt and 15-nt long DNA molecules differ by less than 10%. In the current experiment, however, the mobility of the irregular fraction differed from that of the major fraction by a factor of 3. Furthermore, truncated DNA cannot be detected with the employed qPCR detection method, as shorter sequences would lack the adequate flanking regions required for PCR primer annealing. As a result, all qPCR detected sequences must be of the same length.

The observed velocity profile was analogous to that reported for similar experiments with a highly heterogeneous DNA library containing ~10^{12} different sequences **[48]**. The nature of the velocity profile for the DNA library has never been deciphered, but it was suggested that the conformational heterogeneity of the library was the likely cause of the "heterogeneous" velocity. In my experiments, I studied a single DNA sequence that cannot provide a conformational diversity of a random library, but can still result in multiple thermodynamically stable conformations through inter- or intra-molecular DNA interaction with "itself" **[130]**. To examine whether or not the velocity non-uniformity could be explained by the conformational heterogeneity of DNA, I studied the electromigration of a double-stranded sample of the same DNA sequence (dsDNA). In contrast to ssDNA, dsDNA forms a much smaller variety of structures that primarily consist of the thermodynamically favorable double helix conformation.

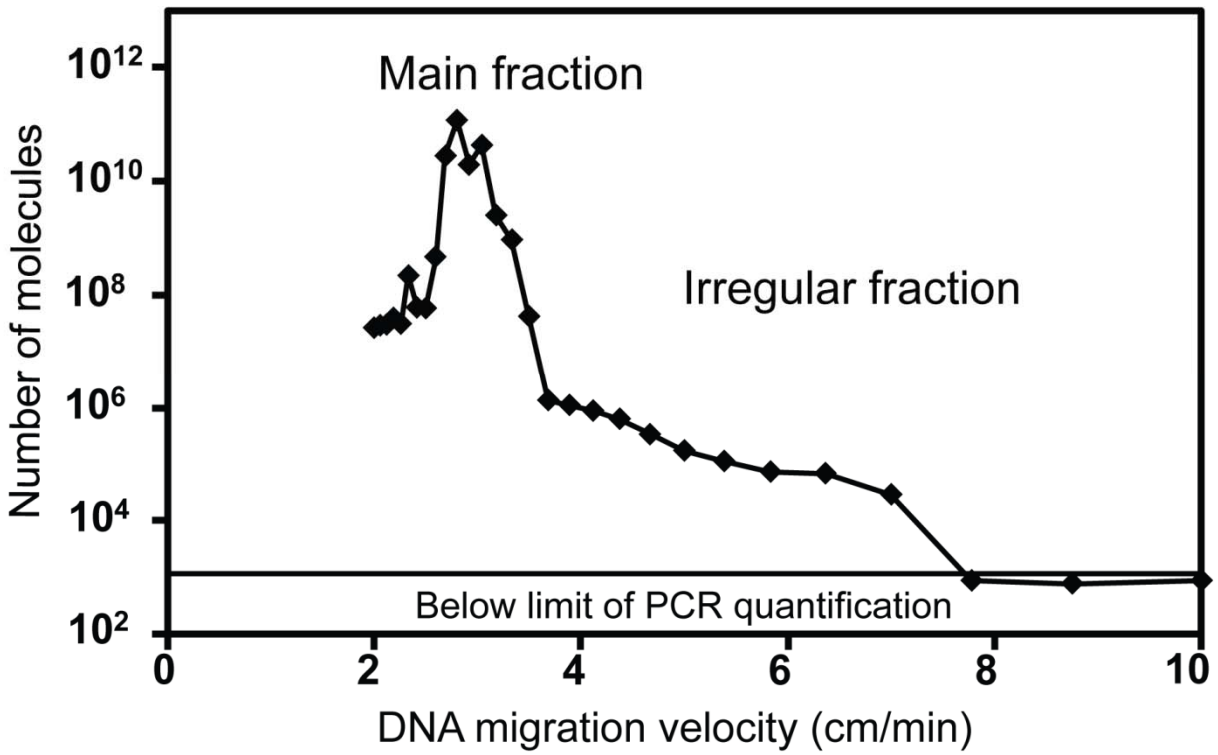

Figure 5.2. Migration profile of dsDNA in a uniform electric field. Despite the decreased heterogeneity of conformations, dsDNA presents a nearly-identical non-uniform velocity profile as ssDNA.

The experiment showed that the velocity profile of dsDNA was also non-uniform (**Figure 5.2**) and nearly identical to that of the ssDNA shown in **Figure 5.1B, bottom row**. This result suggested that non-uniform velocity of DNA cannot be explained by multiple conformations caused by intra- and inter-molecular interactions of DNA.

5.3.3. Interactions of DNA with its counterions

After eliminating all of the alternative explanations for the formation of the irregular fraction, we considered the interaction of DNA with other species in the solution as the possible cause. We noticed that the velocity pattern of DNA in **Figure 5.1, bottom row** resembled those produced in analyses of DNA–protein binding by NECEEM **[50]**. The non-uniform velocity of DNA in NECCEM is caused by the slow dissociation of stable affinity complexes between DNA

and a protein target. While there was no protein present in experiments shown in **Figure 5.1** and **Figure 5.2**, the solution did contain a variety of positive counterions, which could bind to the negatively charged DNA and influence its electrophoretic velocity **[131-135]**. DNA mobility was previously shown to be affected by interactions with cations **[136]**. We thus hypothesized that the observed irregular velocity profile of DNA was caused by dissociation of DNA–counterion complexes.

To produce the observed NECEEM-like velocity profile, a portion of complexes between DNA and counterions must be very stable and slow-dissociating (with half-lives on the order of tens of minutes). DNA can form two general types of complexes with its counterions: diffusely bound complexes and condensation complexes **[132, 137, 138]**. Diffusely bound complexes are unstable, fast-dissociating and predominate in conditions of counterion abundance. In contrast, condensation complexes display higher stability, slower rates of dissociation and occur in conditions of counterion deficiency. The process of DNA counterion condensation was theoretically predicted by Gerald Manning in 1969, and since then has garnered strong experimental support **[138-140]**. Counterion condensation is driven by an excessively high native charge density of DNA molecules. Due to insufficient charge neutralization in conditions of counterion deficiency, the charge density of DNA may exceed a certain threshold value. This causes some counterions to become trapped, or condensed, in close vicinity of DNA, unable to escape through thermal energy alone. The stability of condensation complexes is inversely proportional to their concentration, with the complexes becoming more stable as their concentration decreases. At certain conditions of counterion deficiency, condensation complexes become so strong that their dissociation requires the application of an external force, such as a high-magnitude electric field **[141]**.

Consideration of the counterion condensation theory makes dissociation of DNA-counterion complexes a plausible explanation for the observed non-uniformity of electrophoretic velocity of DNA. Tightly bound metal counterions are likely to be carried over into sample solutions along with DNA from the time of its synthesis. Combined with the more bulky buffer counterions, these create conditions of counterion abundance, in which diffusely bound complexes comprise the vast majority of DNA–counterion interactions. Upon electrophoresis, diffusely bound fast-dissociating counterions separate from DNA in a matter of seconds, resulting in the formation of a large fraction of DNA molecules migrating with uniform electrophoretic velocity. Furthermore, electrophoretic separation of diffusely bound counterions creates counterion deficiency around DNA, leading to the formation of more stable condensation complexes. Electric field-induced slow dissociation of such condensation complexes results in a small fraction of molecules with different mobilities. Importantly, the more condensation complexes dissociate from DNA, the more stable the remaining complexes become. As complex dissociation events are "probability-controlled", dissociation of more stable condensation complexes results in a wide irregular fraction of DNA. As such, this hypothesis is consistent with the shape and the small relative size of the irregular fraction.

We reasoned that if this hypothesis is true, then the extent of DNA velocity heterogeneity can be regulated by the concentration of DNA counterions in the solution. To increase the amount of counterions I simply added NaCl to the solution of DNA and incubated the mixture to reach equilibrium in the interaction between DNA and Na^+. The experiment was similar to NECEEM in a sense that the run buffer did not contain NaCl and the injected DNA-Na^+ complexes were promoted to dissociate when the excess of Na^+ was separated from the complexes.

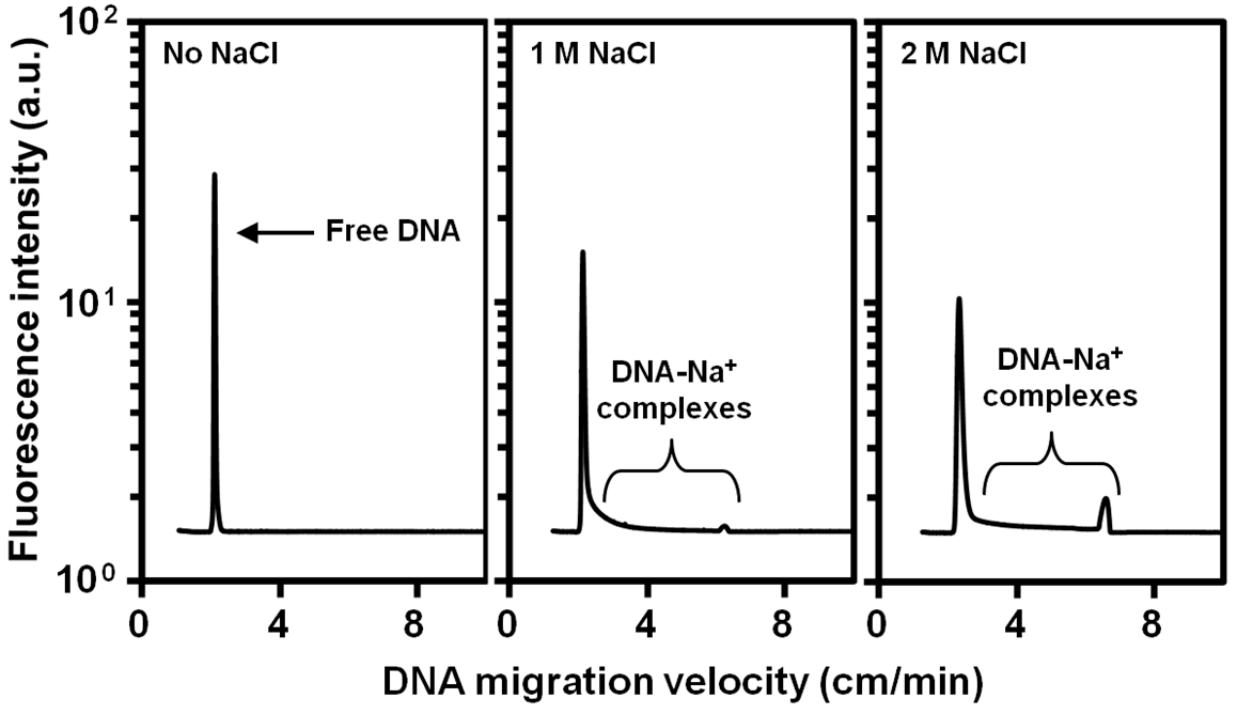

Figure 5.3. Effect of Na^+ counterions on velocity non-uniformity of DNA. At high concentrations of Na^+ counterions, the velocity non-uniformity of DNA increases to the point that it becomes detectable by LIF. Also, note the formation of the highly stable DNA–Na^+ complex.

These experiments clearly showed that the irregular fraction of fast-moving DNA increased with the rising concentration of NaCl in the equilibrium mixture. The increase was so significant that the irregular fraction was detectable by fluorescence, which even showed a peak of stable fast moving DNA-Na^+ complexes (**Figure 5.3**). Similar results were obtained with other monovalent (Li^+ and K^+) and divalent (Mg^{2+}, Ca^{2+} and Ni^{2+}) cations **(**data not shown**)**, confirming that DNA can form stable slow-dissociating complexes with metal counterions.

Next, I tested the effects of decreasing the DNA counterion concentration in solution on the size of the irregular fraction. First, I attempted to decrease the concentration of only the diffusely bound counterions by passive dialysis. I dissolved synthetic fluorescein-labeled 80-nt ssDNA in

deionized water and subjected it to passive dialysis against deionized water for 8 h, with regular replacement of the diluent. After dialysis, the content of the membrane bag was diluted with the run buffer and the resulting solution of DNA was analyzed in CE to measure its velocity profile. The resulting velocity profile was similar to the one of the DNA sample prior to dialysis (**Figure 5.4A, red and blue traces**). This result suggested that the presence of fast-dissociating diffusely bound counterions does not cause the observed DNA velocity heterogeneity.

Decreasing the amount of condensed counterions is not as a trivial task as increasing it. Due to their remarkable stability, DNA–counterion condensation complexes are difficult to dissociate by conventional deionization methods, such as filtration, precipitation, or passive dialysis. However, dissociation of condensed counterion complexes may be achieved more efficiently by application of an external force **[141]**. We designed a simple procedure in which condensed counterions can be first dissociated from DNA by a high-strength electric field and, then, to prevent re-condensation, can be permanently removed from solution by dialysis. An experimental setup is schematically shown in **Figure 5.5A**. The commercially synthesized and desalted DNA sample was dissolved in deionized water, and then placed into a semi-permeable (transparent for small ions, but non-transparent for DNA) dialysis bag (**Figure 5.5B**). The bag was then placed into an electroblotting chamber which incorporated two electrode plates capable of creating a uniform electric field across the chamber. The chamber was filled with deionized water to act as the dialysis diluent. The absence of salts in the sample and diluent allowed me to use a high electric field without overheating the setup. Several 1-min long pulses of a constant electric field of 600 V/cm were then applied across the dialysis bag. The diluent was replaced

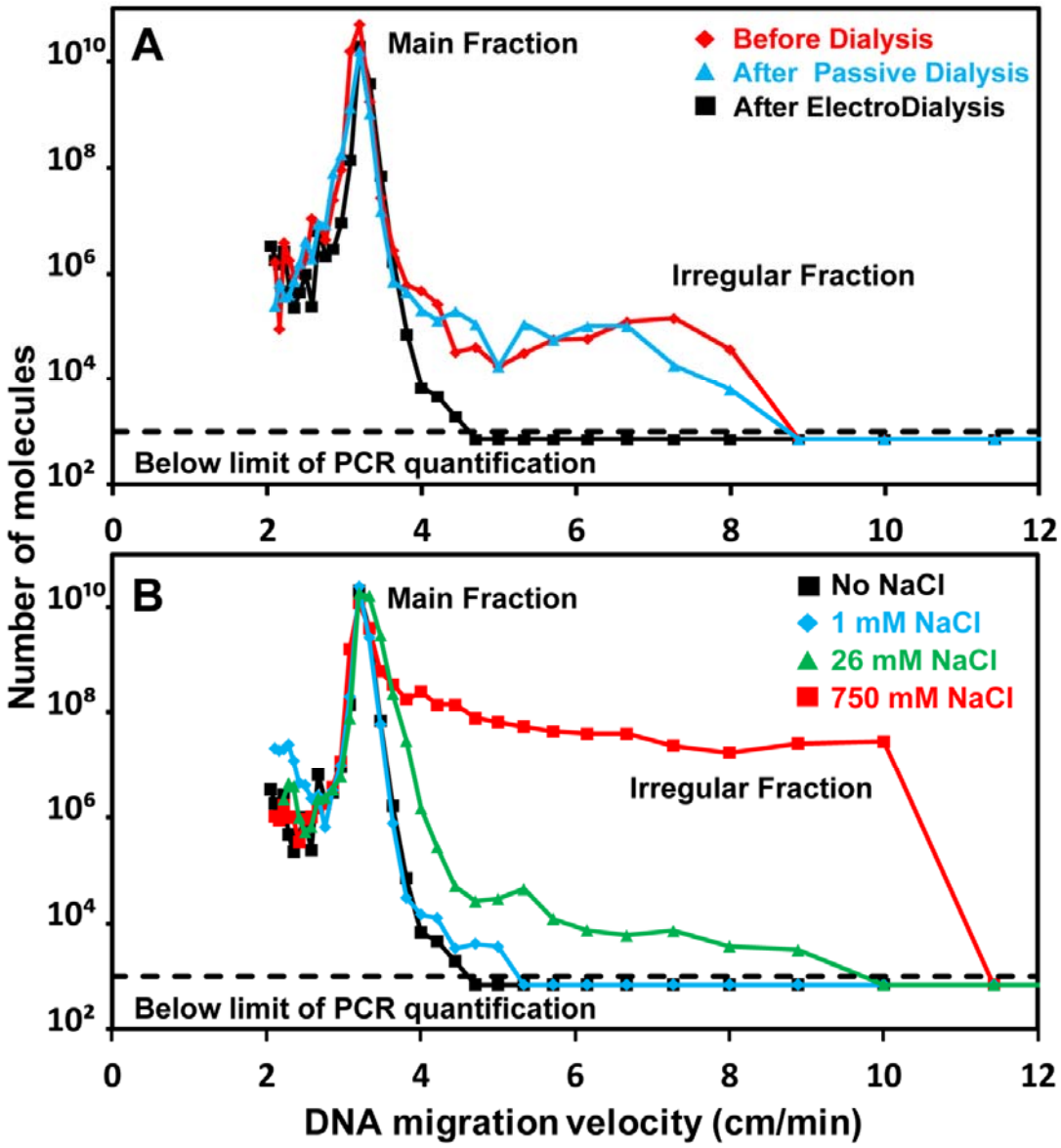

Figure 5.4. The effect of counterion depletion and reintroduction on DNA migration. Panel A: the red trace corresponds to DNA prior to dialysis (control experiment), the blue trace corresponds to DNA that was dialyzed against deionized water in the absence of electric field, and the black trace corresponds to DNA that was dialyzed against deionized water in the presence of electric field of 600 V/cm. **Panel B**: Reintroduction of 0, 1, 26, or 750 mM of NaCl to an electro-dialyzed sample of DNA before analysis by CE with qPCR detection (black, blue, green, and red traces, respectively). Both the size of the DNA fraction with irregular migration velocities and the range of the irregular velocities showed a strong dependence on Na^+ concentration.

with fresh deionized water after each pulse to ensure efficient removal of the dissociated counterions. To our surprise, after just five 1-min pulses of the electric field, we observed a previously unknown phenomenon: DNA precipitated out of solution **(Figure 5.5C)**. This unexpected behavior of DNA will be discussed in detail in the next section.

To test if the removal of the condensed counterions has any influence on the migration pattern of DNA, the electro-dialysis precipitated DNA was resuspended in the run buffer and the resulting solution was analyzed by CE. This time, the result was markedly different: the fraction of irregularly moving DNA was drastically decreased (**Figure 5.4A, black trace**). The same experiment was repeated after incubating DNA at 4°C for 96 hours to ensure that the equilibrium of DNA with buffer ions was established, and the observed DNA velocity profile was not affected by incubation time. The effect of counterion depletion was reversible. Adding NaCl to the electro-dialyzed (*i.e.* counterion deficient) DNA increased the irregular fraction of DNA (**Figure 5.4B**) in a Na^+ concentration-dependent manner. The obtained results allowed us to

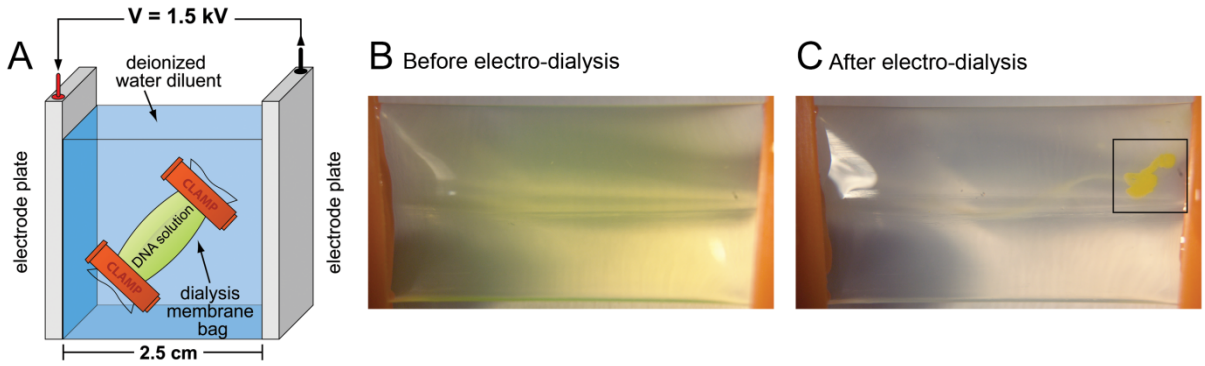

Figure 5.5. Removal of DNA condensed counterions by electro-dialysis. Panel A: schematic diagram of the electro-dialysis setup. **Panels B and C** show photographs of 80-nt synthetic ssDNA in deionized water inside of a semi-permeable membrane bag before and after electro-dialysis, respectively. The green color of the DNA solution and the yellow color of the DNA precipitate are due to the presence of the fluorescein label on the DNA.

make two conclusions. First, that an electric field is required for efficient dissociation of the condensed counterion–DNA complexes. Second, the size of the irregular fraction of fast-moving DNA decreases with the diminishing amount of counterions condensed on to the DNA.

5.3.4. DNA precipitation by electro-dialysis

As mentioned in the previous section, electro-dialysis of DNA had resulted in its precipitation from the solution. This is a highly unusual observation, as removal of counterions is expected to increase the repulsion between DNA molecules by intensifying the charge density of individual chains. The aggregates formed as an amorphous structure on the membrane wall closer to the anode and easily detached from the membrane by brief shaking of the bag after the electric field was turned off. When detached, the aggregates slowly sunk due to gravity. The pH values of the solutions inside and outside the dialysis bag were around 3 and 6 respectively and did not change significantly after the electro-dialysis. With further examination, the obtained DNA precipitates displayed some remarkable properties. The precipitates were stable and did not re-dissolve after the electric field was turned off. Furthermore, the DNA precipitates remained insoluble after they were transferred into a fresh volume of deionized water, even after vigorous mixing and 24-h incubation. However, the precipitates did readily re-dissolve (within 1 min) when placed into a buffer solution or a salt-containing solution **(Figure 5.6).** Well-visible DNA aggregates formed within first 5 min in the electric field, and 85% of the DNA precipitated within 20 min (**Figure 5.7**). No precipitation had occurred in control experiments in which the DNA solution was replaced with either deionized water or a solution of BSA protein. Agarose gel electrophoresis of the buffer-dissolved precipitate showed a single band, with fluorescent properties and a

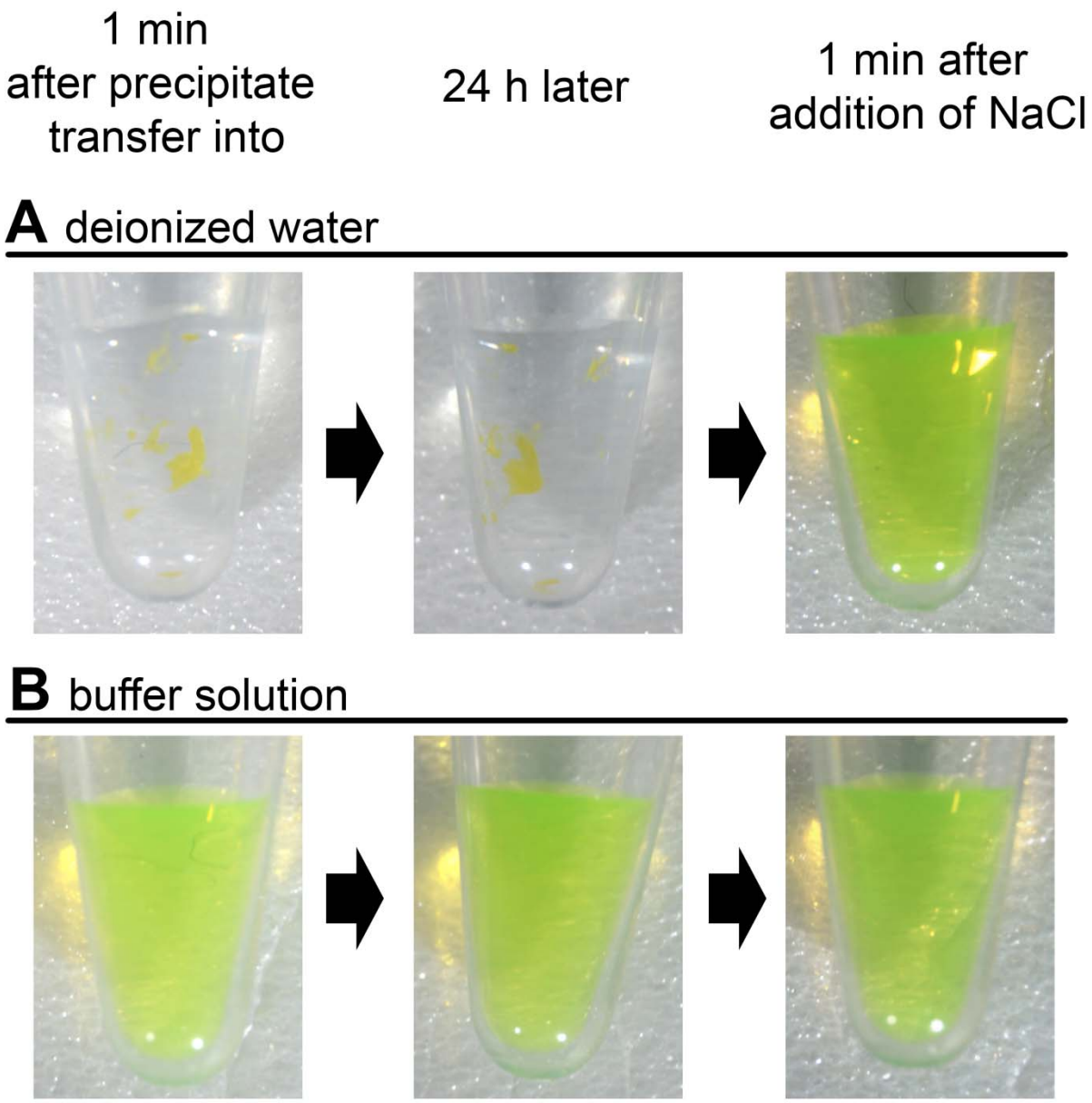

Figure 5.6. Ion-dependent solubility of electro-precipitated DNA. Photographs of two separate DNA precipitates after being transferred into deionized water (**Panel A**) and 50 mM Tris-acetate buffer solution at pH 8.3 (**Panel B**). The photographs were made 1 min and 24 h after the transfer and 1 min after subsequent addition of 1 mM NaCl. The samples were agitated by vortexing after the precipitate transfer and after the NaCl addition. The yellow color of the DNA precipitate and the green color of the DNA solution are due to the fluorescein label on the DNA.

migration pattern identical to those of the original DNA sample (**Figure 5.7**). This result suggests that electro-precipitation did not affect DNA integrity. Precipitation was observed with different types of DNA, including several fluorescently labeled ssDNA of different lengths and sequences, non-labeled ssDNA, dsDNA of various lengths from herring sperm extract, and a purified plasmid DNA.

Interestingly, electro-dialysis of DNA samples has been previously performed by other groups; however, no DNA precipitation was observed in those experiments **[142, 143]**. The reason, most likely, was the use of ion-containing dialysis diluents which would have prevented the permanent depletion of condensed counterions from DNA. Transient DNA aggregation under the influence of a strong electric field has been observed in the past in elegant experiments performed by Maestre et al, and later by Viovy *et al.* and Doyle *et al* **[144-146]**. DNA aggregation occurred only in solutions with low-ionic concentrations and at electric field strengths above a certain threshold. In contrast to our precipitate, the aggregates observed by

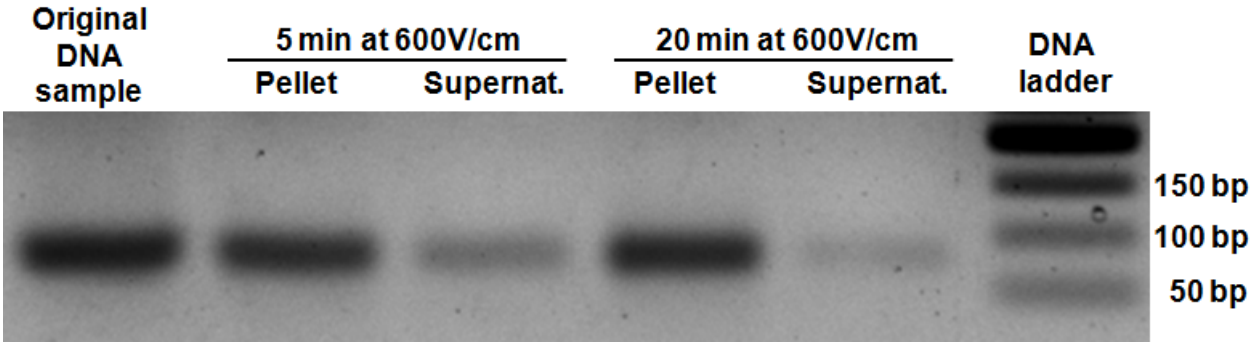

Figure 5.7. Agarose gel electrophoresis analysis of DNA integrity after electro-precipitation. From left to right the samples are: the original solution of 80-nt ssDNA, the precipitate (re-dissolved in 50 mM Tris-acetate buffer), and the supernatant after 5-min and 20-min of electro-dialysis. The rightmost lane shows the migration of DNA molecular weight standards.

Maestre, Viovy and Doyle were only stable in the presence of the electric field and spontaneously dissociated upon its removal. Most likely, the instability of the aggregates was due to re-association of DNA with counterions which could not be permanently removed in the absence of dialysis. Viovy and Doyle proposed mechanisms for the formation of aggregates that involved redistribution of counterions along DNA molecules **[145-148]**, but these mechanisms do not explain the stability of aggregates observed by us.

If viewed purely through simple electrostatic consideration, our results are perplexing and counterintuitive. DNA counterions reduce electrostatic repulsion between DNA molecules by screening the negative charges in their backbones **[149]**. Therefore, their removal should increase repulsion between DNA molecules, making them more soluble. Accordingly, aggregation of DNA is not intuitively expected under the conditions of counterion deficiency. To explain the observed phenomenon a more comprehensive consideration of counterion theory is required. The Oosawa-Manning counterion condensation theory, mentioned in the previous section **[138, 150]**, describes the existence of two distinct subpopulations of DNA counterions: diffusely bound and condensed. Diffusely bound counterions behave as a gas-like cloud, separated from DNA by the entire Debye sphere. Condensed counterions, on the other hand, are much more closely associated with DNA. They occupy a rigid volume within the thickness of the first few shells of DNA hydrating water molecules. The formation of these exceptionally stable counterion complexes is driven by the necessity to reduce the charge density of DNA below a certain threshold value. In solutions with severe counterion deficiency, when the diffusely bound counterion cloud is sparse and its charge-screening effects are low, the condensed counterion layers of two neighboring DNA molecules may exert an influence upon each other. Interestingly, calculations show that this mutual influence of condensed counterion layers may result in

attraction between DNA molecules **[151, 152]**. The precise molecular-level mechanism of this short-range non-electrostatic interaction is yet unknown, but it likely involves entropic stabilization of the DNA−counterion systems through increasing the volume available to condensed counterions. This theory has been proposed to explain the formation of fluid polyion clusters observed in low-salt solutions **[153]**. Our observations seem to be consistent with this theory. In my electro-dialysis experiments, severe counterion deficiency was established to provide the required depletion of the diffusely bound cloud. Furthermore, the electrophoretic movement of DNA increased its local concentration at the dialysis membrane, making it more likely for DNA molecules to come sufficiently close for the short-range forces to take effect. Polarization and reorientation of DNA condensed counterion complexes in the electric field may also play a role in facilitating the formation of the entropy-stabilized attractive forces. The remarkable stability of the aggregates in our experiments, however, suggests that additional physical phenomena may be involved. Combined with the currently reported strong dependence of aggregate solubility on the addition of external ions, it may be suggested that a partial depletion of the condensed counterion layer also takes place during electro-dialysis. It is not clear what happens with DNA if its charge density grows beyond its threshold value, but perhaps, the observed aggregation is one of its manifestations. These conclusions are in line with the effects of electro-dialysis on electrophoretic migration patterns of DNA, described in the previous section, and together these two observations provide strong support for our hypothesis that counterion condensation is responsible for the non-uniform velocity of DNA in high-magnitude electric fields.

5.3.5. Improving partitioning efficiency of NECEEM without desalting

The desalting-based approach for elimination of the non-uniform migration patterns of DNA, described in **Section 5.3.4 on page 115**, helps us understand the cause of the phenomenon, but is impractical for the purposes of aptamer selection. Affinity and specificity of DNA aptamers stem from their ability to take on a variety of stable tertiary structures, such as the G-quadruplex **[154]**. Counterions play a critical role in the formation of these stable structures and thus affect the ability of DNA sequences to bind to a given target. Furthermore, salt ions are often introduced into aptamer selection mixtures from the target samples. As the presence of salt ions is often required to maintain the stability and native conformation of certain targets, such as proteins, their desalting is undesirable. If we cannot avoid the use of salts in aptamer selection, then we cannot eliminate DNA counterion condensation. As a result, achieving the "perfect" NECEEM separation efficiency in the context of aptamer selection may be impossible.

While we may not be able to completely eliminate the non-uniformity of DNA migration, it is still beneficial to decrease the effects of this phenomenon to the best of our abilities. I have devised another approach to reduce the contamination of the complex fractions in aptamer selection by the irregularly migrating DNA, which was inspired by my findings described in Chapter 4. This approach involves the modulation of the velocity of EOF by the addition of NaCl to the running buffer. By increasing the ionic strength of the run buffer, the μ_{EOF} can be slowed down and the separation between the non-bound DNA and the protein-DNA complex (which we need to collect) increased dramatically. While the irregular fraction of DNA will still be present in this context, it will be smeared over a larger span of a capillary, meaning that a smaller amount of molecules will overlap with a narrow complex collection window. To explore this option, I used a model interaction system of MutS protein and its aptamer. NECEEM separation

was performed in a buffer either without additives or with an added 120 mM NaCl. As predicted, the addition of NaCl dramatically improved the separation between the MutS-aptamer complex and the non-bound DNA **(Fig. 5, red traces)** (with NaCl, DNA migrated so slowly that it does not appear on the electropherogram). To estimate the number of DNA molecules traveling as a complex, and as non-bound irregularly migrating molecules in the vicinity of the complex, 6 minute-wide fractions were collected and quantitated using qPCR. The same was done with the samples which excluded the MutS protein (and, thus, the complex) to estimate only the irregularly migrating DNA in the same collection window **(Fig. 5, black traces).** As expected, the qPCR results for the MutS complex-containing fractions were quite similar, with 4×10^7 molecules ± 2% in the fraction collected under conventional separation conditions, and 2×10^7 molecules ± 17% in NaCl-containing buffer. The 2-times difference can be explained by longer residence time of the complex in the capillary (which allows for more of the complex to dissociate), and by the fact that the aptamer might have a lesser affinity for MutS in high salt conditions, given that it was selected in a conventional buffer. The difference between MutS complex-devoid fractions was much more significant, with 2×10^6 molecules ± 3% in the fraction collected under conventional separation conditions, and 8×10^4 molecules ± 2% in NaCl-containing buffer. With a 23-times difference in the amount of collected irregularly migrating DNA molecules, the new approach yielded a 12-time improvement in the partitioning efficiency of the complex and contaminating non-bound DNA. Further improvements in this efficiency may be possible by carefully optimizing the separation conditions, the width of the collection window, and the stringency of selection for a given protein target of choice.

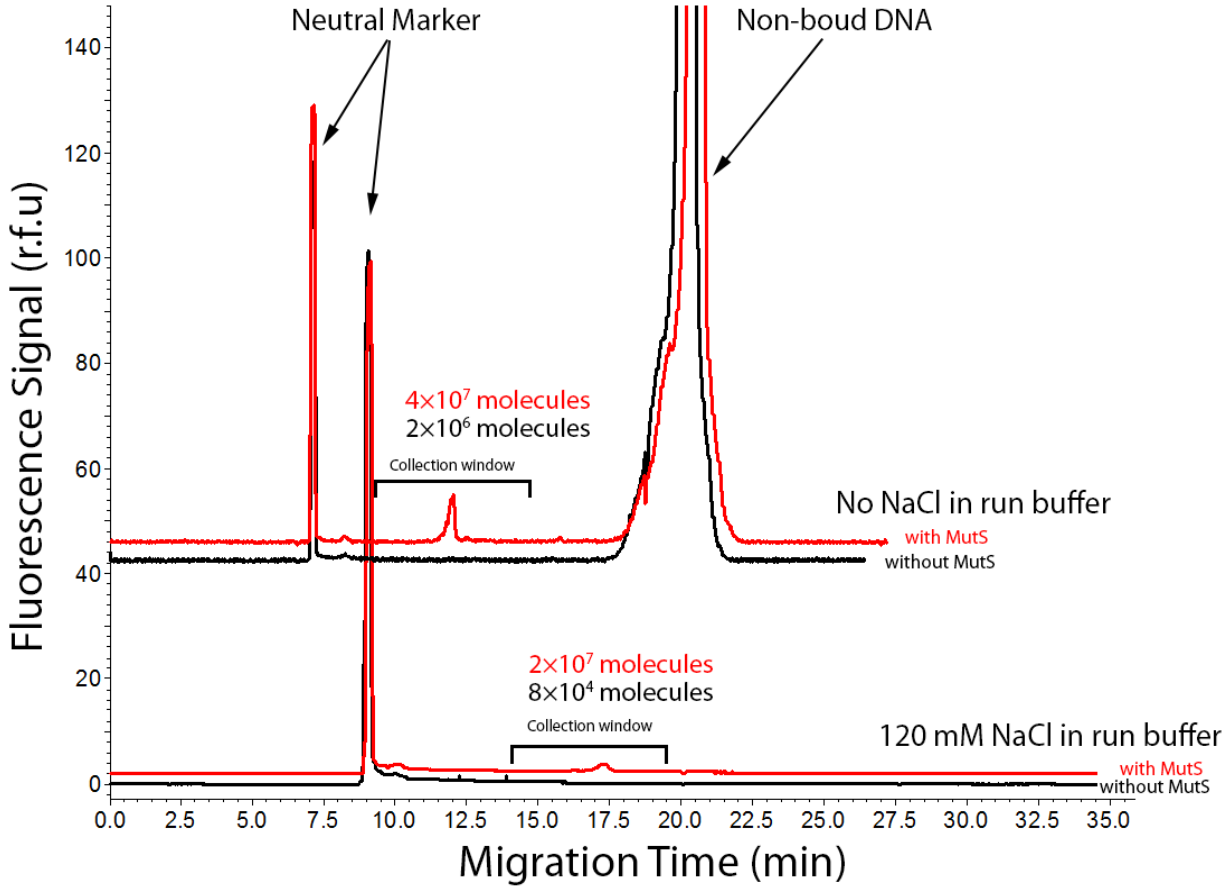

Figure 5.8. Improving the partitioning efficiency of NECEEM by modulation of μ_{EOF}. MutS aptamer was subjected to NECEEM either in the presence (red traces) of absence (black traces) of MutS protein, either in run buffer (top set of traces), or in the same buffer with added 120 mM NaCl (bottom set of traces). Six minute-wide fractions were collected around the expected elution time of the MutS-aptamer complex and quantitated using qPCR. Slowing down of the EOF resulted in a 12-time improvement in partitioning of complex from contaminating DNA in the NaCl-containing buffer.

The newly proposed approach does not yield the coveted "ideal" separation, but it still provides an improvement that can tip the chances of a given SELEX project from failure to success. Evidently, this approach is only suitable for positively charged and neutral targets, and may be detrimental when applied to negatively-charged targets. Modulation of μ_{EOF} by addition of salts also limits the robustness of the method, as we are no longer free to adjust the ionic strength of

the solution to any desired value. Interestingly, the concentration of NaCl used in the described experiment is very close to concentrations in human physiological fluids, making this method suitable for aptamers intended to be used in such environments.

5.4. CONCLUSIONS

In summary, our results strongly suggest that: (i) DNA can form very stable complexes with at least a fraction of counterions, (ii) these complexes can be dissociated by an electric field, and (iii) the observed non-uniform velocity of DNA is caused by electric field-induced slow dissociation of these stable complexes. The first two points are in good agreement with Manning's theory of counterion condensation on polyions **[138, 141]**. Many aspects of this theory were previously difficult to examine due to the lack of suitable experimental techniques. Our results suggest that CE with qPCR detection, as well as electro-dialysis against deionized water, can serve as powerful tools for testing many conclusions of Manning's theory.

The interactions of DNA with ions are of a great importance in nature **[135, 155-157]**. The presence of an electric field in cells in close proximity to DNA has been observed **[158-160]**. Moreover, the shielding effect of both mono- and di-valent cations was shown to modulate the strength of DNA–DNA or protein–DNA interactions occurring in live cells **[156, 158]**. Our finding of electric field influence on DNA–counterion interactions has a potential to be biologically significant, since metal ions on the DNA are involved in DNA biochemistry in living cells.

The non-uniform velocity of homogeneous DNA in a uniform electric field may have many important practical implications, beyond aptamer selection. It may be a potential source of artifacts in DNA probe-based ultra-sensitive analyses of proteins in the presence of electric field, when interactions occur in near-physiological buffers, with relatively high ion content **[161]**. In

pharmaceutics, our technique may help in both analysis and purification of nucleic acid-based drug components, since aptamers are now becoming more widely used in medicine **[162, 163]**.

While our findings show that the maximum efficiency of NECEEM partitioning may not be attainable in the context of aptamer selection, it is still better than most alternative (filter- or HPLC-based) partitioning approaches. Our comprehensive understanding of the causes of non-uniform migration patterns of DNA in CE has allowed us to find an alternative approach to improve the partitioning efficiency of NECEEM by another 10-fold.

Chapter 6. Concluding Statements

6.1. OUTCOMES AND IMMEDIATE APPLICATIONS

Over the course of my research project, I have developed a number of novel and enabling tools that make NECEEM a more practical method. I have uncovered the non-linear dependence of errors in K_d measurements on the initial concentrations of the interacting molecules, and then developed a new approach that allows the optimization of these initial concentrations around the value of K_d without its *a priori* knowledge. Further, I have systematized the complex interrelations between the experimental factors of NECEEM and have developed an original approach to optimization of these factors in an automated manner. In my research, I have also made discoveries regarding some previously unknown behaviors of DNA in the context of strong electric fields, and have applied this newfound knowledge to improve electrophoresis-based analysis of this important biological polymer. I have discovered that the analysis of DNA samples in CE with physiological buffers is affected by a number of detrimental effects that collectively prevent its detection. To address this issue, I have applied a pressure-assisted KCE approach, which has enabled, for the first time, the analysis of DNA interactions by KCE in physiological conditions. I have also discovered that the condensation of counterions on DNA, and their subsequent dissociation under the influence of a strong electric field, is the cause of the unusual electrophoretic migration of DNA in CE. Lastly, I have discovered a new surprising behavior of DNA in the conditions of severe counterion efficiency, in which DNA molecules precipitate out of the solution as a result of counterion-sharing.

KCE methods possess a rare quality in that their underlying processes can be closely described by deterministic physical laws. The ensuing ability to accurately model KCE experiments has empowered us to study the behavior of its systematic errors in great detail. This

is a fairly unique situation, as most other kinetic techniques incorporate physical processes that cannot be modeled deterministically, thus, preventing the use of *in silica* simulations for a comparable level of analysis. Our ability to understand the behavior of the systematic error in NECEEM, and to minimize it in an objective manner is a significant advantage. With the availability of the accuracy-improving approaches, described in **Chapters 2** and **3**, the overall quality of NECEEM data should improve, thus, increasing the confidence of the analytical community in the method. Application of the developed tools in future NECEEM studies will reveal how the ruggedness and robustness of the method is affected. As it stands, NECEEM has the potential to become one of the "benchmark" analytical techniques, which is used for cross-validation of the accuracy of other methods.

Besides improving the accuracy of NECEEM, the optimization approach described in **Chapter 3** also has an interesting added benefit of increasing the versatility of the method. In the past, an interacting pair, such as the one depicted in **Figure 3.6,** would likely be written off as "too slow" for kinetic NECEEM analysis. The new optimization approach enables the users to push the capabilities of their CE instruments to their limits, extending the scope of the studied BMIs with characteristic equilibration times ranging from a few seconds to a few hours. This is in good concordance with the pressure-assisted NECEEM modification, described in **Chapter 4**, which also improves the versatility of NECEEM by enabling the study of BMIs in high-conductivity electrolytes, such as physiological buffers. Poor compatibility of KCE methods with physiological fluids has been a long-standing criticism of the analytical community toward the methodology, which the presented developments should abate. I look forward to the wider application of the developed pressure-assisted modification of NECEEM in studies of BMIs in

more realistic and biologically-relevant conditions, to see if the agreement of the technology with other methods improves as a result.

The most exciting applications of the developed tools, in my view, lie in the field of aptamer selection. The use of the new NECEEM optimization approach in SELEX will enable its users to maximize the CE separation power (making it easier to define collection windows), and will give them the ability to easily define the desired kinetic properties of the resulting aptamers. The use of physiological buffers during selection will likely improve the transition of selected aptamers for usage *in vivo*, for such applications as whole-body imaging and drug delivery. The main benefit to aptamer selection, of course, is expected from the partitioning efficiency-improving strategies, described in **Chapter 5**. As stated previously, the use of physiological concentrations of salt can also be used to our advantage by improving the partitioning efficiency. Over time, with the application of the two presented strategies into CE-SELEX projects, it will become apparent whether the increased partitioning efficiency yields the desired reduction in the number of SELEX cycles, and how the developed approaches affect the stringency of the selection process, and the diversity of the generated aptamer pools. As a separate subject, a better understanding of the DNA precipitation by electro-dialysis phenomenon is required, as well as its potential practical applications in DNA pre-concentration and purification.

6.2. LIMITATIONS AND FUTURE DIRECTIONS

Major factors that still considerably limit the versatility of NECEEM, and other KCE methods, are (i) the lack of efficient strategies for dealing with adhesion of analytes to the surface of capillaries, (ii) poor efficiency of CZE in separating molecules with a small difference in size-to-charge ratio, and (iii) inadequate capacity for analysis of complex interactions.

Adhesion of analytes to the capillary surface affects their migration patterns and can influence the accuracy of K_d and k_{off} measurements by distorting the shapes of electropherogram features or by depleting the effective concentrations of the analytes. In most cases, when an analyte is found to strongly interact with the capillary walls, its analysis by KCE is simply abandoned. As interactions of biomolecules with solid surfaces are common, this tendency restricts the range of BMIs that can be studied by NECEEM. The most efficient way of dealing with the effects of surface adsorption of analytes in CE is by coating the surface of the capillary by a passivating substance. A variety of CE coatings are currently available **[164, 165]**; however, most of these coatings significantly slow down, and even reverse, the velocity of EOF. This often poses a difficulty, as both positive and negative species can no longer be detected in the same experiment. New experimental strategies and extensions to NECEEM theory would have to be developed in order to facilitate kinetic analysis in EOF-supressing coatings. The pressure-assisted modification of NECEEM, described in **Chapter 4**, can be useful in this regard, as it can supplement the decreased velocity of the EOF; however, it is only applicable to "static coatings", which are covalently linked to the surface of the capillary. "Dynamic coatings", which rely on electrostatic or hydrophobic interactions with the capillary surface can be disturbed by the pressure-driven flow during the analysis, and, thus, will require a different approach to be developed. Most likely, such an approach will require the initial sample mixture position to be shifted to the middle of the capillary, followed by precisely timed separation of the oppositely-moving analytes, and their subsequent pressure-propagation to the point of detection. Alternatively, a multi-detector approach may be feasible for such an application as well. Another issue that currently limits the use of coatings in KCE is the lack of a systematic means for selecting an appropriate coating for a given application. Without such a system, the study of

every new BMI model requires a separate, often lengthy, investigation of the suitability of various coatings for its analysis. There is a need for a concentrated effort to systematize such investigations, which will take into account the efficiency of each given coating in passivating the adhesion of different classes of biomolecules, the behavior of the coating in the electric field, its effects on EOF and heat generation, and possible interferences with the measured interactions. To facilitate such investigations, novel efficient CE-based methods for studying surface adsorption of molecules are required, with some progress in this field already achieved **[120, 166, 167]**.

Poor efficiency of CZE in the separation of molecules with a small difference in size-to-charge ratio restricts its applicability to such important types of BMIs as protein-protein and protein-small molecule interactions. While derivatization of interacting molecules by charged or size-increasing groups may help in their separation, this approach can introduce inaccuracies and bias into the measurements, contradicting the premise of KCE methods. The use of mobility-shifting affinity selectors introduces similar issues. A promising approach to addressing the separation efficiency issue lies in the transfer of KCE theory onto other modes of separation. As an example of this, the data-extraction approaches for KCE have recently been adapted for use with size-exclusion chromatography, enabling the analysis of protein–small molecule interactions **[168]**. Conceptually, NECEEM data analysis tools can be adapted to any one-dimensional separation technique, such as capillary gel electrophoresis or analytical ultracentrifugation. One important requirement is that the employed separation technique cannot be based on preferential interactions of the analyte molecules with a stationary phase (*e.g.,* affinity or reverse phase chromatography), because such interactions might interfere with the interactions of the studied molecules. The developed optimization approach, described in

Chapter 3, may be useful in adapting NECEEM theory to new modes of separation as it defines useful analytical performance criteria.

At the current stage, kinetic measurements by NECEEM are mostly applied to simple interacting systems characterized by a 1:1 stoichiometry. As mentioned in **Section 1.3.2**, some progress has been made into expanding the theoretical basis of NECEEM to enable the analysis of higher-order stoichiometry **[56, 57]**. To enable analysis of BMIs in a more holistic and comprehensive manner, the future development of NECEEM, and KCE methodology in general, must include such phenomena as cooperativity and allosteric regulation of binding. In my opinion, the most promising avenue toward achieving this goal lies with further development of numerical modeling tools, such as the one described in **Chapters 2** and **3**. Increasing the complexity of such computational approaches is much simpler than the development of new analytical mathematics-based theory. As a result, a given set of experimental data can be tested against a variety of interaction mechanism models to determine the one that is most suited. Ultimately, in my opinion, the best way to study complex BMIs is by concurrent numerical analysis of data produced by different methods from the KCE toolbox.

LIST OF PUBLICATIONS

14. Kanoatov, M.; Mehrabanfar, S.; Krylov, S.N. Systematic approach to optimization of experimental conditions in nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry*, **2016.** 88 (18), 9300–9308.

13. Kanoatov, M.; Krylov, S.N. Analysis of DNA in phosphate buffered saline using kinetic capillary electrophoresis. *Analytical Chemistry,* **2016.** 88 (14), 7421–7428.

12. 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 (5), 3099–3106.

11. Agostino, F.; Cherney, L.T.; Kanoatov, M.; Krylov, S.N. Reducing pH gradients in free-flow electrophoresis. *Analytical Chemistry*, **2014.** 86 (12), 5656–5660

10. Kanoatov, M.; Cherney, L.T.; Krylov, S.N. Extracting kinetics from affinity capillary electrophoresis (ACE) data: a new blade for the old tool. *Analytical Chemistry,* **2014.** 86, 1298-1305.

9. Ghasemi, F.; Wegman, D.W.; Kanoatov, M.; Yang, B.B.; Liu, S.K.; Yousef, G.M.; Krylov, S.N. Improvements to direct quantitative analysis of multiple microRNAs facilitating faster analysis. *Analytical Chemistry,* **2013.** 85, 10062-10066

8. * Musheev, M.U.; Kanoatov, M.; Retif, C.; Krylov, S.N. Stable DNA aggregation by removal of counterions. *Analytical Chemistry,* **2013.** 85, 10004-10007

7. * Musheev, M.U.; Kanoatov, M.; Krylov, S.N. Non-uniform velocity of DNA in a uniform electric field: the consequence of electric-field-induced slow dissociation of highly-stable DNA-counterion complexes. *Journal of the American Chemical Society,* **2013.** 135, 8041-8046

** – equal contribution with the first author*

6. Kanoatov, M.; Retif, C.; Cherney, L.T.; Krylov, S.N. Peak-shape correction to symmetry for pressure-driven sample injection in capillary electrophoresis. *Analytical Chemistry,* **2011.** 84, 149-154

5. Cherney, L.T.; Kanoatov, M.; Krylov, S.N. Method for determination of peak areas in nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry,* **2011.** 83, 8617-8622

4. Kanoatov, M. and Krylov, S.N. DNA adsorption to the reservoir walls causes irreproducibility in studies of protein-DNA interactions by methods of kinetic capillary electrophoresis. *Analytical Chemistry,* **2011.** 83, 8041–8045

3. Krylova, S.M.; Dove, P.M.; Kanoatov, M.; Krylov, S.N. The slow-dissociation and slow-recombination assumptions in nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry,* **2011.** 83, 7582–7585

2. Kanoatov, M.; Javaherian, S.; Krylov, S. N. Selection of aptamers for a non-DNA binding protein in the context of cell lysate. *Analytica Chimica Acta,* **2010.** 681, 92–97

1. Javaherian, S.; Musheev, M. U.; Kanoatov, M.; Berezovski, M. V.; Krylov, S. N. Selection of aptamers for a protein target in cell lysate and their application to protein purification. *Nucleic Acids Research,* **2009.** 37, e62

REFERENCES

1. Kastritis, P.L. and A. Bonvin, On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *Journal of the Royal Society Interface*, **2013**. *10* (79).
2. Bonetta, L., Protein-protein interactions: Interactome under construction. *Nature*, **2010**. *468* (7325): p. 851-854.
3. Garbett, N.C. and J.B. Chaires, Thermodynamic studies for drug design and screening. *Expert Opinion on Drug Discovery*, **2012**. *7* (4): p. 299-314.
4. Lundahl, P., A. Lundqvist, and E. Greijer, Quantitative Analysis of Biospecific Interactions. **1998**: Harwood Academic Publishers.
5. Verdier, L., J. Gharbi-Benarous, G. Bertho, P. Mauvais, and J.P. Giraud, Measurement of the dissociation-equilibrium constants for low affinity antibiotic binding interaction with bacterial ribosomes by the T-2 (CPMG) and line-broadening methods. *Journal de Chimie Physique et de Physico-Chimie Biologique*, **1999**. *96* (9-10): p. 1616-1623.
6. Green, N.M., Avidin, in *Advances in Protein Chemistry*, J.T.E. C.B. Anfinsen and M.R. Frederic, Editors. **1975**, Academic Press. p. 85-133.
7. Nunez, S., J. Venhorst, and C.G. Kruse, Target-drug interactions: first principles and their application to drug discovery. *Drug Discovery Today*, **2012**. *17* (1-2): p. 10-22.
8. McNaught, A.D. and A. Wilkinson, IUPAC "Gold Book". Compendium of Chemical Terminology. 2nd ed. **1997**, Oxford: Blackwell Scientific Publications.
9. Wong, I. and T.M. Lohman, A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proceedings of the National Academy of Sciences*, **1993**. *90* (12): p. 5428-5432.
10. Bowers, W.F., S. Fulton, and J. Thompson, Ultrafiltration vs equilibrium dialysis for determination of free fraction. *Clinical Pharmacokinetics*, **1984**. *9* (1): p. 49-60.
11. Hellman, L.M. and M.G. Fried, Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, **2007**. *2* (8): p. 1849-1861.
12. Leavitt, S. and E. Freire, Direct measurement of protein binding energetics by isothermal titration calorimetry. *Current Opinion in Structural Biology*, **2001**. *11* (5): p. 560-566.
13. He, J.W., D. Wu, Y.M. Zhai, Q. Wang, X.L. Ma, H.Q. Yang, and H. Li, Interaction of inosine with human serum albumin as determined by NMR relaxation data and fluorescence methodology. *Journal of Molecular Liquids*, **2016**. *219*: p. 547-553.
14. Wienken, C.J., P. Baaske, U. Rothbauer, D. Braun, and S. Duhr, Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, **2010**. *1*: p. 7.
15. Li, L., R. Zhou, H. Geng, L. Yue, F. Ye, Y. Xie, J. Liu, X. Kong, H. Jiang, and J. Huang, Discovery of two aminoglycoside antibiotics as inhibitors targeting the menin–mixed lineage leukaemia interface. *Bioorganic & medicinal chemistry letters*, **2014**. *24* (9): p. 2090-2093.
16. Read, M., R.J. Harrison, B. Romagnoli, F.A. Tanious, S.H. Gowan, A.P. Reszka, W.D. Wilson, L.R. Kelland, and S. Neidle, Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proceedings of the National academy of Sciences of the United States of America*, **2001**. *98* (9): p. 4844-4849.
17. Copeland, R.A., The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety. *Expert Opinion on Drug Discovery*, **2010**. *5* (4): p. 305-310.

18. Kersh, G.J., E.N. Kersh, D.H. Fremont, and P.M. Allen, High-and low-potency ligands with similar affinities for the TCR: the importance of kinetics in TCR signaling. *Immunity*, **1998**. *9* (6): p. 817-826.

19. Pardue, H.L., Kinetic aspects of analytical-chemistry. *Analytica Chimica Acta*, **1989**. *216* (1-2): p. 69-107.

20. Schuck, P., Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annual Review of Biophysics and Biomolecular Structure*, **1997**. *26*: p. 541-566.

21. Concepcion, J., K. Witte, C. Wartchow, S. Choo, D.F. Yao, H. Persson, J. Wei, P. Li, B. Heidecker, W.L. Ma, R. Varma, L.S. Zhao, D. Perillat, G. Carricato, M. Recknor, K. Du, H. Ho, T. Ellis, J. Gamez, M. Howes, J. Phi-Wilson, S. Lockard, R. Zuk, and H. Tan, Label-Free Detection of Biomolecular Interactions Using BioLayer Interferometry for Kinetic Characterization. *Combinatorial Chemistry and High Throughput Screening*, **2009**. *12* (8): p. 791-800.

22. Olson, S.T., K.R. Srinivasan, I. Bjork, and J.D. Shore, Binding of high-affinity heparin to Antithrombin-III - stopped flow kinetic-studies of the binding interaction. *Journal of Biological Chemistry*, **1981**. *256* (21): p. 1073-1079.

23. Schwille, P., U. Haupts, S. Maiti, and W.W. Webb, Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophysical Journal*, **1999**. *77* (4): p. 2251-2265.

24. Petrov, A., V. Okhonin, M. Berezovski, and S.N. Krylov, Kinetic capillary electrophoresis (KCE): A conceptual platform for kinetic homogeneous affinity methods. *Journal of the American Chemical Society*, **2005**. *127* (48): p. 17104-17110.

25. Osmond, R.I.W., W.C. Kett, S.E. Skett, and D.R. Coombe, Protein-heparin interactions measured by BIAcore 2000 are affected by the method of heparin immobilization. *Analytical Biochemistry*, **2002**. *310* (2): p. 199-207.

26. Secundo, F., Conformational changes of enzymes upon immobilisation. *Chemical Society Reviews*, **2013**. *42* (15): p. 6250-6261.

27. Rich, R.L. and D.G. Myszka, Higher-throughput, label-free, real-time molecular interaction analysis. *Analytical Biochemistry*, **2007**. *361* (1): p. 1-6.

28. Schasfoort, R.B. and A.J. Tudos, Handbook of surface plasmon resonance. **2008**: Royal Society of Chemistry.

29. Gomezhens, A. and D. Perezbendito, The stopped-flow technique in analytical chemistry. *Analytica Chimica Acta*, **1991**. *242* (2): p. 147-177.

30. Kalidas, C., Chemical Kinetic Methods : Principles Of Fast Reaction Techniques And Applications. **2005**: New Age International.

31. Meseth, U., T. Wohland, R. Rigler, and H. Vogel, Resolution of fluorescence correlation measurements. *Biophysical Journal*, **1999**. *76* (3): p. 1619-1631.

32. Briandet, R., P. Lacroix-Gueu, M. Renault, S. Lecart, T. Meylheuc, E. Bidnenko, K. Steenkeste, M.-N. Bellon-Fontaine, and M.-P. Fontaine-Aupart, Fluorescence correlation spectroscopy to study diffusion and reaction of bacteriophages inside biofilms. *Applied and Environmental Microbiology*, **2008**. *74* (7): p. 2135-2143.

33. Chen, Y. and J.D. Müller, Determining the stoichiometry of protein heterocomplexes in living cells with fluorescence fluctuation spectroscopy. *Proceedings of the National Academy of Sciences*, **2007**. *104* (9): p. 3147-3152.

34. Meurant, G., Practical Capillary Electrophoresis. **2012**: Elsevier Science.

35. Landers, J.P., Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques, Third Edition. **2007**: CRC Press. 305-335.

36. Sun, L., G. Zhu, Y. Zhao, X. Yan, S. Mou, and N.J. Dovichi, Ultrasensitive and fast bottom-up analysis of femtogram amounts of complex proteome digests. *Angewandte Chemie International Edition*, **2013**. *52* (51): p. 13661-13664.

37. Dada, O.O., B.J. Huge, and N.J. Dovichi, Simplified sheath flow cuvette design for ultrasensitive laser induced fluorescence detection in capillary electrophoresis. *Analyst*, **2012**. *137* (13): p. 3099-3101.

38. Galievsky, V.A., A.S. Stasheuski, and S.N. Krylov, "Getting the best sensitivity from on-capillary fluorescence detection in capillary electrophoresis" – A tutorial. *Analytica Chimica Acta*, **2016**. *935*: p. 58-81.

39. Galievsky, V.A., A.S. Stasheuski, and S.N. Krylov, Capillary Electrophoresis for Quantitative Studies of Biomolecular Interactions. *Analytical Chemistry*, **2015**. *87* (1): p. 157-171.

40. Drossman, H., J.A. Luckey, A.J. Kostichka, J. Dcunha, and L.M. Smith, High-Speed Separations of DNA Sequencing Reactions by Capillary Electrophoresis. *Analytical Chemistry*, **1990**. *62* (9): p. 900-903.

41. Wang, H.L. and T. Li, Organic Osmolyte Mediated Kinetic Capillary Electrophoresis for Study of Protein-DNA Interactions. *Analytical Chemistry*, **2009**. *81* (5): p. 1988-1995.

42. Okhonin, V., S.M. Krylova, and S.N. Krylov, Nonequilibrium capillary electrophoresis of equilibrium mixtures, mathematical model. *Analytical Chemistry*, **2004**. *76* (5): p. 1507-1512.

43. Musheev, M.U., Y. Filiptsev, and S.N. Krylov, Noncooled Capillary Inlet: A Source of Systematic Errors in Capillary-Electrophoresis-Based Affinity Analyses. *Analytical Chemistry*, **2010**. *82* (20): p. 8637-8641.

44. Musheev, M.U., Y. Filiptsev, and S.N. Krylov, Temperature Difference between the Cooled and the Noncooled Parts of an Electrolyte in Capillary Electrophoresis. *Analytical Chemistry*, **2010**. *82* (20): p. 8692-8695.

45. Cherney, L.T., M. Kanoatov, and S.N. Krylov, Method for Determination of Peak Areas in Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures. *Analytical Chemistry*, **2011**. *83* (22): p. 8617-8622

46. Ellington, A. and J. Szostak, Invitro selection of RNA molecules that bind specific ligands. *Nature*, **1990**. *346* (6287): p. 818-822.

47. Wochner, A., M. Menger, D. Orgel, B. Cech, M. Rimmele, and V. Erdmann, A DNA aptamer with high affinity and specificity for therapeutic anthracyclines. *Analytical Biochemistry*, **2008**. *373* (1): p. 34-42.

48. Berezovski, M., A. Drabovich, S.M. Krylova, M. Musheev, V. Okhonin, A. Petrov, and S.N. Krylov, Nonequilibrium capillary electrophoresis of equilibrium mixtures: A universal tool for development of aptamers. *Journal of the American Chemical Society*, **2005**. *127* (9): p. 3165-3171.

49. Drabovich, A.P., M. Berezovski, V. Okhonin, and S.N. Krylov, Selection of smart aptamers by methods of kinetic capillary electrophoresis. *Analytical Chemistry*, **2006**. *78* (9): p. 3171-3178.

50. Berezovski, M. and S.N. Krylov, Nonequilibrium capillary electrophoresis of equilibrium mixtures - A single experiment reveals equilibrium and kinetic parameters of protein-DNA interactions. *Journal of the American Chemical Society*, **2002**. *124* (46): p. 13674-13675.

51. Berezovski, M. and S.N. Krylov, Thermochemistry of protein-DNA interaction studied with temperature-controlled nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry*, **2005**. *77* (5): p. 1526-1529.

52. Musheev, M.U., S. Javaherian, V. Okhonin, and S.N. Krylov, Diffusion as a Tool of Measuring Temperature inside a Capillary. *Analytical Chemistry*, **2008**. *80* (17): p. 6752-6757.

53. Evenhuis, C.J., M.U. Musheev, and S.N. Krylov, Universal Method for Determining Electrolyte Temperatures in Capillary Electrophoresis. *Analytical Chemistry*, **2011**. *83* (5): p. 1808-1814.

54. Patel, K.H., C.J. Evenhuis, L.T. Cherney, and S.N. Krylov, Simplified universal method for determining electrolyte temperatures in a capillary electrophoresis instrument with forced-air cooling. *Electrophoresis*, **2012**. *33* (6): p. 1079-1085.

55. Kanoatov, M., C. Retif, L.T. Cherney, and S.N. Krylov, Peak-Shape Correction to Symmetry for Pressure-Driven Sample Injection in Capillary Electrophoresis. *Analytical Chemistry*, **2012**. *84* (1): p. 149-154.

56. Cherney, L.T., V. Okhonin, A.P. Petrov, and S.N. Krylov, Sequential-dissociation kinetics of non-covalent complexes of DNA with multiple proteins in separation-based approach: General theory and its application. *Analytica Chimica Acta*, **2012**. *724*: p. 111-118.

57. Petrov, A.P., L.T. Cherney, B. Dodgson, V. Okhonin, and S.N. Krylov, Separation-Based Approach to Study Dissociation Kinetics of Noncovalent DNA-Multiple Protein Complexes. *Journal of the American Chemical Society*, **2011**. *133* (32): p. 12486-12492.

58. Zou, D.D., D.P. Zhang, S.Q. Liu, B.L. Zhao, and H.L. Wang, Interplay of Binding Stoichiometry and Recognition Specificity for the Interaction of MBD2b Protein and Methylated DNA Revealed by Affinity Capillary Electrophoresis Coupled with Laser-Induced Fluorescence Analysis. *Analytical Chemistry*, **2014**. *86* (3): p. 1775-1782.

59. Wang, J.H., P.J. Jiang, L. Qiu, C.L. Wang, and J. Xia, Resolving antibody-peptide complexes with different ligand stoichiometries reveals a marked affinity enhancement through multivalency. *Talanta*, **2013**. *115*: p. 394-400.

60. Abd El-Hady, D. and H.M. Albishri, Hyphenated affinity capillary electrophoresis with a high-sensitivity cell for the simultaneous binding study of retinol and retinoic acid in nanomolars with serum albumins. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences*, **2012**. *911*: p. 180-185.

61. Beneito-Cambra, M., J.M. Herrero-Martinez, and G. Ramis-Ramos, Characterization and determination of poly(vinylpyrrolidone) by complexation with an anionic azo-dye and nonequilibrium capillary electrophoresis. *Journal of Chromatography A*, **2009**. *1216* (51): p. 9014-9021.

62. Rauch, J.N., J. Nie, T.J. Buchholz, J.E. Gestwicki, and R.T. Kennedy, Development of a Capillary Electrophoresis Platform for Identifying Inhibitors of Protein-Protein Interactions. *Analytical Chemistry*, **2013**. *85* (20): p. 9824-9831.

63. Sun, J., B. He, Q. Liu, T. Ruan, and G. Jiang, Characterization of interactions between organotin compounds and human serum albumin by capillary electrophoresis coupled with inductively coupled plasma mass spectrometry. *Talanta*, **2012**. *93*: p. 239-244.

64. Drabovich, A.P., M.V. Berezovski, M.U. Musheev, and S.N. Krylov, Selection of Smart Small-Molecule Ligands: The Proof of Principle. *Analytical Chemistry*, **2009**. *81* (1): p. 490-494.

65. Bao, J.Y. and S.N. Krylov, Volatile Kinetic Capillary Electrophoresis for Studies of Protein-Small Molecule Interactions. *Analytical Chemistry*, **2012**. *84* (16): p. 6944-6947.

66. Berezovski, M., R. Nutiu, Y.F. Li, and S.N. Krylov, Affinity analysis of a protein-aptamer complex using nonequilibrium capillary electrophoresis of equilibrium mixtures. *Analytical Chemistry*, **2003**. *75* (6): p. 1382-1386.

67. Mendonsa, S.D. and M.T. Bowser, In vitro evolution of functional DNA using capillary electrophoresis. *Journal of the American Chemical Society*, **2004**. *126* (1): p. 20-21.

68. Berezovski, M., M. Musheev, A. Drabovich, and S.N. Krylov, Non-SELEX selection of aptamers. *Journal of the American Chemical Society*, **2006**. *128* (5): p. 1410-1411.

69. Drabovich, A.P., V. Okhonin, M. Berezovski, and S.N. Krylov, Smart aptamers facilitate multi-probe affinity analysis of proteins with ultra-wide dynamic range of measured concentrations. *Journal of the American Chemical Society*, **2007**. *129* (23): p. 7260-7261.

70. Javaherian, S., M.U. Musheev, M. Kanoatov, M.V. Berezovski, and S.N. Krylov, Selection of aptamers for a protein target in cell lysate and their application to protein purification. *Nucleic Acids Research*, **2009**. *37* (8).

71. Kanoatov, M., S. Javaherian, and S.N. Krylov, Selection of aptamers for a non-DNA binding protein in the context of cell lysate. *Analytica Chimica Acta*, **2010**. *681* (1-2): p. 92-97.

72. Li, S.F.Y., J. Tok, J. Lai, and T. Leung, Selection of aptamers for signal transduction proteins by capillary electrophoresis. *Electrophoresis*, **2010**. *31* (12): p. 2055-2062.

73. Krylov, S.N., S.M. Krylova, A.A. Karkhanina, M.U. Musheev, E.A.L. Bagg, and C.J. Schofield, DNA aptamers for as analytical tools for the quantitative analysis of DNA-dealkylating enzymes. *Analytical Biochemistry*, **2011**. *414* (2): p. 261-265.

74. Ashley, J., K.L. Ji, and S.F.Y. Li, Selection of bovine catalase aptamers using non-SELEX. *Electrophoresis*, **2012**. *33* (17): p. 2783-2789.

75. Tran, D.T., K. Knez, K.P. Janssen, J. Pollet, D. Spasic, and J. Lammertyn, Selection of aptamers against Ara h 1 protein for FO-SPR biosensing of peanut allergens in food matrices. *Biosensors & Bioelectronics*, **2013**. *43*: p. 245-251.

76. Kim, S.E., W. Su, M. Cho, Y. Lee, and W.S. Choe, Harnessing aptamers for electrochemical detection of endotoxin. *Analytical Biochemistry*, **2012**. *424* (1): p. 12-20.

77. Jing, M. and M.T. Bowser, Tracking the Emergence of High Affinity Aptamers for rhVEGF(165) During Capillary Electrophoresis-Systematic Evolution of Ligands by Exponential Enrichment Using High Throughput Sequencing. *Analytical Chemistry*, **2013**. *85* (22): p. 10761-10770.

78. Ashley, J. and S.F.Y. Li, Three-dimensional selection of leptin aptamers using capillary electrophoresis and implications for clone validation. *Analytical Biochemistry*, **2013**. *434* (1): p. 146-152.

79. Riley, K.R., J. Gagliano, J.J. Xiao, K. Libby, S. Saito, G. Yu, R. Cubicciotti, J. Macosko, C.L. Colyer, M. Guthold, and K. Bonin, Combining capillary electrophoresis and next-generation

sequencing for aptamer selection. *Analytical and Bioanalytical Chemistry*, **2015**. *407* (6): p. 1527-1532.

80. Yang, P., Y. Mao, A.W.M. Lee, and R.T. Kennedy, Measurement of dissociation rate of biomolecular complexes using CE. *Electrophoresis*, **2009**. *30* (3): p. 457-464.

81. Noble, J.E., A.E. Knight, A.J. Reason, A. Di Matola, and M.J.A. Bailey, A comparison of protein quantitation assays for biopharmaceutical applications. *Molecular Biotechnology*, **2007**. *37* (2): p. 99-111.

82. Krylov, S.N., Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM): A novel method for biomolecular screening. *Journal of Biomolecular Screening*, **2006**. *11* (2): p. 115-122.

83. Drabovich, A., M. Berezovski, and S.N. Krylov, Selection of smart aptamers by equilibrium capillary electrophoresis of equilibrium mixtures (ECEEM). *Journal of the American Chemical Society*, **2005**. *127* (32): p. 11224-11225.

84. Krylov, S.N., D.A. Starke, E.A. Arriaga, Z.R. Zhang, N.W.C. Chan, M.M. Palcic, and N.J. Dovichi, Instrumentation for chemical cytometry. *Analytical Chemistry*, **2000**. *72* (4): p. 872-877.

85. Wu, S.L. and N.J. Dovichi, High-sensitivity fluorescence detector for fluorescein isothiocyanate derivatives of amino-acids separated by capillary zone electrophoresis. *Journal of Chromatography*, **1989**. *480*: p. 141-155.

86. Cochran, W.G., Some methods for strengthening the common $\chi 2$ tests. *Biometrics*, **1954**. *10* (4): p. 417-451.

87. van de Geer, S.A., Least Squares Estimation, in *Encyclopedia of Statistics in Behavioral Science*, B.S. Everitt and D.C. Howell, Editors. **2005**, John Wiley & Sons, Ltd.: Chichester. p. 1041-1045.

88. Crouch, S.R., A. Scheeline, and E.S. Kirkor, Kinetic determinations and some kinetic aspects of analytical chemistry. *Analytical Chemistry*, **2000**. *72* (12): p. 53R-70R.

89. Shah, N.B. and T.M. Duncan, Bio-layer Interferometry for Measuring Kinetics of Protein-protein Interactions and Allosteric Ligand Effects. *Journal of Visualized Experiments*, **2014** (84): p. e51383-e51383.

90. Ogilvie, I.R.G., V.J. Sieben, M.C. Mowlem, and H. Morgan, Temporal Optimization of Microfluidic Colorimetric Sensors by Use of Multiplexed Stop-Flow Architecture. *Analytical Chemistry*, **2011**. *83* (12): p. 4814-4821.

91. Yang, P., Y. Ma, A.W.M. Lee, and R.T. Kennedy, Measurement of dissociation rate of biomolecular complexes using CE. *Electrophoresis*, **2009**. *30* (3): p. 457-464.

92. Pavski, V. and X.C. Le, Detection of human immunodeficiency virus type 1 reverse transcriptase using aptamers as probes in affinity capillary electrophoresis. *Analytical Chemistry*, **2001**. *73* (24): p. 6070-6076.

93. Mosing, R.K., S.D. Mendonsa, and M.T. Bowser, Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase. *Analytical Chemistry*, **2005**. *77* (19): p. 6107-6112.

94. Long, G.L. and J.D. Winefordner, Limit of detection. A closer look at the IUPAC definition. *Analytical Chemistry*, **1983**. *55* (7): p. 712A-724A.

95. Kanoatov, M., V.A. Galievsky, S.M. Krylova, L.T. Cherney, H.K. Jankowski, and S.N. Krylov, Using Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures (NECEEM) for

Simultaneous Determination of Concentration and Equilibrium Constant. *Analytical Chemistry*, **2015**. *87* (5): p. 3099-3106.

96. Rozet, E., P. Lebrun, B. Debrus, B. Boulanger, and P. Hubert, Design Spaces for analytical methods. *Trac-Trends in Analytical Chemistry*, **2013**. *42*: p. 157-167.

97. ICH Q8, Guidance for Industry, Q8(R2) Pharmaceutical Development. **2006**, International Council for Harmonization: global.

98. Forster, R.E., D.G. Hert, T.N. Chiesl, C.P. Fredlake, and A.E. Barron, DNA migration mechanism analyses for applications in capillary and microchip electrophoresis. *Electrophoresis*, **2009**. *30* (12): p. 2014-2024.

99. Wheeler, D.A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.J. Chen, V. Makhijani, G.T. Roth, X. Gomes, K. Tartaro, F. Niazi, C.L. Turcotte, G.P. Irzyk, J.R. Lupski, C. Chinault, X.Z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D.M. Muzny, M. Margulies, G.M. Weinstock, R.A. Gibbs, and J.M. Rothberg, The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **2008**. *452* (7189): p. 872-876.

100. Liu, Y.J., C.B. Rauch, R.L. Stevens, R. Lenigk, J.N. Yang, D.B. Rhine, and P. Grodzinski, DNA amplification and hybridization assays in integrated plastic monolithic devices. *Analytical Chemistry*, **2002**. *74* (13): p. 3063-3070.

101. Wegman, D.W. and S.N. Krylov, Direct Quantitative Analysis of Multiple miRNAs (DQAMmiR). *Angewandte Chemie-International Edition*, **2011**. *50* (44): p. 10335-10339.

102. Becker, C., A. Hammerle-Fickinger, I. Riedmaier, and M.W. Pfaffl, mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, **2010**. *50* (4): p. 237-243.

103. Lou, B.L., E. Chen, X.Y. Zhao, F. Qu, and J.Y. Yan, The application of capillary electrophoresis for assisting whole-cell aptamers selection by characterizing complete ssDNA distribution. *Journal of Chromatography A*, **2016**. *1437*: p. 203-209.

104. Mironov, G.G., V. Okhonin, N. Khan, C.M. Clouthier, and M.V. Berezovski, Conformational Dynamics of DNA G-Quadruplex in Solution Studied by Kinetic Capillary Electrophoresis Coupled On-line with Mass Spectrometry. *ChemistryOpen*, **2014**. *3* (2): p. 58-64.

105. Clouthier, C.M., G.G. Mironov, V. Okhonin, M.V. Berezovski, and J.W. Keillor, Real-Time Monitoring of Protein Conformational Dynamics in Solution Using Kinetic Capillary Electrophoresis. *Angewandte Chemie-International Edition*, **2012**. *51* (50): p. 12464-12468.

106. Carrasco-Correa, E.J., M. Beneito-Cambra, J.M. Herrero-Martinez, and G. Ramis-Ramos, Evaluation of molecular mass and tacticity of polyvinyl alcohol by non-equilibrium capillary electrophoresis of equilibrium mixtures of a polymer and a dye. *Journal of Chromatography A*, **2011**. *1218* (16): p. 2334-2341.

107. Sloat, A.L., M.G. Roper, X.L. Lin, J.P. Ferrance, J.P. Landers, and C.L. Colyer, Protein determination by microchip capillary electrophoresis using an asymmetric squarylium dye: Noncovalent labeling and nonequilibrium measurement of association constants. *Electrophoresis*, **2008**. *29* (16): p. 3446-3455.

108. Li, T. and H.L. Wang, Organic Osmolyte Mediated Kinetic Capillary Electrophoresis for Study of Protein-DNA Interactions. *Analytical Chemistry*, **2009**. *81* (5): p. 1988-1995.

109. Keefe, A.D., S. Pai, and A. Ellington, Aptamers as therapeutics. *Nature Reviews Drug Discovery*, **2010**. *9* (7): p. 537-550.

110. Xu, C.-X. and X.-F. Yin, Continuous cell introduction and rapid dynamic lysis for high-throughput single-cell analysis on microfludic chips with hydrodynamic focusing. *Journal of Chromatography A*, **2011**. *1218* (5): p. 726-732.

111. Li, S.K., M.R. Liddell, and H. Wen, Effective electrophoretic mobilities and charges of anti-VEGF proteins determined by capillary zone electrophoresis. *Journal of Pharmaceutical and Biomedical Analysis*, **2011**. *55* (3): p. 603-607.

112. Jia, Z.J., T. Ramstad, and M. Zhong, Determination of protein-drug binding constants by pressure-assisted capillary electrophoresis (PACE)/frontal analysis (FA). *Journal of Pharmaceutical and Biomedical Analysis*, **2002**. *30* (3): p. 405-413.

113. Ichiki, T., T. Ujiie, S. Shinbashi, T. Okuda, and Y. Horiike, Immunoelectrophoresis of red blood cells performed on microcapillary chips. *Electrophoresis*, **2002**. *23* (13): p. 2029-2034.

114. Green, L., D. Jellinek, R. Jenison, A. Ostman, C. Heldin, and N. Janjic, Inhibitory DNA ligands to platelet-derived growth factor B-chain. *Biochemistry*, **1996**. *35* (45): p. 14413-14424.

115. Solvent and Environmental Effects, in *Principles of Fluorescence Spectroscopy*, J.R. Lakowicz, Editor. **2006**, Springer US: Boston, MA. p. 205-235.

116. Kanoatov, M. and S.N. Krylov, DNA Adsorption to the Reservoir Walls Causing Irreproducibility in Studies of Protein-DNA Interactions by Methods of Kinetic Capillary Electrophoresis. *Analytical Chemistry*, **2011**. *83* (20): p. 8041-8045.

117. Chang, C.H., Y. Wang, P. Gupta, and D.M. Goldenberg, Extensive crosslinking of CD22 by epratuzumab triggers BCR signaling and caspase-dependent apoptosis in human lymphoma cells. *mAbs*, **2015**. *7* (1): p. 199-211.

118. Panchuk-Voloshina, N., R.P. Haugland, J. Bishop-Stewart, M.K. Bhalgat, P.J. Millard, F. Mao, W.Y. Leung, and R.P. Haugland, Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *Journal of Histochemistry & Cytochemistry*, **1999**. *47* (9): p. 1179-1188.

119. Lim, C.Y., A.E. Lim, and Y.C. Lam, Ionic Origin of Electro-osmotic Flow Hysteresis. *Scientific Reports*, **2016**. *6*.

120. de Jong, S. and S.N. Krylov, Pressure-Based Approach for the Analysis of Protein Adsorption in Capillary Electrophoresis. *Analytical Chemistry*, **2012**. *84* (1): p. 453-458.

121. Kanoatov, M., L.T. Cherney, and S.N. Krylov, Extracting kinetics from affinity capillary electrophoresis (ACE) data: a new blade for the old tool. *Analytical Chemistry*, **2014**. *86* (2): p. 1298-305.

122. Viovy, J.L., Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. *Reviews of Modern Physics*, **2000**. *72* (3): p. 813-872.

123. Mayer, P., G.W. Slater, and G. Drouin, Theory of DNA-sequencing using free-solution electrophoresis of protein-DNA complexes. *Analytical Chemistry*, **1994**. *66* (10): p. 1777-1780.

124. Olivera, B.M., P. Baine, and N. Davidson, Electrophoresis of the nucleic acids. *Biopolymers*, **1964**. *2* (3): p. 245-257.

125. Stellwagen, N.C. and E. Stellwagen, Effect of the matrix on DNA electrophoretic mobility. *Journal of Chromatography A*, **2009**. *1216* (10): p. 1917-1929.

126. Mendonsa, S.D. and M.T. Bowser, In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. *Analytical Chemistry*, **2004**. *76* (18): p. 5387-5392.

127. Zhu, Z., G. Jenkins, W.H. Zhang, M.X. Zhang, Z.C. Guan, and C.Y.J. Yang, Single-molecule emulsion PCR in microfluidic droplets. *Analytical and Bioanalytical Chemistry*, **2012**. *403* (8): p. 2127-2143.

128. Larionov, A., A. Krause, and W. Miller, A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*, **2005**. *6*.

129. Bello, M.S., R. Rezzonico, and P.G. Righetti, Use of taylor-aris dispersion for measurement of a solute diffusion coefficient in thin capillaries. *Science*, **1994**. *266* (5186): p. 773-6.

130. Evans, D.H. and A.R. Morgan, Characterization of Imperfect DNA Duplexes Containing Unpaired Bases and Non-Watson-Crick Base-Pairs. *Nucleic Acids Research*, **1986**. *14* (10): p. 4267-4280.

131. Russo, N., M. Toscano, and A. Grand, Bond energies and attachments sites of sodium and potassium cations to ONA and RNA nucleic acid bases in the gas phase. *Journal of the American Chemical Society*, **2001**. *123* (42): p. 10272-10279.

132. Marincola, F.C., V.P. Denisov, and B. Halle, Competitive Na+ and Rb+ binding in the minor groove of DNA. *Journal of the American Chemical Society*, **2004**. *126* (21): p. 6739-6750.

133. Egli, M., DNA-cation interactions: Quo vadis? *Chemistry & Biology*, **2002**. *9* (3): p. 277-286.

134. Hud, N.V. and M. Polak, DNA-cation interactions: the major and minor grooves are flexible ionophores. *Current Opinion in Structural Biology*, **2001**. *11* (3): p. 293-301.

135. Mohideen-Abdul, K., R. Muhammad, and C.A. Davey, Perturbations in nucleosome structure from heavy metal association. *Nucleic Acids Research*, **2010**. *38* (18): p. 6301-6311.

136. Stellwagen, E. and N.C. Stellwagen, Probing the electrostatic shielding of DNA with capillary electrophoresis. *Biophysical Journal*, **2003**. *84* (3): p. 1855-1866.

137. Perez, A., F.J. Luque, and M. Orozco, Dynamics of B-DNA on the microsecond time scale. *Journal of the American Chemical Society*, **2007**. *129* (47): p. 14739-14745.

138. Manning, G.S., Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions .I. Colligative Properties. *Journal of Chemical Physics*, **1969**. *51* (3): p. 924-933.

139. Manning, G.S., The critical onset of counterion condensation: A survey of its experimental and theoretical basis. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics*, **1996**. *100* (6): p. 909-922.

140. Huber, K. and U. Scheler, New experiments for the quantification of counterion condensation. *Current Opinion in Colloid & Interface Science*, **2012**. *17* (2): p. 64-73.

141. Manning, G.S., Field-Dissociation Relation for Polyelectrolytes with an Application to Field-Induced Conformational-Changes of Polynucleotides. *Biophysical Chemistry*, **1977**. *7* (3): p. 189-192.

142. Rodriguez-Mejia, J.L., C. Martinez-Anaya, J.L. Folch-Mallol, and E. Dantan-Gonzalez, A two-step electrodialysis method for DNA purification from polluted metallic environmental samples. *Electrophoresis*, **2008**. *29* (15): p. 3239-3244.

143. Gobel, U., R. Maas, and A. Clad, Quantitative electroelution of oligonucleotides and large dna fragments from gels and purification by electrodialysis. *Journal of Biochemical and Biophysical Methods*, **1987**. *14* (5): p. 245-260.

144. Mitnik, L., C. Heller, J. Prost, and J.L. Viovy, Segregation in DNA Solutions Induced by Electric-Fields. *Science*, **1995**. *267* (5195): p. 219-222.

145. Tang, J., N. Du, and P.S. Doyle, Compression and self-entanglement of single DNA molecules under uniform electric field. *Proceedings of the National academy of Sciences of the United States of America*, **2011**. *108* (39): p. 16153-16158.

146. Song, L. and M.F. Maestre, Unhooking Dynamics of U-Shaped DNA Molecule Undergoing Gel-Electrophoresis. *Journal of Biomolecular Structure & Dynamics*, **1991**. *9* (1): p. 87-99.

147. Isambert, H., A. Ajdari, J.L. Viovy, and J. Prost, Electrohydrodynamic patterns in macroion dispersions under a strong electric field. *Physical Review E*, **1997**. *56* (5): p. 5688-5704.

148. Magnusdottir, S., H. Isambert, C. Heller, and J.L. Viovy, Electrohydrodynamically induced aggregation during constant and pulsed field capillary electrophoresis of DNA. *Biopolymers*, **1999**. *49* (5): p. 385-401.

149. Bloomfield, V.A., DNA condensation. *Current Opinion in Structural Biology*, **1996**. *6* (3): p. 334-341.

150. Oosawa, F., A Simple Theory of Thermodynamic Properties of Polyelectrolyte Solutions. *Journal of Polymer Science*, **1957**. *23* (103): p. 421-430.

151. Ray, J. and G.S. Manning, An attractive force between 2 rodlike polyions mediated by the sharing of condensed counterions. *Langmuir*, **1994**. *10* (7): p. 2450-2461.

152. Manning, G.S., Counterion condensation theory of attraction between like charges in the absence of multivalent counterions. *European Physical Journal E*, **2011**. *34* (12).

153. Sedlak, M. and E.J. Amis, Dynamics of moderately concentrated salt-free polyelectrolyte solutions - molecular-weight dependence. *Journal of Chemical Physics*, **1992**. *96* (1): p. 817-825.

154. Burge, S., G.N. Parkinson, P. Hazel, A.K. Todd, and S. Neidle, Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*, **2006**. *34* (19): p. 5402-5415.

155. Clark, D.J. and J.O. Thomas, Salt-dependent cooperative interaction of histone H-1 with linear DNA. *Journal of Molecular Biology*, **1986**. *187* (4): p. 569-580.

156. Conlan, L.H. and C.M. Dupureur, Multiple metal ions drive DNA association by PvuII endonuclease. *Biochemistry*, **2002**. *41* (50): p. 14848-14855.

157. Duckett, D.R., A.I.H. Murchie, and D.M.J. Lilley, The Role of Metal-Ions in the Conformation of the 4-Way DNA Junction. *EMBO Journal*, **1990**. *9* (2): p. 583-590.

158. Matzke, A.J.M. and M.A. Matzke, The Electrical-Properties of the Nuclear-Envelope, and Their Possible Role in the Regulation of Eukaryotic Gene-Expression. *Bioelectrochemistry and Bioenergetics*, **1991**. *25* (3): p. 357-370.

159. Matzke, A.J.M., T.M. Weiger, and M. Matzke, Ion Channels at the Nucleus: Electrophysiology Meets the Genome. *Molecular Plant*, **2010**. *3* (4): p. 642-652.

160. Tyner, K.M., R. Kopelman, and M.A. Philbert, "Nanosized voltmeter" enables cellular-wide electric field mapping. *Biophysical Journal*, **2007**. *93* (4): p. 1163-1174.

161. Zhang, H.Q., Z.W. Wang, X.F. Li, and X.C. Le, Ultrasensitive detection of proteins by amplification of affinity aptamers. *Angewandte Chemie-International Edition*, **2006**. *45* (10): p. 1576-1580.

162. Schachat, A.P., New treatments for age-related macular degeneration. *Ophthalmology*, **2005**. *112* (4): p. 531-2.

163. Vallian, S. and M. Khazaei, Medical applications of aptamers. *Research in Pharmaceutical Sciences*, **2009**. *2* (2): p. 59-66.

164. Horvath, J. and V. Dolnik, Polymer wall coatings for capillary electrophoresis. *Electrophoresis*, **2001**. *22* (4): p. 644-655.

165. Righetti, P.G., C. Gelfi, B. Verzola, and L. Castelletti, The state of the art of dynamic coatings. *Electrophoresis*, **2001**. *22* (4): p. 603-611.

166. Cherney, L.T., A.P. Petrov, and S.N. Krylov, One-Dimensional Approach to Study Kinetics of Reversible Binding of Protein on Capillary Walls. *Analytical Chemistry*, **2015**. *87* (2): p. 1219-1225.

167. Cherney, L.T. and S.N. Krylov, Slow-equilibration approximation in studying kinetics of protein adsorption on capillary walls. *Analyst*, **2015**. *140* (8): p. 2797-2803.

168. Bao, J., S.M. Krylova, L.T. Cherney, J.Y. LeBlanc, P. Pribil, P.E. Johnson, D.J. Wilson, and S.N. Krylov, Kinetic Size-Exclusion Chromatography with Mass Spectrometry Detection: An Approach for Solution-Based Label-Free Kinetic Analysis of Protein–Small Molecule Interactions. *Analytical Chemistry*, **2014**. *86* (20): p. 10016-10020.

Appendix A. Mathematical Derivations and Proofs

All mathematical derivations and proofs were done by Dr. Leonid T. Cherney and are presented here for reference only.

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

Derivative of **Eq. 12** in respect to R can be presented in the following form:

$$\Delta K_{\mathrm{d}} \approx \frac{dK_{\mathrm{d}}}{dR}\Delta R \tag{57}$$

Calculations of derivative dK_d/dR using **Eq. 57** give:

$$\frac{dK_{\mathrm{d}}}{dR} = \frac{[\mathrm{L}]_0}{\frac{1}{R}-1} + \frac{[\mathrm{T}]_0 - [\mathrm{L}]_0(1-R)}{\left(\frac{1}{R}-1\right)^2}\frac{1}{R^2} \tag{58}$$

This expression can be simplified and transferred as follows:

$$\begin{aligned}\frac{dK_{\mathrm{d}}}{dR} &= \frac{[\mathrm{L}]_0 R}{1-R} + \frac{[\mathrm{T}]_0 - [\mathrm{L}]_0(1-R)}{(1-R)^2} = \frac{[\mathrm{L}]_0 R}{1-R} + \frac{[\mathrm{T}]_0}{(1-R)^2} - \frac{[\mathrm{L}]_0}{1-R} = \\ &= \frac{[\mathrm{L}]_0 R - [\mathrm{L}]_0}{1-R} + \frac{[\mathrm{T}]_0}{(1-R)^2} = -[\mathrm{L}]_0 + \frac{[\mathrm{T}]_0}{(1-R)^2}\end{aligned} \tag{59}$$

From **Eqs. 57 and 59** we have **Eq. 17** that relates small deviations (or errors) denoted by Δ.

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

At a very small value $[L]_0$ (with respect to $[T]_0$ and K_d) we can approximately transform **Eq. 16** by expanding the square root as follows:

$$R_0 = R_{[\mathrm{L}]_0 \to 0} = -\frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0} + \frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0} \sqrt{1 + \frac{\frac{K_\mathrm{d}}{[\mathrm{L}]_0}}{\left(\frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0}\right)^2}} \approx$$

$$\approx -\frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0} + \frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0} \left(1 + \frac{1}{2} \frac{\frac{K_\mathrm{d}}{[\mathrm{L}]_0}}{\left(\frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{2[\mathrm{L}]_0}\right)^2}\right) = \tag{60}$$

$$= \frac{\frac{K_\mathrm{d}}{[\mathrm{L}]_0}}{\frac{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0}{[\mathrm{L}]_0}} = \frac{K_\mathrm{d}}{K_\mathrm{d} + [\mathrm{T}]_0 - [\mathrm{L}]_0} \approx \frac{K_\mathrm{d}}{K_\mathrm{d} + [\mathrm{T}]_0}$$

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

$$\frac{\partial R_0}{\partial K_\mathrm{d}} = \frac{1}{K_\mathrm{d} + [T]_0} - \frac{K_\mathrm{d}}{\left(K_\mathrm{d} + [T]_0\right)^2} = \frac{[T]_0}{\left(K_\mathrm{d} + [T]_0\right)^2} \tag{61}$$

If we consider a dependence of $\partial R_0/\partial K_\mathrm{d}$ on $[\mathrm{T}]_0$ then we can conclude that $\partial R_0/\partial K_\mathrm{d}$ has a maximum at some value of $[\mathrm{T}]_0 = [\mathrm{T}]^*$. Indeed, according to **Eq. 61** we have $\partial R_0/\partial K_\mathrm{d} = 0$ at $[\mathrm{T}]_0 = 0$, $\partial R_0/\partial K_\mathrm{d} = 0$ at $[\mathrm{T}]_0 = \infty$, and $\partial R_0/\partial K_\mathrm{d} > 0$ in the interval $0 < [\mathrm{T}]_0 < \infty$. In other words, $\partial R_0/\partial K_\mathrm{d}$ is positive inside an interval $(0,\infty)$ and approaches 0 at its ends. Therefore, **Eq. 61** reaches a maximum in some point $[\mathrm{T}]^*$ inside this interval. To find $[\mathrm{T}]^*$ we should calculate the second derivative $\partial(\partial R_0/\partial K_\mathrm{d})\partial\,[\mathrm{T}]_0$:

$$\frac{\partial}{\partial [\mathrm{T}]_0}\left(\frac{\partial R_0}{\partial K_\mathrm{d}}\right) = \frac{1}{\left(K_\mathrm{d} + [\mathrm{T}]_0\right)^2} - \frac{2[\mathrm{T}]_0}{\left(K_\mathrm{d} + [\mathrm{T}]_0\right)^3} \tag{62}$$

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

$$\frac{1}{\left(K_{\mathrm{d}}+[T]^{*}\right)^{2}}-\frac{2[T]^{*}}{\left(K_{\mathrm{d}}+[T]^{*}\right)^{3}}=0 \tag{63}$$

which can be transformed to a simpler one:

$$1-\frac{2[\mathrm{T}]^{*}}{K_{\mathrm{d}}+[\mathrm{T}]^{*}}=0 \tag{64}$$

Solving **Eq. 64** and taking into account **Eq. 60** we finally obtain:

$$[\mathrm{T}]^{*}=K_{\mathrm{d}},\quad \left(R_{0}\right)_{[\mathrm{T}]_{0}=[\mathrm{T}]^{*}}=\frac{1}{2} \tag{65}$$

The second relation in **Eq. 65** can be used to approximately satisfy relation $[\mathrm{T}]_0 = [\mathrm{T}]^*$ in experiments even if K_d and, therefore, $[\mathrm{T}]^*$ are unknown. To do that, one should dilute the target so that a value of R becomes approximately 0.5 at the lowest feasible concentration of L. In this case, R is the most affected by changes in K_d, which, in turn, is the most favorable scenario for fitting theoretical **Eq. 16** into experimental dependence of R on $[\mathrm{L}]_0$. After that, NECEEM experiments should be run at higher $[\mathrm{L}]_0$ which would result in R in the range $0.5 < R < 1$. Finally, fitting dependence **Eq. 16** into such experimental dependence of R on $[\mathrm{L}]_0$ should produce accurate values of K_d and $[\mathrm{T}]_0$.

Appendix B. Extracting Kinetics from Affinity CE Data

The presented material was published previously and reprinted with permission from "Kanoatov, M.; Cherney, L.T.; Krylov, S.N. Extracting kinetics from affinity capillary electrophoresis (ACE) data: a new blade for the old tool. *Analytical Chemistry* 2014, 86, 1298-1305".

B.1. Summary: We describe a mathematical approach that enables extraction of *kinetic rate constants* from thousands of studies conducted over the past two decades with affinity capillary electrophoresis (ACE). Previously, ACE has been used almost exclusively for obtaining *equilibrium constants* of intermolecular interactions. In this article, we prove that there exists an analytical solution of partial differential equations describing mass transfer in ACE. By using an *in-silico* study we demonstrate that the solution is applicable to experimental conditions that are typically used in ACE and found in most historical ACE experiments. The solution was validated by extracting rate constants from previously-published ACE data and closely matching independently obtained results. Lastly, it was used to obtain previously unknown rate constants from historical ACE data. The new mathematical approach expands the applicability of ACE to a wider range of biomolecular interactions, and enables both prospective and retrospective data analysis. The obtained kinetic information will be of significant practical value to the fields of pharmacology and molecular biology.

B.2. Introduction

Development of new data-analysis strategies can improve the performance of existing analytical methods. For example, development of the "second derivative" approach for analysis

of data for quantitative polymerase chain reaction has significantly increased the precision and accuracy of the method.[1] More interesting, however, is the development of data-analysis strategies that can extract previously inaccessible information from both new and old data. In this article, we introduce a simple mathematical approach which allows deconvolution of kinetic rate constants from data produced by affinity capillary electrophoresis (ACE). ACE is a popular method for determining equilibrium constants of affinity interactions between biological molecules.[2] The described mathematical tool enables new and valuable information to be extracted from existing analytical data published in close to 2,000 scientific articles, and to abundant unpublished data from the pharmaceutical industry.[3]

Affinity interactions are involved in regulation of practically all biological processes. Knowing molecular mechanisms that govern these interactions is of extreme importance to our understanding of normal cell function, disease, and drug action. At its basic level, study of intermolecular interactions requires knowledge of their equilibrium constants and kinetics. For this purpose, binding molecules A and B with a formation of complex C can be described by a simple chemical equation:

$$\mathrm{C} \underset{k_+}{\overset{k_-}{\rightleftarrows}} \mathrm{A} + \mathrm{B} \qquad K_\mathrm{d} = k_- / k_+ \tag{1}$$

where k_+ and k_- are rate constants of complex formation and dissociation, respectively, and K_d is the equilibrium dissociation constant. The goal of kinetic and thermodynamic studies is essentially to find k_+, k_-, and K_d.

A wide variety of methods is available that can only measure the K_d value; they can be called equilibrium methods.[4-8] This methodological variety accommodates the study of the vast diversity of biomolecular interactions, as each method offers different benefits and suffers from

different limitations. Availability of robust methods for K_d measurement has made this parameter extremely important in pharmacology, where K_d values are often used as a primary screening criterion for candidate drug compounds. However, it is becoming more evident that knowledge of K_d is not sufficient for characterization of drug candidates, and that their interaction kinetics, characterized by k_+ and k_-, may play a far more important role.[9] This stems from the fact that biological processes rarely occur in equilibrium; thus, knowledge of interaction kinetics allows making more biologically-relevant predictions. The current variety of kinetic methods is much more limited, with only surface plasmon resonance (SPR), bio-layer interferometry (BLI), fluorescence correlation spectroscopy (FCS), and stopped-flow finding practical use.[10-12] SPR and BLI are surface-based methods, as they require immobilization of one of the interaction components, while FCS and stopped-flow are label-based, as they require labeling of at least one of the reaction components. Requirement for immobilization or labeling represents a major limitation of these methods, as modification of components can influence the interaction between them by affecting their conformation or by introducing steric hindrance.[13,14] As a result, relatively few molecular biology studies or drug screening efforts take advantage of kinetic information. Development of solution-based label-free kinetic methods is, thus, highly desirable.

ACE was proposed as a solution-based label-free method for studying affinity interactions in the beginning of 1990's.[15] Since its introduction, it has been acknowledged that ACE data contains equilibrium and kinetic information needed to find K_d, k_+, k_-. However, extracting the convoluted k_+ and k_- necessitated the use of complicated numerical computation.[16] This likely explains the fact that throughout the two decades of existence, ACE has been used almost exclusively as a tool to find K_d values, and not k_+ and k_-. Nevertheless, ACE has found

widespread application in analytical biochemistry and pharmaceutical research with thousands of papers published.[17]

In this article, we demonstrate that there exists an analytical solution of partial differential equations describing mass transfer in ACE. This analytical solution was developed by adapting mathematical equations from a related capillary electrophoresis (CE)-based method called Macroscopic Approach for Studying Kinetics at Equilibrium (MASKE).[18] The developed analytical solution allows accurate deconvolution of kinetic rate constants in a simple and rapid manner, without the need for specialized computer hardware, and can be easily implemented in any of the commonly available spreadsheet or computational software environments. The new mathematical tool expands applicability of ACE to a wider range of biomolecular interactions and allows prospective and retrospective determination of kinetic rate constants. This work shows how a "mathematical blade" can enable an old experimental tool to perform new challenging functions.

B.3. Results and discussion

The concepts of ACE and MASKE

In this section, we provide a conceptual description of ACE and discuss its similarity to MASKE, a related CE method. Conceptual similarities between the two methods suggest that mathematical tools developed for MASKE can be potentially used for analyzing ACE data.

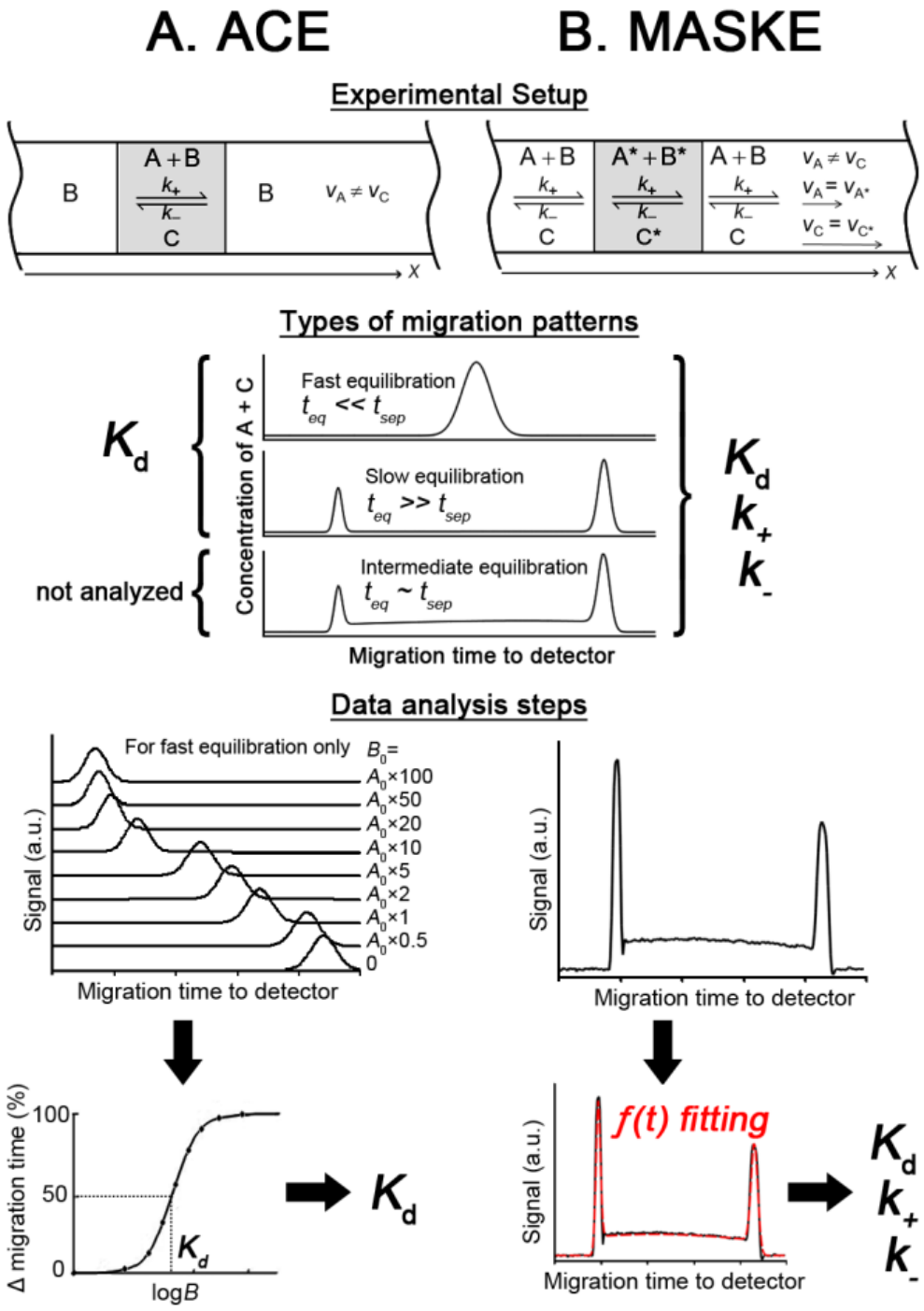

Figure B.1. Comparison of ACE (panel A**) and MASKE (**panel B**) methods. Top row:** schematic representation of initial conditions; **middle row:** schematic representation of general types of migration patterns and the information that is extracted from each type; **bottom row:** schematic representation of steps in data analysis. See text for details.

ACE is electrophoretic separation technique which is used study specific interactions between chemical or biological molecular species.[19] Experimental setup for ACE is conceptually depicted in **Figure B.1 Panel A, top row**. In ACE, a capillary is pre-filled with the run buffer that contains B. Species A is introduced into the capillary as a short plug, often as a part of an equilibrium mixture of A, B, and C (concentrations are denoted by the italicized letters: *A*, *B* and *C*). The total *B* (the sum of *B* and *C*) in the equilibrium plug is the same as *B* in the run buffer. The conditions are chosen so that the velocities of A and C in an electric field (v_A and v_C, respectively) differ. A and C are detected spectroscopically and their cumulative migration pattern is used to retrieve the information about reaction (1). It is important to indicate that equilibrium is not maintained in ACE experiments, as the formation of C causes localized depletion of B and, thus, variation of its concentration along the capillary. To avoid measurement errors caused by these concentration fluctuations, *B* is taken in excess of *A*.

Depending on how fast the equilibrium between A, B, and C is established, ACE electropherograms can present three general types of migration patterns corresponding to fast, slow, and intermediate equilibration. The assignment to one of the three cases is based on the relation between the characteristic equilibration time, t_{eq}, and the characteristic separation time, t_{sep}, which are defined as follows:

$$t_{eq} = 1/(k_+ B + k_-), \quad t_{sep} = w/|v_A - v_C| \qquad (2)$$

where w is the width of the initial zone of the equilibrium mixture. The cases of fast, slow, and intermediate equilibration, correspond to $t_{eq} << t_{sep}$, $t_{eq} >> t_{sep}$, and $t_{eq} \sim t_{sep}$, respectively. One can determine whether the equilibration is fast, slow or intermediate without analyzing equations (2), but by qualitatively analyzing the migration pattern of A and C. The three general cases of migration patterns are depicted in **Figure B.1, middle row**. In the case of $t_{eq} << t_{sep}$, the

equilibration between A and C in reaction (1) occurs much faster than separation of their respective zones, and, as a result, A and C will be moving as a single zone, producing a single peak in an ACE electropherogram. In the case of $t_{eq} >> t_{sep}$, the zones of A and C will be separated before equilibration between them proceeds to a significant extent. Thus, A and C will be moving as separate zones, producing two separate peaks. In the case of $t_{eq} \sim t_{sep}$, equilibration and separation proceed with comparable rates; therefore, A and C will be moving as two zones with a significant overlap between them.

Historically, ACE was applied almost exclusively to cases with $t_{eq} << t_{sep}$. The velocity of the combined A and C zone is defined by concentration-weighted average of v_A and v_C. Thus, gradual shifts in migration time of the combined peak will occur as a function of B. These shifts in velocity of the single peak can be used as the single-parameter response for plotting of an isothermal binding curve, from which K_d can be determined graphically or though Scatchard analysis (**Figure B.1 Panel A, bottom row**).[20] ACE data has been predominantly analysed by this classic single-parameter approach for K_d determination. The k_+ and k_- constants, however, are convoluted within ACE electropherograms, and are more difficult to extract. In 1993, Whitesides and co-authors reported on using a numerical computational approach to extract k_+ and k_- values from ACE data corresponding to the fast equilibration scenario.[16] This numerical approach, however, did not find any further application, most likely due to requirement of a considerable expertise in numerical computational methods and significant computing resources to produce accurate and stable solutions in a reasonable amount of time. More recently, Berezovski and colleagues developed a mathematical tool for approximation of k_+ and k_- from fast-equilibration ACE data by assuming rapid molecular exchange.[21] This tool, however, is only applicable to systems with large K_d values (between 8×10^{-5} mol L^{-1} and 3×10^{-3} mol L^{-1}) and

very fast equilibration times (between 9×10^{-4} s and 0.25 s), which limits its scope. Besides these two examples, ACE has not been used for determination of k_+ and k_-.

Systems that correspond to $t_{eq} >> t_{sep}$ and $t_{eq} \sim t_{sep}$ are generally not analyzed by ACE. Determination of K_d for slow equilibrating systems is possible by analyzing pre-equilibrated sample mixtures and measuring the ratio between the areas of separate peaks of A and C; however, ACE is rarely used for study of such systems as simpler approaches, with no B in the run buffer, can be successfully used for this purpose.[22,23] Intermediate equilibration systems result in intricate migration patterns that have never been analyzed due to a lack of proper deconvolution tools.

Initial and boundary conditions in MASKE resemble those in ACE (**Figure B.1 Panel B, top row)**. To maintain chemical equilibrium, the entire capillary is filled with the equilibrium mixture of A, B, and C. In a short plug, however, A, and accordingly C, are labeled for detection (labeled components are denoted with an asterisk). While A+A*, B, and C+C* are in chemical equilibrium, the label creates informational non-equilibrium and allows one to follow the kinetics in reaction (1). Chemical equilibrium is maintained in MASKE, which has facilitated the development of an analytical solution for its partial differential equations of mass transfer of A* and C*.[18]

The relation between t_{eq} and t_{sep} play similar roles in MASKE and ACE. MASKE experiments that study interacting systems with fast, slow, and intermediate equilibration result in migration patterns similar to those of ACE (**Figure B.1, middle row).** Unlike ACE, however, the analytical solution for MASKE allows k_+, k_-, and K_d to be extracted from all three types of migration patterns. One of the ways to achieve this is by fitting the available analytical solution for

$A^* + C^*$ (t) into experimental migration patterns of the labelled components, while varying k_+ and k_-.

MASKE is a recently developed method that has not yet led to significant accumulation of data. The requirement of MASKE for labeling of A represents the same limitation as found in the other label-based kinetic methods: labeling may affect the interaction between A and B, may be difficult, or expensive. This requirement also precludes the use of MASKE with the most popular label-free detection approach for CE – UV absorption spectroscopy. Simple and versatile experimental setup of ACE is generally preferable to that of MASKE.

The conceptual similarities between ACE and MASKE suggest that mathematical tools developed for MASKE may be adapted for analysis of ACE data. Ability to combine the simple experimental setup of ACE with the simple mathematical tools for MASKE would eliminate the limitations of each individual method. This study was motivated by the insight that the main difference between interaction conditions in ACE and MASKE lies in the concentration profile of B along the capillary: B is constant in MASKE, but not in ACE. However, if there is a sufficient excess of B over A then the deviation of B from its nominal value become insignificant. Consistent with this notion, *in-silico* simulated migration patterns of ACE and MASKE become more similar as B increases over A **(Figure B.2)**. Importantly, this condition is always satisfied in historical ACE experiments, as determination of K_d requires titration of B to achieve saturation in the formation of C.[20]

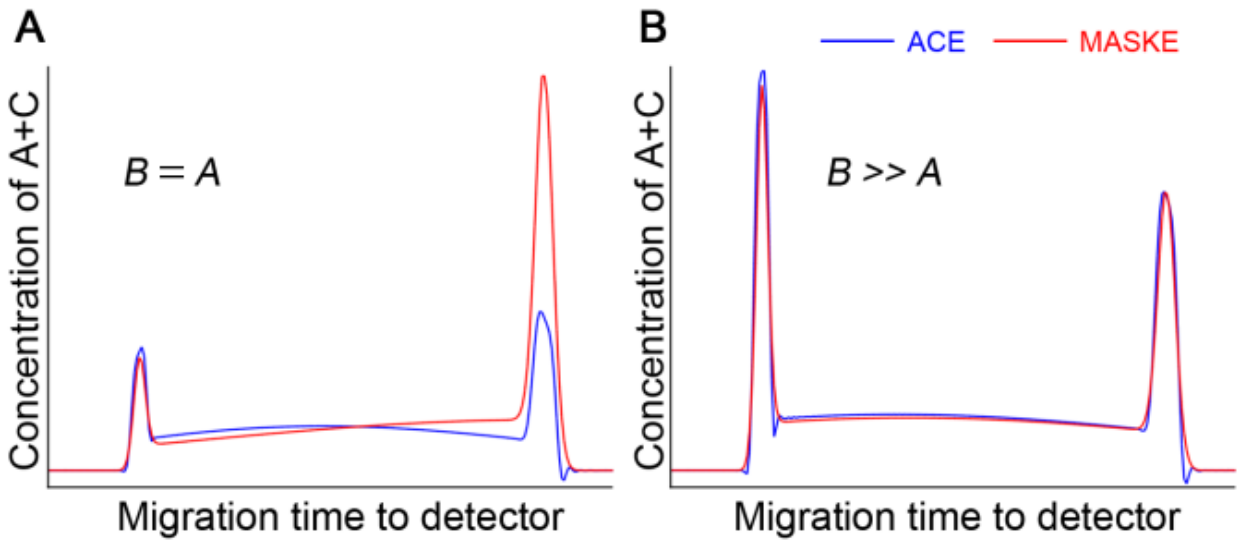

Figure B.2. Effects of concentration of B on deviation between ACE and MASKE results. *In-silico* simulated ACE (blue) and MASKE (red) migration profiles become more similar as total concentration of component B is increased over the concentration of component A. **Panel A:** *A* is equal to *B;* **panel B:** *B* is 10 times higher than *A*. ACE migration profile was simulated numerically, while MASKE migration profile was simulated using the available analytical solution. See text for simulation details. The profiles were simulated with the following values $k_+ = 3 \times 10^4$ $M^{-1}s^{-1}$, $k_- = 3 \times 10^{-3}$ s^{-1}, and $K_d = 100$ nM.

Similarity of differential equations and their solutions in ACE and MASKE

To determine if MASKE analytical solution can be adapted to ACE, we first show that differential equations of mass transfer that describe the two methods are identical if $B >> \max(A, C)$. Diffusion is neglected throughout this consideration. The MASKE equations, which have been derived elsewhere,[18] are the following:

$$\begin{aligned}(\partial_t + v_A\partial_x)A^* &= -k_+A^*B + k_-C^* \\ (\partial_t + v_C\partial_x)C^* &= k_+A^*B - k_-C^*\end{aligned} \qquad (3)$$

where ∂_x and ∂_t are partial derivations by spatial coordinate and time, respectively. In MASKE, A^* and C^* change with time and coordinate while $B = const$ no matter what the relation between

B and *A* is. In ACE, in general, the equation for B must also be considered so that the complete system involves three equations:

$$\begin{aligned}(\partial_t + v_A\partial_x)A &= -k_+AB + k_-C\\ (\partial_t + v_B\partial_x)B &= -k_+AB + k_-C\\ (\partial_t + v_C\partial_x)C &= k_+AB - k_-C\end{aligned} \qquad (4)$$

where v_B is the velocity of B. Conditions at $t = 0$ for system (4) have the following form in the plug

$$A = A_{eq}, \quad B = B_{eq}, \quad C = C_{eq} \quad (t = 0) \qquad (5)$$

Here, A_{eq}, B_{eq}, and C_{eq} are equilibrium concentrations in the plug at $t = 0$. They can be calculated using equations

$$A_{eq} + C_{eq} = A_0, \quad B_{eq} + C_{eq} = B_0, \quad \frac{A_{eq}B_{eq}}{C_{eq}} = K_d \qquad (6)$$

where A_0 and B_0 are concentrations of A and B used in the plug preparation (i.e. concentrations of A and B before formation of C).

By introducing dimensionless variables:

$$\hat{A} = \frac{A}{A_0}, \quad \hat{B} = \frac{B}{B_0}, \quad \hat{C} = \frac{C}{A_0} \qquad (7)$$

we can rewrite (4) as follows:

$$\begin{aligned}(\partial_t + v_A\partial_x)\hat{A} &= -k_+B_0\hat{A}\hat{B} + k_-\hat{C}\\ (\partial_t + v_B\partial_x)\hat{B} &= \lambda(-k_+B_0\hat{A}\hat{B} + k_-\hat{C})\\ (\partial_t + v_C\partial_x)\hat{C} &= k_+B_0\hat{A}\hat{B} - k_-\hat{C}\end{aligned} \qquad (8)$$

Here $\lambda = A_0/B_0$ is a small parameter if B is taken in a sufficient excess to A during the plug preparation. In this case, a solution for equations (6) can be obtained in a form of expansion in λ:

$$\begin{aligned}\hat{A} &= \hat{A}_0 + \lambda\hat{A}_1 + \ldots \\ \hat{B} &= \hat{B}_0 + \lambda\hat{B}_1 + \ldots \\ \hat{C} &= \hat{C}_0 + \lambda\hat{C}_1 + \ldots\end{aligned} \tag{9}$$

Substitution of (7) into the second equation in (6) and into conditions (5) and (6) yields:

$$(\partial_t + v_B\partial_x)\hat{B}_0 = 0, \quad \hat{B}_0 = 1 \quad (t = 0) \tag{10}$$

The first equation (10) has an obvious solution $\hat{B}_0(x,t) = const$. Given the second relation (10), we finally have $\hat{B}(x,t) = \hat{B}_0 = 1$ in the zeroth order of approximation in λ. Substitution of $\hat{B} = 1$ in the first and last equations in (8) reduces system of equations (8) to two equations:

$$\begin{aligned}(\partial_t + v_A\partial_x)\hat{A} &= -k_+B_0\hat{A} + k_-\hat{C} \\ (\partial_t + v_C\partial_x)\hat{C} &= k_+B_0\hat{A} - k_-\hat{C}\end{aligned} \tag{11}$$

After its transformation to the dimensional variables, system (11) is identical to MASKE equations (3) (with $B = B_0$). Thus, differential equations for MASKE (3) and ACE (4) are identical (in the zeroth order of approximation in λ) when $B_0 >> A_0$.

Since the differential equations for ACE (4) and MASKE (3) are approximately identical, their solutions will be similar if $\lambda << 1$ (with an error of the order of λ). The solution for MASKE equations (3) is given elsewhere.[18] It was obtained by transition to the special dimensionless form of equations (3) with subsequent application of Fourier transform in coordinate x. Resulting system of ordinary differential equations (in time t) was then solved for the following specific conditions at t=0 (denoted by subscript 1 or 2)

$$A_1(x,0) = \delta(x), \quad C_1(x,0) = 0 \tag{12}$$

$$A_2(x,0) = 0, \quad C_2(x,0) = \delta(x) \tag{13}$$

where $\delta(x)$ is the Dirac delta function. These solutions have the following form:

$$A_1 = \left(\frac{|\omega|}{2} \sqrt{\frac{\eta}{\mu}} \mathrm{I}_1(\rho) + \delta(\mu) \right) \gamma, \quad C_1 = \frac{|\omega| \varepsilon}{2} \mathrm{I}_0(\rho) \gamma,$$
$$A_2 = \frac{|\omega|}{2\varepsilon} \mathrm{I}_0(\rho) \gamma, \quad C_2 = \left(\frac{|\omega|}{2} \sqrt{\frac{\mu}{\eta}} \mathrm{I}_1(\rho) + \delta(\eta) \right) \gamma \qquad (14)$$

where

$$\varepsilon = \sqrt{k_+ B / k_-}, \quad \omega = 2\sqrt{k_+ k_- B} / (v_\mathrm{A} - v_\mathrm{C})$$
$$\mu = v_\mathrm{A} t - x, \quad \eta = x - v_\mathrm{C} t, \quad \rho = |\omega| \sqrt{\mu \eta}$$
$$\varphi = \omega(\varepsilon - 1/\varepsilon)/2, \quad \psi = \omega(v_\mathrm{C}\varepsilon - v_\mathrm{A}/\varepsilon)/2, \quad \gamma = \exp(t\psi - x\varphi),$$

and I_0, and I_1 are modified Bessel functions of the first kind. Physical meaning of these parameters can be described as follows: ε^2 is a ratio of characteristic times of reverse and forward reactions (1); $1/\omega$ is a characteristic length describing the separation of A and C; relations $\mu = 0$ and $\eta = 0$ give motion laws for peaks of A and C; ρ is a dimensionless combination of coordinate x and time t that appears while solving (3); φ, ψ, and γ are some combinations that arise in transition to special dimensionless form of (3); γ partially describes the effect of reactions (1) on concentrations of A and C. Relations (14) are valid at $\mu\eta > 0$. In the opposite case, when $\mu\eta < 0$, the obtained analytical solution is a trivial one with $A_1 = C_1 = A_2 = C_2 = 0$. Strictly speaking, conditions (12) and (13) at $t = 0$ as well as solutions (14) correspond to the case when concentrations of A and C in the capillary are defined as linear concentrations (i.e. amounts of A and C per unit length). In this case, we can still use equations (3) for linear concentrations A and C if B is defined as a volume concentration.

Solutions (14) represent two different types of Green functions for equations (3) with conditions (12) or (13) defined at $t = 0$. For arbitrary distributions of volume concentrations $A^*(x,0)$ and $C^*(x,0)$ at at $t = 0$ the general solution of (3) can be easily expressed in terms of these Green functions:

$$A^*(x,t) = \int \left(A_1(x-y,t)A^*(y,0) + A_2(x-y,t)C^*(y,0) \right) dy$$
$$C^*(x,t) = \int \left(C_1(x-y,t)A^*(y,0) + C_2(x-y,t)C^*(y,0) \right) dy \qquad (15)$$

Relations (15) combined with $B = B_0$ will also give (at $\lambda << 1$) an approximate solution for A and C in the case of ACE equations (4), if we omit the asterisk on the left-hand side of expressions (15) and replace $A^*(y, 0)$ and $C^*(y, 0)$ with $A_{eq}(y)$ and $C_{eq}(y)$ on the right-hand side of expression (13). After this, the substitution of expressions (14) for A_1, C_1, A_2, and C_2 yields (at $v_C > v_A$):

$$A(x,t) = A_{eq}(x - v_A t)\exp(-k_+ Bt) +$$
$$\int_{x-v_C t}^{x-v_A t} \frac{|\omega|}{2}\sqrt{\frac{\eta_y}{\mu_y}}\mathrm{I}_1\left(|\omega|\sqrt{\mu_y \eta_y}\right)\exp\left(t\psi - (x-y)\varphi\right)A_{eq}(y)dy + \qquad (16)$$
$$\int_{x-v_C t}^{x-v_A t} \frac{|\omega|}{2\varepsilon}\mathrm{I}_0\left(|\omega|\sqrt{\mu_y \eta_y}\right)\exp\left(t\psi - (x-y)\varphi\right)C_{eq}(y)dy$$

$$C(x,t) = C_{eq}(x - v_C t)\exp(-k_- t) +$$
$$\int_{x-v_C t}^{x-v_A t} \frac{|\omega|\varepsilon}{2}\mathrm{I}_0\left(|\omega|\sqrt{\mu_y \eta_y}\right)\exp\left(t\psi - (x-y)\varphi\right)A_{eq}(y)dy + \qquad (17)$$
$$\int_{x-v_C t}^{x-v_A t} \frac{|\omega|}{2}\sqrt{\frac{\mu_y}{\eta_y}}\mathrm{I}_1\left(|\omega|\sqrt{\mu_y \eta_y}\right)\exp\left(t\psi - (x-y)\varphi\right)C_{eq}(y)dy$$

where $\eta_y = x - y - v_C t$ and $\mu_y = v_A t - x + y$. In the case of $v_C < v_A$, integration in expressions (16) and (17) should be taken from $x - v_A t$ to $x - v_C t$.

It should be noted that all of the above equations use concentrations of A, B, and C, while in practice it might be more useful to operate with signal quantities of a given detection system (*e.g.* optical or electrochemical). It is instructive to consider the relations between the concentrations and signals using an example of fluorescence detection, although UV absorbance detection can be handled in the same way. If fluorescence detection is used, values A_{exp} and C_{exp} are related to measured fluorescence signals A_{fs} and C_{fs} as:

$$A_{\text{exp}} = \chi_{\text{A}} \frac{A_{\text{fs}}}{Q_{\text{A}}}, \quad C_{\text{exp}} = \chi_{\text{C}} \frac{C_{\text{fs}}}{Q_{\text{C}}} \tag{18}$$

where χ_A and χ_B are the proportionality coefficients, which depend on fluorophores and detectors used for A and C, and Q_A and Q_C are absolute quantum yields of A and C. In a typical case, a single fluorophore and a single detector are used for both A and C so that $\chi_A = \chi_B = \chi$. If we introduce a calibration coefficient $\alpha = \chi/Q_A$ then relations (16) can be presented in a more practical form:

$$A_{\text{exp}} = aA_{\text{fs}}, \quad C_{\text{exp}} = \alpha \frac{C_{\text{fs}}}{\beta} \tag{19}$$

where $\beta = Q_C/Q_A$ is a relative quantum yield, which is measured much easier than absolute quantum yields.

Finding k_+ and k_-

Relations (16) and (17) can be used to find k_+ and k_- from experimental ACE data in a "pattern-based approach".[24] To do this, one should use non-linear regression in which migration patterns of *A* and *C,* calculated with (16) and (17), are used to fit experimentally obtained A_{exp} and C_{exp} at various sets of k_+ and k_- and fixed known values of initial concentrations and migration velocities. The best fit will give values of k_+ and k_-. If A and C are indistinguishable in detection (*e.g.* spectrally), then the sum of $A + C$ should be calculated from (16) and (17) and fitted into the experimentally obtained sum of $A_{\text{exp}} + C_{\text{exp}}$. Pattern-based approaches for finding k_+ and k_- in ACE are technically similar to those used in SPR or stopped-flow kinetic methods.

Alternatively, simple multi-parameter-based approaches for approximating k_+ and k_- in MASKE have been developed, and are described elsewhere.[18,25,26] Such approaches are equally applicable to ACE. They rely on certain simplifying assumptions that, in turn, allow finding

algebraic equations for k_+ and k_-. The equations include several parameters such as migrations times, peak widths, peak heights, and peak areas. Different assumptions are made for slow, fast, and intermediate equilibration.

Dependence of ACE kinetic measurement accuracy on initial conditions

The mathematical tools described in the previous section can be used to extract k_+ and k_- from ACE electropherograms as long as the chosen initial concentration of B is sufficiently higher than that of A. From the practical standpoint, it would be useful to know how the accuracy of ACE kinetic measurements changes with varying ratios of these concentrations. To allow the end-users to determine suitable conditions for data analysis, we have performed an *in-silico* study of ACE kinetic measurement accuracy for different equilibration-time scenarios and over a wide range of component concentrations.

To simulate ACE migration patterns, we have employed a numerical computational approach, similar to the previously described.[16] We used COMSOL Multiphysics environment, version 4.3a, with the "Transport of Diluted Species" module, which incorporates mass balance equations, chemical reactions, convective and electrophoretic terms, and Fick's law of diffusion to approximate migration patterns of defined species within discrete-space geometry. Direct MUMPS solver was used to approximate concentrations of molecules A and B, and their complex C, in a time dependent manner over a defined 1D geometry that modeled a capillary. One hundred and seventy ACE migration patterns, with different combinations of k_+ and k_-, as well as values of A_0 and B_0, were generated. Analytical equations (16) and (17) were programmed into COMSOL Multiphysics variable set and the "Optimization" module was used to back-calculate the k_+ and k_- values analytically, with numerically-simulated curves set as the

global least-squares objectives. Direct PARADISO solver, using Lavenberg-Marquardt optimization method was used to determine k_+ and k_- values that produced the best fit.

Relative errors between the input and the back-calculated k_+ and k_- values for systems with different characteristic equilibration times appear in **Figure B.3.** To ensure highest sensitivity of measurement, B_0 was chosen to be equal to the K_d value, while A_0 was lowered to achieve various B_0/A_0 values. As expected, the largest error in determination of k_+ was observed when $B_0 = A_0$, with relative error exceeding 1 for systems with fast characteristic equilibration times. The errors reduced significantly when the B_0/A_0 ratio was increased to 2 and 5, while further increases to 10 and 20 produced less significant improvements. Regardless of ratios of concentrations used in the simulations, measurements in systems that had a very slow characteristic equilibration time ($t_{sep}/t_{eq} < 0.01$) produced high errors. This is due to the fact that in such systems few association and dissociation events occur during separation, thus resulting in migration patterns that contain little information on interaction kinetics. In contrast, small errors were observed in measurements of systems with moderately slow ($t_{sep}/t_{eq} \sim 0.1$) and intermediate ($t_{sep}/t_{eq} \sim 1$) equilibration times. This stems from the fact that such migration patterns (two peaks and a connecting bridge region) contain the most features that the curve-fitting software can use to extract unknown parameters. Simulated electropherogram peaks that resulted from analysis of systems with very fast equilibration times ($t_{sep}/t_{eq} > 10$) had shapes which were indistinguishable from the shape of the initial sample plug analyte distribution. As a result, kinetic constants could not be extracted from data obtained for such cases. Accuracy of k_- determination did not change significantly with A_0/B_0 ratios, as dissociation rates do not depend on B. These results suggest that a 5-fold excess of B_0 over A_0 is sufficient to conduct kinetic analysis of ACE data with relative errors under 20%.

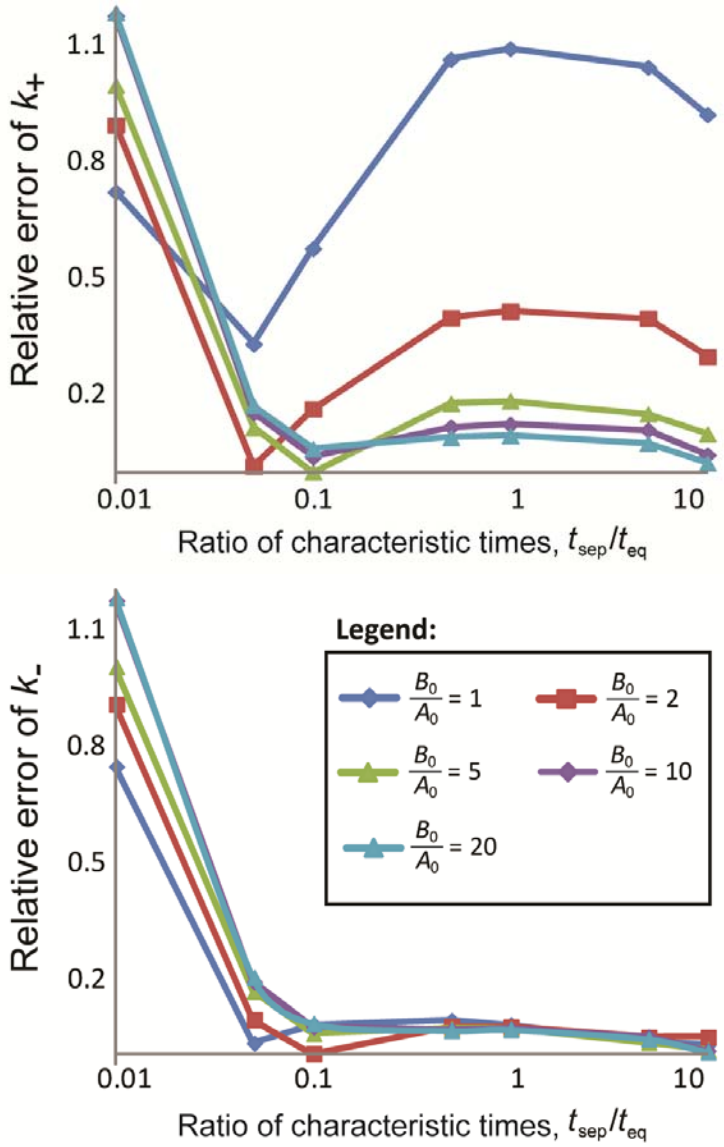

Figure B.3. Accuracy of ACE kinetic measurements at various B_0 and A_0. Top and bottom panels show relative errors of k_+ and k_-, respectively. The errors are defined as ratios between input and back-calculated values of the rate constants. Measurement errors were evaluated for interacting systems with characteristic equilibration times ranging from slow ($t_{sep}/t_{eq} = 0.01$) to fast ($t_{sep}/t_{eq} = 10$) scenarios. For all traces B_0 was chosen to equal K_d, while A_0 was lowered to achieve varying values of B_0/A_0 (see the legend in the figure).

Retrospective analysis of historical ACE data

To demonstrate the power of the new approach in retrospective analysis of ACE data, we applied it to two different historical studies. The first study is the prominent work by Whitesides and co-authors,[16] where a numerical approach was applied for the first time to extract k_+ and k_- values from ACE data. Using this data has enabled us to compare and validate our results against the values produced by the original study. This study investigated the affinity interaction between bovine carbonic anhydrase (CA) and a charged arylsulfonamide ligand (ASL). In this experiment, CA and ASL correspond to molecules A and B, respectively.

To calculate the ACE migration patterns using our analytical solution, the following information was obtained from the "materials and methods" section of the original article: initial concentrations of CA and ASL, applied voltage, total and effective lengths of the capillary, and the length of the initial sample plug. Concentration-to-signal conversion multipliers and electrophoretic mobilities of free CA and CA-ASL complex were approximated from experimental data. Electrophoretic mobility of free ALS was approximated to be 50% slower than that of CA-ASL complex: precise knowledge of this value is not required, as electrophoretic mobility of B does not have a significant effect on migration patterns of systems that undergo fast equilibration ($t_{eq} << t_{sep}$). The experimental electropherograms were digitized from the original publication using open-source Engauge Digitizer software, version 4.1. Equations (16) and (17) were programmed into COMSOL Multiphysics variable set and the "Optimization" module was used in the same way as described in the previous section, except the digitized experimental curves were used as the global least-squares objectives. In the original experiments, CA was maintained at a concentration of 1×10^{-5} mol L^{-1}, while ASL was titrated between 0 and 2×10^{-4} mol L^{-1}. Only the electropherograms with $B_0/A_0 \geq 5$ were used for fitting aiming at

accurate determination of k_+ and k_- **(Figure B.4 Panel A)**. The obtained values, $k_+ = (4.36 \pm 0.09) \times 10^4$ L mol^{-1} s^{-1}, $k_- = 0.31 \pm 0.08$ s^{-1}, and $K_d = (7.03 \pm 0.03) \times 10^{-6}$ mol L^{-1}, were similar to the values reported in the original publication, $k_+ = 1.5 \times 10^4$ L mol^{-1} s^{-1}, $k_- = 0.1$ s^{-1}, and $K_d = 7 \times 10^{-6}$ mol L^{-1}. The errors reported with our values represent the precision of fitting different experimental traces, and are indicative of the high quality of the original experimental data. It should be noted, that according to our *in-silico* study, described in the previous section, a systematic error of 10-20% might be affecting our results, potentially explaining the difference between our values those reported in the original study. Furthermore, this difference may also be explained by the fact that the original study considered only 3 possible combinations of k_+ (1.5×10^5, 1.5×10^4, and 1.5×10^3 L mol^{-1} s^{-1}) and k_- (1, 0.1, and 0.01 s^{-1}) values in their simulations, most likely due to the long-time requirements of numerical computation. Our analytical solution allowed us to simulate hundreds of possible k_+ and k_- combinations within a few minutes, likely yielding more refined values. As predicted by our *in-silico* study, the electropherograms simulated with the obtained k_+ and k_- values did not produce a satisfactory fit with the experimental data when B_0/A_0 was significantly below 5, but improved as the value approached 5 **(Figure B.4 Panel B)**. These findings strongly support the validity of the new mathematical approach.

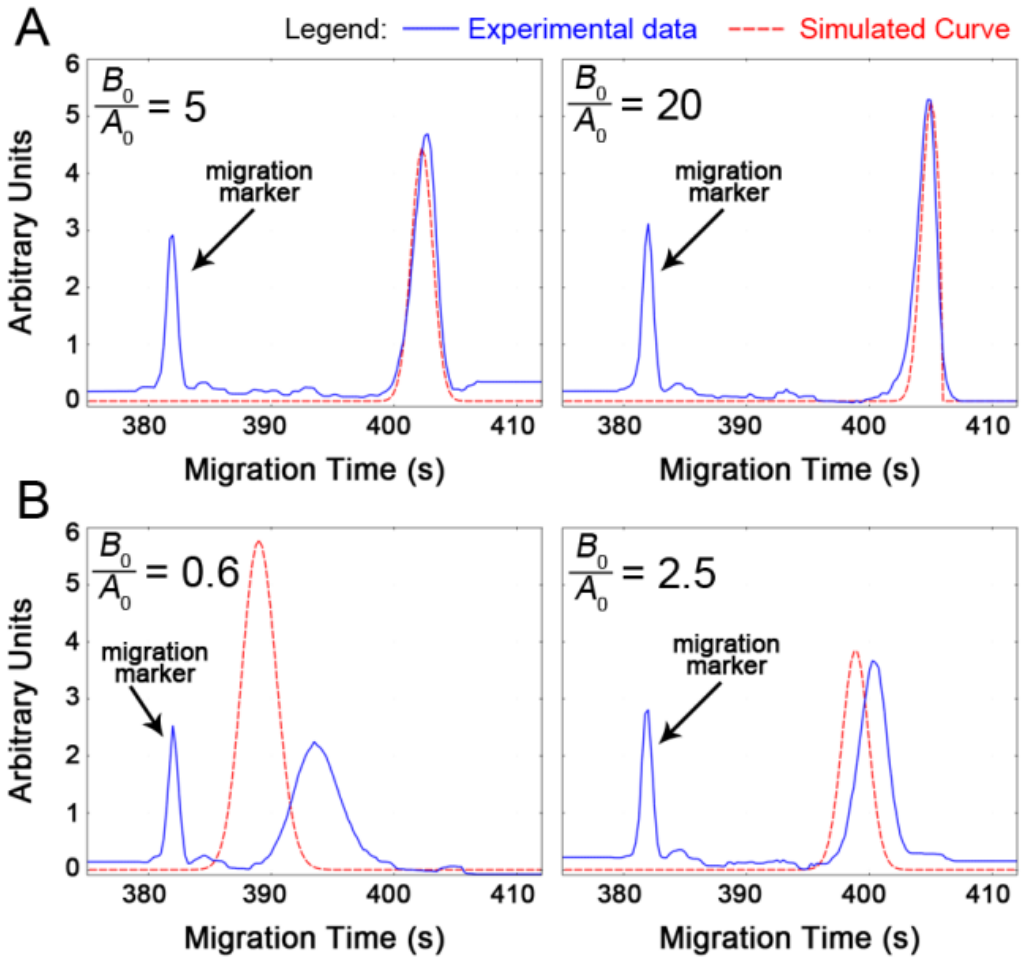

Figure B.4. Fitting of ACE experimental data for interaction between CA and ASL. **Panel A:** k_+, k_- and K_d were determined by fitting the historical data in which $B_0/A_0 \geq 5$. **Panel B:** curves simulated with the obtained parameters poorly matched the experimental data when $B_0/A_0 << 5$, but improved as the ratio approached 5. The peak on the left is a non-interacting migration marker, and was not included in the fitting.

Lastly, the new mathematical tool was applied for retrospective analysis of a recent study by Liu and colleagues.[27] This work simultaneously investigated the affinity interaction between three boronic acids (molecules A) and fructose (molecule B). The boronic acids under investigation were phenylboronic acid (PB), 3-carboxyphenylboronic acid (CP), and 3-carboxybenzoboroxole (CB). Experimental conditions and experimental traces were retrieved

from the original publication and analyzed as described in the first example. The following values were obtained by fitting: for PB $k_+ = 176 \pm 19$ L mol^{-1} s^{-1}, $k_- = 1.10 \pm 0.01$ s^{-1}, $K_d = (6.3 \pm 0.7) \times 10^{-3}$ mol L^{-1}; for CP $k_+ = 19.5 \pm 9.7$ L mol^{-1} s^{-1}, $k_- = 0.11 \pm 0.04$ s^{-1}, $K_d = (6.1 \pm 2.7) \times 10^{-3}$ mol L^{-1}; and for CB $k_+ = 179 \pm 108$ L mol^{-1} s^{-1}, $k_- = 0.6 \pm 0.3$ s^{-1}, $K_d = (4.0 \pm 0.9) \times 10^{-3}$ mol L^{-1} **(Figure B.5)**. To the best of our knowledge, this is the first instance when kinetics of interaction for these molecular pairs have been reported. This example clearly demonstrates the power and capacity of the new approach to produce new and valuable information from existing data. We anticipate that the new mathematical tool will find a wide use in both the analytical and pharmacological communities that employ, or have employed, ACE.

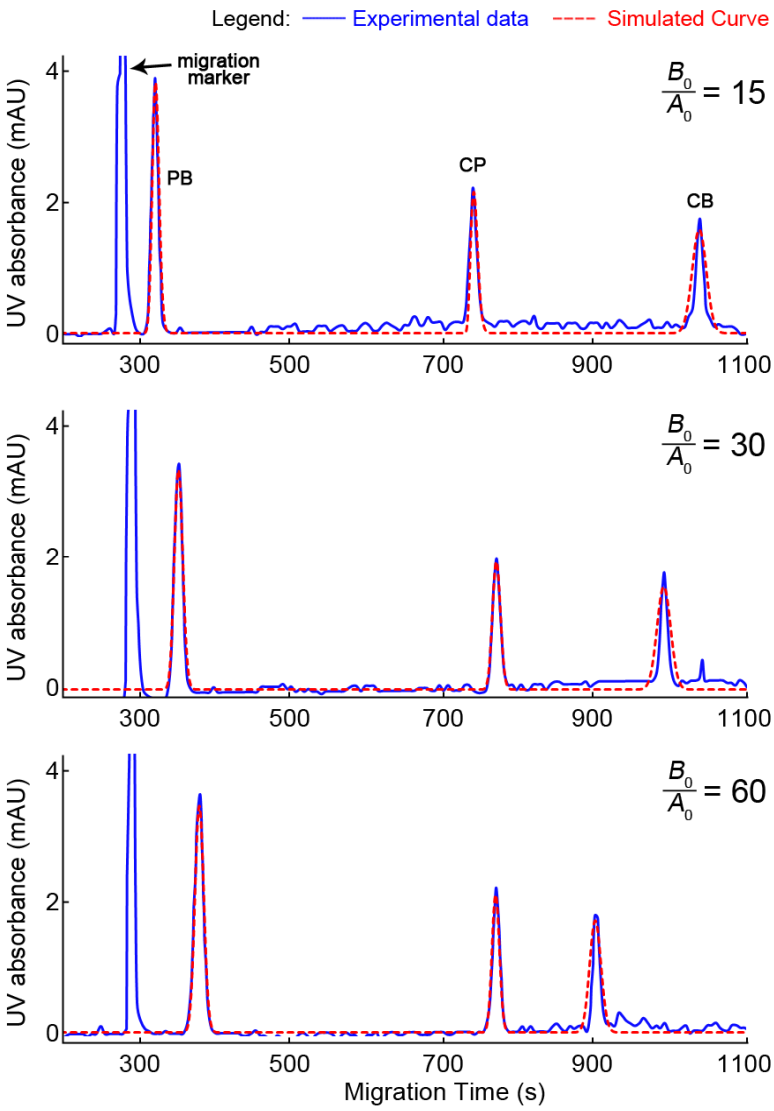

Figure B.5. ACE results for interaction between boronic acids and fructose. Experimental and simulated ACE migration patterns describing interaction between PB, CP and CB (molecules A) and fructose (molecule B). From top to bottom: determination of k_+, k_- and K_d by fitting historical data in which B_0/A_0 were equal to 15, 30 an 60. The peak on the left is a non-interacting migration marker, and was not included in the fitting. Peaks for each boronic acid species was fitted individually, but then combined into a single curve, containing all three peaks, for convenient representation.

B.4. Concluding remarks

We proved theoretically and confirmed through retrospective analysis of previously published data that the propagation patterns of A and C in ACE and A* and C* in MASKE are identical if $B_0 >> A_0$. Choosing B_0 that is at least 5-fold higher than A_0 will satisfy this condition for majority of interaction systems. This allows one to apply general solution (12) to ACE electropherograms. This also enables the use of the simplified approximate methods for finding rate constants from ACE data. Kinetic measurements by ACE become especially useful when labeling of component A is undesirable. Given the fact that in ACE the component B is always taken in excess to component A, the new mathematical tool can be used retrospectively to extract kinetic information from all historical ACE experiments for which the original electropherograms are available.

B.5. Appendix references

1. Zhao, S.; Fernald, R. D. *J. Comput. Biol.* **2005**, *12*, 1047-1064.
2. Chu, Y. H.; Avila, L. Z.; Gao, J. M.; Whitesides, G. M. *Accounts Chem. Res.* **1995**, *28*, 461-468.
3. Neubert, R. H. H.; Schwarz, M. A.; Mrestani, Y.; Platzer, M.; Raith, K. *Pharm. Res.* **1999**, *16*, 1663-1673.
4. Hellman, L. M.; Fried, M. G. *Nat. Protoc.* **2007**, *2*, 1849-1861.
5. Leavitt, S.; Freire, E. *Curr. Opin. Struct. Biol.* **2001**, *11*, 560-566.
6. Chen, H. M.; Puhl, H. L.; Ikeda, S. R. *J. Biomed. Opt.* **2007**, *12*, 9.
7. Ozers, M. S.; Hill, J. J.; Ervin, K.; Wood, J. R.; Nardulli, A. M.; Royer, C. A.; Gorski, J. *J. Biol. Chem.* **1997**, *272*, 30405-30411.
8. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S. *Nat. Commun.* **2010**, *1*, 7.
9. Copeland, R. A. *Expert. Opin. Drug Discov.* **2010**, *5*, 305-310.
10. Schuck, P. *Annu. Rev. Biophys. Biomolec. Struct.* **1997**, *26*, 541-566.
11. Olson, S. T.; Srinivasan, K. R.; Bjork, I.; Shore, J. D. *J. Biol. Chem.* **1981**, *256*, 1073-1079.
12. Schwille, P.; Haupts, U.; Maiti, S.; Webb, W. W. *Biophys. J.* **1999**, *77*, 2251-2265.
13. Osmond, R. I. W.; Kett, W. C.; Skett, S. E.; Coombe, D. R. *Anal. Biochem.* **2002**, *310*, 199-207.
14. Secundo, F. *Chem. Soc. Rev.* **2013**, *42*, 6250-6261.
15. Chu, Y. H.; Avila, L. Z.; Biebuyck, H. A.; Whitesides, G. M. *J. Med. Chem.* **1992**, *35*, 2915-2917.

16. Avila, L. Z.; Chu, Y. H.; Blossey, E. C.; Whitesides, G. M. *J. Med. Chem.* **1993**, *36*, 126-133.

17. Heegaard, N. H. H.; Kennedy, R. T. *Electrophoresis* **1999**, *20*, 3122-3133.

18. Krylov, S. N.; Okhonin, V.; Berezovski, M. V. *J. Am. Chem. Soc.* **2010**, *132*, 7062-7068.

19. Riekkola, M. L.; Jonsson, J. A.; Smith, R. M. *Pure Appl. Chem.* **2004**, *76*, 443-451.

20. Colton, I. J.; Carbeck, J. D.; Rao, J.; Whitesides, G. M. *Electrophoresis* **1998**, *19*, 367-382.

21. Mironov, G. G.; Okhonin, V.; Gorelsky, S. I.; Berezovski, M. V. *Anal. Chem.* **2011**, *83*, 2364-2370.

22. Heegaard, N. H. H.; Nilsson, S.; Guzman, N. A. *J. Chromatogr. B* **1998**, *715*, 29-54.

23. Jiang, C. X.; Armstrong, D. W. *Electrophoresis* **2010**, *31*, 17-27.

24. Petrov, A.; Okhonin, V.; Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 17104-17110.

25. Cherney, L. T.; Krylov, S. N. *Anal. Chem.* **2011**, *83*, 1381-1387.

26. Cherney, L. T.; Krylov, S. N. *Anal. Chem.* **2012**, *137*, 1649-1655.

27. Lu, C. C.; Li, H. Y.; Wang, H. Y.; Liu, Z. *Anal. Chem.* **2013**, *85*, 2361-2369.