NOVEL KINETIC SOLUTION-BASED SEPARATION APPROACHES FOR SMALL MOLECULE DRUG DISCOVERY

JIAYIN BAO

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

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

January 2017

ABSTRACT

The modern pharmaceutical industry has achieved remarkable successes in medicinal chemistry. However, many diseases are incurable due to the difficulty of finding new drugs. *De novo* drug discovery contains two steps: the primary screening focuses on selecting protein (target) binding drug (ligand); the secondary screening concentrates on calculating kinetic binding parameters of target-ligand complex. Conventional methods for the primary screening are typically surface-based, which suffer intensely from nonspecific interactions; the existing methods for secondary screening are either affinity-based or require surface immobilization, both cannot accurately calculate kinetic binding parameters. Hence, this research focuses on the development of the solution-based kinetic platform that facilitates both primary and secondary screenings. We combined kinetic capillary electrophoresis (KCE) with DNA-encoded ligand (DEL) technology to build a solution-based platform for primary screening of ligands. KCE offers high partitioning efficiency but requires the knowledge of electrophoretic mobility of target-ligand complex, and thus, we developed a mathematical model to predict electrophoretic mobility of target-DEL complex. This model was tested by using the targets interacted with 18 artificial DELs that contain various combinations of dsDNA and ssDNA regions, together with 2 DELs manufactured by GlaxoSmithKline. The results confirmed the precision, accuracy, and ruggedness of our model. This model will facilitate the reliable use of KCE-DEL based primary screening. Next, we developed a kinetic size-exclusion chromatography-mass spectrometry (KSEC-MS) as the label-free solution-based platform for calculating kinetic binding parameters of target-ligand interactions in secondary screening. KSEC-MS employs size-exclusion chromatography to separate small molecule ligand from protein target-ligand complex without immobilization and mass spectrometry to detect small molecule without a label. The rate

constants of complex formation and dissociation are calculated from the temporal ligand concentration profile. Methods of KSEC-MS have been developed by using 2 proteins with the corresponding drugs. The resulted kinetic and affinity binding parameters were validated, which confirmed the precision and accuracy of KSEC-MS. We foresee that the KSEC-MS will become a universal approach for the kinetic analysis of target-ligand interactions in secondary screening.

Dedication

To Shane Shane: the sunshine of my life

ACKNOWLEDGEMENTS

I would like to extend my thanks to many people, who generously contributed to the work presented in this dissertation. This work could not have been finished without help from them. It is a privilege and pleasure to acknowledge people who have given me guidance, support, and encouragement.

First and foremost, I would like to thank my supervisor Prof. Sergey Krylov. He offered me an opportunity to participate in this talented research team. He has always been positive and supportive throughout my Ph.D. study. His guidance, knowledge and endless encouragement supported me for the past eight years of research.

Special thanks go to my co-supervisor Prof. Derek Wilson and Prof. Philip Johnson as well as my advisor Prof. Gerald Audette, who offered me valuable discussion, inspiration as well as hands-on support throughout my research. I would like to express my sincere appreciation to Dr. Lana Krylova, Dr. Leonid Cherney, Dr. Yves Le Blanc, and Dr. Oren Reinstein. My work would not be possible without their continuous support and contribution.

I would like to extend my thanks to our industrial collaborators: GlaxoSmithKline, for their valuable contributions in synthesizing DNA-encoded small molecule ligands; and AB Sciex, for providing mass spectrometer and technical support in KSEC studies. Also, I would like to thank our academic collaborators from York University: Wilson group, for providing mass spectrometer and technical assistance in KSEC work; and Johnson group, for providing FPLC, ITC, and technical support in protein purification and affinity analysis. Moreover, I would like to thank all collaborators for their contributions in preparing publications.

My special appreciation goes to my examiners Prof. Kagan Kerman and Prof. Georg Zoidl and also Sven, Alex, Agnesa, who offered great help in editing this dissertation. Also, I would

like to thank my friend and former lab member Fletcher who offered me great support and encouragement during the tough times in my study. I would also like to thank all current and former lab members: Mirzo, Victor, Liang, Dave, Roman… I could not do it without the support from you. Moreover, I want to express my appreciation and blessing to my former undergraduate thesis students Funmi and Kunal, who are brilliant, dedicated and determined young professionals.

Finally, I would like to express my gratitude to my parents, my son Shane and my boyfriend Sven, whose support, patience, and encouragement helped me through the hard times. My deepest appreciation goes to them for their love and understanding. I could not be here today without their blessing and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

COMMONLY USED ABBREVIATIONS

ACZ: acetazolamide

AGP: α1-acid glycoprotein

APCI: atmospheric-pressure chemical ionization

CAII: carbonic anhydrase II

CE: capillary electrophoresis

DEL: DNA encoded ligand

DELSMs: DNA-encoded libraries of small molecules

DHFR: dihydrofolate reductase

dsDNA: double stranded DNA

EM: equilibrium mixture

EOF: electroosmotic flow

ESI-MS: electrospray ionization mass spectrometry

HTS: high throughput screening

ITC: isothermal titration calorimetry

KCE-DEL: kinetic capillary electrophoresis facilitated DNA-encoded ligand selection

KCE-MS: kinetic capillary electrophoresis mass spectrometry

KCE: kinetic capillary electrophoresis

K_d: equilibrium dissociation constant

k_{off}: kinetic rate constants of dissociation

k_{on}: kinetic rate constants of association

KSEC-MS: kinetic seize-exclusion chromatography-mass spectrometry

L: ligand

LIF: laser induced fluorescence

MS: mass spectrometry

MTX: methotrexate

NECEEM: non-equilibrium capillary electrophoresis of equilibrium mixture

P: protein

PCR: polymerase chain reaction

P-SM: protein-small molecule binding complex

peKSEC: pre-equilibration kinetic size-exclusion chromatography

ppKCE-UV: plug-plug kinetic capillary electrophoresis with UV detection

ppKSEC: plug-plug kinetic seize-exclusion chromatography

RSD: relative standard deviation

SA: streptavidin

SEC: size-exclusion chromatography

SM: small molecule

SPR: surface plasmon resonance

SSB: single-stranded DNA binding protein

ssDNA: single stranded DNA

T-L: target-ligand binding complex

T: target

CHAPTER 1: NTRODUCTION TO ANALYSIS OF PROTEIN-SMALL MOLECULE INTERACTIONS

1.1 Overview

The history of pharmaceutical development can be divided intro three eras. The first drug based therapies originated from the 19th mainly came from natural sources and treatments relied on serendipity. The majority of natural compounds were administered without sufficient knowledge of toxicity; therefore, these compounds often failed to serve as useful therapeutics [1]. At the beginning of the 20th century, the modern pharmaceutical industry began to use a variety of new, innovative technologies. For instance, the advancement of X-ray crystallography [2], as well as NMR technologies [3], allowed scientists to study the chemical structures of many therapeutic agents. The birth of recombinant DNA technology [4] facilitated the development of protein-based drug targets. Also, the progression in synthetic chemistry, high throughput screening (HTS), as well as computer-based molecular modeling significantly enhanced the screening measures of pharmaceutical agents. Entering the 21st century, the initiation of "omics" technologies [5] greatly improved the identification of disease targets and also boosted the development of biopharmaceuticals.

Despite these innovations and advances, modern pharmaceutical development still is a long and expensive process. On average, it takes 10-12 years from the initial research to final commercial product and costs approximately US$ 1-1.2 billion to develop a single drug [6]. The entire process of pharmaceutical discovery and development involves four major stages: drug discovery, pre-clinical development, clinical trial, and commercialization. The first stage, drug discovery starts with finding a disease-causing target. The disease-causing targets are commonly

referred to drug targets or biological targets. They are generally the proteins or gene sequences presented in the human body. In the process of drug discovery, the targets are usually used like baits to "lure out" binders from a large pool of ligands, a process typically referred to ligand screening. Ligand screening can be further divided intro primary and secondary screenings. The primary screening focuses on the preliminary selection of all target binders from highly populated combinatorial libraries, which typically contain 10^5 to 10^6 entities. Methods of HTS are usually engaged, which intend to select every binding ligand and rule out any non-binders with the highest efficiency. After primary screening, all selected binders are subjected to secondary screening, in which ligands are investigated by various analytical techniques. Affinity and kinetic parameters are calculated and utilized for ranking. Upon completing the stage of drug discovery, there are typically 10^3 potential leads generated before carrying into the next stage. The second stage is preclinical development, in which potential leads are further tested by extensive *in vitro* and *in vivo* assessments in combination with iterative modifications to optimize the pharmacodynamics, and pharmacokinetics. The third stage is the clinical trial, in which the potential drug candidates are given to human subjects with the dosage form and quantity intended for marketing. The human subjects range from healthy volunteers to patients with various illness levels. The drug efficacy and toxicity are extensively investigated in this stage; many potential leads failed here. The final stage is commercialization, in which the drug products are manufactured, and the application of a new drug has to be filed and approved by a regulatory agency such as the FDA. In addition, the post-approval studies are also conducted for monitoring the safety and performance of the commercialized drugs [7].

Although the modern pharmaceutical industry has grown for over a century, there are still many diseases untreated; likewise, despite the vast amount of known chemical and protein

structures, there are still very limited numbers of structurally different drugs. The Nobel prize-winning J.W. Black once had a famous quote, *"the most fruitful basis for the discovery of a new drug is to start with an old drug"* [8]. For the past decades, research and development in pharmaceutical industry spent more than US$ 50 billion a year to pursue only a limited number of drug targets. Until now, only 10% of disease-related genes have been explored and developed in drug discovery campaigns; similarly, only 6% of drug products on market today are structurally unique [9]. All this information leads to the fact that discovering the new drugs becomes the Holy Grail in today's pharmaceutical industry.

However, identification of novel small molecule ligands remains a practical burden in drug discovery. Conventional methods for ligands screening are inadequate by definition: these methods either suffer intensely from nonspecific interactions or cannot accurately calculate the true kinetic parameters. Therefore, our research focuses on the development of tool sets that contain novel homogeneous solution-based kinetic approaches, which can be used to facilitate both primary and secondary screenings with highest efficiency and accuracy that beyond the stretch of existing methods.

1.2 Targets, Ligands, and Target-Ligand Interactions

1.2.1 Drug Targets

The process of modern drug discovery starts with the identification of disease-causing targets. Disease targets are usually mutant or native proteins from either human body or any parasitic organism, and their functions can be modified upon interacting with the therapeutic agents. While there are multiple ways of finding new targets, research scientists generally follow

two main approaches. Analysis of pathophysiology is a safe and logical approach. This method starts with an understanding of the biochemical pathway that is responsible for the disease phenotype. Afterward, it verifies the particular biochemical step that is amenable to therapeutic intervention. Finally, it selects the key molecule as the target. Alternatively, drug targets can be discovered through a backward procedure. This starts with the analysis of action mechanisms of existing drugs, and then search for the matching targets [7]. In 2006, a comprehensive survey conducted by Imming *et al.* indicated that there were 324 unique protein targets at the time, 266 were human genome derived proteins and remainders came from pathogenic organisms such as bacteria, virus, and fungus [10]. Proteins and glycoproteins, such as enzymes, receptors, ion channels, transport proteins, and antibodies are the most common type of targets. Other disease-causing molecules (for example, gene sequences and ribosomes) have also been discovered and pointed for therapeutic purposes.

1.2.2 Small Molecules Ligands

While there is no strict definition of small molecules, they usually are low molecular weight (less than 1,500 Da) organic molecules [11]. Small molecules are extremely powerful tools as they can quickly penetrate through cell membranes, bind to targets, initiate biomolecular interactions, and then manipulate corresponding biochemical processes. The survey as mentioned earlier published by Imming *et al.* revealed that by 2006, there were 1357 unique drugs, 1204 of these were small molecules and 166 were biotherapeutic agents [10]. Thus, most of the available drugs on the market are small molecules.

1.2.2.1 The Ligand Library

In general, small molecules can be obtained either from the natural sources or chemical synthesis. Natural products such as leaves, shoot, barks, roots, and even marine life forms contain massive amounts of bio-relevant chemical compounds, which have been exploited medicinally for millennia and they still serve as a valuable source of drugs in the modern pharmaceutical industry [2]. One of the most significant benefits of natural products is to diversify the chemical compositions of the existing chemical libraries. However, due to difficulties in sourcing, isolating and identifying bioactive components, production of highly populated natural products library is impractical. Also, analysis and chemical modification of these incredibly complex structures are also time consuming and labor intensive. Therefore, chemical synthesis is considered the most practical approach in manufacturing highly populated small molecule libraries.

Methods of small molecule synthesis can be either target-oriented or diversity-oriented [11]. In target-oriented synthesis, compounds structures are deliberately created to interact with the targets. Accordingly the knowledge of target structures and rational designs are engaged. Diversity-oriented synthesis, on the other hand, aims to maximize the structural diversity without consideration of target structures. The choice is case dependent. For instance, in occasions when dealing with the well-defined targets, the rational design approach is often desirable, in particular, if the structures of nature ligands of such targets are known. Alternatively, when the character of the disease is not well understood, and neither the target nor the ligand was previously established, rational design becomes unrealistic; thus, diversity-oriented synthesis would be the only option.

1.2.2.2 DNA-Encoded Chemical Libraries

Conventional methods for selecting target-binding ligands require high throughput screening of highly populated small molecule libraries, which typically cover up to few million compounds. In HTS, the ligands are usually selected based on their affinity interactions with the target. Sample concentrations can profoundly influence affinity interactions; as a general rule, affinity binding interactions should be conducted with the sample concentration levels at the range of equilibrium dissociation constant K_d. Therefore, high-affinity ligands are usually present in low concentrations, which are tough to isolate and identify.

In contrast to small molecules, identification of biopolymers such as DNA and RNA can be successfully and elegantly accomplished even with very low concentrations. Various display technologies have been developed to achieve the identifications of biopolymer-attached ligands to targets; examples include phage display, yeast display, ribosome display, as well as mRNA display. Inspired by display technologies, in 1992, Brenner and Lerner [12, 13] introduced the concept of associating DNA-based barcode tag with synthetic peptides. Since then, numerous strategies have been implemented for constructing the DNA-encoded ligand (DEL) libraries. Methods of DEL libraries constructions can be mostly classified into two main types. The first type relies on stepwise split-and-pool of multi-step building block assembly, with parallel DNA coding fragments [14]. Such approaches have been widely developed and applied by GlaxoSmithKline (GSK), X-Chem, Nuevolution, and Philochem. The second type is based on DNA-templated synthesis. Pavel Sergeev first introduced the original concept of nucleic acid template chemical synthesis, in which the corresponding DNA-fragment tags hybrid and brought chemically active groups close together, thus, promoting chemical reactions. This approach allows very specific chemical reaction and bypasses the need for using the chemical protecting

groups. DNA-templated synthesis is a popular method for the production of DEL libraries and has been extensively applied by Ensemble Therapeutics, and Vipergen [11].

1.2.3 Target-Ligand Interactions

The potency of any therapeutic agent relies greatly on binding properties towards the corresponding disease target. The binding interactions between targets and ligands are either covalent or non-covalent. Covalent bonds are strong, being formed between interacting pairs of atoms through sharing of electrons. Although covalent drugs have proved to be potent therapeutics, due to the safety concerns, they are rarely considered when initiating a drug discovery process. Pharmaceutical companies fear that covalent drugs were so reactive and permanently bound to the wrong targets leading to toxicity [15]. In most cases, the pharmaceutical industry has, by all means, avoided development of drugs that commit to "marriage" and instead pursues "dating" as the mode of interaction. Alternatively, non-covalent drugs can form affinity interactions with the protein targets. The attraction forces in non-covalent interactions are commonly formed through hydrogen bonding, van der Waals forces, and hydrophobic effects [1].

1.2.3.1 Non-Covalent Target-Ligand Interactions

In biological systems, most of the biomolecular interactions are governed by non-covalent interactions. Likewise, in modern pharmaceutical discovery, most of the drugs are designed to interact with disease target in a non-covalent fashion [10, 16]. Non-covalent interactions can be defined by:

$$T + L \xrightleftharpoons[k_{off}]{k_{on}} T\text{-}L$$

(1-1)

target (T) is typically a macromolecule such as protein or nucleic acid, the ligand (L) is a small

molecule, and binding complex (T-L) is the product formed from interaction between T and L.

Generally, the overall affinity or complex stability is defined by equilibrium dissociation

constant $K_d = k_{off}/k_{on}$ and a small K_d value refers to high binding affinity and complex stability.

The kinetic rate constants k_{off} and k_{on} define the rate of complex dissociation and association

respectively. Assessing the kinetic rate constants is critical in understanding the dynamics of

non-covalent interactions between the target and ligand, which is also essential in determining

drug efficacy and toxicity.

1.2.3.2 Kinetics and Drug-Target Residence Time

For more than a century, the molecular basis of target-ligand interactions has been

predominantly judged by simple binding affinities; interactions between target and ligand were

perceived as structural complimentary in static states. However, in the late 20th century,

advancement in biophysical studies suggested that the target-ligand interactions are often

achieved via adaptive structural changes, in a dynamic manner. Over the past two decades,

various studies demonstrated that the durability and potency of drugs depend rather on the half-

life of target-ligand complex than simple overall complex affinity. Additionally, in the context of

a biological system, the duration of pharmacological effect, or drug-target residence time,

prominently depends on the temporal stability of the target-ligand binding complex. Ideally, in a

closed (*in vitro*) system, where target and ligand concentrations stay constant, the drug-target

residence time depends on the rate constants of both target-ligand association (k_{on}) and

dissociation (k_{off}). Conversely, in the case of open (*in vivo*) system, where active efflux consistently eliminating drug compounds, residence time primarily depends on rate constant of complex dissociation k_{off}. Decisions of selecting desirable drug-target residence time become case dependent. Prolonged residence time often enhance the durability of pharmacological effect, however, might also cause the off-target toxicity [17]. Therefore, the quantitative assessments of kinetic rate constants are critical in understanding drug-target residence time, which is connected to the overall pharmacological and toxicological effects.

1.3 Analytical Methods in Ligand Discovery

In pharmaceutical development, the conventional analytical approaches can be generalized as surface-based or solution-based; each approach can be further defined as either affinity or kinetic method. Surface-based approaches comprise immobilization of targets or ligands onto sensor surface or stationary phase of the chromatography column. Given the extraordinary screening power, surface methods are a rule of thumb in primary screening. Examples of surface-based affinity methods include protein microarray, small-molecule microarray, and affinity chromatography; there are also surface-based kinetic methods such as surface plasmon resonance (SPR) and biolayer interferometer (BLI). Compared to surface-based methods the solution-based approaches are relatively low throughput, as they typically require the purified/identified ligands to work with. However, solution-based methods are renowned for high precision and accuracy in calculating binding parameters, and thus often used in secondary screening. For example, the "gold standard" isothermal titration calorimetry (ITC) as well as electrospray ionization mass spectrometry (ESI-MS) are typically applied in determining the equilibrium dissociation constant K_d, and also in ranking binding affinities for various ligands against one target. Nevertheless,

each and every approach carries inherent systematic biases that often result in false negative or false positive outcomes. Therefore, the trail of analytical methodologies indeed determines the fate of drug discovery.

1.3.1 Surface-based Affinity Methods

Determining the binding affinities of non-covalent target-ligand interactions is important in understanding the therapeutic potencies of lead compounds. Identification of novel and potent ligand from the diverse chemical library is the heart in the early stage of drug discovery. The microarray technology was originated in 1991 when Fodor *et al*. introduced peptide microarray in their seminal work [18]. Significant development has been made over the past two decades, which makes microarray technology a robust and powerful tool in academia and industrial research. The applications of microarray technology cover virtually all types of biomolecules such as nucleic acids, proteins, small molecules, as well as live cells. Among all, protein microarray and small-molecule microarray have been implemented extensively in processes of HTS.

Protein microarray is commonly applied in assessing the functions and activities of proteins on a large scale [19]. Protein microarray (often referred as functional protein microarray) was originally developed for systematic studies of protein bioactivities. Essentially, functional proteins are immobilized onto the optical surface through generic linkers such as His-tag or GST-tag; followed by incubation with the fluorescently labeled ligands solution. Once reaching binding equilibrium, the bound ligands are retained while unbound ligands are partitioned out, and the fluorescent signals are recorded. Methods of protein microarray have been successfully applied in profiling various types of post-translational event, and also

assessing a variety of protein-protein interactions. Also, the latest high-content protein microarray allows simultaneous monitoring of bioactivities of thousands of proteins on a single chip, which significantly enhanced the performance of molecular profiling. Following the same concept, methods of protein microarray have also been applied as a promising HTS platform; numerous studies demonstrated the practicalities of this approach with binding affinities ranging from μM to nM.

Small-molecule microarray played the significant role in the process of drug discovery [11]. However, unlike macromolecules such as protein and DNA, small-molecule immobilization is facing the obstacles of limited structural diversities and hindrance from immobilization. Various fabrication techniques have been developed to circumvent these limitations. For example, the non-covalent approaches by using DNA-directed/encoded ligand immobilization, or using the strong interactions between avidin-coated chips and biotin-labeled small-molecules. Alternatively, the ligand can be immobilized through covalent linkage to achieve a firm and stable bond between the ligands and array surface. Finally, the small-molecule microarrays have also been fabricated via *in situ* chemical synthesis.

The chemical libraries typically contain a vast number of compounds, which range from being achiral to racemic by nature. Screening and selecting of these synthetic racemic compounds is extremely critical in identifying biological activity as well as toxicities. The frontal affinity chromatography based approach with the combination of mass spectrometry detection has been developed for analyzing racemic chemical compounds [1]. In essence, the protein targets are immobilized on solid phase inside of column, followed by a solution of ligands flowing through the column to interact with targets. Each individual ligand has a distinctive eluate volume, which is often referred as the "breakthrough volume" and characterized as a

11

sigmoidal front. Frontal affinity chromatography [20] has been successfully applied to screen mixtures of chemical compounds against a wide range of biological targets. Also, this approach has also been implemented for analyzing binding affinity, such as determining K_d as well as ranking the relative binding affinities of ligand mixtures.

1.3.2 Surface-based Kinetic Methods

Surface-based kinetic methods have been extensively used in studying various biomolecular interactions. Optical biosensors such as SPR are ordinarily used as solid support for analyzing non-covalent target-ligand interactions [21]. Briefly, sample molecules, either targets or ligands, are covalently immobilized onto the gold surface. Gold is the common material for SPR surface sensor as it is highly stable against oxidation and widely compatible with various chemical linkages. The surface attachments of sample molecules impose particular angle in refractive index, which is described as SPR angle. As the binding occurs, the change of refractive index can be monitored using monochromatic light and measured in real time. The final data can be used to calculate the amount of bound analyte, binding affinity as well as the association and dissociation kinetic rate constants.

BLI is a recently developed optical based method [22]. In essence, the sample molecules such as protein targets are immobilized onto the tip of an optic fibre; the molecular layer creates an interference pattern that can be identified. Next, the sample-attached optical fiber is submerged into the solution of ligands; binding of another layer of ligand molecules triggers the shift in the signatures of the interference pattern. Kinetic analysis of SPR and BLI are similar that the signals of association and dissociation are numerically fitted, and then k_{on} and k_{off} can be deconvoluted from the best fitting curves. The main advantages of surface-based techniques are

high sensitivities, little samples consumptions, and high-throughput. Also, these surface-based techniques are pure kinetic methods that rely only on numerical fitting for data processing; binding equilibrium is not required in assumptions and thus allowing the sample concentration of analytes surpass the range of binding affinity (K_d). Despite the renowned advantages, surface-based methods inherently suffer from non-specific surface adsorption, which typically leads to false positive results during primary screening. In addition, the typical experimental setups require chemical immobilization of sample molecules onto the sensing surface, which often disturbs the molecular structures of either targets or ligands. Such disturbance in 3D structures can significantly affect the binding properties of the target-ligand interactions, which is detrimental in determining kinetic binding parameters. Thus, surface-based methods are not ideal for secondary screening.

1.3.3 Solution-based Affinity Methods

In the pharmaceutical industry, one of the most widely used solution-based affinity methods is ITC, which is also referred as "gold standard" in determining target-ligand binding affinity. In essence, ITC measures the time-resolved enthalpy change of the given target-ligand interactions [23]. The quantification of thermodynamic parameters such as enthalpy and entropy, affinity constant K_d, and even binding stoichiometry can be all achieved within a single ITC experiment. Furthermore, ITC can also be used for studying enzymatic kinetics to measure K_{cat} and K_M [24]. Given its true label-free nature, ITC is considered as the method of choice for label-free affinity studies as the analyzed binding parameters reflect the native binding states. However, there are also drawbacks: first, for heat changing measurements, to meet the limit of

13

detection relatively large amount of sample is often required; second, ITC cannot be used for kinetic analysis of non-covalent target-ligand interactions.

Similarly, the conventional ESI-MS and nanoelectrospray ionization mass spectrometry (nESI-MS) have been widely adapted for studying non-covalent complexes [25, 26]. The simple ESI-MS and nESI-MS are considered as "snapshot" methods; essentially the ratio between the complexes and free ligands is recorded to calculate the equilibrium binding constant K_d. The main advantages of such approaches are: (i) a single binding experiment is sufficient to estimate binding affinity; (ii) the minimum sample consumptions. As a result, ESI-MS and nESI-MS have been adopted in studying various target-ligand interactions. However, it is still a debate over the legitimacy of these methods that whether the detection of complexes in gas phase actually reflects the binding ratio in solution phase [27], as the process of ionization can trigger the complex dissociation; also, the ionization efficiencies may be different between the free molecule and binding complex. Finally, both ESI-MS and nESI-MS are affinity methods that cannot be used for kinetic analysis. Conclusively, all conventional label-free solution-based methods can be only used for affinity analysis, but cannot be used for studying the kinetics of target-ligand interactions. However, the kinetic parameters control the drug efficacy and toxicity in the human body.

1.3.4 Solution-based Kinetic Methods

Fluorescence-based bioanalytical assays are renowned for high sensitivity and convenient detection setups. However, such assays are often hindered by intrinsic fluorescence of compounds, present in highly populated libraries. Methods, like fluorescence polarization (FP) and time-resolved Förster resonance energy transfer (TR-FRET), are less affected by intrinsic

fluorescence of ligands during the process of screening [1]. Briefly, FP utilizes the depolarization of emission light to calculate the rate of molecular rotation of fluorescent ligand; the speed of rotation changes upon binding to the target, which offers the basis of target-ligand binding interactions. FRET is also a powerful technique and typically used for studying structures and dynamics of macromolecules. When two matching fluorophores are positioned in close proximity, the excited donor fluorophore transfers energy to the acceptor fluorophore and triggers its emission. Using the same principal, TR-FRET combines a long lifetime donor and a short lifetime acceptor to achieve time-resolved analysis of binding interactions.

As previously discussed, labeling requires covalent modifications that can significantly affect the structures of targets or ligands as well as the binding properties. Therefore, labeled methods are not ideal for kinetic binding studies. Label-free methods, on the other hand, analyze the molecules without chemical modifications; thus the native states of interaction are maintained. Therefore, the solution-based kinetic methods are essential in both finding the target binding ligands and kinetic analysis of corresponding target-ligand interactions. Besides, for kinetic analysis, the label-free approaches are critical in determining the true kinetic parameters for target-ligand interactions.

Kinetic capillary electrophoresis (KCE) is a conceptual platform, which contains a panel of methods that allow separating species to interact during capillary electrophoresis (CE) [28]. The concentration profiles of the interacting species are plotted over time, in which can be used to calculate both equilibrium and kinetic parameters. The basis of CE relies on the separation of species based on the differential charge to size ratios. Thus, KCE-based methods have been substantially adopted for kinetic analysis of protein-DNA interactions as well as DNA-aptamer selections. The detection during KCE-based protein-DNA studies is typically achieved through

fluorescent labeling of DNA molecules. For non-covalent target-ligand interaction studies, we have introduced two different types of KCE methods for label-free solution-based kinetic analyses: (i) kinetic capillary electrophoresis mass spectrometry (KCE-MS) and (ii) plug-plug kinetic capillary electrophoresis with UV detection (ppKCE-UV).

1.3.4.1 Kinetic Capillary Electrophoresis with Mass Spectrometry Detection

KCE-MS is a label-free solution based kinetic approach for analyzing interactions between protein (P) and small molecule (SM). The non-equilibrium capillary electrophoresis of equilibrium mixture (NECEEM) [28] was the method of choice for solution-based kinetic separation. Briefly, the P and SM were incubated to form an equilibrium mixture. The equilibrium mixture contained three distinct species: P, SM, and protein-small molecule complex P-SM. The equilibrium mixture was the sample of injection and during electrophoresis all three species separated from each other according to different charge to size ratios. The charge to size ratios between P and SM usually differ from each other. Thus they migrated into two distinct zones. However the charge to size ratios between P and P-SM are typically similar, thus they co-migrated into a single zone. Also, during electrophoresis, SM separated from P-SM, which led to a disruption of equilibrium and triggered continues dissociation of P-SM. Such continues dissociation is typically featured as a "bridge" connecting the signals between SM and P-SM. Finally, the concentration profiles of SM were recorded over time and utilized in the subsequent kinetic calculations. One of the main advantages of NECEEM is that the signals from single specie are sufficient for calculating binding parameters. In this case, the signals from SM were recorded through the generic label-free mass spectrometry detection.

16

The procedure of KCE-MS [29] comprises two main parts (**Figure 1.1**): first part contains optimization and calibration (Step 1-3), and the second part provides measurements of binding parameters (Step 4-7). The first part, optimization and calibration, is a three-step task. Step 1, analyze and optimize the ESI-MS detection conditions of the pure small molecule ligand. Step 2, establish the serial dilution calibration curve of the ligand using the optimized detection conditions. Step 3, subject the protein target and small molecule ligand individually to determine the three fraction-collection windows (W1, W2, and W3) and complex migration time, $t_{\text{P·SM}}$ (used in k_{off} calculation) using CE with UV light absorption detection. The second part, measurements of the binding parameters, is a four-step procedure. Step 4, an aliquot of the pre-equilibrated protein-small molecule binding mixture is subjected for NECEEM. At the end of NECEEM, three fractions are collected according to pre-established fraction windows (W1, W2, and W3). As schematically illustrated in **Figure 1.1B**, W_1 contains free SM in equilibrium mixture; W_2 contains the SM that dissociated from the complex during NECEEM separation, and W_3 contains the SM that is still within intact binding complexes.

A) Calibration/optimization

Step 1. Identification of small molecule's main ion
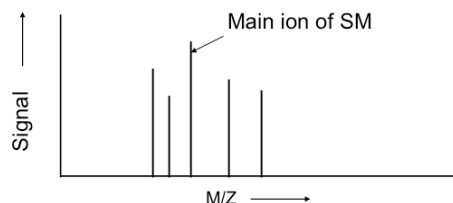
Step 2. Building MS calibration curve
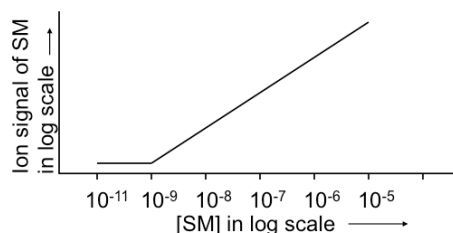
Step 3. Determination of 3 fraction-collection windows (W_1-W_3) and migration time of complex ($t_{P \cdot SM}$)
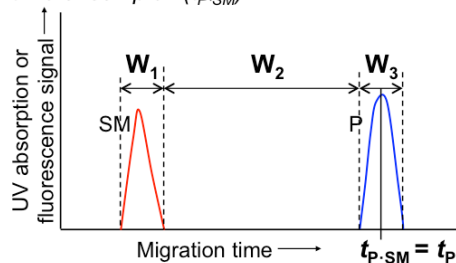
B) Measurements of K_d and k_{off}

Step4. Collection of 3 fractions in blind NECEEM
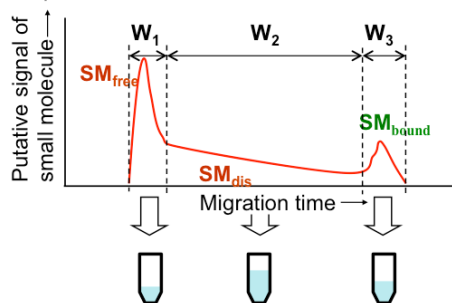
Step 5. Exchange of aqueous buffer for methanol

Step 6. MS measurement of SM concentrations

$[SM]_{free}$ $[SM]_{dis}$ $[SM]_{bound}$

Step 7. Calculation of K_d and k_{off}

$$K_d = F([P]_0, [SM]_0, [SM]_{free}, [SM]_{dis}, [SM]_{bound})$$

$$k_{off} = F([SM]_{dis}, [SM]_{bound}, t_{PgSM})$$

Figure 1.1. Schematic illustration of the practical realization of KCE-MS through off-line interfacing of NECEEM with ESI-TOF MS. Please see text for details.

However, NECEEM uses in a physiological buffer, which is suitable for biomolecular interactions in CE experiments but incompatible with ESI-MS detections. Therefore, each collected fraction is desalted and dissolved in equal volumes of methanol (Step 5). In Step 6, the small molecule ligand concentrations in each fraction are determined by ESI-MS using the previously built calibration curve. Finally, in Step 7, the three concentrations are used to calculate K_d and k_{off} with simple NECEEM formulas:

$$K_d = \frac{[P]_0 - [SM]_0 \left(1 - \dfrac{SM|_{free}}{SX|_{free} + SM|_{dis} + SM|_{bound}}\right)}{\dfrac{SM|_{free} + SM|_{dis} + SX|_{bound}}{SM|_{free}} - 1} \qquad (1\text{-}2)$$

$$k_{off} = \ln\left(\frac{[SM]_{dis} + [SM]_{bound}}{[SM]_{bound}}\right)/t_{P\text{-}SM} \tag{1-3}$$

where $[SM]_{free}$, $[SM]_{dis}$, and $[SM]_{bound}$ indicate the concentrations of free, dissociated, and

bound SM in each fraction.

1.3.4.2 Plug-plug Kinetic Capillary Electrophoresis with UV Detection

KCE-UV is another label-free solution-based kinetic method that previously developed for

analyzing DNA-small molecule interactions [30]. Besides, UV absorption is a generic,

accessible, and practical means of detection: most molecules absorb light in UV, and all

commercial CE instruments are equipped with UV absorption detection. The KCE method of

choice is plug-plug kinetic capillary electrophoresis (ppKCE), which facilitates direct

measurements of kinetic binding parameters k_{on} and k_{off} by first mixing and reacting molecules

and then separating and dissociating the formed complexes. The concept of ppKCE-UV is

schematically presented in **Figure 1.2**. At time zero, small plugs of A and B with concentrations

of [A] and [B] and length of l_A and l_B, respectively, are sequentially injected one after another by

a low-pressure pulse; the components with a lower mobility (*e.g.* A) is injected first. For

examples, the slow moving DNA is A, which is injected first, and a small molecule B is injected

later. Under the presence of a high electric field, both A and B are moved towards the outlet with

different electrophoretic mobilities. When passing through each other, A and B form certain

amount of interaction complex, C. The amount of formed complex is dependent on [A], [B], k_{on},

and time of passage. When the zones of A or B is separated from that of C, C starts dissociating

(the yellow curve illustrates the fitting of dissociation) with a rate depending on k_{off}. A temporal

propagation pattern of one component, for example, B is recorded by a UV detector placed at a

distance of L from the capillary inlet. The values of k_{on} and k_{off} can be determined by using the

signal intensities of B and corresponding migration times from a ppKCE-UV electropherogram:

$$k_{off} = \log\left(\frac{I_{t_2}}{I_{t_1}}\right)\left(\frac{t_1 - t_C}{t_C(t_2 - t_1)}\right) \qquad (1\text{-}4)$$

$$k_{on} = \varepsilon L\left(1/t_B - 1/t_A\right)/([A]l_A) \qquad (1\text{-}5)$$

Here I_{t_1} and I_{t_2} are signal intensities of B at times t_1 and t_2, respectively, in the exponential region

of the trace (shown by the yellow line in Figure 1.2, bottom curve). The migration times of A, B,

and C are represented by t_A, t_B, and t_C, where $t_A = t_C$ since the complex co-migrates with the

DNA. Parameter ε is determined by solving the following non-explicit equation:

$$B_{fr}/B_{tot} = \ln\left\{(exp(\varepsilon) - 1)exp\left(-\varepsilon\left(\frac{[B]l_B}{[A]l_A}\right) + 1\right)\right\}/\varepsilon \qquad (1\text{-}6)$$

In the above equations, B_{fr} is the area of the peak corresponding to unbound ligand. The B_{tot}

represents the integration of total amount of ligand, which is usually measured in a separate

experiment with ligand injection only. l_A and l_B are the injection plug lengths of A and B. The

other parameters have been defined above. The kinetic rate constant of complex dissociation, k_{off},

can be determined by fitting the experimental decay curve.
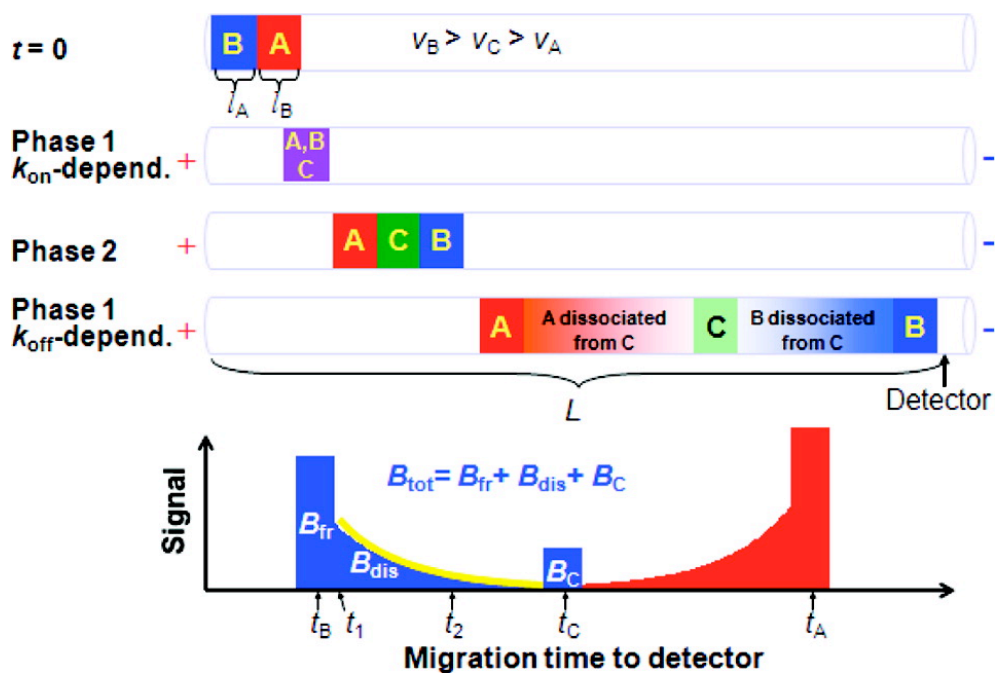
Figure 1.2. Schematic illustration of the ppKCE method. Please see text for details.

Although proven to be robust separation tools, KCE methods alone are not ideal for analyzing protein-small molecule interactions. The separation power of KCE methods depends on the differential charge to size ratios. However, neither protein nor small molecule is highly charged. Methods of KCE, in this case, are not generic separation tools meaning that optimizations are always required for each and every binding pair. Therefore, novel ligand screening methods are still in demand to achieve the solution-based high throughput primary screening; as well as the label-free solution-based kinetic methods for secondary screening.

In this dissertation, we are going to introduce a new approach that combines KCE with DNA-encoded ligand technology to build a generic solution based HTS platform for efficient and accurate small molecule ligand selections (primary screening); we call it kinetic capillary electrophoresis facilitated DNA-encoded ligand selection (KCE-DEL). Furthermore, we are also

developing a kinetic size-exclusion chromatography-mass spectrometry (KSEC-MS) approach for label-free solution-based kinetic analysis of target-ligand interactions (secondary screening). For our methods to be generic and reliable, the following requirements have to be fulfilled. (i) The primary screening methods should be efficient and accurate. (ii) The primary screening methods should also be applicable to wide range of samples. (iii) For secondary screening, the methods of choice should be compatible with generic label-free detection. (iv) For secondary screening, the developed methods should be accurate in assessing the true kinetic parameters. (v) For secondary screening, the developed methods should be robust and applicable for various samples with different conditions. In the following chapters, we will tackle above-mentioned requirements step by step and present KCE-DEL and KSEC-MS as novel kinetic solution-based approaches in facilitating small molecule drug discovery.

CHAPTER 2. PREDICTION OF PROTEIN–DNA COMPLEX MOBILITY IN GEL-FREE CAPILLARY ELECTROPHORESIS

My contribution to the article was: (i) planning all experiments, (ii) designing and synthesising all ligands (iii) performing all experiments, (iv) preparing figures, (v) writing the manuscript.

2.1 Introduction

Selection of protein binders from highly diverse combinatorial libraries (complex mixtures) of molecules is an efficient and economical alternative to traditional screening for discovery of affinity probes and drug leads [31]. The molecules in the most diverse libraries, with only ~ 1–100 copies of every molecule present in a sample, include either DNA or RNA for the purpose of binder identification. The unique property of DNA is that it can be amplified by PCR and sequenced to reveal the binder's identity. RNA, on the other hand, can be easily converted into DNA, which can then be amplified and sequenced. The examples of such libraries are (i) random DNA (or RNA) libraries used for selection of oligonucleotide aptamers [32, 33], (ii) mRNA-display libraries containing chimeras of mRNAs with peptides that they encode and used to select protein-binding peptides [34], and (iii) DNA-encoded libraries of small molecules

used for selection of small-molecule protein binders [35]. For any specific library, the oligonucleotides have identical lengths and are the bulkiest parts of the molecules. They largely define the physical properties of the library molecules such as size and electrical charge, so that other parts, even when present, can be neglected if these physical properties are of major importance. Therefore, for a general consideration of the physical properties of molecules, we can assume that the protein binds DNA and we will use this simplification unless the details are essential.

In the binder selection procedure, the library is mixed with the protein target to allow library molecules to bind the target. The target-bound molecules are partitioned from the target-unbound ones. The collected bound molecules are dissociated from the protein and identified by sequencing their DNA (or DNA complement of RNA). The partitioning step must be very efficient to ensure that the binders are not lost while the nonbinders are removed. Typically, partitioning is done by using surface-based approaches: separation on filters that retain the protein but let DNA (RNA) through or affinity chromatography with the protein immobilized on the stationary phase and retaining the binders [36].

Surface-based techniques suffer from low partitioning efficiency caused by nonspecific binding of the library molecules to the surface of the filter or the stationary phase. The fraction of the library that nonspecifically binds to the surface can be as high as 15% [37]. Such a high background decreases the efficiency of the selection procedure. It is especially detrimental for selection of binders from DNA-encoded libraries of small molecules. Unlike random DNA libraries and mRNA display libraries, the libraries of DNA-encoded small molecules cannot be propagated because small molecules are not amplifiable. Therefore, enrichment of true binders must be achieved within a few rounds of selection, which, in turn, requires high partitioning

24

efficiency of separation methods used in the selection. Failure to successfully select protein binders from the three types of libraries considered here can be caused by low partitioning efficiency of the surface-based separation methods used [38].

Gel-free capillary electrophoresis (CE) is a solution-based separation technique and a highly promising alternative to surface-based techniques for partitioning protein–DNA complexes from the unbound DNA library. The separation in CE is based on different electrophoretic mobilities of DNA and protein–DNA complexes; the protein–DNA complex always has a greater friction coefficient (of the drag force) and a lower negative charge density than unbound DNA. Moreover, all DNA molecules of the same length have similar mobilities and migrate as a single electrophoretic zone. All complexes of the same-length DNA with the same protein also have similar mobilities and migrate as a single electrophoretic zone. When a bare fused silica capillary is used along with a pH-neutral separation buffer, there is always an appreciable electroosmotic flow (EOF) from the positive-electrode end to the negative-electrode end of the capillary. The absolute value of EOF mobility is greater than those of DNA and protein–DNA complexes while the direction is opposite. As a result, DNA and protein–DNA complexes injected at the positive-electrode will move toward the negative-electrode end (despite their overall negative charges) with the complexes moving faster (**Figure 2.1A**). The complexes can be collected at the capillary outlet before the unbound DNA reaches the end. The greater time window between the complexes and unbound DNA will result in a greater partitioning efficiency (**Figure 2.1B**). The background can originate from unbound DNA moving along with the protein–DNA complex [39]. A wide time window between the zones of the complex and unbound DNA guarantees very low background, which must be much lower than that of surface-based methods [40].
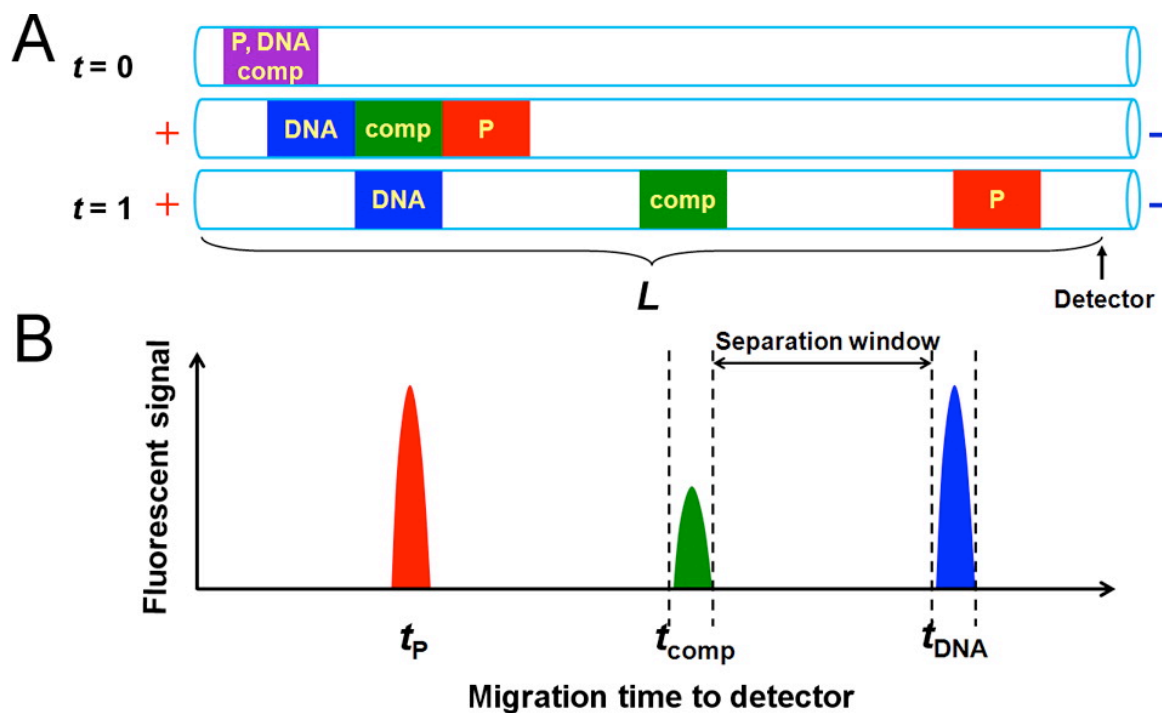
Figure 2.1. Conceptual depiction of migration patterns of DNA, protein, and complex. (A) The sample that contains DNA, protein, and complex is injected into the capillary at $t = 0$. Under high voltage, all three components start to migrate toward the outlet yet separate from each other based on their size to charge ratios. The complete separation is achieved at $t = 1$. (B) The graph illustrates the corresponding migration times of DNA, protein, and complex.

The above advantages of gel-free CE led to its practical use for analytical and preparative separation of protein–DNA complexes. Methods of kinetic capillary electrophoresis (KCE) were successfully utilized for measuring rate constants of complex formation, k_{on}, and dissociation, k_{off}, and equilibrium dissociation constant, K_d [29, 30, 41-43]. KCE methods were also used for selection of protein binders from DNA libraries [44]. In particular, DNA aptamers were selected for a number of proteins [45]. Uniquely, KCE methods allowed selection of aptamers with desirable ranges of k_{off} and K_d values [46]. The library enrichment is typically completed in 1–4 rounds of partitioning in contrast to 10–20 rounds usually required with surface based methods [47]. Such high speed of enrichment is explained by an extremely low level of background of

<0.01% [48]. The use of KCE methods was also suggested for selection of protein binders from libraries of DNA-encoded small molecules, and some performance parameters have been experimentally evaluated for this application [43].

Selection of protein binders from DNA libraries requires collection of a fraction of the intact protein–DNA complex (and/or free DNA that originated from the dissociation of protein–DNA complex) during electrophoresis. Accurate fraction collection requires the knowledge of migration time of protein–DNA complexes. In some instances, adding a great excess of protein to the library leads to creation of nonspecific protein–DNA complexes that can be detectable [49]. However, this approach does not work when the protein does not have a tendency of binding DNA nonspecifically. Blind fraction collection has high odds that either the complex will not be collected or a large amount of "background" DNA will be collected along with the complex. The latter is an indicator of inefficient partitioning that can be detrimental for selection, especially from nonamplifiable libraries of DNA-encoded small molecules. Therefore, it is of great importance for KCE-based selection of protein binders to have a method of accurate prediction of protein–DNA complex mobility. Here we present such a method for complexes of proteins with DNA-encoded small molecules. In this case we use a model of a globular protein with a rigid dsDNA attached to the protein in a single point (**Figure 2.2**).
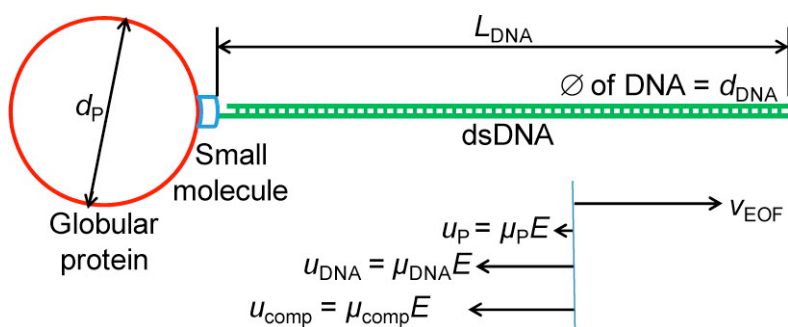
Figure 2.2. Schematic representation of the complex of globular protein and rod-like dsDNA, linked through a small molecule, capable of binding the protein. This model mimics the complex between a protein and a DNA-encoded small molecule with the dsDNA part. The lower part illustrates relative values of velocities of EOF, protein, DNA, and protein–DNA complex.

The model is based on a theory of the thin double layer and corresponding expressions used for the mobilities of a rod-like short oligonucleotide and a sphere-like globular protein. It uses empirical data for mobilities of free DNA and free protein, which can be easily determined experimentally. To test the developed mathematical model, we used binding of streptavidin to biotin-labeled dsDNA of different lengths. The results show that the model can predict the mobility of protein–DNA complex with an error of less than 4% and the travel time of protein–DNA complex to the detector with error less than 6%. It can thus aid selection of protein binders from libraries of DNA-encoded small molecules and advance the use of such libraries in identifying drug leads and diagnostic probes.

2.2 Materials and Methods

2.2.1 Chemicals and Materials

Fused-silica capillary was purchased from Polymicro (Phoenix, AZ). All reagents were dissolved in 50 mM Tris-HCl, pH 8.0. The 40, 80, and 120 dsDNA were synthesized by PCR, and a pETMutS plasmid was used as a template (Addgene plasmid 13245, Cambridge, MA). All DNA primers were purchased from IDT DNA Technology Inc. (Coravile, IA). DNA primer sequences were

Forward primer: FAM 5′-CCGACTACCTCCTCCTCTTC-3′

Reverse primer 40:   Biotin 5′-TCGTAGAAGTCCCCCACCTG-3′

Reverse primer 80:   Biotin 5′-CAGGGCGCGGGCCA-3′

Reverse primer 120: Biotin 5′-TGGTGAAGTCCTTGCTGGTC-3′

All PCR products were subjected onto a 2% agarose gel, and the bands that contained dsDNA were excised and purified by using a QIAquick Gel Extraction Kit (Toronto, ON, Canada). The purified dsDNA were quantified by using fluorescence detection at 520 nm. The FAM-labeled forward primer was used as a concentration standard. The streptavidin (Sigma-Aldrich, Oakville, ON, Canada) was labeled with a fluorogenic dye, Chromeo 488 (Active Motif, Carlsbad, CA) overnight at 4 °C. Bodipy was purchased from Life Technologies Inc. (Burlington, ON, Canada). All other regents were purchased from Sigma-Aldrich (Oakville, ON, Canada). All solutions were made using deionized water filtered through a 0.22 μm filter (Millipore, Nepan, ON, Canada).

2.2.2 Instrumentation

The ABI7300 real time PCR (BioRad, Mississauga, ON, Canada) was used to synthesis dsDNA. The Owl D2 Wide-Gel Electrophoresis System (Thermo Scientific, Wilmington, DE) was used to purify the PCR products. NanoDrop 3300 fluorospectrometer (Thermo Scientific, Wilmington, DE) was used for dsDNA quantification. All CE experiments were carried out with MDQ-PACE instrument (Beckman-Coulter, ON, Canada) equipped with a laser-induced fluorescence (LIF) detector. LIF signal was recorded at 520 nm (for fluorescein, FAM, and bodipy detection) and 605 nm (for chromeo-streptavidin detection) with 4 Hz acquisition rate. The inner diameter of the capillary was 75 μm. The total capillary length was 81.2 cm with 71.2 cm from the injection end to the detection window.

2.2.3 Migration Analysis by CE–LIF

The 50 mM Tris-HCl, pH 8.0 buffer was used for both incubation and separation. The binding mixture was made by incubating 100 nM dsDNA, 1 μM chromeo-labeled streptavidin, 10 nM fluorescein (internal standard), and 5 μM bodipy (neutral marker), at 20 °C for 30 min. The control mixture was the same as binding mixture yet without streptavidin. The capillary was flushed prior to each CE run with 0.1 M HCl, 0.1 M NaOH, ddH$_2$O, and buffer. The sample was injected into the capillary at 0.5 psi for 10 s. The ends of the capillary were inserted into the inlet and outlet reservoirs, and an electric field of 308 V/cm was applied to carry out electrophoresis. The temperature of the capillary was maintained at 15 °C. All experiments were performed in triplicates.

2.3 Results and Discussion

2.3.1 Mathematical Model

In this work, we concentrate on libraries of DNA-encoded small molecules in which the DNA part is dsDNA of ~120 base pairs in length. This case describes a class of practical libraries of DNA encoded small molecules used in the pharmaceutical selection of drug leads [50].

Proteins have been used as tags to cause DNA mobility shift in DNA sequencing [51]. The general separation approach dealing with such molecular chimeras is termed End-Labeled Free-Solution Electrophoresis (ELFSE) of DNA. To aid processing data from ELFSE-of-DNA experiments, theoretical models of ELFSE have been developed [52-58]. Such models usually employ the blob theory that is applicable to DNA, which is sufficiently long to be considered a semiflexible random coil [56, 58]. The polymer can be considered a semiflexible random coil if its contour length L is much larger than the Kuhn length b_K characterizing the polymer stiffness [59, 60]. This assumption is not satisfied for ~120 base pairs long dsDNA for which $L_{DNA} < 41$ nm while $b_{K,DNA} > 100$ nm. Here and below, "DNA" in the subscript indicates that the corresponding parameter describes dsDNA. Thus, the usual ELFSE models, which are based on the blob theory, cannot be used in our case.

Taking into account that L_{DNA} is smaller than $b_{K,DNA}$, we use a different approach assuming that dsDNA (containing ≤120 base pairs) behaves like a rigid rod. The dsDNA diameter, d_{DNA}, can be estimated as 2 nm [61,62]. which is larger than the Debye length for the buffer, λ_D, and the dsDNA length L_{DNA} is many times larger than λ_D. Thus, we can assume that the dsDNA mobility, μ_{DNA}, is estimated by an expression used in a theory of the thin double layer [52,56]:

$$\mu_{DNA} = \frac{\varepsilon_0 \varepsilon_r \zeta_{DNA}}{\eta} \approx \frac{-\sigma_{DNA} \lambda_D}{\eta} \quad \left(\zeta_{DNA} \approx \frac{-\sigma_{DNA} \lambda_D}{\varepsilon_0 \varepsilon_r} \right) \tag{2-1}$$

Here, ε_0 is the vacuum permittivity, ε_r is the relative permittivity of the buffer, ζ_{DNA} is the zeta potential of dsDNA, σ_{DNA} is the surface density of the electric charge in the diffuse part of the double layer around dsDNA (i.e., excluding the Stern layer), and η is the dynamic viscosity of the buffer. Equation 2-1 can be rewritten as follows:

$$\mu_{DNA} = \frac{-\sigma_{DNA} \lambda_D}{\eta} = \frac{q_{DNA} \lambda_D}{\pi \eta d_{DNA}}, \quad \sigma_{DNA} = -\frac{q_{DNA}}{\pi d_{DNA}} \tag{2-2}$$

where q_{DNA} is the charge per unit length of dsDNA. In calculations of q_{DNA}, we should take into account the condensation of the counterions on dsDNA [63,67]. The condensation takes place for cylindrical objects with the linear density electric charge, q, satisfying relations [63]:

$$|q| \geq q_{eff}, \quad q_{eff} = \frac{e}{z_i \lambda_B}, \quad \lambda_B = \frac{e^2}{4\pi \varepsilon_0 \varepsilon_r k_B T} \tag{2-3}$$

Here, e is the charge of proton, z_i is the valence of counterions, λ_B is the Bjerrum length, k_B is the Boltzmann constant, and T is the absolute temperature of the buffer. Usually, dsDNA has two negative charges per 0.34 nm of its length [56] and $\lambda_B = 0.7$ nm for water solutions at room temperature [55, 65]. Thus, equation 2-3 is always satisfied for dsDNA and condensation of counterions reduces the density of the DNA charge q_{DNA} (excluding the Stern layer) to the effective value, $-q_{eff}$, determined by the second equation 2-3 [63]. Since we also consider the Stern layer as a part of the condensed counterion layer, then $|q_{DNA}|$ will be even less than q_{eff}. In this case q_{DNA} can be considered as an adjustable parameter. We should note that the dsDNA mobility has negative values since dsDNA is negatively charged. Equation 2-2 for μ_{DNA} can be also obtained from the balance of electric and hydrodynamic forces, $F_{E,DNA}$ and $F_{H,DNA}$, acting upon dsDNA:

$$F_{E,DNA} + F_{H,DNA} = 0 \tag{2-4}$$

if we assume the following effective values for these forces:

$$F_{E,DNA} = q_{DNA} L_{DNA} E, \quad F_{H,DNA} = -\frac{\pi \eta d_{DNA} L_{DNA}}{\lambda_D} u_{DNA} \tag{2-5}$$

Here, E is the electric field strength and u_{DNA} is a relative velocity of dsDNA with respect to the buffer. Hereafter we use a coordinate system in which electric and hydrodynamic forces have only x-components.

The average diameter d_P of a globular protein with the molecular weight >10 kDa can be estimated as 3 nm [68]. Thus, d_P is significantly larger than λ_D. In this case, the protein mobility μ_P can be determined by expression similar to equation 2-1:

$$\mu_P = \frac{\varepsilon_0 \varepsilon_r \zeta_P}{\eta} \approx \frac{-\sigma_P \lambda_D}{\eta} \quad \left(\zeta_P \approx \frac{-\sigma_P \lambda_D}{\varepsilon_0 \varepsilon_r} \right) \tag{2-6}$$

Here, ζ_P is the zeta potential of the globular protein, and σ_P is the average surface density of the electric charge in the diffuse part of the double layer around the protein (i.e., excluding the Stern layer). Equation 2-6 can be also rewritten as follows:

$$\mu_P = \frac{-\sigma_P \lambda_D}{\eta} = \frac{Q_P \lambda_D}{\pi \eta d_P^2}, \quad \sigma_P = -\frac{Q_P}{\pi d_P^2} \tag{2-7}$$

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

Equation 2-7 for μ_P can also be obtained from the balance of electric and hydrodynamic forces, $F_{E,P}$ and $F_{H,P}$, acting upon the protein molecule:

$$F_{E,P} + F_{H,P} = 0 \tag{2-8}$$

if we assume the following effective values for these forces:

$$F_{E,P} = Q_P E, \quad F_{H,P} = -\frac{\pi\eta d_P^2}{\lambda_D} u_P \tag{2-9}$$

Here, u_P is the relative velocity of the protein with respect to the buffer.

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

$$F_{E,DNA} + F_{E,P} + F_{H,DNA} + F_{H,P} = 0 \tag{2-10}$$

Substitution of equations 2-5 and 2-9 into equation 2-10 gives

$$\left(q_{DNA} L_{DNA} + Q_P\right) E = \left(\frac{\pi\eta d_{DNA} L_{DNA}}{\lambda_D} + \frac{\pi\eta d_P^2}{\lambda_D}\right) u_{comp} \tag{2-11}$$

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

$$\mu_{comp} = \frac{q_{DNA} L_{DNA} + Q_P}{\dfrac{\pi\eta d_{DNA} L_{DNA}}{\lambda_D} + \dfrac{\pi\eta d_P^2}{\lambda_D}} \tag{2-12}$$

Taking into account equations 2-2 and 2-7 for the mobilities of dsDNA and the globular protein, we can rewrite the equation 2-12 as follows:

$$\mu_{comp} = \frac{d_{DNA} L_{DNA} \mu_{DNA} + d_P^2 \mu_P}{d_{DNA} L_{DNA} + d_P^2} \tag{2-13}$$

Using equation 2-13 for the complex mobility, we can readily find the complex travel time to the detector, t_{comp},

$$t_{comp} = \frac{L}{v_{EOF} + \mu_{comp} E} \tag{2-14}$$

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

It should be noted that the final equation 2-13 for complex mobility does not contain charges of dsDNA and protein. We excluded them using equations 2-2 and 2-6 for the mobilities of dsDNA and protein. Thus, we do not need to know the charges of dsDNA and protein to calculate complex mobility since we can experimentally determine the mobilities of dsDNA and protein. In this case, the charges of dsDNA and protein can be back calculated from equations 2-2 and 2-6 using their experimentally found mobilities and, therefore, can be considered as adjustable parameters.

2.3.2 Experimental Validation of Mathematical Model

To validate our mathematical model expressed by equations 2-13 and 2-14, we needed a protein that binds dsDNA at its end and we needed to determine mobilities of free protein and free dsDNA as well as the EOF velocity. We chose streptavidin and biotinylated dsDNA as a binding pair. Streptavidin can bind to biotin with exceptionally high affinity. Three lengths of dsDNA were used (N_{DNA} = 40, 80, and 120 base pairs) to test theory applicability for different DNA lengths. All experiments were performed in triplicates. **Figure 2.3** shows representative electropherograms for the neutral marker (bodipy), free protein, internal standard (fluorescein), free biotinylated dsDNA, and protein–dsDNA mixture. The velocity of EOF was measured and found to be v_{EOF} = 0.1247 ± 0.0002, 0.1249 ± 0.0002, and 0.1193 ± 0.0017 cm/s for experiments with N_{DNA} = 40, 80, and 120 base pairs, respectively. Mobilities of both dsDNA (in the absence of the protein) and the protein (in the absence of dsDNA) were found to be negative, which means that both dsDNA and protein are negatively charged. As a result the complex turned out to be negatively charged and its experimentally measured mobility is negative. Measurements of the dsDNA mobility resulted in the following absolute mobility values: $|\mu_{DNA}|$ = 0.2678 ±

0.0005, 0.2747 ± 0.0014, and 0.2784 ± 0.0002 cm^2/kV s for N_{DNA} = 40, 80, and 120 base pairs,

respectively. Measurements of the protein mobility revealed $|\mu_p|$ = 0.0401 ± 0.0006, $0.0403 \pm$

0.0012, and 0.0384 ± 0.0006 cm^2/kV s for experiments with N_{DNA} = 40, 80, and 120 base pairs,

respectively. Thus, the absolute value of protein mobility is significantly less than that of

dsDNA. The mobilities of the complexes were found to be $|\mu_{comp}|$ = 0.1643 ± 0.0006, $0.2007 \pm$

0.0022, and 0.2195 ± 0.0006 cm^2/kV s for N_{DNA} = 40, 80, and 120 base pairs, respectively. Thus,

a complete set of experimental data required for model validation needs to be obtained.
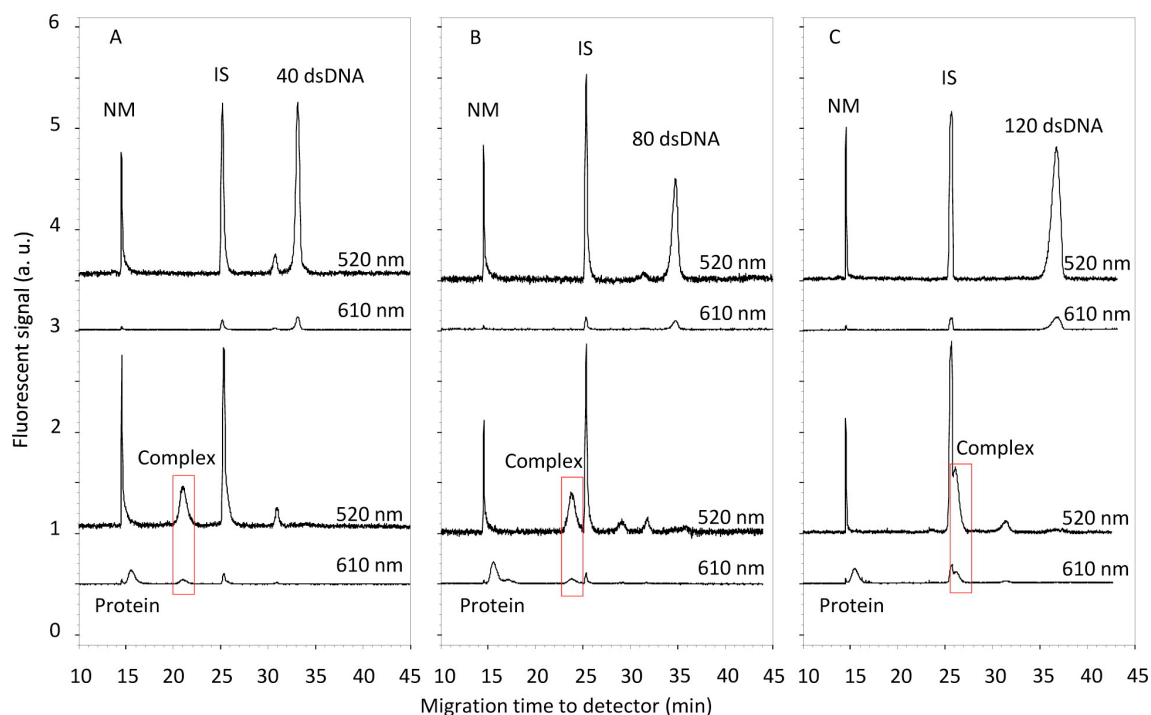
Figure 2.3. Migration information on all components. The migration patterns of 40, 80, and 120 dsDNA are shown in parts A, B, and C, respectively. In each panel, the top two traces represent a control experiment with different detection wavelengths. The control contains 100 nM dsDNA, the neutral marker (NM), and the internal standard (IS). The bottom two traces represent binding, which has the same composition as control plus 1 μM chromeo-streptavidin protein. The binding complex was highlighted with the red box. All experiments were performed in triplicates.

In addition to the described experimental values of mobilities and velocities, we needed to know the hydrodynamic sizes of the streptavidin (d_P) and dsDNA (d_{DNA} and L_{DNA}). We used a value of $d_{DNA} = 2.6$ nm, which includes the hydration shell around dsDNA [69], and a value for the streptavidine molecule diameter, $d_p = 5.3$ nm, determined from crystallographic studies [70]. The dsDNA length was calculated as $L_{DNA} = b_{DNA}N_{DNA}$, where $b_{DNA} = 0.34$ nm is the dsDNA monomer length [56].

We used the described parameters in eqs 2-13 and 2-14 to calculate predicated mobilities and travel times to the detector for protein–dsDNA complexes at different lengths of dsDNA. **Figure 2.4** shows absolute values of the experimental and theoretical mobilities of the protein–DNA complexes. According to the results in **Figure 2.4**, the developed model can predict the mobility of the protein–DNA complex with an error of less than 4% and the travel time of the protein–DNA complex to the detector with error less than 6%. It should be noted that different models were proposed for the mobility of rigid composite objects formed by a rod and a sphere [55, 71]. In these models, only one part (the rod or the sphere) is charged whereas in our case both parts (the rod-like dsDNA and the globular protein) can be charged.
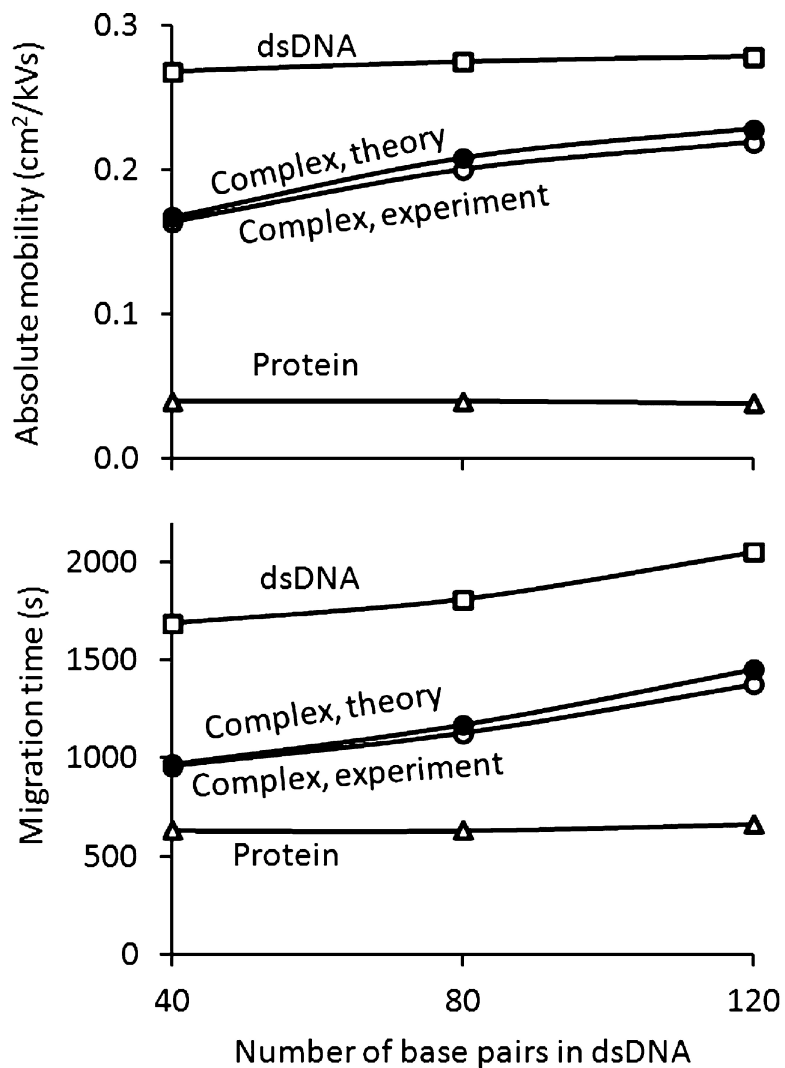
Figure 2.4. Dependences of the protein-dsDNA complex mobility (top) and travel time to the detector (bottom) on the number of base pairs in dsDNA.Mobilities and migration times of dsDNA and protein are also shown. Experimental results are shown by lines with open markers while theoretical results are shown by lines with solid markers. Theoretical values were obtained from equations 2-13 and 2-14.

Extension of these models to our case results in the following equation for the complex mobility

$$\mu_{comp} = \frac{\xi_{DNA}\mu_{DNA} + \xi_P\mu_P}{\xi_{DNA} + \xi_P} \qquad (2\text{-}15)$$

Here ξ_{DNA} and ξ_P are the friction coefficients of a rod and a sphere defined by relations [55, 72]:

$$\xi_{DNA} = \frac{2\pi\eta L_{DNA}}{\ln\dfrac{2L_{DNA}}{d_{DNA}} - \dfrac{1}{2}}, \quad \xi_{P} = 3\pi\eta d_{P} \qquad\qquad\qquad (2\text{-}16)$$

Equation 2-15 differs from equation 2-13 that we obtained and is used to predict the complex mobility. In particular, equation 2-15 depends on d_{DNA} only logarithmically (very weakly). Calculation based on eqs 2-15 and 2-16 give $|\mu_{comp}| = 0.1496$, 0.1748, and 0.1909 cm^2/kV s for $N_{DNA} = 40$, 80, and 120 base pairs, respectively. Comparing these theoretical values of complex mobility to ones experimentally determined above we can conclude that equation 2-15 results in 13% relative error in prediction of complex mobility whereas our equation 2-13 leads to only 4% error.

2.4. Conclusions

To summarize, we developed an approach for accurate estimation of the mobility of protein–dsDNA complex. The approach uses an approximation of a globular protein and a rod-like dsDNA. It will aid in selection and characterization of protein binders from libraries of DNA-encoded small molecules by methods of KCE. The general approach developed here can be utilized to develop similar models for other types of DNA libraries. In this proof of principle study, our model successfully predicted the migration of complex containing pure dsDNA ligand. However, in real selections, most DNA-encoded small molecule libraries comprise the DEL with both ssDNA and dsDNA regions. Therefore in the next chapter, we will introduce a new method for mobility calculations of protein-DNA complexes that contain both ssDNA and dsDNA regions.

CHAPTER 3. PREDICTING ELECTROPHORETIC MOBILITY OF PROTEIN-LIGAND COMPLEXES FOR LIGANDS FROM DNA-ENCODED LIBRARES OF SMALL MOLECULES

The presented material was published previously and reprinted with permission from "Bao, J., Krylova, S. M., Cherney, L. T., Hale, R. L., Belyanskaya, S. L., Chiu, C. H., Shaginian, A., Arico-Muendel, C. C., & Krylov, S. N. (2016). Predicting Electrophoretic mobility of Protein–Ligand complexes for ligands from DNA-Encoded libraries of small molecules. *Analytical Chemistry*, *88*(10), 5498–5506. doi:10.1021/acs.analchem.6b00980" Copyright 2016 American Chemical Society. My contribution to the article was: (i) planning all experiments, (ii) designing and synthesising all ligands except for two ligands provided by GlaxoSmithKline (iii) performing all experiments, (iv) interpreting the results, (v) preparing figures, (vi) writing the manuscript.

3.1 Introduction

In the previous study, we have successfully developed the KCE based ligand screening method, which can be potentially used for efficient and accurate primary drug screening. However the former model was built to predict protein-DNA complex migration with pure dsDNA; in real drug selections, most of the DNA-encoded ligands consists both dsDNA and ssDNA regions. Therefore, in this study, we are introducing a more generic model, which is applicable in real DEL selection.

Finding molecules that can selectively bind therapeutic targets is the initial step in most mainstream approaches of modern drug development [73-75]. Selection of protein binders (ligands) from DNA-encoded libraries of small molecules (DELSMs) is one such approach [76, 77]. DELSMs provide a solution for the main dilemma of selection of ligands from highly diverse mixtures of molecules. On one hand, the probability of finding ligands increases with increasing diversity of the mixture. On the other hand, the increasing diversity decreases the number of copies of unique molecules in the mixture, making their identification impossible by classical structure-analysis methods. In DELSMs, the structure of every small molecule is encoded in its DNA tag and can thus be revealed by amplifying and sequencing the tag. The efficiencies of polymerase chain reaction (PCR) and DNA sequencing are so high [78] that selecting a few copies of each ligand from a DELSM is sufficient for identification of its structures. As a result, DELSMs with diversities of more than 1 billion structures are synthesized and used for drug-lead selection [79].

The concept of DELSM was introduced in 1992 [76], and since then a number of synthetic approaches to the generation of DELSMs have been developed (**Figure 3.1**) [35, 80]. Different synthetic approaches lead to different structures of DNA tags. In general, DNA tags are linear DNA of two types: pure double-stranded DNA (dsDNA) and ds-ssDNA chimeras composed of dsDNA and single-stranded DNA (ssDNA) fragments.
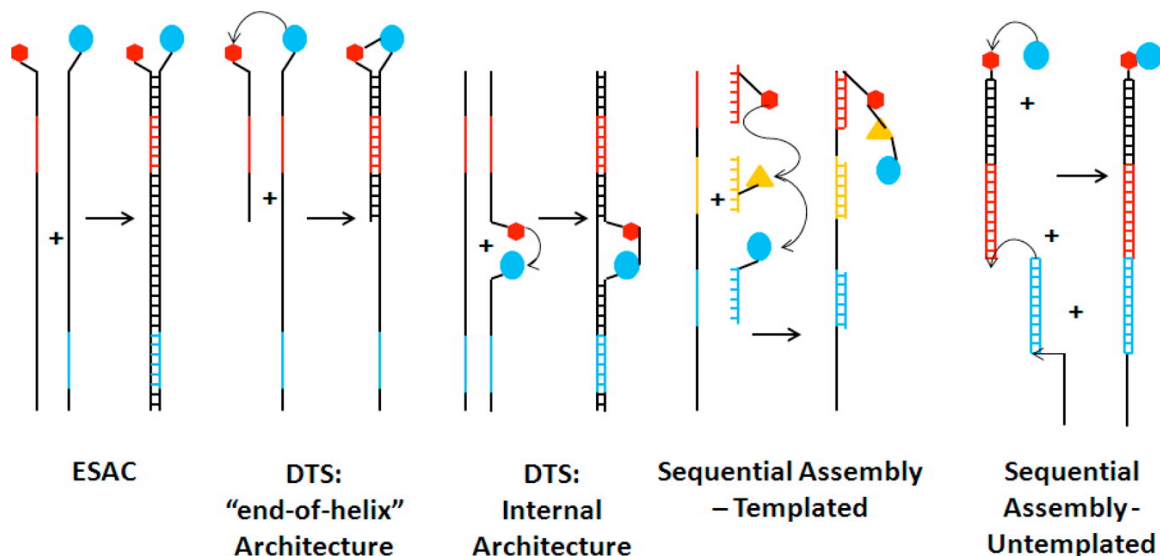
Figure 3.1 Schematic representation of assembly routes and corresponding structures of various DELSMs. (left to right) encoded self-assembling chemical (ESAC) library, DNA-templated synthesis (DTS) "end-of-helix" architecture, DTS internal architecture, sequential assembly templated, and sequential assembly untemplated. Building blocks of the small-molecule head and DNA fragments encoding them are shown by the same colour.

In chapter 2, we have demonstrated that methods of kinetic capillary electrophoresis (KCE) may facilitate highly efficient homogeneous selection of ligands from DELSMs with pure dsDNA structures. Here we present a more generic model that can be equally applied to DELSMs with more sophisticated structures of DNA tags. The new model considers a globular protein attached to the DNA tag at a single point. The thin double layer model is used to find mobilities of protein, dsDNA, and ssDNA. Given these mobilities, effective electric and hydrodynamic forces acting upon protein, dsDNA, and ssDNA are determined. Then the mobility of protein–dsDNA–ssDNA complex is obtained from the equation of balance of all forces acting upon the complex. Finally, complex mobility is expressed in terms of experimentally measurable mobilities of protein and dsDNA–ssDNA chimera.

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

3.2 Materials and Methods

3.2.1 Chemicals and Materials

Fused-silica capillary was purchased from Polymicro (Phoenix, AZ). All reagents were dissolved in 50 mM Tris-HCl at pH 8.0 (unless otherwise specified); the same buffer was used as the CE run buffer. All DNA sequences used for constructing ds-ssDNA chimeras were purchased from IDT DNA Technology Inc. (Coralville, IA). The sequences were as follows: alexa80, 5′-alexa-TGA CTC CCA AAT CGA TGT GTT CCG CAA GAA GCC TGG TAA GCG GAG AAA GGT CGT TTT ACT GCC CGG TCT ACC TGA TGG CG-3′; alexa60, 5′-alexa-TCC GCA AGA AGC CTG GTA AGC GGA GAA AGG TCG TTT TAC TGC CCG GTC TAC CTG ATG GCG-3′; alexa40, 5′-alexa-CGG AGA AAG GTC GTT TTA CTG CCC GGT CTA CCT

GAT GGC G-3′; alexa20, 5′-alexa-GCC CGG TCT ACC TGA TGG CG-3′; bioTEG-anti20, 5′-bioTEG-CGC CAT CAG GTA GAC CGG GC-3′; c1ss10, 5′-AAC GAC CTT T-3′; c2ss10, 5′-CAG GCT TCT T-3′; c3ss10, 5′-TCG ATT TGG G-3′. Alexa is the fluorophore used to label DNA; bioTEG indicates biotin linked to triethyleneglycol; and 80, 60, 40, and 20 indicate the number of nucleotides in each DNA sequence. The DNA sequences are annealed together to make different ds-ssDNA chimeras, detailed structures of which are shown in **Figure 3.2.** Annealing was achieved by incubating corresponding sequences of DNA at 90 °C for 10 min and then gradually cooling them down to the room temperature. Bodipy (4,4-difluoro-4-bora-3a,4a-diaza-s-indacene) was purchased from Life Technologies Inc. (Burlington, ON, Canada). Structural details of dsDNA and chimeric-DNA.
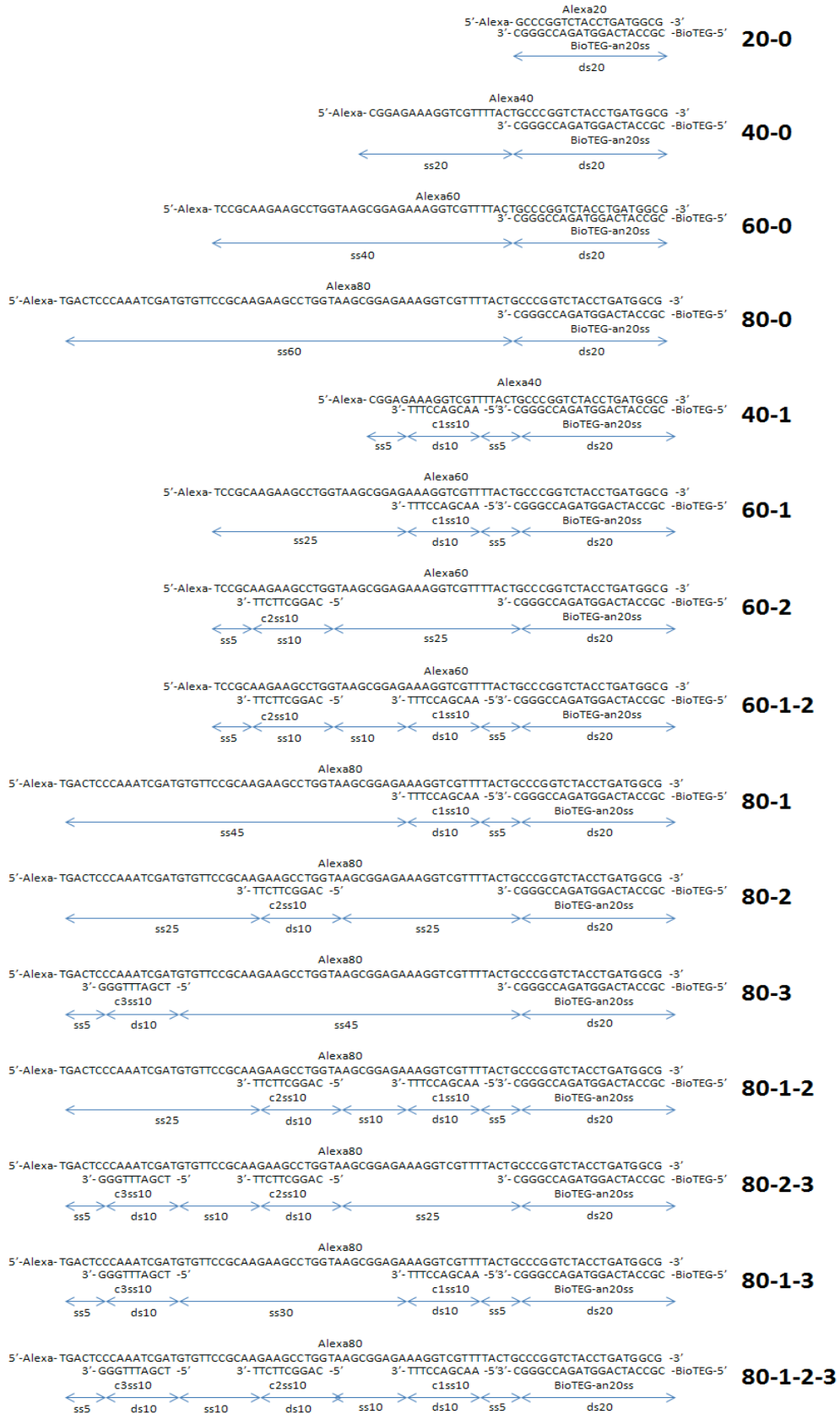
Figure 3.2. Structural details of dsDNA and chimeric-DNA.

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

Biotin and Gly-(l)Leu-4-carboxybenzene sulfonamide (GLCBS-l-leucine) were used as small-molecule heads for binding to SA and CAII, respectively. The DNA-tagged small molecules will be referred to as biotin ligand and GLCBS-l-leucine ligand. Detailed synthetic procedures for these ligands were previously described with a modification of the closing primer ligation method [35]. Klenow polymerization was eliminated and the longer oligo strand was changed to the top, leaving a 31-nucleotide 3′ overhang to provide a noncompetitive priming site for more efficient PCR amplification. All other reagents were purchased from Sigma–Aldrich (Oakville, ON, Canada). All solutions were made in deionized water filtered through a 0.22 µm filter (Millipore, Nepan, ON, Canada).

3.2.2 Instrumentation and Capillary Electrophoresis Conditions

All CE experiments were carried out on MDQ-PACE instrument (Sciex, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detector. LIF signal was recorded at 520 nm for fluorescein, alexa, and bodipy detection and at 610 nm for detection of chromeo-SA and chromeo-CAII. Signal acquisition rate was 4 Hz. Inner diameter of the capillary was 75 µm. Total capillary length was 84.3 cm, with 74.2 cm from the injection end to the detection window. The capillary was flushed prior to each CE run with 20% bleach, 0.1 M HCl, 0.1 M NaOH, deionized H_2O, and run buffer. Sample was injected into the capillary at 0.5 psi for 10 s. The

ends of the capillary were inserted into inlet and outlet reservoirs, and an electric field of 297 V/cm with a positive electrode at the injection end was applied to carry out electrophoresis. Temperature of the capillary was maintained at 15 °C. All experiments were performed in triplicate.

3.2.3 Migration Study of Protein–Ligand Complexes for Mock Ligands

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

3.2.4 Peak Identification of Biotin Ligand

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

3.2.5 Migration Studies of Protein–Ligand Complexes for GlaxoSmithKline Ligands

Two binding systems were tested in this study: SA with biotin ligand and CAII with GLCBS-l-leucine ligand. Both ligands contain the same DNA structure, shown in **Figure 3.3** a combination of two dsDNA (total of 94 bp) and two ssDNA (total of 23 nt) regions. For SA

experiments, the binding mixture was made by incubating 1 µM chromeo-SA, 100 nM biotin ligand, 10 nM fluorescein (internal standard), and 5 µM bodipy (neutral marker), at 20 °C for 30 min; control mixture was the same as the binding mixture but without protein. For CAII experiments, the binding mixture was made by incubating 5 µM chromeo-CAII, 1 µM GLCBS-l-leucine ligand, 10 nM fluorescein (internal standard), and 5 µM bodipy (neutral marker) at 20 °C for 30 min; control mixture was the same as the binding mixture but without protein.
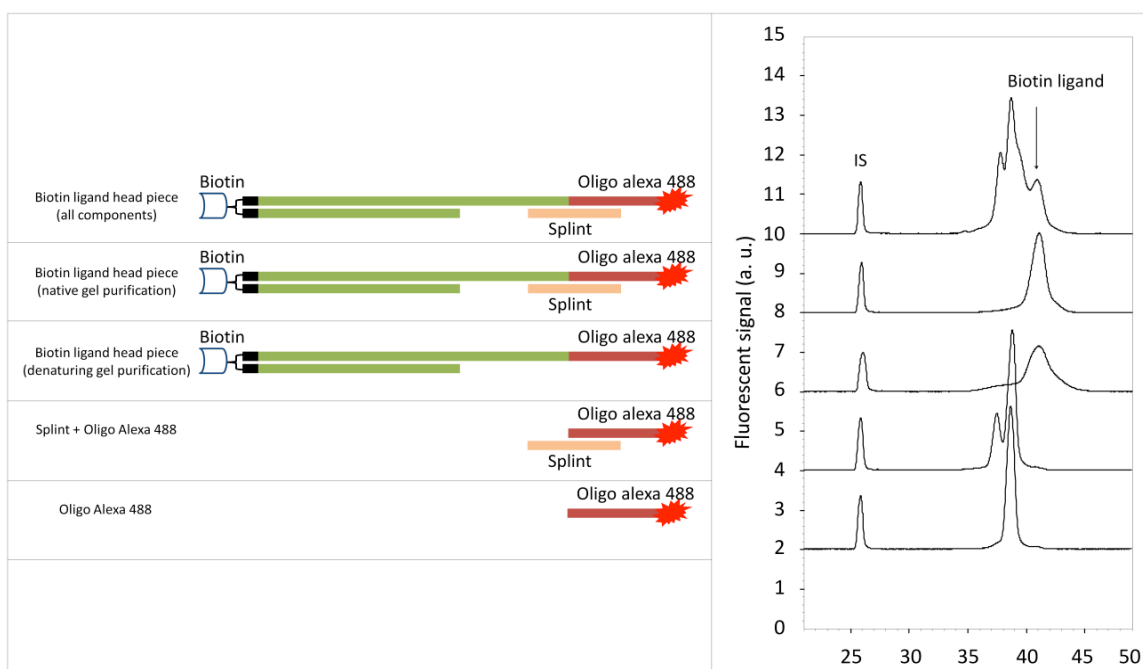
Figure 3.3. Peak identification of biotin-DEL. The schematic structures of components are illustrated on the left panel; the corresponding electropherograms are on the right panel. Each sample contained 100 nM analyzed component with 10 nM internal standard (IS). All experiments were done in triplicates.

3.3 Results and Discussion

3.3.1 Mathematical Model

In this work, we consider mobility of a protein–DNA complex in which the DNA is linear and is either pure dsDNA or a combination of dsDNA and ssDNA. The dsDNA regions are shorter than 72 base pairs, and the ssDNA regions are shorter than 50 nucleotides. This case describes a class of actual DELSMs used by pharmaceutical companies in selection of drug leads [50]. We assume that the protein is attached to one end of the dsDNA region as shown in **Figure 3.4.** This assumption excludes from consideration the "internal architecture" DELSMs (see **Figure 3.1**).

Proteins have been earlier suggested as tags in DNA sequencing based on electrophoretic mobility shift of DNA; the approach is called end-labeled free-solution electrophoresis (ELFSE) of DNA. Although ELFSE-based DNA sequencing has never been advanced beyond proof of principle, the development of ELFSE helped to make significant progress in mobility theory for protein–DNA complexes [52-58]. Such models typically use the blob theory, which is applicable to DNA that is sufficiently long to be considered a semiflexible random coil and has a length significantly greater than the diameter of the protein [56, 58]. The polymer can be considered as a semiflexible random coil if its contour length L is much greater (one or more orders of magnitude) than the Kuhn length b_K characterizing the polymer stiffness [59, 60]. This assumption is not satisfied for dsDNA of fewer than 72 base pairs, for which $L_{dsDNA} < 24$ nm while $b_{K,dsDNA} > 100$ nm. Here and below, dsDNA and ssDNA in the subscript indicate that the corresponding parameters describe dsDNA or ssDNA. Thus, the complex of a protein linked to dsDNA can be considered as a rigid object with a diameter of more than 10 nm (for dsDNA at

least 20 base pairs long and protein diameter ~4 nm). For the contour length of ssDNA of fewer

than 50 nt, we have $L_{ssDNA} < 21$ nm. Thus, L_{ssDNA} is of the same order of magnitude as the

diameter of the protein–dsDNA complex (>10 nm). In this case, blob theory is not applicable.

Moreover, ssDNA itself cannot be considered as a semiflexible random coil since its length is

only three times larger than its Kuhn length, $b_{K,ssDNA}$ ~6 nm [56]. Thus, for DNA of the lengths

considered here, the protein–dsDNA complex is a rigid object and ssDNA cannot be treated as

semiflexible random coil.
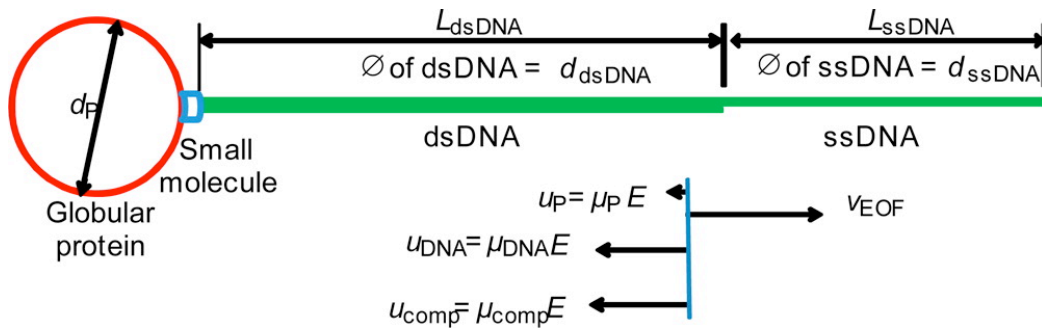
Figure 3.4. Schematic representation of one example of a complex between a globular protein
and a ligand from DELSM. Binding is achieved via protein–small molecule interaction. The
DNA tag in this example is composed of one dsDNA and one ssDNA region. The lower part
illustrates relative values of velocities of EOF, protein, ligand, and protein–ligand complex.

We will study the electrophoretic mobility of a complex formed by a globular protein

attached to the end of a stretch of DNA that contains at least one dsDNA and one ssDNA regions

(see **Figure 3.4**). This model does not describe the case of DTS internal architecture DELSM

(see **Figure 3.1**). We consider globular proteins with a molecular weight of ≥30 kDa. Their

average diameter can be estimated as $d_P \geq 4$ nm [68]. Thus, d_P is larger than the Debye length for

the buffer, $\lambda_D \sim 1$ nm. In this case, the electrophoretic mobility μ_p of the protein can be estimated

by an expression used in thin double layer theory [52, 56]:

$$\mu_P = \frac{\varepsilon_0 \varepsilon_r \zeta_P}{\eta} \approx \frac{-\sigma_P \lambda_D}{\eta}, \quad \zeta_P \approx \frac{-\sigma_P \lambda_D}{\varepsilon_0 \varepsilon_r} \tag{3-1}$$

Here ε_0 is vacuum permittivity, ε_r is relative permittivity of the buffer, ζ_P is ζ potential of the

globular protein, σ_P is average surface density of electric charge in the diffuse part of the double

layer around the protein (i.e., excluding the Stern layer), and η is dynamic viscosity of the buffer.

Equation 3-1 can be rewritten as follows:

$$\mu_P = \frac{-\sigma_P \lambda_D}{\eta} = \frac{Q_P \lambda_D}{\pi \eta d_P^2}, \quad \sigma_P = -\frac{Q_P}{\pi d_P^2} \tag{3-2}$$

where Q_P is electric charge of the protein (including the Stern layer charge). Note that protein

mobility can have both positive and negative values (for positively and negatively charged

proteins, respectively).

Equation 3-2 for μ_P can be also obtained from the balance of electric and hydrodynamic forces,

$F_{E,P}$ and $F_{H,P}$, acting upon the protein molecule:

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

if the following effective values for these forces are assumed:

$$F_{E,P} = Q_P E, \quad F_{H,P} = -\frac{\pi \eta d_P^2}{\lambda_D} u_P \tag{3-4}$$

Here, E is electric field strength, u_P is relative velocity of the protein with respect to buffer, and

$\mu_P = u_P/E$. Hereafter, we use a coordinate system in which both electric and hydrodynamic forces

have only x-components. We will use equation 3-4 in balance of all forces acting upon the

complex (see equation 3-14 below) to find the complex mobility.

Taking into account that L_{dsDNA} is smaller than $b_{K,dsDNA}$, we assume that dsDNA (shorter than 72 bp) behaves like a rigid rod. The dsDNA diameter, d_{dsDNA}, can be estimated as 2 nm [61, 62], which is larger than λ_D, while L_{dsDNA} is many times larger than λ_D. Thus, we can assume that the electrophoretic mobility of dsDNA, μ_{dsDNA}, is determined by an expression used in thin double layer theory [52, 56]:

$$\mu_{dsDNA} = \frac{\varepsilon_0 \varepsilon_r \zeta_{dsDNA}}{\eta} \approx \frac{-\sigma_{dsDNA} \lambda_D}{\eta}, \quad \zeta_{dsDNA} \approx \frac{-\sigma_{dsDNA} \lambda_D}{\varepsilon_0 \varepsilon_r} \qquad (3\text{-}5)$$

Here, ζ_{dsDNA} is ζ potential of dsDNA and σ_{dsDNA} is surface density of the electric charge in the diffuse part of the double layer around dsDNA (i.e., excluding the Stern layer). Equation 3-5 can be rewritten as follows:

$$\mu_{dsDNA} = \frac{-\sigma_{dsDNA} \lambda_D}{\eta} = \frac{q_{dsDNA} \lambda_D}{\pi \eta d_{dsDNA}}, \quad \sigma_{dsDNA} = -\frac{q_{dsDNA}}{\pi d_{dsDNA}} \qquad (3\text{-}6)$$

where q_{dsDNA} is charge per unit length of dsDNA. In calculations of q_{dsDNA}, we should take into account the condensation of counterions on dsDNA [63-67]. The condensation takes place for cylindrical objects with linear density electric charge, q, satisfying the following relationship [63]:

$$|q| \geq q_{eff}, \quad q_{eff} = \frac{e}{z_i \lambda_B}, \quad \lambda_B = \frac{e^2}{4\pi \varepsilon_0 \varepsilon_r k_B T} \qquad (3\text{-}7)$$

Here e is proton charge, z_i is the valence of counterions, λ_B is Bjerrum length, k_B is the Boltzmann constant, and T is absolute temperature of the buffer. Usually, dsDNA has two negative charges per 0.34 nm of its length [56] and $\lambda_B = 0.7$ nm for water solutions at room temperature [55, 65]. Thus, equation 3-7 is always satisfied for dsDNA, and condensation of counterions reduces the density of DNA charge q_{DNA} (excluding the Stern layer) to the effective value $-q_{eff}$, determined by the second relationship in equation 3-7 [63]. Since we consider the

Stern layer as a part of the condensed counterion layer, $|q_{DNA}|$ will be even less than q_{eff}. In this case, q_{DNA} can be considered as an adjustable parameter. We should note that dsDNA mobility has negative values since dsDNA is negatively charged.

Equation 3-6 for μ_{dsDNA} can be also obtained from the balance of electric and hydrodynamic forces, $F_{E,dsDNA}$ and $F_{H,dsDNA}$, acting upon dsDNA:

$$F_{E,dsDNA} + F_{H,dsDNA} = 0 \qquad\qquad (3\text{-}8)$$

if we assume the following effective values for these forces:

$$F_{E,dsDNA} = q_{dsDNA} L_{dsDNA} E, \quad F_{H,dsDNA} = -\frac{\pi \eta d_{dsDNA} L_{dsDNA}}{\lambda_D} u_{dsDNA} \qquad (3\text{-}9)$$

Here u_{dsDNA} is relative velocity of dsDNA with respect to buffer. We will use equation 3-9 (and similar expressions obtained for ssDNA) in the equation of balance of all forces acting upon the complex to find the complex mobility.

Similarly to equation 3-5, we can determine the electrophoretic mobility of ssDNA using an expression from thin double layer theory:

$$\mu_{ssDNA} = \frac{\varepsilon_0 \varepsilon_r \zeta_{ssDNA}}{\eta} \approx \frac{-\sigma_{ssDNA} \lambda_D}{\eta}, \quad \zeta_{ssDNA} \approx \frac{-\sigma_{ssDNA} \lambda_D}{\varepsilon_0 \varepsilon_r} \qquad (3\text{-}10)$$

Here, ζ_{ssDNA} is ζ potential of ssDNA and σ_{ssDNA} is surface density of electric charge in the diffuse part of the double layer around ssDNA. Equation 3-10 can be rewritten as follows:

$$\mu_{ssDNA} = \frac{-\sigma_{ssDNA} \lambda_D}{\eta} = \frac{q_{ssDNA} \lambda_D}{\pi \eta d_{ssDNA}}, \quad \sigma_{ssDNA} = -\frac{q_{ssDNA}}{\pi d_{ssDNA}} \qquad (3\text{-}11)$$

where d_{ssDNA} is ssDNA diameter and q_{ssDNA} is charge per unit length of ssDNA. To find q_{ssDNA}, we also have to take into account condensation of counterions on ssDNA and the Stern layer charge [63-67].

Equation 3-11 can be obtained from the balance of all effective forces acting upon ssDNA:

$$F_{E,ssDNA} + F_{H,ssDNA} = 0 \tag{3-12}$$

where $F_{E,ssDNA}$ and $F_{H,ssDNA}$ are effective electric and hydrodynamic forces acting upon ssDNA.

They are determined by the following relationships similar to equation 3-9:

$$F_{E,ssDNA} = q_{ssDNA}L_{ssDNA}E, \quad F_{H,ssDNA} = -\frac{\pi\eta d_{ssDNA}L_{dsDNA}}{\lambda_D}u_{ssDNA} \tag{3-13}$$

Here u_{ssDNA} is relative velocity of ssDNA with respect to buffer. Equation 3-13 will be used in

equation 3-14.

The electrophoretic mobility of a globular protein attached to the end of dsDNA, the other end of

which is linked to ssDNA, can be found from the balance of all effective forces acting upon such

a complex:

$$F_{E,P} + F_{E,dsDNA} + F_{E,ssDNA} + F_{H,P} + F_{H,dsDNA} + F_{H,ssDNA} = 0 \tag{3-14}$$

Substitution of equation 3-4, equation 3-9, and equation 3-13 into equation 3-14 gives

$$\left(Q_P + q_{dsDNA}L_{dsDNA} + q_{ssDNA}L_{ssDNA}\right)E = \left(\frac{\pi\eta d_P^2}{\lambda_D} + \frac{\pi\eta d_{dsDNA}L_{dsDNA}}{\lambda_D} + \frac{\pi\eta d_{ssDNA}L_{ssDNA}}{\lambda_D}\right)u_{comp} \tag{3-15}$$

By solving this equation with respect to u_{comp} and taking into account that $u_{comp} = \mu_{comp}E$, we

obtain the electrophoretic mobility of the complex, μ_{comp}:

$$\mu_{comp} = \frac{Q_P + q_{dsDNA}L_{dsDNA} + q_{ssDNA}L_{ssDNA}}{\dfrac{\pi\eta d_P^2}{\lambda_D} + \dfrac{\pi\eta d_{dsDNA}L_{dsDNA}}{\lambda_D} + \dfrac{\pi\eta d_{ssDNA}L_{ssDNA}}{\lambda_D}} \tag{3-16}$$

Taking into account equation 3-2, equation 3-6, and equation 3-11 for electrophoretic mobilities

of globular protein, dsDNA, and ssDNA, we rewrite equation 3-16 as follows:

$$\mu_{comp} = \frac{d_P^2\mu_P + d_{dsDNA}L_{dsDNA}\mu_{dsDNA} + d_{ssDNA}L_{ssDNA}\mu_{ssDNA}}{d_P^2 + d_{dsDNA}L_{dsDNA} + d_{ssDNA}L_{ssDNA}} \tag{3-17}$$

In the absence of protein, equation 3-17 reduces to the expression for electrophoretic mobility of ds-ssDNA chimera, $\mu_{\text{ds-ssDNA}}$:

$$\mu_{\text{ds-ssDNA}} = \frac{d_{\text{dsDNA}}L_{\text{dsDNA}}\mu_{\text{dsDNA}} + d_{\text{ssDNA}}L_{\text{ssDNA}}\mu_{\text{ssDNA}}}{d_{\text{dsDNA}}L_{\text{dsDNA}} + d_{\text{ssDNA}}L_{\text{ssDNA}}} \qquad (3\text{-}18)$$

Given equation 3-18, we can express the electrophoretic mobility of complex in terms of the electrophoretic mobilities of protein and ds-ssDNA chimera:

$$\mu_{\text{comp}} = \frac{d_{\text{P}}^2\mu_{\text{P}} + \left(d_{\text{dsDNA}}L_{\text{dsDNA}} + d_{\text{ssDNA}}L_{\text{ssDNA}}\right)\mu_{\text{ds-ssDNA}}}{d_{\text{P}}^2 + d_{\text{dsDNA}}L_{\text{dsDNA}} + d_{\text{ssDNA}}L_{\text{ssDNA}}} \qquad (3\text{-}19)$$

By using equation 3-19 for the electrophoretic mobility of complex, we can readily find the complex migration time to the detector, t_{comp}:

$$t_{\text{comp}} = \frac{L_{\text{capillary}}}{v_{\text{EOF}} + \mu_{\text{comp}}E} \qquad (3\text{-}20)$$

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

Derivation of equation 3-19 for the electrophoretic mobility of complex can be readily generalized for the case of ds-ssDNA molecules containing more than one dsDNA region and more than one ssDNA section. In this case, equation 3-19 will be still valid if we define L_{dsDNA} as total contour length of all dsDNA sections and L_{ssDNA} as total contour length of all ssDNA sections.

It is important to emphasize that equation 3-19 does not contain any empirical parameters except for the diameter of protein, which can typically be found from independent studies or from the literature, and the diameter and length of DNA, which are known. Therefore, no "training set" is required for making equation 3-19 eligible, and its general validity can be tested with a limited set of experimental data. If experimental systems that are poorly described by this

expression are ever found, this would mean that at least one of the following assumptions is not fulfilled: 1, the protein is globular, 2, the protein diameter is greater than Debye length (which requires that its molecular weight be ≥30 kDa), or 3, the DNA tag is rodlike. However, since the utility of the model is to predict an approximate complex mobility for selection of binders, even if the prediction has a systematic error, it can still be useful.

3.3.2 Experimental Validation of Mathematical Model

We used the interaction between SA and biotinylated ds-ssDNA to test our model expressed by equation 3-19 and equation 3-20. Biotin played the role of small molecule. The interaction between SA and biotin is renowned for its exceptionally high affinity ($K_d \approx 10^{-14}$ M). We have tested 14 different constructs of ds-ssDNA together with four dsDNA to ensure the ruggedness of our model. Detailed structural information for DNA tags used is shown in **Figure 3.2.**

As can be seen in equation 3-19 and equation 3-20, finding the mobility and migration time of the protein–DNA complex requires knowledge of the hydrodynamic sizes of protein (d_P), which is SA in this specific example, and DNA (d_{dsDNA}, d_{ssDNA}, L_{dsDNA}, and L_{ssDNA}). We used values of $d_{dsDNA} = 2.6$ nm and $d_{ssDNA} = 1.6$ nm, which include the hydration shells around dsDNA and ssDNA [69], and a value for SA molecule diameter of $d_P = 5.3$ nm determined from crystallographic studies [70]. The dsDNA and ssDNA contour lengths were calculated as $L_{dsDNA} = b_{dsDNA}N_{dsDNA}$ and $L_{ssDNA} = b_{ssDNA}N_{ssDNA}$, where $b_{dsDNA} = 0.34$ nm and $b_{ssDNA} = 0.43$ nm, the lengths of dsDNA and ssDNA monomers [56]. It is worth recalling that the mathematical model was developed with no assumptions on protein or DNA sizes except for the assumption that a protein diameter is larger than a Debye length, which is satisfied for proteins larger than 30 kDa.

Accordingly, the model is applicable to a wide range of molecular sizes provided that the preceding assumption and assumptions of a globular protein and a rodlike DNA are satisfied. In general, shorter DNA tags are beneficial, as they would allow small proteins to introduce great mobility shifts for the ligands

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

SA was labeled with chromeo, a fluorogenic dye that does not change the mobility of protein [81]; we also confirmed that the labeling did not significantly affect protein binding to biotinylated DNA. The protein could, thus, be detected with laser-induced fluorescence (LIF) at 610 nm. The biotinylated DNAs were end-labeled with the alexa dye for LIF detection at 520 nm. Protein–ligand complexes exhibited fluorescence at both wavelengths. Examples of migration patterns of protein, ds-ssDNA, and their complex are shown in **Figure 3.5.** SA is a homotetramer that can bind up to four molecules of biotin, depending on the SA/biotin concentration ratio. The peak next to IS at the right corresponds to a complex of one tetrameric SA with two molecules of biotin-containing ligand. However, in the present study we focus only on the complex with 1:1 stoichiometry.
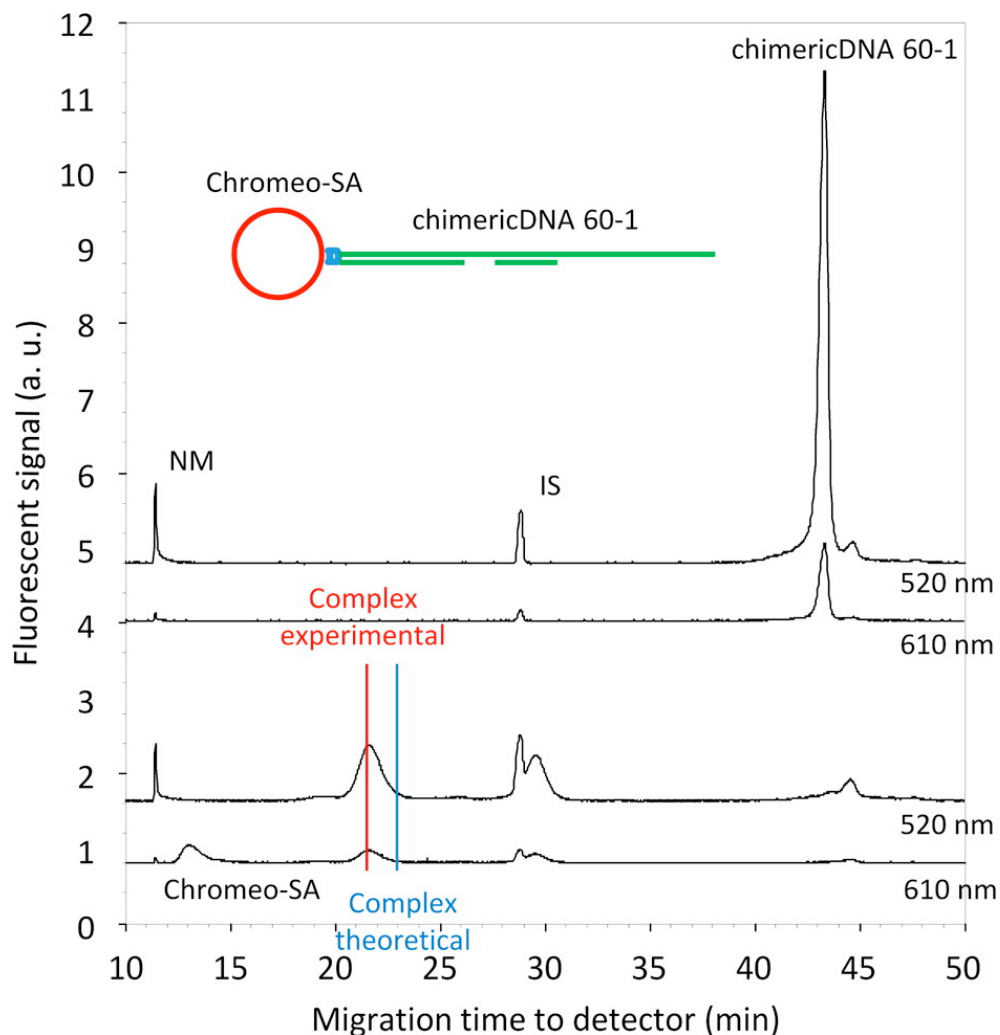
Figure 3.5. Migration analysis of complex between chromeo-SA and ds-ssDNA chimera. The top two traces represent control experiment with different detection wavelengths. The control contains 100 nM ds-ssDNA (60-1), neutral marker (NM), and internal standard (IS). The bottom two traces represent binding, which has the same composition as control plus 1 μM chromeo-SA. Experimental and theoretical positions of the complex are highlighted with red and blue lines, respectively. Traces are offset vertically for clarity. All experiments were performed in triplicate, and representative traces are shown. A schematic illustration of complex used in these experiments is shown in the top panel.

Electrophoretic mobilities of both free DNA (in the absence of protein) and free protein (in the absence of DNA) were found to be negative, which indicated that they were both negatively charged. As a result, the complex was also negatively charged and its experimentally measured

electrophoretic mobility was negative. The absolute value of protein's electrophoretic mobility was found to be significantly less than that of dsDNA. Using the current model, we calculated electrophoretic mobilities (**Table 3.1**) and migration times (**Table 3.2**) of complexes for all DNA tags. The presence of two markers was essential to ensure the precision of measured migration times and calculated mobilities. In our case, RSD was 1% for both mobility and migration times.

ds-ssDNA name	Structures	Experimental complex mobility, mm²/kVs	Theoretical complex mobility, mm²/kVs	Difference between theoretical and experimental complex mobility
20ds		13.39 ± 0.36	12.63 ± 0.39	6%
40ds		16.44 ± 0.08	16.70 ± 0.01	2%
80ds		20.07 ± 0.27	20.80 ± 0.13	4%
120ds		21.95 ± 0.08	22.81 ± 0.01	4%
40-0		15.15 ± 0.16	14.03 ± 0.70	7%
60-0		16.74 ± 0.08	17.48 ± 0.12	4%
80-0		17.76 ± 0.06	18.59 ± 0.07	5%
40-1		15.82 ± 0.02	16.37 ± 0.05	3%
60-1		17.05 ± 0.09	18.38 ± 0.11	8%
60-2		16.87 ± 0.09	18.42 ± 0.04	9%
80-1		17.64 ± 0.11	19.28 ± 0.07	9%
80-2		17.73 ± 0.04	19.30 ± 0.05	9%
80-3		17.63 ± 0.12	19.29 ± 0.07	9%
60-1-2		16.97 ± 0.03	18.66 ± 0.09	10%
80-1-2		17.75 ± 0.03	19.35 ± 0.11	9%
80-2-3		17.65 ± 0.07	19.36 ± 0.08	10%
80-1-3		17.70 ± 0.05	19.49 ± 0.06	10%
80-1-2-3		17.61 ± 0.14	19.50 ± 0.03	11%

Precisions of experimental complex mobility and theoretical complex mobility are presented as one standard deviation of results from the mean value based on three experiments.

Table 3.1. Electrophoretic mobilities of complexes between SA and ds-ssDNA chimeras of different structures.

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

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

Table 3.2. Migration times of complexes of streptavidin with ds-ssDNA chimeras of different structures.

Electrophoretic mobility of complex increased with the overall contour length, which is reasonable, as both DNA and protein are negatively charged; also, as the major contributor to the charge, DNA has the major influence on complex mobility. The dsDNA, however, has higher electrophoretic mobility than ds-ssDNA with similar contour lengths, which is also anticipated as dsDNA has a more rigid rod like shape and hence experiences less friction. On the other hand, ds-ssDNA has patches of more flexible ssDNA, which can form random-coil-like structure and hence experiences greater friction. This is the likely explanation of systematic overestimation for theoretical complex mobility with ds-ssDNA (**Table 3.1**) as the model was based on the assumption of DNA with rigid rod shape; thus, the presence of ssDNA patches introduces flexibility and decreases experimental mobility. By comparing the predicted and experimental

values of complex mobility and migration time, we found the accuracy of prediction. For all tested DNA tags, deviation of predicted from experimental values did not exceed 11%. The same data have also been tested by using the previously developed model, where DNA is considered as a rigid rod shape with constant diameter. The old model resulted in approximately doubled errors for ds-ssDNA.

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

When sampled in CE without proteins, the unbound ligands revealed several peaks, suggesting that, in addition to the ligands, the samples contained impurities. The impurities were identified as the starting material and the intermediates from each step of synthesis used for manufacturing of DELSMs (**Figure 3.3**). The full-length ligand contains the most negative charge and bears the highest electrophoretic mobility. Accordingly, it was identified as the rightmost peak in the electropherogram. In this study, we focused on migration patterns of full-length ligand and the corresponding protein–ligand complex.

Electropherograms for protein–ligand binding experiments are shown in **Figure 3.6.** In each panel, the top two traces are the no-protein control and the bottom two traces correspond to sampling the protein–ligand mixture. SA is built of four subunits and can bind up to four biotin molecules. Accordingly, complexes with different binding stoichiometries are seen in panel A. In this study, we considered only 1:1 binding; the corresponding complex is indicated by the red

line in panel A. By using our mathematical model, we found electrophoretic mobilities of protein–ligand complexes of 21.97 ± 0.09 mm^2(kV·s)$^{-1}$ for biotin ligand and 23.35 ± 0.04 mm^2(kV·s)$^{-1}$ for GLCBS-l-leucine ligand. Deviations between experimentally and theoretically determined complex mobilities were found to be 3% for biotin ligand and 5% for GLCBS-l-leucine ligand. Such accurate prediction will guarantee accurate collection of protein–ligand fraction in selection experiments.
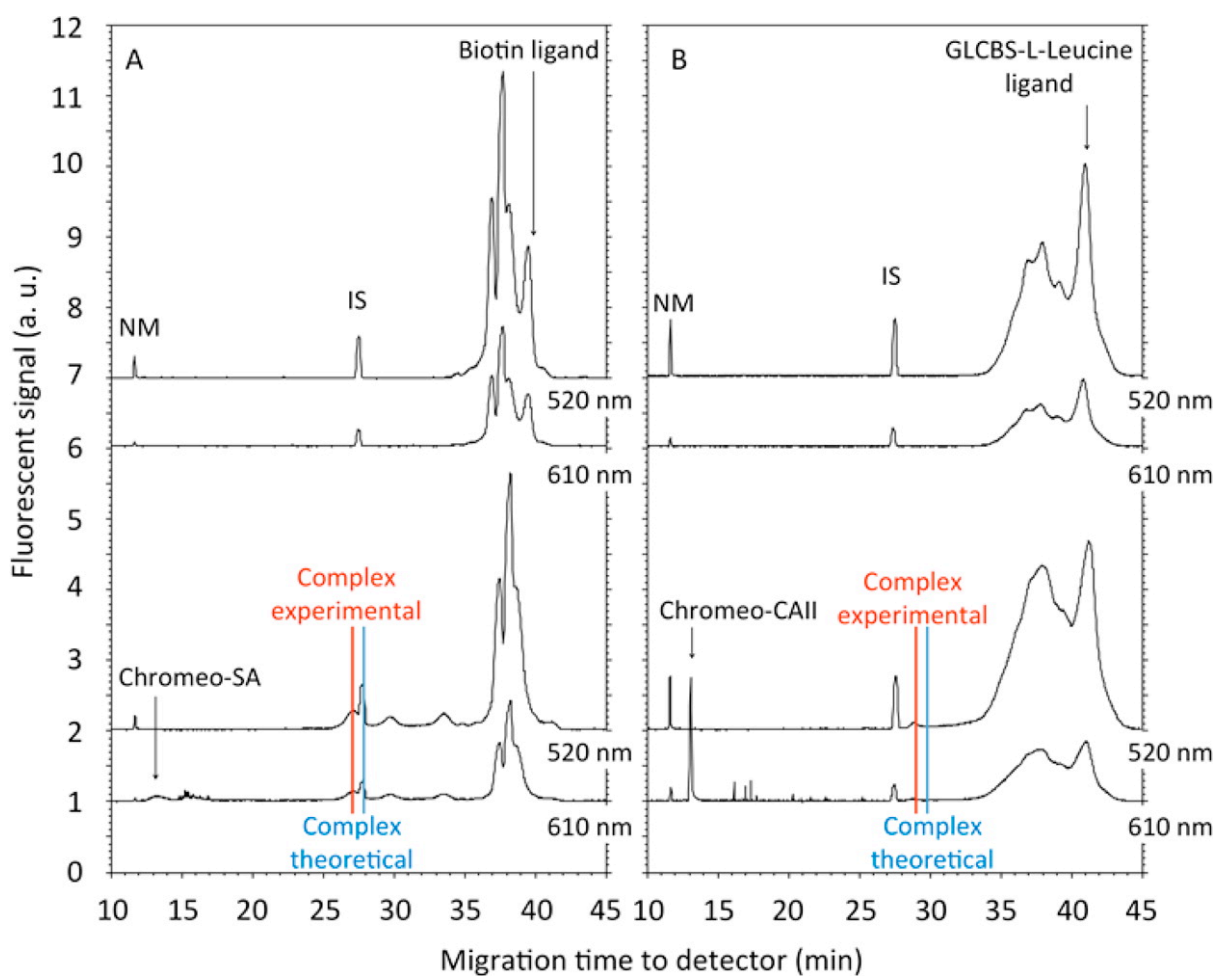
Figure 3.6. Migration study for protein–ligand complex between (A) 1 μM chromeo-SA and 100 nM biotin ligand and (B) 5 μM chromeo-CAII and 1 μM GLCBS-l-leucine ligand. In each panel, the top two traces represent the no-protein control, which contained 100 nM ligand, neutral marker (NM), and internal standard (IS). The bottom two traces correspond to the protein–ligand binding experiment, which had the same composition as the control plus 1 μM protein. Experimental and theoretical positions of the complexes are highlighted with red

and blue lines, respectively. Traces are offset vertically for clarity. All experiments were performed in triplicate, and representative traces are shown.

3.4 Conclusions

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

By now we have successfully developed the primary screening method that is generic and applicable for commercial samples. Therefore, the next logical step would be developing a label-free solution-based method for kinetic analysis of target-ligand interaction. In the former study,

we have built KCE-MS, which involves off-line coupling between solution based KCE methods and label-free mass spectrometry detection. However as we previously revealed, there was an issue of buffer incompatibility. Hence, we took a step back and started looking for the physiological conditions that are compatible for target-ligand interactions as well as generic label-free mass spectrometry detections.

CHAPTER 4. VOLATILE KINETIC CAPILLARY ELECTROPHORESIS FOR STUDIES OF PROTEIN-SMALL MOLECULE INTERACTIONS

The presented material was published previously and reprinted with permission from "Bao, J., & Krylov, S. N. (2012). Volatile kinetic capillary electrophoresis for studies of Protein–Small molecule interactions. *Analytical Chemistry*, *84*(16), 6944–6947. doi:10.1021/ac301829t" Copyright 2012 American Chemical Society. My contribution to the article was: (i) planning all experiments, (ii) performing all experiments, (iii) interpreting the results, (iv) preparing figures, (v) writing the manuscript.

4.1 Introduction

The previous chapters introduced a novel KCE-DEL based approach for primary ligand screening, which focuses on selecting target binders from the large populated combinatorial library. Subsequently, the selected ligands should be subjected to the secondary screening, in which the kinetic parameters are determined for each target-ligand interactions. Evaluations of the true kinetic parameters require a label-free analysis with high detection power. Mass spectrometry (MS) is a powerful and generic method for label-free small molecule detection. This, however, leads to a standard issue of buffer incompatibility with MS-based approaches. The standard MS analyses utilize the volatile buffer systems, which are not physiological to support interactions between protein and a small molecule. In this study, we are developing a

panel of buffer systems that are: (i) volatile and compatible for MS analysis; (ii) physiological thus suitable for protein-small molecule interactions.

Protein–small molecule affinity interactions play an important role in regulatory biological processes [16, 82, 83]. Furthermore, the action of most prospective small-molecule drugs is based on drugs' ability to form affinity complexes with their therapeutic targets, which are typically proteins [10, 84]. The formation and dissociation of an affinity complex, T-L, between protein target, T, and small molecule ligand, L, are characterized by rate constants k_{on} and k_{off} of the forward and reverse processes, respectively:

$$T + L \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} T\text{-}L$$

$$(4\text{-}1)$$

and the stability of the complex is described in terms of the equilibrium dissociation constant K_d = k_{off}/k_{on}. Since the three constants, k_{on}, k_{off}, and K_d, are interconnected, determining any pair of constants will define the third.

Kinetic affinity methods can measure k_{on}, k_{off}, and K_d for protein–small molecule binding and are, thus, essential for understanding the dynamics of biological processes and developing protein-binding small-molecule drugs [85]. Such methods fall into two major categories: heterogeneous and homogeneous [86]. Heterogeneous methods require the immobilization of a small molecule on the surface of a sensor for sensitive detection [87, 88]. The immobilization of a small molecule is usually difficult without affecting its ability to bind the protein [89]. Homogeneous methods do not require the immobilization of any of the binding partners and are, in general, preferred over heterogeneous methods in protein–small molecule studies [90]. However, homogeneous methods often require labeling of a small molecule, which reduces advantages over heterogeneous methods.

Kinetic capillary electrophoresis (KCE) is a toolset of homogeneous kinetic affinity methods with a potential application to protein–small molecule studies [30, 43, 91, 92]. Conceptually, the protein and small molecule are allowed to interact; then, the protein–small molecule complex is separated from the small molecule by capillary electrophoresis, and the small molecule is detected at the end of the capillary (the small molecule is chosen for detection as it experiences a much greater mobility shift than the protein within the affinity complex). The values of k_{on}, k_{off}, and K_d are then determined from the temporal propagation pattern (signal versus migration time) of the small molecule. Most KCE applications have been developed with fluorescence detection to allow high sensitivity and selectivity [28]. Since the small molecule is detected in KCE of protein–small molecule interactions, it should be labeled for fluorescence detection. Fluorescent labeling of small molecules is impractical in most cases.

As an alternative to KCE with fluorescence detection, we have recently suggested KCE with mass-spectrometry (MS) detection which can facilitate label-free analysis of protein–small molecule binding [29]. A similar approach has also been implemented by Sun and coauthors [93]. Our experiments revealed a serious obstacle in the way of making KCE-MS a widely used practical tool. Following the "tradition" of affinity methods, KCE has been always run in near-physiological buffers, such as Tris-acetate and Tris-HCl, when protein–ligand interactions were studied. These buffers are not suitable for MS as they suppress ionization and, thus, lead to a poor limit of detection (LOD). As an illustration, we compared the LOD for electrospray ionization MS of alprenolol dissolved in a nonvolatile Tris-acetate and volatile ammonium acetate and found it to be approximately 30 nM and 300 pM, respectively. It is clear that Tris-acetate cannot be used in KCE-MS of highly stable complexes with low nM affinity as such studies need to be conducted at small molecule concentrations below the LOD. In addition, Tris-

67

acetate leads to solid deposition at the ionization source, which requires its impractically frequent cleaning. We have, thus, been developing approaches to resolve the problem of buffer mismatch in KCE-MS.

While investigating relatively sophisticated modifications to MS, we asked ourselves a very simple question: Can volatile buffers that are suitable for MS be used for KCE involving proteins? On the one hand, we could not find in the literature references to the use of volatile buffers in affinity studies involving proteins. On the other hand, we could not find any reference to the incompatibility of protein–ligand interactions with "non-toxic" volatile buffers. We, thus, set a goal to test a few volatile buffers in KCE analysis of a few protein–ligand pairs. We chose three nontoxic volatile buffers that can be used at neutral pH values: ammonium acetate, ammonium bicarbonate, and ammonium formate. Tris-acetate was used as a near-physiological buffer control. The three protein–ligand pairs investigated were single-stranded DNA binding (SSB) protein with single-stranded DNA (ssDNA), α_1-acid glycoprotein (AGP) protein with bodipy, and MutS protein with its DNA aptamer. Both DNA molecules were fluorescently labeled for detection; bodipy is a fluorescent dye that requires no labeling.

A KCE method termed nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) [94] was used in this work because it allowed us to separately study the influence of buffers on protein–ligand complex stability and on separation of unbound ligand from protein–ligand complex. In NECEEM, an equilibrium mixture of the interactants is first prepared in an incubation buffer; a small volume of mixture is sampled for electrophoresis, and the protein–ligand complex is separated from the unbound ligand in a run buffer. The values of K_d and k_{off} can be determined from peak areas in an electropherogram. Equilibrium is established in the incubation buffer and, thus, K_d corresponds to the incubation buffer conditions. Complex

dissociation, in contrast, occurs in the run buffer; thus, k_{off} is measured under the conditions of the separation buffer. The two buffers may be different or the same.

4.2 Materials and Methods

4.2.1 Chemicals and Materials

Human AGP protein, fluorescein, and all buffer components were obtained from Sigma-Aldrich (Oakville, ON, Canada). Fluorescently labeled ssDNA oligonucleotides 5′-CCC TAT AGT GAG TCG TAT TA-3′ and MutS aptamer 5′-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′ were purchased from IDT DNA Technology Inc. (Coravile, IA, USA). SSB protein from *Escherichia coli* was ordered from Epicentre Biotechnologies (Madison, WI, USA). The UltraTrol dynamic coating was purchased from Target Discovery, Inc. (Palo Alto, CA, USA). MutS protein was purchased from GenScript (Piscataway, NJ, USA). Bodipy was purchased from Invitrogen (Grand Island, NY, USA). The fused-silica capillaries were purchased from Polymicro (Phoenix, AZ, USA). All solutions were made using deionized water filtered through a 0.22 μm filter (Millipore, Nepan, ON, Canada).

4.2.2 Instrumentation

All CE experiments were carried out with an MDQ-PACE instrument (Beckman-Coulter, ON, Canada) equipped with a laser induced fluorescent (LIF) detector. All data were recorded with a 4 Hz acquisition rate. The inner and outer diameters of the capillary were 50 and 360 μm,

respectively. The total capillary length was 38.5 cm with 28.5 cm from the injection end to the detection window.

4.2.3 Kinetic Capillary Electrophoresis (KCE)

NECEEM was used as the model KCE method. Uncoated fused-silica capillaries were used in SSB-ssDNA and AGP-bodipy experiments. The UltraTrol coated capillaries were used in MutS–aptamer experiments. Electrophoresis was run in a 30 kV negative polarity for the coated capillary and 30 kV positive polarity for the uncoated capillary. The nonvolatile buffer was 25 mM Tris-acetate, pH 7.2. There are three volatile buffers: 30 mM ammonium acetate, pH 7.2, 30 mM ammonium bicarbonate, pH 7.8, and 30 mM ammonium formate, pH 7.2. The pre-equilibrated protein–DNA binding mixtures were prepared by mixing 100 nM protein with 100 nM DNA and 100 nM fluorescein as internal standard. The AGP-bodipy mixture was made from 20 μM AGP, 20 μM bodipy with 1 μM fluorescein internal standard. Each control sample contains only ligand (DNA or bodipy) with internal standard. The samples were injected into capillary by a pressure of 0.5 psi for 10 s. Electrophoresis was carried out with a capillary coolant temperature set at 15 °C. All experiments were performed in triplicates.

4.3 Results and Discussion

First, we studied the influence of volatile buffers on complex stabilities. The interacting pairs were incubated in volatile buffers and NECEEM was run in Tris-acetate buffer for all volatile incubation buffers. We found that NECEEM electropherograms for volatile incubation buffers were similar to the electropherograms for Tris-acetate incubation buffer. The data for the MutS–aptamer, SSB-ssDNA, and AGP-bodipy binder interactions are shown in **Figure 4.1**,

70

Figure 4.2, and **Figure 4.3** respectively. These results suggested that volatile buffers did not

significantly change complex stabilities of the three studied protein–ligand pairs.
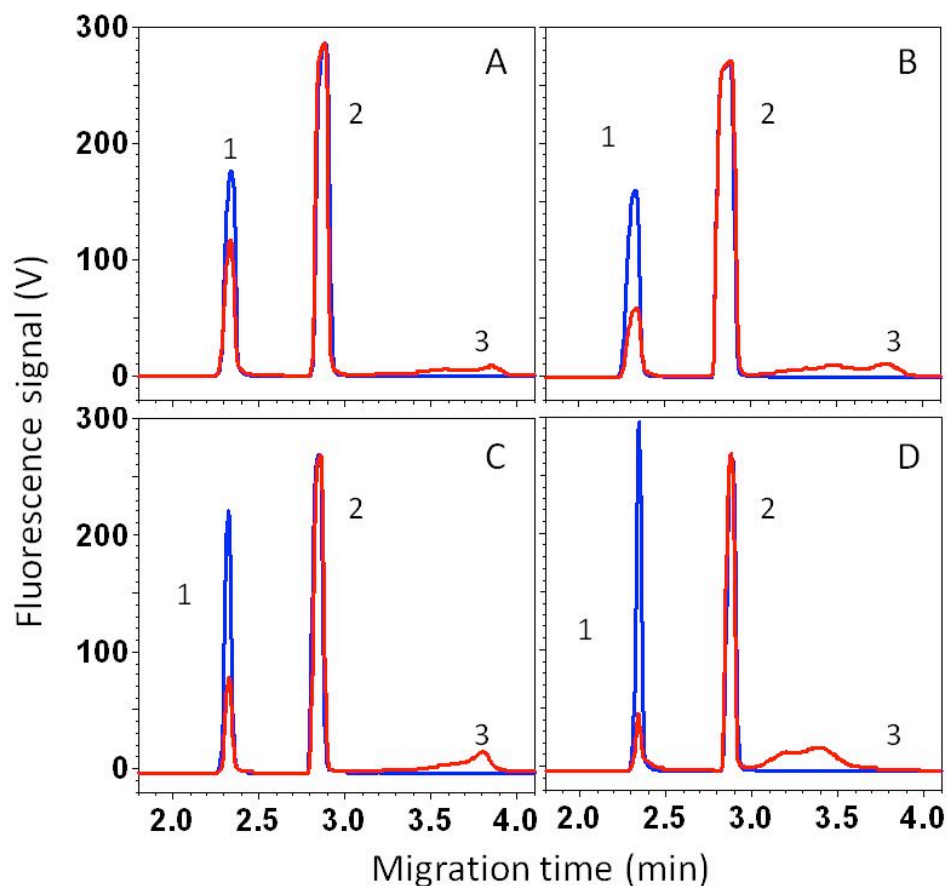
Figure 4.1. NECEEM electropherograms of MutS-aptamer binding analysis under various incubation conditions. 25 mM tris-acetate pH 7.2 (A), 30 mM ammonium acetate pH 7.2 (B), 30 mM ammonium bicarbonate pH 7.8 (C) and 30 mM ammonium formate pH 7.2 (D). The non-volatile 25 mM tris-acetate pH 7.2 was used as separation buffer. Blue and red traces represent the control and binding respectively. The numbers 1, 2 and 3 indicate the signal of free aptamer, internal standard and MutS-aptamer binding complex, respectively.

Figure 4.2. NECEEM electropherograms of SSB-ssDNA binding analysis under various incubation conditions. 25 mM tris-acetate pH 7.2 (A), 30 mM ammonium acetate pH 7.2 (B), 30 mM ammonium bicarbonate pH 7.8 (C) and 30 mM ammonium formate pH 7.2 (D). The non-volatile 25 mM tris-acetate pH 7.2 was used as separation buffer. Blue and red traces represent the control and binding respectively. The number 1, 2 and 3 indicate the signal of free ssDNA, internal standard and SSB-ssDNA binding complex correspondingly.

Figure 4.3. NECEEM electropherograms of AGP-bodipy binding analysis under various incubation conditions. 25 mM tris-acetate pH 7.2 (A), 30 mM ammonium acetate pH 7.2 (B), 30 mM ammonium bicarbonate pH 7.8 (C) and 30 mM ammonium formate pH 7.2 (D). The non-volatile 25 mM tris-acetate pH 7.2 was used as separation buffer. Blue and red traces represent the control and binding respectively. The number 1, 2 and 3 indicate the signal of free bodipy, internal standard and AGP-bodipy binding complex correspondingly.

We then examined how the volatile buffers affected separation of unbound ligands from protein–ligand complexes. In these experiments, the run buffers were identical to the incubation buffers. We found that free ligands could be separated from protein–ligand complexes in all volatile buffers tested. The electropherograms for the MutS–aptamer, SSB-ssDNA, and AGP-bodipy binder interactions are shown in **Figure 4.4**, **Figure 4.5**, and **Figure 4.6** respectively. Separation was the worst in ammonium acetate; it was the best in ammonium bicarbonate (better than in Tris-acetate). Moreover, peak shapes of the DNA aptamer in ammonium bicarbonate and ammonium formate differed from classical Gaussian (Figure 4.4, panels C and D). It is known that DNA aptamer can fold into various secondary structures under different conditions, such as buffer composition, pH, and temperature [95]. Peak shape irregularity is likely caused by the formation of multiple secondary structures of the aptamer in these two buffers. We also found that different buffers could lead to different binding stoichiometries of the affinity complexes (**Figure 4.2** and **Figure 4.5**). SSB protein is a homotetramer, which can bind to more than one ssDNA molecule [96]. Two peaks corresponding to the SSB–DNA complex are seen in the electropherograms, and the ratio between the peaks changes from buffer to buffer. The peak with the shortest migration time most likely corresponds to the complex of SSB with a single DNA molecule. The slower migrating complex is likely SSB with two ssDNA molecules.
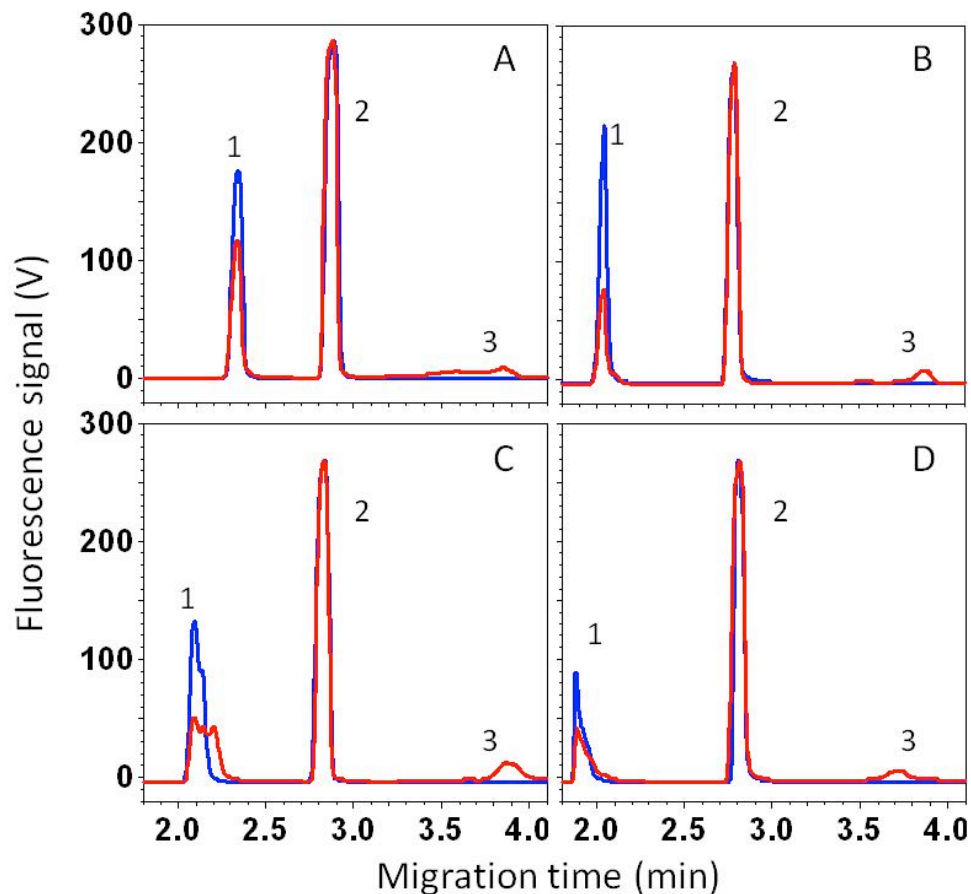
Figure 4.4. NECEEM electropherograms of MutS-aptamer binding analysis by volatile incubation and separation buffers. 25 mM tris-acetate pH7.2 (A), 30 mM ammonium acetate pH 7.2 (B), 30 mM ammonium bicarbonate pH 7.8 (C) and 30 mM ammonium formate pH 7.2 (D). Blue and red traces represent the control and binding respectively. The number 1, 2 and 3 indicate the signal of free aptamer, internal standard and MutS-aptamer binding complex correspondingly.
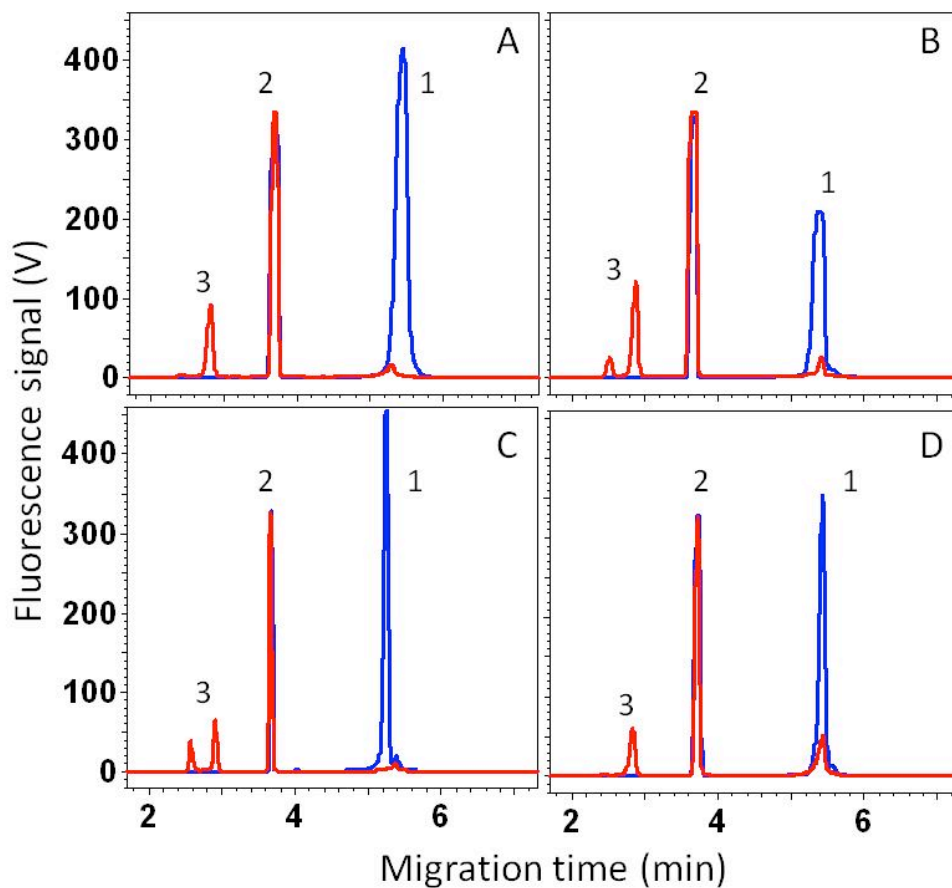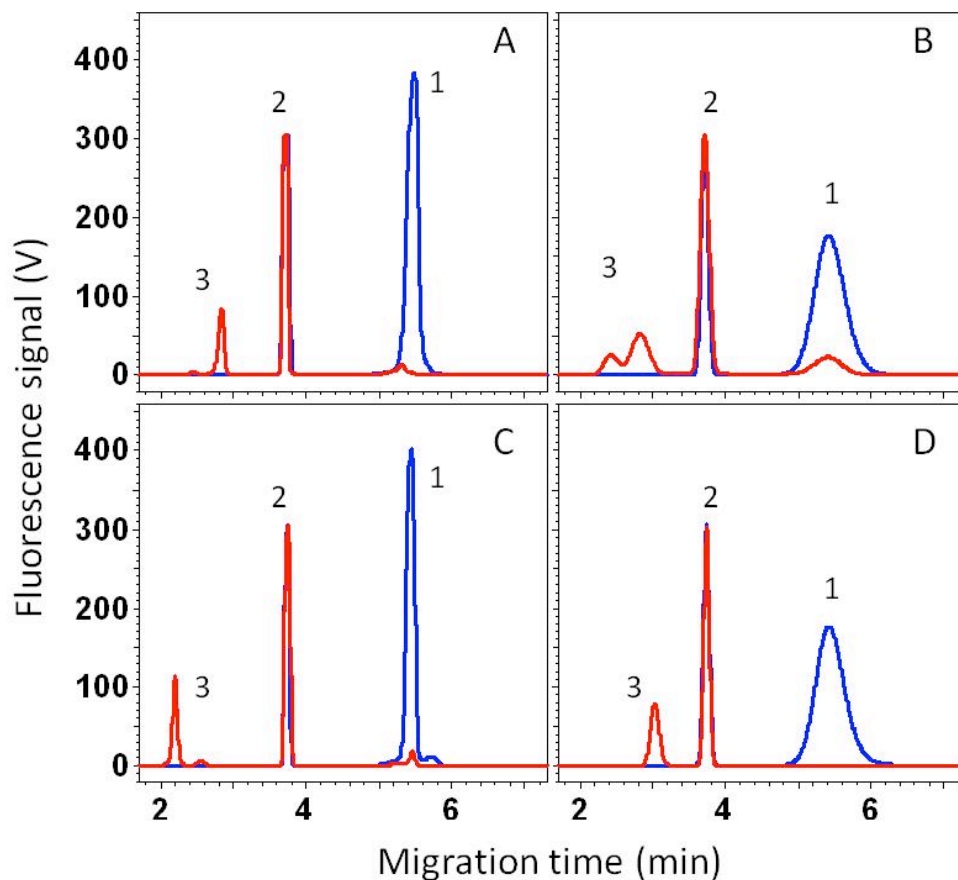
Figure 4.5. NECEEM electropherograms of SSB-ssDNA binding analysis by various incubation and separation buffers. 25 mM tris-acetate pH7.2 (A), 30 mM ammonium acetate pH 7.2 (B), 30 mM ammonium bicarbonate pH 7.8 (C) and 30 mM ammonium formate pH 7.2 (D). Blue and red traces represent the control and binding respectively. The number 1, 2 and 3 indicate the signal of free ssDNA, internal standard and SSB-ssDNA binding complex correspondingly.
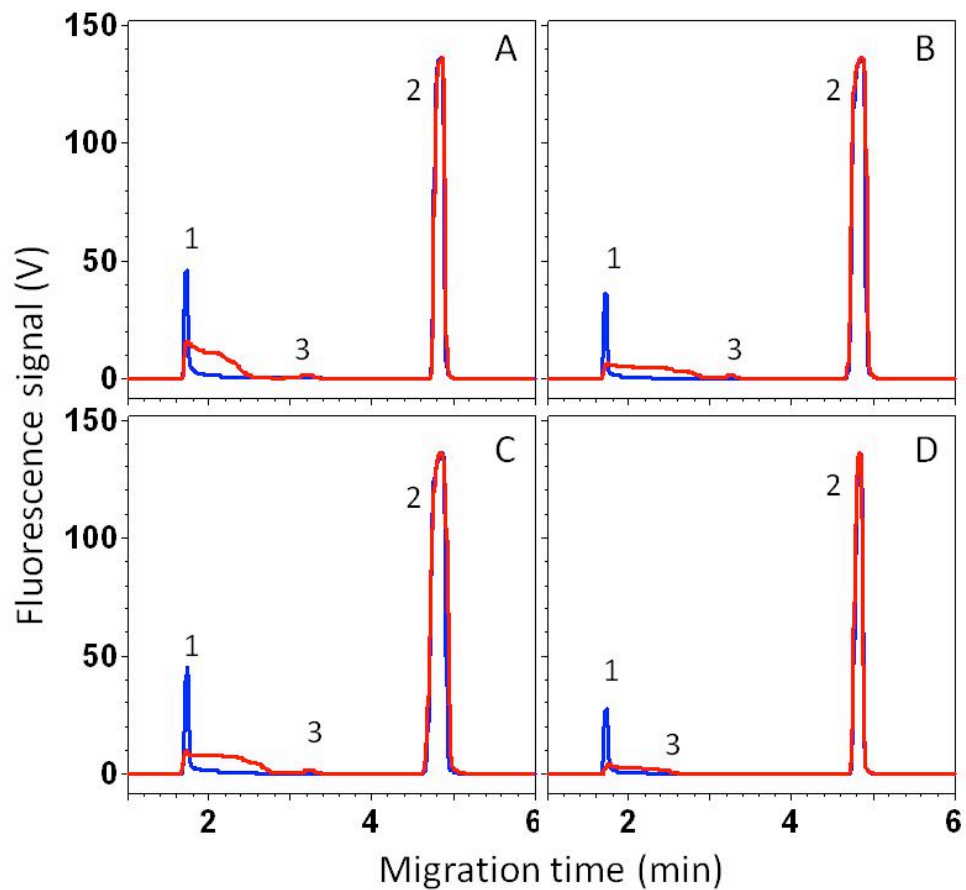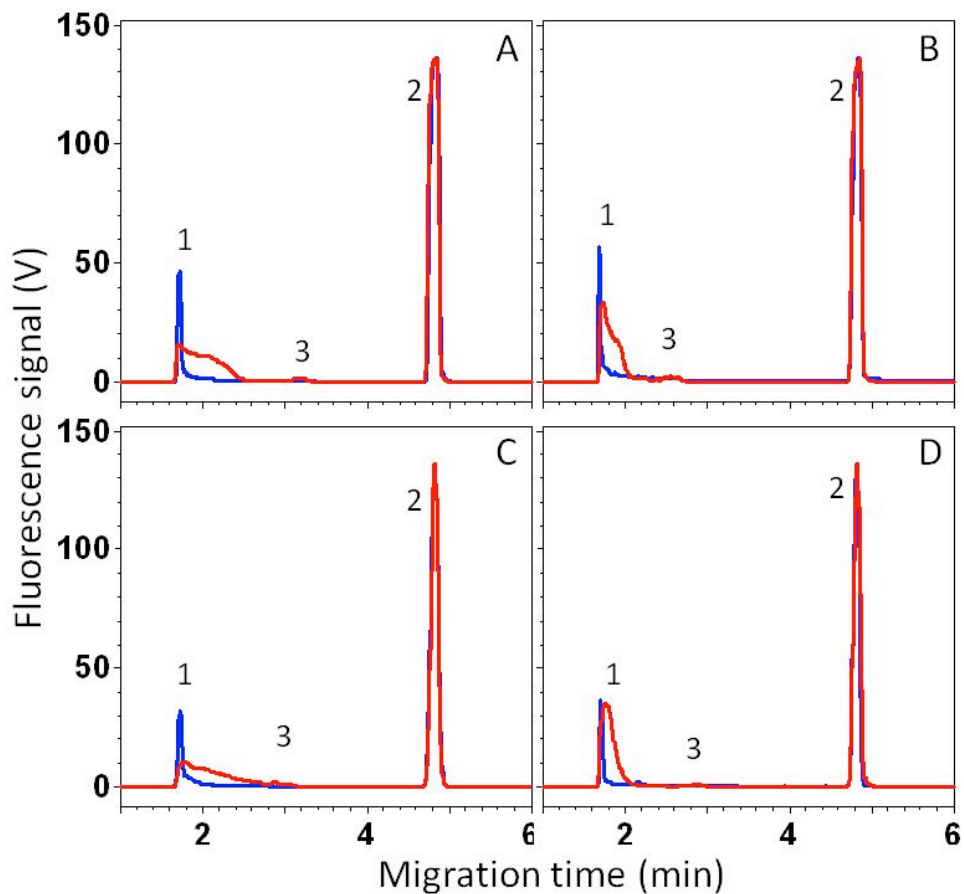
Figure 4.6. NECEEM electropherograms of AGP-bodipy binding analysis by various incubation and separation buffers. 25 mM tris-acetate pH7.2 (A), 30 mM ammonium acetate pH 7.2 (B), 30 mM ammonium bicarbonate pH 7.8 (C) and 30 mM ammonium formate pH 7.2 (D). Blue and red traces represent the control and binding respectively. The number 1, 2 and 3 indicate the signal of free bodipy, internal standard and AGP-bodipy binding complex correspondingly.

Finally, we have determined K_d values for all three interacting pairs in the four buffers studied (**Table 4.1**). While K_d was found to depend on the buffer (which was expected), most values were of the same order of magnitude. The most noticeable effect was that of the ammonium format of SSB–DNA complex: complex stability improved to the level at which K_d was too small to determine its value accurately. The quantitative results unambiguously suggest that volatile buffers did not drastically change complex stability of the three studied protein–ligand complexes.

Table 4.1. Equilibrium dissociation constants (Kd) for 3 pairs of non-covalent protein-ligand complexes measured by NECEEM in 3 different incubation/run buffers.

Buffer	K_d (nM) MutS-aptamer	K_d (µM) AGP-bodipy	K_d (nM) SSB-DNA
Tris-acetate (control)	47 ± 7	5.8 ± 0.8	1.5 ± 0.3
Ammonium-acetate	30 ± 3	12.4 ± 0.5	0.5 ± 0.3
Ammonium-bicarbonate	43 ± 8	17.9 ± 3.3	7.6 ± 6.5
Ammonium-formate	39 ± 9	3.3 ± 2.5	< 0.5

A "$\pm$" indicates the standard deviation of measurements in triplicates

4.4. Conclusions

The goal of this study was to test if volatile buffers could be used for KCE of protein–ligand interactions. This test required that we answered two questions: (1) whether or not volatile buffers significantly affect reversible protein–ligand binding and (2) whether or not volatile buffers can facilitate efficient separation of the protein–ligand complexes from the unbound ligands. These questions could be answered with fluorescently labeled ligands and simple fluorescence detection without relatively sophisticated KCE-MS schemes. Our study strongly suggests that the bias against the use of volatile buffers in studies of protein–ligand interactions is not justified. Specifically, volatile buffers can be used in KCE and will, therefore, facilitate simple coupling of KCE with MS. The tandem of volatile KCE with MS constitutes a homogeneous label-free method that promises to significantly simplify kinetic studies of protein–small molecule interactions.

Up until now, we have developed a panel of volatile buffer conditions that are compatible with biomolecule interaction. These buffer systems can facilitate KCE-MS based label-free solution-based kinetic analysis. However, as previously discussed in Chapter 1, the KCE based approaches are not the best options for protein-small molecule interactions, since neither protein nor small molecule is highly charged. Therefore, the separation condition has to be tailored for each pair of protein-small molecule interaction. This, however, is contradictory to the goal of secondary ligand screening, in which all potential ligands should be evaluated under the same condition. Therefore, KCE based methods are not ideal for secondary ligand screening. Size exclusion chromatography (SEC) or gel-filtration chromatography separate particles solely based on differences in size. Theoretically speaking, SEC is the method of choice for generic solution-based separation between a protein and small molecule. However, SEC has never been

demonstrated for analyzing interactions during separation. Therefore, in the next Chapter, we will introduce a novel label-free solution-based kinetic method for protein-small molecule interactions analysis. We call it kinetic size-exclusion chromatography mass spectrometry, KSEC-MS.

CHAPTER 5. KINETIC SIZE-EXCLUSION CHROMATOGRAPHY WITH MASS SPECTROMETRY DETECTION: AN APPROACH FOR SOLUTION-BASED LABEL-FREE KINETIC ANALYSIS OF PROTEIN–SMALL MOLECULE INTERACTIONS

My contribution to the article was: (i) planning all experiments, (ii) performing all experiments, (iii) preparing figures, (iv) writing the manuscript.

5.1 Introduction

In previous studies, we have developed KCE-DEL based primary ligand screening method that focuses on the preliminary selection of all target-binding molecules from the highly populated combinatorial library. Besides, we have also developed a panel of volatile buffer systems that are physiological to support interaction between the protein target and small molecule. In this study, we are going to introduce a novel label-free solution based kinetic

method for secondary ligand screening, in which the true kinetic parameters of protein-small molecule interactions will be evaluated.

Reversible binding between small molecules and proteins plays an important role in the regulation of various cellular processes [16]. Additionally, such interactions are important in modern drug discovery as small molecule drugs are designed to alter protein functions upon binding [10, 97, 98]. Understanding the dynamics of both cellular regulation by small molecules and the action of small molecule drugs requires knowledge of the kinetics of formation and dissociation of protein–small molecule complexes [85, 99, 100]. Thus, it is important to determine the rate constants, k_{on} and k_{off}, of the following reaction:

$$P + SM \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} P\text{-}SM \tag{5-1}$$

where P is a protein, SM is a small molecule, and P–SM is a protein–small molecule complex. Complex stability is typically characterized by an equilibrium dissociation constant $K_d = k_{off}/k_{on}$ (smaller K_d values correspond to more stable complexes), and determining any two of the three constants will define the third one.

All current methods used for practical measurements of k_{on} and k_{off} in expression 5-1 are either surface-based or label-based. Surface-based methods, such as surface plasmon resonance (SPR) [101, 102] and biolayer interferometry [22, 103], require the immobilization of either P or SM on the surface of a sensor. Label-based methods, such as stopped flow spectroscopy [104, 105], require the modification of either P or SM with a spectroscopically detectable label, typically a fluorophore. Moreover, it is preferable that SM, rather than P, is immobilized or labeled in order to maximize the sensitivity of detection [106, 107]. However, modifications of SM are difficult to achieve without drastically affecting its ability to bind P. Therefore, a solution-based label-free approach would be ideal for simple and accurate measurements of k_{on}

and k_{off}. Here, we propose such an approach, termed kinetic size-exclusion chromatography with mass spectrometry detection (KSEC-MS). Size-exclusion chromatography (SEC) allows generic separation of SM from P–SM without the immobilization of SM or P. Mass spectrometry (MS), in turn, allows the generic detection of SM without labeling it. Instrumentation-wise, SEC is easily integrated with MS, and this combination has been extensively used to study proteins, antibodies, and peptides [108-110]. In KSEC-MS, the migration pattern of SM through the column depends on k_{on} and k_{off}. The rate constants can, thus, be deconvoluted from the temporal pattern of SM elution at the exit of the SEC column.

Here, we present an implementation of KSEC-MS, in which short plugs of SM and P are injected sequentially into a SEC column without the need to premix P and SM outside the column. We call this implementation plug–plug KSEC-MS (ppKSEC-MS) in analogy with plug–plug kinetic capillary electrophoresis [30, 111, 112]. **Figure 5.1A** depicts migration of the species in a SEC column. In the beginning, a short plug of SM is injected into the column followed by injection of a short plug of P, with a small volume of buffer in between, as a spacer, to prevent mixing during the injection. The chromatographic migration is immediately started after injecting P. The molecular size of P is much larger than that of the SM and thus P moves faster than the SM. The P plug passes through the SM plug allowing for P to bind SM and form P–SM, which has a molecular size similar to that of P and thus comigrates with P. When the P/P–SM plug overtakes the SM plug, P–SM starts dissociating into P and SM. The latter is continuously separated from P and P–SM creating a trail of SM behind the P/P–SM plug. The resulting migration pattern is the following. The zone containing P and intact P–SM migrates first. The zone of SM that has not bound P (during the passage of the P plug through the SM plug) migrates the last. The trail of SM that dissociated from P–SM lies between these two

83

zones. **Figure 5.1B** schematically illustrates the detection step. P and P–SM elute first followed by SM that dissociated from P–SM and finally by SM that has not bound P. The eluate is sampled into an MS ionization source; the typical ionization methods are atmospheric-pressure chemical ionization (APCI) and electrospray ionization (ESI). The ionized SM is fragmented and detected by MS/MS, which offers high specificity and signal-to-noise ratio. All intact complexes are destroyed during the ionization so that SM is released from P–SM and also quantitated by MS/MS. In general, the time-dependence of signal from SM (a chromatogram) contains three features merging into one another: (i) a peak corresponding to SM originating from the decay of P–SM during ionization, (ii) a peak of SM that has not bound P, and (iii) a "bridge" between the two peaks that corresponds to SM that dissociated from P–SM during migration in a column (**Figure 5.1B**). The shape of the chromatogram is defined by k_{on} and k_{off} so that, in the final step of analysis, their values are found by fitting the experimental chromatogram with a computer-simulated one while varying k_{on} and k_{off} (**Figure 5.1C**). The best fit of a single chromatogram reveals the values of k_{on} and k_{off}.

To experimentally prove the suggested concept of ppKSEC-MS, we used carbonic anhydrase II (CAII) as P and acetazolamide (ACZ) as SM. CAII catalyzes the interconversion between carbon dioxide and bicarbonate, which is a critical reaction in regulating cellular respiration [113]; ACZ is a known CAII inhibitor [114]. A series of ppKSEC-MS experiments were performed at constant [SM] but varying [P]. The resulting chromatograms are shown by red traces in **Figure 5.2**. The general shape of the chromatograms corresponds to the expected one with two peaks and a bridge between them (**Figure 5.1B**). Increasing [P] led to an anticipated increase of both the leftmost peak, corresponding to intact P–SM eluting from the column, and the bridge, corresponding to SM dissociated from P–SM during its migration through the

84

column. At the same time, the rightmost peak, which corresponds to SM that had not bound P, predictably decreased with increasing [P]. The integral area under the chromatogram, which is proportional to the total amount of SM exiting the column, did not change with increasing [P]. This finding indicates that the P–SM was completely destroyed during the ionization process (which was desirable) and that all SM was accounted for.
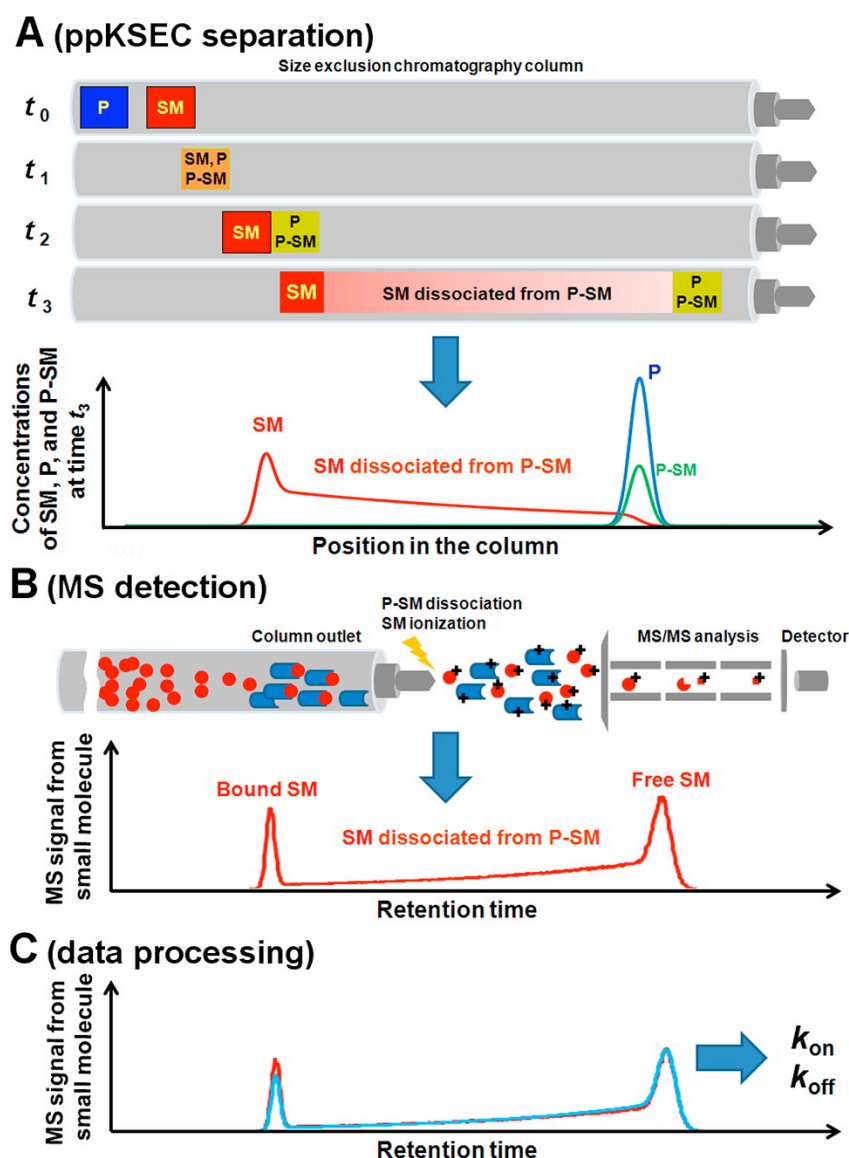
Figure 5.1. Conceptual depiction of ppKSEC-MS. Please see text for details.

5.2. Materials and Methods

5.2.1. Chemicals and Materials

Bio SEC-3 Size exclusion chromatography (SEC) columns were purchased from Agilent (Mississauga, ON, Canada). Carbonic anhydrase II (CAII), acetazolamide (ACZ) and were purchased from Sigma-Aldrich (Oakville, ON, Canada). All reagents were dissolved in 30 mM ammonium formate, pH 7.2. All other reagents were purchased from Sigma-Aldrich (Oakville, ON, Canada). All solutions were made using deionised water filtered through a 0.22 µm filter (Millipore, Nepan, ON, Canada).

5.2.2. Instrumentation

The Shimadzu UFLCXR with Agilent Bio SEC-3 was used for all experiments. The column has 3 µm particle size, 100 Å pore size, 4.6 mm inner diameter and 300 mm length. The AB Sciex QTRAP 6500 with IonDrive Turbo V Source (Concord, ON, Canada) was used for small molecule detection and quantification. Isothermal titration calorimetry (ITC) experiments were performed by using a MicroCal iTC200 system (Northampton, MA, USA).

5.2.3. Plug-plug KSEC with MS detection

The 30 mM ammonium formate, pH 7.2 was used for all sample preparation and separation. An HPLC instrument does not allow multiple injections, therefore, the plug-plug setup was realized by combing two consecutive runs. The first run started with an injection of 10 µL of ACZ, followed by a 2-min long separation with a flow rate of 0.3 mL/min. The second run

began immediately after the first run with an injection of 10 µL of CAII, followed by a 20-min long separation with a flow rate of 0.3 mL/min. Atmospheric-pressure chemical ionization (APCI) was used to ionize ACZ with negative ion mode. The source temperature (TEM) was 300°C, the ionization energy (IS) was -4,500 V, and the de-clustering potential (DP) was -125 V. The MRM mode was used to select the ion of 221/83 (Q1/Q3) by using the collision energy (CE) at of -30 V. All binding experiments were done in triplicates. Fitting the experimental ppKSEC-MS chromatograms with the simulated ones was carried out by using COMSOL Multiphysics 4.3a commercial software (COMSOL Group, Palo Alto, CA).

5.2.4. Isothermal titration calorimetry analysis.

All samples were prepared in 30 mM ammonium formate, pH 7.2. Binding experiments were conducted using 10 µM CAII and 100 µM ACZ at 25°C. The experimental setup consisted of 19 successive 2 µL injections of either ACZ or buffer into CAII every 180 s to a final molar ratio of 2:1. The first injection was 0.2 µL for all experiments. The data were corrected for the heat of dilution of the titrant. Data analysis was carried out with Origin 5.0 software

5.3. Results and Discussion

The values of k_{on} and k_{off} are convoluted into the shape of chromatograms. An analytical solution for their deconvolution does not exist, leaving us with a single option: numerical solution of an inverse problem. In essence, an experimental chromatogram should be fitted with a simulated chromatogram computed using a 1-dimensional mathematical model describing both equation 5-1 and mass transfer in a SEC column.

There are many theoretical works on the separation of polymers by SEC that consider the thermodynamics of distribution of polymer fractions between the mobile and stationary phases [115-120]. They mainly study separation principles and often use fairly complicated mathematics and detailed process descriptions. Two-dimensional hydrodynamic models have also been suggested to describe SEC columns [121-123]. These models employ additional differential equations to take into account diffusion of solutes within the bead pores. On the other hand, a simple one-dimensional hydrodynamic model can be used in our case if this model takes into account the basic features of the described experiments (**Figure 5.1**).

We developed such a model in which a long and narrow cylindrical chromatography column is coaxial with the *x* coordinate. A SEC column is packed with beads that have pores; the solution inside the pores constitutes the stationary phase. We assume that the pores are large enough for SM to enter and reside inside for a significant time and too small for the protein or the protein–small molecule complex to penetrate and be significantly retarded. This is a typical assumption that is confirmed by a significant difference in retention times between SM and P–SM (**Figure 5.2**). We also assume fast re-equilibration between the mobile phase (solution outside the beads) and stationary phase. This is also a typical assumption that is confirmed by narrowness of peaks of P–SM and SM in **Figure 5.2**.
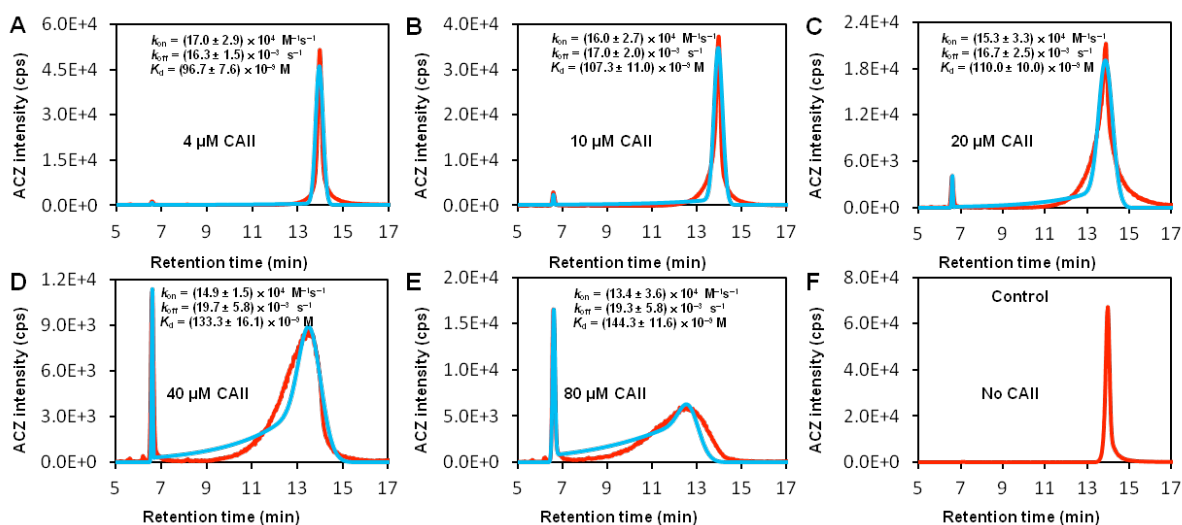
Figure 5.2. Experimental (red) and simulated (blue) ppKSEC-MS chromatograms for kinetic analysis of reversible binding between CAII and ACZ. The concentration of ACZ was 20 μM, and the concentration of CAII varied from 4 μM (A) to 80 μM (E). The control (F) corresponds to a run with a zero concentration of CAII. The ACZ signal was recorded in negative MRM mode for 221/83 (Q1/Q3) *m/z*. Simulated chromatograms were generated from modeling the processes involved in ppKSEC by using COMSOL multiphysics software. The binding parameters were determined from the best fit of the experimental chromatogram by the simulated one and were calculated on the basis of the averages and standard deviations of results obtained in triplicates.

In ppKSEC-MS, a short plug of SM is injected followed by injection of a considerably longer buffer spacer and finally a short plug of P. The injection times, τ, for SM and P are short and equal while the injection time, t_{spc}, for the spacer is much longer, $t_{spc} \gg \tau$, thus eliminating the possibility of mixing between SM and P prior to the start of separation. Since P cannot enter the pores, equation 5-1 can only proceed outside the beads in the so-called free volume. In addition, a hydrodynamic flow of the solution exists only outside the beads. Therefore, we assume that the buffer velocity as well as [P] and [P–SM] are averaged across the column over the area lying outside the beads. Moreover, [SM] outside the beads and inside the beads are averaged across the

89

column over the total area lying outside the beads and inside the pores. Interactions between the species and their mass transfer are described by the following equations:

$$(\partial_t + v_{SM}\partial_x - D_{SM}\partial_x^2)[SM] = \alpha(k_{off}[P\text{-}SM] - k_{on}[SM][P]) \qquad (5\text{-}2)$$

$$(\partial_t + v\partial_x - D_P\partial_x^2)[P] = k_{off}[P\text{-}SM] - k_{on}[SM][P] \qquad (5\text{-}3)$$

$$(\partial_t + v\partial_x - D_P\partial_x^2)[P\text{-}SM] = k_{on}[SM][P] - k_{off}[P\text{-}SM] \qquad (5\text{-}4)$$

$$v_{SM} = \alpha v, \quad \alpha = \frac{\phi_{out}}{\phi_{out} + \phi_{in}}, \quad D_{SM} = \frac{\phi_{out}D_{out} + \phi_{in}D_{in}}{\phi_{out} + \phi_{in}} + \frac{\phi_{out}^2\phi_{in}^2 v^2}{k\left(\phi_{out} + \phi_{in}\right)^3} \qquad (5\text{-}5)$$

Here, v is the average velocity of the hydrodynamic flow in the column; v_{SM} is the average velocity of SM in the column; D_{out} and D_{in} are diffusion coefficients of SM outside the beads and inside their pores, respectively; D_P is the diffusion coefficient of P and P–SM (we consider it to be the same as SM binding P does not significantly affect the molecular size of P); ϕ_{out} and ϕ_{in} are relative volumes (i.e., fractions of the column volume) located outside beads and inside pores, respectively; $k \sim D_{in}/R_{in}^2$ is the kinetic rate constant for a diffusion relaxation between concentrations of small molecules outside the beads and inside their pores; R_{in} is the characteristic size of beads. The average concentrations of SM outside the beads, $[SM]_{out}$, and inside the pores, $[SM]_{in}$, can be considered to be similar due to fast diffusion equilibration between the pores and the outside-the-beads volume ($[SM]_{out} = [SM]_{in} = [SM]$). Indeed, for a characteristic time, t_{in}, of the diffusion relaxation between SM outside the beads and inside their pores, we have $t_{in} \sim R_{in}^2/D_{in} \sim 0.01$ s for typical values of $R_{in} \sim 3$ μm and $D_{in} \sim 10^{-5}$ cm^2/s. Thus, $t_{in} \ll t_{sep} = W/(v - v_{SM})$, where t_{sep} is the separation time which is usually in the order of a few seconds (W is the plug length). It should be noted that a coefficient α depends only on the ratio ϕ_{out}/ϕ_{in} that coincides with the ratio of actual (not relative) volumes located outside beads and inside pores.

In ppKSEC-MS, nonequilibrium boundary conditions at $x = 0$ were used. Such boundary

conditions for equation 5-2 to equation 5-5 can be formulated as follows:

$$[SM] = [SM]_0 \quad (x = 0, \quad 0 < t < \tau)$$
$$[P] = [P]_0 \quad (x = 0, \quad t_{spc} < t < t_{spc} + \tau) \tag{5-6}$$

where $[SM]_0$ and $[P]_0$ are initial concentrations of SM and P injected in the column inlet, τ is the

injection time of SM and P, and t_{spc} is the time interval between injections of SM and P.

Concentrations at $x = 0$ are assumed to be zero for other time intervals. Equation 5-2 to equation

5-6were used to obtain a numerical solution of the problem and to simulate signal $S(t)$ generated

by SM. We assume that all intact P–SM that reaches the end of the column dissociates during

ionization and SM released from this dissociation is detected. As a result, $S(t)$ is proportional to

the total concentration of SM (both unbound and bound to P) at the column exit and g is a

proportionality coefficient:

$$S(t) = g\left([SM](t) + [P\text{-}SM](t)\right) \tag{5-7}$$

The described model was implemented in COMSOL Multiphysics 4.3a (commercial software).

The Transport of Diluted Species module of COMSOL was used in simulations of equation 5-2

to equation 5-6. The program generated simulated ppKSEC-MS chromatograms, $S(t)$. Nonlinear

regression was used to find best fits of the experimental ppKSEC-MS chromatograms (red

traces) by the simulated ones (blue traces) while varying the values of k_{on} and k_{off} (**Figure 5-2**). It

should be noted that parameters v_{SM}, D_{SM}, and g can be determined by fitting the experimental

chromatogram obtained for injecting SM alone (i.e., in the absence of the protein). Similarly,

parameters v and D_P can be found by fitting the experimental chromatogram obtained for

injecting P without SM. Provided that v_{SM} and v are determined, parameter α can be calculated

using the first relation in equation 5-5. As a result, only parameters k_{on} and k_{off} have to be varied

in the fitting procedure involving experimental ppKSEC-MS data obtained for injecting both SM and P.

We varied the concentration of the protein to test if the solutions for k_{on} and k_{off} were stable. When the protein concentration increased 20-fold, the values of k_{on} and k_{off} remained stable: $k_{on} = (15.4 \pm 2.2) \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ and $k_{off} = (17.8 \pm 2.0) \times 10^{-3} \text{ s}^{-1}$ (rules of error propagation were used to find the errors of k_{on} and k_{off}). There was a noticeable trend of monotonic increase in k_{on} and less monotonic increase in k_{off}. This trend indicates that there is a small systematic error in the calculations. The error is most likely due to some minor phenomena in the separation and/or molecular interactions that are not taken into account by the simple mathematical model used to fit the experimental chromatograms. The value of the equilibrium dissociation constant was calculated as k_{off}/k_{on}: $K_d = (117 \pm 16) \times 10^{-9}$ M. To validate our results, we used another solution-based label-free method, isothermal titration calorimetry (ITC). ITC can only determine K_d, but we used this validation method since there is no other label-free kinetic method available for such a validation. ITC experiments revealed $K_d = (76 \pm 5) \times 10^{-9}$ M (**Figure 5.3**). K_d values obtained with ppKSEC-MS (~120 nM) and ITC (~80 nM) are in reasonable agreement considering that the temperatures in the ppKSEC-MS and ITC are difficult to make equal and conceptually different methods can lead to up to several-fold differences in measured equilibrium constants [124]. This agreement indicates that K_d calculated as k_{off}/k_{on} for k_{on} and k_{off} obtained with ppKSEC-MS is correct. Even though there is still a possibility that k_{on} and k_{off} were determined with a similar systematic error, which was canceled upon division of k_{off} over k_{on}, such an error appears to be extremely unlikely. Therefore, we can conclude that ppKSEC-MS correctly determined k_{on} and k_{off} for reversible binding of CAII and ACZ.
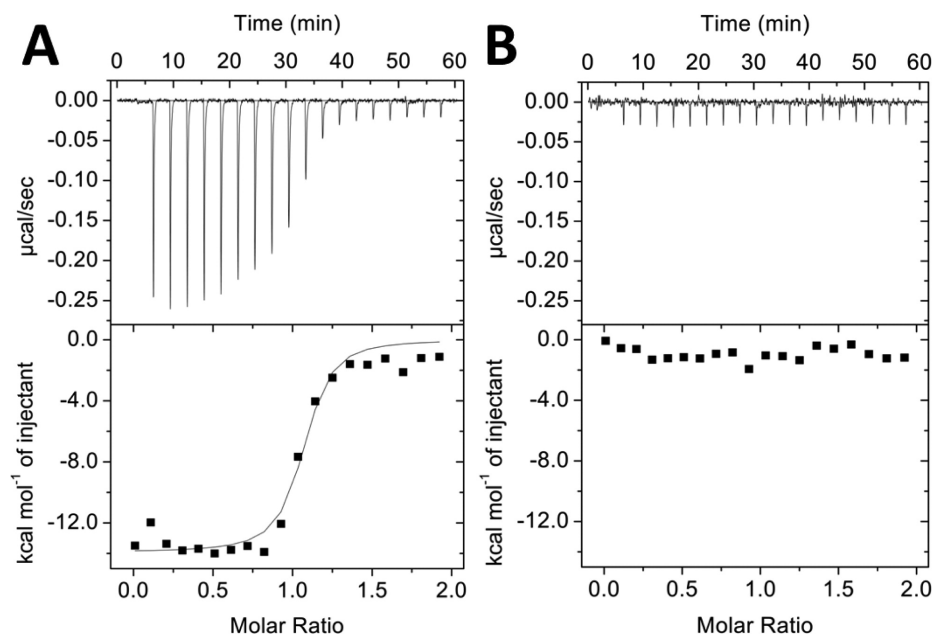
Figure 5.3. Thermograms of binding analysis between CAII and ACZ by ITC. Panel (A) corresponds to titrating CAII with ACZ while panel (B) corresponds to titrating CAII with buffer only. The upper graphs show the raw ITC data in real time, and the lower graphs show the corresponding integrated total heat per injection with respect to molar ratio. The fitting was generated by using Origin software with a single-binding-site model.

5.4. Conclusions

In conclusion, we outline major features of ppKSEC-MS in application to kinetic studies of protein–small molecule interactions. The method relies on generic separates of SM and P–SM by SEC without immobilization and generic quantitative detection of SM by MS without labeling. Any pair of P and SM can be separated by SEC assuming that neither of the molecules adsorbs on the beads material. MS, in turn, can detect any small molecule assuming that suitable ionization conditions are found and the major ion products are known. Advantageously, ppKSEC-MS requires no detection of intact P–SM, which is a very challenging task [125]. Moreover, data processing becomes simpler if the intact P–SM completely dissociates during ionization, which is easy to achieve in APCI [126]. In addition, ppKSEC-MS is a kinetic method that does not require equilibrium to be reached in equation 5-1. As a result, $[P]_0$ and $[SM]_0$ that

significantly differ from K_d [30] can be used, thus relaxing the requirements for the limit of detection in MS [29, 127]. In ppKSEC-MS, SM and P are injected separately and reacted inside the column, thus minimizing sample consumption and making the process easily suitable for automation without the use of sophisticated liquid handlers.

Overall, our results suggest that ppKSEC-MS has a potential to become a generic solution-based label-free platform for kinetic studies of protein–small molecule interactions, but more detailed studies will be needed to understand the advantages and limitations of the method better. In general, a robust analytical platform should be applicable to various samples with different conditions. Therefore, in the next chapter, we will expand the team of KSEC and introduce another type of KSEC-MS analysis. The new method utilizes pre-equilibrated target-ligand binding mixture as the sample of injection, to verify the versatility of KSEC-MS as an analytical platform for secondary ligand screening.

CHAPTER 6. PRE-EQUILIBRATION KINETIC SIZE-EXCLUSION CHROMATOGRAPHY WITH MASS SPECTROMETRY DETECTION (PEKSEC-MS) FOR LABEL-FREE SOLUTION-BASED KINETIC ANALYSIS OF PROTEIN–SMALL MOLECULE INTERACTIONS

The presented material was published previously and reprinted with permission from "Bao, J., Krylova, S. M., Cherney, L. T., Le Blanc, J. C. Y., Pribil, P., Johnson, P. E., Wilson, D. J., & Krylov, S. N. (2015). Pre-equilibration kinetic size-exclusion chromatography with mass spectrometry detection (peKSEC-mS) for label-free solution-based kinetic analysis of protein–small molecule interactions. *The Analyst*, *140*(4), 990–994. doi:10.1039/c4an02232g" Reproduced by permission of The Royal Society of Chemistry. My contribution to the article was: (i) planning all experiments, (ii) synthesized the protein, (iii) performing all experiments, (iv) preparing figures, (v) writing the manuscript.

6.1 Introduction

In the previous study, we have successfully developed the first method of KSEC-MS, in which the small molecule and protein were sequentially injected, followed by the interactions during separation. Plug-plug KSEC-MS is the first generic label-free solution-based kinetic method for analyzing protein-small molecule interactions. However, for secondary ligand screening, the method of choice has to be versatile in analyzing various samples under different

conditions. Therefore, in this study, we will further develop the idea of KSEC and introduce the pre-equilibration kinetic size-exclusion chromatography mass spectrometry (peKSEC-MS).

Recently we have introduced a solution-based, label-free approach for kinetic analysis of non-covalent protein–small molecule interactions called kinetic size-exclusion chromatography mass spectrometry (KSEC-MS) [42]. In a schematic sense, non-covalent protein-small molecule interactions can be illustrated in equation 6-1, which involves a protein (P), a small molecule (SM), and a protein-small molecule complex (P-SM):

$$P + SM \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} P\text{-}SM \qquad\qquad (6\text{-}1)$$

k_{off} and k_{on} are the kinetic rate constants of dissociation and association respectively. The equilibrium dissociation constant can be calculated through $K_d = k_{off}/k_{on}$ (smaller K_d value indicates high affinity binding). In KSEC-MS, generic solution-based kinetic separation is realized in a size-exclusion chromatography (SEC) column; label-free detection of SM is done using tandem mass spectrometry (MS/MS). The shape of the resulting chromatogram, signal (proportional to SM concentration) versus time, is defined by k_{on} and k_{off}. The values of k_{on} and k_{off} can be determined by finding a suitable mathematical model and fitting the experimental chromatogram with simulated ones while varying k_{on} and k_{off}. The best fit reveals the appropriate values of k_{on} and k_{off}. Plug–plug kinetic size-exclusion chromatography mass spectrometry (ppKSEC-MS) was our first practical implementation of the KSEC-MS concept. In essence, short plugs of SM and P are separately injected into the column; SM is followed by P. In a SEC column, P moves faster, and during the plug of P passing through the plug of SM, the binding reaction occurs and the P–SM complex is formed. When P outruns SM, the continuous dissociation of the complex starts. We developed a 1-dimensional numerical model for

96

simulating a ppKSEC-MS chromatogram, and used it to find k_{on} and k_{off} for the interaction of carbonic anhydrase and its inhibitor, acetazolamide.

To further develop the idea of label-free solution-based kinetic measurements, we now introduce the next KSEC-MS method, pre-equilibration kinetic size-exclusion chromatography (peKSEC). The concept of peKSEC is depicted in **Figure 6.1.** An "equilibrium mixture" (EM) is prepared by incubating P with SM to approach the equilibrium shown in equation 6-1. A small volume of the EM (much smaller than free volume of the column) is injected into a SEC column at time t_0 and its components are separated based on their size differences. As soon as SM is separated from P–SM, the latter is no longer at equilibrium and starts dissociating releasing more unbound SM. The dissociation process continues leaving a "tail" of SM. The samples in the column will eventually elute in the following order: (i) the intact P–SM, (ii) the "tail" of SM that dissociated from the complex during separation, and (iii) the unbound SM in the EM. Upon leaving the column the small molecule can be ionized by various ionization methods, such as electrospray ionization (ESI) or atmospheric-pressure chemical ionization (APCI), and detected [128]. During ionization, the intact P–SM is deliberately destroyed, thus SM from the complex is released and also detected by MS/MS. In general, a peKSEC-MS chromatogram contains 3 features: (i) a peak that corresponds to SM that exited the column as a part of the intact P–SM, (ii) a peak of SM that was unbound in EM, and (iii) a bridge between the two peaks that corresponds to SM that was bound to P but dissociated during separation. The shapes and areas of these three features are defined by k_{on} and k_{off}. Accordingly, fitting the chromatogram with a 1-dimensional mathematical model that describes reaction 1 along with mass transfer in the chromatographic column reveals both rate constants k_{on} and k_{off}.
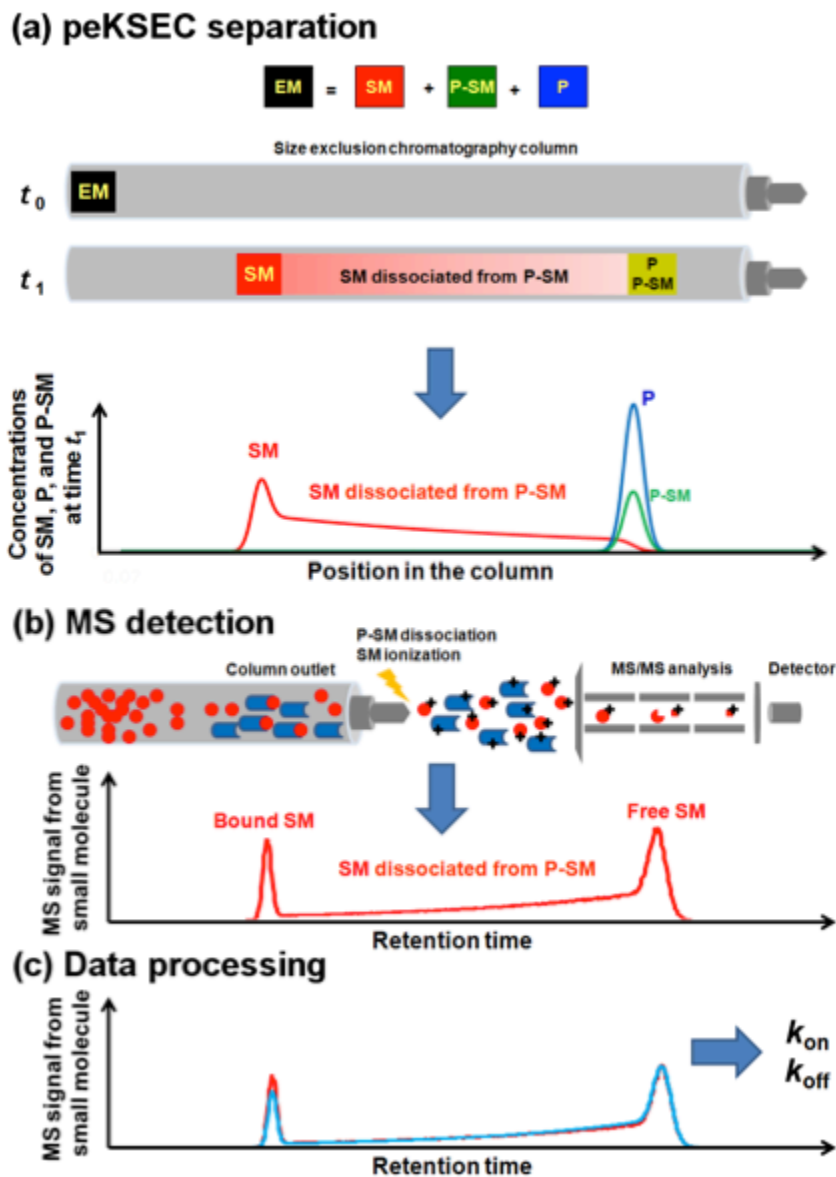
Figure 6.1. Conceptual depiction of separation (a), detection (b), and data processing (c) in peKSEC. Please see text for details.

6.2 Materials and Methods

6.2.1 Chemicals and Materials

Methotrexate (MTX) was purchased from Sigma-Aldrich (Oakville, ON, Canada). The *E. coli* Rossetta-gami™ 2(DE3) competent cells was purchased from EMD Millipore (PA, USA). Sepharose Fast Flow column, Phenyl Sepharose 6 Fast Flow column and Sephacryl S-100 size exclusion column were purchased from GE Healthcare (Toronto, ON, Canada). The 30,000 MW cutoff Amicon® Ultra 15 mL centrifugal filter devices were purchased from EMD Millipore (Nepan, ON, Canada). All other reagents were purchased from Sigma-Aldrich (Oakville, ON, Canada). All solutions were made using deionised water filtered through a 0.22 μm filter (Millipore, Nepan, ON, Canada).

6.2.2 Protein expression and purification.

The plasmid pFW117.1 that contains folA (*E. coli* DHFR) gene was transformed into the competent *E. coli* Rossetta-gami™ 2 (DE3) cells. 1L bacterial culture was growing at 37°C, 260 rpm shaking until the OD_{600} reached 0.8. The culture was then induced with 1.0 mM IPTG and continues growing at 20°C, with 260 rpm shaking overnight. After induction, the cells were collected by centrifugation at 5500 rpm for 1 hour at 4°C. The cell pellet was resuspended in 20 mM Tris-HCl pH 6.2, 0.5 mM EDTA, 0.5 mM DTT , 0.02 % N_aN_3 w/v (buffer A), and then sonicated on ice with 15 s on and 55 s off intervals for 4 min at 60 % amplitude. The lysate was centrifuged at 5500 rpm for 1 hour at 4°C and the supernatant containing DHFR was loaded onto the Q Sepharose Fast Flow column previously equilibrated with buffer A. The DHFR elution was achieved by using a NaCl gradient from 0 to 1.0 M in buffer A, and DHFR eluted at

concentrations of 0.4 M NaCl. The fractions containing DHFR were concentrated and dialyzed

against 50 mM sodium phosphate pH 7.0 and 1.0 M $(NH_4)_2SO_4$, 0.02 % N_aN_3 w/v (buffer B).

The DHFR containing solution was then loaded onto the Phenyl Sepharose 6 Fast Flow column,

which was pre-equilibrated with buffer B. The DHFR elution was achieved using 1.0 to 0 M

$(NH_4)_2SO_4$ gradient in buffer B, and DHFR eluted at 0.5 M $(NH_4)_2SO_4$. The fractions containing

DHFR were concentrated and dialyzed with 100 mM sodium phosphate pH 7.0, 150 mM NaCl,

0.5 mM EDTA, 0.02 % N_aN_3 w/v (buffer C), and then loaded onto a Sephacryl S-100 size

exclusion column. All buffer flow rates were 1 mL/min on an ÄKTA-FPLC system. The purified

DHFR was dialyzed against 30 mM ammonium formate pH 7.2 (buffer D)

6.2.3 Instrumentation

The Allegra 21R centrifuge with S4180 rotor was purchased from Beckman Coulter (ON,

Canada). ÄKTA-FPLC system was purchased from GE Healthcare (Toronto, ON, Canada). The

Shimadzu UFLCXR with Agilent Bio SEC-3 was used for all experiments. Bio SEC-3 Size

exclusion chromatography (SEC) column was purchased from Agilent (Mississauga, ON,

Canada). The column has 3 μm particle size, 150 Å pore size, 4.6 mm inner diameter and 300

mm length. The AB Sciex QTRAP 6500 with IonDrive Turbo V Source (Concord, ON, Canada)

was used for small molecule detection and quantification. Isothermal titration calorimetry (ITC)

experiments were performed by using a MicroCal iTC200 system (Northampton, MA, USA).

6.2.4 Pre-equilibration KSEC with MS detection.

Buffer D was used for all sample preparation and separation. The pre-equilibration binding

mixture was made by incubating 20 nM MTX with different concentrations of DHFR (20 nM –

80 nM) for 10 min at 20°C. 10 µl of binding mixture was injected in to HPLC and separation

was conducted at a flow rate of 0.3 mL/min 20°C. Atmospheric-pressure chemical ionization

(APCI) was used to ionize MTX with positive ion mode. The source temperature (TEM) was

300°C, the ionization energy (IS) was 5,500 V, and the de-clustering potential (DP) was 125 V.

The MRM mode was used to select the ion of 445.2/308.2 (Q1/Q3) by using the collision energy

(CE) at of 28 V. All binding experiments were done in triplicates. Fitting the experimental

peKSEC-MS chromatograms with the simulated ones was carried out by using COMSOL

Multiphysics 4.3a commercial software (COMSOL Group, Palo Alto, CA).

6.2.5 Isothermal titration calorimetry analysis.

Buffer D was used for all experiments. Binding experiments were conducted using 10 µM

DHFR and 100 µM MTX at 25°C. The experimental setup consisted of 19 successive 2 µL

injections of either MTX or buffers into DHFR every 180 s to a final molar ratio of 1:2. The first

injection was 0.2 µL for all experiments. The data were corrected for the heat of dilution of the

titrant. Data analysis was carried out with Origin 5.0 software.

6.3 Results and Discussion

The great strength of peKSEC-MS is that it relies on a generic separation as a small

molecule can always be separated from a large P–SM complex. Moreover, peKSEC-MS uses a

generic detection scheme, as practically any small molecule can be selectively detected by

MS/MS. Here, we chose the interaction between dihydrofolate reductase (DHFR) and

methotrexate (MTX) as a model system. DHFR is an essential enzyme in cell proliferation and

cell growth; it converts dihydrofolic acid to tetrahydrofolic acid, and MTX is its well-known

inhibitor [23, 129, 130]. DHFR and MTX can be easily separated by SEC based on the difference in their sizes. The MS/MS signal intensity of MTX was proportional to the concentration in a range of 10 μM–100 pM; the linear response in the nanomolar range is essential for studying high affinity binding (nanomolar K_d). When EM of DHFR and MTX was sampled for detection of MTX, only a predicted peKSEC-MS chromatogram was obtained in **Figure 6.2**. The signal intensity of the leftmost peak, which corresponds to P–SM increased with increasing protein concentration [P] in EM. A similar trend was also observed for the bridge region, which corresponds to SM that dissociated from P–SM during separation. Meanwhile, as we anticipated, the rightmost peak, free SM, decreased with increasing [P] in EM. The integral of the SM signal over the entire chromatogram remained constant with changing [P] (and the concentration of small molecule [SM] remained constant) suggesting that the intact complex was completely dissociated and all SM was accounted for.

Deconvolution of the kinetic rate constants from a peKSEC-MS chromatogram is not a trivial task. While no analytical solutions are available, we have adapted the numerical approach previously developed for modeling ppKSEC-MS to model processes in peKSEC-MS. It is a 1-dimensional model that considers complex dissociation and complex re-formation during migration of the components through the column. The following setup is used for the 1-dimensional approach in peKSEC-MS. A long and narrow cylindrical chromatography column is used, which is coaxial with the x coordinate. It is filled with beads that constitute the stationary phase. The beads have pores, which are, in the first approximation, large enough for the SM to enter and reside inside for a significant time and too small for the P or the P–SM to be significantly retarded. This is confirmed by the significant difference in retention times between SM and P–SM. Also, the model uses an assumption of fast re-equilibration between the mobile

and stationary phase, which is confirmed by narrow peaks of P and SM. An EM plug containing

SM, P, and P–SM is injected into the column at $t = 0$ (t_0). We assume that the buffer velocity and

concentrations of components P and P–SM are averaged across the column over the area lying

outside the beads. Similarly, the concentrations of SM outside the beads and inside them are

averaged across the column over the area lying outside the beads and inside the pores,

respectively. Mass transfers of SM, P, and P–SM are described by the following equations:

$$(\partial_t + v_A \partial_x - D_A \partial_x^2)[\text{SM}] = \alpha(k_{\text{off}}[\text{P-SM}] - k_{\text{on}}[\text{SM}][\text{P}]) \qquad (6\text{-}2)$$

$$(\partial_t + v \partial_x - D_P \partial_x^2)[\text{P}] = k_{\text{off}}[\text{P-SM}] - k_{\text{on}}[\text{SM}][\text{P}] \qquad (6\text{-}3)$$

$$(\partial_t + v \partial_x - D_P \partial_x^2)[\text{P-SM}] = k_{\text{on}}[\text{SM}][\text{P}] - k_{\text{off}}[\text{P-SM}] \qquad (6\text{-}4)$$

$$v_A = \alpha v, \quad \alpha = \frac{\phi_{\text{out}}}{\phi_{\text{out}} + \phi_{\text{in}}}, \quad D_A = \frac{\phi_{\text{out}} D_{\text{out}} + \phi_{\text{in}} D_{\text{in}}}{\phi_{\text{out}} + \phi_{\text{in}}} \qquad (6\text{-}5)$$

Here, [SM], [P], and [P–SM] are the concentrations of the small molecule, protein, and the

complex, respectively; v is the average velocity of the buffer; D_{out} and D_{in} are diffusion

coefficients of SM outside the beads and inside their pores; D_P is the diffusion coefficient of P

and P–SM (we consider them similar since SM does not significantly affect the size of P upon

binding); ϕ_{out} and ϕ_{in} are relative volumes (i.e. fractions of the column volume) located outside

beads and inside pores, respectively. Average concentrations of SM outside beads and inside

pores are considered to be approximately the same due to fast diffusion equilibration between

pores and outside the bead volume. Indeed, we usually have $t_{\text{in}} \sim R_{\text{in}}^2/D_{\text{in}}$ where t_{in} is the

characteristic time of diffusional relaxation between concentrations of small molecules outside

the beads and inside their pores and R_{in} is the characteristic size of the beads. The relationship t_{in}

$\sim R_{\text{in}}^2/D_{\text{in}}$ follows from the fact that Einstein's characteristic diffusion length $(D_{\text{in}}t)^{1/2}$ should be of

the order of the characteristic size of beads, R_{in}, if we calculate this length for the characteristic

time of diffusional relaxation $t \sim t_{in}$. At $R_{in} \sim 3$ μm and $D_{in} \sim 10^{-5}$ cm^2 s^{-1} calculations give $t_{in} \sim$ 0.01 s. Thus $t_{in} \ll t_{sep} = W/(v - v_A)$ where t_{sep} is the separation time which is usually on the order of a few seconds and W is the plug length. It should be noted that coefficient α depends only on the ratio ϕ_{out}/ϕ_{in} that coincides with the ratio of actual (not relative) volumes located outside beads and inside pores. We have also omitted an additional term proportional to t_{in} [42] in the last equation (6-5) for D_{SM} since we used the fitting procedure to determine D_{SM}.

To formulate initial conditions from equation 6-2 to equation 6-5 take into account that the injection usually satisfies the following conditions: (i) the mixture of SM, P, and P–SM is in equilibrium before the injection; (ii) $t_{inj} \ll t_{eq}$, where t_{inj} is the injection time and $t_{eq} = 1/(k_{on}[P]_0 + k_{off})$ is the equilibration time; and (iii) $t_{in} \ll t_{inj}$. In this case, the concentrations in the injected plug at $t = 0$ (i.e. immediately after injection) are determined by the following relations:

$$[\text{SM}]_0 = \alpha[\text{SM}]_{eq}, \quad [\text{P}]_0 = [\text{P}]_{eq}, \quad [\text{P-SM}]_0 = [\text{P-SM}]_{eq}$$
$$(0 \leq x \leq W, \quad t = 0) \tag{6-6}$$

$$W = \frac{V_{inj}}{\pi \phi_{out} R^2}, \quad \phi_{out} = \frac{V_{free}}{\pi R^2 L} \tag{6-7}$$

Here, $[\text{SM}]_{eq}$, $[\text{P}]_{eq}$, and $[\text{P-SM}]_{eq}$ are concentrations of SM, P, and P–SM in their equilibrium mixture before injection; W is the plug length after injection; V_{inj} is the volume of injected mixture; V_{free} is the free column volume measured by elution of the protein (in the absence of small molecules); R is the inner radius of the column; and L is the column length. Equation 6-2 to equation 6-7 were used to obtain a numerical solution of the problem and to simulate signal $S(t)$ generated by SM. We assume that the intact P–SM that reaches the end of the column dissociates in the mass-spectrometer and SM produced from this dissociation can be detected. As

a result $S(t)$ is proportional to the total concentration of SM (both unbound and bound to P) at the column exit, where g is a proportionality coefficient:

$$S(t) = g([SM](t) + [P\text{-}SM](t)) \hspace{3cm} (6\text{-}8)$$

The model was implemented using COMSOL multi-physics software (4.3a commercial software). The kinetic rate constants k_{on} and k_{off} were convoluted to form a simulated chromatogram. Non-linear regression was used to find the best fit of the experimental peKSEC-MS chromatogram with the simulated one. By fitting the experimental chromatograph to the 1-dimensional model we have calculated the kinetic rate constants: $k_{on} = (57.2 \pm 3.5) \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$, $k_{off} = (47.2 \pm 4.6) \times 10^{-4} \text{ s}^{-1}$, and $K_d = (8.2 \pm 1.3) \times 10^{-9}$ M through k_{off}/k_{on}. The best fits for different chromatograms (**Figure 6.2**) returned similar values of k_{on} and k_{off} suggesting that the solution is stable and also allowing us to estimate the method's precision. To validate our results, we chose another label free solution based method ITC (**Figure 6.3**). ITC is an equilibrium method from which only K_d can be calculated [131]; it was selected due to the lack of other label-free kinetic methods. ITC and peKSEC-MS are conceptually different methods, offering a means of validation with higher stringency. The ITC measured equilibrium dissociation constant $K_d = (10.2 \pm 0.8) \times 10^{-9}$ M, which agreed with peKSEC-MS measured results within acceptable errors, confirmed the accuracy of peKSEC-MS.
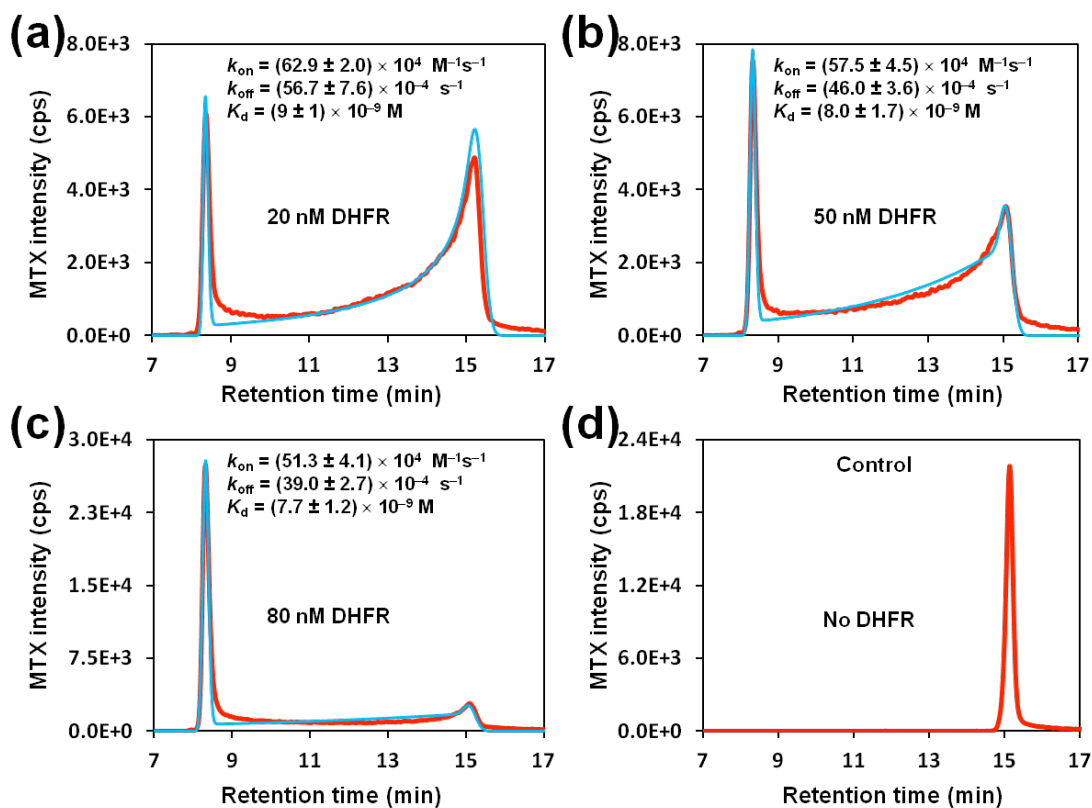
Figure 6.2. The peKSEC analysis of DHFR-MTX binding interactions. There are 20 nM MTX mixed with different concentrations of DHFR (a) to (c). The control (d) contains MTX only. The MTX signal was detected with MS/MS for 455.2/308.2 (Q1/Q3) m/z. Simulation chromatograms (blue) were generated from modeling the processes involved in peKSEC by using COMSOL multi-physics software. The kinetic rate constants were determined from the best fit of the experimental chromatogram (red) by the simulated one, and were calculated based on the average and standard deviation of triplicated results.
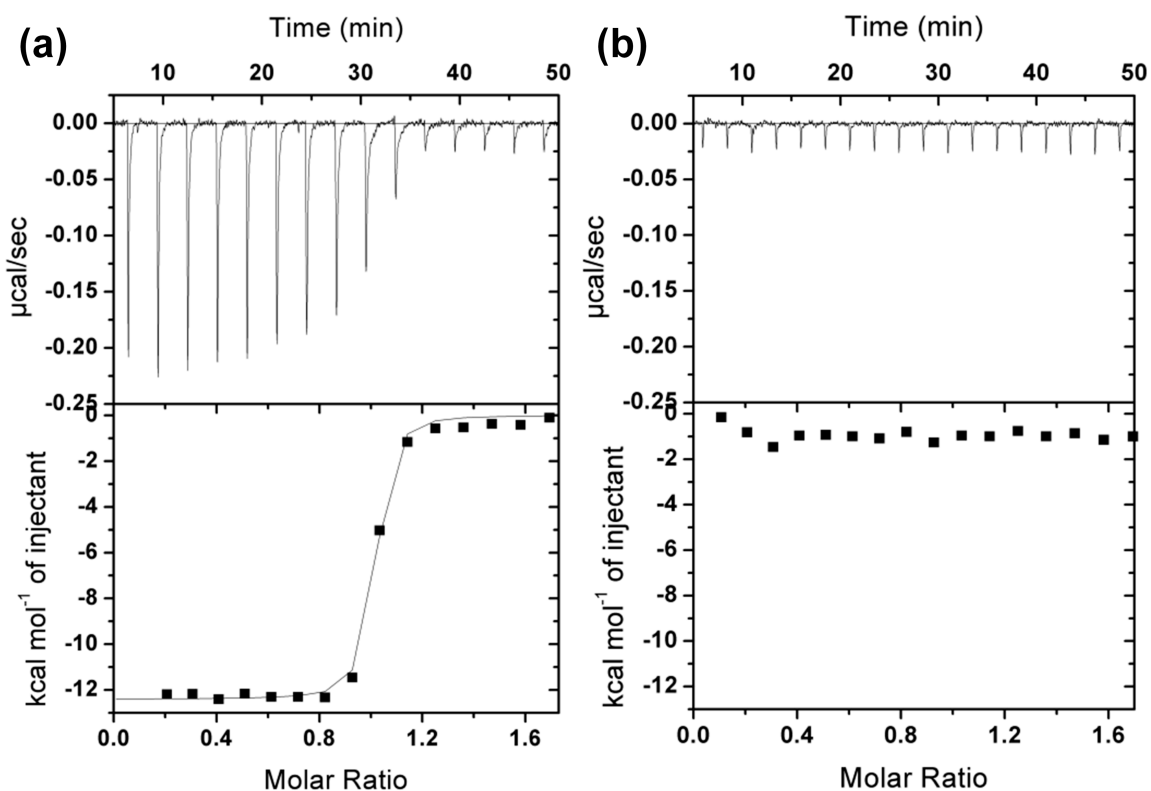
Figure 6.3. Thermograms of binding analysis between DHFR and MTX by ITC. (a): titrating DHFR with MTX and (b): titrating DHFR with buffer only. The upper graphs show the raw ITC data in real time and the lower graphs show the corresponding integrated total heat per injection with respect to molar ratio. The fitting was generated by using Origin 5.0 software with a single-binding-site model.

6.4 Conclusions

To conclude, we have outlined the main features of peKSEC-MS, which is a label-free solution-based method for studying the kinetics of reversible binding between a protein and a small molecule. In peKSEC, the migration pattern of the small molecule through a SEC column is followed by MS detection, and the kinetic parameters are extracted from the MS signal versus time dependence by means of numerical modelling. The numerical model uses two assumptions: (i) complex dissociation is complete during the ionization and (ii) the ionization efficiency of small molecules remains constant. In essence, we assume that the linear response of MS to small

molecule concentration is retained throughout the analysis. The requirement of complete dissociation is easily satisfied, as it is difficult to keep non-covalent complexes intact during the ionization process. The second assumption may not be always satisfied; for example, the ionization efficiency of a small molecule could be affected when it co-elutes with the protein [132]. Therefore, it is essential to confirm the method's validity by comparing the integrated small molecule signals among all tests – it should remain constant for constant small molecule concentration and not depend on the concentration of the protein. Moreover, measuring fast reactions with low K_d values will require low concentrations of interacting molecules and accordingly lower detection limit of MS [29, 30, 41]. The sub-nanomolar K_d measurements will require a mass spectrometer with a sub-nanomolar detection limit. Instrumentation for peKSEC-MS used in this study can measure small molecule concentration of 100 pM and the best contemporary MS instruments have limits of detection in the zeptamolar range [133]. Advantageously, peKSEC-MS does not require MS detection of an intact protein–small molecule complex, which may be very challenging. The ability of MS to rapidly scan through a wide mass range can potentially facilitate simultaneous analysis of one protein with several small molecules that potentially can be used for the rapid screening of panels of drug leads. The peKSEC method allows tight control of binding conditions such as incubation time and temperature, presence of cofactors, etc. Furthermore, as peKSEC relies on the established equilibrium prior to injection, we can now analyze the protein–small molecule interactions with slow association rates. Whereas in ppKSEC, the on-column incubation of a slow interacting pair is impractical, and it can also induce sample diffusion and peak broadening. We foresee that peKSEC-MS can become a generic solution-based label-free platform for kinetic studies of protein–small molecule interactions.

LIMITATIONS

We have developed kinetic solution-based ligands selection methods: KCE-DEL and KSEC-MS. The primary screening involves KCE-DEL based selection, which focuses on the preliminary selection of all target binders. The secondary screening involves KSEC-MS, which is a label-free solution-based approach that focuses on evaluating the true kinetic binding parameters and ranking the potential ligands.

For KCE-DEL based primary ligand screening, one of the most significant limitations is the contamination from synthetic intermediates. These synthetic intermediates have the very similar charge to size ratio and thus migrate very closely to the full ligand. Besides, these intermediates are all PCR amplifiable and thus can significantly affect the subsequent decoding procedure. Fortunately, the library of ligands can be purified before the primary screening. Although library purification leads to potential loss of ligands, the benefits from the elimination of contaminations far outweigh this disadvantage.

Moreover, in KCE-DEL approach, the DNA tag provides means of ligand identification and also contributes to the electrophoretic mobility. However, as we discussed previously, any tag or label affects binding; therefore, the selected DEL may not truly reflect the interaction between the target and the ligand (without DNA tag). For example, on a microscopic level, only a particular portion of a ligand molecule participates in interaction with the binding site of a target protein; thus, a single DEL-target interaction cannot provide a whole picture of binding interactions. To overcome this, it is required to use a diversified library for ligand screening for discovery of all possible interacting fragments from the DEL library, and mapping an entire landscape for ligand-target interactions. Furthermore, there are also DEL molecules with

multiple ligands attached to a single DNA strand, which can be used to probe various binding sites of a protein target.

For KSEC-MS based ligand selection, the major issue comes from the limit of detection (LOD) for ligands in MS. In general, high detection power allows searching for ligands with high target-binding affinity, which is often the most desirable property when selecting a potential drug molecule. Although each molecule has unique properties such as size and ionization efficiency that are contributing to the LOD, certain types of mass analyzers have higher detection power than others. Our current KSEC-MS methods were developed using triple quadruple mass analyzer, which has a LOD range about 10^{-10} to 10^{-11} M. This LOD range allows analyzing the target-ligand interactions with pM K_d. There are also more sensitive mass analyzers such as Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) with zeptomole detection limit.

Ideally, samples in KSEC should travel through the size-exclusion column without interacting with the solid phase. However, depending on the hydrophobicity, the small molecule can sometimes stick with certain types of column surfaces. Likewise, the non-specific surface adsorptions are also problematic in many other analytical techniques. Most target proteins, as well as protein-small molecule complexes, are typically not sticky in the size-exclusion columns. However since the profiles of small molecules are essential in determining kinetic binding parameters, it is crucial to find the appropriate column and optimize the separation conditions before KSEC-MS analysis.

CONCLUDING REMARKS

The main attention in current pharmaceutical industry focuses on finding new drugs. Discovery of novel drug ligands significantly relies on screening methods. In general, the process of drug discovery involves primary and secondary screening steps. The existing primary screenings are mainly surface-based and inherently suffer from non-specific surface adsorption. Therefore, we developed a KCE based DEL selection, to overcome this issue. During the development of KCE-DEL, we built a model to accurately predict migration times of binding complexes with various DNA constructs that allow us to collected binding complex and select target-binding ligand. We have successfully demonstrated the efficiency and accuracy by using two proteins, SA and (CAII), and two ligands with tag structures identical to those in actual GSK libraries. Moreover, our model can be potentially used for other types of electrophoresis-based protein-DNA binding analysis.

The secondary screening is conducted to investigate the binding parameters of target-ligand interactions, and to rank ligands based on affinity and kinetic parameters. The existing methods are either surface-based kinetic methods or solution-based affinity methods, and neither of them was able to calculate the true kinetic binding parameters by definition. We have developed the first label-free solution-based kinetic platform, KSEC-MS, for analyzing target-ligand interactions. Two types of methods, ppKSEC and peKSEC have been established and applied for assessing CAII-ACZ and DHFR-MTX binding interactions. They are also the first cases of calculating kinetic parameters k_{on} and k_{off} in a real label-free solution-based system. Furthermore, we were able to detect the binding affinity in the range of nM K_d.

To conclude, we have developed KCE-DEL and KSEC-MS for kinetic solution-based drug selections. We believe that in the very near future, our methods will emerge in the field of pharmaceutical discovery and facilitate as well as accelerate the therapies of many diseases.

FUTURE DIRECTIONS

UP until now, the proof of principle studies of KCE-DEL have been conducted and demonstrated as the solution-based approach for primary drug ligand screening. The next step will be developing KCE-DEL towards real ligand selection from the combinatorial libraries of DEL against therapeutic relevant protein targets. There are two potential problems that we may encounter when dealing with practical combinatorial libraries of DEL. First, the contaminations from synthetic intermediates will likely present in the library, which would interfere with the complex formations and subsequent PCR amplifications. Methods of CE are general separation approaches for DNA-based molecules and to solve such problem we can add a step of CE based DEL purification prior to the primary screening. Second, the binding conditions for many therapeutic relevant proteins require high salt conditions, which often trigger problems of high current and overheating during CE experiments. Therefore we need to find the proper combinations of salt concentrations, capillary dimensions, and the voltage supply, to support both target-ligand interactions and reliable CE separation. Also, we want to conduct KCE-based ligand selection in parallel with the conventional surface-based approach such as SPR, to compare the efficiencies and the systematic biases of both approaches, and also gain a in-depth understanding of these theoretically different analyses.

KSEC is a novel concept that developed for generic separation between protein targets and small molecule ligands. Since KSEC-MS utilizes standard LC-MS system, it is ready to be used for ligand ranking in the pharmaceutical industry. The kinetic parameter calculations at the current stage require modeling of target-ligand binding interactions inside of SEC column, which is cumbersome to perform and also very time-consuming. Therefore, the next step will be developing a fast and straightforward analytical solution for kinetic analysis of KSEC-MS data.

Furthermore, we have analyzed target-ligand interactions with 1:1 binding ratio, meaning a single target bind to a single ligand. In future, we want to explore stoichiometric binding such as a single target protein molecule binds to multiple ligands (of the same type). Also, we are interested in testing the simultaneous interactions of a single target against multiple types of ligands to study the synergistic effect of different drug ligands. Additionally, we plan to develop a ppKSEC based in situ enzymatic reaction analysis, in which the enzymatic reactions are performed and monitored in-column. Finally, we want to exploit other types of MS analyzers such as FT-ICR, to study high-affinity target-ligand interactions.

LIST OF PUBLICATIONS

1.   **Bao, J.**, Krylova, S. M., Wilson, D. J., Reinstein, O., Johnson, P. E., & Krylov, S. N. (2011). Kinetic capillary electrophoresis with Mass-Spectrometry detection (KCE-MS) facilitates label-free solution-based kinetic analysis of protein-small molecule binding. *ChemBioChem*, *12*(17), 2551–2554. doi:10.1002/cbic.201100617

2.   **Bao, J.**, Krylova, S. M., Reinstein, O., Johnson, P. E., & Krylov, S. N. (2011). Label-free solution-based kinetic study of Aptamer–Small molecule interactions by kinetic capillary electrophoresis with UV detection revealing how Kinetics control equilibrium. *Analytical Chemistry*, *83*(22), 8387–8390. doi:10.1021/ac2026699

3.   **Bao, J.**, & Krylov, S. N. (2012). Volatile kinetic capillary electrophoresis for studies of Protein–Small molecule interactions. *Analytical Chemistry*, *84*(16), 6944–6947. doi:10.1021/ac301829t

4.   **Bao, J.**, Krylova, S. M., Cherney, L. T., LeBlanc, J. C. Y., Pribil, P., Johnson, P. E., Wilson, D. K., & Krylov, S. N. (2014). Kinetic size-exclusion chromatography with mass Spectrometry detection: An approach for solution-based label-free kinetic analysis of Protein–Small molecule interactions. *Analytical Chemistry*, *86*(20), 10016–10020. doi:10.1021/ac503391c

5.   **Bao, J.**, Krylova, S. M., Cherney, L. T., Hale, R. L., Belyanskaya, S. L., Chiu, C. H., Arico-Muendel, C. C., & Krylov, S. N. (2015). Prediction of Protein–DNA complex mobility in gel-free capillary electrophoresis. *Analytical Chemistry*, *87*(4), 2474–2479. doi:10.1021/ac504504c

6.   **Bao, J.**, Krylova, S. M., Cherney, L. T., Le Blanc, J. C. Y., Pribil, P., Johnson, P. E., Wilson, D. J., & Krylov, S. N. (2015). Pre-equilibration kinetic size-exclusion chromatography with mass spectrometry detection (peKSEC-mS) for label-free solution-based kinetic analysis of protein–small molecule interactions. *The Analyst*, *140*(4), 990–994. doi:10.1039/c4an02232g

7.   **Bao, J.**, Krylova, S. M., Cherney, L. T., Hale, R. L., Belyanskaya, S. L., Chiu, C. H., Shaginian, A., Arico-Muendel, C. C., & Krylov, S. N. (2016). Predicting Electrophoretic mobility of Protein–Ligand complexes for ligands from DNA-Encoded libraries of small molecules. *Analytical Chemistry*, *88*(10), 5498–5506. doi:10.1021/acs.analchem.6b00980

REFERENCES

1.   Pina, A. S., Hussain, A., & Roque, A. C. A. (2010). *Ligand-macromolecular interactions in drug discovery: Methods and protocols (methods in molecular biology)*. New York, NY: Springer Verlag.

2.   Hauptman, H. A. (1990). History of x-ray crystallography. *Structural Chemistry*, *1*(6), 617–620. doi:10.1007/bf00674136

3.   Marion, D. (2013). An introduction to biological NMR spectroscopy. *Molecular & Cellular Proteomics*, *12*(11), 3006–3025. doi:10.1074/mcp.o113.030239

4.   Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*(172), . doi:10.3389/fmicb.2014.00172

5.   Debnath, M., Prasad, G. B. K. S., & Bisen, P. S. (2009). Omics technology. In *Molecular Diagnostics: Promises and Possibilities* (pp. 11–31). doi:10.1007/978-90-481-3261-4_2

6.   Ng, R. (2015). *Drugs: From discovery to approval*. Hoboken, NJ, United States: John Wiley & Sons.

7.   Hill, R. G., Rang, H. P., & Vallance, P. (2012). *Drug discovery and development: Technology in transition* (2nd ed.). Edinburgh: Elsevier Health Sciences.

8.   Bradley, D. (2005). Why big pharma needs to learn the three 'R's. *Nature Reviews Drug Discovery*, *4*(6), 446–446. doi:10.1038/nrd1766

9.   Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews Drug Discovery*, *5*(12), 993–996. doi:10.1038/nrd2199

10.   Imming, P., Sinning, C., & Meyer, A. (2006). Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery*, *5*(10), 821–834. doi:10.1038/nrd2132

11.   Trabocchi, A., & Schreiber, S. L. (2013). *Diversity-oriented synthesis: Basics and applications in organic synthesis, drug discovery, and chemical biology*. United States: John Wiley & Sons.

12.   Mannocci, L., Leimbacher, M., Wichert, M., Scheuermann, J., & Neri, D. (2011). 20 years of DNA-encoded chemical libraries. *Chemical Communications*, *47*(48), 12747. doi:10.1039/c1cc15634a

13.   Mannocci, L., Zhang, Y., Scheuermann, J., Leimbacher, M., De Bellis, G., Rizzi, E., Dumelin, C., Melkko, S., & Neri, D. (2008). High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proceedings of the National Academy of Sciences*, *105*(46), 17670–17675. doi:10.1073/pnas.0805130105

14.   Mullard, A. (2016). DNA tags help the hunt for drugs. *Nature*, *530*(7590), 367–369. doi:10.1038/530367a

15.   Guterman, L. (2011). Covalent drugs form long-lived ties, irreversible inhibitors may provide unique benefits in drug development. *Chemical & Engineering News*, pp. 19–26

16.   Firestone, A. J., & Chen, J. K. (2010). Controlling destiny through chemistry: Small-molecule regulators of cell fate. *ACS Chemical Biology*, *5*(1), 15–34. doi:10.1021/cb900249y

17.   Keserü, G. M., & Swinney, D. C. (Eds.). (2015). *Thermodynamics and Kinetics of drug binding, volume 65*. Germany: Wiley-Vch.

18.   Fodor, S., Read, J., Pirrung, M., Stryer, L., Lu, A., & Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, *251*(4995), 767–773. doi:10.1126/science.1990438

19.   Wilson, D. S., & Nock, S. (2002). Functional protein microarrays. *Current Opinion in Chemical Biology*, *6*(1), 81–85. doi:10.1016/s1367-5931(01)00281-2

20.   Ng, E. S. M., Chan, N. W., Lewis, D. F., Hindsgaul, O., & Schriemer, D. C. (2007). Frontal affinity chromatography—mass spectrometry. *Nature Protocols*, *2*(8), 1907–1917. doi:10.1038/nprot.2007.262

21.   Boozer, C., Kim, G., Cong, S., Guan, H., & Londergan, T. (2006). Looking towards label-free biomolecular interaction analysis in a high-throughput format: A review of new surface plasmon resonance technologies. *Current Opinion in Biotechnology*, *17*(4), 400–405. doi:10.1016/j.copbio.2006.06.012

22.   Wartchow, C. A., Podlaski, F., Li, S., Rowan, K., Zhang, X., Mark, D., & Huang, K.-S. (2011). Biosensor-based small molecule fragment screening with biolayer interferometry. *Journal of Computer-Aided Molecular Design*, *25*(7), 669–676. doi:10.1007/s10822-011-9439-8

23.   Batruch, I., Javasky, E., Brown, E. D., Organ, M. G., & Johnson, P. E. (2010). Thermodynamic and NMR analysis of inhibitor binding to dihydrofolate reductase. *Bioorganic & Medicinal Chemistry*, *18*(24), 8485–8492. doi:10.1016/j.bmc.2010.10.048

24.   Demarse, N. A., Killian, M. C., Hansen, L. D., & Quinn, C. F. (2013). Determining enzyme Kinetics via isothermal titration Calorimetry. *Methods in Molecular Biology*, *978*(3), 21–30. doi:10.1007/978-1-62703-293-3_2

25.   Hilton, G. R., & Benesch, J. L. P. (2012). Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *Journal of The Royal Society Interface*, *9*(70), 801–816. doi:10.1098/rsif.2011.0823

26.   Kitova, E. N., El-Hawiet, A., Schnier, P. D., & Klassen, J. S. (2012). Reliable determinations of Protein–Ligand interactions by direct ESI-MS measurements. Are we there yet? *Journal of The American Society for Mass Spectrometry*, *23*(3), 431–441. doi:10.1007/s13361-011-0311-9

27.   Kaltashov, I. A., & Eyles, S. J. (2005). *Mass Spectrometry in biophysics: Conformation and dynamics of Biomolecules*. New York, NY, United States: Wiley-Interscience.

28.   Petrov, A., Okhonin, V., Berezovski, M., & Krylov, S. N. (2005). Kinetic capillary electrophoresis (KCE): A conceptual platform for kinetic homogeneous affinity methods. *Journal of the American Chemical Society*, *127*(48), 17104–17110. doi:10.1021/ja056232l

29.   Bao, J., Krylova, S. M., Wilson, D. J., Reinstein, O., Johnson, P. E., & Krylov, S. N. (2011). Kinetic capillary electrophoresis with Mass-Spectrometry detection (KCE-MS) facilitates label-free solution-based kinetic analysis of protein-small molecule binding. *ChemBioChem*, *12*(17), 2551–2554. doi:10.1002/cbic.201100617

30.   Bao, J., Krylova, S. M., Reinstein, O., Johnson, P. E., & Krylov, S. N. (2011). Label-free solution-based kinetic study of Aptamer–Small molecule interactions by kinetic capillary electrophoresis with UV detection revealing how Kinetics control equilibrium. *Analytical Chemistry*, *83*(22), 8387–8390. doi:10.1021/ac2026699

31.   Mayr, L. M., & Bojanic, D. (2009). Novel trends in high-throughput screening. *Current Opinion in Pharmacology*, *9*(5), 580–588. doi:10.1016/j.coph.2009.08.004

32.   Oliphant, A. R., Brandl, C. J., & Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of yeast GCN4 protein. *Molecular and Cellular Biology*, *9*(7), 2944–2949. doi:10.1128/mcb.9.7.2944

33.   Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*(4968), 505–510. doi:10.1126/science.2200121

34.   Roberts, R. W., & Szostak, J. W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences*, *94*(23), 12297–12302. doi:10.1073/pnas.94.23.12297

35.   Clark, M. A., Acharya, R. A., Arico-Muendel, C. C., Belyanskaya, S. L., Benjamin, D. R., Carlson, N. R., Centrella, P. A., Chiu, C. H., Creaser, S. P., Cuozzo, J. W., Davie, C. P., Ding, Y., Franklin, G. J., Franzen, K. D., Gefter, M. L., Hale, S. P., Hansen, N. J., Israel, D. I., Jiang, J., Kavarana, M. J., Kelley, M. S., Kollmann, C. S., Li, F., Lind, K., Mataruse, S., Medeiros, P. F., Messer, J. A., Myers, P., O'Keefe, H., Oliff, M. C., Rise, C. E., Satz, A. L., Skinner, S. R., Svendsen, J. L., Tang, L., van Vloten, K., Wagner, R. W., Yao, G., Zhao, B., & Morgan, B. A. (2009). Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature Chemical Biology*, *5*(9), 647–654. doi:10.1038/nchembio.211

36.   Biroccio, A., Hamm, J., Incitti, I., De Francesco, R., & Tomei, L. (2002). Selection of RNA Aptamers that are specific and high-affinity ligands of the hepatitis C virus RNA-Dependent RNA polymerase. *Journal of Virology*, *76*(8), 3688–3696. doi:10.1128/jvi.76.8.3688-3696.2002

37.   Papoulas, O. (2001). Rapid separation of protein-bound DNA from free DNA using nitrocellulose filters. *Current Protocols in Molecular Biology*, *12*, Unit 12.8. doi:10.1002/0471142727.mb1208s36

38.   Wang, J., Rudzinski, J. F., Gong, Q., Soh, H. T., & Atzberger, P. J. (2012). Influence of target concentration and background binding on in vitro selection of affinity reagents. *PLoS ONE*, *7*(8), e43940. doi:10.1371/journal.pone.0043940

39.   Berezovski, M., Musheev, M., Drabovich, A., & Krylov, S. N. (2006). Non-SELEX selection of Aptamers. *Journal of the American Chemical Society*, *128*(5), 1410–1411. doi:10.1021/ja056943j

40.   Tok, J., Lai, J., Leung, T., & Li, S. F. Y. (2010). Selection of aptamers for signal transduction proteins by capillary electrophoresis. *Electrophoresis*, *31*(12), 2055–2062. doi:10.1002/elps.200900543

41.   Berezovski, M., & Krylov, S. N. (2002). Nonequilibrium capillary electrophoresis of equilibrium mixtures − A single experiment reveals equilibrium and kinetic parameters of Protein−DNA interactions. *Journal of the American Chemical Society*, *124*(46), 13674–13675. doi:10.1021/ja028212e

42.   Bao, J., Krylova, S. M., Cherney, L. T., LeBlanc, J. C. Y., Pribil, P., Johnson, P. E., Wilson, D. K., & Krylov, S. N. (2014). Kinetic size-exclusion chromatography with mass Spectrometry detection: An approach for solution-based label-free kinetic analysis of Protein–Small molecule interactions. *Analytical Chemistry*, *86*(20), 10016–10020. doi:10.1021/ac503391c

43.   Drabovich, A. P., Berezovski, M. V., Musheev, M. U., & Krylov, S. N. (2009). Selection of smart small-molecule ligands: The proof of principle. *Analytical Chemistry*, *81*(1), 490–494. doi:10.1021/ac8023813

44.   Javaherian, S., Musheev, M. U., Kanoatov, M., Berezovski, M. V., & Krylov, S. N. (2009). Selection of aptamers for a protein target in cell lysate and their application to protein purification. *Nucleic Acids Research*, *37*(8), e62–e62. doi:10.1093/nar/gkp176

45.   Hamula, C., Guthrie, J., Zhang, H., Li, X., & Le, X. (2006). Selection and analytical applications of aptamers. *TrAC Trends in Analytical Chemistry*, *25*(7), 681–691. doi:10.1016/j.trac.2006.05.007

46.   Drabovich, A. P., Okhonin, V., Berezovski, M., & Krylov, S. N. (2007). Smart Aptamers facilitate multi-probe affinity analysis of proteins with ultra-wide dynamic range of measured concentrations. *Journal of the American Chemical Society*, *129*(23), 7260–7261. doi:10.1021/ja072269p

47.   Berezovski, M. V., Musheev, M. U., Drabovich, A. P., Jitkova, J. V., & Krylov, S. N. (2006). Non-SELEX: Selection of aptamers without intermediate amplification of candidate oligonucleotides. *Nature Protocols*, *1*(3), 1359–1369. doi:10.1038/nprot.2006.200

48.   Musheev, M. U., Kanoatov, M., & Krylov, S. N. (2013). Non-uniform velocity of homogeneous DNA in a uniform electric field: Consequence of electric-field-induced slow dissociation of highly stable DNA–Counterion complexes. *Journal of the American Chemical Society*, *135*(21), 8041–8046. doi:10.1021/ja402257x

49.   Beinoravičiūtė-Kellner, R., Lipps, G., & Krauss, G. (2005). In vitro selection of DNA binding sites for ABF1 protein from Saccharomyces cerevisiae. *FEBS Letters*, *579*(20), 4535–4540. doi:10.1016/j.febslet.2005.07.009

50.   Kleiner, R. E., Dumelin, C. E., & Liu, D. R. (2011). Small-molecule discovery from DNA-encoded chemical libraries. *Chemical Society Reviews*, *40*(12), 5707–5717. doi:10.1039/c1cs15076f

51.   Kim, J., Bhinge, A. A., Morgan, X. C., & Iyer, V. R. (2004). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nature Methods*, *2*(1), 47–53. doi:10.1038/nmeth726

52.   Viovy, J. L. (2000). Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. *Reviews of Modern Physics*, *72*(3), 813–872. doi:10.1103/revmodphys.72.813

53.   Mayer, P., Slater, G. W., & Drouin, G. (1994). Theory of DNA Sequencing using free-solution electrophoresis of Protein-DNA complexes. *Analytical Chemistry*, *66*(10), 1777–1780. doi:10.1021/ac00082a029

54.   Hubert, S. J., & Slater, G. W. (1995). Theory of capillary electrophoretic separations of DNA-polymer complexes. *Electrophoresis*, *16*(1), 2137–2142. doi:10.1002/elps.11501601345

55.   Desruisseaux, C., Long, D., Drouin, G., & Slater, G. W. (2001). Electrophoresis of composite molecular objects. 1. Relation between friction, charge, and ionic strength in free solution. *Macromolecules*, *34*(1), 44–52. doi:10.1021/ma0002702

56.   Meagher, R. J., Won, J. I., McCormick, L. C., Nedelcu, S., Bertrand, M. M., Bertram, J. L., Drouin, G., Barron, A. E., & Slater, G. W. (2005). End-labeled free-solution electrophoresis of DNA. *Electrophoresis*, *26*(2), 331–350. doi:10.1002/elps.200410219

57.   Meagher, R. J., McCormick, L. C., Haynes, R. D., Won, J.-I., Lin, J. S., Slater, G. W., & Barron, A. E. (2006). Free-solution electrophoresis of DNA modified with drag-tags at both ends. *Electrophoresis*, *27*(9), 1702–1712. doi:10.1002/elps.200500554

58.   Long, D., Dobrynin, A. V., Rubinstein, M., & Ajdari, A. (1998). Electrophoresis of polyampholytes. *The Journal of Chemical Physics*, *108*(3), 1234–1244. doi:10.1063/1.475485

59.   Teraoka, I. A. (2002). *Polymer solutions: An introduction to physical properties*. New York: Wiley, John & Sons.

60.   Strobl, G. R. (2006). *The physics of polymers: Concepts for understanding their structures and behavior* (3rd ed.). Berlin: Springer-Verlag Berlin and Heidelberg GmbH & Co. K.

61.   Simmel, F. C., & Dittmer, W. U. (2005). DNA Nanodevices. *Small*, *1*(3), 284–299. doi:10.1002/smll.200400111

62.   Krishnan, Y., & Simmel, F. C. (2011). Nucleic acid based molecular devices. *Angewandte Chemie International Edition*, *50*(14), 3124–3156. doi:10.1002/anie.200907223

63.   Manning, G. S. (1969). Limiting laws and Counterion condensation in Polyelectrolyte solutions I. Colligative properties. *The Journal of Chemical Physics*, *51*(3), 924–933. doi:10.1063/1.1672157

64.   Manning, G. S. (1981). Limiting laws and counterion condensation in polyelectrolyte solutions. 7. Electrophoretic mobility and conductance. *The Journal of Physical Chemistry*, *85*(11), 1506–1515. doi:10.1021/j150611a011

65.   Barrat, J.-L., & Joanny, J.-F. (1996). Theory of Polyelectrolyte Solutions. *Advances in Chemical Physics*, *94*, 1–66.

66.   Anik, N., Airiau, M., Labeau, M.-P., Vuong, C.-T., Reboul, J., Lacroix-Desmazes, P., Gerardin, C., & Cottet, H. (2009). Determination of polymer effective charge by indirect UV detection in capillary electrophoresis: Toward the characterization of macromolecular architectures. *Macromolecules*, *42*(7), 2767–2774. doi:10.1021/ma8025095

67.   Ibrahim, A., Koval, D., Kašička, V., Faye, C., & Cottet, H. (2013). Effective charge determination of Dendrigraft Poly- l -lysine by capillary Isotachophoresis. *Macromolecules*, *46*(2), 533–540. doi:10.1021/ma302125f

68.   CalcTool: Protein size calculator. (2008b). Retrieved from http://www.calctool.org/CALC/prof/bio/protein_size

69.   Schneider, B., Patel, K., & Berman, H. M. (1998). Hydration of the phosphate group in Double-Helical DNA. *Biophysical Journal*, *75*(5), 2422–2434. doi:10.1016/s0006-3495(98)77686-6

70.   Hendrickson, W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A., & Phizackerley, R. P. (1989). Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proceedings of the National Academy of Sciences*, *86*(7), 2190–2194. doi:10.1073/pnas.86.7.2190

71.   Long, D., & Ajdari, A. (1996). Electrophoretic mobility of composite objects in free solution: Application to DNA separation. *Electrophoresis*, *17*(6), 1161–1166. doi:10.1002/elps.1150170628

72.   Happel, J., & Brenner, H. (1983). *Low Reynolds number hydrodynamics. With special applications to particulate media*. Leiden: Martinus Nijhoff Publishers, The Hague.

73.   Hughes, J., Rees, S., Kalindjian, S., & Philpott, K. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–1249. doi:10.1111/j.1476-5381.2010.01127.x

74.   Jorgensen, W. L. (2009). Efficient drug lead discovery and optimization. *Accounts of Chemical Research*, *42*(6), 724–733. doi:10.1021/ar800236t

75.   Keserű, G. M., & Makara, G. M. (2006). Hit discovery and hit-to-lead approaches. *Drug Discovery Today*, *11*(15-16), 741–748. doi:10.1016/j.drudis.2006.06.016

76.   Brenner, S., & Lerner, R. A. (1992). Encoded combinatorial chemistry. *Proceedings of the National Academy of Sciences*, *89*(12), 5381–5383. doi:10.1073/pnas.89.12.5381

77.   Melkko, S., Scheuermann, J., Dumelin, C. E., & Neri, D. (2004). Encoded self-assembling chemical libraries. *Nature Biotechnology*, *22*(5), 568–574. doi:10.1038/nbt961

78.   Pfaffl, M. W., & Hageleit, M. (2001). *Biotechnology Letters*, *23*(4), 275–282. doi:10.1023/a:1005658330108

79.   Franzini, R. M., Neri, D., & Scheuermann, J. (2014). DNA-Encoded chemical libraries: Advancing beyond conventional small-molecule libraries. *Accounts of Chemical Research*, *47*(4), 1247–1255. doi:10.1021/ar400284t

80.   Li, X., & Liu, D. R. (2004). DNA-Templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angewandte Chemie International Edition*, *43*(37), 4848–4870. doi:10.1002/anie.200400656

81.   de Jong, S., & Krylov, S. N. (2011). Protein labeling enhances Aptamer selection by methods of kinetic capillary electrophoresis. *Analytical Chemistry*, *83*(16), 6330–6335. doi:10.1021/ac201242r

82.   He, X., Ding, Y., Li, D., & Lin, B. (2004). Recent advances in the study of biomolecular interactions by capillary electrophoresis. *Electrophoresis*, *25*(45), 697–711. doi:10.1002/elps.200305727

83.   Fenteany, G., & Zhu, S. (2003). Small-molecule inhibitors of Actin dynamics and cell Motility. *Current Topics in Medicinal Chemistry*, *3*(6), 593–616. doi:10.2174/1568026033452348

84.   Backes, A., Zech, B., Felber, B., Klebl, B., & Müller, G. (2008). Small-molecule inhibitors binding to protein kinases. Part I: Exceptions from the traditional pharmacophore approach of type I inhibition. *Expert Opinion on Drug Discovery*, *3*(12), 1409–1425. doi:10.1517/17460440802579975

85.   Swinney, D. C. (2010). Influence of drug binding kinetics on pharmacodynamic properties. *Drug Discovery*, *7*, 53–57.

86.   Newton, P., Harrison, P., & Clulow, S. (2008). A novel method for determination of the affinity of protein: Protein interactions in homogeneous assays. *Journal of Biomolecular Screening*, *13*(7), 674–682. doi:10.1177/1087057108321086

87.   Abdiche, Y., Malashock, D., Pinkerton, A., & Pons, J. (2008). Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Analytical Biochemistry*, *377*(2), 209–217. doi:10.1016/j.ab.2008.03.035

88.   Rich, R. L., & Myszka, D. G. (2011). Survey of the 2009 commercial optical biosensor literature. *Journal of Molecular Recognition*, *24*(6), 892–914. doi:10.1002/jmr.1138

89.   Shiau, A., Massari, M., & Ozbal, C. (2008). Back to basics: Label-free technologies for small molecule screening. *Combinatorial Chemistry & High Throughput Screening*, *11*(3), 231–237. doi:10.2174/138620708783877807

90. Jamieson, E. R., Jacobson, M. P., Barnes, C. M., Chow, C. S., & Lippard, S. J. (1999). Structural and kinetic studies of a Cisplatin-modified DNA Icosamer binding to HMG1 domain B. *Journal of Biological Chemistry*, *274*(18), 12346–12354. doi:10.1074/jbc.274.18.12346

91. Kim, S.-E., Su, W., Cho, M., Lee, Y., & Choe, W.-S. (2012). Harnessing aptamers for electrochemical detection of endotoxin. *Analytical Biochemistry*, *424*(1), 12–20. doi:10.1016/j.ab.2012.02.016

92. Sloat, A. L., Roper, M. G., Lin, X., Ferrance, J. P., Landers, J. P., & Colyer, C. L. (2008). Protein determination by microchip capillary electrophoresis using an asymmetric squarylium dye: Noncovalent labeling and nonequilibrium measurement of association constants. *Electrophoresis*, *29*(16), 3446–3455. doi:10.1002/elps.200700808

93. Sun, J., He, B., Liu, Q., Ruan, T., & Jiang, G. (2012). Characterization of interactions between organotin compounds and human serum albumin by capillary electrophoresis coupled with inductively coupled plasma mass spectrometry. *Talanta*, *93*, 239–244. doi:10.1016/j.talanta.2012.02.024

94. Krylov, S. N. (2005). Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM): A novel method for Biomolecular screening. *Journal of Biomolecular Screening*, *11*(2), 115–122. doi:10.1177/1087057105284339

95. Neves, M. A. D., Reinstein, O., Saad, M., & Johnson, P. E. (2010). Defining the secondary structural requirements of a cocaine-binding aptamer by a thermodynamic and mutation study. *Biophysical Chemistry*, *153*(1), 9–16. doi:10.1016/j.bpc.2010.09.009

96. Lohman, T. M., & Ferrari, M. E. (1994). Escherichia coli single-stranded DNA-Binding protein: Multiple DNA-Binding modes and Cooperativities. *Annual Review of Biochemistry*, *63*(1), 527–570. doi:10.1146/annurev.bi.63.070194.002523

97. Marcotte, D., Zeng, W., Hus, J.-C., McKenzie, A., Hession, C., Jin, P., Bergeron, C., Lugovskoy, A., Enyedy, I., Cuervo, H., Wang, D., Atmanene, C., Roecklin, D., Vecchi, M., Vivat, V., Kraemer, J., Winkler, D., Hong, V., Chao, J., Lukashev, M., & Silvian, L. (2013). Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorganic & Medicinal Chemistry*, *21*(14), 4011–4019. doi:10.1016/j.bmc.2013.04.019

98. Vintonyak, V. V., Waldmann, H., & Rauh, D. (2011). Using small molecules to target protein phosphatases. *Bioorganic & Medicinal Chemistry*, *19*(7), 2145–2155. doi:10.1016/j.bmc.2011.02.047

99. Lu, H., & Tonge, P. J. (2010). Drug–target residence time: Critical information for lead optimization. *Current Opinion in Chemical Biology*, *14*(4), 467–474. doi:10.1016/j.cbpa.2010.06.176

100. Copeland, R. A. (2010). The dynamics of drug-target interactions: Drug-target residence time and its impact on efficacy and safety. *Expert Opinion on Drug Discovery*, *5*(4), 305–310. doi:10.1517/17460441003677725

101.   Cooper, M. A. (2002). Optical biosensors in drug discovery. *Nature Reviews Drug Discovery*, *1*(7), 515–528. doi:10.1038/nrd838

102.   Myszka, D. (2004). Analysis of small-molecule interactions using Biacore S51 technology. *Analytical Biochemistry*, *329*(2), 316–323. doi:10.1016/s0003-2697(04)00260-x

103.   Fang, Y. (2012). Ligand–receptor interaction platforms and their applications for drug discovery. *Expert Opinion on Drug Discovery*, *7*(10), 969–988. doi:10.1517/17460441.2012.715631

104.   Murugan, R., & Mazumdar, S. (2006). Effect of alcohols on binding of camphor to cytochrome P450cam: Spectroscopic and stopped flow transient kinetic studies. *Archives of Biochemistry and Biophysics*, *455*(2), 154–162. doi:10.1016/j.abb.2006.09.007

105.   Rajagopalan, P. T. R., Zhang, Z., McCourt, L., Dwyer, M., Benkovic, S. J., & Hammes, G. G. (2002). Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics. *Proceedings of the National Academy of Sciences*, *99*(21), 13481–13486. doi:10.1073/pnas.172501499

106.   Kanoh, N., Kyo, M., Inamori, K., Ando, A., Asami, A., Nakao, A., & Osada, H. (2006). SPR imaging of photo-cross-linked small-molecule arrays on gold. *Analytical Chemistry*, *78*(7), 2226–2230. doi:10.1021/ac051777j

107.   Kooyman, R. P. H. (2008). *Handbook of surface plasmon resonance*. doi:10.1039/9781847558220-00015

108.   Brady, L. J., Valliere-Douglass, J., Martinez, T., & Balland, A. (2008). Molecular mass analysis of antibodies by on-line SEC-MS. *Journal of the American Society for Mass Spectrometry*, *19*(4), 502–509. doi:10.1016/j.jasms.2007.12.006

109.   Liu, H., Gaza-Bulseco, G., & Chumsae, C. (2009). Analysis of reduced monoclonal antibodies using size exclusion chromatography coupled with mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *20*(12), 2258–2264. doi:10.1016/j.jasms.2009.08.015

110.   Schmidt, A., Fahlbusch, B., & Otto, M. (2009). Size exclusion chromatography coupled to electrospray ionization mass spectrometry for analysis and quantitative characterization of arsenic interactions with peptides and proteins. *Journal of Mass Spectrometry*, *44*(6), 898–910. doi:10.1002/jms.1563

111.   Okhonin, V., Petrov, A. P., Berezovski, M., & Krylov, S. N. (2006). Plug−Plug kinetic capillary electrophoresis: Method for direct determination of rate constants of complex formation and dissociation. *Analytical Chemistry*, *78*(14), 4803–4810. doi:10.1021/ac060108i

112.   Saito, K., Nakato, M., Mizuguchi, T., Wada, S., Uchimura, H., Kataoka, H., Yokoyama, S., Hirota, H., & Kiso, Y. (2014). Application of plug-plug technique to ACE experiments for discovery of peptides binding to a larger target protein: A model study of calmodulin-binding fragments selected from a digested mixture of reduced BSA. *Electrophoresis*, *35*(6), 846–854. doi:10.1002/elps.201300339

113.   Teppema, L. J., Bijl, H., Gourabi, B. M., & Dahan, A. (2006). The carbonic anhydrase inhibitors methazolamide and acetazolamide have different effects on the hypoxic ventilatory response in the anaesthetized cat. *The Journal of Physiology*, *574*(2), 565–572. doi:10.1113/jphysiol.2006.110528

114.   Ho, Y. T., Purohit, A., Vicker, N., Newman, S. P., Robinson, J. J., Leese, M. P., Ganeshapillai, D., Woo, L. W., Potter, B. V., & Reed, M. J. (2003). Inhibition of carbonic anhydrase II by steroidal and non-steroidal sulphamates. *Biochemical and Biophysical Research Communications*, *305*(4), 909–914. doi:10.1016/s0006-291x(03)00865-9

115.   Casassa, E. F. (1967). Equilibrium distribution of flexible polymer chains between a macroscopic solution phase and small voids. *Journal of Polymer Science Part B: Polymer Letters*, *5*(9), 773–778. doi:10.1002/pol.1967.110050907

116.   Casassa, E. F., & Tagami, Y. (1969). An equilibrium theory for exclusion chromatography of branched and linear polymer chains. *Macromolecules*, *2*(1), 14–26. doi:10.1021/ma60007a003

117.   Casassa, E. F. (1971). Theoretical models for peak migration in gel permeation chromatography. *The Journal of Physical Chemistry*, *75*(26), 3929–3939. doi:10.1021/j100695a003

118.   Netopilík, M. (2002). Relations between the separation coefficient, longitudinal displacement and peak broadening in size exclusion chromatography of macromolecules. *Journal of Chromatography A*, *978*(1-2), 109–117. doi:10.1016/s0021-9673(02)01339-0

119.   Song, M. S., Hu, G. X., Li, X. Y., & Zhao, B. (2002). Study on the concentration effects in size exclusion chromatography. *Journal of Chromatography A*, *961*(2), 155–170. doi:10.1016/s0021-9673(02)00347-3

120.   Sun, T., Chance, R. R., Graessley, W. W., & Lohse, D. J. (2004). A study of the separation principle in size exclusion chromatography. *Macromolecules*, *37*(11), 4304–4312. doi:10.1021/ma030586k

121.   Li, Z., Gu, Y., & Gu, T. (1998). Mathematical modeling and scale-up of size-exclusion chromatography. *Biochemical Engineering Journal*, *2*(2), 145–155. doi:10.1016/s1369-703x(98)00027-8

122.   Venkata Saritha, N., & Madras, G. (2001). Modeling the chromatographic response of inverse size-exclusion chromatography. *Chemical Engineering Science*, *56*(23), 6511–6524. doi:10.1016/s0009-2509(01)00302-5

123.   Zelic, B., & Nesek, B. (2006). Mathematical modeling of size exclusion chromatography. *Engineering in Life Sciences*, *6*(2), 163–169. doi:10.1002/elsc.200620903

124.   Fuchs, H., & Gessner, R. (2001). The result of equilibrium-constant calculations strongly depends on the evaluation method used and on the type of experimental errors. *Biochemical Journal*, *359*(2), 411–418. doi:10.1042/bj3590411

125.   Wortmann, A., Jecklin, M. C., Touboul, D., Badertscher, M., & Zenobi, R. (2008). Binding constant determination of high-affinity protein–ligand complexes by electrospray ionization mass spectrometry and ligand competition. *Journal of Mass Spectrometry*, *43*(5), 600–608. doi:10.1002/jms.1355

126.   Cristoni, S., Bernardi, L. R., Biunno, I., & Guidugli, F. (2002). Analysis of protein ions in the range 3000-12000?Th under partial (no discharge) atmospheric pressure chemical ionization conditions using ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry*, *16*(12), 1153–1159. doi:10.1002/rcm.693

127.   Giorgianni, F., Cappiello, A., Beranova-Giorgianni, S., Palma, P., Trufelli, H., & Desiderio, D. M. (2004). LC−MS/MS analysis of peptides with methanol as organic modifier: Improved limits of detection. *Analytical Chemistry*, *76*(23), 7028–7038. doi:10.1021/ac0493368

128.   Ackermann, B. L., Berna, M. J., Eckstein, J. A., Ott, L. W., & Chaudhary, A. K. (2008). Current applications of liquid chromatography/mass Spectrometry in pharmaceutical discovery after a decade of innovation. *Annual Review of Analytical Chemistry*, *1*(1), 357–396. doi:10.1146/annurev.anchem.1.031207.112855

129.   Summerfield, R. L., Daigle, D. M., Mayer, S., Mallik, D., Hughes, D. W., Jackson, S. G., Sulek, M., Organ, M. G., Brown, E. D., & Junop, M. S. (2006). A 2.13 Å structure of *E. Coli* Dihydrofolate Reductase bound to a novel competitive inhibitor reveals a new binding surface involving the M20 loop region. *Journal of Medicinal Chemistry*, *49*(24), 6977–6986. doi:10.1021/jm060570v

130.   Wani, M. A., Xu, X., & Stambrook, P. J. (1994). Increased methotrexate resistance and dhfr gene amplification as a consequence of induced ha-ras expression in NIH 3T3 cells. *Cancer Research*, *54*(9), 2504–2508.

131.   Liu, C. T., Hanoian, P., French, J. B., Pringle, T. H., Hammes-Schiffer, S., & Benkovic, S. J. (2013). Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proceedings of the National Academy of Sciences*, *110*(25), 10159–10164. doi:10.1073/pnas.1307130110

132.   Matuszewski, B. K., Constanzer, M. L., & Chavez-Eng, C. M. (2003). Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC−MS/MS. *Analytical Chemistry*, *75*(13), 3019–3030. doi:10.1021/ac020361s

133.   Belov, M. E., Gorshkov, M. V., Udseth, H. R., Anderson, G. A., & Smith, R. D. (2000). Zeptomole-sensitivity electrospray ionization−fourier transform ion cyclotron resonance mass Spectrometry of proteins. *Analytical Chemistry*, *72*(10), 2271–2279. doi:10.1021/ac991360b