To my parents
# TABLE OF CONTENTS

**LIST OF TABLES** .......................................................... vii

**LIST OF FIGURES** ......................................................... viii

**ABSTRACT** ....................................................................... xv

**CHAPTER 1. Introduction** .................................................. 1
  1.1 Biological background ..................................................... 1
    1.1.1 The SELEX experiment ............................................. 1
    1.1.2 Aptamer .................................................................... 2
  1.2 Mathematical models for SELEX ........................................... 4
  1.3 Outline of the thesis .......................................................... 5

**CHAPTER 2. A mathematical analysis of multiple-target (positive) SELEX** 8
  2.1 Introduction ................................................................. 8
  2.2 Formulation of a single round of multiple target SELEX ............... 9
    2.2.1 Mass action kinetics .................................................. 9
    2.2.2 Formulation of the selection process .............................. 10
    2.2.3 Overall dissociation and association constants and efficiencies ... 12
  2.3 The SELEX iteration scheme .............................................. 14
    2.3.1 Algorithm of the SELEX iteration scheme ...................... 14
    2.3.2 Optimal choice for the target concentration .................... 15
  2.4 Limiting states of the SELEX iteration scheme ....................... 16
    2.4.1 Limiting values of NA fractions as the round number becomes large and their asymptotic stability .............................. 16
2.4.2 Convergence of nucleic acid fractions ........................................ 17
2.4.3 Limiting free target fractions and target efficiencies ....................... 19

2.5 The geometry of the SELEX problem .............................................. 20

2.6 Solution of the SELEX problem ...................................................... 22
  2.6.1 The solution of the SELEX problem at infinite NA dilution ([NA] ≈ 0) 22
  2.6.2 The solution of the SELEX problem at finite NA dilution ([NA] > 0) . 22
  2.6.3 Selected indices of the SELEX iteration scheme: The maximal target
       affinity function ............................................................................ 23
  2.6.4 A system of nonlinear SELEX equations: The relation of final nucleic
       acid fractions and final target fractions ........................................... 26

2.7 The connection between the limiting efficiencies and chemical potentials 28

2.8 The stability and uniqueness theorems for multiple target SELEX ............ 30
  2.8.1 Statements of the stability and uniqueness theorems for multiple target
       SELEX, at infinite target dilution .................................................... 30
  2.8.2 Proof of Theorem 2.8.1.1 and Theorem 2.8.1.2 ............................. 31

2.9 Geometric properties of the family of convex hulls generated by a single free
     target vector \( \hat{\omega} \) ........................................................................ 37

2.10 The initial target fraction relationship to the final free target and nucleic acid
     fractions ......................................................................................... 38

2.11 A special solution of the stationary SELEX equations .......................... 40

2.12 Simulation results for multiple-target SELEX .................................... 43
  2.12.1 Final fractions using the SELEX iteration scheme .......................... 43
  2.12.2 Maximal target affinity function, initial target decomposition and chemi-
          cal potential functions using the iteration scheme ....................... 46
  2.12.3 Comparison of stationary SELEX solutions with those obtained by iter-
           ation and stability properties of stationary solutions ................. 47
  2.12.4 Stability properties of stationary solutions. ................................. 47
CHAPTER 3. A mathematical analysis of alternate SELEX against multiple targets

3.1 Introduction ........................................................................................................ 57
3.2 Formulation of a single round of multiple-target positive SELEX .................. 58
3.3 Formulation of a single round of multiple-target negative SELEX ............... 60
3.4 Alternate SELEX iteration scheme: Positive SELEX followed by negative SELEX 63
3.5 Limiting behavior of the alternate SELEX iteration scheme and maximal target
     affinity function ............................................................................................... 66
3.6 SELEX efficiency for alternate SELEX .............................................................. 69
3.7 Limiting behavior of alternate SELEX when the grand round number becomes
     large: Specificity ............................................................................................. 73
3.8 The set of stationary solutions, specificity of the dynamical system on the chem-
     ical potential energy surfaces ......................................................................... 77
3.9 Simulation results for alternate SELEX .............................................................. 79
     3.9.1 Limiting NA fractions, efficiencies and target affinity function values
             using the SELEX iteration scheme ............................................................ 79
     3.9.2 Specificity, the limiting target specificity function values and the set of
             indices for specificity ................................................................................. 81
     3.9.3 Chemical potential surfaces and the relationship between the limiting
             chemical potential curves and the steepest descent curves ....................... 82

CHAPTER 4. Summary and future work ................................................................. 93

4.1 Results .............................................................................................................. 93
     4.1.1 Multiple-target (positive) SELEX ............................................................ 93
     4.1.2 Alternate SELEX: Positive selection followed by negative selection ...... 93
4.2 Future work ....................................................................................................... 94

CHAPTER 5. Appendix ......................................................................................... 97

5.1 Appendix A. Notation ..................................................................................... 97
5.2 Appendix B. Convergence of nucleic acid fractions ....................................... 97
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Chemical Symbol Notation</td>
<td>97</td>
</tr>
<tr>
<td>5.2</td>
<td>Vector Notation</td>
<td>98</td>
</tr>
<tr>
<td>5.3</td>
<td>Data for Figure 2.9, panel (a). $\hat{\Omega} = (0.3, 0.6, 0.1)$</td>
<td>100</td>
</tr>
<tr>
<td>5.4</td>
<td>Data for Figure 2.9, panel (b). $\hat{\Omega} = (0.15, 0.3, 0.55)$</td>
<td>100</td>
</tr>
<tr>
<td>5.5</td>
<td>Data for Figure 2.9, panel (c). $\hat{\Omega} = (0.65, 0.15, 0.2)$</td>
<td>100</td>
</tr>
<tr>
<td>5.6</td>
<td>Data for Figure 2.9, panel (d). $\hat{\Omega} = (0.4, 0.1, 0.5)$</td>
<td>101</td>
</tr>
<tr>
<td>5.7</td>
<td>Data for Figure 2.10, panel (a). $\hat{\Omega} = (0.35, 0.4, 0.25)$</td>
<td>101</td>
</tr>
<tr>
<td>5.8</td>
<td>Data for Figure 2.10, panel (b). $\hat{\Omega} = (0.1, 0.6, 0.3)$</td>
<td>101</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 The steps of SELEX iteration process. The single-stranded DNA selection protocols and the RNA selection protocols are followed by the solid arrows and the dashed arrows, respectively. (Adapted from [15]) . . . . . . . . 5

2.1 Panels (a,b): With \([NA] = [T] = 1\mu M\), the index sets for the selected nucleic acids are (a) \(\{8, 9, 10, 12, 16\}\) and (b) \(\{8, 9, 10, 16\}\). The initial target vectors are (a) \(\hat{\Omega}_a = \langle 0.1374, 0.1346, 0.4090, 0.1844, 0.1346 \rangle\) and (b) \(\hat{\Omega}_b = \langle 0.2376, 0.1453, 0.1145, 0.2821, 0.2205 \rangle\) with the limiting NA fraction vectors given by \(F_a = \langle F_8, F_9, F_{10}, F_{12}, F_{16} \rangle = \langle 0.1956, 0.2794, 0.0498, 0.3843, 0.0908 \rangle\) and \(F_b = \langle F_8, F_9, F_{10}, F_{16} \rangle = \langle 0.3060, 0.0670, 0.2629, 0.3640 \rangle\). These limiting vectors agree with values from the formula (2.63) to at least eight significant figures (not shown). Panels (c,d): The final free target fractions in cases (a,b), after 200 rounds, not shown, are: (c) \(\tilde{\omega} = \langle 0.1617, 0.1241, 0.3139, 0.2420, 0.1583 \rangle\) and (d) \(\tilde{\omega} = \langle 0.1690, 0.1288, 0.2965, 0.2476, 0.1581 \rangle\), with the limiting association constants \(K_a = 3207.45(\mu M)^{-1}\) and \(K_a = 3256.05(\mu M)^{-1}\), respectively. 49

2.2 The individual nucleic acid efficiencies after 40 rounds. The bar graphs are not identical because, in Figure 2.1(a), selection is for indices \(\{8, 9, 10, 12, 16\}\) while, in Figure 2.1(b), selection is for indices \(\{8, 9, 10, 16\}\). . . . . . . . . . 50
2.3 After 125 rounds it would appear from panel (a) that the selected indices are \{1, 6, 8, 9, 10, 12, 16\} whereas from panel (b) after 400 rounds the selected indices would appear to be \{1, 8, 9, 10, 12, 16\}. Compare with Figure 2.1(a). Thus, the importance of reducing the total concentration of target from round to round in a systematic way is even more pronounced in the multiple target problem.

2.4 In panel (a), the dependence of the number of selected nucleic acids on the nucleic acid pool size is illustrated. In panel (b), $pK_a$ decreases with decreasing pool size, i.e., $K_a$ decreases with increasing pool size.

2.5 Nonmonotonicity of $pK_a$ (or, equivalently, $pK_d$) as a function of round number. The left-hand panel traces $pK_a(r)$ for various pool sizes up to 120 rounds. The right-hand panel is a continuation of these curves for an expanded scale on the vertical axis.

2.6 Panel (a) is a graph of the maximal target affinity function when its domain has been projected onto the plane. Using the SELEX iteration scheme for 2,000 rounds, the panel (c) was generated from a fixed initial pool of nucleic acids $\hat{F} = (F_1, F_2, F_3, F_4, F_5) = (0.3002, 0.0731, 0.1917, 0.1535, 0.2815)$. (The number of rounds is large because the rate of convergence of SELEX program slows as the nucleic acid pool size decreases.) In panels (c,d), the indicated regions are labeled with the indices of the nucleic acids that will be selected when the initial target is selected from the interior of the indicated region. The red triangle in panel (b) generates the red hexagon in panel (d). (There is a corresponding hexagon for panel (c) that is omitted in the interests of clarity.) The initial target vectors $\langle \Omega_1, \Omega_2, \Omega_3 \rangle$ were generated by setting $\Omega_3 = 1 - \Omega_1 - \Omega_2$, where for $j = 1, \ldots, 199$, $i = 1, \ldots, 200 - j$ and $\Omega_1 = \Omega_1(i, j) = j/200, \Omega_2 = \Omega_2(i, j) = i/200$ or about 20,000 initial target vectors.
Panels (a) and (b) illustrate the graphs of chemical potentials defined as $Q$ in (2.56), at infinite target dilution. The minimum values of these functions are indicated by $\ast$'s. Then the corresponding final NA fractions are (a) $\hat{F} = (0.2400, 0.2400, 0.5200)$ with the minimum value of $-7.9328$ and (b) $\hat{F} = (0.6900, 0.0100, 0.3000)$ with the minimum value of $-7.9992$. Using the SELEX iteration scheme the final NA fractions are (c) $\hat{F} = (0.2449, 0.2355, 0.5196)$ with $-\ln K_a = -7.9369$ after 40 rounds, and (d) $\hat{F} = (0.6984, 0.0000, 0.3016)$ with $-\ln K_a = -8.0001$ after 30 rounds. The values of $\hat{F}$ at the critical points in panels (a,b) are in quite good agreement with the values of $\hat{F}$, found by iteration using the SELEX scheme, in panels (c,d), respectively.

The panels (a,c), were generated in the same manner as panels (c,d) in Figure 2.6. In all four panels, the subregions are labeled with the indices of the nucleic acids that will be selected when the initial target vector is in the indicated region. We refer to the case illustrated in panels (c,d) as improper because the minimum of the maximal target affinity function, $\varphi$, is defined by the intersection of any three of the four planes that define its graph.

These figures show that whatever convex hull in Figure 2.8, panel (a) (or panel (b)), to which the initial target fraction vector belongs, the SELEX process converges to a unique set of final nucleic acid fractions independently of the distribution of starting nucleic acids in the starting pool. Each of the five curves in each panel correspond to an independent trial for the starting pool of nucleic acids. (The vertical axis notation $|F_{i} - F_{j}|_1$ is shorthand for $|\hat{F}_{i}^{(r)} - \hat{F}_{j}^{(r)}|_1 = \sum_{i=1}^{5} |F_{i}^{(r)} - F_{i,j}^{(r)}|$. Here $j = 2, \ldots, 6$ for five of the six random starting vectors. The starting ordinates of the norms $|F_{i} - F_{j}|_1$ for each of the curves in panels (a)-(d) are recorded in the fourth column of Tables 5.3-5.6, in Appendix C, respectively.)
The graphs show that the SELEX process is not globally asymptotically stable when the initial target fraction belongs to the convex hull of an improper face (panel (a)) and is asymptotically stable when it belongs to a convex hull corresponding to a proper face (panel (b)). Each of the five curves in both panels correspond to an independent trial for the starting pool of nucleic acids. (The vertical axis notation $|F_1 - F_j|_1$ is shorthand for $|\hat{F}_1^{(r)} - \hat{F}_j^{(r)}|_1 = \sum_{i=1}^{5} |F_{i,1}^{(r)} - F_{i,j}^{(r)}|$. Here $j = 2, \ldots, 6$ for five of the six random starting vectors. The starting ordinates of the norms $|F_1 - F_j|$ for each of the curves in panels (a)-(d) are recorded in the fourth column of Tables 5.7, 5.8, in Appendix C, respectively.)

3.1 The steps of the alternate SELEX iteration process: One can go around one loop $R_s$ times and around the other $R_n$ times for one grand round, $R = R_s + R_n$.

3.2 Each panel illustrates the distribution of nucleic acid fractions for negative SELEX only with different choices of target concentrations. $[NA]$ is fixed as $1\mu M$ for all three cases.

3.3 Corresponding efficiencies to the cases in Figure 3.2 are described. $[NA] = 1\mu M$.

3.4 Comparison of positive SELEX with alternate SELEX was made. During negative selection the fifth target component was absent. In panels (b) and (c), the grand round number $R = 50$ was taken. By comparing the panel (a) with (b), we notice that the 9th NA species has disappeared, while the fraction of 8th NA species has increased. From panel (d) we see that 13th NA is the poorest binder to the target with the fifth component deleted. In panels (b-d), $[T_n] = 100\mu M$. 

[56, 59, 83, 84, 85]
Alternate SELEX beginning with negative selection was performed to check the dependency on the order of positive and negative selections. In panel (a), alternate SELEX with two negative rounds (yellow) followed by three positive rounds (dark red) was performed. \( \lambda = 0.6 \). Compare with Figure 3.4, panel (b). In panel (b), alternate SELEX with three negative rounds (yellow) followed by two positive rounds (dark red) was performed. \( \lambda = 0.4 \). Compare with Figure 3.4, panel (c).

Target affinity function values, (equation (3.33) after 50 rounds) are shown. In panel (a), selected NA indices are \{8, 10, 12, 16\}. See Figure 3.4, panel (b). In panel (b), selected NA indices are \{17, 20\}. See Figure 3.4, panel (c). They have significant differences in magnitude between the two values in panel (a) and (b). The values of the target affinity functions that correspond to the selected NA indices are (a) \( \langle \varphi_8, \varphi_{10}, \varphi_{12}, \varphi_{16} \rangle = \langle 0.6955, 0.6962, 0.6952, 0.6959 \rangle \) and (b) \( \langle \varphi_{17}, \varphi_{20} \rangle = \langle 0.4113(10^{-4}), 0.4112(10^{-4}) \rangle \).

Efficiencies for alternate SELEX are illustrated. These are the efficiency plots for the corresponding panels illustrated in Figure 3.4. In panels (b,c) the curves designated by ‘\( \Diamond \)’ denote the cumulative efficiencies using equation (3.54). The insets in these panels indicate the asymptotic behavior of these efficiencies.

Effect of increasing the multiplier \( m \) on specificity is described.
Alternate SELEX for a large number of positive selection rounds ($mR_s >> 1$) followed by a few rounds of negative selection was performed to examine the specificity of the process. Five target components case is taken from multiple-target positive SELEX in Figure 2.1 panel (a). Each plot shows nucleic acid fractions for corresponding indices \{8, 9, 10, 12, 16\} performed during negative selection, as a function of round number. The starting NA fractions for the first round above are the final NA fractions at selection when only positive selection was performed, and they are updated during the negative selection process. In each case, the number of selected nucleic acid species are decreased to the single NA species that binds to the desired target (absent during negative selection) with high affinity and specificity.

Limiting target specificity function values are presented here. In panel (a) we calculated $\psi_j(\lambda, \tilde{\omega}_\lambda^s)$ as a function of $j = 1, \ldots, 20$, for $\lambda = 0.6$. Notice that the maximum occurs precisely for $j = 8$. In panel (b) we calculated $\psi_j(\lambda, \tilde{\omega}_1^s)$ for $\lambda = 0.6$. While they are not identical, the two panels indicate that at least for $\lambda > 1/2$, we may use $\tilde{\omega}_1^s$ as a reasonable approximation for $\tilde{\omega}_\lambda^s$.

Selected indices are shown from various approximation approaches. In panel (a), for each $\lambda \in [0, 1]$, the members of $\mathcal{L}_\lambda$ are indicated in heavy dots. For example, when $\lambda = 0.5$, $\mathcal{L}_\lambda = \{8, 17\}$. In panel (b), nucleic acid indices that maximize the target specificity function values, i.e., that maximize $\psi_j(\lambda, \tilde{\omega}_\lambda^s)$, $j \in \mathcal{N}$ from equations (3.61) and (3.62) are plotted. In panel (c), approximation to panel (b) was made from the indices that maximize $\psi_j(\lambda, \tilde{\omega}_1^s)$, $j \in \mathcal{N}$. The advantage is for this approximation that it only requires the information of the limiting free target vector obtained from positive SELEX only ($\lambda = 1.0$).
3.12 Limiting chemical potentials for the three target case are shown. The pool consisted of seven nucleic acid species. The set $L_{\lambda,s} = 1, 4, 6$. Here, $F_6 = 1 - F_1 - F_4$. The limiting curve in panel (a) (roughly the case when $m = 160$), is orthogonal to the level sets of $R_s$ when these are referred its graph over the simplex $S_{L_{\lambda,s}}$. However, even when referred to the projection of this simplex in the plane, this curve is very nearly orthogonal to the level sets of $R_s$. 

3.13 “Hysteresis” loop for the limiting nucleic acid fraction vectors $\hat{F}_s, \hat{F}_n$ is shown for the case $m = 10$ and $m = 160 (= \infty)$ in panels (a,b), respectively. 

3.14 Plots of steepest descent curves and the point sets $P_s, P_n$ (or the curves $C_s(\lambda), C_n(\lambda)$ defined by the point sets, respectively) in alternate selection are shown over the chemical potential contours.
This thesis develops a mathematical model of the biological procedure SELEX (Systematic Evolution of Ligands by EXponential Enrichment). The procedure is an \textit{in vitro} method for identifying nucleic acid (NA) molecules that have an ability to bind tightly and specifically to target species of interest, such as small organic molecules, peptides or proteins.

We explore two main algorithms: multiple target (positive) SELEX and alternate SELEX. The schemes are considered as \textit{discrete time dynamical systems}, and the limiting (steady-state) behaviors of the processes are characterized by the initial parameters of each system: concentration of total targets, concentration of a pool of nucleic acids and fractional distributions of NA and target molecules. To gain a better understanding of our systems, we also construct (simplex-based) geometric structures of the limiting states, in terms of chemical thermodynamics.

We find that the dynamical system defined by the multiple target (positive) SELEX process is globally asymptotically stable, when and only when a certain family of chemical potentials at infinite target dilution has at most one critical point. That is, the SELEX iteration scheme converges to a unique subset of nucleic acids and does not depend on the distribution of nucleic acids present in the pool, but does depend upon the total nucleic acid concentration, the initial fractional distribution of the targets and the overall limiting equilibrium association constant.

We also present a mathematical model for “alternate SELEX”. The goal here is to minimize an enrichment of non-specifically binding nucleic acids against multiple targets by alternating two processes: positive selection and negative selection, which in combination result in the selection of nucleic acids with high “selectivity” and “specificity”.

\textbf{ABSTRACT}

This thesis develops a mathematical model of the biological procedure SELEX (Systematic Evolution of Ligands by EXponential Enrichment). The procedure is an \textit{in vitro} method for identifying nucleic acid (NA) molecules that have an ability to bind tightly and specifically to target species of interest, such as small organic molecules, peptides or proteins.

We explore two main algorithms: multiple target (positive) SELEX and alternate SELEX. The schemes are considered as \textit{discrete time dynamical systems}, and the limiting (steady-state) behaviors of the processes are characterized by the initial parameters of each system: concentration of total targets, concentration of a pool of nucleic acids and fractional distributions of NA and target molecules. To gain a better understanding of our systems, we also construct (simplex-based) geometric structures of the limiting states, in terms of chemical thermodynamics.

We find that the dynamical system defined by the multiple target (positive) SELEX process is globally asymptotically stable, when and only when a certain family of chemical potentials at infinite target dilution has at most one critical point. That is, the SELEX iteration scheme converges to a unique subset of nucleic acids and does not depend on the distribution of nucleic acids present in the pool, but does depend upon the total nucleic acid concentration, the initial fractional distribution of the targets and the overall limiting equilibrium association constant.

We also present a mathematical model for “alternate SELEX”. The goal here is to minimize an enrichment of non-specifically binding nucleic acids against multiple targets by alternating two processes: positive selection and negative selection, which in combination result in the selection of nucleic acids with high “selectivity” and “specificity”.
CHAPTER 1. Introduction

1.1 Biological background

1.1.1 The SELEX experiment

A systematic evolution methodology, SELEX (or in vitro selection, [29], [9]), has become a frequently used approach in molecular biology. It is a method to isolate nucleic acid (NA) molecules having biological functions for specific target species of interest. This process consists of several cycles of selection and amplification steps of nucleic acid molecules (deoxyribonucleic acid (DNA) or ribonucleic acid (RNA)). The binding target molecules could be protein, peptide, inorganic or small organic molecules, as well as complex targets like target mixtures or whole cells.

Once target molecules are exposed to a random library\(^1\) of nucleic acids, target-NA complexes are obtained and separated (or partitioned) from the unbound nucleic acids. The binding NA molecules are eluted (selection step) and amplified by Polymerase Chain Reaction \(^2\) (PCR, [16]) (amplification step). Each selection-amplification round of the SELEX process is repeated using the purified pool of nucleic acids that bind better to targets from the previous round. By successive rounds, the pool is enriched with nucleic acids that bind tightly to specific target molecules. In this thesis we call this process SELEX or positive SELEX and define “negative” SELEX as follows.

Negative SELEX also involves selection and amplification steps. Instead of eluting bound nucleic acids from the NA mixture, however, unbound (or free) NA molecules are collected

---

\(^1\)A DNA library is a collection of cloned DNA fragments.

\(^2\)PCR is an in vitro method for synthesizing a very large number of copies of a specific DNA sequence, approximately \(2^n\) copies after \(n\) cycles of replication. When RNAs are eluted from the selection step, they are reverse-transcribed to complementary DNA (cDNA) and then amplified. (See [9], [11], [29] and Figure 1.1.)
during negative SELEX. The unbound NA species are amplified by PCR to initiate the next round of the selection process. While positive SELEX alone determines NAs that bind tightly to targets, the negative SELEX process removes NA species that bind readily to target molecules other than the desired target. Here, we introduce the “alternate” SELEX process with multiple targets by incorporating negative selection into a (positive) SELEX protocol. After several rounds of positive selection with the desired target as part of the mix of targets to select the NAs, the selected NA species are likely to contain the NA molecules that bind to the undesired targets (such as the support or matrix material) as well as those that bind the desired target. Thus, one can implement the negative selection procedure in such a way that the NA molecules are mixed only with the undesired targets, and the binding NA species (having a high affinity to the undesired targets) are now removed. By combining these two procedures the NA molecules that bind to the desired target(s) with high selectivity can be obtained.

1.1.2 Aptamer

Antibodies are proteins that allow the immune system in humans and animals to identify foreign molecules or organisms such as antigens, bacteria and viruses. They are used as the molecular identifiers in clinical diagnostic tests for infections or drug use. However, there are a limited number of naturally produced antibodies, and they require physiological conditions\(^3\) to retain their properties. Various kinds of \textit{in vitro} selection techniques have been developed to produce such “antibody-like” molecules for the use of therapeutic and diagnostic applications.

Aptamers\(^4\) are DNA or RNA molecules that fold into unique three dimensional structures, allowing them to bind tightly to their target molecules. It is known that the binding affinities of aptamers to their target molecules are often similar to or higher than the affinities of antibody-antigen interactions. More importantly, aptamers can be selected \textit{in vitro} by the SELEX process to bind various target species of interest. The relatively small size of aptamers allows them to penetrate tissues more effectively than antibodies, so many aptamers are being developed for detecting targets such as cancer cells, and also for detecting analytes such

---

\(^3\) The physiological conditions needed here are rather simple—just pH and salt concentration.

\(^4\) The name was termed by [9] derived from the Latin word “aptus” which means “to fit”.
as blood proteins and drug concentrations, and for imaging cellular processes. Anti-VEGF\(^5\) aptamers ([18]) for the cure of AMD\(^6\) and aptamers against Amyloid β–protein ([22]) for the recognition of Alzheimer’s disease are only a couple of examples of aptamers that have been developed for medical applications.

Aptamers are highly specific, and certain aptamers are capable of discriminating between very similar functional groups on molecules and bind only to specific target molecules ([9], [10]). This binding affinity and specificity\(^7\) of aptamers can be achieved by methods that involve selecting for the desired targets and against the undesired targets. For example, an aptamer to theophylline, a drug used to treat asthma, was selected using “counter-SELEX” in which the undesired target (caffeine molecules that are chemically similar to theophylline) was used in a competitive manner to remove aptamers to which it bound ([12]). While this approach results in an aptamer with a selective binding property, it can only be done if the structure of the undesired target is known.

In cases in which there is more than one undesired target and their structures or even their identities are unknown, then another approach is necessary. Examples of this situation are when two cell types are compared (for example, a normal and cancer cell) in which the one cell (perhaps the cancer cell) expresses the desired target that is not expressed on the normal cell. However, many other targets on these cells are unknown. Also, when a (target) protein is captured on a filter, there are many unknown targets that are part of the filter. In these last two circumstances, multiple-target (positive) SELEX or alternate SELEX is appropriate for obtaining an aptamer with specificity for the desired target. See also [3] and review articles [4], [5], [14], [19], [20], [21] and [25] for applications of aptamers. In [5] and [25], various forms of SELEX techniques (SELEX process modifications, Table 2, [25]) and the separation methods (separation of unbound nucleic acids from its bound product) are described.

\(^5\)Anti-vascular endothelial growth factor.
\(^6\)Age-related macular degeneration, a disease causing vision losses.
\(^7\)Specificity has been defined many different ways. In [31], general descriptions of specificity of DAN-protein interactions are illustrated, and the review article [26] addresses concerns about measuring specific binding constants for DAN-protein interactions.
1.2 Mathematical models for SELEX

Even though numerous SELEX techniques have been developed over the last 20 years in laboratories, not many mathematical and computational analyses have been published. A mathematical approach can contribute to further understanding of challenging biological problems. In this thesis, the goal is to understand the novel process of SELEX mathematically.

The SELEX experiment is an iterative procedure of selection and enrichment. Figure 1.1 illustrates the SELEX iterative protocols. Based on a kinetic and equilibrium binding analysis, the problem was first analyzed computationally in [11]. In [15], a mathematical analysis for the SELEX process against single target species appears, and it was proven that the SELEX process converges to a population of single nucleic acid species that bind specifically to the target with the highest affinity. The authors of [27] and [28] also considered single-target SELEX and used a branching process approach to model both selection and amplification procedures. In [28], the number of rounds of SELEX cycles necessary to perform successful SELEX experiments is shown to be closely related to the concentration of the target.

In many cases, however, most target molecules, including proteins, have multiple binding sites, and all their interactions can be considered specific and saturable. Moreover, some targets may not be purified or may be included with other target species in a mixture, so it can be difficult to obtain a desired single target component of interest from a mixture with multiple target species. It is also known that aptamers can be isolated against multiple target components using the SELEX process ([17], [23], [24], [30]). We refer to this process as multiple target SELEX.

In [30], the authors extended the single target model in [11] to a multiple target model. Optimal selection strategies were tested by considering various experimental conditions, such as partitioning efficiencies and relative target concentrations. The authors in [30] suggested the method of subpooling nucleic acid species with similar affinities against each target. This subpooling strategy allows them to identify, simultaneously, a mixture of high affinity nucleic acid species.
acids against all target species. The resulting pool can then be used to isolate refined nucleic acid species against each individual target considering as a single target SELEX process. A log-normal distributional assumption was used to choose numerical values for nucleic acid affinities against each target. The subpooling method requires $N^M$ subpools yielding $N^M$ equations, where $N$ is the number of subpools of nucleic acids and $M$ is the number of targets. Instead, the authors in [7] and [8] suggested a condensed subpooling method. With this method, the size of nucleic acid subpools can be reduced optimally to $N \times M$ instead of $N^M$. See Figure 2 in [7], for a comparison of these two subpooling methods. Stochastic simulations of the complex-target SELEX process appearing in [6] characterize the evolution dynamics of the process under the influence of random effects such as point mutations.

1.3 Outline of the thesis

In this thesis, we reexamine the method in [30] and analyze the multiple target SELEX process. It is assumed all species are present in sufficient amounts so the law of large numbers can be applied to use average values for species concentrations. The primary goal is to charac-
terize the limiting values of the nucleic acid concentrations in terms of the initial and limiting experiment parameters when selection occurs. We note here that, in a laboratory, the SELEX process is usually taken through fewer than 25 rounds (see [30], page 580). Even though we do not restrict the use of a large number of rounds to reach the limiting states, we believe that our mathematical approach will provide a good understanding of the multiple-target SELEX process for experimentalists.

Suppose that we have $M$ target components and an initial pool of $N$ nucleic acids defining $M \times N$ affinity matrix. We show that the SELEX process always converges to some limiting distribution of nucleic acid fractions and some limiting vector of free target fractions. The dynamical system defined by multiple target SELEX depends upon several experiment parameters: the initial fractional target distribution, the total nucleic acid concentration, the initial fractional nucleic acid distribution and the overall limiting equilibrium association constant. We then give a geometric condition on the affinity matrix so that the dynamical system defined by the multiple target SELEX process converges to a unique subset of nucleic acids, whose individual concentrations do not depend upon the initial distribution of nucleic acid fractions as long as all $N$ species are present in the given pool. We refer to this outcome as the “proper” case of the multiple-target SELEX process. The condition is also closely related to the geometric properties of a family of chemical potentials against the entire pool at infinite target dilution (i.e., when the total target concentration approaches zero). That is, every member of a certain family of chemical potentials at infinite target dilution can have at most one critical point. The outcome is then analogous to the result in the single target case ([15]) that the dynamical system defined by single target SELEX will always converge to the single best binding nucleic acid present in the initial pool.

We also present an algorithm for the alternate SELEX process that involves doing positive SELEX followed by negative SELEX. The negative SELEX process can be accomplished by incubating the pool with a target of the same (or different) type as that used for positive selection, except that one or more of the target components are missing. During negative selection, we retain the unbound NA species to be subjected to PCR amplification. By al-
ternating the two SELEX procedures, we obtain a refined population of NA molecules with high selectivity. Here, we give a formulation of the negative selection process and define its efficiency in a meaningful way. We then provide an alternate SELEX iteration scheme and analyze its selectivity and specificity mathematically and computationally. Also, we employ chemical potentials to see how the alternate SELEX process leads to the selection of NA with more specific interactions when the ratio of the numbers of positive and negative selection rounds changes.

Finally, we propose several conjectures related to this procedure for further study.
CHAPTER 2. A mathematical analysis of multiple-target (positive) SELEX

The contents of this chapter have appeared in [23] published in the *Bulletin of Mathematical Biology*.

2.1 Introduction

In this chapter we model the multiple target SELEX process. Based on the law of mass action, we formulate a single step of the selection procedure and propose an iterative scheme of the SELEX experiment as a function of a round number. Using this algorithm, we analyze how the dynamical system behaves when the round number becomes large in terms of system’s convergence and asymptotic stability. By convergence, we mean the property that the system solution converges to a limit point consisting of nucleic acid fractions, when the system reaches selection. Meanwhile, by asymptotic stability, we mean a certain condition exists so some limit points do not depend on the system’s initial condition, where the initial condition is defined by the starting nucleic acid fractions in a given pool.

The SELEX process typically begins with a random library of up to $10^{15}$ DNA or RNA molecules ([30]). Once target-NA bound products are obtained and partitioned from unbound NA molecules, the bound NA molecules are eluted and amplified using a method such as PCR. A single round of the SELEX process consists of one selection-amplification step. The surviving NA molecules from the first round constitute a new library and are subjected to a selection process in the next round. The entire process is repeated until the pool is enriched with nucleic acids\footnote{The selected oligonucleotide (DNA or RNA) molecules are then cloned and sequenced. The frequently occurring sequences are presumed to be the selected NA and these oligonucleotides are synthesized and tested} with high affinity and specificity for multiple target species.
Here, we consider the case for which there is no background loss either through the support or via free nucleic acids left on the support. We also assume that each target has a single NA binding site so that only one NA molecule binds to a target molecule at a time.

Notation for species concentrations and fractions used here is provided in Table 5.1 followed by their (unit) vector notation located in Table 5.2 in Appendix A.

2.2 Formulation of a single round of multiple target SELEX

2.2.1 Mass action kinetics

Consider a pool of \( N \) nucleic acid species, \( NA_j \) for \( j \in \{1, 2, 3, \ldots, N\} = \mathcal{N} \), and \( M \) target species, \( T_i \) for \( i \in \{1, 2, 3, \ldots, M\} = \mathcal{M} \). A chemical reaction by which each target binds to each nucleic acid is given as follows:

\[
\{T_i : NA_j\} \xrightleftharpoons[k_{-i,j}]{k_{i,j}} Tf_i + Na_fj,
\]

(2.1)

where \( k_{i,j} \) and \( k_{-i,j} \) are on and off reaction rate constants, respectively, and the symbol \( \xrightleftharpoons{} \) indicates the reaction is reversible. Here, \( \{T_i : NA_j\} \), \( Tf_i \) and \( Na_fj \) are the bound products, the \( i^{th} \) free (or unbound) target species and the \( j^{th} \) free (or unbound) nucleic acids, respectively. The stoichiometric coefficients of each reactant and product species are one, that is, one molecule (or mole) of each target reacts with one molecule (or mole) of NA to produce one molecule (or mole) of the bound product. By the law of mass action, the rate of change in concentration of the bound product is

\[
\frac{d\{T_i : NA_j\}}{dt} = -k_{-i,j}\{T_i : NA_j\} + k_{i,j}[Tf_i][Na_fj],
\]

where \( \{T_i : NA_j\} \), \( [Tf_i] \) and \( [Na_fj] \) denote the concentrations of the \( i^{th} \) target bound to the \( j^{th} \) nucleic acids, \( i^{th} \) free target and \( j^{th} \) free nucleic acids, respectively, at time \( t \). Assuming each reaction to be at equilibrium, (that is, \( \frac{d\{T_i : NA_j\}}{dt} \approx 0 \)), the dissociation constant for the \( j^{th} \) nucleic acid species bound to the \( i^{th} \) target is given by

\[
K_{ij} = \frac{k_{-i,j}}{k_{i,j}} = \frac{[Tf_i][Na_fj]}{\{T_i : NA_j\}},
\]

(2.2)

against the targets to identify potential aptamer candidates.
The corresponding equilibrium association constant, as a measure of binding affinity, for the reactions is defined by

\[ A_{ij} = \frac{1}{K_{ij}} = \frac{\{T_i : NA_j\}}{[Tf_i][NAf_j]}, \quad (2.3) \]

We then introduce the affinity matrix

\[
A = \begin{bmatrix}
1/K_{11} & \ldots & 1/K_{1N} \\
1/K_{21} & \ldots & 1/K_{2N} \\
\vdots & \ddots & \vdots \\
1/K_{M1} & \ldots & 1/K_{MN}
\end{bmatrix} = \begin{bmatrix}
A_{11} & \ldots & A_{1N} \\
A_{21} & \ldots & A_{2N} \\
\vdots & \ddots & \vdots \\
A_{M1} & \ldots & A_{MN}
\end{bmatrix}, \quad (2.4)
\]

and denote the row vectors of this matrix by \( \overrightarrow{A}_i \) for \( i \in \mathcal{M} \) and the column vectors by \( \overrightarrow{A}_j \) for \( j \in \mathcal{N} \).

2.2.2 Formulation of the selection process

The dynamical system defined by the multiple-target SELEX process is governed by mass conservation laws as follows:

\[
\begin{align*}
\{T : NA_j\} & = \sum_{i=1}^{M} \{T_i : NA_j\}, \\
\{T : NA\} & = \sum_{j=1}^{N} \{T_i : NA_j\},
\end{align*} \quad (2.5)
\]

\[
\begin{align*}
[NA_j] & = [NAf_j] + \sum_{i=1}^{M} \{T_i : NA_j\}, \\
[Ti] & = [Tf_i] + \sum_{j=1}^{N} \{T_i : NA_j\},
\end{align*} \quad (2.6)
\]

\[
\begin{align*}
[NA] & = \sum_{j=1}^{N} [NA_j], \\
[T] & = \sum_{i=1}^{M} [T_i].
\end{align*} \quad (2.7)
\]

During the selection process, several methods of partitioning can be used for an effective separation of bound products from unbound NA molecules, such as nitrocellulose filtration or gel-shift analysis. (See [25], p. 389, for more details.) The ratio of bound \( NA_j, \{T : NA_j\} \), to free \( j^{th} \) NA, \([NAf_j]\), can be measured from the gel shift experiment ([26]) and defined as

\[
D_{j,f} = \frac{\{T : NA_j\}}{[NAf_j]} = \sum_{i=1}^{M} \frac{\{T_i : NA_j\}}{[NAf_j]} = \sum_{i=1}^{M} [Tf_i] A_{ij} = [Tf] \cdot \overrightarrow{A}_j, \quad \text{for} \quad j \in \mathcal{N}, \quad (2.8)
\]

where the equation (2.3) is used to express in term of concentrations of the free targets. We also define the fraction \( F_j \) of nucleic acid \( NA_j \) as

\[
F_j = \frac{[NA_j]}{[NA]},
\]
and the unit vector of NA fractions as $\vec{F} = \langle F_1, F_2, \ldots, F_N \rangle$, in the $L^1$ norm sense, i.e., $|\vec{F}| = \sum_{j=1}^{N} F_j = 1$ and $F_j$'s are all nonnegative, as well as the unit target and free target fraction vectors as

$$
\hat{\Omega} = \left\langle \frac{T_1}{T}, \ldots, \frac{T_M}{T} \right\rangle = (\Omega_1, \ldots, \Omega_M) \quad \text{and} \quad \hat{\omega} = \left\langle \frac{[Tf_1]}{[Tf]}, \ldots, \frac{[Tf_M]}{[Tf]} \right\rangle = (\omega_1, \ldots, \omega_M).
$$

Since the first equation in (2.6) and the equation (2.3) yield $[NA_j] = [NAf_j] \left( 1 + \sum_{i=1}^{M} [Tf_i]A_{ij} \right)$, the NA fractions $F_j$ can be written as

$$
F_j = \frac{[NA_j]}{[NA]} = \frac{[NAf_j]}{[NA]} \left( 1 + \sum_{i=1}^{M} [Tf_i]A_{ij} \right) = \frac{[NAf_j]}{[NA]} (1 + D_{j,f}). \quad (2.9)
$$

The total free nucleic acid $[NAf]$ is then computed as follows:

$$
[NAf] = \sum_{j=1}^{N} [NAf_j] = [NA] \sum_{j=1}^{N} \frac{F_j}{1 + D_{j,f}}, \quad (2.10)
$$

Once the bound NA species are eluted, the ratio of $j^{th}$ nucleic acids bound to the $i^{th}$ target species to the total nucleic acid is given by

$$
\frac{[[T_i : NA_j]]}{[[T : NA]]} = \frac{[NAf_j][Tf_i]A_{ij}}{[NA]} = \frac{F_j[Tf_i]A_{ij}}{1 + \sum_{i=1}^{M} [Tf_i]A_{ij}} = \frac{F_j[Tf_i]A_{ij}}{1 + D_{j,f}}. \quad (2.11)
$$

In addition, we can define the updated fractions $F'_j$ of the bound $NA_j$, for $j \in \mathcal{N}$, as

$$
F'_j \equiv \frac{[[T : NA_j]]}{[[T : NA]]} = \frac{[NA]}{[[T : NA]]} \left( \frac{D_{j,f}}{1 + D_{j,f}} \right) F_j = \frac{\sum_{l=1}^{N} \frac{F_jD_{j,f}}{1 + D_{j,f}} F_j}{\sum_{j=1}^{N} \sum_{i=1}^{M} F_jD_{j,f}} F_j, \quad (2.12)
$$

where we have used the formula from (2.11) that $[[T : NA]] = \sum_{j=1}^{N} \sum_{i=1}^{M} [[T_i : NA_j]] = [NA] \sum_{j=1}^{N} \frac{F_jD_{j,f}}{1 + D_{j,f}}$. As stated in [15], “the ratios $[[T : NA_j]]/[[T : NA]]$ do not change under PCR”, so after amplification, we can retain the updated NA fractions and use them for the next round of the selection step. Note also that $\sum_{j=1}^{N} F'_j = 1$.

To quantify the updated fractions in (2.12) we need to compute $D_{j,f} = \sum_{i=1}^{M} [Tf_i]A_{ij}$. That is, the $M$ free target concentrations $[Tf_i]$ are needed to be computed. The equation in (2.11) combined with the second equation in (2.6) then provides us a nonlinear system of equations.
for the $M$ free target concentrations to compute as follows:

$$[T_i] = [Tf_i] + \sum_{j=1}^{N}([T_i : NA_j]) = [Tf_i]\left(1 + [NA]\sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}}\right), \text{ for } i \in \mathcal{M}, \quad (2.13)$$

where $D_{j,f} = \sum_{l=1}^{M}[Tf_l]A_{lj}$. By taking the sum on both sides of the equations in (2.13) over all $M$ species, we have

$$[T] = [Tf]\left(1 + \frac{[NA]}{[Tf]} \sum_{j=1}^{N} \frac{F_j D_{j,f}}{1 + D_{j,f}}\right). \quad (2.14)$$

Using the unit target fraction vector notations, the equations (2.13) and (2.14) can also be written as

$$\Omega_i = \omega_i\left(1 + [NA]\sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}}\right)\frac{[Tf]}{[T]} = \omega_i\left(1 + [NA]\sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + [Tf] A_j \cdot \hat{\omega}}\right)\frac{[Tf]}{[T]} \quad (2.15)$$

for $i = 1, \ldots, M$, and

$$[T] = [Tf]\left(1 + [NA] \sum_{j=1}^{N} \frac{F_j \hat{A}_j} {1 + [Tf] A_j \cdot \hat{\omega}}\right). \quad (2.16)$$

Notice that, to this point, we have not used amplification in our formulation of the SELEX equations. However, when we reformulate the equations (2.12)-(2.16) as an iterative scheme, we utilize an amplification protocol such as PCR (in the case of nucleic acids) to restore the pool to its original size. Mathematically, this means that the nucleic acid pool size does not change from round to round.

### 2.2.3 Overall dissociation and association constants and efficiencies

Binding of nucleic acids against multiple target species or a target with multiple binding sites can often lead to very complex biochemical behaviors. Thus we consider the following overall reaction as well as overall sub reactions for better understanding of the multiple-target SELEX dynamical system:

$$\{T : NA\} \rightleftharpoons Tf + NAf, \quad (2.17)$$

$$\{T : NA_j\} \rightleftharpoons Tf + NAf_j, \quad j \in \mathcal{N}, \quad (2.18)$$

$$\{T_i : NA\} \rightleftharpoons Tf_i + NAf, \quad i \in \mathcal{M}. \quad (2.19)$$
We define the overall equilibrium dissociation constant for the overall reaction (2.17) as

\[ K_d = \frac{1}{K_a} = \frac{[T_f][NAf_j]}{[\{T_i : NA\}]} = \frac{[T_f] \sum_{j=1}^{N} \frac{F_j}{1 + D_{j,f}}}{\sum_{j=1}^{N} \frac{F_j D_{j,f}}{1 + D_{j,f}}} = \frac{\sum_{j=1}^{N} \frac{F_j}{1 + D_{j,f}}}{\sum_{j=1}^{N} \frac{F_j A_j \cdot \tilde{\omega}}{1 + D_{j,f}}} \quad (2.20) \]

where \( K_a \) is the corresponding overall association constant. The overall dissociation and association constants for the \( N \) individual sub reactions (2.18) of each nucleic acid with the overall target is given by

\[ K_{d,i} = \frac{1}{K_{a,i}} = \frac{[T_f][NAf_j]}{[\{T_i : NA\}]} = \frac{[T_f] \sum_{j=1}^{N} [NAf_j]}{\sum_{j=1}^{N} [\{T_i : NA\}]} = \frac{\sum_{j=1}^{N} \frac{F_j}{1 + D_{j,f}}}{\sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}}} \quad (2.21) \]

Last but not least, the overall dissociation and association constants for the \( M \) sub reactions (2.19) for each target component with the overall pool of nucleic acids can be written as

\[ \kappa_{d,i} = \frac{1}{\kappa_{a,i}} = \frac{[T_f][NAf_j]}{[\{T_i : NA\}]} = \frac{[T_f] \sum_{j=1}^{N} [NAf_j]}{\sum_{j=1}^{N} [\{T_i : NA\}]} = \frac{\sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}}}{\sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}}} \quad (2.22) \]

The quantity that is experimentally useful, as a measure of binding curve or saturation, is the fraction of target-NA binding. In relation to this quantity we now define overall target efficiency for the reaction (2.17) as follows:

\[ E = \frac{[T : NA]}{[T]} = \frac{[T] - [T_f]}{[T]} = 1 - \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j A_j \cdot \tilde{\omega}}{1 + [T_f] A_j \cdot \tilde{\omega}} \right)^{-1} \]

\[ = \frac{[NA] \sum_{j=1}^{N} \frac{F_j A_j \cdot \tilde{\omega}}{1 + D_{j,f}}}{[NA] \sum_{j=1}^{N} \frac{F_j A_j \cdot \tilde{\omega}}{1 + D_{j,f}} + 1} = \frac{[NA]}{K_d + [T_f] + [NA]} \quad (2.23) \]

The overall efficiency will also attain its maximum value when the available free target concentration reaches zero. It also tells us that, given a small value of \([T_f]\), \((E \approx \frac{[NA]}{K_d + [NA]})\), the higher the \( K_d \) (lower affinity), the weaker the binding \((E \approx 0)\). Similarly, the overall efficiencies of each nucleic acid for the sub reactions (2.18) can be defined by

\[ E_j = \frac{[T : NA_j]}{[NA_j]} = \frac{[NA_j] - [NAf_j]}{[NA_j]} = \frac{D_{j,f}}{1 + D_{j,f}} = \frac{[T_f] K_{a,i}}{1 + [T_f] K_{a,i}} \quad (2.24) \]

as well as the overall target efficiencies of each subtarget for the sub reactions (2.19) by

\[ E_i = \frac{[T_i : NA]}{[T_i]} = \frac{[T_i] - [T_{fi}]}{[T_i]} = 1 - \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + [T_f] A_j \cdot \tilde{\omega}} \right)^{-1} \]

\[ = \frac{[NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}}}{[NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}} + 1} = \frac{[NA]}{\left( \kappa_{a,i} \sum_{j=1}^{N} \frac{F_j}{1 + D_{j,f}} \right)} \quad (2.25) \]
These quantities of efficiencies are important thermodynamic parameters and are also referred to as binding probabilities ([8]) or as a measure of thermodynamic free energy or chemical potential. We will discuss the relationship between chemical potentials and equilibrium constants or overall efficiencies of these reactions later.

We also notice, from the equations (2.20), (2.22), (2.23) and (2.25), that the overall association constant $K_a$ and the overall efficiency $E$ can be written as the weighted averages of $M$ individual $\kappa_{a,i}$’s and $E_i$’s, i.e.,

$$K_a = \sum_{i=1}^{M} \omega_i \kappa_{a,i} \quad \text{and} \quad E = \sum_{i=1}^{M} \Omega_i E_i$$

(2.26)
since each weight sums to 1. From a thermodynamics point of view, these results are very reasonable. The overall association constant, $K_a$, is contributed by not only each individual association constant from each target component but also the corresponding free target species. Meanwhile, the overall target free energy is contributed by each individual efficiency of sub targets, as well as the corresponding target fractions. In addition, the target and free target fractions have the following relationship, in terms of the overall target efficiencies:

$$\omega_i = \left( \frac{1 - E_i}{1 - E} \right) \Omega_i.$$  

(2.27)

### 2.3 The SELEX iteration scheme

#### 2.3.1 Algorithm of the SELEX iteration scheme

We now propose the SELEX iteration scheme. To achieve this, we provide formulas at every round of iteration, indexed by the round number, $r$, and indicate what the formulas will be used for the $r^{th}$ round of selection and amplification steps. Computationally, the amplification allows us to choose the size of the total concentration of nucleic acids at every round.

First, we start with an initial target vector $\overrightarrow{T^{(1)}} = \langle [T_1^{(1)}], \ldots, [T_M^{(1)}] \rangle$ and give a pool of nucleic acids with the concentration of $[NA]$ and the initial NA fraction vector, $\hat{F}^{(1)} = \langle F_1^{(1)}, \ldots, F_N^{(1)} \rangle$. (Note that if we are given the total NA concentration in a pool, then the

---

2The ability for chemical systems to work. In chemical process, energy is required to break chemical bonds, and it is a measurable quantity in the laboratory when molecules absorb or release the energy as heat.
(fractional) distribution of nucleic acids and the values of dissociation constants can be “at least estimable”. See [15], p. 15.) Now, we compute the free target vector $\vec{T}_{f}^{(1)}$ by solving the nonlinear equation in (2.13), using fixed point iteration or Newton’s method. The updated NA fractions can be computed using the equation in (2.12), and these fractions will be available for the next round of selection. To perform the amplification mathematically, we fix the total concentration of nucleic acid molecules from round to round so that $[NA]^{(r)} = [NA]$. After repeating several rounds of selection-amplification, we have a target vector $\vec{T}^{(r)} = ([T_{1}^{(r)}], \ldots, [T_{M}^{(r)}])$ and the refined NA fraction vector $\vec{F}^{(r)} = (F_{1}^{(r)}, \ldots, F_{N}^{(r)})$ at the $r$th round.

The free target vector $\vec{T}_{f}^{(r)}$ and the updated NA fraction vector $\vec{F}^{(r+1)}$ for the next round, are again computed as follows:

$$[T_{i}^{(r)}] = [T_{f_{i}}^{(r)}] \left( 1 + [NA] \sum_{j=1}^{N} F_{j}^{(r)} A_{ij} \right), \quad \text{for} \quad i \in M, \quad (2.28)$$

and

$$F_{j}^{(r+1)} = \frac{D_{j,f}^{(r)}}{\sum_{l=1}^{N} D_{l,f}^{(r)}} F_{l}^{(r)}, \quad \text{for} \quad j \in N, \quad (2.29)$$

where $D_{j,f}^{(r)} = \sum_{i=1}^{M} [T_{l}^{(r)}] A_{ij}$. After successive rounds the pool is enriched with nucleic acids that bind to the targets with high affinity.

### 2.3.2 Optimal choice for the target concentration

In [15], the optimal strategy for choosing the total target concentration was discussed for the single target case. It was shown that reducing the target concentration (the target dilution) from round to round achieves the maximum target efficiency for the process. Quoting from [15], (p. 16), “maximum probability for binding the best binder occurs when the free target is small while the probability of binding the poorest binder will be at a minimum when the free target is small”. It is also mentioned “the fraction of best binding molecules in the pool should increase relative to the others because of the greater likelihood that they will be bound to the target than those of lower affinity. But, one might miss the best binder altogether as one lowers the target.” Here we adopt the same strategy for multiple-target SELEX. The fraction
of each target component, however, does not change from round to round. That is, we take
\[ T^{(r+1)} = (1 - s_r)T^{(r)}, \] where \( 0 \leq s_r < 1 \) and \( \hat{\Omega}^{(r)} = \hat{\Omega}. \) This allows that \( T^{(r)} \to 0 \) as the round number \( r \) increases. Therefore, \( Tf_i^{(r)} \to 0 \) for \( i \in \mathcal{M} \) as well as \( Tf^{(r)} \to 0. \) We will use this choice of the target dilution through the analysis of the (positive) SELEX process and make a series of observations, at the limiting states. Simulation results also show that if the size of total target is fixed at every round, the iteration process requires many more rounds to achieve the selection than a case of target dilution, for multiple targets. See Figure 2.3, panels (a) and (b).

### 2.4 Limiting states of the SELEX iteration scheme

#### 2.4.1 Limiting values of NA fractions as the round number becomes large and their asymptotic stability

The goal of analyzing our iteration scheme is to characterize the nucleic acids that survive or become eliminated from the pool after successive rounds of the SELEX process, that is, which NA species bind to targets with high affinity and specificity.

First, we observe the ratio of the \( j^{th} \) nucleic acid fraction to the \( l^{th} \), for \( j \neq l \), after \( r \) rounds,
\[
\frac{F_j^{(r+1)}}{F_l^{(r+1)}} = \frac{D_{j,l}^{(r)} (1 + D_{l,j}^{(r)}) F_j^{(r)}}{D_{l,j}^{(r)} (1 + D_{j,l}^{(r)}) F_l^{(r)}}, \quad \text{for} \quad j, l \in \mathcal{N}, \tag{2.30}
\]
using the equation (2.29). This ratio will be smaller after one round of SELEX than before the round if and only if \( D_{j,l}^{(r)} > D_{l,j}^{(r)} \), or equivalently, \( \bar{A}^l \cdot \hat{\omega}^{(r)} > \bar{A}^j \cdot \hat{\omega}^{(r)}. \) We will use this observation to obtain more information about which indices correspond to nucleic acids that are eventually eliminated from the pool and those that survive.

Using the iteration scheme the ratio can also be written in terms of the initial vector of nucleic acids, \( \hat{F}^{(1)} \),
\[
\frac{F_j^{(r+1)}}{F_l^{(r+1)}} = \left( \prod_{p=1}^r \frac{D_{j,p}^{(p)} (1 + D_{l,j}^{(p)})}{D_{l,j}^{(p)} (1 + D_{j,l}^{(p)})} \right) \frac{F_j^{(1)}}{F_l^{(1)}}, \tag{2.31}
\]
In order to have selection, i.e., convergence of the SELEX scheme to a unique set of final NA fractions, independently of the initial fractional distribution of nucleic acids such that
$F_j^{(1)} > 0$, for all $j \in \mathcal{N}$, we need the notion of global asymptotic stability with respect to such pools. We say that the SELEX process is globally asymptotically stable if for every $\hat{\Omega}$ there exists a limiting set of nucleic acids $\hat{F}$ such that $\lim_{r \to \infty} |\hat{F}(r) - \hat{F}| = 0$, independently of the choice of $\hat{F}^{(1)}$ (i.e., the initial pool with all nucleic acids present).

### 2.4.2 Convergence of nucleic acid fractions

First, we need to show that all the (uniformly bounded) sequences $\{F_j^{(r)}\}_{r=1}^{\infty}$ converge to some limiting value, not all of which can be zero. Here, we give a proof of convergence. A different approach to the convergence proof is provided in [23]. (See Appendix B for the proof of convergence provided in [23].)

We prove the convergence of the sequences $\{F_j^{(r)}\}_{r=1}^{\infty}$ from a thermodynamic point of view. First, it follows from the normalization condition that the sequences $0 \leq F_j^{(r)} \leq 1$ and $\sum_{j=1}^{N} F_j^{(r)} = 1$, for every $r$. The usual physical assumption for the reactions, in thermodynamics, is that they proceed in such a way as to maximize the target efficiencies, both overall and for each target. In mathematical terms, this means that the following limits do exist:

$$E = \lim_{r \to \infty} E^{(r)} = 1 - \lim_{r \to \infty} \frac{[Tf^{(r)}]}{[T^{(r)}]} = \lim_{r \to \infty} \frac{[NA]}{K_a^{(r)} + [NA]} = \lim_{r \to \infty} \frac{[NA]K_a^{(r)}}{1 + [NA]K_a^{(r)}} = 1 - W \quad (2.32)$$

where we have used that $[Tf^{(r)}] \to 0$ as $r \to \infty$, with the equation (2.23), and

$$E_i = \lim_{r \to \infty} E_i^{(r)} = 1 - \lim_{r \to \infty} \frac{[Tf_i^{(r)}]}{[T_i^{(r)}]} = 1 - V_i. \quad (2.33)$$

Recall from the previous section that each target is diluted in such a way that $\overrightarrow{T^{(r+1)}} = (1 - s_r)\overrightarrow{T^{(r)}}$ where $0 \leq s_r < 1$ so that $[T_i^{(r)}] \to 0$ as the round number, $r$, increases, as well as $[Tf_i^{(r)}] \to 0$. Also, $[T_i^{(r)}] = \prod_{p=1}^{r-1} (1 - s_p)[T_i^{(1)}]$ for $i \in \mathcal{M}$.

**Theorem 2.4.2.1.** Suppose that the limits, $\lim_{r \to \infty} \frac{[Tf_i^{(r)}]}{[T_i^{(r)}]} = V_i$ exist for all $i \in \mathcal{M}$. Then each nucleic acid fraction $F_j^{(r)}$ converges for every $j \in \mathcal{N}$. 

Proof. Since \( \lim_{r \to \infty} \frac{[Tf_i^{(r)}]}{[T_i^{(r)}]} = \nu_i \), we can write
\[
[Tf_i^{(r)}] \approx \nu_i[T_i^{(r)}] = \nu_i \cdot \prod_{p=1}^{r-1} (1 - s_p)[T_i^{(1)}] \quad \text{where } 0 \leq s_p < 1 \text{ and given } [T^{(1)}],
\]
for large \( r \). Then \( D_{j,f}^{(r)} \) can be approximated as follows, for large \( r \) and all \( j \in \mathcal{N} \):
\[
D_{j,f}^{(r)} = [Tf^{(r)}] \cdot \overline{A}j = [Tf_1^{(r)}, Tf_2^{(r)}, \ldots, Tf_M^{(r)}] \cdot \overline{A}j
\]
\[
\approx \prod_{p=1}^{r-1} (1 - s_p)[\nu_1T_1^{(1)}, \nu_2T_2^{(1)}, \ldots, \nu_M T_M^{(1)}] \cdot \overline{A}j \equiv \prod_{p=1}^{r-1} (1 - s_p)a_j,
\]
where \( a_j = [\nu_1T_1^{(1)}, \nu_2T_2^{(1)}, \ldots, \nu_M T_M^{(1)}] \cdot \overline{A}j, \quad j \in \mathcal{N} \).

Without loss of generality we can assume that \( a_j \)'s are ordered: \( a_1 \leq a_2 \leq \cdots \leq a_N \), and \( a_N > 0 \). Then for large \( r \),
\[
D_{1,f}^{(r)} \leq D_{2,f}^{(r)} \leq \cdots \leq D_{N,f}^{(r)},
\]
which implies,
\[
\frac{D_{1,f}^{(r)}}{1 + D_{1,f}^{(r)}} \leq \frac{D_{2,f}^{(r)}}{1 + D_{2,f}^{(r)}} \leq \cdots \leq \frac{D_{N,f}^{(r)}}{1 + D_{N,f}^{(r)}}.
\]
Thus,
\[
F_{N}^{(r+1)} = \frac{\sum_{l=1}^{N} \frac{D_{l,f}^{(r)}}{1 + D_{l,f}^{(r)}} F^{(r)}_l}{\sum_{l=1}^{N} \frac{D_{l,f}^{(r)}}{1 + D_{l,f}^{(r)}} F^{(r)}_l} F^{(r)}_N \geq F^{(r)}_N
\]
for large \( r \). Since \( F_N^{(r)} \) is increasing in \( r \) and bounded above, \( F_N^{(r)} \to F_N > 0 \).

For \( j \in \{1, 2, 3, \ldots, N - 1\} \), we have, from the equation (2.29), that
\[
F_j^{(r+1)} = \frac{D_{j,f}^{(r)}}{1 + D_{j,f}^{(r)}} F^{(r)}_j = \frac{b_j^{(r)} F_j^{(r)}}{b_N^{(r)} F_N^{(r)}},
\]
where \( b_j^{(r)} = \frac{D_{j,f}^{(r)}}{1 + D_{j,f}^{(r)}}. \) Then for large \( r \)
\[
\frac{F_j^{(r+1)}}{F_N^{(r+1)}} = \left[ \frac{b_j^{(1)} b_j^{(2)} \cdots b_j^{(k-1)}}{b_N^{(1)} b_N^{(2)} \cdots b_N^{(k-1)}} \right] \cdot \left[ \frac{b_j^{(k)} \cdots b_j^{(r)}}{b_N^{(k)} \cdots b_N^{(r)}} \right] \cdot \frac{F_j^{(1)}}{F_N^{(1)}} = P \cdot C_r \cdot \frac{F_j^{(1)}}{F_N^{(1)}}
\]
where \( C_r = \frac{b_j^{(k)} \cdots b_j^{(r)}}{b_N^{(k)} \cdots b_N^{(r)}} \) and for some fixed \( P = \frac{b_j^{(1)} b_j^{(2)} \cdots b_j^{(k-1)}}{b_N^{(1)} b_N^{(2)} \cdots b_N^{(k-1)}} \). Moreover, \( 0 < C_r \leq 1 \) since \( b_j^{(k)} \leq b_N^{(k)}, \ldots, b_j^{(r)} \leq b_N^{(r)} \). Therefore \( \frac{F_j^{(r)}}{F_N^{(r)}} \) is decreasing in \( r \) and it converges. Since \( F_N^{(r)} \) converges, \( F_j^{(r)} \) converges for all \( 1 \leq j \leq N - 1 \). Thus \( F_j^{(r)} \) converges for all \( j \in \mathcal{N} \). \( \square \)

### 2.4.3 Limiting free target fractions and target efficiencies

Because the sequences \( \{ F_j^{(r)} \}_{r=1}^\infty \) converge, let \( F_j = \lim_{r \to +\infty} F_j^{(r)} \) and partition the set of indices \( \mathcal{N} \) into a set \( \mathcal{L}' \) for which the limit is not zero and its complement, \( \mathcal{J}' \), for which the limit is zero, i.e., \( \mathcal{N} = \mathcal{L}' \cup \mathcal{J}' \). We will discuss determining these sets at the limiting states later.

Using the equation (2.28), we obtain:
\[
\frac{\omega_i^{(r)} [Tf^{(r)}]}{[T^{(r)}]} = \frac{\Omega_i}{1 + [NA] \hat{A}_i \cdot \hat{F}^{(r)}},
\]
where we have used the fact that \( [Tf_1^{(r)}] \to 0 \) if and only if \( [Tf^{(r)}] \to 0 \). Now, let \( \mathcal{W}^{(r)} = \frac{[Tf^{(r)}]}{[T^{(r)}]} \) and \( \mathcal{W}_i^{(r)} = \omega_i^{(r)} \mathcal{W}^{(r)} \). The existence of the limiting NA fraction vector allows that the limit of the right-hand side of equation (2.34) does exist. Hence the following limits
\[
\lim_{r \to +\infty} \mathcal{W}_i^{(r)} = \lim_{r \to +\infty} \omega_i^{(r)} \mathcal{W}^{(r)} = \frac{\Omega_i}{1 + [NA] \hat{A}_i \cdot \hat{F}} = \mathcal{W}_i, \quad i \in \mathcal{M},
\]
also exist. Because \( \sum_{i=1}^M \mathcal{W}_i^{(r)} = \mathcal{W}^{(r)} \), the limit
\[
\lim_{r \to +\infty} \mathcal{W}^{(r)} = \sum_{i=1}^M \frac{\Omega_i}{1 + [NA] \hat{A}_i \cdot \hat{F}} = \mathcal{W}
\]
exists. The limiting free target fractions are then given by
\[
\lim_{r \to +\infty} \omega_i^{(r)} = \frac{\mathcal{W}_i}{\mathcal{W}} \equiv \omega_i.
\]

Note that the assumption of the existence of the following limits,
\[
E = \lim_{r \to +\infty} E^{(r)} = 1 - \lim_{r \to +\infty} \frac{[Tf^{(r)}]}{[T^{(r)}]} = 1 - \mathcal{W} \quad \text{and} \quad \lim_{r \to +\infty} \omega_i^{(r)} = \omega_i,
\]

(2.38)
does not necessarily imply the uniqueness of the limiting NA fractions from (2.34) because the affinity matrix \( A \) may not be of full rank. Notice also from the equation (2.27) that the existence of any two of the limits, among \( E, \omega_i, \) and \( E_i \), implies the existence of the third.

Finally, for the limiting value of the overall association constant \( K_a \) we have:

\[
K_a = \lim_{r \to +\infty} K^{(r)}_a = \lim_{r \to +\infty} \sum_{i=1}^{M} \omega_i^{(r)} \kappa_{a,i}^{(r)} = \frac{1}{\mathbb{W}} \sum_{i=1}^{M} \frac{\Omega_i \bar{A}_i \cdot \bar{F}}{1 + [NA] \bar{A}_i \cdot \bar{F}} \quad \text{since} \quad \kappa_{a,i}^{(r)} \to \bar{A}_i \cdot \bar{F}. 
\] (2.39)

Also, we have:

\[
\lim_{r \to +\infty} \frac{D_{j,f}^{(r)}}{D_{l,f}^{(r)}} = \lim_{r \to +\infty} \frac{K_{a,j}^{(r)}}{K_{a,l}^{(r)}} = \frac{\bar{A}_j \cdot \bar{W}}{\bar{A}_l \cdot \bar{W}},
\] (2.40)

where \( \bar{W} = (\mathbb{W}_1/\mathbb{W}, \ldots, \mathbb{W}_M/\mathbb{W}) \). As mentioned in section 2.4.1, the limiting behavior of the quantities \( D_{j,f} \) or \( \bar{A}_j \cdot \bar{W} \) contributes to the nucleic acids selection as well as the binding affinities of the overall sub reactions.

### 2.5 The geometry of the SELEX problem

Before we discuss the issue of determining the set of selected nucleic acid indices at the limiting states, we mention the underlying geometry of the SELEX problem here.

First, define the two simplices for the sets of all possible fraction vectors of the target (or free target) and nucleic acids:

\[
S = \{ \bar{\omega} = (\omega_1, \ldots, \omega_M) | \omega_i \geq 0, \ i = 1, \ldots, M \quad \text{and} \quad \sum_{i=1}^{M} \omega_i = 1 \}, \]

\[
S_F = \{ \bar{f} = (f_1, \ldots, f_N) | f_i \geq 0, \ i = 1, \ldots, N \quad \text{and} \quad \sum_{i=1}^{N} f_i = 1 \}. \] (2.41)

Define also the open subset of \( S \), say \( S_0 \), as follows:

\[
S_0 = \{ \bar{\omega} \in S \ | \ \omega_i > 0, \ i = 1, \ldots, M \}. \] (2.42)

To consider sub simplices associated with limiting nucleic acid fraction vectors, for any subset \( \mathcal{L} \subset \mathcal{N} \), define the simplex,

\[
S_{F,\mathcal{L}} = \{ \bar{f} \in S_F \ | \ f_i \geq 0 \ \text{for} \ i \in \mathcal{L}, \ f_i = 0 \ \text{if} \ i \in \mathcal{N} - \mathcal{L} \}. \] (2.43)
For each positive number $\tau_f$ and each pair of vectors $(\hat{\Omega}, \hat{f}) \in S \times S_F$ the equations (2.15), (2.16) and (2.12) can be written as the following $N + M + 1$ functions:

$$
\tau(\tau_f, \hat{\omega}, \hat{f}) = \tau_f \left(1 + [N\Lambda] \sum_{j=1}^{N} \frac{f_j \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}}}{1 + \tau_f \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}}}\right),
$$

$$
\Omega_i(\tau_f, \hat{\omega}, \hat{f}) = \omega_i \left(1 + [N\Lambda] \sum_{j=1}^{N} \frac{A_{ij} f_j}{(1 + \tau_f \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}})} \right) \frac{\tau_f}{\tau}, \quad i = 1, \ldots, M,
$$

$$
\hat{f}_j(\tau_f, \hat{\omega}, \hat{f}) = \left(\frac{\tau_f \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}} f_j}{1 + \tau_f \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}}} \right) \left(\sum_{l=1}^{N} \frac{\tau_f \overrightarrow{A^l} \cdot \overrightarrow{\hat{\omega}}}{1 + \tau_f \overrightarrow{A^l} \cdot \overrightarrow{\hat{\omega}}} f_l\right)^{-1}, \quad j = 1, \ldots, N.
$$

This defines a mapping from $\mathbb{R} = [0, \infty) \times S \times S_F$ into itself. (Notice that $\sum_i \omega_i = \sum_j \hat{f}_j = 1$.)

In the SELEX problem, given $(\tau, \hat{\Omega}, \hat{f})$, we seek the selection vector $(\tau_f, \hat{\omega}, \hat{f})$. Here we approach this problem by solving the first $M + 1$ equations to find $(\tau_f, \hat{\omega})$ and using them to compute $\hat{f}$.

We consider the final values of the SELEX iteration scheme as the fixed points of the above map.

Let $\hat{f}$ be a fixed point of the above map (i.e. a SELEX solution). Fix $[N\Lambda]$ and $\hat{\Omega}$, and let $\tau \to 0$ in such a way that the ratio $\frac{\tau f}{\tau} \to W < 1$ for some positive number $W$. (Note that $\tau_f$ also approaches 0.) Then

$$
\omega_i \to \frac{\Omega_i}{W(1 + [N\Lambda] \overrightarrow{A_i} \cdot \overrightarrow{\hat{f}})},
$$

and the updated fractions

$$
\hat{f}_j \to \frac{f_j \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}}}{\sum_{l=1}^{N} f_l \overrightarrow{A^l} \cdot \overrightarrow{\hat{\omega}}},
$$

Since we are assuming that the vector $\hat{f}$ is a fixed point, $\hat{f}_j = f_j$ for all indices. If $f_j > 0$ (i.e., $j^{th}$ NA is selected) then

$$
\overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}} = 1 \quad \text{or} \quad \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}} = \sum_{l=1}^{N} f_l \overrightarrow{A^l} \cdot \overrightarrow{\hat{\omega}} = E,
$$

where the efficiency of the process $E = \frac{\tau - \tau_f}{\tau} = \frac{[N\Lambda] \sum_j f_j \overrightarrow{A^j} \cdot \overrightarrow{\hat{\omega}}}{[N\Lambda] \sum_l \hat{f}_l \overrightarrow{A^l} \cdot \overrightarrow{\hat{\omega}} + 1}$. This tells us that, for the affinity vectors $\overrightarrow{A^j}$ corresponding to the $j^{th}$ selected nucleic acids (i.e. $f_j > 0$), they all have the same projection onto $\overrightarrow{\hat{\omega}}$. This argument, however, does not tell us the final fraction values, but it provides an important characteristic of the selection indices, which we will discuss from the next section.
2.6 Solution of the SELEX problem

2.6.1 The solution of the SELEX problem at infinite NA dilution \([NA] \approx 0\)

In chemistry, “infinite dilution” is a concept of the amount (concentration) of the solvent and solute\(^3\) in a solution or a mixture. As one adds more solvent to a solution, it becomes more dilute and approaches infinite dilution. By completing this, one can see whether a certain property of the solution originates from the solute or not. By examining the SELEX equations (2.28) and (2.29), when \([NA]\) reaches zero at the \(r^{th}\) round we have:

\[
[T_i^{(r)}] = [Tf_i^{(r)}], \text{ for } l \in \mathcal{M}, \text{ and thus } [T^{(r)}] = [Tf^{(r)}].
\]

This implies that \(\omega_i^{(r)} = \Omega_i^{(r)} = \Omega_i\) since we keep the target fractions the same from round to round. Equations (2.32) and (2.33) also tell us that the total and individual efficiencies after \(r\) rounds will be zero. Thus, the bindings become weaker as the total NA pool size decreases. Since \(D_{j,f}^{(r)} = [Tf^{(r)}] \hat{A}_j \cdot \hat{\omega}^{(r)} = [T^{(r)}] \hat{A}_j \cdot \hat{\Omega}\), the ratios \(\frac{D_{j,f}^{(r)}}{D_{l,f}^{(r)}} = \frac{\hat{A}_j \cdot \hat{\Omega}}{\hat{A}_l \cdot \hat{\Omega}}\) and do not depend on the round number any more. This means, if \(l \in \mathcal{L}\) and \(j \notin \mathcal{L}\), where \(\mathcal{L} = \{l \in \mathcal{N} | \hat{A}_l \cdot \hat{\Omega} = \max\{\hat{A}_k \cdot \hat{\Omega}, k \in \mathcal{N}\}\}\),

\[
\frac{F_j^{(r+1)}}{F_l^{(r+1)}} = \frac{D_{j,f}^{(r)}}{D_{l,f}^{(r)}} \frac{1 + D_{j,f}^{(r)}}{1 + D_{j,f}^{(r)}} \frac{F_j^{(r)}}{F_l^{(r)}} = \left(\frac{\hat{A}_j \cdot \hat{\Omega}}{\hat{A}_l \cdot \hat{\Omega}}\right)^r \frac{F_j^{(1)}}{F_l^{(1)}} \quad (2.45)
\]

approaches zero and therefore \(F_j^{(r)} \to 0\) for \(j \notin \mathcal{L}\). If both indices, \(j\) and \(l\), are in the set \(\mathcal{L}\), then \(F_j^{(r+1)} = F_l^{(1)} F_j^{(1)} / F_l^{(1)} \) after \(r\) rounds. This means that the characteristic of the initial pool would remain the same as the final pool of nucleic acids. That is, for a small value of NA concentration such that \([NA] << [T^{(1)}]\), the SELEX process may require many rounds to reach its selection. See Figure 2.5 for an illustration of values of \(K_a^{(r)}\) versus round numbers when the NA pool size becomes very small.

2.6.2 The solution of the SELEX problem at finite NA dilution \([|NA| > 0]\)

When the nucleic acid concentration is positive, the problem becomes mathematically more complicated than in the case of infinite NA dilution. Based on the observations made in the

\(^3\)The substance dissolved in solvent.
preceding sections, consider the following set, at every round,
\[ C_f^{(r)} = \{ \vec{A}^1 \cdot \hat{W}^{(r)}, \vec{A}^2 \cdot \hat{W}^{(r)}, \ldots, \vec{A}^N \cdot \hat{W}^{(r)} \} , \]
and denote the limiting set as
\[ C = \{ \vec{A}^1 \cdot \hat{W}, \vec{A}^2 \cdot \hat{W}, \ldots, \vec{A}^N \cdot \hat{W} \} , \]
where the limit values are defined in (2.35). Then, there is a subset, depending on the target fractions and the NA concentration, \( C(\hat{\Omega}, [NA]) \subset C \), say. Define the set of indices in \( C(\hat{\Omega}, [NA]) \) as \( \mathcal{L}(\hat{\Omega}, [NA]) \) and its complementary set as \( \mathcal{J}(\hat{\Omega}, [NA]) \).

Suppose, for some fixed \( \delta > 0 \), that \( \vec{A}^l \cdot \hat{W} \geq (1 + \delta) \vec{A}^j \cdot \hat{W} \). From (2.40), for all sufficiently large round numbers \( r \),
\[ \frac{D_{j,f}^{(r)}}{D_{l,f}^{(r)}} \approx \frac{\vec{A}^j \cdot \hat{W}}{\vec{A}^l \cdot \hat{W}} . \]
This again leads us that, for \( l \in \mathcal{L}(\hat{\Omega}, [NA]) \) and \( j \in \mathcal{J}(\hat{\Omega}, [NA]) \),
\[ \lim_{r \to \infty} F_j^{(r)} = 0 . \]
The result does not imply that all nucleic acid indices in the set \( \mathcal{L}(\hat{\Omega}, [NA]) \) are necessarily selected at the limiting state, but the indices in the set \( \mathcal{J}(\hat{\Omega}, [NA]) \) will not be selected. That is, we cannot say that \( \lim_{r \to \infty} F_l^{(r)} = F_l > 0 \) for all \( l \in \mathcal{L}(\hat{\Omega}, [NA]) \). The set \( \mathcal{L}' \) defined in section 2.4.3, however, is a subset of \( \mathcal{L}(\hat{\Omega}, [NA]) \).

2.6.3 Selected indices of the SELEX iteration scheme: The maximal target affinity function

Finally, we establish the maximal set of the SELEX indices for the study of final nucleic acid species selected. The same definitions and notations in [23] are also used here.

From the equation (2.40), the numbers \( \vec{A}^j \cdot \hat{\omega} \) can be interpreted as possible overall association constants for each nucleic acid relative to the target vector, i.e., as values of \( K_{a,j} = 1/K_{d,j} \).

**Definition 2.6.3.1.** On the simplex \( S \) defined in (2.41), we define the (continuous) convex function as follows:
\[ \varphi(\hat{\omega}) = \max \left\{ \vec{A}^j \cdot \hat{\omega} \mid j \in \mathcal{N} \right\} \quad \text{for} \quad \hat{\omega} \in S . \] (2.46)
We call this function the maximal target affinity function.

By defining this function, we can interpret the value \( \varphi(\hat{\omega}) \) as the affinity of each selected nucleic acid for the limiting target vector as a function of the individual components of \( \hat{\omega} \) on the simplex \( S \). The minimum value of \( \varphi \) can then be interpreted as the smallest possible overall affinity at selection. Let \( \varphi_{\text{min}} \) denote this positive minimum value and \( \varphi_{\text{max}} \) its maximum value on the simplex. Corresponding to each target vector \( \hat{\omega} \) on the simplex \( S \), there is a unique set of nucleic acid indices \( \mathcal{L}(\hat{\omega}) \subset \mathcal{N} \) for which \( \varphi(\hat{\omega}) = \vec{A}_j \cdot \hat{\omega} \) for \( j \) in this set.

We denote the number of elements of indices in \( \mathcal{L}(\hat{\omega}) \) by \( L(\hat{\omega}) \) and its complement by \( J(\hat{\omega}) \) corresponding to the complementary set \( J(\hat{\omega}) \).

Now define the level sets of \( \varphi \) as follows. For each positive number \( K_a \in [\varphi_{\text{min}}, \varphi_{\text{max}}] \), the set

\[
\mathcal{L}_{K_a} = \{ l \in \mathcal{N} \mid K_a = \vec{A}_l \cdot \hat{\omega} = \varphi(\hat{\omega}) \} = \bigcup_{\varphi(\hat{\omega}) = K_a} \mathcal{L}(\hat{\omega})
\]

is the maximum possible set of nucleic acid indices that can be selected when \( \varphi = K_a \). Thus if we define the set of integers,

\[
\mathcal{L}_0 = \bigcup_{\varphi_{\text{min}} \leq K_a \leq \varphi_{\text{max}}} \mathcal{L}_{K_a} = \bigcup_{\hat{\omega} \in S} \mathcal{L}(\hat{\omega}),
\]

this tells that a nucleic acid, \( NA_j \) is selectable for some initial target vector if and only if \( j \in \mathcal{L}_0 \).

We note that, for two different limiting free target vectors \( \hat{\omega} \) and \( \hat{\omega}' \), there is a possibility that \( K_a = \vec{A}_l \cdot \hat{\omega} = \varphi(\hat{\omega}) \) as well as \( K_a = \vec{A}_j \cdot \hat{\omega}' = \varphi(\hat{\omega}') \), for \( l, j \in \mathcal{L}_{K_a} \). If the limiting free target vector \( \hat{\omega} \) corresponds to the \( l^{th} \) nucleic acid such that \( F^{(r)}_l \to F_l > 0 \), and \( \vec{A}_l \cdot \hat{\omega} > \vec{A}_j \cdot \hat{\omega} \), then the \( j^{th} \) nucleic acid will not be selected for such a choice of \( \hat{\omega} \). Just because an index is in \( \mathcal{L}_{K_a} \) does not mean that the nucleic acid corresponding to it will ultimately be selected, even if it is present in the initial pool.

The graph of the convex function \( \varphi \), on the simplex \( S \), constructs closed faces from the intersections of hyperplanes (points, lines or higher dimensional planes). Let us now define closed faces of the graph using the maximal target affinity function.
**Definition 2.6.3.2.** Let $\mathcal{L}$ be an increasingly ordered subset of $\mathcal{N}$ with $L$ elements. Define an $L$ face of the graph of $\varphi$ by the set

\[
\Phi(\mathcal{L}) = \{ (\tilde{\omega}, \varphi(\tilde{\omega})) \mid \overline{A^l} \cdot \tilde{\omega} = \varphi(\tilde{\omega}) \text{ where } \tilde{\omega} \in \mathcal{S} \text{ and } l \in \mathcal{L} \}
\]  

(2.48)

where the over line denotes the closure of the set below it (including its limit points) and where the set of vectors $\{ A^l \mid l \in \mathcal{L} \}$ is linearly independent.

We say an $L$ face is *proper* if $\Phi(\mathcal{L}) = \Phi(\mathcal{L}')$ implies $\mathcal{L} = \mathcal{L}'$, that is, the set of indices that describes it is uniquely determined. If an $L$ face is proper, then the set $\mathcal{L}$ is maximal with respect to the defining property. However, a set can be maximal with respect to this property without being unique if the face is not proper. We refer to this as an improper face.

As a simple example of improper faces, consider four planes in 3 dimensions whose normals have only positive components and with the property that the normals of any three of them are linearly independent. Suppose they intersect at a common point in $\mathcal{S}$. Then they are 4 ordered subsets of three integers that describe this point. This point is an improper 3 face. Likewise, suppose three such planes intersect at a single point. Suppose a fourth plane contains a line of intersection of two of the other planes but does not pass through the dihedral angle made by them. Then the line of intersection is an improper two face also. See Figure 2.8, panel (c) for an example of the improper case.

**Definition 2.6.3.3.** If $\varphi$ has the property that every $L$ face is proper for $L = 1, 2, \ldots M$, we say that the maximal target affinity function is proper. Otherwise we say the maximal target affinity function is improper.

Again when the maximal target affinity function $\varphi$ is improper, uniqueness of the indices fails. That is, the SELEX process will no longer converge to a unique set of final nucleic acids for each choice of the initial target vector $\hat{\Omega}$ and any initial pool of nucleic acids with all nucleic acid species present. Equivalently, the system will converge to a pool that depends on the starting nucleic acid fractions. This loss of uniqueness can also lead to the possibility that the number of elements, $L_0$, in the set $\mathcal{L}_0$ can be larger than the number of target species, $M$. 

This means, for the multiple target case, it is possible to select for more nucleic acids than targets. Again, see Figure 2.8 panels (c, d) for the illustration of improperness.

Next we define submatrices, corresponding to the selected nucleic acid indices, of the affinity matrix $A$ as follows:

**Definition 2.6.3.4.** Given the affinity matrix $A$, suppose that there is a final free target vector $\hat{\omega}$ such that $\varphi(\hat{\omega}) = K_a \geq \varphi_{\min}$ and belongs to some $L$ face. Let $\mathcal{L}$ be the index set for this $L$ face. Let $A_L$ be the sub matrix of $A$ that consists only of those columns whose column index is in $\mathcal{L}$. Let its complement denote $\mathcal{J}$ and $A_J$ have a similar meaning. Then we call the matrix $A_L$ the affinity selection matrix (ASM) for the face with index set $\mathcal{L}$ and the matrix $A_J$ the complement of this matrix (CASM).

Using this definition the affinity matrix $A$ can be reordered as:

$$A = [A_L, A_J]$$

indicating that the first $L$ columns correspond to the selected nucleic acid species.

We also summarize the (limiting) sets of nucleic acid indices defined above using the following inclusion relationship:

$$\mathcal{L}' \subseteq L(\hat{\Omega}, [NA]) \subset \mathcal{L}.$$

While the condition of the indices $l \in \mathcal{L}$ is necessary for nucleic acids to be selected for a given pool, the set $L(\hat{\Omega}, [NA])$ such that $F_l > 0$ is the selected set of indices for a given initial target vector and a total NA concentration.

### 2.6.4 A system of nonlinear SELEX equations: The relation of final nucleic acid fractions and final target fractions

Suppose that the only nucleic acids that can be selected are those whose indices belong to the set $\mathcal{L}$. Then from (2.35) we have

$$W_i = \frac{\Omega_i}{1 + [NA] \sum_{j=1}^{N} F_j A_{ij}} = \frac{\Omega_i}{1 + [NA] \sum_{j \in \mathcal{L}} F_j A_{ij}} = \frac{\Omega_i}{1 + [NA] \sum_{j \in \hat{\mathcal{L}(\hat{\Omega}, [NA])}} F_j A_{ij}}, \quad (2.49)$$
where \( j \in \mathcal{L} \) is one of the selected nucleic acid indices. Equivalently, we can write (2.49) as

\[
[NA] \vec{A}_i \cdot \vec{F} = \frac{\Omega_i}{\tilde{W}_i} - 1, \quad i \in \mathcal{M}.
\] (2.50)

Let \( \varphi(\tilde{W}) \) denote the maximal target affinity function such that \( \varphi(\tilde{W}) = \max\{\vec{A}_j \cdot \tilde{W} \mid j \in \mathcal{L}\} \) where \( \tilde{W} = \tilde{W}/W \). Given a target vector \( \tilde{\Omega} \) and a concentration of nucleic acid \([NA]\), the SELEX process selects a vector \( \vec{F} \in S_{F,\mathcal{L}} \) such that

\[
\vec{A}_j \cdot \vec{F} = \varphi(\tilde{W})(\vec{F}) \quad \text{where} \quad \tilde{W}_i(\vec{F}) = \frac{\Omega_i}{1 + [NA]\vec{A}_i \cdot \vec{F}}.
\] Then multiplying both sides of this equation by \( F_j \geq 0 \) (but \( F_j > 0 \) on \( \mathcal{L}(\tilde{\Omega}, [NA]) \)), summing over \( j \in \mathcal{L} \), expressing the dot product as a sum over \( i \) and interchanging the order of summation on the left, we find that

\[
\sum_{i=1}^{M} \frac{[NA]\Omega_i \vec{A}_i \cdot \vec{F}}{1 + [NA]\vec{A}_i \cdot \vec{F}} = [NA]\varphi(\tilde{W}(\vec{F}))\tilde{W}(\vec{F}).
\] (2.51)

Using equations (2.50) and (2.51) we derive that

\[
[NA]\tilde{W}(\vec{F})\varphi(\tilde{W}(\vec{F})) = 1 - \tilde{W}(\vec{F}) \quad \text{and} \quad \tilde{W}(\vec{F}) = \frac{1}{1 + [NA]\varphi(\tilde{W}(\vec{F}))}.
\] (2.52)

Thus

\[
\frac{[NA]K_a}{1 + [NA]K_a} = E = 1 - \tilde{W} = \frac{[NA]\varphi(\tilde{W}(\vec{F}))}{1 + [NA]\varphi(\tilde{W}(\vec{F}))},
\] (2.53)

which implies \( \varphi(\tilde{W}(\vec{F})) = K_a \) as expected. (See also (2.32).) This means, at selection (after infinitely many rounds), the overall final affinity of each selected nucleic acid for the target set is the same and has the value

\[
\vec{A}_{\mathcal{L}} \cdot \tilde{W}(\vec{F}) = K_a\tilde{W} = \frac{K_a}{1 + [NA]K_a},
\] (2.54)

where we have defined the mean target affinities for each target component on \( \mathcal{L} \) as the row averages of \( A_{\mathcal{L}} \), i.e., the vector \( \vec{A}_{\mathcal{L}} \) has components \( A_{\mathcal{L},i} = \frac{1}{L} \sum_{l \in \mathcal{L}} A_{il} \).

**Remark 2.6.4.1.** In a single target case [15], the quantity \( \frac{[NA]K_a}{1 + [NA]K_a} \) in (2.54) is referred to as “the maximum bound target efficiency”, and it is quoted from [32] that “maximizing the bound target efficiency is the same thing as minimizing \( K_d \) and this is in turn equivalent to minimizing the free energy”. (Here, \( K_d = 1/K_a \).) We also notice that as \([NA]\) increases we expect \( K_a \) to decrease. This suggests that the overall efficiency, \( E = [NA]K_a/(1 + [NA]K_a) \)
should decrease because poorer binders will have a better chance to bind with the target components. See Figure 2.4 for illustrations of these arguments.

Finally, a system of \( L + M \) nonlinear equations that provides the relation of final (limiting) fractions:

\[
\begin{align*}
\overrightarrow{A} \cdot \hat{\omega} &= K_a, \quad l \in \mathcal{L}, \\
\overrightarrow{A}_i \cdot \hat{F} &= \frac{1}{[NA]} \left( \frac{\Omega_i}{\omega_i W} - 1 \right), \quad i \in \mathcal{M},
\end{align*}
\]

where \( W = \frac{1}{1 + [NA]K_a} \) and \( W_i = \omega_i W \), under the constraint that \( F_l > 0 \) for \( l \in \mathcal{L}(\hat{\Omega}, [NA]) \) such that \( \sum_l F_l = 1 \). We also refer to this system as the SELEX stationary system.

It is not hard to show from (2.55) by summing \((\overrightarrow{A} \cdot \hat{\omega})F_l = K_a F_l\) over \( l \) and then summing the equations \( \Omega_i = W \omega_i (1 + [NA] \overrightarrow{A}_i \cdot \hat{F}) \) over \( i \) that \( W \left( \sum_i \omega_i - \sum_l F_l \right) = 1 - \sum_l F_l \), and therefore, the constraint conditions \( \sum_i \omega_i = 1 \) and \( \sum_l F_l = 1 \) are equivalent.

### 2.7 The connection between the limiting efficiencies and chemical potentials

In this section we include a brief discussion of the chemical potential and relate the SELEX process with the chemical potential at the limiting states.

The Second Law of Thermodynamics states that the total entropy\(^4\) of a system and its surroundings always increase for a spontaneous process. In a chemical system a reaction can occur spontaneously only if the (Gibbs) free energy change of a reaction is negative. Since entropy varies with the concentration of molecules, the free energy change of a chemical reaction also depends on the concentrations of both its reactants and their products, as well as equilibrium constants. Here we use the concepts of partial molar Gibbs free energy (Gibbs free energy change per mole of substance), also known as the chemical potential, which concerns the contribution of each species to the energy in a mixture. (See [2] and [32].) The associated chemical potentials, \( \mu^a \) and \( \mu^a_i \), of the reactions (2.17) and (2.19), at equilibrium, can then be

\(^4\)Entropy is a measure of the level of randomness or disorder.
expressed as follows:

\[ \mu^a = -RT \ln(K_a/K_0) \quad \text{and} \quad \mu_i^a = -RT \ln(\kappa_{a,i}/K_0), \]

where \( R \) is the gas constant and \( T \) is the Kelvin temperature, and the quantity \( K_0 \) is a reference value for the non-dimensionality. Let \( K_0 \) be one.

Since \( -RT \ln K_a \) represents the change in free energy, the larger the \( K_a \) (or the smaller the \( K_d \)), the more negative the change in free energy and the more spontaneous the reaction. Thus, the reaction proceeds most readily in the direction of which \( K_a \) attains its maximum or the target efficiency is close to one. (Note that \( E = \frac{[NA]K_a}{1+}[NA][K_a] \).) The negative log of the overall association constant \( pK_a = -\log_{10} K_a = -pK_d \), however, is not necessarily a monotone decreasing function of the round number. (See Figure 2.5.)

Now for reactions (2.17), (2.18) and (2.19), we have the following limiting values at infinite target dilution:

\[ K_a = \sum_{l \in \mathcal{L}} F_l K_{a,l} = \sum_{i=1}^{M} \omega_i \kappa_{a,i} = \hat{\omega}^t \cdot \hat{A} F, \]

\[ K_{a,j} = \hat{A}^j \cdot \hat{\omega}, \quad \text{and} \]

\[ \kappa_{a,i} = \hat{F} \cdot \hat{A}_i \]

from the equations in (2.20), (2.21), (2.22), respectively. The corresponding limiting efficiencies for equations (2.23) and (2.25), at infinite target dilution, become:

\[ E = [NA]K_a/(1 + [NA]K_a) \quad \text{and} \quad E_i = [NA]\kappa_{a,i}/(1 + [NA]\kappa_{a,i}). \]

From these relationships we see that any function of the chemical potentials can be viewed as a function of the overall association constants \( K_a \) and \( \kappa_{a,i} \), or as a function of the efficiencies \( E \) and \( E_i \). The functions defined below are equivalent at infinite target dilution (i.e., small \([T]\)) and provide a relationship between the chemical potentials and the individual efficiencies.
or the association constants:

\[
Q(\vec{E}) = -RT \sum_{i=1}^{M} \Omega_i \ln \left[ \frac{1}{1 - E_i} \right],
\]

\[
Q(\vec{F}) = -RT \sum_{i=1}^{M} \Omega_i \ln \left( 1 + [NA] A_i \cdot \vec{F} \right),
\]

\[
Q(\vec{\kappa}_a) = -RT \sum_{i=1}^{M} \Omega_i \ln \left( 1 + [NA] \kappa_{a,i} \right),
\]

\[
Q(\vec{\mu}^a) = -RT \sum_{i=1}^{M} \Omega_i \ln \left( 1 + [NA] e^{-\mu_a^i / RT} \right),
\]

(2.56)

where \( R \) is the gas constant and \( T \) is the Kelvin temperature.

From the last equation in (2.56), \( Q(\vec{\mu}^a) \approx \sum_{i=1}^{M} \Omega_i \mu_a^i - RT \ln [NA] \), for large \( [NA] \). That is, the function \( Q \) can be viewed as the weighted chemical potential, \( \mu_a \), when \( [NA] \) is very large but \( [T] \) is very small. One can also interpret this result by saying, in the \( M \) sub reactions in (2.19), a large nucleic acid pool can actually be considered as interacting with individual target species independently. In addition, minimizing \( Q \) is equivalent to minimizing the (weighted) chemical potential or (weighted) free energy for the entire system of \( MN \) reactions at equilibrium, for large concentrations of nucleic acids.

### 2.8 The stability and uniqueness theorems for multiple target SELEX

In the case of single target SELEX, the iterative process converges to a population of single nucleic acid species that bind tightly and specifically to the target. In multiple-target SELEX, the situation is more complicated as we have seen so far. The notion of “proper” as defined above, however, supports the uniqueness and stability of the dynamical system. The following statements hold.

#### 2.8.1 Statements of the stability and uniqueness theorems for multiple target SELEX, at infinite target dilution

**Theorem 2.8.1.1.** The target affinity function \( \varphi \) is proper in the sense that we have defined above if and only if for every \( \vec{\Omega} \) of starting target fractions, the SELEX iteration scheme converges to a unique final free target, \( \vec{\omega} \), and a unique set of final fractions \( \vec{F} \) that is independent
of the starting pool \( \hat{F}^{(1)} \in S_F \) as long as this starting vector is an interior point of the simplex \( S_F \). That is, the SELEX process is globally asymptotically stable.

**Theorem 2.8.1.2.** The target affinity function \( \varphi \) is proper in the sense that we have defined above if and only if the chemical potential at infinite target dilution defined by each \( L \) face has a unique minimum point.

The proof of Theorem 2.8.1.1 is given below. Theorem 2.8.1.2 is a consequence of the proof of Theorem 2.8.1.1. Theorem 2.8.1.1 says that if the maximal target affinity function is improper, i.e., some face is improper, the SELEX process will no longer always converge to a unique set of final nucleic acids for each choice of the initial target vector, \( \hat{\Omega} \), and will depend on the starting pool of nucleic acids with all species present. Figure 2.8 illustrates the aspects of proper and improper faces. While Figure 2.8 (a) is constructed from a target affinity function in which all the faces are proper, Figure 2.8 (c) is constructed from a target affinity function for which there is one improper face. We see that if the initial target fraction vector lies in the quadrilateral labeled \( \{1, 2, 3, 4\} \), we can select for four nucleic acids although there are only three target components. In Figures 2.9 and 2.10, we also illustrate the stability of the SELEX scheme for the cases in Figure 2.8.

**2.8.2 Proof of Theorem 2.8.1.1 and Theorem 2.8.1.2**

To prove Theorem 2.8.1.1, we make a few preliminary observations based on the system (2.55).

**Observation 2.8.2.1.** From (2.39) together with (2.54) we have

\[
K_a W = \frac{K_a}{1 + [NA]K_a} = \frac{\sum_{i=1}^{M} \Omega_i \bar{A}_i \cdot \bar{F}}{1 + [NA] \bar{A}_i \cdot \bar{F}}. 
\]

The system of equations (2.55) can be written as follows:

\[
\bar{A}^l \cdot \tilde{\omega} = K_a \quad \text{for} \quad l \in \mathcal{L}, \quad \omega_i = \frac{\Omega_i}{W(1 + [NA] \bar{A}_i \cdot \bar{F})} \quad \text{for} \quad i \in \mathcal{M},
\]

where \( W = 1 - E = 1/(1 + [NA]K_a) \), \( \tilde{\omega} \in \mathcal{S} \) and \( \bar{F} \in S_{F,L}^5 \). Suppose we have a solution of \( S_{F,L}^5 = \{ \hat{f} = (f_1, f_2, \ldots, f_N) | f_j \geq 0 \ \text{for} \ j \in \mathcal{L}, \ f_j = 0 \ \text{if} \ j \in \mathcal{N} - \mathcal{L} \ \text{and} \ \sum_{j=1}^{N} f_j = 1 \} \).
the system of equations (2.58), at infinite target dilution. Then we can say the following:

\[ W = \sum_{i=1}^{M} \frac{\Omega_i}{1 + [NA]A_i \cdot \hat{F}} = \frac{1}{K_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il}}{1 + [NA]A_i \cdot \hat{F}}, \quad \text{for } l \in \mathcal{L}. \quad (2.59) \]

Conversely, if we have a solution \( \hat{F} \in \mathcal{S}_{F, \mathcal{L}} \) that satisfies the system (2.59), then we can define \( \hat{\omega} \) using the second set of equations (2.58).

**Observation 2.8.2.2.** Here we prove the existence of a solution \( \hat{F} \in \mathcal{S}_{F, \mathcal{L}} \), in the system of equations,

\[ W = \frac{1}{K_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il}}{1 + [NA]A_i \cdot \hat{F}}, \quad \text{for } l \in \mathcal{L} \text{ and any } \hat{\Omega} \in \mathcal{S}, \]

where \( W = \frac{1}{1 + [NA]K_a} \).

**Proof.** One can easily observe, for \( l \in \mathcal{L} \), that

\[ \frac{K_a}{1 + [NA]K_a} = \frac{1}{1 + [NA]K_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il}}{1 + [NA]A_i \cdot \hat{F}} = \frac{M}{1 + [NA]A_i \cdot \hat{F}}. \]

This implies that

\[ \sum_{i=1}^{M} \frac{\Omega_i A_{il} F_l}{1 + [NA]A_i \cdot \hat{F}} = \sum_{i=1}^{M} \frac{\Omega_i A_{il} F_l}{1 + [NA]A_i \cdot \hat{F}}, \]

and therefore,

\[ \sum_{l=1}^{N} F_l = 1. \]

Let \( g_l(\hat{F}) = \frac{1}{WK_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il} F_l}{1 + [NA]A_i \cdot \hat{F}} \), for \( l \in \mathcal{L} \), and \( g_l(\hat{F}) = 0 \), for \( l \in \mathcal{N} - \mathcal{L} \). Now we want to show that there exists \( \hat{F} \) such that \( g(\hat{F}) = \hat{F} \) where \( g = (g_1, g_2, \ldots, g_N) \) and \( \hat{F} \in [0, 1]^N \).

Since \( 0 \leq F_l \leq 1 \), we have \( 0 \leq g_l(\hat{F}) = \frac{1}{WK_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il} F_l}{1 + [NA]A_i \cdot \hat{F}} \leq 1 \). Thus, by the Schauder fixed point theorem, since \( \mathcal{B} = [0, 1]^N \) is a compact and convex set in a Banach space \( \mathbb{R}^N \) and \( g \) is a continuous mapping of \( \mathcal{B} \) into itself (i.e., \( g : [0, 1]^N \rightarrow [0, 1]^N \)), \( g \) has a fixed point, i.e., there exists \( \hat{F} \in [0, 1]^N \) such that \( g(\hat{F}) = \hat{F} \).

**Observation 2.8.2.3.** Using Lagrange multipliers we can characterize the above system of equations. Let, for any set of nonnegative numbers \( F_l \) (say, the component of a vector \( \vec{F} \), not
necessarily a unit vector) and $l$ in any subset $L$ of nucleic acid indices,

$$R(\vec{F}, L) = Q(\vec{F}, L) + \lambda \left( \sum_{l \in L} F_l - 1 \right),$$

where $Q(\vec{F}, L) = -RT \sum_{i=1}^{M} \Omega_i \ln(1 + [NA] \vec{A}_i \cdot \vec{F})$ using the second equation in (2.56). By taking the partial derivative of both sides with respect to $F_l$, we have

$$\frac{\partial R}{\partial F_l} = -[NA] \sum_{i=1}^{M} \frac{\Omega_i A_{il}}{1 + [NA] \vec{A}_i \cdot \vec{F}} + \lambda.$$ 

If we rewrite

$$W(\vec{F}) = 1 - E(\vec{F}) = \sum_{i=1}^{M} \frac{\Omega_i}{1 + [NA] \vec{A}_i \cdot \vec{F}},$$

$$K_a(\vec{F}) = \frac{1 - W}{[NA] W},$$

$$\omega_i(\vec{F}) = \frac{\Omega_i}{W(\vec{F}) (1 + [NA] \vec{A}_i \cdot \vec{F})},$$

for any vector of nonnegative numbers $F_l$, and compute $\vec{A}^l \cdot \hat{\omega}$, we have

$$\frac{\partial R}{\partial F_l} = -[NA] W \vec{A}^l \cdot \hat{\omega} + \lambda.$$ 

Thus, any solution of the system $\partial_F R = 0$ is an extremal of $Q$ subject to the constraint that $\sum_{l \in L} F_l = 1$ if and only if $[NA] W \vec{A}^l \cdot \hat{\omega} = \lambda$, or $\vec{A}^l \cdot \hat{\omega} = \frac{\lambda}{E} K_a$. Moreover, the indices $l$ will correspond to those for the maximal target affinity function, $\varphi$, defined in (2.46) for this extremal if and only if $\lambda = E$ and the value of $K_a$ is a given number in the range of this function. This is an important observation from a thermodynamic point of view because, in statistical physics, the lagrange multiplier can often be viewed as the chemical potential, when maximizing the entropy.

**Observation 2.8.2.4.** Here we prove the following:

The eigenvalues of the Hessian matrix for $R$ (or $Q$) are all strictly positive if and only if the set of vectors $\{\vec{A}^l\}$ defining the $L$ face is linearly independent, i.e., the $L$ face is proper.

**Proof.** The Hessian matrix for $R$ (or $Q$) with respect to $\vec{F}$ has the bilinear form:

$$\sum_{k,l \in L} \xi_k \frac{\partial^2 R}{\partial F_k \partial F_l} \xi_l = [NA]^2 \sum_{i=1}^{M} \frac{(\vec{\xi} \cdot \vec{A}_i)^2 \Omega_i}{(1 + [NA] \vec{A}_i \cdot \vec{F})^2}.$$

(2.60)
First, note that \( \sum_{l \in L} \xi_l \vec{A}^t = (\vec{\xi} \cdot \vec{A}_1, \ldots, \vec{\xi} \cdot \vec{A}_M)^t \). If we let \( H = \left[ \frac{\partial^2 R}{\partial F_k \partial F_l} \right]_{k,l \in L} \), then we have \( \vec{\xi}^t H \vec{\xi} \geq 0 \) for all \( \vec{\xi} \) in the L face.

(\( \Rightarrow \)) Suppose the set of vectors \( \{\vec{A}^t\} \) defining the L face is dependent. Then there exists \( \vec{\xi} \) such that \( \sum_{l \in L} \xi_l \vec{A}^t = \vec{0} \) and \( \xi_l \neq 0 \) for some \( l \). Now we have \( \vec{\xi} \cdot \vec{A}_j = 0 \) for \( j = 1, 2, \ldots, M \), and \( \vec{\xi}^t H \vec{\xi} = 0 \) for \( \vec{\xi} \neq \vec{0} \). Therefore, \( H \) is not strictly positive definite, and this implies that \( H \) is not strictly convex. Thus some eigenvalues of \( H \) are not strictly positive.

(\( \Leftarrow \)) We have \( \vec{\xi}^t H \vec{\xi} \geq 0 \) for \( \vec{\xi} \neq \vec{0} \). If \( H \vec{\xi} = \lambda \vec{\xi} \), \( \vec{\xi}^t H \vec{\xi} = \lambda \vec{\xi}^t \vec{\xi} \geq 0 \), so \( \lambda \geq 0 \). Therefore, the eigenvalues of \( H \) are nonnegative. Suppose now that some eigenvalues of the Hessian are not strictly positive. (That is, there exists \( \lambda = 0 \) such that \( H \vec{\xi} = \lambda \vec{\xi} = 0 \), for \( \vec{\xi} \neq \vec{0} \).) Then \( \vec{\xi}^t H \vec{\xi} = 0 \), and \( \vec{\xi} \cdot \vec{A}_j = 0 \) for \( j = 1, \ldots, M \). Thus, \( \sum_l \xi_l \vec{A}^t = \vec{0} \) for \( \vec{\xi} \neq \vec{0} \). Hence the set of vectors \( \{\vec{A}^t\} \) defining the L face is dependent.

This then implies that the surface defined by \( R \) over the simplex \( S_{F,L} \) must be strictly convex if and only if the L face of the maximal target affinity function \( \varphi \) is proper. Hence, there cannot be more than one critical point in \( S_{F,L} \), i.e., there is at most one solution of (2.58) in \( S_{F,L} \) whenever the L face is proper. Moreover, if \( \varphi \) is proper, the limiting SELEX solution does not depend on the initial nucleic acid fractions, but does depend on the value of the limiting overall association constant, \( K_a \), and the initial target vector, \( \hat{\Omega} \).

The proof of Theorem 2.8.1.2 is the statement that an L face is proper if and only if the chemical potential over that face has a unique minimum.

Next we argue that if \( \varphi \) is improper, the SELEX solution will depend on the starting nucleic acid fractions as well as the association constant \( K_a \), the initial target vector \( \hat{\Omega} \) and the corresponding final free target vector \( \hat{\omega} \) that belongs to an improper face of the graph of \( \varphi \). Whether the face is improper or not, the initial target fractions, \( \Omega_i \), the total nucleic acid concentration, \([NA]\), and the final fractions have the following relationship:

\[
\Omega_i = \frac{\omega_i (1 + [NA] \vec{A}_i \cdot \hat{F})}{\sum_{j=1}^M \omega_j (1 + [NA] \vec{A}_j \cdot \hat{F})} = \frac{\omega_i (1 + [NA] \vec{A}_i \cdot \hat{F})}{1 + [NA]K_a},
\]  

(2.61)

where \( \vec{A}_i \cdot \hat{F} = \sum_{i \in L} F_i A_{il} \).
Equations (2.61) in vector form are
\[
\tilde{\Omega} = \mathcal{W} \left( \sum_{l \in \mathcal{L}} F_l \{ \tilde{\omega} + [NA]\tilde{A}^l \omega \} \right),
\]
(2.62)
where we have used the fact that \( \mathcal{W} = \frac{1}{1 + [NA]K_a} \). We see that the starting fraction vector can be expressed as a convex combination of the unit vectors \( \hat{V}^l = \hat{V}^l(\tilde{\omega}) \equiv \mathcal{W}(\tilde{\omega} + [NA]\tilde{A}^l \omega) \), where we define the vectors \( \tilde{A}^l \) as \( \tilde{A}^l = \langle A_{1j} \omega_1, A_{2j} \omega_2, \ldots, A_{Mj} \omega_M \rangle^t \), for each index \( j = 1, \ldots, N \) and each vector \( \tilde{\omega} \in \mathcal{S} \). The set of vectors \( \{ \tilde{\omega} + [NA]\tilde{A}^l \omega, l \in \mathcal{L} \} \) is linearly independent if and only if the set \( \{ \tilde{A}^l \omega, l \in \mathcal{L} \} \) is linearly independent. Therefore once we know the final components of the free target vector, the final NA fractions can be uniquely determined from \( \tilde{\Omega} \) and \( K_a \) if and only if \( \{ \tilde{A}^l \omega, l \in \mathcal{L} \} \) is linearly independent. The linear independence of \( \{ \tilde{A}^l \omega, l \in \mathcal{L} \} \) is also equivalent to the linear independence of the set of vectors \( \{ \tilde{A}^1, \tilde{A}^2, \tilde{A}^3 \} \) that define the columns of the affinity selection matrix (ASM) \( A \mathcal{L} \) for \( \tilde{\omega} \in \mathcal{S}_0 \). Then the final fractions are given by:
\[
\hat{F} = \begin{pmatrix}
\hat{V}^{l_1} \cdot \hat{V}^{l_1} & \hat{V}^{l_1} \cdot \hat{V}^{l_2} & \ldots & \hat{V}^{l_1} \cdot \hat{V}^{l_L} \\
\hat{V}^{l_2} \cdot \hat{V}^{l_1} & \hat{V}^{l_2} \cdot \hat{V}^{l_2} & \ldots & \hat{V}^{l_2} \cdot \hat{V}^{l_L} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{V}^{l_L} \cdot \hat{V}^{l_1} & \hat{V}^{l_L} \cdot \hat{V}^{l_2} & \ldots & \hat{V}^{l_L} \cdot \hat{V}^{l_L}
\end{pmatrix}^{-1}
\begin{pmatrix}
\hat{\Omega} \cdot \hat{V}^{l_1} \\
\hat{\Omega} \cdot \hat{V}^{l_2} \\
\vdots \\
\hat{\Omega} \cdot \hat{V}^{l_L}
\end{pmatrix}.
\]
(2.63)
The inverse of the Grammian\(^6\) matrix on the right-hand side of the equation exists if and only if the vectors \( \hat{V}^l \) are linearly independent. Moreover, if the unit target vector \( \hat{\Omega} \) is in the convex hull of the \( \hat{V}^l \)'s, then \( F_l > 0 \) for \( l \in \mathcal{L} \) such that \( \sum_l F_l = 1 \) and vice versa.

Now if \( \tilde{\omega} \) belongs to an improper face, then the final NA fractions cannot be determined from the stationary system (2.55) even if we know the final free target vector. It means, when \( \varphi \) is improper, the final NA fractions will depend on the starting NA pool as well as the initial target fractions. We argue this statement as follows.

Suppose, for example, that we have two target components \( (M = 2) \) and the set of vectors \( \{ \tilde{A}^1, \tilde{A}^2, \tilde{A}^3 \} \), where the column vectors are defined in the affinity selection matrix (ASM)

---

\(^6\)Grammian matrix (or Gram matrix or Gramian) \( G \) of a set of vectors \( \hat{V}^l, l \in \mathcal{L} \) in an inner product space is the symmetric matrix of inner products, whose entries are given by \( G_{ij} = \langle \hat{V}^i, \hat{V}^j \rangle \).
\(A_L\). Suppose also that the vectors \(\overrightarrow{A}^1, \overrightarrow{A}^2, \overrightarrow{A}^3\) are linearly dependent, but pairwise linearly independent. Then the unit vectors \(\hat{V}^1, \hat{V}^2, \hat{V}^3\) are also pairwise linearly independent. For a given single vector \(\hat{\omega}\), let us assume that \(K_a = \varphi(\hat{\omega}) = \overrightarrow{A}^1 \cdot \hat{\omega} = \overrightarrow{A}^2 \cdot \hat{\omega} = \overrightarrow{A}^3 \cdot \hat{\omega}\), but for some \(\hat{\omega}' \neq \hat{\omega}\), \(\varphi(\hat{\omega}') = \max\{\overrightarrow{A}^1 \cdot \hat{\omega}', \overrightarrow{A}^2 \cdot \hat{\omega}'\} > \overrightarrow{A}^3 \cdot \hat{\omega}'\). Then from the definition in subsection 2.6.3 we have \(L(\hat{\omega}) = \{1, 2, 3\}\). There are two 1 faces defined by the indices 1, 2, respectively. However, the pair \((\hat{\omega}, \varphi(\hat{\omega}))\) defines a two face. This the two face is not proper because the set of indices that describes it is not unique. That is, there are three sets \(\{1, 2\}, \{1, 3\}, \{2, 3\}\) that define its indices and all are proper subsets of \(\{1, 2, 3\} = L\).

We can then write \(\hat{V}^3 = \lambda_1 \hat{V}^1 + \lambda_2 \hat{V}^2\), where \(\lambda_1 + \lambda_2 = 1\), since \(\hat{V}^3\) is the unit vector, as well as \(F_1 + F_2 + F_3 = 1\). Then the equation (2.62) becomes

\[
\hat{\Omega} = F_1 \hat{V}^1 + F_2 \hat{V}^2 + F_3 \hat{V}^3 = (F_1 + \lambda_1 F_3) \hat{V}^1 + (F_2 + \lambda_2 F_3) \hat{V}^2 \equiv g_1 \hat{V}^1 + g_2 \hat{V}^2.
\]

Since \(\hat{V}^1\) and \(\hat{V}^2\) are independent, \(g_1\) and \(g_2\) are uniquely determined. We can also find that \(g_1 + g_2 = 1\), and therefore, at least one of the \(g_i\) is positive. Then,

\[
g_1 + g_2 = (F_1 + \lambda_1 F_3) + (F_2 + \lambda_2 F_3) = 1 + (\lambda_1 + \lambda_2 - 1) F_3 = 1.
\]

That is, \(F_3\) cannot be determined from the stationary equations even when the final free target vector \(\hat{\omega}\) is known. Moreover, all the stationary solutions are of the form \(\hat{F}_s(t) \equiv (g_1 - \lambda_1 t, g_2 - \lambda_2 t, t) = (F_1, F_2, F_3)\) for \(t\) in some subinterval of \((0, 1)\), namely that subinterval for which \(g_1 - \lambda_1 t \geq 0\) and \(g_2 - \lambda_2 t \geq 0\).

The argument can be easily expanded to the general case. So, for a single free target vector, if we have a \(L\) face that is improper, then there is a proper subset with \(L'\) elements such that \(L' < L\) and \(L' = L - 1\), say. Then we have \(L\) unknown final fractions for a system of \(L - 1\) equations, which must be degenerate.

If the \(L\) face is improper, the Hessian of the chemical potential corresponding to it must have its smallest eigenvalue vanish. Therefore, each set of stationary solutions (in our example, the vector family \(\{\hat{F}_s(t)\}\)) minimizes the chemical potential at infinite target dilution, i.e., is a realizable thermodynamic state. One consequence of this result is that, in such a limiting degenerate case, the final NA fractions, for a given \(\hat{\Omega}\), change under an appropriate small
perturbation of initial NA fractions. That is, if we start with such a state as an initial state, with the given $\hat{\Omega}$, we obtain another state as a final state, indicating the failure of asymptotic stability.

2.9 Geometric properties of the family of convex hulls generated by a single free target vector $\hat{\omega}$

In the preceding section we defined the unit vectors as $\hat{V}^l = \hat{V}^l(\hat{\omega}) \equiv W(\hat{\omega} + [NA]\hat{A}_l\omega)$. We now form the convex hull\(^7\) of the vectors $\hat{V}^l$ as $\mathcal{H} = \mathcal{H}([NA], \hat{\omega}) \subset S$. The study of this geometric structure of convex hull provides us to understand how the SELEX system, at selection, is affected by changes of the size of nucleic acids.

Assuming that the vectors $\hat{V}^i$ are linearly independent, the convex hull of the set of vectors $\{\hat{V}^l, l \in L\}$ defined above has the dimension of $L - 1$. Therefore, the largest (in dimensionality) sets of initial targets come from those $\hat{\omega}$ that yield an $M \times M$ affinity selection matrix with full rank $M$, which corresponds to the $M$ selected nucleic acids.

The diameter of the set of vectors $\hat{V}^l$ is equal to the diameter of its convex hull. The diameter of the set is defined to be the maximum (or supremum) of the distances between pairs of points in the set. We now define the diameter of the convex hull $\mathcal{H}$ and denote the number as

$$(\mathcal{H}) \equiv \max\{|\hat{V}^l - \hat{V}^m|, l, m \in L\} = [NA] \max\{|\hat{A}_l\omega - \hat{A}_m\omega|, l, m \in L\}/(1 + [NA]K_a),$$

for all indices $i = 1, \ldots, M$. Clearly, this diameter is an increasing function of $[NA]$ and takes a value of zero when $[NA] = 0$. If $[NA] \to 0$ then $\hat{\Omega} = \hat{\omega}$, i.e., the convex hull, $\mathcal{H} = \{\hat{\omega}\}$.

Moreover, as $[NA] \to \infty$, $\hat{V}^l \to \hat{A}_l\omega$, for each $l \in L$,

and

$$(\mathcal{H}) \to \max\{|\hat{A}_l\omega_l - \hat{A}_m\omega_m|, l, m \in L\}/K_a.$$

\(^7\)The convex hull of a set of points in two dimensions is the smallest convex polygon that encloses all of the points. (It is the minimal convex set.) See Chapter 1 in [1], for examples of convex hulls.
Here we have used the fact that the vectors
\[
\overrightarrow{A_l}\omega/K_a = \overrightarrow{A_l}\omega/\varphi(\omega) = \overrightarrow{A_l}\omega/(\overrightarrow{A_l} \cdot \hat{\omega}) = \overrightarrow{A_l}\omega/|\overrightarrow{A_l}\omega| = \hat{\overrightarrow{A_l}}\omega
\]
are unit vectors with positive components. Then we can rewrite (2.61) as
\[
\hat{\Omega} = W\left(\hat{\omega} + [NA]\sum_{l \in L} F_l \overrightarrow{A_l}\omega\right) = W\hat{\omega} + [NA]K_aW\sum_{l \in L} F_l \overrightarrow{A_l}\omega/K_a = W\hat{\omega} + (1 - W)\sum_{l \in L} F_l \hat{\overrightarrow{A_l}}\omega
\]
(2.64)
where \(\hat{F} \in S_{F, L}\). Therefore, every starting target fraction vector \(\hat{\Omega}\) that can reach \(\hat{\omega}\) can be expressed as a convex combination of a free target vector, \(\hat{\omega}\), and the unit affinity vectors, \(\hat{\overrightarrow{A_l}}\omega\), with \(l \in L\).

If each vertex of the convex hull \(H([NA], \hat{\omega})\) depends on the size of nucleic acids, \([NA]\), then \(\hat{V}^l = \hat{V}^l([NA]) = W\hat{V}^l([0]) + (1 - W)\hat{V}^l(\infty) = W([NA])\hat{\omega} + (1 - W([NA]))\hat{\overrightarrow{A_l}}\omega\). Therefore, the family of convex hulls forms an increasing family of sets. That is, for two different values of nucleic acids, \([NA]\) and \([NA]'\) such that \(0 < [NA] < [NA]'\), we have
\[
\{\hat{\omega}\} \subset H([NA], \hat{\omega}) \subset H([NA]', \hat{\omega}) \subset S.
\]
Then from the equations (2.61), (2.62) and (2.63), we have, as \([NA] \to \infty\),
\[
\Omega_i = \frac{\omega_i(1 + [NA]A_i \cdot \hat{F})}{1 + [NA]K_a} \to \frac{\omega_i A_i \cdot \hat{F}}{K_a}, \quad \text{and therefore,}
\]
\[
\hat{\Omega} = W\left(\sum_{l \in L} F_l\{\hat{\omega} + [NA]A_l \cdot \hat{\omega}\}\right) \to \sum_{l \in L} F_l \hat{\overrightarrow{A_l}}\omega, \quad \text{where}
\]
\[
\hat{F} \to \begin{pmatrix}
\hat{A}_1 \omega \cdot \hat{A}_1 \omega & \hat{A}_1 \omega \cdot \hat{A}_2 \omega & \cdots & \hat{A}_1 \omega \cdot \hat{A}_L \omega \\
\hat{A}_2 \omega \cdot \hat{A}_1 \omega & \hat{A}_2 \omega \cdot \hat{A}_2 \omega & \cdots & \hat{A}_2 \omega \cdot \hat{A}_L \omega \\
\vdots & \vdots & \ddots & \vdots \\
\hat{A}_L \omega \cdot \hat{A}_1 \omega & \hat{A}_L \omega \cdot \hat{A}_2 \omega & \cdots & \hat{A}_L \omega \cdot \hat{A}_L \omega
\end{pmatrix}^{-1}
\begin{pmatrix}
\hat{\Omega} \cdot \hat{A}_1 \omega \\
\hat{\Omega} \cdot \hat{A}_2 \omega \\
\vdots \\
\hat{\Omega} \cdot \hat{A}_L \omega
\end{pmatrix}.
\]
(2.65)

2.10 The initial target fraction relationship to the final free target and nucleic acid fractions

Based on the observations made in the two preceding sections, here we suggest an algorithm of finding nucleic acids fractions, given that we know some experimental parameters such as
the initial target fraction vector, $\hat{\Omega}$, the overall association constant, $K_a$, and the concentration of nucleic acids, $[NA]$.

The given affinity matrix $A$ defines two functions, $\varphi : S \rightarrow (0, \infty)$ and $L : S \rightarrow P(N)$, where $\varphi$ is convex and continuous on the simplex $S$ and $P(N)$ is the set of all subsets of $N = \{1, 2, \ldots, N\}$. For a given free target vector $\hat{\omega}$, the function $W = W([NA], \hat{\omega}) = 1/(1 + [NA]\varphi(\hat{\omega}))$ and unit vectors in $S$, $\hat{V}^l = W\hat{\omega} + (1 - W)A^l\hat{\omega}$ for $l \in L$ are defined. We denoted the convex hull of this set of vectors by $H([NA], \hat{\omega})$. If the final free target fraction vector is the vector $\hat{\omega}$, then the starting target fractions must come from this set and vice versa. In order to find the final nucleic acid fraction vector $\hat{F}$ when $\hat{\Omega} \in H([NA], \hat{\omega})$, we set $F_j = 0$ if $j \notin L$ and use (2.63) to evaluate the $F_j$ when $j \in L$.

If $\varphi$ is proper, then the set $\{\hat{V}^l | l \in L\}$ is a linearly independent set, i.e., the rank of the Grammian in (2.63) is $L$ and the final nucleic acid fractions are uniquely determined by $\hat{\Omega}$ and $\hat{\omega}$. However if $\varphi$ is improper and the set $L$ does not uniquely determine the $L$ face that corresponds to the pair $(\hat{\omega}, \varphi(\hat{\omega}))$, then the Grammian, in section 2.8, will not be invertible and two different starting nucleic acid fraction vectors $\hat{F}$ will determine two different sets of final nucleic acid fractions. See Figure 2.10, panel (a).

If we know the distribution of nucleic acid fractions in the given initial pool, we can determine both $\hat{\omega}$ and $\hat{F}$ simultaneously, whether $\varphi$ is proper or not, using the SELEX iteration scheme. In the laboratory, however, one could ask the following question: How can we find the final nucleic acid fractions if we know the initial target fraction vector $\hat{\Omega}$, the overall association constant $K_a$ and the concentration of nucleic acids $[NA]$? Let us consider the laboratory case when the initial pool of nucleic acids is unknown. We proceed as follows:

1. Calculate $\varphi$ and determine its faces. The simplex $S_0$ can be projected onto the interior of the unit cube in $R^{M-1}$ via the transformation $\omega_1 = 1 - s_1, \omega_2 = s_2(1 - s_1), \ldots, \omega_{M-1} = s_1 \cdots s_{M-2}(1 - s_{M-1}), \omega_M = s_1 \cdots s_{M-2}s_{M-1}$. A rectangular grid can then be imposed on this cube. The pointwise evaluation of $\varphi$ is then carried out over this grid.

2. Use equation (2.62) to compute the convex hull of the set of vectors $\{\hat{V}^l(\hat{\omega})|(\hat{\omega}, \varphi(\hat{\omega})) \in \Phi(L)\}$. Call this hull $H([NA], \Phi(L))$. Then the simplex $S$ of target fractions $\hat{\Omega}$ can be
written as
\[ S = \cup \{ \mathcal{H}([NA], \Phi(\mathcal{L})) \mid \Phi(\mathcal{L}) \text{ is a face of the graph of } \varphi \} \] (2.66)

3. Suppose that \( \hat{\Omega} \in \mathcal{H}([NA], \Phi(\mathcal{L})) \). Then \( \varphi(\hat{\omega}) = K_a \) for some free target vector \( \hat{\omega} \) and \( \Omega_i = (1 + [NA] \vec{A}_i \cdot \vec{F}) \omega_i \mathcal{W} \) where \( \mathcal{W} = 1/(1 + [NA] K_a) \) and where \( F_j = 0 \) if \( j \notin \mathcal{L} \).

4. If the face in question is proper, the system of \( L \) equations (2.59) has one and only one solution. If it is not proper, then there will be a several parameter family of stationary final fraction vectors satisfying (2.59). They will be stable but not asymptotically stable.

5. In the proper case, components of the unique free target are then found from \( \omega_i = \Omega_i/(1 + [NA] \vec{A}_i \cdot \vec{F}) \mathcal{W} \). In the improper case, there will be a several parameter family of free targets corresponding to the family of solutions of (2.59).

### 2.11 A special solution of the stationary SELEX equations

We now consider the following question: Is there a starting target fraction vector, \( \hat{\Omega} \), and a final free target vector, \( \hat{\omega} \), such that for every pool size, \([NA]\), \( \hat{\Omega} \in \mathcal{H}([NA], \hat{\omega}) \)? That is, is there a target vector with the property that the SELEX iteration scheme must converge to the same final free target vector independently of the size of the initial pool as well as the initial nucleic acid distribution?

Another way of formulating this question is to ask whether or not there is a choice of \( \hat{\Omega} \) such that the right-hand side of (2.63) does not depend upon the nucleic pool size although the vectors \( \vec{V}_i \) depend on it.

If \([NA] = 0\), it is clear that \( \hat{\Omega} \in \mathcal{H}(0, \hat{\omega}) \) for any \( \hat{\omega} \). For other values of \([NA]\), this choice holds for \( \hat{\Omega} \) if and only if
\[ \vec{A}_i \cdot \vec{F} = K_a \text{ for all indices } i = 1, \ldots, M. \] (2.67)

This system arises naturally if we try to minimize the chemical potential in the form given in the first equation in (2.56) with respect to the partial energies \( E_i \) subject to the constraint that the total (weighted) energy is fixed. We see this as follows. Using Lagrange multipliers to
minimize \( Q \) subject to the constraint that \( E = \sum \Omega_i E_i \) is fixed yields \( E_i = E \). This extreme point is unique, and therefore, \( E_i = [NA] \bar{A}_i \cdot \bar{F} / (1 + [NA] \bar{A}_i \cdot \bar{F}) = E \) and \( [NA] \bar{A}_i \cdot \bar{F} = E/(1 - E) \). Now suppose that our set of final fractions \( \bar{F} \in S_{F, L} \) satisfies (2.67). Then \( E = [NA] K_a / (1 + [NA] K_a) \) and \( W = 1 - E = 1/(1 + [NA] K_a) \). Therefore we must have \( \omega_i = \Omega_i / (W(1 + [NA] K_a)) \), and hence \( \omega_i = \Omega_i \).

Since if \( L < M \) the system is over determined, we assume \( L = M \) and \( M \leq N \). Suppose that \( \mathcal{L} = \{l_1, \ldots, l_M\} = \{1, \ldots, M\} \) and \( \{\bar{A}^1, \ldots, \bar{A}^M\} \) is a linearly independent set. We have \( \varphi(\bar{\omega}) = \bar{A}^1 \cdot \bar{\omega} = \cdots = \bar{A}^M \cdot \bar{\omega} = K_a > \bar{A}^j \cdot \bar{\omega} \) for \( j \neq 1, \ldots, M \). It is not hard to show from Cramer’s rule for solving linear systems that

\[
\omega_j = K_a \frac{\text{det} \left( \begin{array}{cccc}
A_{11} & \ldots & A_{j-1,1} & 1 & A_{j+1,1} & \cdots & A_{M1} \\
A_{12} & \ldots & A_{j-1,2} & 1 & A_{j+1,2} & \cdots & A_{M2} \\
\vdots & \cdots & \vdots & \cdots & \cdots & \cdots & \vdots \\
A_{1,M-1} & \ldots & A_{j-1,M-1} & 1 & A_{j+1,M-1} & \cdots & A_{M,M-1} \\
A_{1M} & \ldots & A_{j-1,M} & 1 & A_{j+1,M} & \cdots & A_{M,M} \\
\end{array} \right)}{\text{det}(A_{LM})} \tag{2.68}
\]

for \( j = 1, \ldots, M \). The numerator is clearly the column sum of the elements of the \( j \)th column of the classical adjoint of the matrix \( A_{LM} \). Thus, \( \bar{\omega} \) will be well defined if and only if

a. All the column sums of the classical adjoint of the SELEX matrix \( A_{LM} \) have the same sign as the determinant of this matrix.

b. The overall association constant is given by

\[
K_a = \frac{\text{det}(A_{LM})}{\Delta A} \tag{2.69}
\]

where \( \Delta A \equiv \sum_{j=1}^{M} \sum_{i=1}^{M} A_{ij} \) denotes the sum of the cofactors \( A_{ij} \) of \( A_{LM} \).

Using the equations (2.67) and again invoking Cramer’s rule as we did above with \( A_{LM} \) replacing \( A'_{LM} \), we obtain

\[
F_i = K_a \frac{\sum_{j=1}^{M} A_{ij}}{\text{det}(A_{LM})} = \frac{\text{det}(A_{LM})}{\Delta A} \frac{\sum_{j=1}^{M} A_{ij}}{\text{det}(A_{LM})} = \frac{\sum_{j=1}^{M} A_{ij}}{\Delta A} \tag{2.70}
\]
These sum to unity. However, this will be the case if the column sums of the classical adjoint are nonnegative. Therefore, there is a starting target fraction vector, \( \hat{\Omega} \), and a final free target fraction vector, \( \hat{\omega} \), that does not depend on the total pool concentration, \([NA]\), if and only if the SELEX matrix has the property that the row sums and column sums of its classical adjoint all have the same sign as its determinant. In this case \( \hat{\Omega} = \hat{\omega} \).

For any finite value of \([NA]\), the efficiency is

\[
E_{\text{final}} = 1 - \sum_{i=1}^{M} W_i = \frac{[NA] \det(A_{LM})}{\Delta A + [NA] \det(A_{LM})} = \frac{[NA]}{\Delta A/ \det(A_{LM}) + [NA]}. \tag{2.71}
\]

**Remark 2.11.0.1.** When \( M = 2 \) the classical adjoint of \( A_{L_2}^t \) is

\[
\begin{pmatrix}
A_{22} & -A_{21} \\
-A_{12} & A_{11}
\end{pmatrix}.
\]

Suppose for the moment that \( A_{L_2}^t \) has a positive determinant. Then a necessary and sufficient condition for the row and column sums of the classical adjoint to be positive is that \( \max\{A_{12}, A_{21}\} \leq \min\{A_{11}, A_{22}\} \).

The above result becomes

\[
\hat{\Omega} = \begin{pmatrix}
\Omega_1 \\
\Omega_2
\end{pmatrix} = \begin{pmatrix}
\omega_1 \\
\omega_2
\end{pmatrix} = \frac{1}{\Delta A} \begin{pmatrix}
A_{22} - A_{21} \\
A_{11} - A_{12}
\end{pmatrix} \quad \text{with}
\]

\[
\hat{F} = \begin{pmatrix}
F_1 \\
F_2
\end{pmatrix} = \frac{1}{\Delta A} \begin{pmatrix}
A_{22} - A_{12} \\
A_{11} - A_{21}
\end{pmatrix}. \tag{2.72}
\]

**Remark 2.11.0.2.** There is a geometric meaning to the condition that the row sums as well as the column sums of the classical adjoint of the SELEX matrix are positive. This condition is equivalent to the condition that the column sums of the classical adjoint of the matrix \( A_{LM}^t \) have the same sign as the determinant of this matrix. Suppose, without loss of generality that \( \det(A_{LM}) > 0 \). The \( M \) hyperlines that form the one dimensional edges of the surface \( z = \varphi(\hat{\omega}') \) near the minimum \( \hat{\omega} \) are given by the parametric equations (for fixed \( j = 1, \ldots, M \)) \( t = (\omega_i - \omega_{i,m})/A_{ij} \) for \( i = 1, \ldots, M \) where \( t \) is the free parameter for the line and where \( A_{ij} \) is defined above. That is, the vector \( \vec{B}_j = \langle A_{1j}, \ldots, A_{Mj} \rangle \) is (up to a scalar) the direction vector for the \( j^{th} \) edge. The vector \( \vec{I} = \langle 1, \ldots, 1 \rangle/\sqrt{M} \) is the outer normal to \( S \). The condition that the minimum of \( \varphi \) occur at \( \hat{\omega}_m \) is equivalent to the condition that \( \vec{B}_j \cdot \vec{I} > 0 \), i.e., that the
row sums of the classical adjoint of $A^t_{LM}$ are all positive. This tells us that the row sums of the classical adjoint $A^t_{LM}$ are positive if and only if the vector $\hat{\omega}$ whose components are given in (2.68), are well defined and are such that $\varphi(\hat{\omega}) = \varphi_{\text{min}}$.

Therefore, there is starting target fraction, $\hat{\Omega}$, in $\mathcal{H} = \mathcal{H}([NA],\hat{\omega})$ for all values of the nucleic acid pool and some final free target fraction vector $\hat{\omega}$ if and only if $\hat{\Omega} = \hat{\omega}$ where $\varphi(\hat{\omega}) = \varphi_{\text{min}}$.

### 2.12 Simulation results for multiple-target SELEX

In this section, a number of simulation results relating to the SELEX iteration schemes are discussed. Numerical values and the affinity matrices used for the simulations are reported in Appendix C. In a laboratory, molarity ($M$, moles of solute per liter of solution) is the most commonly used measure of concentration. In [11] the range of $10^{-9}$ to $10^{-7} M$ was used for dissociation constants $K_d$, or equivalently, the range of $10^7$ to $10^9 M^{-1}$ for association constants, as well as $[NA] = 10^{-5} M = 10 \mu M$. Here, we use the range of between $10^2$ to $10^4 (\mu M)^{-1}$ for the values of association constants, and fix $[NA] = 1 \mu M$ at every round to employ a PCR amplification. We also take the parameter value $[T] = 1 \mu M$. The effects of selection by diluting nucleic acids or target species are also described in this section. One of our goals in the simulations is to visualize the behavior of the multi-parameter dynamical system. Here, three and five target cases are illustrated to provide more geometric insights of our dynamical system. The initial nucleic acid fractions are chosen by using the random number generator in MATLAB.

#### 2.12.1 Final fractions using the SELEX iteration scheme

The simulation procedure is performed as follows. At each round, given a target vector $[\overrightarrow{TF}(r)] = (1 - s_r)[\overrightarrow{TF}(r-1)]$ and a nucleic acid fraction vector $\hat{F}^{(r)}$, we first solve the $M$ nonlinear equations (2.28) to compute the free target vector $[\overrightarrow{TF}(r)]$, using either Newton’s method or fixed point iteration method. Starting from an initial guess as the zero vector, we solve the nonlinear equations and repeat the process until the relative error $|(\overrightarrow{TF}_i^{(r,k+1)})/|\overrightarrow{TF}_i^{(r,k)}| - 1|$
becomes smaller than a specified tolerance. (Here, \( k \) is the iteration number for this procedure.) Since the functions on the right-hand side of (2.28) are increasing in each \( [T^f_i(r)] \) and take values in \((0, \infty)\), the iteration scheme converges. Once we compute the free target vector, the quantities, \( D_{i,j}^{(r)} = [T^f(r)]A^j \cdot \hat{\omega}^{(r)} \) for \( j = 1, \ldots, N \) can be computed. Then, the updated NA fractions for the next \((r + 1)^{th}\) round can be determined from the equation (2.29). The entire process is then repeated until the scheme reaches selection. As mentioned at the beginning of this chapter and also shown in [15] for a single target case, we take \([\overrightarrow{T}(r+1)] = (1-s_r)[\overrightarrow{T}(r)]\) where \(s_r = 1/(r+1)\) for the choice of target concentration so that \([T(r)] \to 0\), to reach an infinite dilution of the probe\(^8\). The figures here were generated using this iterative approach.

1a. Figure 2.1(a,c,e,f). If we use the final free target vector found by iteration and given in the caption of Figure 2.1 to evaluate \(\varphi(\hat{\omega})\) from its definition, we obtain the value \(\varphi(\hat{\omega}) = 3207.45(\mu M)^{-1}\), as well as the set of indices \(\{8, 9, 10, 12, 16\}\). To check that this is the minimum value, we computed \(\varphi(\hat{V}_i)\) where \(\hat{V}_i = (\hat{\omega} + \epsilon \hat{e}_i)/(1 + \epsilon)\) where \(\hat{e}_i\) is one of the standard basis vectors \(\hat{e}_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{i5})\) and where \(\epsilon > 0\) was small. We found that \(\varphi(\hat{V}_i) > \varphi(\hat{\omega})\). Because the graph of \(\varphi\) is convex, this strongly suggests that \(\varphi(\hat{\omega}) = \varphi_{min} = K_a\). Using the above set of indices, \(\hat{\omega}\) and the starting target fraction vector \(\hat{\Omega}_a = \langle 0.1374, 0.1346, 0.4090, 0.1844, 0.1346 \rangle\), we computed the final nucleic acid fractions corresponding to these vectors from (2.63) and found good agreement. This is confirmed by calculating row and column sums of the classical adjoint of the \(5 \times 5\) matrix whose columns are indexed by \(\{8, 9, 10, 12, 16\}\) and is given in (5.6), a task that can be computationally intensive for large matrices. For this matrix, the classical adjoint is given in (5.7). The column vector of row sums is \(10^{15}\langle 1.1182, 0.8577, 2.1713, 1.6744, 1.0952 \rangle^t\) and the row vector of column sums is \(10^{15}\langle 1.5082, 0.9144, 1.3998, 1.8013, 1.2930 \rangle\). The value for \(K_a = 3207.45(\mu M)^{-1}\) given in the caption of Figure 2.1 was found by iteration.

The same value is found by using formula (2.69).

1b. Figure 2.1(b,d,e,f). We have the set of indices \(\{8, 9, 10, 16\}\) with the limiting \(K_a = \)

---

\(^8\)A probe is a piece of DNA or RNA used to detect specific nucleic acid sequences by hybridization (binding of two nucleic acid chains by base pairing).
$3256.05(\mu M)^{-1} = \varphi(\hat{\omega}) > \varphi_{\text{min}}$, for $\hat{\Omega}_b = (0.2376, 0.1453, 0.1145, 0.2821, 0.2205)$. The limiting NA fraction vector given in the caption of Figure 2.1 agrees with the final fractions computed from the formula (2.63). The overall dissociation constant and the overall target efficiency as a function of round number are shown in panel (e,f) together with the former case in Figure 2.1(a).

2. Figure 2.2. We calculated the individual nucleic acid efficiencies in terms of the total nucleic acid efficiency at the end of 40 rounds using the formulas $E_{j,r} = D_{j,f}^{(r)}/(1 + D_{j,f}^{(r)})$ with $M_{j,r} = E_{j,r}/\sum_{l=1}^{N} E_{l,r}$. (After 40 rounds, $[Tf] \approx 10^{-6}M$ so that $1 + K_{a,j}[Tf] \approx 1.0$ for the values used.)

3. Figure 2.3. We repeated the same calculation as used for Figure 2.1(a,b), but did not reduce the total target concentration from round to round. Convergence to the final nucleic acid pools was much slower.

4. Figure 2.4. We fixed the starting target vector, $\hat{\Omega} = \hat{\Omega}_a$ and several values of $\log_{10}[NA] = 0, -1, -2, -3, -4, -5$. In panel (a), we see how the number of indices that correspond to the selected nucleic acids varies with pool size. Also note, there are corresponding nucleic acid indices. For example, when three indices are selected, they will correspond to nucleic acid indices 8, 9, 12. In panel (b) we see that $pK_a$ appears to be a monotone decreasing function of $pNA$.

5. Figure 2.5. We plotted $(pK_a)^{(r)}$ as a function of round number for various pool sizes using the same initial target vector as in Figure 2.4. In panel (a), the association constant appears to be decreasing with a round number. However, in panel (b), this is not the case when the round number is increased beyond 100 and hence no statement can be made as to the monotonicity of $(pK_a)^{(r)}$. This is a stark difference between single and multiple target SELEX because in the single target case, the overall $K_a$ increases as a function of a round number.
2.12.2 Maximal target affinity function, initial target decomposition and chemical potential functions using the iteration scheme

In Figures 2.1 ∼ 2.5 we compute the limiting parameters: NA fractions, free target fractions, overall association constants and efficiencies, using the SELEX iteration scheme for the five targets case. However, we are uncertain that the maximal target affinity function is proper or not, even though the results do not require any procedure to check properness. In this section, we now take an example where there are three targets and five nucleic acid species for the experiments, so that we can investigate and compare our iteration scheme more visually with the theoretical results.

In Figure 2.6, panels (a) and (b), we present the graph of the maximal target affinity function and its contour plot over the free target simplex. It shows here that the maximal target affinity function \( \varphi \) for this case is proper. In panels 2.6 (c) and (d) of Figure 2.6, we decompose the limiting SELEX indices over a grid in the initial target space. Each grid corresponds to each initial target vector (in total, there are about 20,000 grids), and given the initial target we simulated the SELEX iterative procedure until we reached a selection. Then, we partition the initial target space into the various hulls based on the final NA indices that correspond to the initial target vector. To see how the size of nucleic acids affects the final NA indices we performed the experiments for the cases when (c) \([NA] = 10^{-4} \mu M = 100 \mu M\) and (d) \([NA] = 1 \mu M\). As we discussed in section 2.9, the size (diameter) of convex hull changes as the size of total nucleic acids changes. By comparing panel (c) with panel (d), we see as \([NA]\) increases, the size of the regions where two or three nucleic acids can be selected increases, while the size as of convex hulls, where one nucleic acid species can be selected, decreases as \([NA]\) increases.

In Figure 2.7 we plot the chemical potential functions \(Q(\tilde{F})\) defined in the second equation in (2.56): \(Q(\tilde{F}) = -RT \sum_{i=1}^{M} \omega_i \ln(1 + [NA] \vec{A}_i \cdot \tilde{F})\). In Theorem 2.8.1.2, we have proven that the target affinity function \(\varphi\) is proper if and only if the chemical potential at infinite target dilution defined by each \(L\) face has a unique minimum point. Using the proper case in Figure 2.6, we plot the chemical potential functions \(Q(\tilde{F})\) over the simplex \(S_{\mathcal{F},\mathcal{L}}\) when the
initial targets \( (a,c) \hat{\Omega} = \langle 0.2, 0.42, 0.38 \rangle \) and \( (b,d) \hat{\Omega} = \langle 0.5, 0.2, 0.3 \rangle \), with \([NA] = 1\mu M\). The corresponding final NA index sets are \( \{1, 2, 3\} \) and \( \{1, 3\} \), respectively. The minimum values are designated by ‘\( \star \)’ and the corresponding NA fraction vectors \( \hat{F} \) in panels (a, b) have a good agreement with the final NA fractions in panels (c,d), respectively.

2.12.3 Comparison of stationary SELEX solutions with those obtained by iteration and stability properties of stationary solutions

In Figure 2.8, we illustrate the SELEX partitions over the initial target spaces for the two cases when \( \varphi \) is proper (panels (a,b)) and \( \varphi \) is improper (panels (c,d)). Panel (a) is the case where the maximal target function \( \varphi \) is proper and has a unique minimum point (corresponding to the index set \( \{3, 4, 5\} \)) and another point (corresponding to the index set \( \{1, 4, 5\} \)) on the graph of \( \varphi \) defined by exactly three intersecting planes. Panel (b) concerns the case where the function \( \varphi \) is improper with a unique point (corresponding to the index set \( \{1, 2, 3, 4\} \)) common to all four hyperplanes. That is, there are four ordered subsets of three integers that describe this point: \( \{1, 2, 3\} \), \( \{1, 2, 4\} \), \( \{1, 3, 4\} \) and \( \{2, 3, 4\} \). Panels (a) and (c) were generated by the SELEX program for many points in the initial target simplex in the same manner as described in the preceding section. Panels (b) and (d) were generated by computing the convex hull of the set of vectors \( \hat{V}^l \) as discussed in section 2.10. In Figures 2.9 and 2.10, we examine the stability properties for these three target problems.

2.12.4 Stability properties of stationary solutions.

To illustrate the asymptotic stability properties of the SELEX process, we consider two cases for which \( M = 3 \) and \( N = 5 \). In the first case, the maximal target affinity function is proper (in panels (a, b), Figure 2.8), while in the second case, it is improper (in panels (c, d), Figure 2.8). The asymptotic stability of the SELEX scheme was tested for several choices of initial target fractions one from each of the regions labeled \( \{4\} \), \( \{3, 4\} \), \( \{1, 4, 5\} \) and \( \{3, 4, 5\} \) from Figure 2.8, panel (a) (or (b)), and the regions labeled \( \{1, 2, 3, 4\} \) and \( \{2, 3\} \) from Figure 2.8, panel (c) (or panel (d)). We label these sets as \( S(\{4\}), S(\{3, 4\}), S(\{1, 4, 5\}), S(\{3, 4, 5\}) \)
and $\mathcal{S}([1, 2, 3, 4]), \mathcal{S}([2, 3])$, respectively.

In each of the four regions indicated, a value of the starting target fraction vector $\hat{\Omega}$ was selected. In the simplex of the nucleic acid fractions, $\mathcal{S}_F$, we generated six random vectors $\{\hat{F}_1^{(0)}, \ldots, \hat{F}_6^{(0)}\}$, computed the one norms, $|\hat{F}_1^{(r)} - \hat{F}_j^{(r)}|_1$, for $j = 2, \cdots, 6$, and plotted as a function of the round number, $r$.

In all cases in Figure 2.9 the final NA fractions does not depend on the distribution of initial NA fractions, since all quantities of $|\hat{F}_1^{(r)} - \hat{F}_j^{(r)}|_1$ converge to zero for $j = 2, \cdots, 6$ for all four cases (a-d). Therefore, the SELEX process is globally asymptotically stable, whenever the initial target belongs to any convex hull of these proper faces.

In Figure 2.10 panel (a), these quantities $|\hat{F}_1^{(r)} - \hat{F}_j^{(r)}|_1$ do not converge to zero, thus illustrating the failure of global asymptotic stability at an improper face, while in Figure 2.10 panel (b), these quantities do converge to zero.
Figure 2.1  Panels (a,b): With $[NA] = [T] = 1 \mu M$, the index sets for the selected nucleic acids are (a) $\{8, 9, 10, 12, 16\}$ and (b) $\{8, 9, 10, 16\}$. The initial target vectors are (a) $\hat{\Omega}_a = \langle 0.1374, 0.1346, 0.4090, 0.1844, 0.1346 \rangle$ and (b) $\hat{\Omega}_b = \langle 0.2376, 0.1453, 0.1145, 0.2821, 0.2205 \rangle$ with the limiting NA fraction vectors given by $F_a = (F_8, F_9, F_{10}, F_{12}, F_{16}) = (0.1956, 0.2794, 0.0498, 0.3843, 0.0908)$ and $F_b = (F_8, F_9, F_{10}, F_{16}) = (0.3060, 0.0670, 0.2629, 0.3640)$. These limiting vectors agree with values from the formula (2.63) to at least eight significant figures (not shown). Panels (c,d): The final free target fractions in cases (a,b), after 200 rounds, not shown, are: (c) $\hat{\omega} = \langle 0.1617, 0.1241, 0.3139, 0.2420, 0.1583 \rangle$ and (d) $\hat{\omega} = \langle 0.1690, 0.1288, 0.2965, 0.2476, 0.1581 \rangle$, with the limiting association constants $K_a = 3207.45(\mu M)^{-1}$ and $K_a = 3256.05(\mu M)^{-1}$, respectively.
(a) Individual nucleic acid efficiencies, $r = 40$, for Figure 2.1(a).

(b) Individual nucleic acid efficiencies, $r = 40$, for Figure 2.1(b).

Figure 2.2 The individual nucleic acid efficiencies after 40 rounds. The bar graphs are not identical because, in Figure 2.1(a), selection is for indices $\{8, 9, 10, 12, 16\}$ while, in Figure 2.1(b), selection is for indices $\{8, 9, 10, 16\}$.

Figure 2.3 After 125 rounds it would appear from panel (a) that the selected indices are $\{1, 6, 8, 9, 10, 12, 16\}$ whereas from panel (b) after 400 rounds the selected indices would appear to be $\{1, 8, 9, 10, 12, 16\}$. Compare with Figure 2.1(a). Thus, the importance of reducing the total concentration of target from round to round in a systematic way is even more pronounced in the multiple target problem.
Figure 2.4 In panel (a), the dependence of the number of selected nucleic acids on the nucleic acid pool size is illustrated. In panel (b), $pK_a$ decreases with decreasing pool size, i.e., $K_a$ decreases with increasing pool size.

Figure 2.5 Nonmonotonicity of $pK_a$ (or, equivalently, $pK_d$) as a function of round number. The left-hand panel traces $pK_a^{(r)}$ for various pool sizes up to 120 rounds. The right-hand panel is a continuation of these curves for an expanded scale on the vertical axis.
Figure 2.6  Panel (a) is a graph of the maximal target affinity function when its domain has been projected onto the plane. Using the SELEX iteration scheme for 2,000 rounds, the panel (c) was generated from a fixed initial pool of nucleic acids \( \hat{F} = (F_1, F_2, F_3, F_4, F_5) = (0.3002, 0.0731, 0.1917, 0.1535, 0.2815) \). (The number of rounds is large because the rate of convergence of the SELEX program slows as the nucleic acid pool size decreases.) In panels (c,d), the indicated regions are labeled with the indices of the nucleic acids that will be selected when the initial target is selected from the interior of the indicated region. The red triangle in panel (b) generates the red hexagon in panel (d). (There is a corresponding hexagon for panel (c) that is omitted in the interests of clarity.) The initial target vectors \( \langle \Omega_1, \Omega_2, \Omega_3 \rangle \) were generated by setting \( \Omega_3 = 1 - \Omega_1 - \Omega_2 \), where for \( j = 1, \ldots, 199 \), \( i = 1, \ldots, 200 - j \) and \( \Omega_1 = \Omega_1(i,j) = j/200, \Omega_2 = \Omega_2(i,j) = i/200 \) or about 20,000 initial target vectors.
Figure 2.7  Panels (a) and (b) illustrate the graphs of chemical potentials defined as $Q$ in (2.56), at infinite target dilution. The minimum values of these functions are indicated by $\star$’s. Then the corresponding final NA fractions are (a) $\hat{F} = \langle 0.2400, 0.2400, 0.5200 \rangle$ with the minimum value of $-7.9328$ and (b) $\hat{F} = \langle 0.6900, 0.0100, 0.3000 \rangle$ with the minimum value of $-7.9992$. Using the SELEX iteration scheme the final NA fractions are (c) $\tilde{F} = \langle 0.2449, 0.2355, 0.5196 \rangle$ with $-\ln K_a = -7.9369$ after 40 rounds, and (d) $\tilde{F} = \langle 0.6984, 0.0000, 0.3016 \rangle$ with $-\ln K_a = -8.0001$ after 30 rounds. The values of $\hat{F}$ at the critical points in panels (a,b) are in quite good agreement with the values of $\tilde{F}$, found by iteration using the SELEX scheme, in panels (c,d), respectively.
Figure 2.8 The panels (a,c), were generated in the same manner as panels (c,d) in Figure 2.6. In all four panels, the subregions are labeled with the indices of the nucleic acids that will be selected when the initial target vector is in the indicated region. We refer to the case illustrated in panels (c,d) as improper because the minimum of the maximal target affinity function, $\varphi$, is defined by the intersection of any three of the four planes that define its graph.
Figure 2.9 These figures show that whatever convex hull in Figure 2.8, panel (a) (or panel (b)), to which the initial target fraction vector belongs, the SELEX process converges to a unique set of final nucleic acid fractions independently of the distribution of starting nucleic acids in the starting pool. Each of the five curves in each panel correspond to an independent trial for the starting pool of nucleic acids. (The vertical axis notation $|F_1 - F_j|$ is shorthand for $|\hat{F}_1^{(r)} - \hat{F}_j^{(r)}|_1 = \sum_{i=1}^5 |F_{i,1}^{(r)} - F_{i,j}^{(r)}|$. Here $j = 2, \ldots, 6$ for five of the six random starting vectors. The starting ordinates of the norms $|F_1 - F_j|$ for each of the curves in panels (a)-(d) are recorded in the fourth column of Tables 5.3-5.6, in Appendix C, respectively.)
Figure 2.10 The graphs show that the SELEX process is not globally asymptotically stable when the initial target fraction belongs to the convex hull of an improper face (panel (a)) and is asymptotically stable when it belongs to a convex hull corresponding to a proper face (panel (b)). Each of the five curves in both panels correspond to an independent trial for the starting pool of nucleic acids. (The vertical axis notation $|F_1 - F_j|_1$ is shorthand for $|\hat{F}_1^{(r)} - \hat{F}_j^{(r)}|_1 = \sum_{i=1}^{5} |F_i^{(r)} - F_{i,j}^{(r)}|$. Here $j = 2, \ldots, 6$ for five of the six random starting vectors. The starting ordinates of the norms $|F_1 - F_j|$ for each of the curves in panels (a)-(d) are recorded in the fourth column of Tables 5.7, 5.8, in Appendix C, respectively.)
CHAPTER 3. A mathematical analysis of alternate SELEX against multiple targets

3.1 Introduction

At an early stage of the body’s developmental process, cells differentiate to become more specialized cell types and then become fixed in their specialty, losing the ability to change. Cancer cells, however, are abnormal cells that divide without control and are able to invade other tissues. Nucleic acid aptamers that are capable of distinguishing tumor cells from normal cells, or cells in various stages of differentiation, can be used to detect tumor cells in cancer research and other clinical studies. It is shown in [33] that nucleic acid aptamers are able to discriminate between differentiated cells and undifferentiated cells (parental cells) by the SELEX method in a selection procedure called “subtraction”. That is, before each round of positive SELEX, nucleic acids that have bound to undifferentiated cells are removed during the subtraction procedure. The nucleic acid species that remain unattached are then selected and used for the positive SELEX process. By repeating this process, the pool of nucleic acids is enriched with aptamers specific for differentiated cells. The subtraction process is equivalent to what we define as negative SELEX.

In this chapter we propose a mathematical model for the alternate SELEX process. We call the SELEX procedure illustrated in the preceding chapter positive SELEX. Negative SELEX refers to a procedure performed to remove NA species that are not desired so that one can minimize enrichment of non-specifically binding nucleic acids and nucleic acids with lower affinity for the desired target over the remaining possible targets (or equivalently, maximize enrichment of nucleic acids with higher affinity for the desired target over those not desired). Alternate SELEX is iteratively done by doing positive SELEX followed by negative SELEX.
We also use the terms “positive targets” and “negative targets” to distinguish between the target species, when they are bound to nucleic acid molecules during positive SELEX and negative SELEX, respectively. Here, negative targets (undesired targets) can be the same type of species as the ones used during the positive selection steps, except that one or more of the target components are missing during negative SELEX. (The missing target species during negative SELEX are the desired target species.) The difference between positive SELEX and negative SELEX is in the selection step. During the selection step in negative SELEX, unbound nucleic acid species are retained instead of the bound NA species being retained as is done in positive SELEX. We refer to this step as the “negative selection step”.

Once target-NA bound products are separated from free (unbound) nucleic acids, the binding nucleic acids are eluted and amplified to increase the size of the pool of nucleic acid species that bound preferentially to the positive targets. The enriched population of nucleic acids is then exposed to the negative target species. At this point, we remove the bound product, retain the free nucleic acid species and amplify them. By alternating the two processes, we obtain a refined population of nucleic acids, and this results in selectivity and specificity. See [8], [12] and [30] for various aspects of performing the negative selection processes. (The SELEX procedure that involves negative selection (or subtractive selection) is also referred to as subtractive SELEX.) Numerical analysis and its simulation results of these processes are also illustrated in [8].

We present a number of mathematical and experimental arguments and discuss specificity and selectivity. Figure 3.1 illustrates the alternate selection steps starting with positive SELEX followed by negative SELEX or vice versa.

### 3.2 Formulation of a single round of multiple-target positive SELEX

For a single round of positive SELEX, we use the formulation from the multiple-target SELEX in Chapter 2, by replacing the notation for the target, \( T \), and free target, \( T_f \), by the positive target, \( T_s \), and the free positive target, \( T_{sf} \), respectively. The following formulas are
Figure 3.1 The steps of the alternate SELEX iteration process: One can go around one loop $R_s$ times and around the other $R_n$ times for one grand round, $R = R_s + R_n$.

then available:

$$[T_{s,i}] = [T_{s,f}] + \sum_{j=1}^{N} \left( \{T_{s,j} : NA_j\} \right) = [T_{s,f}] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{s,j}}\right), \text{ for } i \in \mathcal{M} = \{1, 2, \cdots, M\},$$

(3.1)

where $D_{j,f} = \sum_{l=1}^{M} [T_{s,f}] A_{lj}$. Using the unit positive target fractions $\Omega_{s,i}$ and free target fractions $\omega_{s,i}$, as well as the vector notations, we have a system of nonlinear equations for free targets:

$$\Omega_{s,i} = \omega_{s,i} \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{s,j}}\right) [T_{s,f}] = \omega_{s,i} W_s \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + [T_{s,f}] A^{j} \cdot \hat{\omega}_{s}}\right)$$

(3.2)

for $i = 1, \ldots, M$, where $W_s \equiv \frac{[T_{s,f}]}{[T_{s}]}$, and

$$[T_s] = [T_{s,f}] \left(1 + \frac{[NA]}{[T_{s,f}]} \sum_{j=1}^{N} \frac{F_j D_{s,j}}{1 + D_{s,j}}\right) = [T_{s,f}] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A^{j} \cdot \hat{\omega}_{s}}{1 + [T_{s,f}] A^{j} \cdot \hat{\omega}_{s}}\right).$$

(3.3)

The updated nucleic acid fractions bound to targets are given as:

$$F'_j = \frac{\left\{T : NA_j\right\}}{[T : NA]} = \frac{D_{s,j} F_j}{\sum_{l=1}^{N} D_{s,l} F_l}, \text{ for } i \in \mathcal{N} = \{1, 2, \cdots, N\}.$$  

During the positive selection steps, we have used the optimal choice of infinite target dilution to maximize the efficiencies. Here, we adopt the same strategy to the iteration scheme.
That is, as the round number, \( r \), becomes large, \([T_{s,i}^{(r)}]\) → 0, for \( i \in \mathcal{M} \), and \([T_{s}^{(r)}]\) → 0. This results in \([T_{s,i}^{(r)}]\) → 0 for \( i \in \mathcal{M} \) as well as \([T_{s}^{(r)}]\) → 0. We can use the following approximations:

\[
\Omega_{s,i} \approx \omega_{s,i} W_s \left( 1 + [N A] \sum_{j=1}^{N} F_j A_{ij} \right), \quad i \in \mathcal{M},
\]

and

\[
\frac{F_j'}{F_l'} \approx \left( \frac{\hat{A}^j \cdot \hat{\omega}_s}{\hat{A}^l \cdot \hat{\omega}_s} \right) \frac{F_j}{F_l}, \quad j, l \in \mathcal{N}.
\]

Recall from Chapter 2 that we defined the maximal target affinity function by

\[
\varphi(\hat{\omega}) = \max \left\{ \hat{A}^j \cdot \hat{\omega} \mid j \in \mathcal{N} \right\} \text{ for } \hat{\omega} \in \mathcal{S},
\]

where the simplex \( \mathcal{S} = \{ \hat{\omega} = (\omega_1, \ldots, \omega_M) \mid \omega_i \geq 0, \ i = 1, \ldots, M \ \text{and} \ \sum_{i=1}^{M} \omega_i = 1 \} \). Then \( \varphi(\hat{\omega}) \) can be considered as the affinity of the selected nucleic acid, as a function of the individual components of \( \hat{\omega} \).

### 3.3 Formulation of a single round of multiple-target negative SELEX

Here, we use the subscript “\( n \)” for the target parameters to indicate that the targets are used during the negative selection steps.

A chemical reaction governed by the negative SELEX process is:

\[
\{T_{n,i} : NA_j\} = T_{n,f_i} + NA_{f_j},
\]

where \( T_{n,i} \) and \( T_{n,f_i} \) are the \( i^{th} \) negative target and free target species, and \( NA_j \) and \( NA_{f_j} \) are the \( j^{th} \) bound and unbound nucleic acid species, respectively. Again, by employing the mass conservation laws, we can construct a nonlinear system of equations for the free negative target components \( T_{n,f_i} \), for \( i \in \mathcal{M} \), as follows:

\[
[T_{n,i}] = [T_{n,f_i}] \left( 1 + [N A] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + \sum_{l=1}^{M} [T_{n,f_i}] A_{lj}} \right) = [T_{n,f_i}] \left( 1 + [N A] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{n,j}} \right),
\]

where \( D_{n,j} = \sum_{l=1}^{M} [T_{n,f_i}] A_{lj} = [T_{n,f}] \cdot \hat{A}^j \). We also have

\[
[T_{n}] = [T_{n,f}] \left( 1 + \frac{[N A] [T_{n,f}]}{[T_{n,f}]} \sum_{j=1}^{N} \frac{F_j D_{n,j}}{1 + D_{n,j}} \right),
\]
as well as
\[ \Omega_{n,i} = \omega_{n,i} W_n \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + [T_n f] A_j \cdot \hat{\omega}_n} \right), \quad (3.9) \]
where \( W_n \equiv \frac{[T_n f]}{[T_n]} \), and \( \omega_{n,i} \) and \( \hat{\omega}_n \) are the free (negative) target fractions and the free target vector, respectively.

Instead of obtaining binding nucleic acids, we now retain the free nucleic acids during the negative selection steps. The updated nucleic acid fractions are now computed from:
\[ F'_j = \frac{[NAf_j]}{[NAf]} = \frac{[NA] \frac{F_j}{1 + D_{n,j}}}{[NA] \sum_{l=1}^{N} \frac{F_l}{1 + D_{n,l}}} = \frac{F_j}{\sum_{l=1}^{N} \frac{F_l}{1 + D_{n,l}}}, \quad j \in \mathcal{N}, \quad (3.10) \]
where we have used the fact that \([NAf] = \sum_{j=1}^{N} [NAf_j] = [NA] \sum_{j=1}^{N} \frac{F_j}{1 + D_{n,j}} \). Also \( \sum_{j=1}^{N} F'_j = 1 \).

The ratios of the nucleic acid fractions, for \( j, l \in \mathcal{N} \), are
\[ \frac{F'_j}{F'_l} = \frac{(1 + D_{n,l}) F_j}{(1 + D_{n,j}) F_l} = \frac{(1 + [T_n f] \overline{A} \cdot \hat{\omega}_n) F_j}{(1 + [T_n f] A_j \cdot \hat{\omega}_n) F_l}. \quad (3.11) \]
This determines if we want to use a choice of infinite target dilution, the initial pool and the limiting pool of nucleic acids would agree. Thus it takes more rounds for this system to have selection. Here, we choose to keep the total (negative) target concentration fixed, as well as the fractions, such that \([NA]/[T_n] \) is very small at every round. In this way, we can adopt the approximations of \([T_n f] \approx [T_n] \) and \( W_n \approx 1 \), by observing the equations (3.8) and (3.9).
Moreover, \( \Omega_{n,i} \approx \omega_{n,i} \). This choice of target concentration is reasonable, because during the negative selection we want to remove the bound target product as much as possible, and this can be achieved by increasing the size of a pool of target species. The more the (negative) target molecules are available, the higher the chances of the non-specifically binding nucleic acids are removed. We could then consider the following two options:

1. One could start with a pool of nucleic acids such that \([NA]/[T_n] \) is fairly small (i.e., take the \([T_n] \) relatively larger than \([NA] \)) and then fix both values of concentration from round to round. The use of PCR is needed to bring both pool sizes to some fixed value.

2. With the target size \([T_n] \) fixed and a given \([NA] \), use the pool size of \([NAf] \) instead of \([NA] \) from round to round. Then since \([NAf] \) becomes smaller (see the remark below)
from round to round the ratio $[NA]/[T_n]$ becomes small. This can be completed without using PCR.

Here, we used the first strategy for our simulations. See panels in Figures 3.2 and 3.3, and panel (d) in Figures 3.4 and 3.7 for the optimal choice of negative target concentration. For all four cases, we fixed $[NA] = 1$ at every round.

Using the choice of $[NA]/[T_n]$ small, we can make the following approximation:

$$
\frac{F'_j}{F'_l} \approx \left(1 + \frac{[T_n] A^j \cdot \hat{\omega}_n}{1 + [T_n] A^j \cdot \hat{\omega}_n}\right) \frac{F_j}{F_l} \approx \left(1 + \frac{[T_n] A^l \cdot \hat{\Omega}_n}{1 + [T_n] A^l \cdot \hat{\Omega}_n}\right), \quad \text{for all } j, l \in \mathcal{N}. \quad (3.12)
$$

If $A^j \cdot \hat{\Omega}_n > A^l \cdot \hat{\Omega}_n$, then the fraction $F_j$ decreases relative to $F_l$ from round to round during negative selection. If we consider the quantity $A^j \cdot \hat{\Omega}_n$ as the $j^{\text{th}}$ affinity vector, more contribution in affinity to the $j^{\text{th}}$ nucleic acids will actually result in less contribution in its selection.

**Remark 3.3.0.1.** One of the reasons we take the first strategy for the simulation is that, we realized the fairly huge drop in the total NA concentration in a few rounds of negative selection. The reason for this can be seen using the following approximation:

$$
\frac{[NAf]}{[NA]} = \sum_{j=1}^{N} \frac{F_j}{1 + D_{nj}} \leq \frac{1}{1 + [T_{fn}] \min\{A^j \cdot \hat{\omega}_n, \ j = 1, \ldots, N\}} \leq \frac{1}{1 + [T_{fn}] \min\{A_{ij}, \ i = 1, \ldots, M, \ j = 1, \ldots, N\}}.
$$

From (3.8) we have, for $i \in \mathcal{M}$, that

$$
W_n \equiv \frac{[T_{nf}]}{[T_n]} = \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A^j \cdot \hat{\omega}_n}{1 + [T_{nf}] A^j \cdot \hat{\omega}_n}\right)^{-1}.
$$

(3.13)

Suppose that the size of the affinities $A_{ij}$ are larger than $10^3 (\mu M^{-1})$, as used in our simulations, and the target is in excess, i.e., $[T_n]$ is relatively larger than $[NA]$, say $\frac{[NA]}{[T_n]} \leq \frac{1}{2}$. Then, from (3.13) we have $[T_n] - [NA] \leq [T_{nf}] \leq [T_n]$, and thus $2[NA] < [T_n]$ implies that $[NA] \leq [T_n] - [NA] \leq [T_{fn}]$. Therefore

$$
\frac{[NAf]}{[NA]} \leq \frac{1}{1 + 10^3 |NA|}.
$$
This tells us that with each negative selection step, the drop is by a factor of $10^{-3}$ or more. Considering that we use the affinity range of order $10^3$ per micromolar, after seven consecutive rounds of negative selection, we would have less than one nucleic acid molecule per liter.

### 3.4 Alternate SELEX iteration scheme: Positive SELEX followed by negative SELEX

We now formulate the alternate SELEX procedure. Here, we perform the alternate SELEX procedure by beginning with positive SELEX followed by negative SELEX. The round numbers of the positive and negative SELEX processes can be chosen in many different ways. Therefore, the following definitions are useful to analyze when we alternate several different rounds of positive and negative selections.

**Definition 3.4.0.1.** Let $R = R_s + R_n$, where $R_n$ and $R_s$ are nonnegative integers. When the alternate SELEX procedure is performed in such a way that one starts with $R_s$ rounds of positive SELEX followed by $R_n$ rounds of negative SELEX, we call $R$ the grand round number.

If $R_n = 0$, then we only perform the positive SELEX, while if we take $R_s = 0$ then the performance of negative SELEX alone will be made. Since the numbers of positive and negative selection procedures can be different we also define the ratio of the two round numbers.

**Definition 3.4.0.2.** Let $\lambda = \frac{R_s}{R}$ be the fraction of round numbers, which corresponds to positive selection. Then we call $\lambda$ the selection fraction. Also $1 - \lambda = \frac{R_n}{R}$.

Again, when $\lambda = 1$, we only perform positive SELEX, while, only negative SELEX is performed for $\lambda = 0$.

To make a transition rule for the SELEX iteration scheme, from positive selection to negative selection, we first look at the round number transitions. By iterating $R_s$ rounds of positive selection followed by $R_n$ rounds of negative selection, the round number $r$ is defined as follows, for $k = 0, 1, 2, \ldots$:

$$
r = kR + \begin{cases} 
  l, & \text{if } l = 1, \ldots, R_s \text{ and } r \text{ is a positive selection round;} \\
  R_s + l, & \text{if } l = 1, \ldots, R_n \text{ and } r \text{ is a negative selection round}
\end{cases}. \quad (3.14)
$$
Therefore, if \( kR < r \leq kR + R_s \), we are performing positive selection, and if \( kR + R_s < r \leq (k + 1)R \), then negative selection is performed. Now if we apply this transition rule to the positive and negative selection equations in the preceding two sections, for \( kR < r \leq kR + R_s \), we have

\[
\Omega^{(r)}_{s,i} = \omega^{(r)}_{s,i} W^{(r)}_s \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j^{(r)} A_{ij}}{1 + [T_s f]^{(r)} A j \cdot \hat{\omega}^{(r)}_s} \right)
\]

(3.15)

for \( i \in M \), where \( W^{(r)}_s = \frac{T_s f}{{T_s}^r} \),

\[
[T_s]^{(r)} = [T_s f]^{(r)} \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j^{(r)} \overrightarrow{A j} \cdot \hat{\omega}^{(r)}_s}{1 + [T_s f]^{(r)} \overrightarrow{A j} \cdot \hat{\omega}^{(r)}_s} \right),
\]

(3.16)

and

\[
F_j^{(r+1)} = \frac{\frac{D_{s,j}^{(r)}}{1 + D_{s,j}^{(r)}} F_j^{(r)}}{\sum_{l=1}^{N} \frac{D_{s,l}^{(r)}}{1 + D_{s,l}^{(r)}} F_l^{(r)}}, \quad \text{for } i \in N.
\]

(3.17)

For \( kR + R_s < r \leq (k + 1)R \), we have

\[
\Omega^{(r)}_{n,i} = \omega^{(r)}_{n,i} W^{(r)}_n \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j^{(r)} A_{ij}}{1 + [T_n f]^{(r)} A j \cdot \hat{\omega}^{(r)}_n} \right),
\]

(3.18)

\[
[T_n]^{(r)} = [T_n f]^{(r)} \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j^{(r)} \overrightarrow{A j} \cdot \hat{\omega}^{(r)}_n}{1 + [T_n f]^{(r)} \overrightarrow{A j} \cdot \hat{\omega}^{(r)}_n} \right),
\]

(3.19)

and

\[
F_j^{(r+1)} = \frac{\frac{D_{n,j}^{(r)}}{1 + D_{n,j}^{(r)}} F_j^{(r)}}{\sum_{l=1}^{N} \frac{D_{n,l}^{(r)}}{1 + D_{n,l}^{(r)}} F_l^{(r)}}, \quad \text{for } j \in N.
\]

(3.20)

The optimal target strategies for positive and negative selection mentioned above can be applied to this transition rule as follows: for the positive target,

\[
[T_s]^{(kR+l)} = \begin{cases} 
[T_s]^{(0)}, & \text{if } l = 1, \ldots, R_s; \\
0, & \text{if } l = R_s + 1, \ldots, R,
\end{cases}
\]

(3.21)

and for the negative target,

\[
[T_n]^{(kR+l)} = \begin{cases} 
0, & \text{if } l = 1, \ldots, R_s; \\
[T_n]^{(0)}, & \text{if } l = R_s + 1, \ldots, R,
\end{cases}
\]

(3.22)
where $[T_s]^{(0)}$ and $[T_n]^{(0)}$ denote the initial target concentrations. For the total nucleic acid pool size $[NA]$ during positive and negative selection, we fix both $[NA]$ at every round with the use of PCR. For the total NA concentration during negative selection, however, we fix $[NA]$ small such that $[T_n]/[NA]$ is large.

To examine the iterative process of alternate SELEX we now make a transition for the nucleic acid fractions using the equations (3.17) and (3.20). Since the transitions from positive selection to negative selection are made at round when $r = 0, R, 2R, 3R, \ldots$, and the transition from negative selection to negative selection is always made at round numbers $R_s, R + R_s, 2R + R_s, 3R + R_s, \ldots$, we can write, for $k = 0, 1, 2, \ldots$:

$$
\frac{F_j^{(r+1)}}{F_l^{(r+1)}} = \frac{F_j^{(r)}}{F_l^{(r)}} \left\{ \begin{array}{ll}
(1+D^{(r)}_s D^{(r)}_n) & \text{for } r = kR + p, \ p = 1, \ldots, R_s; \\
1 + D^{(r)}_n & \text{for } r = kR + R_s + q, \ q = 1, \ldots, R_n
\end{array} \right. 
$$

(3.23)

We can then consider the following sequences of parameters for each of the positive and negative selection processes: for fixed $p \in \{1, \ldots, R_s\}$, the $R_s$ positive selection sequences are denoted by

$$
\{ \hat{F}_s^{(p + kR)}, \hat{\omega}_s^{(p + kR)}, [T_s f]^{(p + kR)} \}_{k=0}^{\infty} = \{ \hat{F}_s^{(p,kR)}, \hat{\omega}_s^{(p,kR)}, [T_s f]^{(p,kR)} \}_{k=0}^{\infty}, 
$$

(3.24)

and for fixed $q \in \{1, \ldots, R_n\}$, the $R_n$ negative selection sequences are denoted by

$$
\{ \hat{F}_n^{(q + kR)}, \hat{\omega}_n^{(q + kR)}, [T_n f]^{(q + kR)} \}_{k=0}^{\infty} = \{ \hat{F}_n^{(q,kR)}, \hat{\omega}_n^{(q,kR)}, [T_n f]^{(q,kR)} \}_{k=0}^{\infty}, 
$$

(3.25)

where we have now used the subscripts “$s$” and “$n$” to indicate nucleic acid fractions for positive and negative selections, respectively. Note, as shown in Chapter 2, the sequences in (3.24) converge.

Then the transition rules for every first round of positive selection and negative selection are followed by:

$$
\hat{F}_n^{(1,kR)} = \hat{F}_s^{(R_s + 1,kR)} \text{ for } k \geq 0
$$

and

$$
\hat{F}_s^{(1,(k+1)R)} = \hat{F}_n^{(R_n + 1,kR)} \text{ for } k \geq 0.
$$
For example, for \( k = 0 \), after first \( R_s \) rounds of positive selection, \( r = R_s \), the updated nucleic acid fractions for the next round \( F_{s,j}^{(r+1)} \) are computed. Then, they are used as the starting nucleic acid fractions for negative selection, by setting \( F_{n,j}^{(1)} = F_{s,j}^{(R_s+1)} \), et cetera.

Moreover, the transition using the equation becomes for \( k \geq 0 \),

\[
\frac{F_{n,j}^{(1,kR)}}{F_{n,l}^{(1,kR)}} = \frac{F_{s,j}^{(R_s+1,kR)}}{F_{s,l}^{(R_s+1,kR)}} = \left( \prod_{p=1}^{R_s} \frac{D_{s,j}^{(p,kR)} (1 + D_{s,l}^{(p,kR)})}{D_{s,l}^{(p,kR)} (1 + D_{s,j}^{(p,kR)})} \right) \frac{F_{s,j}^{(1,kR)}}{F_{s,l}^{(1,kR)}}
\]

and

\[
\frac{F_{s,j}^{(1,kR)}}{F_{s,l}^{(1,kR)}} = \left( \prod_{q=1}^{R_s} \frac{1 + D_{n,j}^{(q,kR)}}{1 + D_{n,l}^{(q,kR)}} \right) \frac{F_{n,j}^{(1,kR)}}{F_{n,l}^{(1,kR)}},
\]

in terms of the nucleic acid fractions at the beginning of each round of positive and negative selection. The ratios of nucleic acid fractions for \( p \in \{2, \ldots, R_s \} \) and \( q \in \{2, \ldots, R_n \} \), are then

\[
\frac{F_{s,j}^{(p,kR)}}{F_{s,l}^{(p,kR)}} = \left( \prod_{p' = 1}^{p-1} \frac{D_{s,j}^{(p',kR)} (1 + D_{s,l}^{(p',kR)})}{D_{s,l}^{(p',kR)} (1 + D_{s,j}^{(p',kR)})} \right) \frac{F_{s,j}^{(1,kR)}}{F_{s,l}^{(1,kR)}},
\]

\[
\frac{F_{n,j}^{(q,kR)}}{F_{n,l}^{(q,kR)}} = \left( \prod_{q' = 1}^{q-1} \frac{1 + D_{n,j}^{(q',kR)}}{1 + D_{n,l}^{(q',kR)}} \right) \frac{F_{n,j}^{(1,kR)}}{F_{n,l}^{(1,kR)}},
\]

Therefore

\[
F_{s,l}^{(p,kR)} = \frac{\left( \prod_{p' = 1}^{p-1} \frac{D_{s,j}^{(p',kR)} (1 + D_{s,l}^{(p',kR)})}{D_{s,l}^{(p',kR)} (1 + D_{s,j}^{(p',kR)})} \right) F_{s,l}^{(1,kR)}}{\sum_{j=1}^{N} \left( \prod_{p' = 1}^{p-1} \frac{D_{s,j}^{(p',kR)} (1 + D_{s,l}^{(p',kR)})}{D_{s,l}^{(p',kR)} (1 + D_{s,j}^{(p',kR)})} \right) F_{s,j}^{(1,kR)}} \quad \text{for} \quad p = 2, \ldots, R_s,
\]

\[
F_{n,l}^{(q,kR)} = \frac{\left( \prod_{q' = 1}^{q-1} \frac{1 + D_{n,j}^{(q',kR)}}{1 + D_{n,l}^{(q',kR)}} \right) F_{n,l}^{(1,kR)}}{\sum_{j=1}^{N} \left( \prod_{q' = 1}^{q-1} \frac{1 + D_{n,j}^{(q',kR)}}{1 + D_{n,l}^{(q',kR)}} \right) F_{n,j}^{(1,kR)}} \quad \text{for} \quad q = 2, \ldots, R_n.
\]

### 3.5 Limiting behavior of the alternate SELEX iteration scheme and maximal target affinity function

Now, consider the limiting behavior of the alternate SELEX iteration scheme. Denote the limiting values of the sequences in (3.24) and (3.25) by

\[
\{ \tilde{F}_{s}^{(p)}, \tilde{\omega}_{s}^{(p)}, [T_{s}f]^{(p)} \}, \quad \{ \tilde{F}_{n}^{(q)}, \tilde{\omega}_{n}^{(q)}, [T_{n}f]^{(q)} \},
\]
for each \( p \in \{1, \cdots, R_s\} \) and \( q \in \{1, \cdots, R_n\} \). When \( p = 1 \), the limiting positive selection fractions are given by

\[
\frac{F_{s,j}^{(1)}}{F_{s,l}^{(1)}} = \left( \prod_{q=1}^{R_n} \frac{1 + D_{n,l}(q)}{1 + D_{n,j}(q)} \right) \left( \prod_{p=1}^{R_s} \frac{D_{s,j}(p)(1 + D_{s,l}(p))}{D_{s,l}(p)(1 + D_{s,j}(p))} \right) \frac{F_{s,j}^{(1)}}{F_{s,l}^{(1)}}.
\]

Then, using the approximation made with optimal target choices in (3.6) and (3.12), we have

\[
\left( \prod_{q=1}^{R_n} \frac{1 + D_{n,l}(q)}{1 + D_{n,j}(q)} \right) \left( \prod_{p=1}^{R_s} \frac{D_{s,j}(p)(1 + D_{s,l}(p))}{D_{s,l}(p)(1 + D_{s,j}(p))} \right) \approx \frac{1 + [T^n(0) \tilde{A}^j \cdot \tilde{\Omega}_n]}{1 + [T^n(0) \tilde{A}^l \cdot \tilde{\Omega}_n]} \left( \prod_{p=1}^{R_s} \frac{\tilde{A}^l \cdot \tilde{\omega}_s^{(p)}}{\tilde{A}^l \cdot \tilde{\omega}_s^{(p)}} \right).
\]

By the similar argument made in the preceding chapter for positive selection, the approximation on the right-hand side of the above equation tells us that the quantities will also serve as the affinities for each nucleic acid and contribute to a determination of the set of selected nucleic acid indices. Then we can also write the approximation, for each nucleic acid index \( j \), as

\[
\left( \prod_{p=1}^{R_s} \frac{\tilde{A}^j \cdot \tilde{\omega}_s^{(p)}}{\tilde{A}^j \cdot \tilde{\omega}_s^{(p)}} \right) \frac{(1 + [T^n(0) \tilde{A}^j \cdot \tilde{\Omega}_n])^{-R_n}}{(1 + [T^n(0) \tilde{A}^l \cdot \tilde{\Omega}_n])^{(1-\lambda)/\lambda}} = \frac{R_n}{1 - \lambda} = \frac{R_n}{R_s}.
\]

Let us now define the \( j^{th} \) column (vector) of the matrix \( C_{\lambda} \) by

\[
\tilde{C}^j_{\lambda} = [1 + [T^n(0) \tilde{A}^j \cdot \tilde{\Omega}_n])^{-\lambda/(1-\lambda)} \tilde{A}^j.
\]

**Definition 3.5.0.3.** We call the matrix \( C_{\lambda} \) the specificity matrix for alternate selection with selection fraction \( \lambda \).

Then

\[
\tilde{C}^j_{\lambda} \cdot \tilde{\omega} = \frac{\tilde{A}^j \cdot \tilde{\omega}}{[1 + [T^n(0) \tilde{A}^j \cdot \tilde{\Omega}_n])^{(1-\lambda)/\lambda}}.
\]

We can consider this dot product as the affinity of the \( j^{th} \) nucleic acid by rescaling from positive selection by negative selection. Moreover, \( [T^n(0) \tilde{A}^j \cdot \tilde{\Omega}_n] \) are known here.

Now, we define the **maximal target affinity function** as follows:

\[
\varphi_R(\lambda, \tilde{\omega}_s^{(1)}, \ldots, \tilde{\omega}_s^{(R_s)}) = \max \left\{ \varphi_j(\lambda, \tilde{\omega}_s^{(1)}, \ldots, \tilde{\omega}_s^{(R_s)}) \mid j \in \mathcal{N} \right\},
\]

where the target affinity functions are defined for each nucleic acid index \( j \in \mathcal{N} \), as

\[
\varphi_j(\lambda, \tilde{\omega}_s^{(1)}, \ldots, \tilde{\omega}_s^{(R_s)}) = \left( \prod_{p=1}^{R_s} \tilde{C}^j_{\lambda} \cdot \tilde{\omega}_s^{(p)} \right)^{1/R_s}, \text{ for } \tilde{\omega}_s^{(p)} \in \mathcal{S} \text{ and } p = 1, \ldots, R_s.
\]
We also define the NA index set corresponding to the limiting nucleic acids, by
\[
\mathcal{L}_R(\lambda, \widehat{\omega}_s^{(1)}, \widehat{\omega}_s^{(2)}, \ldots, \widehat{\omega}_s^{(R_s)}) = \left\{ l \in \mathcal{N} \mid \varphi_R(\widehat{\omega}_s^{(1)}, \ldots, \widehat{\omega}_s^{(R_s)}) = \varphi_l(\lambda, \widehat{\omega}_s^{(1)}, \ldots, \widehat{\omega}_s^{(R_s)}) \right\}.
\]

Notice here that for \( \lambda = 1 \), we have \( \overline{C}_\lambda \cdot \widehat{\omega} = \overline{A} \cdot \widehat{\omega} \) and \( \varphi_R(\lambda, \widehat{\omega}_s^{(1)}, \ldots, \widehat{\omega}_s^{(R_s)}) = \varphi(\widehat{\omega}) \) in the case of positive SELEX. See Section 2.6.3 for more details.

The limiting nucleic acid fractions then satisfy
\[
F_{s,l}^{(p)} = \frac{\overline{A}^l \cdot \widehat{\omega}^{(p-1)} F_{s,l}^{(p-1)}}{\sum_{j=1}^{N} \overline{A}^l \cdot \widehat{\omega}^{(p-1)} F_{s,j}^{(p-1)}} = \frac{\gamma_{l}^{(p-1)} F_{s,l}^{(1)}}{\sum_{j=1}^{N} \delta_{j}^{(p-1)} F_{n,j}^{(1)}} = \frac{\delta_{j}^{(q-1)} F_{n,j}^{(1)}}{\sum_{j=1}^{N} \delta_{j}^{(q-1)} F_{n,j}^{(1)}} \quad \text{for} \quad p = 2, \ldots, R_s,
\]
\[
F_{n,l}^{(q)} = \frac{\prod_{n=1}^{N} \frac{1}{1 + [T_n(0)] A^l \Omega_n} F_{n,l}^{(q-1)}}{\sum_{j=1}^{N} \frac{1}{1 + [T_n(0)] A^l \Omega_n} F_{n,j}^{(q-1)}} = \frac{\delta_{n}^{(q-1)} F_{n,l}^{(1)}}{\sum_{j=1}^{N} \delta_{n}^{(q-1)} F_{n,j}^{(1)}} = \frac{\delta_{n}^{(q-1)} F_{n,j}^{(1)}}{\sum_{j=1}^{N} \delta_{n}^{(q-1)} F_{n,j}^{(1)}} \quad \text{for} \quad q = 2, \ldots, R_n,
\]
where we have defined the following:
\[
k_{s}^{(p-1)} = \sum_{j=1}^{N} \gamma_{j}^{(p-1)} F_{s,j}^{(1)}, \quad k_{n}^{(q-1)} = \sum_{j=1}^{N} \delta_{j}^{(q-1)} F_{n,j}^{(1)},
\]
\[
\delta_{n}^{(q-1)} = (1 + [T_n(0)] A^l \cdot \widehat{\omega}_s^{(p'-1)})^{-q+1} \quad \text{and} \quad \gamma_{l}^{(p-1)} = \prod_{p'=1}^{p-1} \overline{A}^l \cdot \widehat{\omega}_s^{(p')}.
\]

By the use of the transition rule, at the liming states, we have
\[
F_{s,l}^{(1)} = F_{s,l}^{(R_s+1)} = \frac{\delta_{l}^{(R_s)} F_{s,l}^{(1)}}{\sum_{j=1}^{N} \delta_{j}^{(R_s)} F_{n,j}^{(1)}} = \frac{\delta_{l}^{(R_s)} F_{s,l}^{(1)}}{k_{s}^{(R_s)}}, \quad \text{and}
\]
\[
F_{n,l}^{(1)} = F_{n,l}^{(R_s+1)} = \frac{\gamma_{l}^{(R_s)} F_{s,l}^{(1)}}{\sum_{j=1}^{N} \gamma_{j}^{(R_s)} F_{s,j}^{(1)}} = \frac{\gamma_{l}^{(R_s)} F_{s,l}^{(1)}}{k_{n}^{(R_s)}}.
\]
Therefore,
\[
F_{s,l}^{(1)} = \frac{\delta_{l}^{(R_s)} F_{s,l}^{(1)}}{k_{l}^{(R_s)} k_{s}^{(R_s)}} \iff \delta_{l}^{(R_s)} \gamma_{l}^{(R_s)} = k_{n}^{(R_s)} k_{s}^{(R_s)}
\]
for the indices, such that \( F_{s,l}^{(1)} \)’s are nonzero. Then, for \( j \in \mathcal{L}_R(\lambda, \widehat{\omega}_s^{(1)}, \widehat{\omega}_s^{(2)}, \ldots, \widehat{\omega}_s^{(R_s)}) \), we define the overall limiting association constant, \( K_a \), for alternate SELEX by
\[
\varphi_R(\lambda, \widehat{\omega}_s^{(1)}, \ldots, \widehat{\omega}_s^{(R_s)}) = (\gamma_{l}^{(R_s)} \delta_{l}^{(R_s)})^{1/R_s} = (k_{n}^{(R_s)} k_{s}^{(R_s)})^{1/R_s} \equiv K_a \quad \text{for} \quad l \in \mathcal{L}_R.
\]

If \( k_{s}^{(R_s)} \) and \( k_{n}^{(R_s)} \) are measurable and thus they are known, we can determine \( K_a \).
3.6 SELEX efficiency for alternate SELEX

Before we propose an issue of the specificity of the alternate SELEX problem, we define the efficiency for alternate SELEX.

For positive selection, we have defined the overall target efficiencies as follows (see also Section 2.2.3):

\[
E_s = \frac{[T_s : N.A]}{[T_s]} = 1 - \frac{[T_s f]}{[T_s]} = \frac{[N.A]K_s}{1 + [N.A]K_s} = 1 - W_s,
\]

(3.37)

where we have used the positive target notations, \(T_s\) and \(T_s f\), and \(K_s\) is the overall association constant (in Chapter 2, we denoted \(K_a\) as the overall association constant) for positive selection only. From (2.20),

\[
K_s = \sum_{j=1}^{N} \frac{F_{s,j} \tilde{A}^j \cdot \tilde{\omega}_s}{1 + D_{s,j}} \approx \sum_{j=1}^{N} \frac{F_{s,j} \tilde{A}^j \cdot \tilde{\omega}_s},
\]

(3.38)

where the approximation was made by the infinite target dilution.

For negative selection, the following definition would be appropriate:

\[
\tilde{E}_n = \frac{[T_n f]}{[T_n]} = \frac{[T_n] - [T_n : N.A]}{[T_n]} = \tilde{W}_n.
\]

(3.39)

However, when we perform the negative selection procedures, we remove one or more negative target components. Now, call such target components that are removed or remain subtargets. So, for example, if we have multiple (positive) targets with five components and remove one of them, we have two subtargets with one component and with four components. Since each target component contributes to binding nucleic acids differently, we need to define the negative target efficiency by considering two negative selection experiments: performing negative selection with all target species \([T]\) present and with the subtargets considered as negative targets, \([T_n]\).

Let \(\overrightarrow{T} = \overrightarrow{U}_n + \overrightarrow{T}_n = [U_n] \hat{u} + [T_n] \hat{t} = [T] \hat{\Omega}\) where \(\hat{\Omega}, \hat{u}\) and \(\hat{t}\) are unit vectors, i.e., their entries are positive and sum to unity. The vectors, \(\overrightarrow{U}_n\) and \(\overrightarrow{T}_n\), are then subtargets of the target vector \(\overrightarrow{T}\).

First, we perform the negative selection with \(\overrightarrow{T}\) as the negative target vector, and then perform another experiment with \(\overrightarrow{T}_n\) as the negative target vector. Then, we measure the
negative selection efficiency as follows:

\[ E_n = \frac{([T_n] - [NA : T_n])}{([T_n] - [NA : T_n]u)/[T_n]} = \frac{[T_n f]}{[T_n f] u}, \]  

where \([NA : T_n]_u = [NA : T] - [NA : U]\) is the concentration of bound nucleic acid due to \(T_n\) when \(U_n\) is present, and the subscript \(u\) denotes the concentration of the product bound to \(T_n\) when \(U_n\) is present. When \(U_n\) is not a part of the negative target components, the concentration of bound nucleic acid is \([NA : T_n]\), which results in \([NA : T_n] > [NA : T_n]u\). Therefore, \(E_n \in (0, 1)\). In both cases we use the same negative target concentration \([T_n]\) to rule out possible effects of the size of \([T_n]\), whether \(U_n\) is present or not.

Now, to evaluate the efficiency, we present the equations for computing the numerator and the denominator of (3.40) below. To compute the numerator, when \(U_n\) is absent, we use (3.7) to find the free targets \([T_n f1]\), and compute the \(D_{n,j} = D_{T_n f,j} = \sum_{i=1}^{M}[T_n f1]A_{ij}\). Then, \([T_n f]\) can be found from

\[ [T_n f] = \sum_{i=1}^{M}[T_n f1] = \sum_{i=1}^{M}[T_{n,i}] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + \sum_{i=1}^{M}[T_n f1] A_{ij}} \right)^{-1} \]

\[ = \sum_{i=1}^{M} \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + \sum_{i=1}^{M}[T_n f1] A_{ij}} \right) [T_{n,i}] \]  

(3.41)

To compute the denominator, when \(U_n\) is present, we again compute \([T f1]\) from (3.7), and we set \(D_{n,j} = D_{T f,j} = \sum_{i=1}^{M}[T f1]A_{ij}\). Then, we calculate \([T_n f]_u\) by writing

\[ [T_{n,i}] = [T_n f1]_u + [NA : T_{n,i}]_u = [T_n f1]_u + \sum_{j=1}^{N} [NA_j : T_{n,i}]_u \]

\[ = [T_n f1]_u \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{T f,j}} \right) \]

so that

\[ [T_n f]_u = \sum_{i=1}^{M}[T_n f1]_u = \sum_{i=1}^{M} \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{T f,j}} \right) [T_{n,i}] \]  

(3.42)

Now let \(\mathbf{\tilde{f}} = \mathbf{\Omega_n}\) be a unit subtarget vector of \(\mathbf{\Omega_s}\) and \(\mathbf{\tilde{u}} = \mathbf{\Omega_u}\) be its complimentary unit vector, and \(\mathcal{M}_1 \subset \mathcal{M} = \{1, 2, \ldots, M\}\). Then, define

\[ t_i = \Omega_{n,i} = \begin{cases} 0, & \text{if } i \notin \mathcal{M}_1; \\ \Omega_{s,i} / (\sum_{j \in \mathcal{M}_1} \Omega_{s,j}), & \text{if } i \in \mathcal{M}_1. \end{cases} \]
Define

\[ u_i = \Omega_{u,i} = \begin{cases} 
\Omega_{s,i}/(\sum_{j \notin M_1} \Omega_{s,j}), & \text{if } i \notin M_1; \\
0, & \text{if } i \in M_1. 
\end{cases} \]

If we perform negative selection with the vector \([T_n]\Omega_n\), then we use the vector \([T]\Omega_s = [U_n]\hat{u} + [T_n]\hat{t}\) with

\[ [T] = \frac{[T_n]}{\sum_{j \in M_1} \Omega_{s,j}} \quad \text{and} \quad [U_n] = \frac{[T_n] \sum_{j \notin M_1} \Omega_{s,j}}{\sum_{j \in M_1} \Omega_{s,j}}. \]

Then define the following ratio:

\[ \theta \equiv \frac{[U_n]}{[T_n]} = \frac{\sum_{j \notin M_1} \Omega_{s,j}}{\sum_{j \in M_1} \Omega_{s,j}}. \quad (3.43) \]

Notice that \(\theta\) does not depend on both \([T]\) and \([T_n]\), but does depend on the initial target fractions, \(\Omega_{s,j}\) and the subset, \(M_1\). Moreover, \(1 + \theta = \frac{1}{\sum_{j \in M_1} \Omega_{s,j}}\) and \((1 + \theta)\Omega_{s,j} = \Omega_{n,j}\).

We now calculate \(E_n\) for this choice of negative targets. When \([NA]/[T_n]\) is small, we have \([T_n] \approx [T_n]\Omega_n\), so that \(\omega_{n,i} \approx \Omega_{n,i}\). Then, \([NA]/[T]\) also becomes small, so we can have \(\omega_i \approx \Omega_{s,i}\).

The efficiency (3.40) is computed as follows:

\[ E_n = \frac{\sum_{i=1}^{M} \frac{l_i}{1+[NA] \sum_{j=1}^{N} F_j A_{ij}(1+[T_n]A^j \cdot \Omega_n)^{-1}} \Omega_{s,i}}{\sum_{i=1}^{M} \frac{l_i}{1+[NA] \sum_{j=1}^{N} F_j A_{ij}(1+[T_n]A^j \cdot \Omega_n)^{-1}}} = \frac{\sum_{i \in M_1} \frac{\Omega_{s,i}}{1+[NA] \sum_{j=1}^{N} F_j A_{ij}(1+[T_n]A^j \cdot \Omega_n)^{-1}} \Omega_{n,i}}{\sum_{i \in M_1} \frac{\Omega_{s,i}}{1+[NA] \sum_{j=1}^{N} F_j A_{ij}(1+[T_n]A^j \cdot \Omega_n)^{-1}}}. \quad (3.44) \]

If we consider the case that only one component is absent during negative selection, then \(M_1 = \{1, 2, \ldots, M - 1\}\). Suppose the numbers \(A_{Ml}\) are large relative to the numbers, \(A_{il}\) for \(i = 1, \ldots, M - 1\) and \(l\) in \(L_R\). Then the equation (3.44) has the following approximation available:

\[ E_n \approx \sum_{i \in M_1} \frac{\Omega_{s,i}(1 + \theta)}{1 + [NA] \sum_{l \in L_R} F_l A_{il}(1 + [T_n]A^l \cdot \Omega_n)^{-1}} \Omega_{n,i} = \sum_{i \in M_1} \frac{\Omega_{n,i}}{1 + [NA] \sum_{l \in L_R} F_l B_{il}}. \quad (3.45) \]

Let

\[ W_n = E_n = \sum_{i \in M_1} \frac{\Omega_{n,i}}{1 + [NA] \hat{F} \cdot \hat{B}_i} = \frac{1}{1 + [NA] \hat{K}_n}. \quad (3.46) \]

Then

\[ \hat{K}_n = \sum_{l \in L_R} F_l \hat{B}_l^T \cdot \hat{\Omega}_n. \quad (3.47) \]
and by comparing with the $K$ in (3.38) for positive selection, the quantity $K_n (\approx \sum_{l \in \mathcal{L}_R} F_l \vec{B}^l \cdot \hat{\omega}_n)$ can be considered as the limiting overall (equilibrium) constant for positive selection with matrix $B$.

Let us now look at the overall association constant, $\tilde{K}_n$. By definition, the overall dissociation constant for negative selection

$$\tilde{K}_n = \frac{[\{T_n : NA\}]}{[T_n f] [N A f]} = \frac{\sum_{j=1}^N F_{n,j} D_{T_n,f,j}/(1 + D_{T_n,f,j})}{[T_n f] \sum_{j=1}^N F_{n,j}/(1 + F_{n,j} D_{T_n,f,j})} \approx \frac{\sum_{j=1}^N F_{n,j} \vec{A}^j \cdot \hat{\Omega}_n/(1 + [T_n] \vec{A}^j \cdot \hat{\Omega}_n)}{\sum_{j=1}^N F_{n,j}/(1 + [T_n] \vec{A}^j \cdot \hat{\Omega}_n)}.$$  

(3.48)

Moreover,

$$\frac{1}{1 + [T_n] \tilde{K}_n} = \sum_{j=1}^N \frac{F_{n,j}}{1 + [T_n] \vec{A}^j \cdot \hat{\Omega}_n} \quad (3.49)$$

and

$$\frac{\tilde{K}_n}{1 + [T_n] \tilde{K}_n} = \sum_{j=1}^N \frac{F_{n,j} \vec{A}^j \cdot \hat{\Omega}_n}{1 + [T_n] \vec{A}^j \cdot \hat{\Omega}_n} \approx \sum_{j=1}^N \frac{F_{n,j} \vec{B}^j \cdot \hat{\Omega}_n}{1 + [T_n] \vec{A}^j \cdot \hat{\Omega}_n} \quad (3.50)$$

where the elements of matrix $B$ are defined by $B_{ij} = A_{ij}/(1 + [T_n] \vec{A}^j \cdot \hat{\Omega}_n)$. Then the efficiency defined in (3.39) becomes

$$\tilde{E}_n = \tilde{W}_n = \frac{[T_n f]}{[T_n]} = \frac{1 + [T_n f] \tilde{K}_n}{1 + ([N A] + [T_n f]) \tilde{K}_n} \approx \frac{1 + [T_n] \tilde{K}_n}{1 + ([N A] + [T_n]) \tilde{K}_n}$$  

(3.51)

We also have

$$\frac{\tilde{K}_n}{1 + [T_n] \tilde{K}_n} = K_n = \sum_{l \in \mathcal{L}_R} F_l \vec{B}^l \cdot \hat{\Omega}_n.$$  

(3.52)

$$E_n = \frac{1}{1 + [N A] K_n} = \frac{1}{1 + [N A]} \frac{K_n}{1 + [T_n] K_n} = \frac{1 + [T_n] \tilde{K}_n}{1 + ([N A] + [T_n]) \tilde{K}_n}.$$  

This tells us a very interesting relationship that given the measurable quantity, $\tilde{K}_n$, we can determine the two-step heat of reaction corresponding to the two-step efficiency, $E_n$, when the affinities, $A_{Ml}$ (or $A_{ml}$, $m \notin M_1$), corresponding to the selected nucleic acids for the neglected target components are very large.

Finally we measure the alternate selection efficiency of $R = R_s + R_n$ rounds, consisting of $R_s$ positive selection rounds followed by $R_n$ negative selection rounds, after having completed $kR$ rounds, as follows:

$$E^{(k)} = \left( \prod_{l=1}^{R_s} E_s^{(kr+l)} \prod_{l=R_s+1}^{R} E_n^{(kr+l)} \right)^{1/R} \quad (3.53)$$
where $k$ denotes the index that tells us which set of $R$ rounds are involved. More generally, in any process consisting of $P + Q$ rounds of which $P$ rounds are positive SELEX and $Q$ rounds are negative SELEX, we define the efficiency as

$$E^{(P+Q)} = \left( \prod_{l=p}^{P} E^{(p)}_{s} \prod_{p=1}^{Q} E^{(q)}_{n} \right)^{1/(P+Q)}.$$  

(3.54)

See Figure 3.7 for the examples of efficiencies.

### 3.7 Limiting behavior of alternate SELEX when the grand round number becomes large: Specificity

Here, we examine the effects of the multiplier, $m$, i.e., the limiting behavior of alternate selection when the grand round number, $R$, becomes large, for a fixed $\lambda$. When the positive and negative round numbers, $R_s$ and $R_n$, are replaced by $mR_s, mR_n$, where $m$ positive integer, the selection fraction, $\lambda$, does not change. For example, for a fixed $\lambda = mR_s/mR_n = 2/3$, say, there are many possibilities we can chose for the grand round numbers, $R = mR_s + mR_n$, such as $(mR_s, mR_n) = (2, 3) = (4, 6) = (10, 15)$. From the transition rules for limiting fractions combined with simulation results, when the grand round number is finite, the selected nucleic acid indices, as well as the final fractions do not depend on the order of positive selection and negative selection. See also Figure 3.5. However, simulation results indicate that for fixed $\lambda$, if we increase the grand round number (i.e., increase $m$), the limiting nucleic acid indices are more likely specified to the index set of nucleic acids that best bind to the target species absent during negative selection. For example, with $\lambda = 0.6$, from Figure 3.4, panel (b) and Figure 3.8, panels (b,d,f,h) we see that $L_R = \{8, 10, 12, 16\}$ for $m = 1, 2, 4, 8, 16$. On the other hand, notice that in Figure 3.4, panel (b) and Figure 3.8, panels (a,c,e,g), the set of indices $l$ for which the nucleic acid fractions $F_l > 0$ at the end of eighty rounds consists only of the single nucleic acid species, $l = 8$ for $m > 8$ say. This tells us that, while positive SELEX alone gives us “selectivity” to select the nucleic acid species that binds tightly to specific targets of interest, the alternate SELEX process provides a resulting pool of nucleic acid species with “specificity” to specify (or discriminate) the ones from the other species that bind preferentially.
to the negative targets.

We now consider the following thought experiment: Suppose the alternate selection process is performed in such a way that after a large number of positive selections (so that the system can nearly reach the limiting state or when the selection profile is stabilized), we do a few rounds of negative selection, without repeating both processes, and observe the selected nucleic acid indices. See Figure 3.9. After a large number of positive selection rounds the system is able to obtain the final nucleic acid fractions which correspond to the same indices that we can have when we only perform positive selection. Starting with the final NA fractions from the results of positive selection, several rounds of negative selection are performed. From Figure 3.9, we also see that by adding up the negative selection process (with one or more target components removed), the system reaches to the set consisting of nucleic acid species that best bind to the desired target species, which had been removed during the negative SELEX rounds.

Consider the following limits, for \( p = 1, 2, \ldots \),

\[
\lim_{m \to +\infty} \hat{F}_s^{(p,m)} = \hat{F}_s^{(p,\lambda)}, \quad \lim_{m \to +\infty} \hat{F}_n^{(q,m)} = \hat{F}_n^{(q,\lambda)}
\]

(3.55)

and assume they exist. If these vectors exist for each \( p, q \) respectively, then, we can also assume the limit

\[
\lim_{m \to +\infty} \hat{\omega}_s^{(p,m)} = \hat{\omega}_s^{(p,\lambda)}
\]

(3.56)

exists. Furthermore, we have

\[
\lim_{m \to +\infty} \frac{1}{m} \sum_{m' = 1}^{m} \hat{F}_s^{(p,m')} = \hat{F}_s^{(p,\lambda)}, \quad \lim_{m \to +\infty} \frac{1}{m} \sum_{m' = 1}^{m} \hat{F}_n^{(q,m')} = \hat{F}_n^{(q,\lambda)},
\]

\[
\lim_{m \to +\infty} \frac{1}{m} \sum_{m' = 1}^{m} \hat{\omega}_s^{(p,m')} = \hat{\omega}_s^{(p,\lambda)}, \quad \lim_{m \to +\infty} \frac{1}{m} \sum_{m' = 1}^{m} \hat{\omega}_n^{(p,m')} = \hat{\omega}_n^{(p,\lambda)}.
\]

(3.57)

Suppose that we perform alternate SELEX, for fixed \( \lambda \) by first doing \( mR_{s,0} \) rounds of positive selection followed by \( mR_{n,0} \) rounds of negative selection and the process is iterative. We can consider this situation as the oscillation between two different manifolds of the dynamical processes. Now at the end of \( mR_{s,0} \) rounds when \( m \) is large, there will be a subset of integers \( P_1 \subset N \) that can be obtained from positive SELEX only. Then \( \sum_{j \in P_1} F_j \) should be very close to unity. That is, we should be very near the limit for the pure positive selection
process before we begin the negative selection steps. If we then begin negative SELEX and perform \( m R_{n,0} \) rounds, we will find a subset \( P'_1 \subset N \) for which \( \sum_{j \in P'_1} F_j \) should be very close to unity. This does not tell us that \( P'_1 \) will not contain all of \( P_1 \) but will contain elements of the complementary set \( N - P_1 \equiv P'_1 \), the set of poorest binders to the positive targets. When we now start the positive SELEX again, we will find another subset \( P_2 \) of indices for which \( \sum_{j \in P_2} F_j \) should be very close to unity. If we then initiate negative SELEX and perform \( m R_{n,0} \) rounds, we will find a subset \( P'_2 \) for which \( \sum_{j \in P'_2} F_j \) should be very close to unity. After a finite number of grand rounds \( K \) say, it must be the case that \( P_K = P_{K+1} = P_{K+2} = \ldots \) and \( P'_K = P'_{K+1} = P'_{K+2} = \ldots \). Thus we expect,

**Conjecture 3.7.1.** For any fixed \( \lambda \), the set \( L_R = L_{m R_{0,\lambda}} \) does not depend on \( m \) and in fact \( L_R = P_K \cup P'_K \). We call this set \( L_\lambda \).

Note that when we define the maximal target affinity function using the specificity matrix, \( C_\lambda \), this matrix depends upon the selection fraction, \( \lambda \), but does not depend on the multiplier \( m \), while the maximal target affinity function depends on \( m \). In Figure 3.8, panels (b,d,f,h), with \( \lambda = 0.6 \), notice that \( L_\lambda = \{8, 10, 12, 16\} \) no matter what choice of \( m \) was made.

For large \( m \), we then should be able to compute the following limits for each process: pure positive and pure negative selection.

\[
\lim_{p \to +\infty} \hat{F}_s^{(p,\lambda)}, \quad \lim_{q \to +\infty} \hat{F}_n^{(q,\lambda)}, \quad \lim_{p \to +\infty} \hat{\omega}_s^{(p,\lambda)}, \quad \lim_{q \to +\infty} \hat{\omega}_n^{(q,\lambda)}
\]

Then we can make the following conjecture by considering the averages of the limiting values of parameters:

**Conjecture 3.7.2.** If we begin with positive or negative SELEX and let \( m \to +\infty \) for fixed
Conjecture 3.7.3. We always have $\mathcal{L}_{\lambda,s} \subset \mathcal{L}_{\lambda}$. 

Given the conjecture 3.7.2 we can conclude the following as well:

$$
\lim_{m \to +\infty} \left( \prod_{m'=1}^{m} \prod_{p=1}^{m'} \frac{C_{\lambda}^{l} \cdot \hat{\omega}_{s}^{(p,m')}}{C_{\lambda}^{j} \cdot \hat{\omega}_{s}^{(p,m')}} \right)^{2/m(m+1)R_{s,0}} = \frac{C_{\lambda}^{l} \cdot \hat{\omega}_{s}^{\lambda}}{C_{\lambda}^{j} \cdot \hat{\omega}_{s}^{\lambda}},
$$

(3.59)

This then leads us to define the target specificity functions and the maximal target specificity function as follows: First, let $\vec{\Psi}(\lambda, m)$ denote the vector consisting of all the free target vectors $\vec{\omega}_{s}^{(p,m')}$ as $p, m'$ vary over the positive integers such that $1 \leq p \leq m' \leq m$. Then define the target specificity functions and the maximal target specificity function by

$$
\psi_{j}(\lambda, m, \vec{\Psi}(\lambda, m)) \equiv \left( \prod_{m'=1}^{m} \prod_{p=1}^{m'} \frac{C_{\lambda}^{l} \cdot \hat{\omega}_{s}^{(p,m')}}{C_{\lambda}^{j} \cdot \hat{\omega}_{s}^{(p,m')}} \right)^{2/m(m+1)R_{s,0}} \text{ for } j \in \mathcal{N}
$$

and

$$
\psi_{\lambda,m}(\lambda, m, \vec{\Psi}(\lambda, m)) \equiv \max\{ \psi_{j}(\vec{\Psi}(\lambda, m)), \ j \in \mathcal{N}\}.
$$

(3.60)

Now define the limiting target specificity functions and the limiting maximal target specificity function for $\lambda \in [0, 1]$ and $\hat{\omega} \in \mathcal{S}_{M}$ by

$$
\psi_{j}(\lambda, \hat{\omega}) \equiv \frac{C_{\lambda}^{j}}{C_{\lambda}^{l}} \cdot \hat{\omega} \text{ for } j \in \mathcal{N},
$$

(3.61)

$$
\psi_{\lambda}(\hat{\omega}) \equiv \max\{ \psi_{j}(\lambda, \hat{\omega}) \text{ for } j \in \mathcal{N}\}
$$

respectively. We define, for the limiting vectors, $\hat{\omega}_{s}^{\lambda}$, the set

$$
\mathcal{L}_{\lambda,s} = \{ l \in \mathcal{N} | \psi_{l}(\hat{\omega}_{s}^{\lambda}) = \psi_{l}(\lambda, \hat{\omega}_{s}^{\lambda}) \}.
$$

(3.62)
The set \( L_{\lambda,s} \) as well as the set of indices that maximize \( \psi_j(\lambda, \hat{\omega}^1) \) can help us to determine the set of nucleic acid indices that best binds to targets absent during the negative selection process. Here, \( \hat{\omega}^1 \) is the limiting free target vector when we only perform positive SELEX. See Figures 3.10 and 3.11 for this aspect of simulation results.

### 3.8 The set of stationary solutions, specificity of the dynamical system on the chemical potential energy surfaces

In this section, we study the relationship between the limiting target fractions and the chemical potential. When we only perform positive selection, we define the following function of the individual efficiencies at infinite target dilution in Chapter 2:

\[
R(\vec{E}) = -RT \sum_{i=1}^{M} \Omega_i \ln[1/(1 - E_i)],
\]

where \( R \) is the gas constant and \( T \) is the Kelvin temperature where the \( E_i \) are the limiting efficiencies of each target component.

To understand the behavior of the dynamical system of alternate SELEX in terms of thermodynamics as discussed in Chapter 2, we first take a set of limiting selected indices, \( L = L_\lambda \) and define the simplex \( S_{L_\lambda,\vec{F}} = \{ \vec{F} \in R^N | 0 \leq F_i \leq 1, F_i = 0 \text{ if } i \notin L_\lambda, \sum F_i = 1 \} \), in Euclidian \( N \) space \( R^N \). Then, based on the observations made in Section 3.6, define two chemical potentials as follows:

\[
R_s(\vec{F}, \lambda, \hat{\Omega}_s) = -RT \sum_{i=1}^{M} \Omega_{s,i} \ln(1 + [NA]A_i \cdot \vec{F}),
\]

\[
R_n(\vec{F}, \lambda, \hat{\Omega}_n) = RT \sum_{i \in M_t} \Omega_{n,i} \ln(1 + [NA]B_i \cdot \vec{F}),
\]

the first for the limiting positive SELEX NA fractions and the second for the limiting negative SELEX NA fractions. We define the chemical potential for negative selection to be positive, which is the opposite sign from the chemical potential defined for positive selection, because we are interested in maximizing the unbound species in negative selection.

The functions in (3.64) (or their restrictions on the sub simplex) also define chemical potential surfaces, \( P_s, P_n \) over \( S_{L_\lambda,\vec{F}} \). The surface \( P_s \) forms a convex graph over this simplex (as we have seen in Chapter 2), while the graph of the surface \( P_n \) is concave.
For each fixed \( m = 1, 2, 3, \ldots \), our interest is in the arrangement of the point set

\[
P_s(m) = \{(F_s^{(p,m)}, R_s(F_s^{(p,m)}, \lambda)) \mid p = 1, \ldots, R_s\}
\]

on the former surface and arrangement of the point set by

\[
P_n(m) = \{(F_n^{(q,m)}, R_n(F_n^{(q,m)}, \lambda)) \mid q = 1, \ldots, R_n\}
\]

on the latter surface. If for each fixed \( p, q \) (each of which is ultimately smaller than \( mR_s, mR_n \) respectively), the limits

\[
\lim_{m \to +\infty} F_s^{(p,m)} = F_s^{(p,\lambda)}, \quad \lim_{m \to +\infty} F_n^{(q,m)} = F_n^{(q,\lambda)}
\]

exist, these point sets should approach point sets defined as follows:

\[
P_s(\lambda) = \{(F_s^{(p,\lambda)}, R_s(F_s^{(p,\lambda)}, \lambda)) \mid p = 1, \ldots, \infty\} \tag{3.65}
\]

for the former and the point set

\[
P_n(\lambda) = \{(F_n^{(q,\lambda)}, R_n(F_n^{(q,\lambda)}, \lambda)) \mid q = 1, \ldots, \infty\} \tag{3.66}
\]

for the latter. More precisely, for example, we should expect that

\[
\lim_{m \to +\infty} \sum_{p=1}^{mR_s,0} (|F_s^{(p,m)} - F_s^{(p,\lambda)}| + |R_s(F_s^{(p,m)}, \lambda) - R_s(F_s^{(p,\lambda)}, \lambda)|) = 0.
\]

In turn, each of the point sets in (3.65), (3.66) should constitute a set of points which lie on a piecewise smooth curve on the corresponding chemical potential surface. Call these curves \( C_s(\lambda) \), \( C_n(\lambda) \), respectively.

We illustrate these concepts and the point sets graphically in Figures 3.12 and 3.13 for a nucleic acid pool with seven nucleic acids and a target with three components, one of which was absent during negative selection. In Figure 3.14, we plot the limiting point sets defined above along with steepest descent curves over each of the chemical potential surfaces for each fixed \( m = 10, 40, 80, 160 \). We interpret this result that the set of points on the chemical potential surface defined by the limiting states of the positive SELEX process appears to lie along a smooth curve of steepest descent constrained by the negative SELEX process and vice versa.
3.9 Simulation results for alternate SELEX

In Figures 3.2-3.11 we used the same 5 by 20 affinity matrix \( A \) used in the multiple target case (see Appendix C), and the initial nucleic acid fractions are randomly generated for each experiment unless otherwise stated. The affinity matrix (3 by 7) used for Figures 3.12-3.14 is also provided in Appendix C. Based on the computer simulation limitations on approximating the limiting values, especially when the nucleic acid fractions are very small, say less than \( 10^{-17} \), it would be reasonable for us to set a population of zero molecules of the species in this case. This still provides us a good agreement with our theoretical results for the limiting parameter values.

3.9.1 Limiting NA fractions, efficiencies and target affinity function values using the SELEX iteration scheme

1. In Figure 3.2 and Figure 3.4(d), we first observed the optimal strategy for choosing the target concentration from round to round during the negative selection steps. Here we only performed negative SELEX and computed NA fractions using an iteration scheme given in (3.20). The results show for negative selection, that it is better to fix the total target concentration from round to round rather than reduce it as one does in iterative positive selection. In each panel in Figure 3.2 and Figure 3.4, panel (d), we completed four different experiments: negative selection with (a) a target held fixed, \([T_n]^{(r)} = 1 \mu M\), from round to round, (b) decreasing target from round to round, like \([T_n]^{(r)} = \frac{1}{r}[T_n]^{(1)}\) starting with \([T_n]^{(1)} = 1 \mu M\), (c) target reduction \([T_n]^{(r)} = T_n^{(1)} \frac{1}{r}[T_n]^{(1)}\) with \([T_n]^{(1)} = 100 \mu M\), from round to round, and (d) in Figure 3.3, fixed target \([T_n]^{(r)} = 100 \mu M\) at every round. Here we fixed the nucleic acid pool size, \([NA] = 1 \mu M\) from round to round (PCR) for all four experiments. In panels (a-c) of Figure 3.3 and panel (d) in Figure 3.7, we calculated the negative selection efficiencies, \(E_n^{(r)}\), using the definition in (3.40). The comparison of Figure 3.3 panel (a) with Figure 3.7 panel (d) suggests that it is more efficient to fix \([NA]/[T_n]\) small, for fixed target, from round to round during negative selection to obtain the poorest binder, \(NA_{13}\) rather than to fix \([NA]/[T_n] = 1\).
2. In Figure 3.4, we illustrated four experiments of alternate selection procedures by taking the selection fractions, $\lambda = 0, 0.4, 0.6, 1$. Theoretical PCR amplification was performed to keep the pool size of $[NA] = 1\mu M$ from round to round. The initial target vector for the positive selection steps was taken from Figure 2.1 panel (a), $\hat{\Omega}_s = \langle \Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5 \rangle = \langle 0.1374, 0.1346, 0.4090, 0.1844, 0.1346 \rangle$, and the total positive target concentration $[T_s]$ was reduced from unity from round to round. For the negative selection steps, the negative target vector $\hat{\Omega}_n = \langle \Omega_1, \Omega_2, \Omega_3, \Omega_4, 0 \rangle / (\sum_{i=1}^{4} \Omega_i) = \langle 0.1588, 0.1555, 0.4726, 0.2131, 0 \rangle$ was used, and the total target concentration was fixed as $[T_n] = 100\mu M$ at each round, i.e., $\overrightarrow{[T_n]} = \langle 15.88, 15.55, 47.26, 21.31, 0 \rangle$. In panels (a) and (d) we plotted NA fractions using the iteration scheme for positive selection only ($\lambda = 1$) and negative selection only ($\lambda = 0$), respectively. In panel (a) the limiting NA fraction vector, after 50 rounds (also shown in 2.1 panel (a)), $\hat{F} = \langle F_8, F_9, F_{10}, F_{12}, F_{16} \rangle = \langle 0.1956, 0.2794, 0.0498, 0.3843, 0.0908 \rangle$. In panel (d), $F_{13} = 1$.

For panels (b) and (c) we alternated positive selection with negative selection, beginning with positive selection. In panel (b), for $\lambda = 0.6$, we started with three rounds of positive selection followed by two rounds of negative selection and repeated this sequence 10 times. After 150 rounds (first 50 rounds are shown in panel (b)), we have selection with the set of nucleic acid indices $\{8, 10, 12, 16\}$, and the limiting NA fraction vector $\hat{F} = \langle F_8, F_{10}, F_{12}, F_{16} \rangle = \langle 0.3754, 0.1496, 0.2581, 0.2169 \rangle$ at the end of negative selection. (The panel (b) illustrates the effect that by removing the bound products (with the negative (sub)target) during negative SELEX, a pool of nucleic acids is enriched with better binders to the target component deleted.) Here we notice that 8th nucleic acid binds best to the 5th target component, which was absent during negative SELEX, and the fraction $F_8$ has increased in passing from panel (a) to (b), while $F_9$ has disappeared. In panel (c), for $\lambda = 0.4$, after several iterations of two positive rounds followed by three negative rounds, the selected NA indices are $\{17, 20\}$, and the limiting NA fraction $\hat{F} = \langle F_{17}, F_{20} \rangle = \langle 0.7915, 0.2085 \rangle$ after 150 rounds (first 50 rounds shown in panel (c)).

3. In Figure 3.5, we performed alternate selection in the same manner as in Figure 3.4,
panels (b) and (c), except that we started with the negative selection steps followed by the positive selection steps. Here, after 50 rounds, we proceeded for several more rounds of negative selection (52 rounds total for panel (a) and 53 rounds total for panel (b)) to compare the final NA fractions at the end of negative selection step (or the first positive selection step) with panels (b) and (c) in Figure 3.4. Although the order does affect the iterative outcome during the first several rounds of alternate selection, the limiting NA fractions of panels (a,b) agree with the above limiting values for Figure 3.4 (b,c), respectively.

4. In Figure 3.6, individual target affinities for 20 nucleic acids in Figure 3.4 panels (b) and (c) were computed using equation (3.33) after 50 rounds. In panels (a,b), the maximum indicates the selected indices \{8, 10, 12, 16\} for \(\lambda = 0.6\), and \{17, 20\} for \(\lambda = 0.4\), respectively. In panel (a) the limiting \(K_a = 0.6964M^{-1} = \varphi_R(0.6, \hat{\omega}_s^{(1)}, \hat{\omega}_s^{(2)}, \hat{\omega}_s^{(3)})\) where the limiting free target vectors are \(\hat{\omega}_s^{(1)} = \langle 0.1218, 0.0791, 0.5145, 0.2158, 0.0688 \rangle\), \(\hat{\omega}_s^{(2)} = \langle 0.1449, 0.1117, 0.4058, 0.2425, 0.0951 \rangle\) and \(\hat{\omega}_s^{(3)} = \langle 0.1574, 0.1307, 0.3515, 0.2496, 0.1108 \rangle\). In panel (b) the limiting \(K_a = 0.4113(10^{-4})M^{-1} = \varphi_R(0.4, \hat{\omega}_s^{(1)}, \hat{\omega}_s^{(2)})\) where the limiting free target vectors are \(\hat{\omega}_s^{(1)} = \langle 0.1525, 0.1811, 0.4208, 0.2083, 0.0373 \rangle\) and \(\hat{\omega}_s^{(2)} = \langle 0.1545, 0.1810, 0.4199, 0.2075, 0.0371 \rangle\).

5. In Figure 3.7, we plotted target efficiencies as a function of the round number for corresponding four panels in Figure 3.4. We used formulas (3.37) for panel (a), (3.40) for panel (d) and (3.54) for panels (b,c).

### 3.9.2 Specificity, the limiting target specificity function values and the set of indices for specificity

1. Figure 3.8 demonstrates the effect of the multiplier, \(m\), i.e., the effect of increasing the grand round number. The selection fraction used here is \(\lambda = 0.6\). Panels (a,c,e,g) illustrate the nucleic acid fractions for 80 rounds, for \(m=2, 4, 8, 16\), respectively. In panel (d), for example, alternate selection with 48 rounds of positive selection followed by 32 rounds of negative selection was performed, \(m = 16\). Panels (b,d,f,h) describe the
corresponding target affinity function values for each nucleic acid after 80 rounds. \( K_a \) has units of \((\mu M)^{-1}\). As \( m \) increases, the limiting set is more specified to the \( 8^{th} \) NA species that best binds to the target absent during negative selection. Notice that for each of the four panels, the maximum of the bar graphs occurs at the indices \{8, 10, 12, 16\}.

2. In Figure 3.9, we performed alternate SELEX for \( mR_s \gg 1 \). In this way we have a selection at the end of positive SELEX. By performing negative SELEX with the final NA fractions (that result from positive selection only), we reached the set of indices that correspond to the NA species that specifically bind to targets absent during negative selection.

3. In Figure 3.10, we present the limiting target specificity function values, and this result indicates that for \( \lambda > 0.5 \) we have a good approximation for the set for specificity.

4. In Figure 3.11, different limiting sets of NA indices were computed for the use of approximation with specificity. This also suggests that for \( \lambda > 0.5 \), we have good approximation for estimating the sets of “specific” indices.

3.9.3 Chemical potential surfaces and the relationship between the limiting chemical potential curves and the steepest descent curves

1. In Figure 3.12, the limiting chemical potentials for the three targets case are shown for positive and negative selection.

2. In Figure 3.13, the “hysteresis” loops for the cases, \( m = 10 \) and \( m = 160 \) (= “\( \infty \)”), are shown in panels (a) and (b). If one begins on the lower \((P_s)\) surface in panel (a) at the point furthest from the surface minimum, after 90 positive selection rounds, the last round will correspond to the point in \( P_s(10) \) closest to the minimum of this surface. The transition to the upper surface takes one to a point on the upper \((P_n)\) surface. If one completes 10 negative selection rounds, the last round will correspond to the point in \( P_n(10) \) closest to the minimum of this surface. The orientation on each point set is always toward the minimum point on the surface. Panel (b) shows the form of the limiting loop.
(a) $[T_n] = 1 \mu M$ is fixed for each round.

(b) Initially $[T_n] = 1 \mu M$ and decreases like $1/r$.

(c) Initially $[T_n] = 100 \mu M$ and decreases like $100/r$.

Figure 3.2 Each panel illustrates the distribution of nucleic acid fractions for negative SELEX only with different choices of target concentrations. $[NA]$ is fixed as $1 \mu M$ for all three cases.

as approximated by the case when $m = 160$. In both panels, the loop is traversed in such a way that the direction of travel is toward the minimum on the surface.

3. In Figure 3.14, the plots of the point sets $P_s(m)$, $P_n(m)$ and $P_s(\lambda)$, $P_n(\lambda)$ along with the steepest decent curves over two chemical potential surfaces.
(a) $[T_n] = 1 \mu M$ is fixed for each round.

(b) Initially $[T_n] = 1 \mu M$ decreases like $1/r$.

(c) Initially $[T_n] = 100 \mu M$ and decreases like $100/r$.

Figure 3.3 Corresponding efficiencies to the cases in Figure 3.2 are described. $[NA] = 1 \mu M$. 
(a) Positive SELEX (no negative rounds). $\lambda = 1.0$. Selected indices are $\{8, 9, 10, 12, 16\}$.

(b) Three positive rounds (dark red) followed by two negative rounds (yellow). $\lambda = 0.6$. Selected indices are $\{8, 10, 12, 16\}$.

(c) Two positive rounds (dark red) followed by three negative rounds (yellow). $\lambda = 0.4$. Selected indices are $\{17, 20\}$.

(d) Negative (Subtractive) SELEX (no positive rounds). $\lambda = 0.0$. Selected index is $\{13\}$.

Figure 3.4 Comparison of positive SELEX with alternate SELEX was made. During negative selection the fifth target component was absent. In panels (b) and (c), the grand round number $R = 50$ was taken. By comparing the panel (a) with (b), we notice that the 9th NA species has disappeared, while the fraction of 8th NA species has increased. From panel (d) we see that 13th NA is the poorest binder to the target with the fifth component deleted. In panels (b-d), $[T_n] = 100\mu M$. 
Alternate SELEX beginning with negative selection was performed to check the dependency on the order of positive and negative selections. In panel (a), alternate SELEX with two negative rounds (yellow) followed by three positive rounds (dark red) was performed. $\lambda = 0.6$. Compare with Figure 3.4, panel (b). In panel (b), alternate SELEX with three negative rounds (yellow) followed by two positive rounds (dark red) was performed. $\lambda = 0.4$. Compare with Figure 3.4, panel (c).

**Figure 3.5**

Target affinity function values, $\lambda = 0.6$, $m = 1$.

(a) Negative SELEX first.

(b) Negative SELEX first.

**Figure 3.6**

Target affinity function values, (equation (3.33) after 50 rounds) are shown. In panel (a), selected NA indices are $\{8, 10, 12, 16\}$. See Figure 3.4, panel (b). In panel (b), selected NA indices are $\{17, 20\}$. See Figure 3.4, panel (c). They have significant differences in magnitude between the two values in panel (a) and (b). The values of the target affinity functions that correspond to the selected NA indices are (a) $\langle \varphi_8, \varphi_{10}, \varphi_{12}, \varphi_{16} \rangle = \langle 0.6955, 0.6962, 0.6952, 0.6959 \rangle$ and (b) $\langle \varphi_{17}, \varphi_{20} \rangle = \langle 0.4113(10^{-4}), 0.4112(10^{-4}) \rangle$. 

(a) Target affinity function values, $\lambda = 0.6$, $m = 1$.

(b) Target affinity function values, $\lambda = 0.4$, $m = 1$. 
Figure 3.7  Efficiencies for alternate SELEX are illustrated. These are the efficiency plots for the corresponding panels illustrated in Figure 3.4. In panels (b,c) the curves designated by ‘⋄’ denote the cumulative efficiencies using equation (3.54). The insets in these panels indicate the asymptotic behavior of these efficiencies.
Figure 3.8 Effect of increasing the multiplier $m$ on specificity is described.
(a) First target component was absent during negative selection

(b) Third target component was absent during negative selection

(c) Fourth target component was absent during negative selection

(d) Fifth target component was absent during negative selection

Figure 3.9 Alternate SELEX for a large number of positive selection rounds ($mR_s >> 1$) followed by a few rounds of negative selection was performed to examine the specificity of the process. Five target components case is taken from multiple-target positive SELEX in Figure 2.1 panel (a). Each plot shows nucleic acid fractions for corresponding indices \{8, 9, 10, 12, 16\} performed during negative selection, as a function of round number. The starting NA fractions for the first round above are the final NA fractions at selection when only positive selection was performed, and they are updated during the negative selection process. In each case, the number of selected nucleic acid species are decreased to the single NA species that binds to the desired target (absent during negative selection) with high affinity and specificity.
Figure 3.10 Limiting target specificity function values are presented here. In panel (a) we calculated $\psi_j(\lambda, \hat{\omega}_s^\lambda)$ as a function of $j = 1, \ldots, 20$, for $\lambda = 0.6$. Notice that the maximum occurs precisely for $j = 8$. In panel (b) we calculated $\psi_j(\lambda, \hat{\omega}_s^1)$ for $\lambda = 0.6$. While they are not identical, the two panels indicate that at least for $\lambda > 1/2$, we may use $\hat{\omega}_s^1$ as a reasonable approximation for $\hat{\omega}_s^\lambda$.

Figure 3.11 Selected indices are shown from various approximation approaches. In panel (a), for each $\lambda \in [0, 1]$, the members of $\mathcal{L}_\lambda$ are indicated in heavy dots. For example, when $\lambda = 0.5$, $\mathcal{L}_\lambda = \{8, 17\}$. In panel (b), nucleic acid indices that maximize the target specificity function values, i.e., that maximize $\psi_j(\lambda, \hat{\omega}_s^\lambda)$, $j \in \mathcal{N}$ from equations (3.61) and (3.62) are plotted. In panel (c), approximation to panel (b) was made from the indices that maximize $\psi_j(\lambda, \hat{\omega}_s^1)$, $j \in \mathcal{N}$. The advantage is for this approximation that it only requires the information of the limiting free target vector obtained from positive SELEX only ($\lambda = 1.0$).
Figure 3.12 Limiting chemical potentials for the three target case are shown. The pool consisted of seven nucleic acid species. The set $L_{\lambda, \nu} = 1, 4, 6$. Here, $F_0 = 1 - F_1 - F_4$. The limiting curve in panel (a) (roughly the case when $m = 160$), is orthogonal to the level sets of $R_s$ when these are referred its graph over the simplex $S_{L_{\lambda, \nu}}$. However, even when referred to the projection of this simplex in the plane, this curve is very nearly orthogonal to the level sets of $R_s$. 
Figure 3.13  “Hysteresis” loop for the limiting nucleic acid fraction vectors $\hat{F}_s, \hat{F}_n$ is shown for the case $m = 10$ and $m = 160$ (= $\infty$) in panels (a,b), respectively.

Figure 3.14  Plots of steepest descent curves and the point sets $P_s(m)$ and $P_n(\lambda)$. Positive selection. (b) Plots of steepest descent curves (red) from several points in the point sets, $P_n(m)$ and $P_n(\lambda)$. Negative selection.
CHAPTER 4. Summary and future work

4.1 Results

4.1.1 Multiple-target (positive) SELEX

In Chapter 2, we present a SELEX iteration scheme, considered as a discrete time dynamical system, for multiple-target (positive) SELEX. The SELEX process always converges to some limiting vector of nucleic acid fractions and free target fractions, when the concentration of the total target pool is varied by dilution. We provide a geometric condition on the affinity matrix so that, independently of the composition of the initial pool of nucleic acids present, the process determines a unique subset of indices corresponding to the nucleic acids that bind tightly to targets. The geometric condition allows the SELEX system to have a proper maximal target affinity function (defined on the simplex of possible free target vectors $\hat{\omega}$), providing a unique subset of nucleic acid species that bind to the targets with high affinity. When this is the case, the SELEX process is globally asymptotically stable. This, in turn, is equivalent to the geometric condition (that defines a “proper” face) that the chemical potential has a unique minimum point at infinite target dilution.

4.1.2 Alternate SELEX: Positive selection followed by negative selection

In Chapter 3, we present a SELEX iteration scheme, considered as a discrete time dynamical system, for alternate SELEX. The process consists of two processes: positive selection and negative selection. By alternating these two processes, we obtain a refined pool of nucleic acids that not only bind tightly to the positive targets during positive SELEX (“selectivity”), but also bind specifically to the target component absent during negative SELEX (“specificity”).
We construct two chemical potential energy surfaces at limiting states and describe a transition state of free energies. The set of points on the chemical potential surface defined by the limiting states of positive SELEX lies along a smooth curve whose position on the surface is constrained by negative SELEX and vice versa. We also examine the effect of the multiplier, $m$, when the selection fraction, $\lambda$, is fixed. By completing this, the alternate SELEX process provides a means for obtaining nucleic acids with high selectivity and specificity for the desired target.

### 4.2 Future work

There are a number of open questions that have arisen as a consequence of my research, regarding the SELEX iteration scheme for multiple targets. These are interesting and important questions that will be pursued in the future.

1. In biological experiments, there are experimental losses during the process. When the nucleic acids are passed through a support material, $S$ (e.g., the nitrocellulose filter), some fractions, $b_{ij}$, of free nucleic acid species can be captured by this support material, considered as non-specific binding. On the other hand, when the binding NA species are separated (or partitioned) from the target-NA complex, some fractions, $a_{ij}$, of the complex can be lost. This aspect of selection behavior was analyzed for the single-target case in [11] and [15], and for the multiple-target case in [30] extended from the single target case in [11].

   In the multiple target problems in Chapter 2, the term $[T_i : NA_j]$ is then replaced by $a_{ij}[T_i : NA_j] + b_{ij}[NA_f_j]$, where $0 < a_{ij}, b_{ij} < 1$ and $[NA_f_j] = [NA_j] - \{S : NA_j\} - \{T : NA\}$. The correction factors can usually be taken independently of the particular nucleic acid and target component, but this need not always be the case. This was achieved in [8], p. 753. The effect of these parameters is to slow the rate of convergence of the iteration scheme. See [11, 30] for definitions and [15] for a detailed analysis and simulations illustrating their effect on the rate of convergence. In [11] it was also mentioned that with effective partitioning techniques to avoid non-specific binding against the background (or through the support), $b_{ij}$ should converge to zero and $a_{ij}$ should converge to unity.

2. In [15], a continuous analog of the single-target SELEX iteration scheme was presented...
by considering the round number, \( r \), as a continuous (time) parameter value and the “ratio” \( (F_j^{(r+1)} - F_j^{(r)})/1 \) as a difference quotient approximation to \( dF_j/dr \), using ordinary differential equations. There the ratio \( \frac{dF_j}{dr} = (E_j(r) - 1)F_j(r) \), where \( F_j^{(r)} \) was replaced by the continuous notation \( F_j(r) \) and \( E_j(r) = \frac{F_j(r + 1)}{F_j(r)} \). It was also assumed that the dissociation constant for each of the \( N \) nucleic acids is ordered as \( 0 < K_{d1} < \cdots < K_{dN} \). Moreover, \( \frac{dT}{dr} = -s(r)[T](r) \), where \( [T](r) \) and \( s_r \) were replaced by \( [T](r) \) and \( s(r) \in [0, 1) \). Then \( \frac{F_j(r)}{F_1(r)} = \frac{F_j(1)}{F_1(1)} \exp \left( - \int_1^r (E_1(s) - E_j(s)) \, ds \right) \) and \( [T](r) = [T](1) \exp \left( - \int_1^r s(\rho) \, d\rho \right) \). It was shown that \( F_j(r) \to 0 \) for \( j \geq 2 \) as \( r \to +\infty \) (i.e., the pool of nucleic acids converges to a pool consisting only of the best binding nucleic acid) if and only if \( \int_1^\infty \exp \left( - \int_1^r s(\rho) \, d\rho \right) \, dr = +\infty \). This can be achieved for a single target case with the choice of \( s(r) = 1/r \) (this, in turn, is the case that \( \int_1^\infty s(\rho) \, d\rho = +\infty \)). Moreover, maximum bound target efficiency \( \frac{[NA]}{K_{d1} + [NA]} \) can be obtained at the limiting state by the choice of infinite target dilution. A continuous analog of the dynamical system defined by multiple-target SELEX can be considered for the case of infinite target dilution.

3. The strong ergodic property of the alternate SELEX problem for the “improper” case should also be considered. In the multiple-target (positive) SELEX, the dynamical system that defines a “proper” maximal target affinity function has a strong ergodic property. That is, when it is allowed to run for a long time, the behavior of the dynamical system defined by the proper multiple-target SELEX process, statistically speaking, “forgets” its initial nucleic acid state. Then, one may ask the following question: Does alternate SELEX change the status of improperness to properness when alternating the two processes? For example, suppose we perform the positive SELEX process whose maximal target affinity function is improper (the distribution of the final NA fractions depends upon the initial pool of nucleic acids). Then, we perform negative SELEX with a pool of NA resulting from positive SELEX. Can the maximal target affinity function defined by the alternate SELEX process be considered proper? That is, does the alternate SELEX process always provide a same “specified” set of nucleic acid species (that best bind to the desired target) without depending upon the distribution of nucleic acids at the beginning of negative SELEX? Several simulation results show (not shown in this thesis)
the alternate SELEX process is strongly ergodic, when the grand round number is large and when we alternate positive SELEX followed by negative SELEX once. This ergodic property for alternate SELEX must be proven. This also would provide insight into the system’s (global) asymptotic stability for the alternate SELEX process.

4. In alternate SELEX, before the targets are subjected to the negative selection procedure we removed one target component from the multiple targets. Then, one can ask what happens when we remove more than one target component. The target component absent during negative selection can be considered as the desired target. So far, we have seen, for large $m$, the alternate SELEX process can select for a single NA species that best binds to the single desired target. Then, this can be considered as a single-target case for finding best binding NA species to the single (desired) target (absent during negative selection). With the several desired target components, we should be able to apply the idea from the multiple-target case in Chapter 2. From simulation results (not shown in this thesis), when we remove two target components, for large $m$ at the end of negative SELEX, only one or two NA species remain that specifically bind to the targets absent during negative selection. Perhaps this results from the relationship with the overall (not just single binding affinity to each target) association equilibrium constant against multiple targets deleted.

5. We iterate the alternate SELEX process by performing $R_s$ rounds of positive SELEX followed by $R_n$ rounds of negative SELEX. However, the selection fraction, $\lambda$, does not need to be the same from grand round to grand round. It can be a function of one of the experimental parameters or can be changed in a systematic manner. An optimal choice for the number of positive and negative selections can also be determined to achieve specificity. This can be a challenging problem, but it can be a possible question asked by experimentalists.

6. Further work needs to be done to make precise mathematical arguments for the relationship between the free energy curves and the steepest descent curves, as well as for the convergence of the alternate SELEX iteration scheme. The conjectures presented in Chapter 3 also must be proved.
CHAPTER 5. Appendix

5.1 Appendix A. Notation

<table>
<thead>
<tr>
<th>Species</th>
<th>Concentration or fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>target, $T$</td>
<td>$[T]$</td>
</tr>
<tr>
<td>$i^{th}$ target component, $T_i$</td>
<td>$[T_i]$</td>
</tr>
<tr>
<td>$i^{th}$ target fraction, $[T_i]/[T]$</td>
<td>$\Omega_i$</td>
</tr>
<tr>
<td>free target, $Tf$</td>
<td>$[Tf]$</td>
</tr>
<tr>
<td>$i^{th}$ free target, $[Tf_i]$</td>
<td>$[Tf_i]$</td>
</tr>
<tr>
<td>$i^{th}$ free target fraction, $[Tf_i]/[Tf]$</td>
<td>$\omega_i$</td>
</tr>
<tr>
<td>nucleic acid, $NA$</td>
<td>$[NA]$</td>
</tr>
<tr>
<td>$j^{th}$ nucleic acid, $NA_j$</td>
<td>$[NA_j]$</td>
</tr>
<tr>
<td>$j^{th}$ nucleic acid fraction, $[NA_j]/[NA]$</td>
<td>$F_j$</td>
</tr>
<tr>
<td>free nucleic acid, $NAf$</td>
<td>$[NAf]$</td>
</tr>
<tr>
<td>free $j^{th}$ nucleic acid, $NAf_j$</td>
<td>$[NAf_j]$</td>
</tr>
<tr>
<td>bound $j^{th}$ nucleic acid, ${T : NA_j}$</td>
<td>$[{T : NA_j}]$</td>
</tr>
<tr>
<td>bound $j^{th}$ nucleic acid with $i^{th}$ target, ${T_i : NA_j}$</td>
<td>$[{T_i : NA_j}]$</td>
</tr>
<tr>
<td>bound nucleic acid ${T : NA}$</td>
<td>$[{T : NA}]$</td>
</tr>
</tbody>
</table>

5.2 Appendix B. Convergence of nucleic acid fractions

For every pair of indices $j, l$, $j \neq l$ we have

$$\frac{F_j^{(r+1)}}{F_l^{(r+1)}} = \left( \prod_{p=1}^{r} \frac{D_{j,p}^{(p)} (1 + D_{l,f}^{(p)})}{D_{l,f}^{(p)} (1 + D_{j,f}^{(p)})} \right) \frac{F_j^{(1)}}{F_l^{(1)}}. \quad (5.1)$$

The partial products converge to zero (the infinite product is then said to diverge to zero) if and only if

$$\sum_{p=1}^{\infty} \frac{|D_{l,f}^{(p)} - D_{j,f}^{(p)}|}{D_{l,f}^{(p)} (1 + D_{j,f}^{(p)})} = \sum_{p=1}^{\infty} \frac{|1 - D_{j,f}^{(p)}/D_{l,f}^{(p)}|}{(1 + D_{j,f}^{(p)})} \quad (5.2)$$
is divergent. Examining the tail-end of this series and noting that $D_{j,l}^{(p)} \to 0$ uniformly in $j$, we see that the divergence of the product is equivalent to the divergence of the series:

$$\sum_{p=1}^{\infty} |1 - D_{j,l}^{(p)} / D_{l,l}^{(p)}|.$$  \hspace{1cm} (5.3)

To establish that all of the sequences $\{F_{j}^{(r)}\}_{r=1}^{\infty}$ converge, it suffices to establish that one sequence does.

To see the latter statement, fix some arbitrary index $l \in \mathcal{N}$. If for all $j \neq l$, the series in (5.3) diverged, then the sequences $\{F_{j}^{(r)}\}_{r=1}^{\infty}$ would all converge to zero. But then it follows from the normalization condition that the sequence $\{F_{l}^{(r)}\}_{r=1}^{\infty}$ converges to unity.

Suppose instead that for this $l$ the set of all indices $\mathcal{L}_l$ for which this series in (5.3) is convergent, is not empty. We claim that $\lim_{r \to \infty} F_{l}^{(r)} > 0$. We write

$$1 = \sum_{j=1}^{N} F_{j}^{(r+1)} = F_{l}^{(r+1)} + \sum_{j \in \mathcal{L}_l} F_{j}^{(r+1)} + \sum_{j \notin \mathcal{L}_l} F_{j}^{(r+1)}.$$  \hspace{1cm} (5.4)

The second sum on the right must converge to zero as $r \to +\infty$. On the other hand, we write $F_{j}^{(r+1)} = c^{(r)}(j,l) F_{l}^{(r+1)}$ if $j \in \mathcal{L}_l$ where $c^{(r)}(j,l)$ denotes the right-hand side of (5.1). By hypothesis, the sequences $\{c^{(r)}(j,l)\}_{r=1}^{\infty}$ converge to nonzero limits. Therefore,

$$F_{l}^{(r+1)} = \frac{1 - \sum_{j \notin \mathcal{L}_l} F_{j}^{(r+1)}}{1 + \sum_{j \in \mathcal{L}_l} c^{(r)}(j,l)}$$

and hence $\{F_{l}^{(r)}\}_{r=1}^{\infty}$ converges to a nonzero value. Consequently, so do the sequences $\{F_{j}^{(r)}\}_{r=1}^{\infty}$ for $j \in \mathcal{L}_l$. This completes the convergence proof.

### Table 5.2 Vector Notation

<table>
<thead>
<tr>
<th>Species concentration vector</th>
<th>Concentration or fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>target, $T$</td>
<td>$[T] = \langle[T_1], \ldots, [T_M]\rangle = [T](\Omega_1, \ldots, \Omega_M) = [T]\Omega$</td>
</tr>
<tr>
<td>selection target, $T_s$</td>
<td>$[T_s] = \langle[T_{s,1}], \ldots, [T_{s,M}]\rangle = [T_s](\Omega_{s,1}, \ldots, \Omega_{s,M}) = [T_s]\Omega_s$</td>
</tr>
<tr>
<td>negative selection target, $T_n$</td>
<td>$[T_n] = \langle[T_{n,1}], \ldots, [T_{n,M}]\rangle = [T_n](\Omega_{n,1}, \ldots, \Omega_{n,M}) = [T_n]\Omega_n$</td>
</tr>
<tr>
<td>free target, $T_f$</td>
<td>$[T_f] = \langle[T_{f,1}], \ldots, [T_{f,M}]\rangle = [T_f](\omega_1, \ldots, \omega_M) = [T_f]\omega$</td>
</tr>
<tr>
<td>free selection target, $T_s f$</td>
<td>$[T_s f] = \langle[T_{s,1} f_1], \ldots, [T_{s, M} f_M]\rangle = [T_s f](\omega_{s,1}, \ldots, \omega_{s,M}) = [T_s f]\omega_s$</td>
</tr>
<tr>
<td>free negative selection target, $T_n f$</td>
<td>$[T_n f] = \langle[T_{n,1} f_1], \ldots, [T_{n, M} f_M]\rangle = [T_n f](\omega_{n,1}, \ldots, \omega_{n,M}) = [T_n f]\omega_n$</td>
</tr>
<tr>
<td>nucleic acid, $NA$</td>
<td>$[NA] = \langle[NA_1], \ldots, [NA_N]\rangle = [NA](F_1, \ldots, F_N) = [NA]\widehat{F}$</td>
</tr>
</tbody>
</table>
5.3 Appendix C. Numerical values used to generate the figures

In order to generate figures in Chapters 2 and 3, for the case of\( M = 5\) and \(N = 20\), we used the affinity matrix \(A = A(1 : 5, 1 : 20)\) given in (5.4), (5.5) below:

\[
A(1 : 5, 1 : 10) = \begin{bmatrix}
822.37 & 618.81 & 521.92 & 984.25 & 759.88 & 1938.0 & 3164.6 & 1623.4 & 4629.6 & 2403.8 \\
4203.8 & 1091.7 & 1396.6 & 659.63 & 521.92 & 1225.5 & 706.21 & 8620.7 & 4629.6 & 1623.4 \\
4629.6 & 759.88 & 521.92 & 706.21 & 582.75 & 3164.6 & 984.25 & 550.66 & 1938.0 & 1396.6 .
\end{bmatrix}
\]

\[
A(1 : 5, 11 : 20) = \begin{bmatrix}
896.06 & 550.66 & 1225.5 & 1396.6 & 496.03 & 8620.7 & 659.63 & 706.21 & 582.75 & 1091.7 \\
759.88 & 896.06 & 496.03 & 1938.0 & 984.25 & 822.37 & 618.81 & 8620.7 & 4629.6 & 550.66 \\
896.06 & 706.21 & 659.63 & 822.37 & 1623.4 & 958.98 & 496.03 & 618.81 & 550.66 & 1396.6 \\
706.21 & 822.37 & 550.66 & 496.03 & 2403.8 & 1623.4 & 3164.6 & 1396.6 & 521.92 & 1938.0 .
\end{bmatrix}
\]

The values for the first row were chosen randomly over the range that corresponds to the range of values used in [11]. Each of the remaining rows were obtained from the first by doing a random reordering of the values of the first row.

The affinity selection matrix for the above matrix (the submatrix of columns 8,9,10,12,16) is:

\[
A_{\mathcal{S}} = \begin{bmatrix}
1623.4 & 4629.6 & 2403.8 & 550.66 & 8620.7 \\
8620.7 & 4629.6 & 1623.4 & 896.06 & 822.37 \\
550.66 & 1938.0 & 1396.6 & 8620.7 & 1623.4 \\
1396.6 & 4629.6 & 8620.7 & 706.21 & 1938.0 \\
8620.7 & 984.25 & 582.75 & 822.37 & 4629.6
\end{bmatrix}
\]

The classical adjoint of the preceding matrix is:

\[
adj(A_{\mathcal{S}}) = 10^{15} \begin{bmatrix}
-1.1585 & 0.7249 & -0.2051 & 0.0800 & 2.0669 \\
2.4628 & 4.6087 & -0.0722 & -1.2136 & -4.8713 \\
-1.52016 & -2.1133 & -0.1316 & 3.2899 & 1.8750 \\
-0.5970 & -0.3036 & 2.6343 & -0.2150 & 0.3426 \\
-1.9310 & -1.9990 & -0.0542 & -0.2669 & 1.6821
\end{bmatrix}
\]

For Figures 2.6 and 2.7 (three target case), we used the matrix:

\[
A = \begin{bmatrix}
4629.6 & 1396.6 & 1623.4 & 3164.6 & 1091.7 \\
3164.6 & 4629.6 & 1938.0 & 1623.4 & 1225.5 \\
1091.7 & 1623.4 & 4629.6 & 1225.5 & 2403.8
\end{bmatrix}
\]

For Figure 2.8 (three target case), panels (a,b) and Figure 2.9, we used

\[
A = \begin{bmatrix}
4629.6 & 1396.6 & 1623.4 & 3386.5 & 4420.2 \\
3164.6 & 4629.6 & 1938.0 & 4800 & 1925.5 \\
1091.7 & 1623.4 & 4630 & 2445 & 2103.8
\end{bmatrix}
\]

For Figure 2.8 (three target case), panels (c,d) and Figure 2.10, we used

\[
A = \begin{bmatrix}
9355 & 5529 & 1987 & 7468 & 846.2 \\
916.9 & 8132 & 3038 & 6451 & 5252 \\
4993.4 & 990 & 9722 & 931.8 & 2026
\end{bmatrix}
\]
For Figures 3.12-3.14 (three target case), we used

\[
A = \begin{bmatrix}
5538 & 2289 & 7513 & 6702 & 2435 & 2518 & 4791 \\
7265 & 3991 & 5896 & 8119 & 2925 & 8698 & 6062 \\
8641 & 3603 & 7822 & 6315 & 3663 & 3795 & 1420
\end{bmatrix}.
\]  

(5.11)

Table 5.3  Data for Figure 2.9, panel (a). \( \hat{\Omega} = (0.3, 0.6, 0.1) \).

| \( i \) | \( F_1^{(0)} \) | \( F_1^{(200)} \) | \( |F_1^{(0)} - F_1^{(1)}|_1 \) | \( |F_1^{(200)} - F_1^{(200)}|_1 \) |
|-----|---------|---------|---------|---------|
| 1   | (0.3062, 0.0731, 0.1917, 0.1535, 0.2815) | (0.0000, 0.0000, 0.0000, 1.0000, 0.0000) | 0        | 0        |
| 2   | (0.3129, 0.0746, 0.1934, 0.1529, 0.2699) | (0.0000, 0.0000, 0.0000, 1.0000, 0.0000) | 0.0301   | 0.0000   |
| 3   | (0.2846, 0.0587, 0.2099, 0.1599, 0.2935) | (0.0000, 0.0000, 0.0000, 1.0000, 0.0000) | 0.0600   | 0.0000   |
| 4   | (0.3161, 0.0430, 0.2083, 0.1386, 0.2940) | (0.0000, 0.0000, 0.0000, 1.0000, 0.0000) | 0.0901   | 0.0000   |
| 5   | (0.3008, 0.1325, 0.1614, 0.1350, 0.2703) | (0.0000, 0.0000, 0.0000, 1.0000, 0.0000) | 0.1201   | 0.0000   |
| 6   | (0.2542, 0.0947, 0.2451, 0.1481, 0.2579) | (0.0000, 0.0000, 0.0000, 1.0000, 0.0000) | 0.1500   | 0.0000   |

Table 5.4  Data for Figure 2.9, panel (b). \( \hat{\Omega} = (0.15, 0.3, 0.55) \).

| \( i \) | \( F_1^{(0)} \) | \( F_1^{(200)} \) | \( |F_1^{(0)} - F_1^{(1)}|_1 \) | \( |F_1^{(200)} - F_1^{(200)}|_1 \) |
|-----|---------|---------|---------|---------|
| 1   | (0.3062, 0.0731, 0.1917, 0.1535, 0.2815) | (0.0000, 0.0000, 0.4583, 0.5417, 0.0000) | 0        | 0        |
| 2   | (0.2991, 0.0833, 0.2015, 0.1536, 0.2625) | (0.0000, 0.0000, 0.4583, 0.5417, 0.0000) | 0.0401   | 0.4(10^{-12}) |
| 3   | (0.3329, 0.0474, 0.1846, 0.1513, 0.2838) | (0.0000, 0.0000, 0.4583, 0.5417, 0.0000) | 0.0700   | 0.3(10^{-12}) |
| 4   | (0.2943, 0.1093, 0.2011, 0.1583, 0.2370) | (0.0000, 0.0000, 0.4583, 0.5417, 0.0000) | 0.1008   | 0.4(10^{-12}) |
| 5   | (0.2561, 0.1227, 0.1708, 0.1660, 0.2844) | (0.0000, 0.0000, 0.4583, 0.5417, 0.0000) | 0.1301   | 1.1(10^{-12}) |
| 6   | (0.2693, 0.1212, 0.2436, 0.1338, 0.2321) | (0.0000, 0.0000, 0.4583, 0.5417, 0.0000) | 0.2901   | 3.6(10^{-12}) |

Table 5.5  Data for Figure 2.9, panel (c). \( \hat{\Omega} = (0.65, 0.15, 0.2) \).

| \( i \) | \( F_1^{(1)} \) | \( F_1^{(200)} \) | \( |F_1^{(1)} - F_1^{(1)}|_1 \) | \( |F_1^{(200)} - F_1^{(200)}|_1 \) |
|-----|---------|---------|---------|---------|
| 1   | (0.3002, 0.0731, 0.1917, 0.1535, 0.2815) | (0.1112, 0.0000, 0.0000, 0.4097, 0.4881) | 0        | 0        |
| 2   | (0.2958, 0.0880, 0.1912, 0.1433, 0.2817) | (0.1110, 0.0000, 0.0000, 0.4096, 0.4884) | 0.0301   | 0.0005   |
| 3   | (0.3065, 0.0717, 0.1631, 0.1652, 0.2935) | (0.1105, 0.0000, 0.0000, 0.4097, 0.4888) | 0.0601   | 0.0013   |
| 4   | (0.2833, 0.1022, 0.2062, 0.1284, 0.2839) | (0.1104, 0.0000, 0.0000, 0.4096, 0.4890) | 0.0801   | 0.0016   |
| 5   | (0.3115, 0.0841, 0.1432, 0.1470, 0.2942) | (0.1114, 0.0000, 0.0000, 0.4096, 0.4880) | 0.1100   | 0.0003   |
| 6   | (0.2817, 0.0689, 0.1344, 0.2239, 0.2911) | (0.1091, 0.0000, 0.0000, 0.4098, 0.4901) | 0.1600   | 0.0043   |
Table 5.6  Data for Figure 2.9, panel (d). $\hat{\Omega} = \langle 0.4, 0.1, 0.5 \rangle$.

| $i$ | $\hat{F}_i^{(1)}$ | $\hat{F}_i^{(200)}$ | $|\hat{F}_i^{(1)} - \hat{F}_i^{(200)}|_1$ |
|-----|-----------------|-----------------|----------------|
| 1   | (0.3092, 0.0731, 0.1917, 0.1535, 0.2815) | (0.0000, 0.0000, 0.3974, 0.4161, 0.1865) | 0.0000 |
| 2   | (0.2952, 0.0730, 0.1922, 0.1570, 0.2826) | (0.0000, 0.0000, 0.3974, 0.4162, 0.1864) | 0.0000 |
| 3   | (0.3120, 0.0746, 0.1934, 0.1520, 0.2680) | (0.0000, 0.0000, 0.3974, 0.4162, 0.1864) | 0.0000 |
| 4   | (0.3028, 0.0558, 0.1855, 0.1759, 0.2800) | (0.0000, 0.0000, 0.3974, 0.4169, 0.1858) | 0.0000 |
| 5   | (0.2920, 0.1193, 0.1959, 0.1386, 0.2548) | (0.0000, 0.0000, 0.3975, 0.4158, 0.1867) | 0.0000 |
| 6   | (0.3033, 0.0935, 0.2282, 0.1283, 0.2466) | (0.0000, 0.0000, 0.3975, 0.4155, 0.1870) | 0.0000 |

Table 5.7  Data for Figure 2.10, panel (a). $\hat{\Omega} = \langle 0.35, 0.4, 0.25 \rangle$.

| $i$ | $\hat{F}_i^{(1)}$ | $\hat{F}_i^{(200)}$ | $|\hat{F}_i^{(1)} - \hat{F}_i^{(200)}|_1$ |
|-----|-----------------|-----------------|----------------|
| 1   | (0.2384, 0.1693, 0.1056, 0.1075, 0.3792) | (0.1334, 0.4464, 0.2506, 0.1696, 0.0000) | 0.0000 |
| 2   | (0.2411, 0.1641, 0.1061, 0.1090, 0.3797) | (0.1318, 0.4427, 0.2514, 0.1741, 0.0000) | 0.0000 |
| 3   | (0.2440, 0.1637, 0.1048, 0.1122, 0.3753) | (0.1307, 0.4403, 0.2519, 0.1771, 0.0000) | 0.0000 |
| 4   | (0.2282, 0.1738, 0.1104, 0.1025, 0.3851) | (0.1353, 0.4509, 0.2497, 0.1641, 0.0000) | 0.0000 |
| 5   | (0.2460, 0.1845, 0.0939, 0.0942, 0.3814) | (0.1427, 0.4680, 0.2462, 0.1431, 0.0000) | 0.0000 |
| 6   | (0.2391, 0.1875, 0.0744, 0.1237, 0.3753) | (0.1337, 0.4470, 0.2505, 0.1688, 0.0000) | 0.0000 |

Table 5.8  Data for Figure 2.10, panel (b). $\hat{\Omega} = \langle 0.1, 0.6, 0.3 \rangle$.

| $i$ | $\hat{F}_i^{(1)}$ | $\hat{F}_i^{(200)}$ | $|\hat{F}_i^{(1)} - \hat{F}_i^{(200)}|_1$ |
|-----|-----------------|-----------------|----------------|
| 1   | (0.2384, 0.1693, 0.1056, 0.1075, 0.3792) | (0.0000, 0.6020, 0.3980, 0.0000, 0.0000) | 0.0000 |
| 2   | (0.2327, 0.1685, 0.1156, 0.1074, 0.3758) | (0.0000, 0.6020, 0.3980, 0.0000, 0.0000) | 0.0000 |
| 3   | (0.2166, 0.1661, 0.1134, 0.1081, 0.3956) | (0.0000, 0.6020, 0.3980, 0.0000, 0.0000) | 0.0000 |
| 4   | (0.2643, 0.1767, 0.0905, 0.1093, 0.3592) | (0.0000, 0.6200, 0.3980, 0.0000, 0.0000) | 0.0000 |
| 5   | (0.2378, 0.1562, 0.1545, 0.1090, 0.3425) | (0.0000, 0.6020, 0.3980, 0.0000, 0.0000) | 0.0000 |
| 6   | (0.1768, 0.1582, 0.0778, 0.2035, 0.3837) | (0.0000, 0.6020, 0.3980, 0.0000, 0.0000) | 0.0000 |
5.4 Appendix D. MATLAB codes for generating alternate SELEX:
Positive SELEX followed by negative SELEX

Alternate.m

% Alternate SELEX iteration scheme
function Alternate(T1,T2,T3,jj,jj2,m)

format long;
jj=jj*m; % jj is the number of positive selection steps, \( R_s \), \( m \) is a multiplier.
jj2=jj2*m; % jj2 is the number of negative selection steps, \( R_n \).

% The selection fraction \( \lambda = \frac{jj}{jj + jj2} \).

r1=10; % The number of iterations of alternate SELEX.

% In each round, \( jj = mR_s \) rounds of positive selection and \( jj2 = mR_n \) rounds of negative selection are performed.

M2=3; % Number of positive target components, \( M \).
N=5; % Number of nucleic acid species, \( N \).
NA2=1; % Total NA concentration, [NA], and it is fixed as one at every round to apply PCR amplification.

% Randomly selected initial NA fractions, \( \widehat{F}^{(1)} \).
F=rand(1,N);
F=F/sum(F); % Normalization.

Fn(:,1)=F';

% Initial positive target fraction vector, \( \widehat{\Omega}_s^{(1)} \).
TA(1,1)=T1; TA(1,2)=T2; TA(1,3)=T3;

% Affinity matrix.
A=POSData3;
for i=1:M2
for j=1:N
K(i,j)=1/A(i,j); % $K_{ij}$ is the dissociation equilibrium constant for the $j^{th}$ NA species bound to the $i^{th}$ target species.
end
end

% Here we have used the fixed point iteration scheme to solve for each of the free target species, $[T_{f_i}], i = \{1, 2, \cdots, M\}$ using the nonlinear system of equations, for both positive selection and negative selection. This iteration scheme is less labor intensive than Newton’s method and converges rapidly.
% Main loop starts here.
for r=1:r1

% Positive SELEX begins. This subprogram can be taken to use the multiple target (positive) SELEX simulations in Chapter 2.
for i=2:jj
TA(i,:)=(1-1/i)*TA(i-1,:); % Infinite positive target dilution, $[T_{a}^{(r+1)}] = (1 - s_r)[T_{a}^{(r)}]$, where $s_r = \frac{1}{r}$.
end

Fnn(:,1)=Fn(:,r); % This will be the starting NA fractions for positive SELEX updated after one or several rounds of negative selection.
for p=1:jj
Nr2=FixedPoint3(TA(p,:),K,Fnn(:,p),NA2,M2,N,A); % Fixed point iteration for solving the nonlinear system of equations for $[T_{s,f}]$.
for i=1:M2  
  Tf2(p,i)=Nr2(i);  % Free target vector $\overrightarrow{T_{sf}}$ computed from the fixed point iteration.  
end  

for i=1:M2  
  Omegaf2(p,i)=Tf2(p,i)/sum(Tf2(p,:));  % Free target fraction vector $\overrightarrow{\omega_s}$ normalized from the free target vector $\overrightarrow{T_{sf}}$.  
end  

for j=1:N  
  ss5=0;  
  for i=1:M2  
    ss5=ss5+Tf2(p,i)*A(i,j);  
  end  
  D2(j,p)=ss5;  % Compute $D_{s,j} = \sum_{i=1}^{M} [T_{sf}]_{ij} A_{ij}$ for each $j \in \mathcal{N}$.  
end  

ss6=0;  
for l=1:N  
  ss6=ss6+Fnn(l,p)*D2(l,p)/(1+D2(l,p));  % Compute the sum $\sum_{l=1}^{N} \frac{D_{s,l}}{1+D_{s,l}} F_{s,l}$.  
end  

for j=1:N  
  Fnn(j,p+1)=Fnn(j,p)*(D2(j,p)/(1+D2(j,p)))/ss6;  % Compute updated NA fractions $F'_{s,j}$.  
end  

end

% Positive selection ends.

Fs(:,r)=Fnn(:,jj+1);  % Keep the updated nucleic acid fractions and use them as a stating pool of nucleic acids for negative selection.

TA(1,:)=TA(jj,:);
% Negative SELEX begins.

M1=3;

% Negative target vector $\mathbf{T}_n$ with the size of total target concentration of $[T_n] = 100$.
TB(1,1)=100*T1/(T1+T3); TB(1,2)=100*0; TB(1,3)=100*T3/(T1+T3);
for i=2:jj2
TB(i,:)=TB(1,:); % Total negative target concentration is fixed from round to round.
end

NA(r,1)=1; % Starting nucleic acid concentration, $[NA]$, for negative SELEX.
Fss(:,1)=Fs(:,r); % Set the updated NA fractions after positive selection as the initial
                % NA fractions for negative selection.

for p=1:jj2
Nrb=FixedPoint3(TB(1,:),K1,Fss(:,p),NA(r,p),M1,N,A); % Fixed point iteration
        % to solve for $\mathbf{T}_n^f$.
        % Nrb=Newton2(TB(1,:),K1,Fss(:,p),NA(r,p),M1,N,A); % Newton’s method for solving
        % the nonlinear system of equations for $\mathbf{T}_n^f$.
        for i=1:M1
Tf3(p,i)=Nrb(i); % Free target vector $\mathbf{T}_n^f$ resulting from fixed point iteration.
end

for i=1:M1
Omegafb3(p,i)=Tf3(p,i)/sum(Tf3(p,:)); % Free target fraction vector $[\omega_n]$ normalized
                                       % from $\mathbf{T}_n^f$.
end

for j=1:N
s2=0;
for i=1:M1
s2=s2+Tf3(p,i)*A(i,j);
end
D1(j,p)=s2; % Compute $D_{n,j} = \sum_{l=1}^{M} [T_n^f_l] A_{lj}$.
end
s3=0;
for l=1:N
s3=s3+Fss(l,p)/(1+D1(l,p)); % Compute \[ NAf = \sum_{j=1}^{N} \frac{F_j}{1 + D_{n,j}} \]. Take \[ NA = 1. \]
end
for j=1:N
Fss(j,p+1)=(Fss(j,p)/(1+D1(j,p)))/s3; % Compute updated nucleic acids fractions \( F'_{n,j} \).
end
NA(r,p+1)=NA(r,1); % \[ NA \] is fixed from round to round for PCR amplification.
end
% Negative selection ends.
Fn(:,r+1)=Fss(:,jj2+1); % Updated NA fractions after \( R_n \) rounds of negative selection.
end

POSData3.m

% Affinity Matrix \( A \) : 3 by 5.
function A=POSData3
   A(1,1)=4629.6; A(1,2)=1396.6; A(1,3)=1623.4; A(1,4)=3164.6; A(1,5)=1091.7;
   A(2,1)=3164.6; A(2,2)=4629.6; A(2,3)=1938.0; A(2,4)=1623.4; A(2,5)=1225.5;
   A(3,1)=1091.7; A(3,2)=1623.4; A(3,3)=4629.6; A(3,4)=1225.5; A(3,5)=2403.8;
end

FixedPoint3.m

% Fixed point iteration method.
function Tfr=FixedPoint3(T,K,F,NA,M,N,A)
   format long;
   for i=1:M
      Tfr(1,i)=0; % Start with an initial guess as a zero vector.
   end
for r=1:10 % Number of iterations for fixed point.
∧ (20)
for j=1:N
s1(j)=0;
end
for j=1:N
for l=1:M
s1(j)=s1(j)+Tf(r,l)*A(l,j); % Compute $D_{j,f}$.
end
end
for i=1:M
s5=0;
for j=1:N
s5=s5+F(j)*A(i,j)/(1+s1(j)); % Compute $\sum_{j=1}^{N} F_j A_{ij} / (1 + D_{j,f})$.
end
Tf(r+1,i)=T(i)*(1+NA*s5)^(-1);
end
% Compute the free target vector $[Tf^{(k+1)}]$, where $k$ is the iteration number for fixed point iteration.

if norm(Tf(r+1,:)-Tf(r,:),inf) < 10^(-15)
m=r+1;
break
end
% A tolerance of $10^{-15}$ for convergence of fixed point iteration was used here. The iteration will be repeated until the relative error becomes smaller than $10^{-15}$.

**Remark 5.4.1.** Fixed point iteration method for solving the system of nonlinear equations
above is defined by a formula of the form, at $r^{th}$ round:

$$[Tf_i^{(r,k+1)}] = g([Tf_i^{(r,k)}]) \quad \text{for } i \in \mathcal{M} = \{1, 2, \ldots, M\} \quad \text{and } k > 1,$$

where $k$ is the number of iteration rounds and

$$g([Tf_i^{(r,k)}]) = \left[T_i^{(r,k)}\right] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j^{(r,k)} A_{ij}}{1 + \sum_{l=1}^{M} [Tf_l^{(r,k)}] A_{lj}}\right)^{-1}$$

using the equation (2.28). Starting with an initial vector, $[Tf(r,1)]$, compute, for every $i \in \mathcal{M}$,

$$[Tf_i^{(r,2)}] = g([Tf_i^{(r,1)}]) \quad \text{and } \quad ||[Tf_i^{(r,2)}] - [Tf_i^{(r,1)}]||.$$

Iterate the procedure until $||[Tf_i^{(r,k+1)}] - [Tf_i^{(r,k)}]|| < \epsilon$ for an arbitrary small $\epsilon$ and every $i$, and here we take $\epsilon = 10^{-15}$.

% Newton's method.

function Tfr=Newton2(T,K,F,NA,M,N,A)
format long;
for i=1:M
Tf(1,i)=0;
end

% Iteration starts from here.
for r=1:10 ^ (20)
for i=1:N
s1(i)=0;
end
for j=1:N
for l=1:M
s1(j)=s1(j)+Tf(r,l)*A(l,j);
end
end
for k1=1:M
for k2=1:M
if k1==k2
s2=0;
for j=1:N
s2=s2+F(j)*A(k1,j)/(1+s1(j));
end
s3=0;
for j=1:N
s3=s3+F(j)*A(k1,j)*A(k1,j)/(1+s1(j))^2;
end
J(k1,k1)=1+NA*s2-Tf(r,k1)*NA*s3;
else
s4=0;
for j=1:M
s4=s4+F(j)*A(k1,j)*A(k2,j)/(1+s1(j))^2;
end
J(k1,k2)=-Tf(r,k1)*NA*s4;
end
end
end
end
for k=1:M
s5=0
for j=1:N
s5=s5+F(j)*A(k,j)/(1+s1(j));
end
G(k)=-T(k)+Tf(r,k)*(1+NA*s5);
end
b=-G;
H=Gaussian2(M,J,b);      % Solve $JH = -G$ for $H = \text{inv}(J)^*b'$.

for i=1:M
    Tf(r+1,i)=Tf(r,i)+H(i);
end

if norm(Tf(r+1,:)-Tf(r,:),inf) < 10^(-15)
    j=r+1;
    Tfr=Tf(j,:);
    break
else
    Tfr=Tf(r+1,:);
end
end

function H=Gaussian2(M,J,b)      % Gaussian elimination.
for i=1:M
    p(i)=i;
    max1=0;
    for j=1:M
        if J(i,j)>=max1
            max1=abs(J(i,j));
        end
    end
    s(i)=max1;
end
for k=1:M-1
    max2=0;
    for i=k:M
        if abs(J(p(i),k))/s(p(i))>=max2
            max2=abs(J(p(i),k))/s(p(i));
            p(i)=k;
        end
    end
end

for i=1:M
    if J(i,p(i))>=max1
        J(i,:)=J(p(i),:);
        p(i)=p(i);
    end
end

for i=1:M
    if s(i)>=max1
        s(i)=max1;
    end
end

for k=1:M-1
    max2=0;
    for i=k:M
        if abs(J(p(i),k))/s(p(i))>=max2
            max2=abs(J(p(i),k))/s(p(i));
            p(i)=k;
        end
    end
end

for i=1:M
    if J(i,p(i))>=max1
        J(i,:)=J(p(i),:);
        p(i)=p(i);
    end
end
\begin{align*}
\max_{2} &= \text{abs}(J(p(i),k)) / s(p(i)); \\
j &= i; \\
end \\
end \\
Q &= p(j); \\
p(j) &= p(k); \\
p(k) &= Q; \\
\text{for } i &= k+1: M \\
z &= J(p(i),k) / J(p(k),k); \\
J(p(i),k) &= z; \\
\text{for } j &= k+1: M \\
J(p(i),j) &= J(p(i),j) - z \cdot J(p(k),j); \\
end \\
end \\
\text{for } k &= 1: M-1 \\
\text{for } i &= k+1: M \\
b(p(i)) &= b(p(i)) - J(p(i),k) \cdot b(p(k)); \\
end \\
end \\
H(M) &= b(p(M)) / J(p(M),M); \\
\text{for } k &= 2: M \\
s &= 0; \\
t &= M+2-k; \\
\text{for } j &= t: M \\
s &= s + J(p(M+1-k), j) \cdot H(j); \\
end \\
H(M+1-k) &= (b(p(M+1-k)) - s) / J(p(M+1-k), M+1-k); 
\end{align*}
Remark 5.4.2. For Newton’s method we let, for $k \in \mathcal{M},$
\[
g_k([Tf_1], [Tf_2], \cdots, [Tf_M]) = -[T_k] + [Tf_k] \left( 1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{kj}}{1 + \sum_{l=1}^{M} [Tf_l] A_{lj}} \right) = 0.
\]
Then the system of equations can be expressed
\[G(X) = 0,
\]
where $X = ([Tf_1], [Tf_2], \cdots, [Tf_M])'$ and $G = (g_1, g_2, \cdots, g_M)'.$ Suppose that $X$ is an approximate solution of $G(X) = 0.$ Next we compute the correction vector $H = (h_1, h_2, \cdots, h_M)'$ so that $G(X + H)$ will be a better approximate solution of the system of equations defined by $0 = G(X + H) \approx G(X) + G'(X)H,$ using the Taylor expansion. Here $G'(X)$ is the $M$ by $M$ Jacobian matrix with elements $\frac{\partial g_i}{\partial [Tf_j]}:
\[
G'(X) = \begin{pmatrix}
\frac{\partial g_1}{\partial [Tf_1]} & \cdots & \frac{\partial g_1}{\partial [Tf_M]}
\frac{\partial g_2}{\partial [Tf_1]} & \cdots & \frac{\partial g_2}{\partial [Tf_M]}
\vdots & \ddots & \vdots \\
\frac{\partial g_M}{\partial [Tf_1]} & \cdots & \frac{\partial g_M}{\partial [Tf_M]}
\end{pmatrix}.
\]
In practice, $H$ can be obtained from the equation $G'(X)H = -G(X)$ using Gaussian elimination. Thus Newton’s method is then given by
\[X^{(k+1)} = X^{(k)} + H^{(k)},
\]
where $G'(X^{(k)})H^{(k)} = -G(X^{(k)}).$ In the Matlab code above, we have used the matrix $J$ with the elements as follows:
\[
J(k, k) = \frac{\partial g_k}{\partial [Tf_k]} = 1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{kj}}{1 + \sum_{l=1}^{M} [Tf_l] A_{lj}} - [Tf_k] [NA] \sum_{j=1}^{N} \frac{F_j A_{kj} A_{kj}}{(1 + \sum_{l=1}^{M} [Tf_l] A_{lj})^2},
\]
and for $i \neq k$
\[
J(k, i) = \frac{\partial g_k}{\partial [Tf_i]} = -[Tf_k] [NA] \sum_{j=1}^{N} \frac{F_j A_{kj} A_{ij}}{(1 + \sum_{l=1}^{M} [Tf_l] A_{lj})^2}.
\]


ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, I would like to express my sincere gratitude to my advisors, Dr. Howard Levine (Dept. of Mathematics) and Dr. Marit Nilsen-Hamilton (Dept. of Biochemistry, Biophysics and Molecular Biology) for their expert advice, support and patience throughout my graduate career. Not only did they help me improve both mathematical and biological understandings for my research, but also their insights and encouragement kept me motivated and made me work harder. I would also like to thank my committee members Drs. Ralph Ackerman, Khalid Boushaba and Ryan Martin for serving on my committee.

I would like to thank the Department of Mathematics for its financial support during my graduate study. I would also like to thank my three-year officemate, Tracy, my colleagues and classmates at Iowa State for being such good friends to me. Special thanks goes to Jin-young Go for his assistance and guidance in graduate studies during my first two years at Iowa State.

I would like to express my deep and sincere thanks to Dr. Jerold Mathews and his wife, Eleanor Mathews, who continue to provide invaluable support, friendship and love. They have always treated me as part of their family and made me feel very much at home in Ames. They always believed that I would accomplish my goals and encouraged me to continue. I cannot express enough appreciation for their help and support for me.

Finally, I would like to express my deepest thanks to my parents, my sister, and my brother in South Korea for their unconditional love, patience and moral support. Without their support, I could not have made it this far.