Published in final edited form as:
Comput Biol Chem. 2007 February ; 31(1): 11–35.

A MATHEMATICAL ANALYSIS OF SELEX

Howard A. Levine and
Department of Mathematics, halevine@iastate.edu

Marit Nilsen-Hamilton
Department of Biochemistry, Biophysics and Molecular Biology, marit@iastate.edu, Iowa State University, Ames, Iowa, 50011, United States of America

Abstract

SELEX (Systematic Evolution of Ligands by Exponential Enrichment) is a procedure by which a mixture of nucleic acids can be separated into pure components with the goal of isolating those with specific biochemical activities.

The basic idea is to combine the mixture with a specific target molecule and then separate the target-NA complex from the resulting reaction. The target-NA complex is then separated by mechanical means (for example by nitrocellulose filtration), the NA is then eluted from the complex, amplified by PCR (polymerase chain reaction) and the process repeated. After several rounds, one should be left with a pool of [NA]that consists mostly of the species in the original pool that best binds to the target. In Irvine et al. (1991) a mathematical analysis of this process was given.

In this paper we revisit Irvine et al. (1991). By rewriting the equations for the SELEX process, we considerably reduce the labor of computing the round to round distribution of nucleic acid fractions. We also establish necessary and sufficient conditions for the SELEX process to converge to a pool consisting solely of the best binding nucleic acid to a fixed target in a manner that maximizes the percentage of bound target. The assumption is that there is a single nucleic acid binding site on the target that permits occupation by no more than one nucleic acid. We analyze the case for which there is no background loss, (no support losses and no free [NA] left on the support.) We then examine the case in which such there are such losses. The significance of the analysis is that it suggests an experimental approach for the SELEX process as defined in Irvine et al. (1991) to converge to a pool consisting of a single best binding nucleic acid without recourse to any a-priori information about the nature of the binding constants or the distribution of the individual nucleic acid fragments.

1. Introduction

In this paper we present an alternative approach to that used in Irvine et al. (1991) to analyze mathematically the process of SELEX. Our goal is to simplify the mathematical analysis and to thereby provide the experimentalist a means of improving upon the success of this process.

First we provide a detailed description of the SELEX process as it is performed in the laboratory. Then we develop a mathematical framework to describe and analyze this process by which nucleic acids with new functions can be selected from a large random pool of nucleic acid sequences.[1] The plan of the paper is as follows:

- Section 2: The SELEX process is introduced and a mathematical overview of the paper is given.
- Section 3: Here the notation and the equilibrium equations are given. The notion of the efficiency of the selection process is defined.
- Section 4: The SELEX process is defined mathematically as an iteration scheme.
- Section 5: A necessary and sufficient condition for the convergence of this iteration scheme is given in the case that there are no losses of products through the support or binding of free nucleic acid to the support (partitioning). This case is the mathematical ideal.
- Section 6: Here partitioning is precisely defined as in Irvine et al. (1991).
- Section 7: The two major theorems on the convergence of the SELEX process are given when there are losses through the support or free nucleic acid binding to the support are stated. These theorems give necessary and sufficient conditions for the convergence of the SELEX process. Although they are asymptotic results, in concrete cases they give practical information as is shown in the simulations.
- Section 8: We give upper and lower bounds on the number of rounds needed to raise the concentration fraction of the best binding nucleic acid from a very small fraction of the total pool (as little as one molecule in 10^{12} for example) to a very large fraction of the total pool.
- Section 9: In this section, a number of simulations are given based on a very simple Matlab program that illustrate the theorems and approximations discussed in the preceding sections.
- Section 10: A discussion of SELEX from a geometric point of view is given.
- Section 11: The proofs of Theorems 1, 2, and 3 are given.
- Section 12: The simple Matlab programs are given. There is one main program and three small function subprograms.
- Section ??: In this section, mostly out of curiosity, we replace the discrete iteration scheme by an analogous system of ordinary differential equations. The results analogous to Theorems 1, 2, and 3 are deduced from the solution of the system of ordinary differential equations.

2. The SELEX process and mathematical overview

2.1. The SELEX Process

Antibodies have served medical science extremely well for diagnostics and, in some special cases, as medications. More recently it has been discovered that certain single-stranded nucleic acids can adopt similar properties to antibodies in having high affinity and high specificity for their target molecule. Although they were only discovered in 1990, aptamers are already being developed as analytical agents (Tombelli et al,, (2005)) and for clinical treatments (Cerchia et al. (2002)). One aptamer, that recognizes vascular endothelial growth factor, is now in clinical use to treat macular degeneration (Zhou et al.(2006)). Among the many advantages of aptamers over antibodies are the stability of aptamers for diagnostics and their lack of immunogenicity for clinical treatments. Another important characteristic of aptamers is that they can be selected

[1] The term ”ligand” is sometimes used interchangeably with the term ”nucleic acid” although it is more general than nucleic acid. In a reaction $A + B \leftrightarrows C$ the smaller molecular weight molecule of A and B is generally called the ligand while the larger is called the target. However, in SELEX, the target is sometimes smaller than the NA. However, throughout this paper we will always use the term ligand to mean the nucleic acid.

in vitro by a process called SELEX. Although most frequently depicted in the double helical structure of chromosomal DNA, nucleic acids (NA) are capable of forming many alternate structures; to whit the ribosome, transfer RNAs, ribosomes and aptamers. Aptamers are short single-stranded nucleic acids that behave like antibodies, binding their target molecules with high affinity and high specificity. However, antibodies and aptamers differ substantially in their stability and in the means by which they are obtained. Aptamers are prepared synthetically whereas antibodies still require an animal for their production.

Aptamers are selected by a selection process called SELEX (systematic evolution of ligands by exponential enrichment) an (Ellington et al. (1990) and Teurk et al. (1990)). This is a reiterative process of selection and amplification that can be combined with mutagenesis to expand the pool of possible NA for selection. Here we will deal only with the selection aspects of this process, which starts with a randomized pool of nucleic acids that has been prepared synthetically. Each molecule in the pool is of the same length, but varies in an internal sequence (generally 40–80 bases long) in which positions along the polymer are randomly assigned to one of the four bases (A, G, C, T/U)[2]. Although the technology for producing and amplifying the pool differs depending on whether the molecules in the pool are RNA or DNA, the same basic steps are performed to isolate aptamers that bind a target (T) with high affinity and specificity (Figure 1).

The first step in SELEX is to use T attached to a solid support (S) such as a filter or a column to select molecules with sequences that promote their folding into structures that bind T. The interaction between T and NA is assumed to be at equilibrium and thus can be represented as $T + NA_i \leftrightarrows T{:}NA_i$ in which NAi is the ith NA in the pool. The equilibrium constant (Kd) for each NAi is different and characteristic of the NAi sequence. Because the use of S is often technically necessary to achieve the separation, there is also the possibility that certain NA sequences will fold to structures that bind S. Thus, another set of equilibria that occurs in every incubation of T and S with NA is $S + NAi \leftrightarrows S{:}NA_i$ with a variety of Kd’s that are characteristic of the individual NAi.

In each round of SELEX, the goal is to select for the NAi with the highest affinity (lowest Kd) for T. Therefore, after incubating T and NA the T:NA complexes are separated from T and NA, generally with the aid of S. The S:NAi is retained and captured together with the T:NAi. In some selection protocols, T:NA is then separated from S and S:NA. The bound NA is then extracted from T and S. When T:NA cannot be first separated from S and S:NA, the extracted pool contains NA that was bound to T (the desired aptamers) and NA that was bound to S (undesired background). Thus, part of the SELEX process is to minimize the number of background molecules and maximize the number of desired aptamer molecules.

Three general approaches are used to eliminate background in SELEX. The most common approach is to remove the background NA by incubating with S alone then discarding NA:S (Conrad, (1994)). Another approach is to associate T to S through a reversible linkage that can be broken prior to extracting NA from T:NA (Bock et al. (1992)). A third approach, that has more recently been developed, is to dispense with S by using capillary electrophoresis to separate T from T:NA (German et al. (1998)). Thus in some cases, one can dispense with S and, hence, as in Irvine et al. (1991), we will not include equilibrium $S{+}NA_i \leftrightarrows S{:}NA_i$ in our analysis.

After NA has been extracted from T (and S) this new NA population is amplified by polymerase chain reaction (PCR) to make more NA of the same sequences. PCR utilizes a heat stable DNA polymerase and the predefined sequences that are present at the termini of each NA molecule

[2] When referring to bases in NA sequences, T (thymine) is the base in DNA and U (uracil), is the equivalent base in RNA.

in the pool. With primers that are complementary to the predefined sequences, and by going through multiple cycles of annealing, polymerization and melting, the PCR protocol grows the population to a size that is equal to or larger than the original SELEX population. This amplified population is then used for a new round of SELEX in which the binding species are again selected from the population as just described.

Once it is determined that a binding population has evolved (by measuring K_D and the bound fraction [$T{:}NA$]/[NA]) the population is cloned, which produces a sample set of NA from the population. Each molecule in the sample set is sequenced and all the sequences are aligned in a search for identities. The presence of identical sequences amongst the sample set of groups identifies members of the population that have likely been selected through the process. If the population contained two or more molecules with similar K'_ds then two or more sub populations will be found in the sample set. Putative aptamer sequences identified in this way are chemically synthesized and tested for their ability to bind the target.

Although it is a matter of luck that the original NA population contains one or more NA sequences that have a high affinity for T, some aspects of SELEX protocols can be optimized for successful selection of an aptamer from the pool. Examples of these factors are the concentrations of T and NA and their ratios. Success in SELEX is also influenced by background binding NA:S, which should be as low as possible. This paper presents a mathematical analysis of SELEX with the intent of providing practical guidance for SELEX experiments in the laboratory.

2.2. Mathematical overview

We show that, under ideal conditions, selection will occur in all cases. The target concentration also tends to zero with increasing round number in an ideal selection. However, if the selection conditions are not ideal and some bound target passes through the support, or some unbound nucleic acid binds to the support (nonspecific binding), the selection will fail if the decrease in target from round to round is not done within a range of increments that can be defined mathematically.

The underlying goal of the mathematical analysis is to give a formula for the number of rounds needed to raise the concentration of a pool of nucleic acids that consists of at least one molecule per unit volume of the best binding NA to a pool that consists of some specified percentage of the best binding NA. Such an ideal formula would depend on (1), the desired percentage; (2), the ratio of target concentration to total pool concentration, and (3), the errors or losses in passing from round to round, i.e. the fraction of NA molecules that bind to the support and on the capture fraction by the support of the bound target-nucleic acid complex; (4), the initial distribution of nucleic acids in the pool and finally (5); the dissociation constants themselves, which, like the distribution of nucleic acids in the original pool, may not be known, or known only approximately. (In the latter situation, one may have some idea of the ratio of the largest to the smallest dissociation constant in the pool.)

Precise conditions for a successful non-ideal SELEX experiment are given in this paper. Theorems 2 and 3 provide the basis for an experimental SELEX protocol that requires little prior knowledge of the nature of the binding constants or the numerical distribution of the concentrations of each nucleic acid component in the pool.

In Irvine et al. (1991), the authors resort to solving a large nonlinear system of equations numerically to illustrate the mathematical underpinnings of the SELEX method. We show that one needs only to solve a single nonlinear equation in the free target for its sole positive root. Once this is known, it is a simple matter to calculate the bound target from the total target and to then to track how the concentration ratios [NA_i]/[NA_1] vary from round to round where

$[NA_i]$ denotes the concentration of the i^{th} nucleic acid species. The assumption here is that the first species binds better to the target than all the other species in the pool. We also give some upper and lower bounds for the round number needed to reach a specified pool fraction of the best binding nucleic acid.

Other modeling approaches have been made to the SELEX problem, (Djordjevic et al. (2006), Levitan (1997) and Sun et al.(1994)), but we believe our theoretical and computational approaches offer the advantages of simplicity and ease of applicability for the practitioner as it rests on mass action considerations (i.e. the law of large numbers) rather than individual probabilistic considerations. One approach, based on probability arguments is given in (Sun et al. (1994)) in the case in which there is no loss through the support of captured target and no nonselective retention of nucleic acids. If the optimal nucleic acid is very rare in the first round of SELEX, one may miss it entirely. Thus there is a very real need for a probabilistic model that goes beyond that of Sun et al. (1994)). In this paper, the assumption is that we are operating in the range of the law of large numbers so that we may use the Law of Mass Action with impunity.

We believe however, that our results provide a practical algorithm for carrying out the SELEX process in the laboratory. This is especially important because the individual binding constants are generally not known, although free energy considerations were used to estimate them in some special cases in (Sun et al. (1994)) for example.

Finally we remark that the SELEX process is, in some ways, mathematically analogous to to multicomponent distillation processes. See McCAbe et al. (2001).

3. Chemistry

Here we establish the following equivalence: Selection will be approached at maximum target efficiency if and only if the overall dissociation constant converges to the smallest dissociation constant and the concentration of the total target converges to zero. This equivalence is established near the end of this section in subsection 3.2. In order to do this, we need to define our terms and our problem carefully. (For example, allowing the total target to approach zero in the continuum sense is not a physical notion any more than the terminology ”infinite dilution” is.)

3.1. Notation and Mathematical overview

The notation is given in Table 1.

Here we frame the underlying chemistry of a single SELEX round in terms of chemical equilibria. Following Irvine et al. (1991), we exclude the possibility of nucleic acid binding to the support S. We envisage an initial pool of N nucleic acids, NA_i, for $i = 1, 2, \ldots N$. Here NA stands for nucleic acid which could be DNA or RNA. These are called nucleic acid ligands. They bind to a target molecule T via the dissociation-association:

$$\{T : NA_i\} \underset{k_i}{\overset{k_{-i}}{\rightleftharpoons}} Tf + NAf_i, \tag{3.1}$$

assumed to be in equilibrium. The dissociation constant for each of the N nucleic acids is given by:

$$K_{di} = \frac{k_{-i}}{k_i} = \frac{[NAf_i][Tf]}{[\{T : NA_i\}]} \tag{3.2}$$

where

$$[NA_i] = [NAf_i] + [\{T : NA_i\}]. \tag{3.3}$$

Thus, solving for the bound target:

$$[\{T : NA_i\}] = \frac{[NA_i][Tf]}{K_{di} + [Tf]} = [NA]\frac{[F_i][Tf]}{K_{di} + [Tf]}. \tag{3.4}$$

where we have set

$$F_i = \frac{[NA_i]}{[NA]},$$

the fraction of the i^{th} nucleic acid. It is assumed that the dissociation constants are ordered: $0 < K_{d1} < K_{d2} \cdots < K_{dN}$. Otherwise, they are to be regarded as unknown. Ordering them is done simply for mathematical convenience. Any set of N distinct numbers can be ordered.

In addition there is the overall dissociation constant given by

$$K_d = \frac{[NAf][Tf]}{[\{T : NA\}]} \tag{3.5}$$

where

$$\begin{aligned}[NA] &= \sum_{i=1}^{N}[NA_i],\\ [NAf] &= \sum_{i=1}^{N}[NAf_i],\\ [\{T : NA\}] &= \sum_{i=1}^{N}[\{T : NA_i\}]\end{aligned} \tag{3.6}$$

denote the total NA, the total free NA and the total bound target respectively. The total bound target can be determined under the stoichiometric assumption that there is only one NA bound to a target molecule, an assumption made here and in Irvine et al. (1991). In a given round of the SELEX process, one begins with a pool of nucleic acids for which one knows the initial total concentration of nucleic acids, the initial concentration of binding target, and the overall dissociation constant. Thus

$$\begin{aligned}[NA] &= [\{T : NA\}] + [NAf],\\ [T] &= [\{T : NA\}] + [Tf].\end{aligned} \tag{3.7}$$

Thus, using (3.5), (3.6) and (3.7)

$$\frac{[NA][Tf]}{K_d + [Tf]} = [\{T : NA\}] = \sum_{i=1}^{N}[\{T : NA_i\}] = [Tf][NA]\sum_{i=1}^{N}\frac{F_i}{K_{di} + [Tf]}. \tag{3.8}$$

Thus

$$\frac{1}{K_d + [Tf]} = \sum_{i=1}^{N}\frac{F_i}{K_{di} + [Tf]} = \mathscr{H}(\overrightarrow{F}, [Tf]) \tag{3.9}$$

where $\mathscr{H}$ is defined by the left hand side and where $\vec{F} = (F_1, F_2, \ldots, F_N)$. Thus the overall constant K_d depends only on the free target, the individual dissociation constants, and the fractions of each nucleic acid in the pool. Note also that $\sum_{i=1}^{N} F_i = 1$. Because $\sum_i F_i = 1,\ 1/(K_{d1} + [Tf]) > \mathscr{H}(\vec{F}, [Tf]) = 1/([Tf] + K_d) > 1/(K_{d_N} + [Tf])$ it follows that

$$K_{d1} < K_d(\vec{F}, [Tf]) < K_{d_N}, \tag{3.10}$$

i. e., the overall dissociation constant must lie between the largest and smallest such constants.

The overall constant K_d is also a function of the total target, the total nucleic acid and the free target in the given pool:

$$K_d = \frac{([NA] + [Tf] - [T])[Tf]}{[T] - [Tf]}. \tag{3.11}$$

Thus, one can eliminate K_d between (3.9), (3.11) to obtain a single nonlinear equation for the free target. This is easily found as follows: From the second equation in (3.7) and the far right hand expression for the bound target as a sum in equation (3.8) one finds

$$[T] = [Tf] + [Tf]\sum_{i=1}^{N} \frac{[NA_i]}{K_{di} + [Tf]} = [Tf] + [Tf][NA]\sum_{i=1}^{N} \frac{F_i}{K_{di} + [Tf]}. \tag{3.12}$$

The extreme ends of this equation give a single nonlinear equation for the free target. The bound target concentration is then

$$[T] - [Tf] = [\{T : NA\}] = [Tf][NA]\sum_{i=1}^{N} \frac{F_i}{K_{di} + [Tf]},$$

the maximum concentration of nucleic acid available for amplification by PCR.

Turning to the individual fractions, new concentration fractions of nucleic acids are related to the old via

$$F'_i = \frac{[\{T : NA_i\}]}{[\{T : NA\}]} = \frac{[NA]}{[\{T : NA\}]} = \frac{[Tf]}{K_{di} + [Tf]}\frac{[NA_i]}{[NA]} = \frac{K_d([Tf]) + [Tf]}{K_{di} + [Tf]}F_i. \tag{3.13}$$

From a mathematical point of view, one only has to follow the ratios F_i/F_1, i.e.

$$\frac{F'_i}{F'_1} = \frac{[\{T : NA_i\}]}{[\{T : NA_1\}]} = \frac{K_{d1} + [Tf]}{K_{di} + [Tf]}\frac{F_i}{F_1}.$$

The beauty of PCR from the chemist's point of view is that the ratios $[\{T{:}N\,A_i\}]/[\{T{:}NA\}]$ do not change under PCR. Therefore we can adjust (at least in principle) the concentration of the new pool to be the same as the concentration of the original pool *without changing the ratio* F'_i/F'_1. Thus the concentration of $[NA]$ can be regarded as constant from round to round.

Because the dissociation constants increase in *i* the ratio in $[Tf]$ is smaller than unity and is a minimum at $[Tf] = 0$. This formula needs to be modified when there is nonselective binding of nucleic acids by the support, or losses of bound target (Irvine et. al., 1991). We revisit it in Section 6.

Unlike the procedure followed in Irvine et al. (1991), we adopt a different approach. Equation (3.12) is a single nonlinear equation of the form $F([Tf], [NA]) = [T]$. If the pool concentration $[NA]$ is given, the fractional distributions of the nucleic acids and the values of the dissociation constants are known (or at least estimable) then, given the target concentration $[T]$, it is a simple matter to use Newton's method (for example) to calculate $[Tf]$. Once this is found, all the new ratios are easily computed. (Notice that

$$F_1'\left(1+\sum_{i=2}^{N}\frac{F_i'}{F_1'}\right)=1$$

so that if one knows $F_1, \ldots, F_N$ and $F_2'/F_1', \ldots, F_N'/F_1'$, then one knows all the fractions at the next round.)

In the laboratory, one usually fixes $[NA]$ and takes $[T] \to 0$ as the round number increases. What justifies such a protocol? The ratios

$$R_j=\frac{[T:NA_j]}{[T:NA]}=\frac{[Tf][NA]\frac{F_j}{K_{dj}+[Tf]}}{[Tf][NA]\sum_{i=1}^{N}\frac{F_i}{K_{di}+[Tf]}}=\frac{1}{1+\sum_{i\neq j}^{N}\frac{(F_i/F_j)(K_{dj}+[Tf])}{K_{di}+[Tf]}}$$

represent the fraction of bound NA_j to total bound NA. (These can also be viewed as the relative likelihood of binding one NA type to the binding of any type.) One sees that when $j = 1$, this ratio will be a maximum at $[Tf] = 0$ because

$$\frac{d}{d[Tf]}\sum_{i=2}^{N}\frac{(F_i/F_j)(K_{d1}+[Tf])}{K_{di}+[Tf]}=\sum_{i=2}^{N}\frac{(F_i/F_j)(K_{di}-K_{d1})}{(K_{di}+[Tf])^2}$$

is strictly positive unless we are at selection. Hence R_1 is decreasing in $[Tf]$ and has its maximum at $[Tf] = 0$. Likewise, if we compute $d[R_N]/d[Tf]$ we see that this ratio is strictly increasing in $[Tf]$ and hence has its maximum when $Tf] = [T] = +\infty$.[3] This justifies the protocol. It also says that 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. (The concept is closely related to the concept of maximum bound target efficiency as defined below.)

The argument above does **NOT** say that $R_1 > R_N$. To take an extreme example, if we have only one target molecule, a pool consisting of two species of nucleic acids, one that bind with an affinity of only 1/100 that of the the other but the concentration of the poorer binder is 10^6 times that of the better binder, the interaction of the pool with the target is going to lead to the target bound to the poorer binder far more often than to the to the target bound to the better binder. (For the example, $R_1 = 10^{-4}$ and $R_2 \approx 0.999998$ when $Tf \approx 0$. The reader should keep in mind that we are talking about equilibrium thermodynamics here and not kinetics.)

In theory as one decreases the target from round to round, 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, as the above example shows, one might

[3] The values of the free target for which the other ratios are maximized can be found, if they exist, by solving the nonlinear equations

$$\sum_{i=1}^{j-1}\frac{F_i(K_{dj}-K_{di})}{(K_{di}+[Tf])^2}=\sum_{i=j+1}^{N}\frac{F_i(K_{dj}-K_{di})}{(K_{di}+[Tf])^2}$$

for $j = 2,\ldots, N-1$.

miss the the best binder altogether as one lowers the target. Another manifestation of this can be seen in Figure 8. We see that as the initial target is decreased, the round number to achieve a fixed level of selection first decreases and then increases. The decreasing of the round number reflects the the improved opportunity given to the best binder while the increasing of the round number as the initial target level continues to fall reflects the fact that R_1 is much smaller than R_N (at zero free target) and more rounds are needed to change this inequality.

The fundamental issue remains. How do we choose the target from round to round? The theorems we develop here tell us that in the absence of information about the dissociation constants, there is, at least in principle, a way to reduce the target concentration from round to round, fixing the total pool size, in such a way as to insure that selection occurs. This is the subject of Section 5 and Section 7.

We sometimes suppress the argument $\vec{F}$ in $K_d(\vec{F}, [Tf])$ and in $[Tf](\vec{F}, [T])$ in the interest of readability.

3.2. Efficiency and selection

Operating under the assumption that at most one nucleic acid binds to a single target, the SELEX process can be monitored by following either the relative concentration of bound NA or the overall dissociation constant and the free NA. To see this define the fraction of bound target as $[\{T{:}NA\}]/[T] = ([T] - [Tf])/[T]$. Then

$$[T]_b \equiv \frac{[T] - [Tf]}{[T]} = \frac{[NA]\mathscr{H}(\overrightarrow{F}, [Tf])}{1 + [NA]\mathscr{H}(\overrightarrow{F}, [Tf])} = \frac{[NA]}{K_d(\overrightarrow{F}, [Tf]) + [Tf] + [NA]}. \tag{3.14}$$

We can write:

$$K_d([T], [T]_b) = (1 - [T]_b)([NA] - [T][T]_b)/[T]_b. \tag{3.15}$$

Equation (3.15) tells us that if we monitor $[T]$, $[T]_b$, we can monitor the overall dissociation constant. From (3.14) we see that $\lim_{[Tf]\to 0}[Tf]/[T] = 1/(1 + [NA]\mathscr{H}(\overrightarrow{F}, 0))$ and thus $[Tf] \to 0$ if and only if $[T] \to 0$ when $[NA]$ is fixed.

Consequently

$$\frac{K_{d1}}{K_{d1} + [NA]} < \lim_{[Tf]\to 0} \frac{[Tf]}{[T]} < \frac{K_{dN}}{K_{dN} + [NA]}$$

equality holding at one side or the other according as $\vec{F} = (1, 0, \ldots, 0, 0)$ or $\vec{F} = (0, 0, \ldots, 0, 1)$.

From (3.14), because the ratio on the right is increasing in $\mathscr{H}$ and $\mathscr{H}$ is decreasing in $[Tf]$, the ratio is a maximum when $[Tf] = 0$. Whatever the value of $[Tf]$, the maximum value of the relative concentration must occur at $F = (1, 0, \ldots, 0)$. Thus

$$\max\{[T]_b | 0 \le [Tf] \le [T] < \infty\} = \tfrac{[NA]}{K_d(0)+[NA]} \text{ and}$$
$$\max\left\{[T]_b | \sum_{i=1}^{N} F_i = 1, F_i \ge 0\right\} = \tfrac{[NA]}{K_{d1}+[Tf]+[NA]} \tag{3.16}$$

while

$$\max\left\{[T]_b \middle| 0 \le [Tf] \le [T] < \infty, \sum_{i=1}^{N} F_i = 1, F_i \ge 0\right\} = \frac{[NA]}{K_{d1} + [NA]}. \tag{3.17}$$

We call $\frac{[NA]}{K_{d1}+[NA]}$ the maximum bound target efficiency.

Thus, we approach selection at maximum bound target efficiency (i. e. at the maximum value of the bound fraction) if and only if $K_d \to K_{d1}$ and $[Tf] \to 0$ (or $[T] \to 0$).

4. The selection process as an iterative scheme

The sequential process, selection, PCR, selection …, can be written an iterative scheme. To do this, we introduce notation that suitably represents this process. For the initial step, we have *NA* fractions, $\overrightarrow{F^{(1)}} = \left\{F_1^{(1)}, \ldots, F_N^{(1)}\right\}$, with $\sum_i F_i^{(1)} = 1$ and a starting concentration of target $[T]_1$. After the initial pool is exposed to the target (in the presence or absence of a support), we obtain as output, new *NA* fractions, $\overrightarrow{F^{2}} = \left\{F_1^{(2)}, \ldots, F_N^{(2)}\right\}$ and some free target that is then discarded. (The free target can be viewed as output from the first round. However, it is notationally simpler to call it $[Tf]_1$.) We then select a new target $[T]_2$. More generally, we are given a fixed sequence of target fractions $\{[T]_r\}_{r=1}^{\infty}$ with $[T]_1 \le [NA]$. We make any assumptions on this sequence that can be realized in the laboratory. At the r^{th} step we have *NA* fractions, $\overrightarrow{F^{(r)}} = \left\{F_1^{(r)}, \ldots, F_N^{(r)}\right\}$, with $\sum_i F_i^{(r)} = 1$. We obtain a new pool, $\overrightarrow{F^{(r+1)}} = \left\{F_1^{(r+1)}, \ldots, F_N^{(r+1)}\right\}$ defined as follows: First we compute the free target left over from the reaction at the r^{th} step by solving

$$[T]_r = [Tf]_r(1 + [NA]\mathscr{H}(\overrightarrow{F^{(r)}}, [Tf]_r)) \tag{4.1}$$

for $[Tf]_r$ in terms of $[T]_r$. This value is then used to compute the fractions in the new pool from those in the old pool by evaluating the right hand sides of

$$F_i^{(r+1)} = \frac{K_d(\overrightarrow{F^{(r)}}, [Tf]_r) + [Tf]_r}{K_{di} + [Tf]_r} F_i^{(r)} \tag{4.2}$$

for $i = 1, \ldots, N$. This is much simpler than the procedure described in (Irvine et al. 1991).

5. Convergence of the selection process in the case of no background interference

The proof of Theorem 1 is given in Appendix B (Section 11).

Theorem 1

Assume that there is no loss through the support, that $F_1^{(1)} > 0$ *and* $[T]_1 \ge [T]_r$ *for* $r \ge 2$. *Then the iterative scheme will converge to a pool consisting only of the best binding nucleic acid and*

$$\lim_{r \to +\infty} K_d(\overrightarrow{F^{(r)}}, [Tf]_r) = K_{d1}. \tag{5.1}$$

The two conclusions above are equivalent. The convergence to selection, when it occurs, will be at maximum target efficiency if and only if $[Tf]_r \to 0$. *(See subsection 3.2).*

Remark 1

From the proof of Theorem 1 one sees that the convergence to selection is very rapid. Indeed, from equation (11.7) *in Appendix 11 one has for* $N \geq i \geq 2$

$$\frac{F_i^{(r+1)}}{F_1^{(r+1)}} \div \frac{F_i^{(1)}}{F_1^{(1)}} = \frac{\prod_{k=1}^{r}(K_{d1} + [Tf]_k)}{\prod_{k=1}^{r}(K_{di} + [Tf]_k)} < \left(\frac{K_{d1} + [Tf]_1}{K_{d2} + [Tf]_1}\right)^r = e^{-rQ} < 1$$

where $Q = \ln(K_{d2} + [Tf]_1)/(K_{d1} + [Tf]_1)$. *Thus the decay to zero of the mole fractions of all except the best binding aptamer is at least exponentially fast. This will be the case if* $K_d^{(r)}$ *is close to* K_{d1} *and* $[Tf]_r$ *is small. Thus it is important to monitor* K_d *approach selection at maximum bound target efficiency.*

Remark 2

Given a sequence of input targets, $\{[T]_r\}$ *with* $[T]_r < [T]_1$ *for* $r \geq 2$, *the corresponding sequence of overall dissociation rate constants will converge to the dissociation constant of the best binding nucleic acid and the concentrations of the nucleic acid pool will approach that of a pool consisting solely of the best binding nucleic acid. However, the approach will be optimal (at maximum target efficiency) if and only if* $[T]_r \to 0$.

6. Partitioning

In practice, there are experimental losses. When the sample is passed through a support, some free NA will be bound to the support. Also, some of the product will be lost through the support. Following (Irvine et al. 1991), we say that the NA pool has been partitioned. Again following (Irvine et al. 1991), we express the individual NA relative concentrations in the form:

$$[\{T:NA_i\}]^{part} = b_g[NAf_i] + c_p[\{T:NA_i\}] = b_g F_i[NA] + (c_p - b_g)[\{T:NA_i\}]. \tag{6.1}$$

where, in the author's notation, c_p is the percent of captured target caught by the i^{th} NA species that is eluted from the support and b_g is the percent of background free NA_i that is used for PCR by being nonselectively trapped by the support. In principle c_p and b_g should be species dependent. However, at the outset, following (Irvine et al. 1991), we assume they are not because it is difficult to measure them individually. Then summing (6.1) over all species, we have

$$[\{T:NA\}]^{part} = b_g[NA] + (c_p - b_g)[\{T:NA\}]. \tag{6.2}$$

In order to compute the percent of NA_i available for *PCR* we now define $\delta = b_g/(c_p - b_g)$ and $\varepsilon = \delta/(1+\delta) = b_g/c_p$:

$$F_i' = \frac{[\{T:NA_i\}]^{part}}{[\{T:NA\}]^{part}} = \frac{\delta F_i[NA] + [\{T:NA_i\}]}{\delta[NA] + [\{T:NA\}]} = F_i \frac{\delta + [Tf]/(K_{di} + [Tf])}{\delta + [Tf]\mathscr{H}(\vec{F}, [Tf])} \tag{6.3}$$

where again $[\{T:NA\}] = [T] - [Tf] = [Tf]\mathscr{H}(\vec{F}, [Tf])$ and set (suppressing the arguments in $[Tf](F, [T])$ and in $K_d(F, [Tf])$ on the right hand side)

$$E_i([Tf], \delta) = \frac{F_i'}{F_i} = \frac{\delta + [Tf]/(K_{di} + [Tf])}{\delta + [Tf]/(K_d + [Tf])} = \left(\frac{\varepsilon K_{di} + [Tf]}{K_{di} + [Tf]}\right)\left(\frac{K_d + [Tf]}{\varepsilon K_d + [Tf]}\right). \tag{6.4}$$

Notice that the last term consists of the product of two factors, the first is always less than unity (when 0< ε < 1 and $[T f] > 0$) while the second is always larger than unity for this range of ε. Notice that $1 < E_i([T f], \delta) < E_i([T f], 0)$ if and only if $K_{di} < K_d$. Thus, it is better to use

$$\frac{F_i'}{F_1'} = \frac{[\{T : NA_i\}]}{[\{T : NA\}]} = \left(\frac{K_{d1} + [T f]}{K_{di} + [T f]}\right)\left(\frac{\varepsilon K_{di} + [T f]}{\varepsilon K_{d1} + [T f]}\right)\frac{F_i}{F_1} \tag{6.5}$$

When $\delta > 0$ we see that as $[T f] \to 0$ or as $[T f] \to +\infty$, the ratio $E_i/E_1 \to 1$. Thus the extreme values of E_i/E_1 must occur for nonzero values of the free target. It is an easy exercise in calculus to show that each ratio has unique minimum value of

$$\left(\frac{\sqrt{\varepsilon} + \sqrt{K_{d1}/K_{di}}}{1 + \sqrt{\varepsilon K_{d1}/K_{di}}}\right)^2$$

which occurs at $[T f] = \sqrt{\varepsilon K_{d1} K_{di}}$.

7. Convergence of the selection process in the case of NA partitioning

There are, as when ε= 0, zero, a number of fixed points for the scheme, each having the form $F_i^j = \delta_{ij}$ with $K_d = K_{dj}$ for $j = 1, 2, \ldots N$. (Here $\delta_{ij} = 1$ or $\delta_{ij} = 0$ according as $i = j$ or $i \neq j$.) The goal is to determine necessary and sufficient conditions for the convergence of the iterative sequence to converge to the fixed point corresponding to the case $j = 1$.

We establish two theorems. In the first theorem, we assume that $[T]_1 \geq [T]_r \geq [T]_0 > 0$ with round number. In the second, it is assumed that $[T]_r \to 0$ with round number.

Theorem 2

Suppose, in the selection process we define input target concentrations $[T]_r$ recursively by the rule $[T]_{r+1} = (1 - s_r)[T]_r = \prod_1^r (1 - s_k)[T]_1$. Suppose also that $[T]_r \to [T]_0 > 0$. That is, the series $\Sigma_r s_r$ is convergent. Suppose also that $F_1^{(1)} > 0$. Then the iterative scheme will converge to a pool consisting only of the best binding nucleic acid and

$$\lim_{r \to +\infty} K_d(\overrightarrow{F^{(r)}}, [T f]_r) = K_{d1}. \tag{7.1}$$

The two conclusions are equivalent. The convergence to selection, when it occurs, will fail to be at maximum bound target efficiency because $\{[T f]_r\}$ is bounded below by a positive constant.

Theorem 3

Suppose, in the selection process we define input target concentrations $[T]_r$ recursively by the rule $[T]_{r+1} = (1 - s_r)[T]_r = \prod_1^r (1 - s_k)[T]_1$. Suppose also that $[T]_r \to 0$ with round number. (Equivalently, $\Sigma_r s_r$ is a divergent series.) Then a necessary and sufficient condition for the SELEX method to converge to the best binding nucleic acid is that the series

$$\sum_{r=1}^{\infty}\left[\prod_{k=1}^{r}(1 - s_k)\right] \tag{7.2}$$

is divergent. Moreover, if the series is divergent: The convergence of the iterative scheme to a pool consisting only of the best binding nucleic acid and

$$\lim_{r\to+\infty} K_d(\overrightarrow{F^{(r)}}, [T\,f]_r) = K_{d1} \tag{7.3}$$

are equivalent statements. The convergence to selection, when it occurs, will be approach maximum bound target efficiency because $[T\,f]_r \to 0$. *(See subsection 3.2.)*

A useful corollary is the following:

Corollary 1

If $\{z_r\}_{r=0}^{\infty}$ *satisfies*

$$z_r \geq z_{r+1} > 0 \textit{ and } \lim_{r\to\infty} z_r = 0,$$

with

$$\sum_{r=1}^{\infty} z_r = +\infty,$$

then

$$\{s_r\}_{r=1}^{\infty} = \left\{1 - \frac{z_r}{z_{r-1}}\right\}_{r=1}^{\infty}$$

satisfies the conditions of Theorem 3. Conversely, if $\{s_r\}_{r=1}^{\infty}$ *is a sequence such that this theorem holds, then the sequence given by recursively by* $z_0 = 1$, $z_{r+1} = s_{r+1} z_r$ *satisfies the above conditions.*

Thus it is relatively easy to generate sequences for which one can satisfy the conditions of the theorem.

For example, if $z_r = 1/(r + 1)$, then $s_r = 1/(r + 1)$, (the harmonic sequence) then $\Sigma_r\, s_r = \Sigma_r\, 1/(r + 1)$ is a divergent series. Furthermore, the series in (11.11) reduces to this same series and hence selection will take place. The harmonic sequence $\{1/(r + 1)\}$ is not the only one with this property. For example, $z_r = 1/(r+1)\ \ln(r+2))$ will give a sequence with $s_r = 1 - r \ln r/(r+1) \ln(r+1)) \approx 1/r$ for large r also satisfies the conditions of the theorem. Thus, in the absence of any information about the dissociation constants, the harmonic sequence is a good choice for target reduction in each round in SELEX.

However, if the input target is reduced by a fixed fraction $1 - c$ at each step, then the series in (11.11) is a convergent geometric series and selection is not possible. (That is, it is not possible in the mathematical sense although clearly, the more slowly the (11.11) converges, i.e., the closer c is to unity, the more likely we are to get something approaching perfect selection.

In Section 9 we illustrate these results with numerical simulations.

8. Partial selection - Likelihood of success

Here we want to consider how many rounds will be needed to achieve a concentration of the best binding NA that is a large multiple σ of the other nucleic acid concentrations in pool. Our approach to this problem is somewhat different than that of (Irvine et al. 1991). We can write

$$\frac{F_i^{(r)}}{F_1^{(r)}} = \frac{F_i^{(1)}}{F_1^{(1)}} \prod_{k=1}^{r} \frac{(K_{d1} + [Tf]_k)(\varepsilon K_{di} + [Tf]_k)}{(K_{di} + [Tf]_k)(\varepsilon K_{d1} + [Tf]_k)} = \frac{F_i^{(1)}}{F_1^{(1)}} \prod_{k=1}^{r} \left(1 - \frac{(K_{di} - K_{d1})(1-\varepsilon)[Tf]_k}{(K_{di} + [Tf]_k)(\varepsilon K_{d1} + [Tf]_k)}\right) = P_{i,r} \frac{F_i^{(1)}}{F_1^{(1)}} \tag{8.1}$$

where $P_{i,r}$ denotes the indicated product.

Notice that the products $P_{i,r}$ satisfy

$$P_{2,r} > P_{3,r} > \cdots > P_{N,r}.$$

Because $\sum_1^N F_1^i = 1$ it follows that

$$\begin{aligned} F_1^{(r)}(1 + \Theta P_{N,r}) \le & \; 1, \\ F_1^{(r)}(1 + \Theta P_{2,r}) \ge & \; 1 \end{aligned} \tag{8.2}$$

where we have set

$$\Theta = \frac{\sum_{i=2}^{N} F_i^{(1)}}{F_1^{(1)}} = \frac{1 - F_1^{(1)}}{F_1^{(1)}}. \tag{8.3}$$

Thus

$$\frac{1}{1 + \Theta P_{2,r}} \le F_1^{(r)} \le \frac{1}{1 + \Theta P_{N,r}} \tag{8.4}$$

We want good upper bounds for $P_{2,r}$ (in order to get good lower bounds for $F_1^{(r)}$) and good lower bounds for $P_{N,r}$.

To get a good upper bound on $P_{2,r}$ note that

$$\begin{aligned} &\frac{(K_{d1}+[Tf]_k)(\varepsilon K_{d2}+[Tf]_k)}{(K_{d2}+[Tf]_k)(\varepsilon K_{d1}+[Tf]_k)} = 1 - \frac{(K_{d2}-K_{d1})(1-\varepsilon)[Tf]_k}{(K_{d2}+[Tf]_k)(\varepsilon K_{d1}+[Tf]_k)} \\ &\approx 1 - \frac{(K_{d2}-K_{d1})(1-\varepsilon)}{([Tf]_k+K_{d2})} \\ &\le 1 - (1 - K_{d1}/k_{d2})(1-\varepsilon) \equiv (1 - \Lambda_2). \end{aligned}$$

when we assume that $K_{d2} \gg [Tf]_k \gg \varepsilon K_{d1}$. If $[Tf]_k \approx \sqrt{K_{d1} K_d}$, this inequality will hold if $K_{d2}^2/(\varepsilon K_{d1} > K_d)$ and $K_d > \varepsilon K_{d1}$. The latter inequality is always true since $K_d > K_{d1}$. The former will be true if $K_{d2} > \sqrt{\varepsilon K_{d1} K_{dN}}$, a claim that will always hold if the background is small enough. On the other hand, it may take a number of preliminary rounds in order to get to the level for which $K_{d2} \gg [Tf]_k \gg \varepsilon K_{d1}$.

In this case, $P_{2,r} \le (1 - \wedge_2)^{(r)}$.

To get a good lower bound on $P_{N,r}$ we note that for any value of $[Tf]_r$

$$\frac{(K_{d1} + [Tf]_k)(\varepsilon K_{dN} + [Tf]_k)}{(K_{dN} + [Tf]_k)(\varepsilon K_{d1} + [Tf]_k)} \ge \left(\frac{\sqrt{\varepsilon} + \sqrt{K_{d1}/K_{dN}}}{1 + \sqrt{\varepsilon K_{d1}/K_{dN}}}\right)^2 \equiv (1 - \lambda_N)^2$$

where

$$\lambda_N = \frac{(1-\sqrt{\varepsilon})(1-\sqrt{K_{d1}/K_{dN}})}{1+\sqrt{\varepsilon K_{d1}/K_{dN}}}. \tag{8.5}$$

Therefore

$$\frac{1}{1+\Theta(1-\Lambda_2)^{(r)}} \le F_1^{(r)} \le \frac{1}{1+\Theta(1-\lambda_N)^{2r}} \tag{8.6}$$

or

$$\Theta(1-\Lambda_2)^{(r)} \le \frac{1-F_1^{(r)}}{F_1^{(r)}} \le \Theta(1-\lambda_N)^{2r}. \tag{8.7}$$

Suppose that $0 < \sigma < 1$. Then we can be sure that $F_1^{(r)} \ge \sigma$ if

$$r \ge r_U = \frac{\ln[\sigma\Theta/(1-\sigma)]}{\ln[1/(1-\Lambda_2)]} = \frac{\ln\left\{(\sigma/F_1^{(1)})[(1-F_1^{(1)})/(1-\sigma)]\right\}}{\ln[1/(1-\Lambda_2)]} \tag{8.8}$$

Whereas $F_1^{(r)} \le \sigma$ provided

$$r \le r_L = \frac{1}{2}\frac{\ln[\sigma\Theta/(1-\sigma)]}{\ln[1/(1-\lambda_N)]}. \tag{8.9}$$

Thus we define the interval of uncertainty as the interval (of integers) (r_L, r_U) where the value of the round number must belong in order for $F_1^{(r)}$ to achieve the value σ. *It is important to keep in mind that* (8.8) *holds only under the hypothesis that* $K_{d2} \gg [T\, f]_k \gg \varepsilon K_{d1}$. *Consequently, the number* r_U *may understate the number of rounds needed for* $F_1^{(r)} \ge \sigma$. *That is, we must allow for a certain number K of rounds say to take place before we can assert that* $K_{d2} \gg [T\, f]_k \gg \varepsilon K_{d1}$. *Thus, the interval of uncertainty is* $(r_L, r_U + K)$.

Notice that as $\varepsilon \to 0^+$,

$$\frac{r_U}{r_L} \to \frac{\ln(K_{dN}/K_{d1})}{\ln(K_{d2}/K_{d1})}$$

which ratio is unity when $N = 2$. Thus at least one of the two numbers r_U, r_L cannot give the required minimum number of rounds needed for $F_1^{(r)}$ to achieve the value σ unless there are only two nucleic acids present in the initial pool and $\varepsilon = 0$.

Now suppose in our initial pool we have M molecules per unit volume of [NA]. We are going to look at some distribution scenarios. We compute the interval of uncertainty with data from (Irvine et al. 1991).[4] First, suppose also that all but 1 of them are of the poorest binding type while the sole exception is of the best binding type. That is $F_1^{(1)} = 1/M$ and $F_N^{(1)} = (M-1)/M$ while none of the intermediate binders are present. Then $\Theta = M - 1$.

[4] The values of the dissociation constants above were reported in (Irvine et al. 1991) based on ”the observed correlation between nucleic acid information content and free energy of binding”. The authors refer to Berg et al. (1986), Stormo et al. (1991), and von Hippel et al. (1986) for details.

The number nucleotides, with distinct binding constants is taken as $N = 5$. The pool size is $[NA] = 3(10^{-5})M$. In order to take $[NA] = 1$, the dissociation constants have to be rescaled to this concentration. $K_{d1} = 4.8(10^{-9})M/[NA] = 1.6(10^{-4})$, $K_{d2} = 12.0(10^{-9})M/[NA] = 3(10^{-4})$, $K_{d3} = 17.0(10^{-9})M/[NA] = 5.7(10^{-4})$, $K_{d4} = 27.0(10^{-9})M/[NA] = 9(10^{-4})$, $K_{d5} = 3.2(10^{-7})/[NA] = 1.6(10^{-2})$ where $\varepsilon \approx 0.1/80 = 1.25(10^{-3})$. The input or target concentration, $[T] = [NA]10^{-3} = 3(10^{-8})M = [T]_1[NA]$. Hence $[T]_1 = 1.0(10^{-3})$. If the initial distribution is such that $F_1^{(1)} = 1/65536$ with $F_2^{(1)} = F_3^{(1)} = F_4^{(1)} = 0,\ F_5^{(1)} = 65535/65536$, then $(1 - F_1^{(1)})/F_1^{(1)} \approx 65535$. In order to find $[T f]$ we need to solve the equation arising from (3.9)

$$[T]_1 = [Tf]\left(1 + \frac{F_1^{(1)}}{K_{d1} + [Tf]} + \frac{1 - F_1^{(1)}}{K_{dN} + [Tf]}\right)$$

which, in this case leads to a cubic in $[T f]$. However, using the values for $[T]_1$, K_{d1}, K_{dN}, $F_1^{(1)}$ we can easily estimate the value of $[T f]$ as $[T f] \approx 1.6(10^{-5})$. Thus $K_{dN} \gg K_{d1} > [T f] \gg \varepsilon K_{d1} \approx 2.0(10^{-7})$. If one seeks a pool consisting of 84% of the best binding nucleic acid, then $\sigma = 0.84$ Then $\sigma/(1 - \sigma) = 5.25$. With $M = 65536$, $\ln(\Theta\, \sigma/(1 - \sigma)) = 12.7485$. We find that $\ln[1/(1 - \wedge_2)] \approx \ln[K_{d2}/K_{d1}] = \ln[1.875] = 0.628$ and this gives $r_U \approx 12.7485/0.628 \approx 21.0$. On the other hand $\sqrt{K_{d1}/K_{dN}} = 0.1$ while $\varepsilon = 1.25(10^{-3})$ so that$1 - \lambda_5 = 0.135/(1 + 0.00354) \approx 0.135$. Thus $2 \ln(1/(1 - \lambda_5)) = 4.04$ and hence $r_L \approx 3.18$. Thus we obtain 84% selectivity in not less than three nor more than 20 rounds.

Notice that if we only demand a 50% pool of the best binding aptamer, then $\sigma = 0.5$ and $\ln(\Theta\sigma/(1 - \sigma)) = \ln(65535)$ so that $r_U \approx 18$ while $r_L = 2.7$.

Using pubmed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed) as the search engine, a review of the recent literature (2003 through mid-2006) revealed 26 publications describing successful SELEX experiments (Boyce et al.(2006), Chen et al.(2003), Cerchia et al. (2005), Cui et al. (2004), DeStefano et al. (2004), Eulberg et al. (2005), Fan et al. (2004), Gening et al. (2006), Gopinath et al. (2006), Jarosch et al. (2006), Kim et al. (2003), Kulbachinskiy et al. (2004), Lee et al. (2004), Lee et al. (2005), Mi et al. (2005), Mochizuki et al. (2005), Moreno et al. (2003), Mori et al. (2003), Ogawa et al. (2004), Pileur et al. (2003), Rhie et al. (2003), Skrypina et al. (2004), Surugiu-Warnmark et al. (2005), Vo et al. (2003), Yang et al. (2006) and White et al. (2001)). In all instances the targets were proteins. The number of rounds prior to cloning varied from 7 to 22 with a mean of 12 ± 4 and a median of 12. These results identify the round at which each group of investigators identified binding activity of the aptamer(s) in the oligonucleotide pool and decided to clone the sequences. The decision to clone can vary depending on the results obtained from previous rounds of the SELEX experiment and does not indicate that a certain percentage of the oligonucleotides in the pool are aptamers of the highest a, nity form. Because it is prohibitively time consuming to test every oligonucleotide sequence in the cloned pool, the information regarding percent best binding aptamer sequences in the pool is usually sketchy at best. However, the collective results from a number of SELEX experiments should provide a view of the number of rounds it generally takes to obtain a population with measurable (greater than about 10%) binding activity. With the understanding that the experimental data is not uniform, the results from the mathematical model are consistent with the experimental data. The concordance of experimental results with mathematical predictions that are based only on chemical equilibria suggest that, in most SELEX experiments, the binding equilibrium is the major factor determining selection, whereas the evolution enabled by in vitro mutagenesis might not have a major impact on the rate of aptamer selection.

As a second example, suppose that we compare the best binding nucleic acid with the worst binding nucleic acid. Then $\wedge_2 = \wedge_N = (1 - \varepsilon)\, K_{d1}/K_{dN}$ and λ_N is given in (8.5). Suppose we have a pool consisting of $10^{(1)}2$ nucleic acids, 10^k of which are the best binder and the rest are of the worst binding type. Then $F_1^{(1)} = 10^{k-12}$ and $\Theta = 10^{12-k} - 1$. If we seek a pool of 50% of the best binding aptamer, we have $K_{d1}/K_{dN} = 10^{-2}$. Then $\ln(\sigma\, \Theta/(1 - \sigma)) \approx (12 - k) \ln 10 = 2.303(12 - k)$ while $-\ln(1 - \wedge_N) = -\ln(1 - 0.95(.99)) = -\ln(0.0595) = 2.821$ and $\lambda_N = (1 - \sqrt{(}0.05))(1 - 0.1)/(1 + \sqrt{(}0.05(.01)) = 0.69875/1.02236 = 0.682468$ and $-\ln(1 - \wedge_N) = 1.14766$. Thus $0.5(1.14766)2.303(12 - k) \geq r \geq 2.303(12 - k)/2.821$ or

$$1.32(12 - k) \geq r \geq 0.816(12 - k). \tag{8.10}$$

Thus when $k = 0$ we should need not fewer than 10 rounds nor no more than 16 rounds to get a pool consisting of 50%. of the best binding aptamer. If we have 10 molecules of the best binder so that $k = 1$, then $9 \leq r \leq 15$. See Figures 11, 12.

9. Simulations

In this section, we present some simulations. We take a fixed number $N = 15$ of nucleic acids and a fixed linear ordering of the dissociation constants. In Figures 2–8 we use $K_{di} = (1.6+2.2(i - 1))10^{-4}$, $i = 1,\ldots, N$ (rescaled to a fixed pool size of $[NA] = 3(10^{-5}M)$. We started with a nucleic acid pool generated by using a random number generator. Once the pool is selected, it is fixed for the Figures 3–9.

In Figures 11–16 we increased the spread of the dissociation constants by a factor of 10, i. e. we used This range is consistent with the data used in (Irvine et al. 1991), figure 4. We also looked at the worst case pool distribution, i. e. $F_1^{(1)} = \ldots F_N^{(1)} = 10^{-12}$ and $F_N^{(1)} = 1$.

Because we do not need to solve a large system of equations, the Matlab program we use runs very rapidly. Figures 3–11 are organized as follows:

1. In the first set of experiments, we take $\varepsilon = 0.05$, $[T]_1 = 1.0$ and vary the reduction sequence. The choices for $\{s_r\}_{r=1}^{\infty}$ are $\left\{1/(r+1)^{5/2}\right\}_{r=1}^{\infty}$ (to illustrate Theorem 2), $\left\{r^2/(r^2+1)\right\}_{r=1}^{\infty}$ (Theorem 3, series in (7.2) is convergent, no selection) and $\{1/(2r+1)\}_{r=1}^{\infty}$ (Theorem 3, series in (7.2) is divergent, selection). We give the graphs of K_d, $[T f]$, $[T]_b$, as functions of round number.

   In Figures 3 and 5 we have selection. However, while the panels 2 in both figures indicate that the overall dissociation constant, K_d, is converging to the smallest such constant, K_{d1} with round number, the convergence is faster in the case of Figure 4. Also, there is much less free target left and more efficient binding in the case of Figure 5 as compared to Figure 3 (compare panels 3, 4 in both figures).

2. For the second set of experiments, Figure 6, with $[T]_1 = 1.0$, we examine the case that $\{s_r = s\}_{r=1}^{\infty}$, a series of constants. We compare slow reduction $s = 0.1$, 0.4, 0.6 and $s = 0.95$. While we are led to geometric series in (7.2) in all four cases, the rate of convergence of the series accelerates as s increases toward unity.

   Notice how, as we move from panel to panel in Figure 6, the number of rounds for which the poorer binders survive increases. Notice also that nearly perfect selection of the first nucleic acid becomes impossible to achieve in less than twenty rounds when $s > 0.5$.

3. In the third set of experiments, we start simulations with $[T]_1 = 0.1$ and with ε variable over the four values 0.05, 0.20, 0.40, 0.70 $s_r = 0.1$ for all round numbers r. In the panels in Figure 7, we see the effect of increasing $\varepsilon \approx b_b/c_p$ on selection.

4. In the fourth set of experiments, we start simulations with $\varepsilon = .05$ and with $[T]_1 = 4^{-k}$ for $k = -2, -1, 0, 1, 2, \ldots 2, 9$. We reduced the target by 10%($s_r = 0.1$) in every case. See Figure 8.

   We see from Figures 8 that there is an optimal starting value for $[T]_1$ lying in the interval (1/16, 1/4) for which the round number leading to selection will be minimal. Much of the discussion in (Irvine et al. 1991), pages 749–753, is concerned with estimating this optimal starting value.

5. If we use the formula $[Tf] = \sqrt{\delta K_{d1} K_d(0)}$ in every round, we are led to fixing $[T](r) \approx 0.1451$ for every round. That is, we are setting $s_r = 0$. In Figure 9 we have given the nucleic acid fractions for a random pool with this fixed value for the input target with $\varepsilon = 0.05$.

   Although in this case there is no reduction in initial target from round to round, one needs to have a reasonable idea of the background and capture fractions as well as the geometric mean, $\sqrt{K_{d1} K_{dN}}$, of the smallest and largest dissociation constants in order to implement this in the laboratory. Notice also that in this case all the free target is not consumed nor is the binding fraction as close to unity as they are in Figure 5. (Compare panels 3 in Figures 5, 9 and panels 4 in Figures 5, 9.) Because we do not have convergence of the free target to zero (Figure 9, panel 3) we do not obtain convergence of the overall dissociation constant to K_{d1}. (Compare Figure 5, panel 2 with Figure 6, panel 2.)

6. In Figures 11 and 12 we use the values $K_{di} = (1.6 + 22\,(i - 1))10^{-4}$, $i = 1, \ldots, N$ and the starting value $[T]_1 = 1$. We chose $s_r = 2/(2r + 1)$ so that selection is assured. Here the pool is chosen in such a way that, $F_1^{(1)} = 10^{k-12}$ for $k = 1, 2, \ldots 5$ and $F_N^{(1)} = 1$ while $F_j^{(1)} = 10^{-12}$ for $2 \le j \le N - 1$. Figure 9 illustrates how the increase in the number of best binding molecules in the initial pool affects the overall dissociation constant as a function of round number. Figure 11 illustrates how the increase in the number of best binding nucleic acid in the initial pool affects the number of rounds need to bring the pool to a size consisting of 50% or more of best binding nucleic acid.

7. In Figure 13, we have taken $K_{d\mathrm{N}}/K_{d1} = 93$ and $M^b = 99$ in the first column of figures. We took $F_r^{(1)} = 10^{-12}$ if $n < 15$ and $F_N^{(1)} = 1$ if $N = 15$ as the initial pool distribution in all cases. $M_b/(1 + M_b)$ is the probability of binding one molecule of $[NA_1]$. (See (Irvine et. al., 1991) for details.)

   We follow the strategy of (Irvine et al. 1991) in that K_d is updated from round to round while $[T]_1 = M_b K_{d1} + M_b K_{d1}[NA]/(M_b K_{d1} + K_d([Tf]))$ is fixed in every round. This choice gives $[T]_1 = 0.5304$ as the initial target value. In the second column, we used this as a starting ratio along with $s_r = 2r/(2r + 1)$. In the first case we do not achieve even a 50-50 pool until after around 45 rounds while in the second case, we achieve this pool in less than 20 rounds. On the other hand, in Figure 14, we took $K_{dN}/K_{d1} = 9300$ and $M_b = 99$. Here $[T](1) = 0.1117$. We see that this time it is better to follow the strategy of (Irvine et al. 1991). Notice the shapes of the curves for K_d, $[T]_b$ are very similar. It appears from the second panel in the first row as though Theorem 3 is violated. In fact, selection does occur here also but it takes many more than 30 rounds to achieve it because the initial target value $[T](1)$ is so small. See Figure 8.

8. Figures 16 and 17 indicate that either the use of Theorem 3 (with $s_r = 2/(2r + 1)$ here) or use of equation $[Tf] = \sqrt{\varepsilon K_{d1} K_d([Tf])}$ of (Irvine et al. 1991) to select the free target from round to round, leads to very nearly the same round number for the nucleic acid pool to consist of 50% of the best binding molecule when there is only one initially present. The agreement is better, the larger the ratio K_{dN}/K_{d1} is. When this ratio is relatively small, of order 10 or less, it is probably better to resort to some other method for discriminating between aptamers such as cloning unless one has sufficient information about K_{d1} and K_d in order to be able to invoke $[Tf] = \sqrt{\varepsilon K_{d1} K_d([Tf])}$. Both methods for computing the round number at 50% lead to larger and larger values for the round number but once at least one of the methods gives an estimate 20 or more rounds, one should perhaps consider whether the time and expense of using the SELEX method is worth the expected outcome.

Acknowledgements

The authors thank Hans Weinberger for a number of useful discussions and comments that improved an earlier version of this paper.

The first author thanks the Institute for Pure and Applied Mathematics at UCLA for partial support of this research. Both authors acknowledge the support of NIH grant R42 CA110222.

References

Berg OG, von Hippel PH. Selection of DNA binding sites by regulatory proteins: statistical mechanical theory and application to operators and promoters. J Mol Biol 1986;193:723–750. [PubMed: 3612791]

Bock LC, Griffen LC, Latham JA, Vermass EH, Toole JJ. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. Nature 1992;355:564–566. [PubMed: 1741036]

Boyce M, Scott F, Guogas LM, Gehrke L. Base-pairing potential identified by in vitro selection predicts the kinked RNA backbone observed in the crystal structure of the alfalfa mosaic virus RNA-coat protein complex. J Mol Recognit 2006;19:68–78. [PubMed: 16312015]

Chen CH, Chernis GA, Hoang VQ, Landgraf R. Inhibition of heregulin signaling by an aptamer that preferentially binds to the oligomeric form of human epidermal growth factor receptor-3. Proc Natl Acad Sci U S A 2003;100:9226–31. [PubMed: 12874383]

Cerchia L, Duconge F, Pestourie C, Boulay J, Aissouni Y, Gombert K, Tavitian B, de Franciscis V, Libri D. Neutralizing aptamers from whole-cell SELEX inhibit the RET receptor tyrosine kinase. PLoS Biol 2005;3:e123. [PubMed: 15769183]

Cerchia L, Hamm J, Libri D, Tavitian B, De Franciscis V. Nucleic acid aptamers in cancer medicine. FEBS Lett 2002;528:12–16. [PubMed: 12297271]

Conrad R, Keranen LM, Ellington AD, Newton AC. Isozyme-specific inhibition of protein kinase C by RNA aptamers. J Biol Chem 1994;269:32051–32054. [PubMed: 7528207]

Cui Y, Rajasethupathy P, Hess GP. Selection of stable RNA molecules that can regulate the channel-opening equilibrium of the membrane-bound gamma-aminobutyric acid receptor. Biochemistry 2004;43:16442–9. [PubMed: 15610038]

DeStefano JJ, Cristofaro JV. Selection of primer-template sequences that bind human immunodeficiency virus reverse transcriptase with high affinity. Nucleic Acids Res 2006;34:130–9. [PubMed: 16397296]

Djordjevic M, Sengupta AM. Quantitative modeling and data analysis of SELEX experiments. Physical Biology 2006;3(13):13–28. [PubMed: 16582458]

Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific nucleic acids. Nature 1990;346:818–822. [PubMed: 1697402]

Eulberg D, Buchner K, Maasch C, Klussmann S. Development of an automated in vitro selection protocol to obtain RNA-based aptamers: identification of a biostable substance P antagonis. Nucleic Acids Res 2005;22:e45. [PubMed: 15745995]

Fan X, Shi H, Adelman K, Lis JT. Probing TBP interactions in transcription initiation and reinitiation with RNA aptamers that act in distinct modes. Proc Natl Acad Sci U S A 2004;101:6934–9. [PubMed: 15103022]

Gening LV, Klincheva SA, Reshetnjak A, Grollman AP, Miller H. RNA aptamers selected against DNA polymerase beta inhibit the polymerase activities of DNA polymerases beta and kappa. Nucleic Acids Res 2006;34:2579–86. [PubMed: 16707660]

German I, Buchanan DD, Kennedy RT. Aptamers as nucleic acids in affinity probe capillary electrophoresis. Anal Chem 1998;70:4540–4545. [PubMed: 9823713]

Gopinath SC, Misono TS, Kawasaki K, Mizuno T, Imai M, Odagiri T, Kumar PK. An RNA aptamer that distinguishes between closely related human influenza viruses and inhibits haemagglutinin-mediated membrane fusion. J Gen Virol 2006;87:479–87. [PubMed: 16476969]

Irvine D, Tuerk C, Gold L. SELEXION. Systematic evolution of nucleic acids by exponential enrichment with integrated optimization by non-linear analysis. J Mol Biol 1991;222:739–761. [PubMed: 1721092]

Buchner, Jarosch K.; Klussmann, S. Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX. Nucleic Acids Res 2003;31:e130. [PubMed: 14576330]

von Hippel PH, Berg OG. On the specificity of DNA-protein interactions. Proc Nat Acad Sci USA 1986;83:1608–1612. [PubMed: 3456604]

Kim YM, Choi KH, Jang YJ, Yu J, Jeong S. Specific modulation of the anti-DNA autoantibody-nucleic acids interaction by the high affinity RNA aptamer. Biochem Biophys Res Commun 2003;300:516–23. [PubMed: 12504114]

Kulbachinskiy A, Feklistov A, Krasheninnikov I, Goldfarb A, Nikiforov V. Aptamers to Escherichia coli core RNA polymerase that sense its interaction with rifampicin, sigma-subunit and GreB. Eur J Biochem 2004;271:4921–31. [PubMed: 15606780]

Lee SK, Park MW, Yang EG, Yu J, Jeong S. An RNA aptamer that binds to the beta-catenin interaction domain of TCF-1 protein. Biochem Biophys Res Commun 2005;327:294–9. [PubMed: 15629461]

Lee SY, Jeong S. In vitro selection and characterization of TCF-1 binding RNA aptamers. Mol Cells 2004;17:174–9. [PubMed: 15055546]

Levitan B. Models and Search Strategies for Applied Molecular Evolution. Ann rep Comb Chem and Mol Div 1997;1:1–72.

McCabe, WL.; Smith, JC.; Harriott, P. Unit Operations of Chemical Engineering. 5. McGraw-Hill; NY: 2001.

Mi J, Zhang X, Giangrande PH, McNamara JO 2nd, Nimjee SM, Sarraf-Yazdi S, Sullenger BA, Clary BM. Targeted inhibition of alphavbeta3 integrin with an RNA aptamer impairs endothelial cell growth and survival. Biochem Biophys Res Commun 2005;338:956–63. [PubMed: 16256939]

Mochizuki K, Oguro A, Ohtsu T, Sonenberg N, Nakamura Y. High affinity RNA for mammalian initiation factor 4E interferes with mRNA-cap binding and inhibits translation. RNA 2005;11:77–89. [PubMed: 15611299]

Moreno M, Rincon E, Pineiro D, Fernandez G, Domingo A, Jimenez-Ruiz A, Salinas M, Gonzalez VM. Selection of aptamers against KMP-11 using colloidal gold during the SELEX process. Biochem Biophys Res Commun 2003;308:214–8. [PubMed: 12901856]

Mori T, Oguro A, Ohtsu T, Nakamura Y. RNA aptamers selected against the receptor activator of NF-kappaB acquire general affinity to proteins of the tumor necrosis factor receptor family. Nucleic Acids Res 2004;32:6120–8. [PubMed: 15562003]

Ogawa A, Tomita N, Kikuchi N, Sando S, Aoyama Y. Aptamer selection for the inhibition of cell adhesion with fibronectin as target. Bioorg Med Chem Lett 2004;4:4001–4. [PubMed: 15225715]

Pileur F, Andreola ML, Dausse E, Michel J, Moreau S, Yamada H, Gaidamakov SA, Crouch RJ, Toulme JJ, Cazenave C. Selective inhibitory DNA aptamers of the human RNase H1. Nucleic Acids Res 2003;31:5776–88. [PubMed: 14500841]

Pollard, J.; Bell, SD.; Ellington, AD. Generation and Use of Combinatorial Libraries. In: Ausubel, GFM.; Brent, R.; Kingston, RE.; Moore, DD.; Seidman, JG.; Smith, JA.; Struhl, K., editors. Current Protocols in Molecular Biology. 4. New York, NY., USA: Greene Publishing Associates and John Wiley Liss & Sons, Inc.; 2000. p. 24.21.21-24.25.34.

Rhie A, Kirby L, Sayer N, Wellesley R, Disterer P, Sylvester I, Gill A, Hope J, James W, Tahiri-Alaoui A. Characterization of 2′-fluoro-RNA aptamers that bind preferentially to disease-associated conformations of prion protein and inhibit conversion. J Bio Chem 2003;278:39697–705. [PubMed: 12902353]

Stormo GD, Yoshioka M. Specificity of the mnt protein determined by binding to randomized operators. Proc Nat Acad Sci USA 1991;88:5699–5703. [PubMed: 2062848]

Skrypina NA, Savochkina LP, Beabealashvilli R. In vitro selection of single-stranded DNA aptamers that bind human pro-urokinase. Nucleosides Nucleotides Nucleic Acids 2004;23:891–3. [PubMed: 15560078]

Sun F, Galas D, Waterman MS. A mathematical analysis of *in vitro* molecular selection-amplification. J Mol Biol 1996;258(4):650–60. [PubMed: 8636999]

Surugiu-Warnmark I, Warnmark A, Toresson G, Gustafsson JA, Bulow L. Selection of DNA aptamers against rat liver X receptors. Biochem Biophys Res Commun 2005;332:512–7. [PubMed: 15910755]

Tombelli S, Minunni M, Mascini M. Analytical applications of aptamers. Biosens Bioelectron 2005;20:2424–2434. [PubMed: 15854817]

Tuerk C, Gold L. Systematic evolution of nucleic acids by exponential enrichment: RNA nucleic acids to bacteriophage T4 DNA polymerase. Science 1990;249:505–510. [PubMed: 2200121]

Wall, FT. Chemical Thermodynamics. W. H. Freeman; San Francisco and London: 1958.

Vo NV, Oh JW, Lai MM. Identification of RNA ligands that bind hepatitis C virus polymerase selectively and inhibit its RNA synthesis from the natural viral RNA templates. Virology 2003;307:301–16. [PubMed: 12667800]

White RR, Shan S, Rusconi CP, Shetty G, Dewhirst MW, Kontos CD, Sullenger BA. Inhibition of rat corneal angiogenesis by a nuclease-resistant RNA aptamer specific for angiopoietin-2. Proc Natl Acad Sci U S A 2003;100:5028–33. [PubMed: 12692304]

Yang C, Yan N, Parish J, Wang X, Shi Y, Xue D. RNA aptamers targeting the cell death inhibitor CED-9 induce cell killing in Caenorhabditis elegans. J Biol Chem 2006;281:9137–44. [PubMed: 16467303]

Zhou B, Wang B. Pegaptanib for the treatment of age-related macular degeneration. Exp Eye Res 2006;83:615–619. [PubMed: 16678158]

10. Appendix A. Geometric observations

The entire SELEX iteration scheme can be viewed to take place in the Cartesian product of two sets $\mathcal{T} \times \mathcal{S}$. The set $\mathcal{T}$ is given by

$$\mathcal{T} = \left\{ \overrightarrow{F} \in R^N | F_i \geq 0 \ \text{ and } \ \sum_{i=1}^{N} F_i = 1 \right\}, \tag{10.1}$$

the simplectic triangle in Euclidian N space.

The set $\mathcal{S}$ can be described as follows: In the three dimensional orthant determined by the inequalities $[Tf] \geq 0$, $[NA] \geq 0$, $[T] \geq 0$, there are two surfaces S_1, S_N say, defined by the equations

$$[T] = [Tf] + \frac{[NA][Tf]}{K_{d1} + [Tf]} \ \text{ and } \ [T] = [Tf] + \frac{[NA][Tf]}{K_{dN} + [Tf]}$$

respectively. Then

$$\mathcal{S} = \left\{ ([Tf], [T], [NA]) \ \ || \ \ [Tf] \geq 0, \ [NA] \geq 0, \ [T] \geq 0, \ \frac{[NA][Tf]}{K_{dN} + [Tf]} \leq [T] - [Tf] \leq \frac{[NA][Tf]}{K_{d1} + [Tf]} \right\} \tag{10.2}$$

is the region between and including the two surfaces S_1, S_N.

The surface S defined by

$$[T] = [Tf] + \frac{[NA][Tf]}{K_d(\overrightarrow{F}, [Tf]) + [Tf]}$$

must be between these two surfaces because $K_{d1} < K_d < K_{dN}$. Likewise the remaining N-2 surfaces S_i defined by $[T] = [Tf]+([NA][Tf])/(K_{di} +[Tf])$ for $i = 2, \ldots, N$-1 sit between these two surfaces. All N +1 surfaces intersect along the straight line $([Tf], [NA], [T]) = (0, [NA], [0])$. When one fixes $[NA] = [NA]_0 >$ say, the surfaces S_i intersect this plane in curves C_i, which are branches of hyperbolae. These curves are asymptotic to lines parallel to $[T] = [Tf]$ as $[Tf]$ becomes large. They have different limiting slopes $[T]'(0) = 1 + [NA]/K_{di}$ at $[Tf] = 0$, . The surface S has limiting slope $1 + [NA]/K_d(F, 0)$.

Depending on how one chooses $[T]_r \to 0$ and determines $[Tf]_r$ (from 3.12) and $\overrightarrow{F^{(r)}}$ (from (6.5)), one can obtain a limiting ratio $[T]_r/[Tf]_r$ that is different from $1 + [NA]/K_{d1}$, the desired limiting ratio for selection. (This cannot happen when $s = 0$.) The theorems give necessary and sufficient conditions on the sequence $\{[T]_r\}$ in order to obtain the correct limit. Equation (3.12) defines a functional dependence of $[Tf]$ on $[NA]$, $[T]$ as independent variables because the left hand side is a strictly increasing function of $[Tf]$. By means of implicit differentiation:

$$\frac{\partial[Tf]}{\partial[NA]} = -\frac{[\{T:NA\}]}{[NA]}\left(1 + [NA]\sum_{i=1}^{N}\frac{F_iK_{di}}{(K_{di}+[Tf])^2}\right)^{-1} < 0,$$
$$\frac{\partial[Tf]}{\partial[T]} = \left(1 + [NA]\sum_{i=1}^{N}\frac{F_iK_{di}}{(K_{di}+[Tf])^2}\right)^{-1} > 0. \qquad (10.3)$$

Therefore there is no extreme value for the free target in the region determined by $[NA] > 0$ and $[T] > 0$.

Likewise, using (3.9)

$$\frac{\partial K_d}{\partial[Tf]} = \left[\sum_{i=1}^{N}\frac{F_i}{(K_{di}+[Tf])^2} - \left(\sum_{i=1}^{N}\frac{F_i}{K_{di}+[Tf]}\right)^2\right](K_d + [Tf])^2 = S^2(K_d + [Tf])^2. \qquad (10.4)$$

Viewing $K_d = K_d([NA], [T])$ after elimination of $[Tf]$ from (3.9) and implicit differentiation again, we find

$$\frac{\partial K_d}{\partial[NA]} = S^2(K_d + [Tf])^2\frac{\partial[Tf]}{\partial[NA]}, \quad \frac{\partial K_d}{\partial[T]} = S^2(K_d + [Tf])^2\frac{\partial[Tf]}{\partial[T]}. \qquad (10.5)$$

It follows from (10.4), (10.5) and Schwarz's inequality[5] that the extreme value for K_d in $\mathcal{S}$ occur if and only if one of the fractions F_i vanish and the exception is unity, i.e. when F is a vertex of $\mathcal{T}$. When this happens, $K_d([NA], [T]) = K_{di}$ for some i. The smallest value of $K_d([NA], [T])$ occurs when $i = 1$. In this case $([Tf], [T], [NA])$ must be a point on S_1, one of the boundary surfaces, i. e. $([T] - [Tf])/[T] = [NA]/(K_{d1} + [Tf] + [NA])$. The maximum value of this expression, the maximum target efficiency, occurs at $([Tf], [T], [NA]) = (0, 0, [NA])$ for fixed $[NA]$ and increases to unity as $[NA] \to \infty$.

[5] Schwarz's inequality asserts that if x, y are two Euclidian vectors, then the magnitude of their scalar product cannot exceed the product of their Euclidian lengths and can equal this product if and only if the two vectors are collinear. In this case, the two vectors are $x = (\sqrt{F_1}, \sqrt{F_2} \ldots, \sqrt{F_N})$ and $y = (\sqrt{F_1}/(K_{d1} + [Tf]), \sqrt{F_2}/(K_{d2} + [Tf]), \ldots, \sqrt{F_N}/(K_{dN} + [Tf]))$.

11. Appendix B.Proofs of Theorems

Because the value of $[NA]$ plays no role in the proofs of the theorems, we take $[NA] = 1$ in this section.

11.1. Proof of Theorem 1

If we strike the ratio $F_i^{(r+1)}/F_1^{(r+1)}$ we see that

$$\frac{F_i^{(r+1)}}{F_1^{(r+1)}} = \frac{K_{d1} + [Tf]_r F_i^{(r)}}{K_{di} + [Tf]_r F_1^{(r)}}. \tag{11.6}$$

We see that for $i \geq 2$,

$$\frac{F_i^{(r+1)}}{F_1^{(r+1)}} \div \frac{F_i^{(1)}}{F_1^{(1)}} = \frac{\prod_{k=1}^{r}(K_{d1} + [Tf]_k)}{\prod_{k=1}^{r}(K_{di} + [Tf]_k)} < \frac{\prod_{k=1}^{r}(K_{d1} + [Tf]_k)}{\prod_{k=1}^{r}(K_{d2} + [Tf]_k)} < \left(\frac{K_{d1} + [T]_1}{K_{d2} + [T]_1}\right)^r < 1 \tag{11.7}$$

Because $K_{d2} \leq K_{di}$ and the ratio $(a + x) / (b + x)$ is increasing in x when a, b, x are all positive and $a < b$, except when $i = 1$, the coefficient of $F_i^{(r)}/F_1^{(r)}$ is bounded above by $(K_{d1} + [T]_1) / (K_{di} + [T]_1) < 1$. Hence, $i \neq 1$, $\lim_{r\to+\infty}$ $\lim_{r\to+\infty} F_i^{(r)} = 0$. since the sums $\sum_i F_i^{(r)} = 1$, this implies that $\lim_{r\to+\infty} F_1^{(r)} = 1$.

The convergence of the overall dissociation constants then follows from:

$$K_d(\overrightarrow{F^{(r)}}, [Tf]_r) - K_{d1} = \frac{\sum_{i=2}^{N}(K_{di} - K_{d1})F_i^{(r)}/(K_{di} + [Tf]_r)}{\sum_{i=1}^{N} F_i^{(r)}/(K_{di} + [Tf]_r)} \tag{11.8}$$

since the denominators on the right hand side are all bounded away from zero by K_{d1} and above by K_{dN} +$[T]_1$.

Conversely, if $\lim_{r\to+\infty} K_d(\overrightarrow{F^{(r)}}, [Tf]_r) - K_{d1} = 0$, we must have $lim_{r\to\infty} F_i^{(r)} = 0$ for $i \geq 2$. This establishes the equivalence.

11.2. Proof of Theorem 2

Again we strike the ratio $F_i^{(r+1)}/F_1^{(r+1)}$ to find

$$\frac{F_i^{(r+1)}}{F_1^{(r+1)}} \div \frac{F_i^{(1)}}{F_1^{(1)}} = \prod_{k=1}^{r} \frac{(K_{d1} + [Tf]_k)(\varepsilon K_{di} + [Tf]_k)}{(K_{di} + [Tf]_k)(\varepsilon K_{d1} + [Tf]_k)} \tag{11.9}$$

If $\{[T]_r\}$ is a convergent sequence with a nonzero limit, then the same is true of the sequence $\{[Tf]_r\}$. Thus we can assume that $0 \leq [T]_0 \leq [Tf]_r \leq [T]_1$. Because the function

$$f_i(x) = \frac{(K_{d1} + x)(\varepsilon K_{di} + x)}{(K_{di} + x)(\varepsilon K_{d1} + x)}$$

satisfies $f_i(x) < 1$ for $0 < x < \infty$ if $0 < \varepsilon < 1$, we know that on $[[T]_0, [T]_1]$ there is a constant ℓ_i such that $f_i(x) \leq \ell_i < 1$. Consequently, we have $\lim_{r\to\infty} F_i^{(r)} = 0$ if $i > 1$. This implies that $\lim_{r\to\infty}$

$K_d(\vec{F}, [Tf]_r) = K_{d1}$ as before. Likewise, if this limit holds, then from (11.8) it follows that $\lim_{r\to\infty} F_i^{(r)} = 0$ for $i \geq 2$.

11.3. Proof of Theorem 3

If $[T]_r \to 0$, then $[Tf]_r \to 0$ and the functions $f_i([Tf]_r)$ converge to unity. *Hence we cannot assume that we have selection in this case.* Thus the selection of the sequence $\{[T]_r\}$ is more delicate. We write $[T]_{r+1} = [T]_r(1 - s_r)$.

First we show that the sequence of vectors $\left\{\overrightarrow{F^{(r)}}\right\}$ converges to some vector and that the sequence $\left\{K_d(\overrightarrow{F^{(r)}}, [Tf]_r) \equiv K^{(r)}\right\}$ converges to some limit, L.

We have again

$$\frac{F_i^{(r+1)}}{F_1^{(r+1)}} \div \frac{F_i^{(1)}}{F_i^{(1)}} = \prod_{k=1}^{r} \frac{(K_{d1} + [Tf]_k)(\varepsilon K_{di} + [Tf]_k)}{(\varepsilon K_{d1} + [Tf]_k)(K_{di} + [Tf]_k)} = G_{i,r}.$$

The k^{th}factor in $G_{i,r}$ can be written in the form

$$\frac{(K_{d1} + [Tf]_k)(\varepsilon K_{di} + [Tf]_k)}{(\varepsilon K_{d1} + [Tf]_k)(K_{di} + [Tf]_k)} = 1 + \frac{(1-\varepsilon)(K_{d1} - K_{di})[Tf]_k}{(\varepsilon K_{d1} + [Tf]_k)(K_{di} + [Tf]_k)}. \tag{11.10}$$

A theorem of analysis says that if $|b_r| < 1$, then the infinite product $\prod_{r=1}^{\infty}(1 + b_r)$ converges to a *non zero* constant if and only if the series $\sum_{r=1}^{\infty}|b_r|$ is convergent. (This follows from the inequality $\ln(1 + |b_r|) \leq |b_r| \leq \ln(1 + 2|b_r|)$ valid for $0 \leq |b_{(r)}| \leq 1$.)

Recall that $\lim_{r\to\infty}[Tf]_r/[T]_r$ is positive and finite and suppose first that $\sum_{r=1}^{\infty}\prod_{k=1}^{r}(1 - s_k) < \infty$, i.e. the numbers $[Tf]_r$ form the terms of a convergent series or equivalently,

$$\sum_{r=1}^{\infty} \frac{(1-\varepsilon)(K_{di} - K_{d1})[Tf]_r}{(K_{di} + [Tf]_r)(\varepsilon K_{d1} + [Tf]_r)} < \infty. \tag{11.11}$$

Then for each i, $\lim_{r\to+\infty} F_i^{(r)}$ exists and is not zero.(The series in (11.11) will not converge if $\varepsilon = 0$.) Setting $\lim_{r\to+\infty} F_i^{(r)} = B_i(\varepsilon) > 0$, it follows that

$$\lim_{r\to+\infty} K_d(\overrightarrow{F^{(r)}}, [Tf]_r) = \left\{\sum_{i=1}^{N} \frac{B_i(\varepsilon)}{K_{di}}\right\}^{-1} > K_{d1}. \tag{11.12}$$

Hence selection does not occur in this case.

Suppose next that $\sum_{r=1}^{\infty}\prod_{k=1}^{r}(1 - s_k)$ diverges. Because the coefficients of $F_i^{(1)}/F_1^{(1)}$ are $G_{i,r}$ and the series (11.11) is now divergent, we conclude that the infinite products $G_{i,r}$ diverge to zero. Hence $F_i^{(r)} \to 0$ if $i \geq 2$ and consequently $K_d(\overrightarrow{F^{(r)}}, [Tf]_r) \to K_{d1}$. Thus selection occurs in this case.

Remark 3

It is of some mathematical interest to examine the total derivative of K_d as a function of $\vec{F}$, $[T f]$ along the iteration trajectory in $\mathcal{T} \times \mathcal{S}$. We show that:

$$\frac{\partial K_d(\overrightarrow{F}, [T f])}{\partial [T f]} \Delta [T f] + \sum_{i=1}^{N} \frac{\partial K_d(\overrightarrow{F}, [T f])}{\partial F_i} \Delta F_i = -s[T f] \left(\frac{S}{\mathscr{H}} \right)^2 - [T f] \left(\frac{S}{\mathscr{H}} \right)^2 \left(\frac{(1-\varepsilon)(K_d + [T f])}{\varepsilon K_d + [T f]} \right). \qquad (11.13)$$

where

$$S^2 = \left(\sum_{i=1}^{N} \frac{F_i}{(K_{di} + [T f])^2} \right) - \left(\sum_{i=1}^{N} \frac{F_i}{K_{di} + [T f]} \right)^2,$$

and use this to establish a relationship between the terms of the series $\sum_{r=1}^{\infty} \prod_{k=1}^{r} (1 - s_k)$ *and the rate of convergence of the sequence* $\left\{ K_d((\overrightarrow{F^{(r)}}, [T f]_r)) \right\}$ *to its limit.*

*We see from (*11.13*) that the first term describes how the differential changes in $\mathcal{S}$. The second term describes how this differential changes in $\mathcal{T}$. However, the change is being driven by how* $[T f] \to 0$ *at a rate that clearly depends on the background parameter ε. The closer ε is to unity, the less influential changes in of the F_i in $\mathcal{T}$ are on K_d.*

In our iteration scheme, a sequence $\left\{ \overrightarrow{F^{(r)}} \right\}$ *is generated using the formulas involving the products $G_{i,r}$ which tells us how to calculate the vector c→ and gives us specific information on the rule for determining* $\Delta \overrightarrow{F^{(r)}} = \overrightarrow{F^{(r+1)}} - \overrightarrow{F^{(r)}}$. *Suppose therefore that* $\overrightarrow{F^{(r)}} \to \overrightarrow{C} \in \mathcal{T}$ *and* $[T]_r \to 0$. *Then*

$$\frac{1}{K^{(r+1)} + [T f]_{r+1}} \to \sum_{i=1}^{N} \frac{c_i}{K_{di}} \equiv \mathscr{H}(\overrightarrow{c}, 0) \qquad (11.14)$$

as r → +∞. Likewise, $\left\{ K_d(\overrightarrow{F^{(r)}}), [T f]_r \right\} \equiv K^{(r)} \to 1/\mathscr{H}(\overrightarrow{c}, 0) \equiv L$.

Using the shorthand $\mathscr{H}^{(r)} = \mathscr{H}(\overrightarrow{F^{(r)}}), [T f]_r)$, *we have*

$$[T f]_r (1 + \mathscr{H}^{(r)}) = [T]_r = [T]_{r+1} + s_r [T]_r = [T f]_{r+1} (1 + \mathscr{H}^{(r+1)}) + s_r [t f]_r (1 + \mathscr{H}^{(r)}).$$

Thus

$$\frac{[T f]_{r+1}}{[T f]_r} = (1 - s_r) \frac{1 + \mathscr{H}^{(r)}}{1 + \mathscr{H}^{(r+1)}} = (1 - s_r) \frac{([T f]_{r+1} + K^{(r+1)})([T f]_r + K^{(r)} + 1)}{([T f]_r + K^{(r)})([T f]_{r+1} + K^{(r+1)} + 1)}$$

and hence, for any index m

$$\frac{[T f]_{r+m}}{[T f]_r} = \prod_{l=1}^{m} \frac{[T f]_{r+l}}{[T f]_{r+l-1}} = \frac{([T f]_{r+m} + K^{(r+m)})([T f]_r + K^{(r)} + 1)}{([T f]_r + K^{(r)})([T f]_{r+1} + K^{(r+m)} + 1)} \prod_{l=1}^{m} (1 - s_{l+r-1}).$$

The overall dissociation constants satisfy $K_d(\overrightarrow{F^{(r)}}, [Tf]_r) \equiv K^{(r)} \to L \geq K_{d1}$ *as* $[Tf]_r \to 0$. *Hence for all m and all sufficiently large r,*

$$\frac{[Tf]_{r+m}}{[Tf]_r} \approx \prod_{l=1}^{m} (1 - s_{l+r-1}) \tag{11.15}$$

We abandon the round number index r temporarily for readability. We approximate ΔK_d *to first order in* ΔF_i, $\Delta[Tf]$ *directly from the equation* $K_d = -[Tf] + 1/(\mathscr{H}(\vec{F}, [Tf]))$. *The components of the gradient of* K_d *in these variables are:*

$$\frac{\partial K_d}{\partial [Tf]} = \frac{\mathscr{H}^2 + \mathscr{H}_{[Tf]}}{\mathscr{H}^2} = \frac{S^2}{\mathscr{H}^2} \quad and \quad \frac{\partial K_d}{\partial F_i} = \frac{-1}{(K_{di} + [Tf])\mathscr{H}^2}$$

where S^2 *is defined in (*10.3*) and is positive unless one of the* $F_i = 1$ *and all the others vanish.*

We write $s = s_r$, $[Tf] = [Tf]_r$, $[Tf]_{r+1} = [Tf] + \Delta[Tf]$, $\overrightarrow{F^{(r)}} = \vec{F}$, $\overrightarrow{F^{(r+1)}} = \vec{F} + \Delta\vec{F}$ *in order to calculate the total differential of* K_d. *We need formulas for* ΔF_i *and* $\Delta[Tf]$. *Recalling from (*6.4*) the definition of* E_i *there results:*

$$\Delta F_i = (E_i - 1)F_i = \frac{[Tf](1-\varepsilon)}{(\varepsilon K_d + [Tf])} \frac{(K_d - K_{di})}{(K_{di} + [Tf])} F_i.$$

We have

$$K_d(\vec{F} + \Delta\vec{F}, [Tf] + \Delta[Tf]) - K_d(\vec{F}, [Tf]) \approx \frac{\partial K_d(\vec{F}, [Tf])}{\partial [Tf]} \Delta[Tf] + \sum_{i=1}^{N} \frac{\partial K_d(\vec{F}, [Tf])}{\partial F_i} \Delta F_i$$

*Hence from equation (*11.15*) with m* = 1,

$$\begin{aligned}
&\mathscr{H}^2(\vec{F}, [Tf])\Delta K_d = S^2 \Delta[Tf] - \tfrac{[Tf](1-\varepsilon)}{\varepsilon K_d + [Tf]} \sum_{i=1}^{N} \tfrac{K_d - K_{di}}{(K_{di} + [Tf])^2} F_i, \\
&= -sS^2[Tf] - \tfrac{[Tf](1-\varepsilon)}{\varepsilon K_d + [Tf]} \sum_{i=1}^{N} \tfrac{(-[Tf] + 1/\mathscr{H}) - K_{di}}{(K_{di} + [Tf])^2} F_i, \\
&= -sS^2[Tf] - \tfrac{[Tf](1-\varepsilon)}{\varepsilon K_d + [Tf]} \left\{ \tfrac{1}{H(\vec{F}, [Tf])} \sum_{i=1}^{N} \tfrac{F_i}{(K_{di} + [Tf])^2} - \sum_{i=1}^{N} \tfrac{F_i}{(K_{di} + [Tf])} \right\},
\end{aligned}$$

Finally, returning to the index notation:

$$K^{(r+1)} - K^{(r)} = -[Tf]_r \left(\frac{S^{(r)}}{\mathscr{H}^{(r)}} \right)^2 \left(\frac{(1-\varepsilon)(K^{(r)} + [Tf]_r)}{\varepsilon K^{(r)} + [Tf]_r} + s_r \right). \tag{11.16}$$

Thus the terms of the sequence $\{K^{(r)}\}$ decreases to L. Therefore for sufficiently large r

$$K^{(r)} - K^{(r+m)} = L^2 [Tf]_r \sum_{l=1}^{m} \left(\frac{1}{\varepsilon} - (1 - s_{l+r}) \right) \left(\frac{S^{(r+l)}}{\mathscr{H}^{(r+l)}} \right)^2 \left[\prod_{j=1}^{l} (1 - s_{j+r-1}) \right]$$

Letting m + ∞

$$\frac{K^{(r)}-L}{[Tf]_r}=L^2\sum_{l=1}^{\infty}\left(\frac{1}{\varepsilon}-(1-s_{l+r})\right)\left(\frac{S^{(r+l)}}{\mathscr{H}^{(r+l)}}\right)^2\left[\prod_{j=1}^{l}(1-s_{j+r-1})\right]. \tag{11.17}$$

Thus we have an expression for $K_d(\overrightarrow{F^{(r)}},[Tf]_r)-L$ *in terms of* $[Tf]_r$*. The first coefficient on the right in* (11.17) *is bounded above by* 1/ε *and below by* (1 − ε)/ε*. Thus*

$$\sum_{l=r+1}^{\infty}\left(\frac{S^{(l)}}{\mathscr{H}^{(l)}}\right)^2\prod_{j=r}^{l-1}(1-s_j)=\sum_{l=r+1}^{\infty}\frac{\partial K^{(l)}}{\partial[Tf]}\prod_{j=r}^{l-1}(1-s_j)$$

must be convergent for all large indices r and hence (since $1-s_k\le 1$*) for every index r.*

The sequence of coefficients $\left\{S^{(l)}/\mathscr{H}^{(l)}\right\}$ *is convergent since* $\vec{F}^{(r)}\to c\rightarrow$*. Thus, if the series* $\sum_{r=1}^{\infty}\prod_{k=1}^{r}(1-s_k)$ *is divergent,* $S^{(l)}L\approx S^{(l)}/\mathscr{H}^{(l)}\to 0$*. That is, the divergence of the series forces the convergence of the iteration scheme to one of the vertices of* $\mathcal{T}$ *along one of the hyperbolic curves* C_i *defined in Section 10,* Appendix A.

The role of ε *on the absolute convergence of* (11.17) *is easily seen. When* s_r = 1/(r + 1) *and* ε = 1, *the series on the right in* (11.17) *will converge whether or not* $S^{(l)}/\mathscr{H}^{(l)}\to 0$*. (The reason is that the coefficients* $S^{(l)}/\mathscr{H}^{(l)}$ *are bounded above and the series* $\sum l=r^{\infty}\frac{1}{l(l+r)l}$ *is convergent.) As* ε *decreases from unity, the partial sums* $\sum_{l=r}^{m}[1/\varepsilon-(l+r-1)/(l+r)]1/l$ *increase. Thus the coefficients* $S^{(l)}/\mathscr{H}^{(l)}$ *in the partial sums of* (11.17) *decrease more rapidly and hence, for the entire series, decrease more rapidly to zero as* ε *decreases. From Theorem 3* c_1 = 1 *and* $L = K_{d1}$.

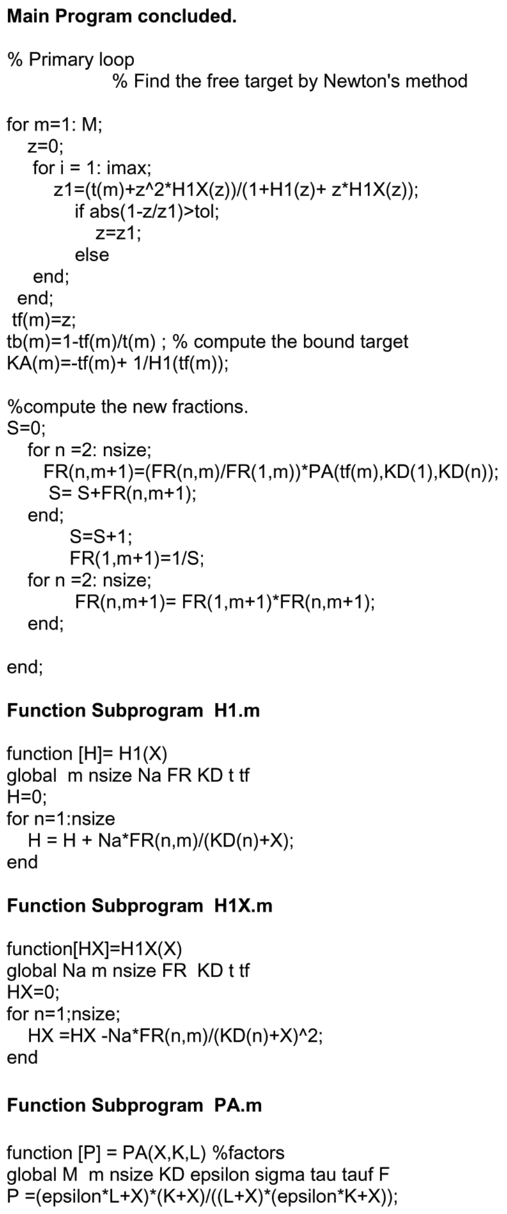
If the series $\sum_{r=1}^{\infty}\prod_{k=1}^{r}(1-s_k)$ *is convergent, nothing can be said about L or* S^2 *from* (11.17)*. This is to be expected since from Theorem 3, selection cannot occur.*

12. Appendix C. Matlab code

We include the programs we used here. Notice that only a single nonlinear equation is to be solved by Newton's method.

Main Program   selex.m

```
global  M m nsize  KD epsilon sigma t tf FR
% Passed function subprograms
%  H1(X)=[NA]sum(F_i/(Kd_i+X))
%  H1X= H1'(X)
%  PA=(epsilon*L+X)*(K+X)/((L+X)*(epsilon*K+X))
%
%
Na = 1;                    % concentration of total nucleic acid
M=  40;                    % number of SELEX rounds
nsize =15;                 % number of separate  NA species.
imax= 100;                 % number of iterations in Newton's method
KD = zeros(nsize,1);       %  vector of normalized dissassociation constants KD(i)=Kd_i
FR= ones(nsize,M);         % matrix of NA fractions after m rounds.
KA=zeros(M,1) ;            %  overall dissassociation constants
t=Na*ones(M,1) ;               % target inputs
tf=zeros(M,1) ;            %  free target values
tb=zeros(M,1) ;            %  bound target fractions
s = zeros(M,1) ;           % target reduction fractions
epsilon =   0.05,          % background partition (ratio bg/cp)
tol= .00001;               % degree of accuracy demanded for convergence of Newton's method.
% set the starting variables:
K=72/nsize;
S=0;
%Y=rand(size(ones(nsize)));       %random vector used for generating random
                                  %Kd's or random initial fractions.

for  n=1: nsize;                  % loop to generate Kd's
   KD(n,1)= Na*(1.6+K*2.2*(n-1))*10^(-4);  % based on Irvine et al p. 747
end

for  n=1: nsize;                   % Loop to generate initial fractions.  This can be done several
                                  ways.
           FR(n,1)=1;
end

 FR(nsize,1)=10^12;

for n=1:nsize;                       %normalize the first row.
  FR(n,1)=FR(n,1)/sum(FR(:,1));
end

for m = 1:M;
  s(m)=2/(2*m+1);
end
   for m = 2:M;
      t(m) =(1-s(m-1))*t(m-1);
   end;
```



Main Program concluded.

```
% Primary loop
                % Find the free target by Newton's method

for m=1: M;
   z=0;
    for i = 1: imax;
       z1=(t(m)+z^2*H1X(z))/(1+H1(z)+ z*H1X(z));
          if abs(1-z/z1)>tol;
             z=z1;
          else
    end;
  end;
 tf(m)=z;
tb(m)=1-tf(m)/t(m) ; % compute the bound target
KA(m)=-tf(m)+ 1/H1(tf(m));

%compute the new fractions.
S=0;
   for n =2: nsize;
     FR(n,m+1)=(FR(n,m)/FR(1,m))*PA(tf(m),KD(1),KD(n));
      S= S+FR(n,m+1);
   end;
         S=S+1;
         FR(1,m+1)=1/S;
   for n =2: nsize;
          FR(n,m+1)= FR(1,m+1)*FR(n,m+1);
   end;

end;
```

Function Subprogram H1.m

```
function [H]= H1(X)
global  m nsize Na FR KD t tf
H=0;
for n=1:nsize
   H = H + Na*FR(n,m)/(KD(n)+X);
end
```

Function Subprogram H1X.m

```
function[HX]=H1X(X)
global Na m nsize FR  KD t tf
HX=0;
for n=1;nsize;
   HX =HX -Na*FR(n,m)/(KD(n)+X)^2;
end
```

Function Subprogram PA.m

```
function [P] = PA(X,K,L) %factors
global M  m nsize KD epsilon sigma tau tauf F
P =(epsilon*L+X)*(K+X)/((L+X)*(epsilon*K+X));
```

13. Appendix D. A continuous analog of the SELEX iteration scheme

The mathematical and scientific literature abounds with examples of continuous time processes being modeled as the limit of discrete time processes as a time step is allowed to go to zero. Conversely, continuous processes are frequently approximated as discrete time processes.

In that spirit, we can think of the round number as a continuous parameter (time). Our goal is to determine the dynamical system of ordinary differential equations that corresponds to the selection process. We replace the discrete time notation $F_i^{(r)}$, s_r, $[T]_r$, $[Tf]_r$, $K_d^{(r)}$ by the continuous time notation $F_i(r)$, $s(r)$, $[T](r)$, $[Tf](r)$, $K_d(r)$ and convert differences to time derivatives by replacing" difference quotients" of the form $(F_i^{(r+1)} - F_i^{(r)})/1$ by

$(F_i^{(r+\Delta r)} - F_i^{(r)})/\Delta r$ and let $\Delta r \to 0$. Thus, we should expect to have, for the continuous dynamics, the following:

$$\begin{aligned}\frac{dF_i}{dr} &= (E_i(r) - 1)F_i(r)\\ \frac{dT}{dr} &= -s(r)[T](r)\end{aligned} \tag{13.18}$$

where

$$E_i(r) = \frac{\varepsilon K_{di} + [Tf](r)}{K_{di} + [Tf](r)} \frac{K_d(r) + [Tf](r)}{\varepsilon K_d(r) + [Tf](r)}, \quad K_d(r) = \frac{([T](r) - [Tf](r) + 1)[Tf](r)}{[T](r) - [Tf](r)}, \quad 1 = \sum_{i=1}^{N} F_i(r)$$

and where

$$\frac{1}{K_d(r) + [Tf](r)} = \sum_{i=1}^{N} \frac{F_i(r)}{K_{di} + [Tf](r)} = \frac{[T](r) - [Tf](r)}{[Tf](r)[NA]} = \mathscr{H}(r).$$

Then

$$\frac{F_i(r)}{F_1(r)} = \frac{F_i(1)}{F_1(1)} \exp\left(-\int_1^r [E_1(s) - E_i(s)]ds\right). \tag{13.19}$$

Because the disassociation constants are ordered, $[Tf](r) \le [T](r) \le [Tf](r)(1 + [NA]/K_{d1})$ and $[T](r) \le [T](1)$ and one can show that $L[Tf](r) \le E_1(r) - E_i(r) \le U[Tf](r)$ where L, U are constants given by

$$L = \frac{(K_{d2} - K_{d1})(1 - \varepsilon)}{(K_{d2} + [T](1))(K_{d1} + [T](1))} \text{ and } U = \frac{1 - \varepsilon}{\varepsilon} \frac{K_{dN} - K_{d1}}{K_{dN} K_{d1}}$$

From these simple inequalities and the fact that $[T](r) = [T](1)\exp\left(-\int_1^r s(\rho)d\rho\right)$ it follows immediately that $F_i(r) \to 0$ for $i \ge 2$ as $r \to +\infty$ if and only if

$$\int_1^\infty \exp\left(-\int_1^r s(\rho)d\rho\right) dr = +\infty. \tag{13.20}$$

Two cases obtain:

1. $\int_1^\infty s(\rho)d\rho < \infty$ and (13.20) holds. Then $K_d(r) + [Tf](r) \to K_{1d} + [Tf](r)$ so $K_d(r) \to K_{d1}$. Consequently, $[T](r) \to T_\infty > 0$ and $[Tf](r) \to [Tf]_\infty$ where $K_{d1}(T_\infty - [Tf]_\infty) = [Tf]_\infty (1 + T_\infty - [Tf]_\infty)$, a quadratic easily solved for $T_\infty > 0$. In this case
$$\lim_{r \to +\infty} \frac{[T](r) - [Tf](r)}{[T](r)} = \frac{[NA]}{[Tf]_\infty + K_{d1} + [NA]} < \frac{[NA]}{K_{d1} + [NA]},$$
i.e. maximum bound target efficiency is not obtained.

2. $\int_1^\infty s(\rho)d\rho = +\infty$. In this case we still must require that (13.20) holds. Then $T_\infty = 0 = [Tf]_\infty$ and
$$\lim_{r \to +\infty} \frac{[T](r) - [Tf](r)}{[T](r)} = \frac{[NA]}{K_{d1} + [NA]},$$
i.e. maximum bound target efficiency is obtained.

If we take $s(r) = s_0/r^2$ where $s_0 \in (0, 1)$ then $\int_1^\infty s(r)dr = s_0 < 1$ and we have the first case. If we take $s(r) = 1/r$ we are in the second case. In both cases, (13.20) holds. Notice that when $s(r) = s_0$ where $s_0 \in (0, 1)$ the result says that selection cannot occur.

Finally, a calculation shows that

$$\frac{dK_d(r)}{dr} = -[Tf](r)[K_d(r) + [Tf](r)]^2(S(r))^2\left(\frac{(1-\varepsilon)(K_d(r) + [Tf](r))}{\varepsilon K_d(r) + [Tf](r)} + s(r)\right)$$

where

$$S^2(r) = \left[\sum_{i=1}^{N}\frac{F_i(r)}{(K_{di} + [Tf](r))^2} - \left(\sum_{i=1}^{N}\frac{F_i(r)}{K_{di} + [Tf](r)}\right)^2\right].$$

This tells us that $K_d'(r)/[Tf](r) \to 0$ if and only if selection occurs.

Finally, after a little algebra we find

$$\frac{F_1(1)}{[1 - F_1(1)]\exp\left(-L\int_1^r[Tf](s)ds\right) + F_1(1)} \le F_1(r) \le \frac{F_1(1)}{[1 - F_1(1)]\exp\left(-U\int_1^r[Tf](s)ds\right) + F_1(1)}.$$

From these inequalities it is possible to get upper and lower bounds on how large r must be in order that $F_1(r)$ reach a fixed fraction. Notice that as ε increases to unity, these upper and lower bounds on r must recede to infinity as L, $U \to 0$ with $\varepsilon \uparrow 1$.

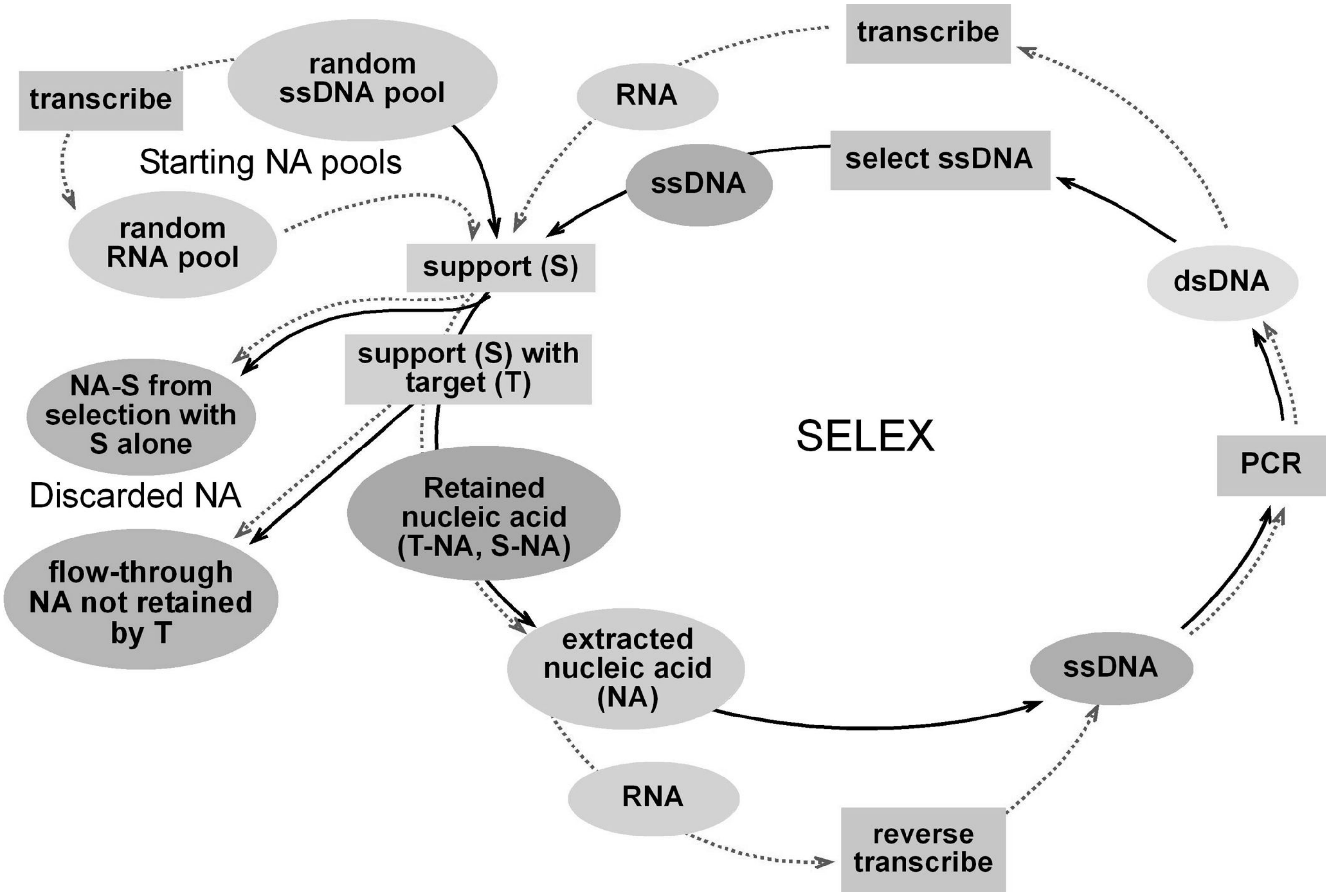

transcribe
random ssDNA pool
Starting NA pools
random RNA pool
support (S)
RNA
ssDNA
select ssDNA
transcribe
dsDNA
support (S) with target (T)
NA-S from selection with S alone
Discarded NA
flow-through NA not retained by T
Retained nucleic acid (T-NA, S-NA)
SELEX
PCR
extracted nucleic acid (NA)
ssDNA
RNA
reverse transcribe

Figure 1.
The steps of SELEX are demonstrated in this figure. Starting in the top left corner of the figure, the blue and pink ovals represent the initial NA pools. SELEX can be done for RNA or single stranded DNA (ssDNA) molecules. Both protocols are represented here. The RNA selection protocols can be followed by the red dashed arrows and the ssDNA protocols by the black arrows. The square yellow selection step [support (S) with or without target (T)] is used to select the S-NA complex in combination with or without the T-NA complex. The S-NA complex, selected in the absence of T is discarded as is the NA that flows through the support-target combination. Retained extracted NA is taken through a SELEX round that includes the PCR amplification step and that generates the next NA pool, which is again selected against the support and or support plus target. SELEX protocols can vary greatly, depending on the desired characteristics of the selected aptamer. Not all rounds of SELEX include an initial selection against support, although this is a recommended practice (Pollard et al. (2000)).

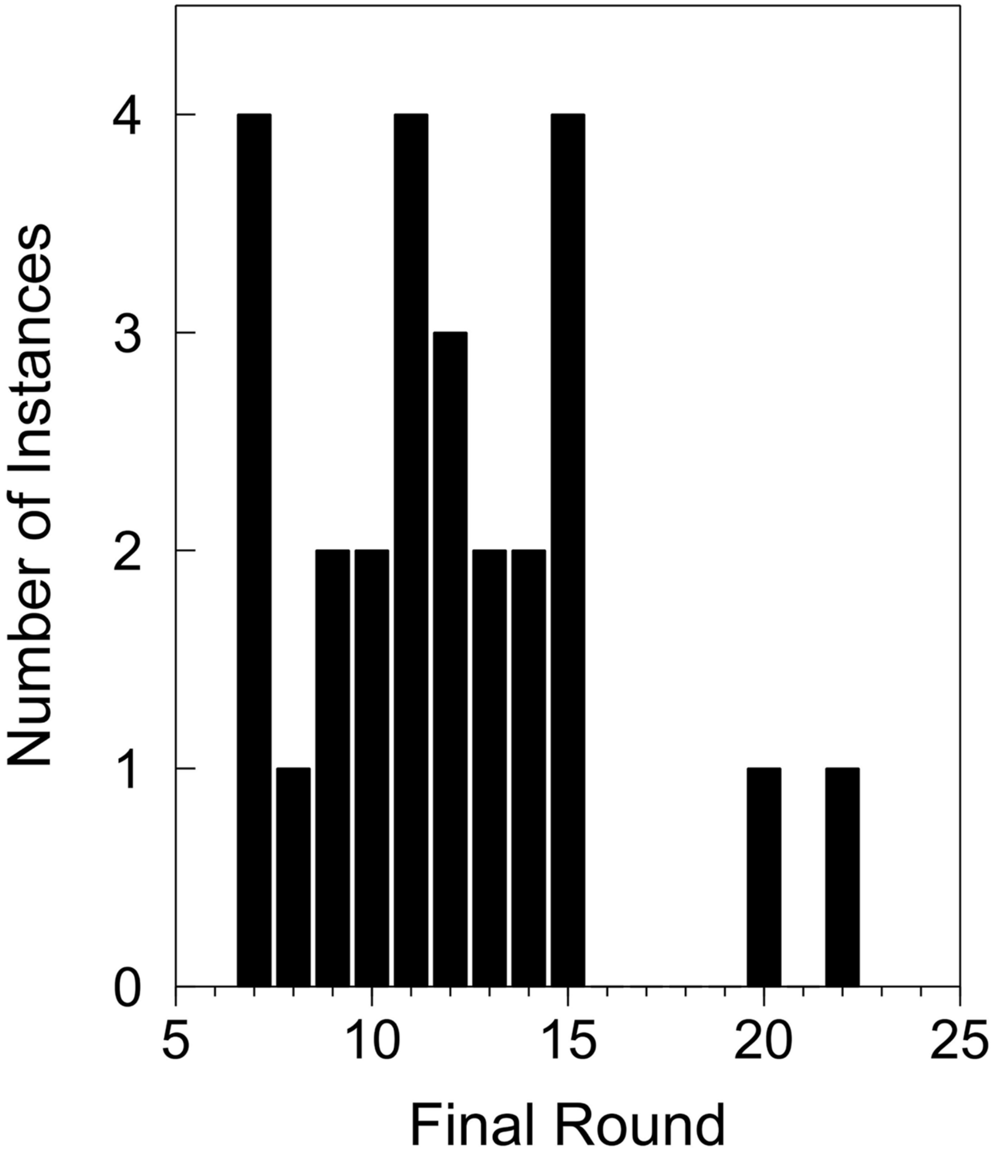


Figure 2.

Survey of Rounds to Completion in SELEX Experiments. Plotted is a summary of 26 publications from 2003 to mid-2006 in which SELEX experiments were reported that resulted in the cloning of one of more aptamers. The number of rounds performed before the aptamers were cloned was determined for each instance and the number of instances is plotted against the number of rounds prior to cloning. The number of rounds prior to cloning varied from 7 to 22 with a mean of 12 ± 4 and a median of 12.

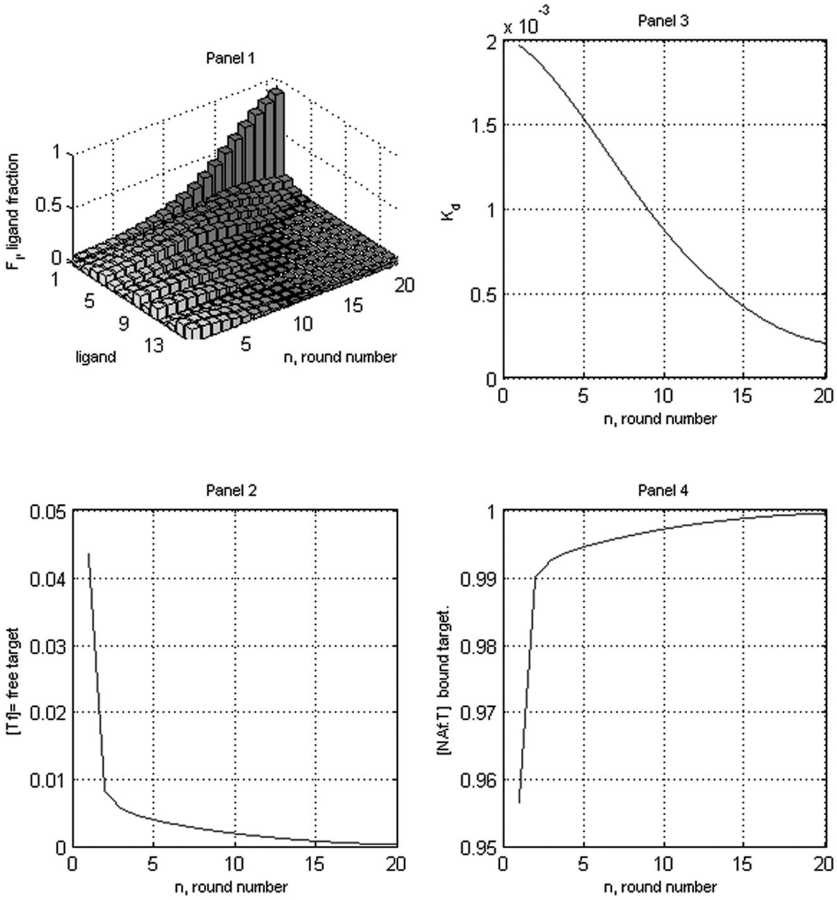


Figure 3.

The decrease in target concentration from round to round is very slow but nevertheless, selection is occurring, nearly all but the first and second nucleic acids being essentially gone after 10 rounds. In panels 2, 4 the plots begin at round numbers 2 and 3. This was done for convenience of scale. In particular, in Panel 4, we see that the maximum target efficiency is 1/(1 + K_{d1} + [Tf](20)) < 1/(1 + K_{d1}) at twenty rounds.

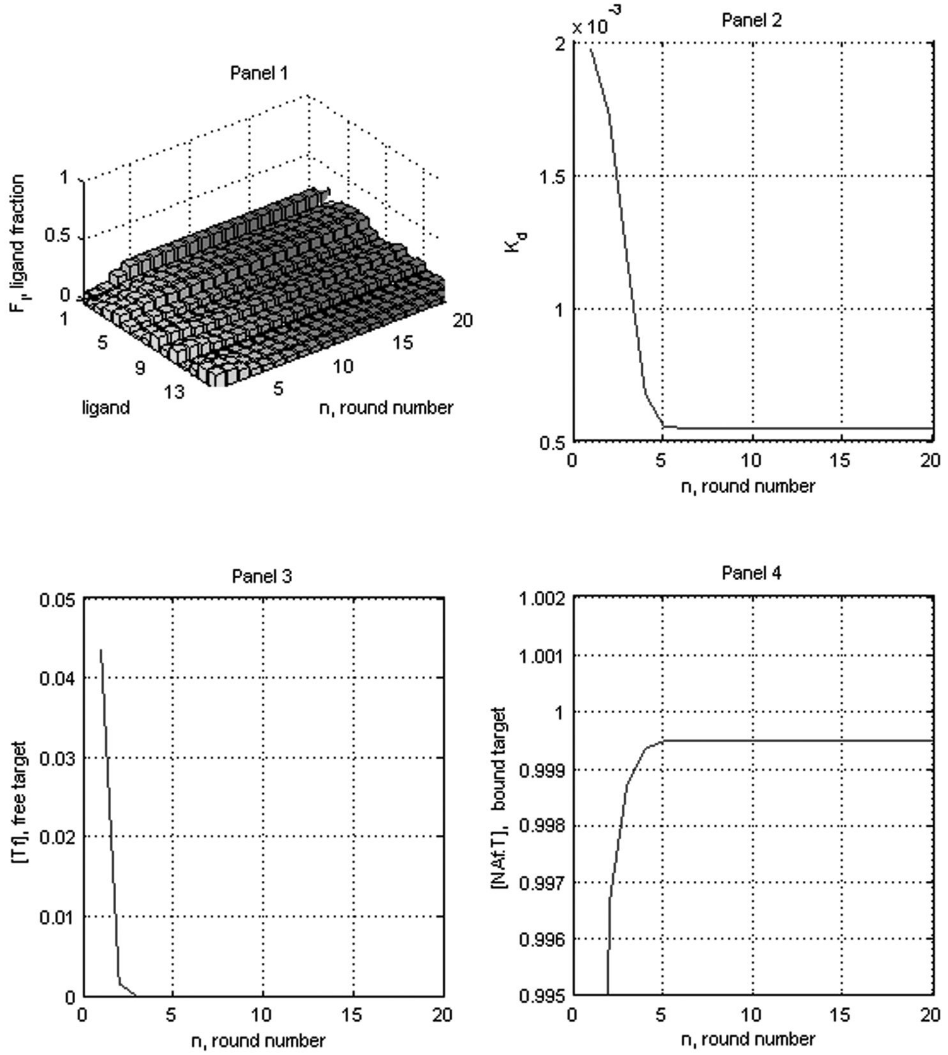


Figure 4.

The decrease in target concentration from round to round is such that the series in (7.2) is convergent. Clearly selection is not taking place. Notice the scale on the vertical axis in Panel 1. Notice also that almost all of the free target is used up after four rounds. Panel 2 (incorrectly) suggests that we have achieved selection as the overall (rescaled) dissociation constant has fallen to 0.6(10^{-3}). If we didn't have other information we might be inclined to conclude that this value is K_{d1}. In fact, for this experiment, K_{d1} = 1.6(10^{-4}). Although all of the free target is exhausted, the maximum bound target fraction is not unity, but rather 0.9995 ≈ 1/(1.00006), a number smaller than the maximum target efficiency, $1/(1 + K_{d1})$.

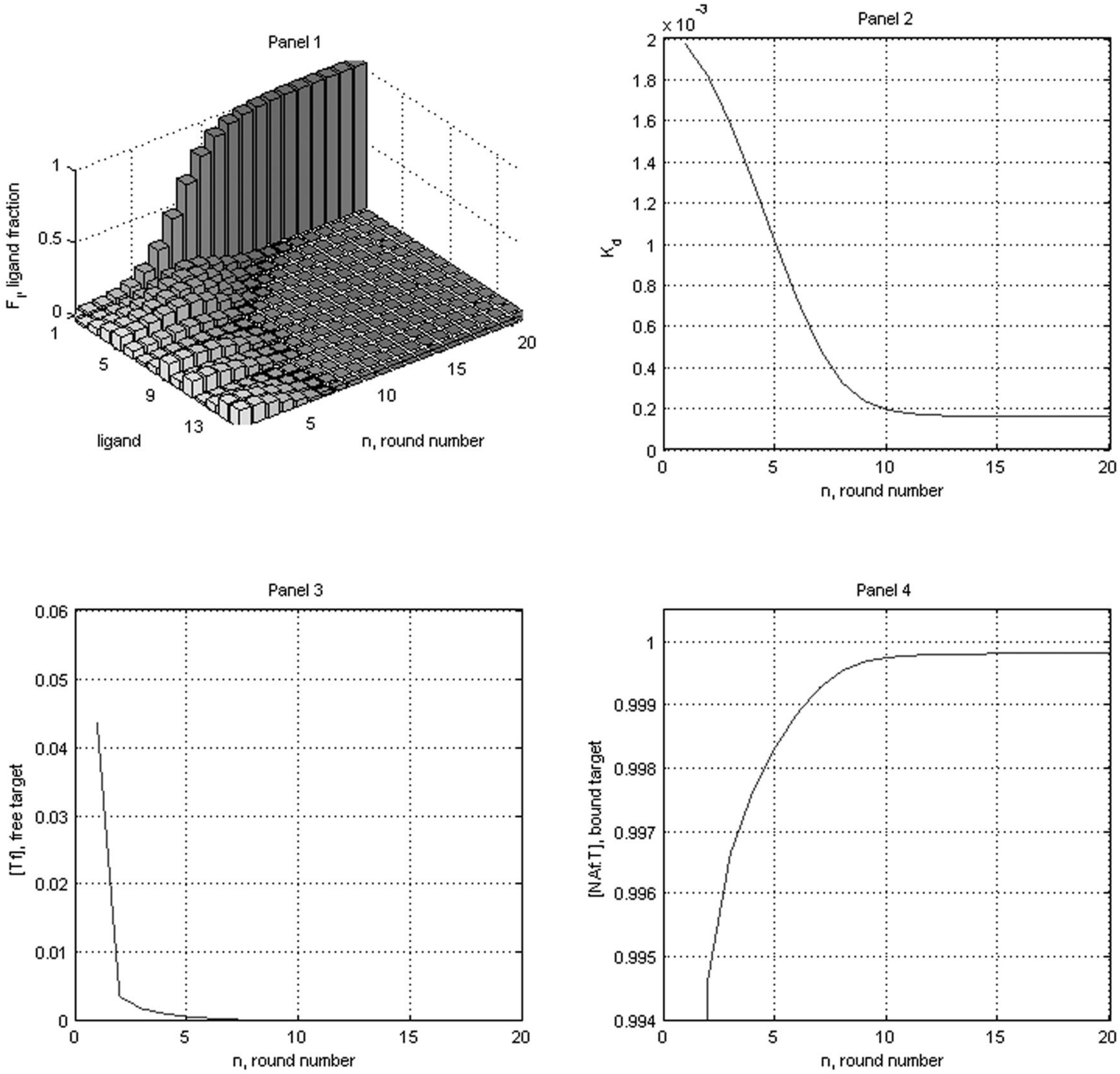


Figure 5.

The decrease in target concentration from round to round is such that the series in (7.2) is divergent. Almost all of the target is gone after seven rounds and that only the best binding nucleic acid remains in the pool after eight or nine rounds. Now we see from Panel 4, that the maximum target efficiency (≈ 0.9998) has been attained.

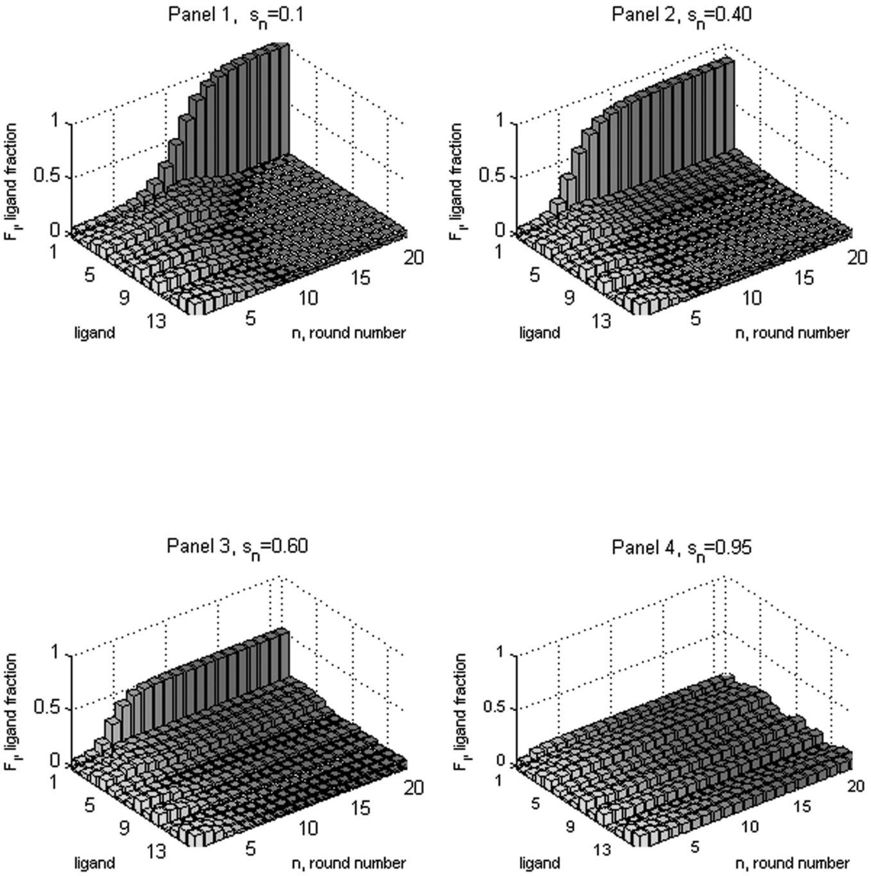


Figure 6.

The results of uniform reduction from round to round. Selection becomes harder to achieve if we reduce the starting target from round to round too quickly.

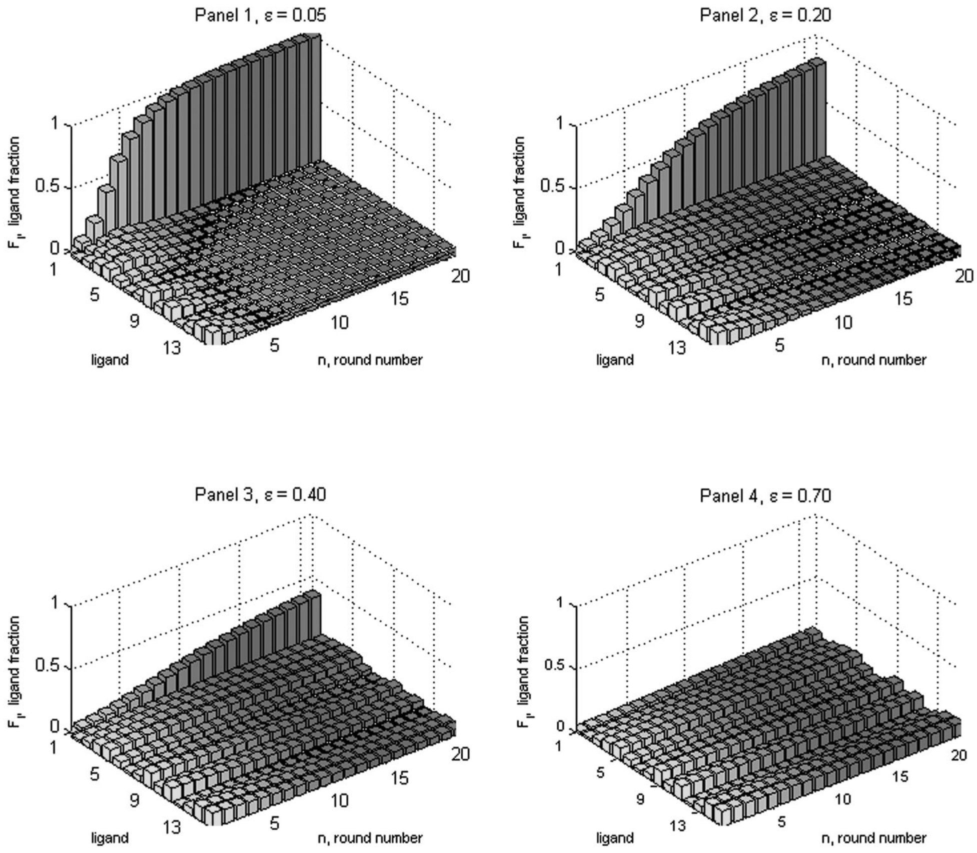


Figure 7.

The effects of partitioning (losses). As the loss fraction (*ε*), increases from 0 to 1, it becomes harder to achieve selection.

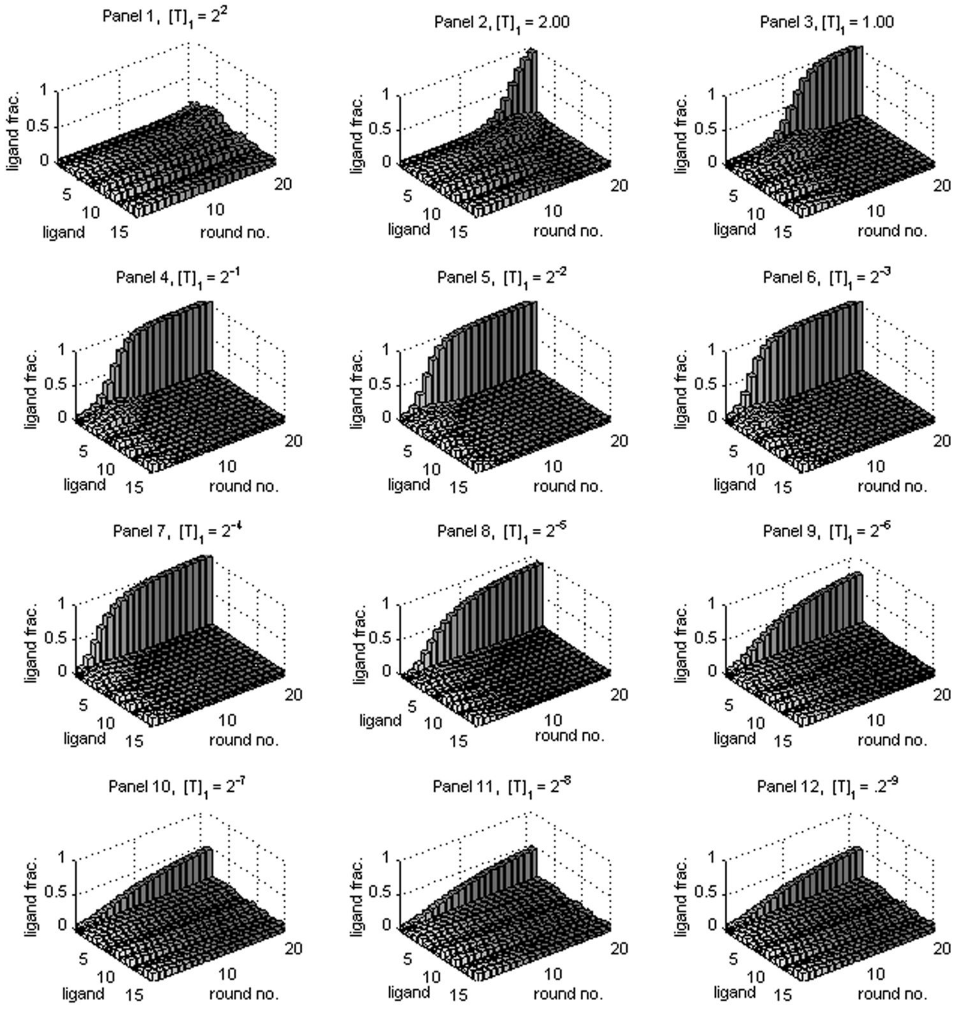


Figure 8.

As the initial target is decreased progressively from panel 1 through panel 6, selection takes fewer rounds to achieve. Further increases in the initial target result in increases in the round number (the number of rounds required to reach a fixed percentage of ligand 1), begins to increase. *This illustrates the point that simply increasing target over the concentration of the initial pool or else reducing it considerably will not necessarily decrease round number.*

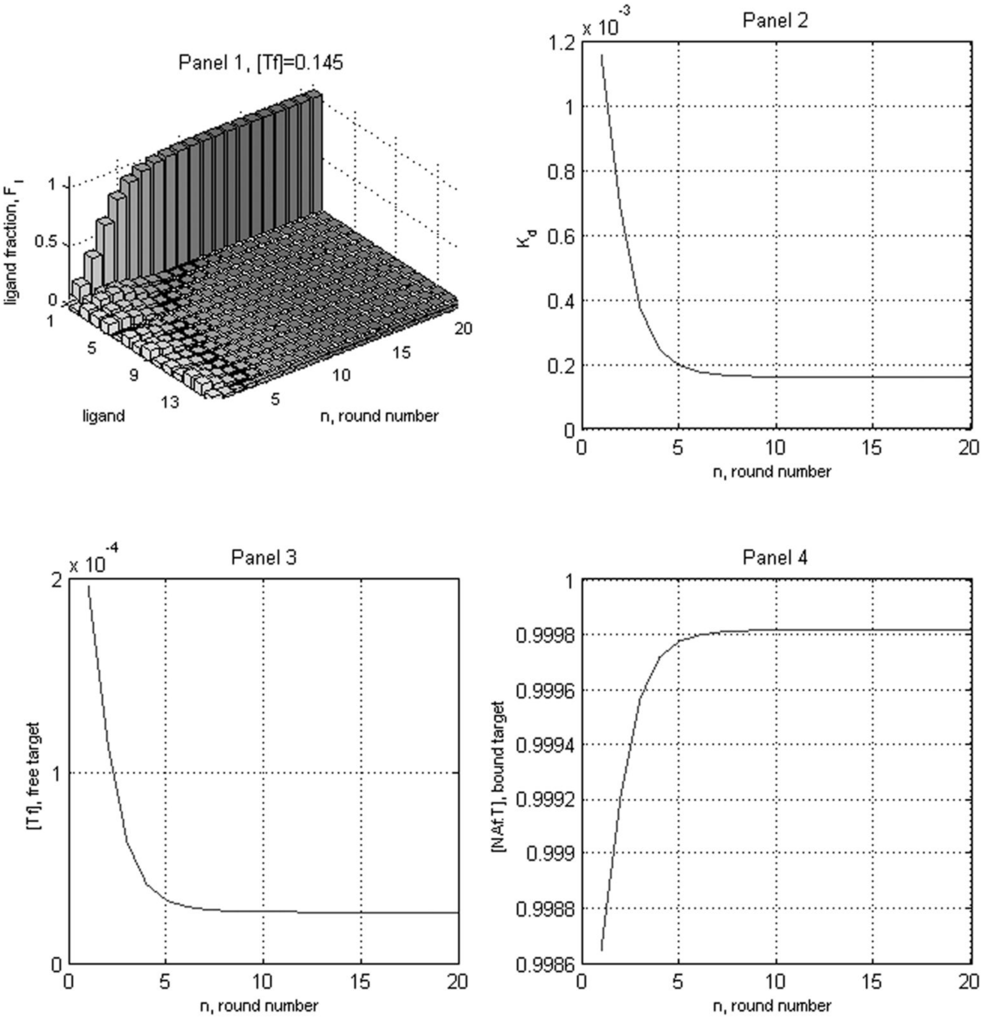


Figure 9.

We use formula $[Tf] = \sqrt{\delta K_{d1} K_d(0)}$ for the initial target in every round. That is, $s_r = 0$. The initial pool is again random and $\varepsilon = 0.05$. The first panel demonstrates improved selection over all the panels in Figure 7.

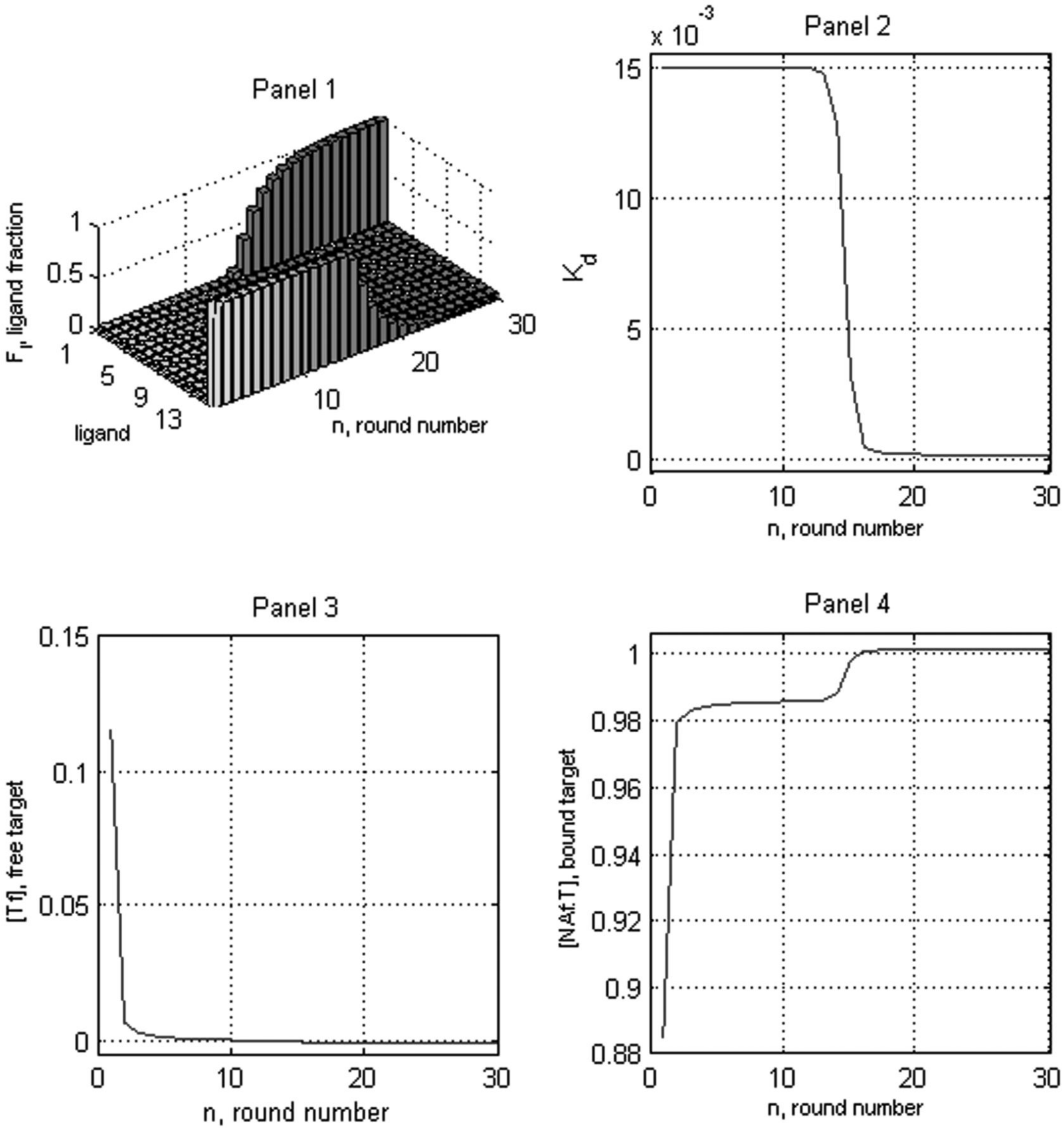


Figure 10.

A plot of the best and poorest binding fractions as a function of round number with only one molecule of each nucleic acid present except the poorest binder and there = 10^{-2} molecules of it. There are fifteen nucleic acid types. Notice the unusual kink in the graph in Panel 2. It occurs at about the value of the round number for which the pool size is roughly evenly divided.

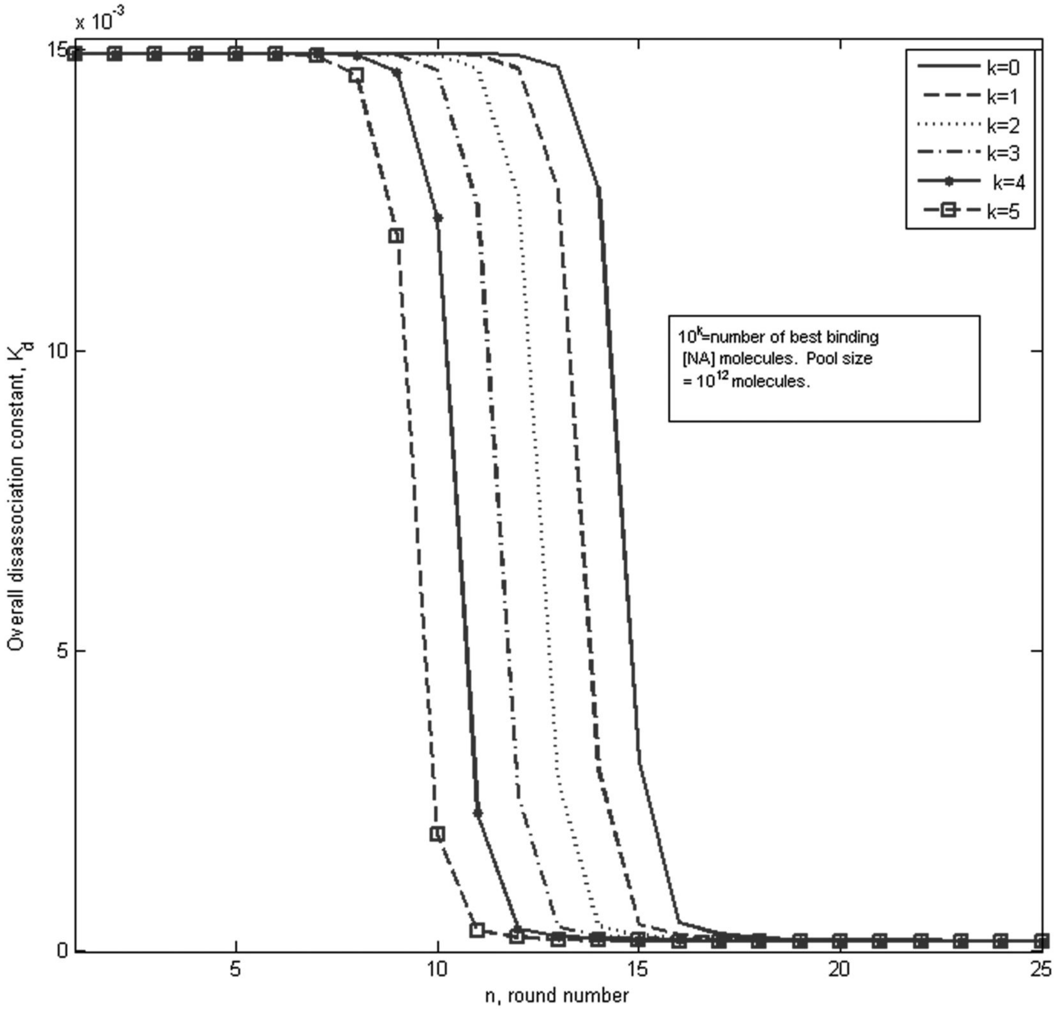


Figure 11.

A plot of the overall dissociation constant as a function of round number for six different initial fractions of best binding nucleic acid. 10^k=number of best binding [NA] molecules in a pool of = 10^{12} molecules. There are fifteen nucleic acid types.

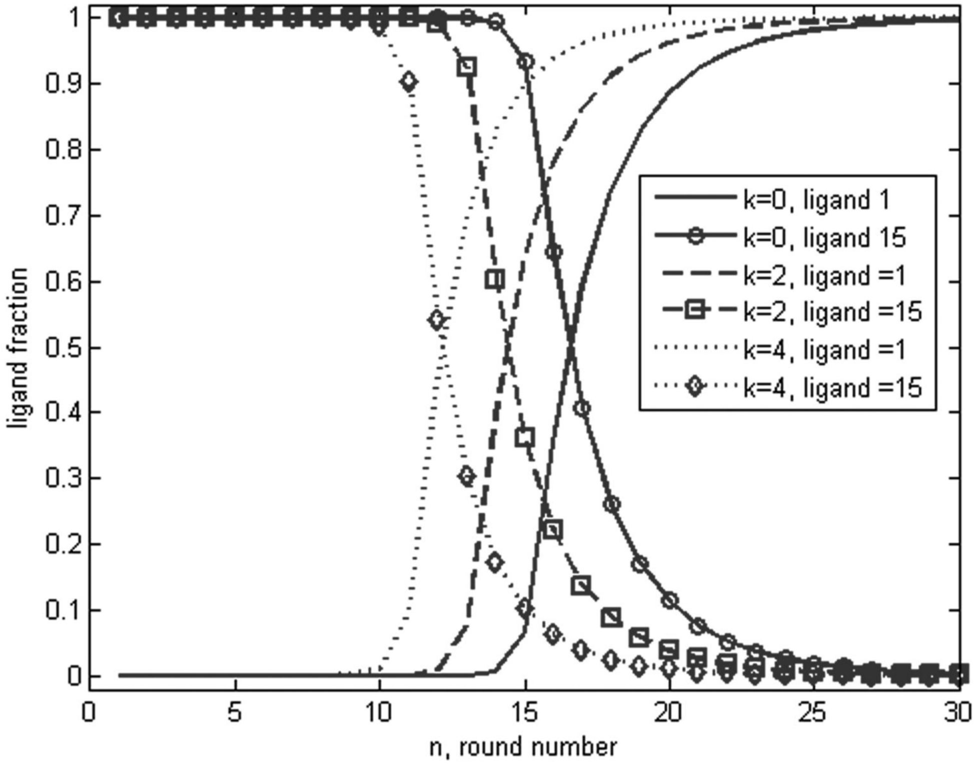


Figure 12.

A plot of the best and poorest binding fractions as a function of round number for three different initial fractions of best binding nucleic acid. Here 10^k=number of best binding [NA] molecules in a pool of = 10^{12} molecules. There are fifteen nucleic acid types. Clearly the round number at which the nucleic acid fractions of the best and worst binders are each 1/2 of the pool falls with in the range predicted by the inequalities in (8.10)

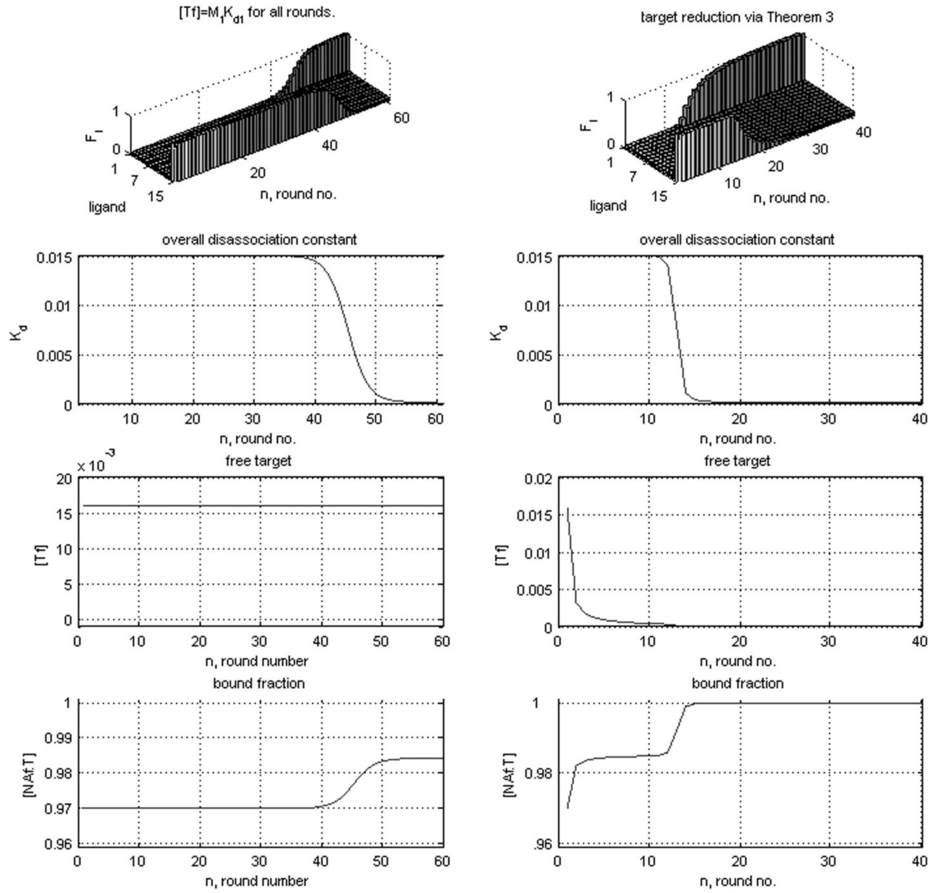


Figure 13.

In this set of figures the starting value for the target is taken as the starting value of the target dictated by demanding a probability of 0.99 for one molecule of the best binding nucleic acid to bind in order to generate the starting target value as dictated by (Irvine et al. 1991). The ratio $K_{dN}/K_{d1} \approx 100$ was used for these figures. Again, we need to interpret the bound target graphs carefully. The maximum value in the bottom panel in the first column here is clearly smaller than unity, as it should be. In the bottom panel in the second column, it appears to reach unity, but is in fact, smaller than unity, being approximately $1/(1 + K_{d1})$ as the free target is nearly zero near the last few rounds.

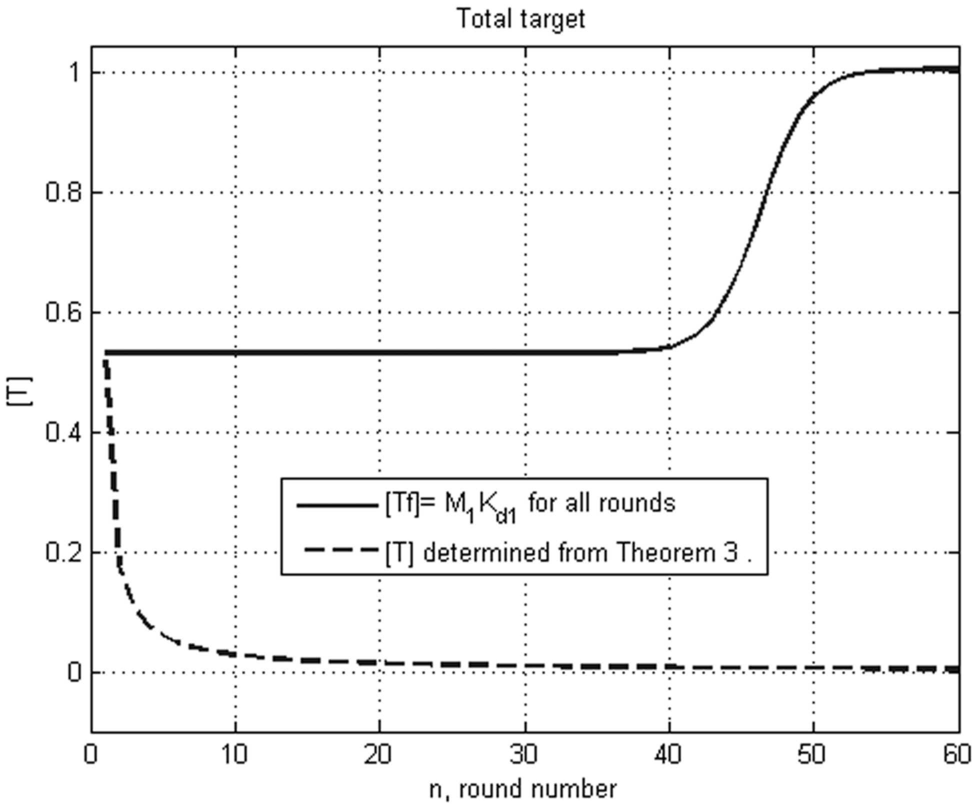


Figure 14.
In this figure we plot the total target as a function of round number for the two cases illustrated in Figure 13.

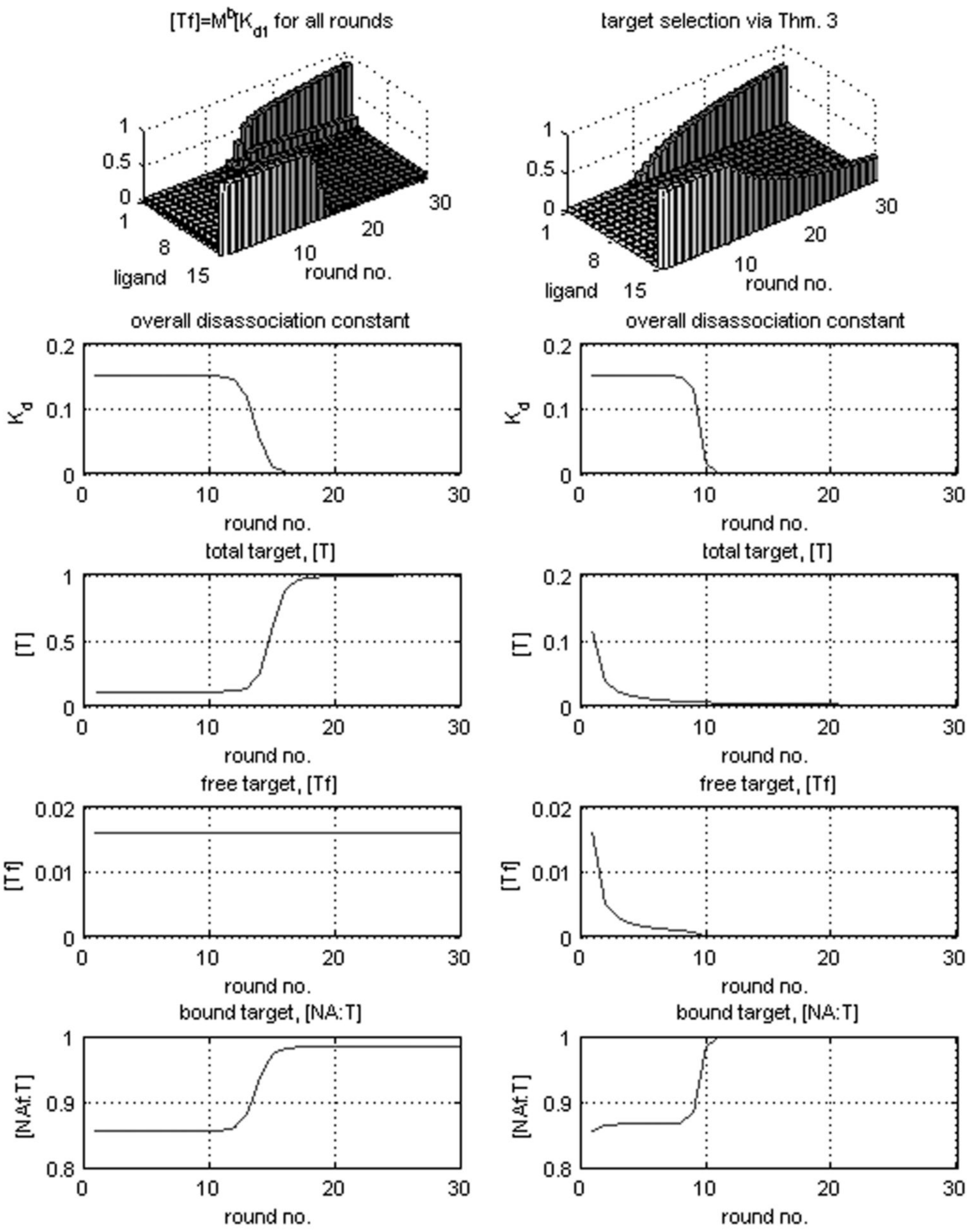


Figure 15.

The effect of using the small starting value of the target dictated by demanding a probability of 0.99 for one molecule of the best binding nucleic acid to bind in order to generate the starting target using Theorem 3. To generate this figure the ratio $K_{dN} / K_{d1} \approx 10^4$ was used. The same comments concerning the graphs of $[T]_b$ in the figure caption for Figure 13 apply here also.

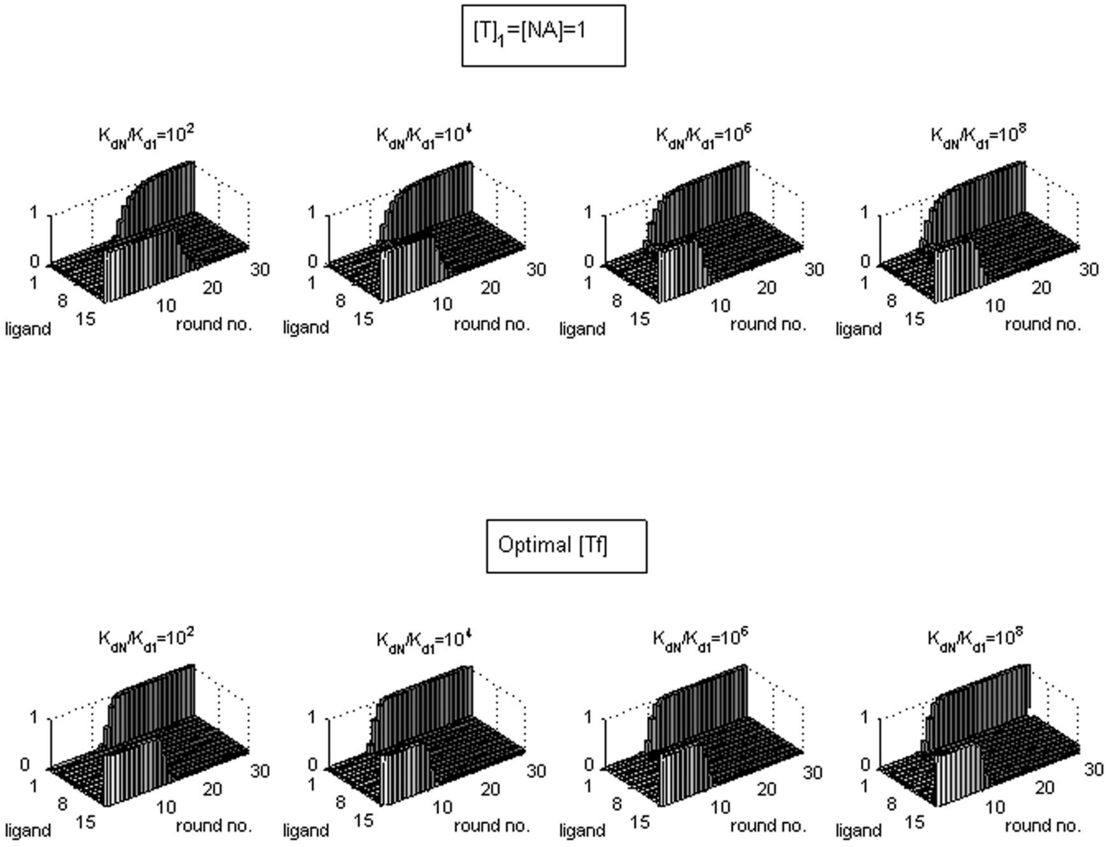


Figure 16.

The top row of figures illustrates the effect of increasing the ratio K_{dN} / K_{d1} on the round number at which selection becomes significant. The the round number for which 50% selection is achieved decreases from 14 to around 8 over six orders of magnitude. The input target at subsequent rounds was dictated by Theorem 3. The bottom row of figures was generated by using the solution $[Tf] = \sqrt{\varepsilon K_{d1} K_d([Tf])}$ of to generate the free target at each round.

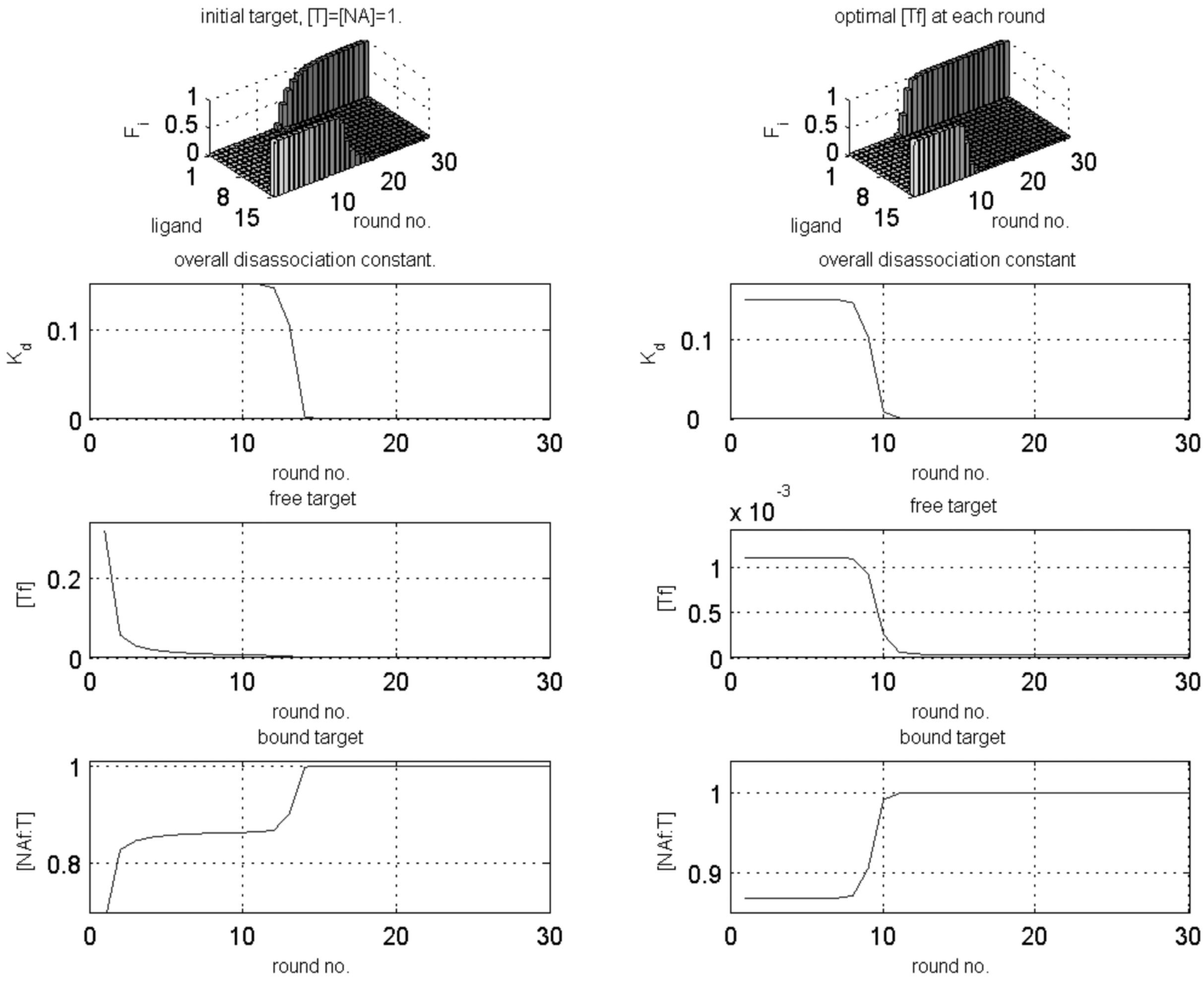


Figure 17.

Here all the relevant plots are given for the case $K_{dN} / K_{d1} = 10^3$, a case not included in Figure 16. The same comments concerning the graphs of $[T]_b$ in the figure caption for Figure 13 apply here also.

Table 1

Notation and problem formulation for a single SELEX round

We extend the notation of Irvine et al. (1991) to permit a more general discussion. Thus the protein (P) is replaced by a target (T) and RNA by NA (nucleic acid).

species	quantity (See (Irvine et al. 1991).)
starting target	$[T]$
starting NA_i	$[NA_i]$
starting NA	$[NA]$
free NA_i	$[NAf_i]$
free NA	$[NAf]$
bound NA_i	$[\{T{:}NA_i\}]$
free target	$[Tf]$
bound NA (max.avail for PCR)	$[T{:}NA]$