Bulletin of Mathematical Biology (2010) 72: 1623–1665
DOI 10.1007/s11538-009-9491-x

ORIGINAL ARTICLE

A Mathematical Analysis of Multiple-Target Selex

Yeon-Jung Seo[a], Shiliang Chen[b], Marit Nilsen-Hamilton[c], Howard A. Levine[a,*]

[a] *Department of Mathematics, Iowa State University, Ames, IA 50011, USA*
[b] *Department of Mathematics, Princeton University, Princeton, NJ 08544, USA*
[c] *Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA*

Received: 5 May 2009 / Accepted: 24 November 2009 / Published online: 14 January 2010

Abstract SELEX (Systematic Evolution of Ligands by Exponential Enrichment) is a procedure by which a mixture of nucleic acids can be fractionated with the goal of identifying those with specific biochemical activities.

One combines the mixture with a specific target molecule and then separates the target-NA complex from the resulting reactions. The target-NA complex is separated from the unbound NA by mechanical means (such as by filtration), the NA is eluted from the complex, amplified by PCR (polymerase chain reaction), and the process repeated. After several rounds, one should be left with the nucleic acids that best bind to the target. The problem was first formulated mathematically in Irvine et al. (J. Mol. Biol. 222:739–761, 1991). In Levine and Nilsen-Hamilton (Comput. Biol. Chem. 31:11–25, 2007), a mathematical analysis of the process was given.

In Vant-Hull et al. (J. Mol. Biol. 278:579–597, 1998), multiple target SELEX was considered. It was assumed that each target has a single nucleic acid binding site that permits occupation by no more than one nucleic acid. Here, we revisit Vant-Hull et al. (J. Mol. Biol. 278:579–597, 1998) using the same assumptions. The iteration scheme is shown to be convergent and a simplified algorithm is given. Our interest here is in the behavior of the multiple target SELEX process as a discrete "time" dynamical system. Our goal is to characterize the limiting states and their dependence on the initial distribution of nucleic acid and target fraction components. (In multiple target SELEX, we vary the target component fractions, but not their concentrations, as fixed and the initial pool of nucleic acids as a variable starting condition.)

Given N nucleic acids and a target consisting of M subtarget component species, there is an $M \times N$ matrix of affinities, the (i, j) entry corresponding to the affinity of the jth

*Corresponding author.
E-mail addresses: syeonj@iastate.edu (Yeon-Jung Seo), stc2118@columbia.edu (Shiliang Chen), marit@iastate.edu (Marit Nilsen-Hamilton), halevine@iastate.edu (Howard A. Levine).
In honor of Hans Weinberger on the occasion of his 80th birthday.
The research of S. Chen was supported by an NSF grant # DMS-0353880 and that of H.A. Levine, Y.-J. Seo, and M. Nilsen-Hamilton was supported by NIH grant #R43 CA-110222.

nucleic acid for the ith subtarget. We give a structure condition on this matrix that is equivalent to the following statement: *For any initial pool of nucleic acids such that all N species are represented, the dynamical system defined by the multiple target SELEX process will converge to a unique subset of nucleic acids, each of whose concentrations depend only upon the total nucleic acid concentration, the initial fractional target distribution (both of which are assumed to be the same from round to round), and the overall limiting association constant. (The overall association constant is the equilibrium constant for the system of MN reactions when viewed as a composite single reaction.) This condition is equivalent to the statement that every member of a certain family of chemical potentials at infinite target dilution can have at most one critical point.* (The condition replaces the statement for single target SELEX that the dynamical system generated via the process always converges to a pool that contains only the nucleic acid that binds best to the target.) This suggests that the effectiveness of multiple target SELEX as a separation procedure may not be as useful as single target SELEX unless the thermodynamic properties of these chemical potentials are well understood.

Keywords SELEX · Chemical potential · Fractionation · Discrete dynamical system · Asymptotic stability

1. Introduction

1.1. Biochemical background

The alternatives to antibodies for selective and high affinity recognition are peptide or nucleic acid aptamers. Aptamers are selected for their binding properties from a "library" of peptides (in the context of a scaffold protein) or of nucleic acids in which each constituent molecule contains a region of its sequence that has been created by the insertion of randomly chosen precursors. For almost all peptide and nucleic acid libraries used for selection, the number of constituent molecules is far fewer than the number of possible unique sequences. Barring multiple sequences created as a result of the probabilistic nature of the processes used to synthesize the libraries, each constituent molecule in these libraries is unique. For both types of aptamer, the individual molecules with high affinity and selectivity are captured from the library by a series of alternating selection and amplification steps in which a subset of molecules are first selected for a binding property to a target, then the selected population is expanded (amplified) to increase the number of copies of each selected molecule. The expanded pool is then again used as the basis for selection of a subset of binders and the rounds of selection and expansion continued. Typically, after about twelve selection rounds, the remaining molecules are evaluated for the existence in the pool of an expanded representation of high affinity binders.

With the ability to detect and isolate molecules with important roles in biological systems comes increased knowledge of these molecules in their in vivo context, which contributes to a fundamental understanding of the mechanisms by which these molecules perform their biological role. Consequently, the development of probes such as antibodies and aptamers that recognize specific molecules is important for the further development of biological investigations. Most aptamer selections today are performed with purified protein targets. Although these preparations contain a single molecular species,

each molecule presents more than one target for aptamer binding. Using the immunological nomenclature, these targets would be referred to as epitopes. Probes are also needed for many proteins that have not yet been adequately purified. Thus, it is anticipated that experimentalists will more frequently utilize mixed targets (e.g., impure proteins or intact cells or organelles) for selection. To optimize these selection protocols, it will be important to understand the nature of the selection process, which is the purpose of the current work that considers nucleic acid aptamers, commonly referred to as aptamers.

1.2. Mathematical overview

Because the selection process is iterative, we view multiple target SELEX as a discrete dynamical system. As in Vant-Hull et al. (1998), we employ a mean field (deterministic) model. That is, we assume all species are present in sufficient quantities to invoke the law of large numbers and use average values for the species concentrations. We do not impose any distribution rule on the affinity matrix, taking its entries to be randomly generated numbers within a specified physically meaningful interval. Likewise, we take the initial pool of nucleic acid fractions to be random positive numbers summing to unity. A number of papers and the varied approaches to this subject are worth mentioning. Complex-target SELEX models were considered in Chen (2007), Chen and Kuo (2007), Chen et al. (2007), Vant-Hull et al. (1998) based on mean field theory. In Vant-Hull et al. (1998), the authors generalized the complex-target model from the single-target model in Irvine et al. (1991). There they introduced the method of ligand (nucleic acid) subpools with similar affinities for each target. Within each subpool, it was assumed that the target affinities satisfied a log Gaussian distribution for the ligand affinities. They organized the nucleic acid pool and the subtarget proteins as follows (using our notation below). If we have N nucleic acids with concentrations $[NA_j]$ with indices j in the set $\mathcal{N} = \{1, 2, \ldots, N\}$, we call this set the full nucleic acid pool. Any ordered subset of $\mathcal{N}$ with the same nucleic acid concentrations for the indices in this subset is called a subpool of the given pool $\mathcal{N}$. Now consider a target (vector) composed of proteins $T_1, \ldots, T_M$. Consider for subtarget T_1, the subpool S_{11} of nucleic acids that bind best to this target and call its affinity $a_{1,1}$. Let S_{21} be the subpool of nucleic acids that bind second best to T_1 with affinity $a_{2,1} < a_{1,1}$. Clearly S_{11} and S_{21} are disjoint. Eventually, we arrive at a subpool $S_{n_1,1}$ with poorest affinity $a_{n_1,1} < a_{n_1-1,1} < \cdots < a_{2,1} < a_{1,1}$. Repeating this process for the remaining $M-1$ targets, we arrive at a collection of subpools $C_i = \{S_{j_i,i} |\ j_i = 1, \ldots, n_i\}$ for each subtarget T_i, $1 \le i \le M$. The subpools in each collection are pairwise disjoint. The subpools of interest in Vant-Hull et al. (1998) are defined as follows: For each M-tuple of indices $(j_1, \ldots, j_M)$ with $1 \le j_i \le n_i$, let $S_{(j_1 j_2 \cdots j_M)} = \bigcap_{i=1}^{M} S_{j_i,i}$. The meaning of the definition is that a nucleic acid in one of these subpools is the j_1st best binder to T_1, the j_2th best binder to T_2, etc. There are $n_1 n_2 \cdots n_M$ such subpools. The number of these subpools can be quite large as was remarked in Vant-Hull et al. (1998). In our example below, $n_i = N = 20$ and $M = 5$ so that the number of subpools is $20^5 = 3.2(10^6)$, far more than the number of nucleic acid species. A more typical but somewhat less dramatic illustration was given in Vant-Hull et al. (1998), i.e., with 16 proteins and 10 binding constants per protein, the number of such pools would be 10^{16} but a typical value for $N \approx 10^{15}$.[1] Although, as is remarked

[1]The mathematical meaning of this is that some of the sets $S_{j_1 j_2 \cdots j_M}$ must be empty. For example, if the pool of nucleic acids is represented as $\{a_1, a_2, a_3, a_4\}$ and the subpools of nucleic acids for the first protein

in Vant-Hull et al. (1998), for reasons cited there, this can be viewed as "merely a mathematical technicality," from the point of view of the computational scientist, it can cause programming headaches and strains on computing resources (memory and running time), especially as the addition of a single target increases the number of variables by a factor of N. (This was noted also in Chen, 2007, p. 199 and in Chen and Kuo, 2007, p. 1,017.)

In view of these and related issues, the authors in Chen and Kuo (2007), Chen et al. (2007) developed a condensed subpooling model for complex-target SELEX as well as for subtractive SELEX (Vant-Hull et al., 1998, p. 594) to optimally reduce the size of ligand subpools considered in Vant-Hull et al. (1998). Stochastic simulations of ligand evolution were carried out in Chen (2007) to characterize the evolution dynamics under the influence of random effects such as point mutations. (In Chen, 2007, the author modeled the binding and selection probability for complex-target SELEX. Conditions and evolution trajectories of ligands were examined for the aptamer enrichment and ligand dynamics of complex SELEX.) In Chen (2007, p. 198), the author also argued that "Missing aptamers for some targets in the fully enriched library is conceivable for real experiments, but in theory it will never be predicted by mean-field model-based simulations in Chen and Kuo (2007), Vant-Hull et al. (1998)."

Our primary goal is to characterize the limiting values of the nucleic acid concentrations in terms of the initial target components. Although an experimenter would eschew SELEX procedures that involve more than 20 rounds (iterations), the computational scientist is not limited by this restriction, but only by the computational power of the computer employed. Nevertheless, dealing with 10^{16} subpools as variables can strain the resources of a desktop computer. It was shown in Sun et al. (1996), for single target SELEX, that for successful SELEX experiments, the number of rounds of SELEX cycles should be closely tied to the concentration of the target. In Levine et al. (2007), it was shown for single target SELEX that in the absence of other information the optimal way to proceed from round to round was to reduce the target concentration by a factor of $1/m$ in passing from the $(m-1)$th to the mth, and that the convergence rate also depended geometrically on the difference between the largest affinity and the second largest affinity. (Such a choice always maximizes the limiting binding probability (target efficiency).) Use of the same round reduction strategy in the multiple target case also leads to very rapid convergence (under 25 rounds). (See Fig. 1(a), (b) and compare with Fig. 3.) Based on these observations, we were strongly motivated to consider the multiple target problem from the dynamical systems point of view.

In Vant-Hull et al. (1998, p. 585), it is stated that "It is not possible for any single paper to fully cover the immense parameter space associated with this model..." We audaciously attempt to do this, at least at a theoretical level. This is possible because we organize the calculation in different manner than that described above.

Consider a library of nucleic acids and a library of target proteins. The results of Levine et al. (2007) suggested an experimental approach for the SELEX process to converge to a pool consisting of a single best binding nucleic acid without recourse to any a priori

in order from best to worst binders are $S_{11} = \{a_1, a_4\}$, $S_{21} = \{a_2\}$, $S_{31} = \{a_3\}$ and the pools for the second protein are $S_{12} = \{a_3\}$, $S_{22} = \{a_1, a_2\}$, $S_{32} = \{a_4\}$, then there is no nucleic acid that is the poorest binder to both the first and the second protein, i.e., there is no nucleic acid in the set $S_{31} \cap S_{32}$. Even if the a_i represent large concentrations of nucleic acids with similar binding properties, or if there are more a_i than nine, it is still quite possible to have a pool for which no member binds most poorly to both proteins.

information about the nature of the binding constants or the distribution of the individual nucleic acid fragments. With a single target, the definition of what it means to be the best binding nucleic acid in a pool of nucleic acids is clear, namely the nucleic acid with the highest affinity with respect to the target. Moreover, given any pool of nucleic acids, the SELEX iteration scheme will always converge to the single best binding nucleic acid present in the initial pool.

However, for the multiple target problem, one has an $M \times N$ matrix, A, of affinities, where M is the number of target components and N is the number of nucleic acid species present in the pool. The relative proportions of each target component are fixed, but the concentration of the total target pool can be varied by dilution. It is not a priori clear that the final distribution of nucleic acid fractions will be independent of the distribution of nucleic acid fractions in the initial pool of nucleic acid fractions, even assuming that all nucleic acids are present in the initial pool. Moreover, it is not even clear what the condition on this matrix should replace the statement for single target SELEX that the SELEX process always converges to the best binding nucleic acid in the pool. No statistical assumption is made about the distribution of equilibrium constants for each protein.

We present a necessary and sufficient condition on the matrix of affinities that ensures, from a theoretical point of view, that a single set of final fractions of nucleic acids is obtained for a fixed target distribution *independently of the initial fractional distribution of nucleic acids as long as all N species are initially present.* This structure condition is closely tied to the geometric properties of a family of chemical potentials against the entire pool at infinite target dilution.

Perhaps the simplest way to formulate this geometric condition is as follows. To each nucleic acid, we associate a vector whose components are its affinities for each target. This vector can be thought of as the target affinity vector for this nucleic acid. (The reader may find the two or three target cases the easiest to visualize.) We form the dot product of each target affinity vector with a candidate vector of free target fractions (i.e., a weighted target affinity for each nucleic acid) and take the largest such dot product. We call this *the maximal target affinity function*. Its graph (over the space of possible free target fraction vectors) is a convex polyhedral surface. This surface will have a minimum value over this space.

Whenever any nucleic acid has the property that any dot product of its target affinity vector with a free target fraction vector is smaller than this minimum, the nucleic acid in question cannot be one of the limiting selection products. In other words, a nucleic acid cannot be selected if every weighted target affinity corresponding to it is smaller than the minimum of the maximal target affinity function. (See Remark 3.)

The faces of this graph are subsets of hyperplanes varying dimensions (points, lines, planes, and higher dimensional "planes"). One can think of these faces as being the intersection of a number of hyperplanes defined by the free-target fraction vectors. A face is said to be proper if the set of target affinity vectors that determine the intersecting hyperplanes is linearly independent, i.e., if the corresponding set of indices uniquely determines the face. If every face is proper, we say the maximal target affinity function is proper. (This is the condition that replaces the condition that each nucleic acid is defined by a unique target affinity in single target SELEX. There we agreed that if two nucleic acids had the same affinity for the target, then with respect to the target they were the "same" nucleic acid.)

Table 1 Notation

Species	Concentration or fraction
Target, T	$[T]$
ith target, T_i	$[T_i]$
ith target fraction, $[T_i]/[T]$	Ω_i
Free target, Tf	$[Tf]$
ith free target, Tf_i	$[Tf_i]$
ith free target fraction, $[Tf_i]/[Tf]$	$\Omega_{f,i}$
Nucleic acid, NA	$[NA]$
jth nucleic acid, NA_j	$[NA_j]$
jth nucleic acid fraction, $[NA_j]/[NA]$	F_j
Free nucleic acid, NAf	$[NAf]$
Free jth nucleic acid, NAf_j	$[NAf_j]$
Bound jth nucleic acid, $\{T : NA_j\}$	$[\{T : NA_j\}]$
Bound jth nucleic acid with ith target, $\{T_i : NA_j\}$	$[\{T_i : NA_j\}]$
Bound nucleic acid $\{T : NA\}$	$[\{T : NA\}]$

In the proper case, it is possible to partition the simplex of initial target fractions into polyhedra with pairwise disjoint interiors in such a way that, given *any set of initial target fractions*, we can uniquely determine the final free target fractions and final nucleic acid fractions if we also know the overall dissociation constant for the limit of the SELEX rounds. In the improper case, the lack of uniqueness for the minima of the chemical potentials means that we can only assert that for every minimizer of the chemical potential there is a free target vector. *The decomposition allows one to determine which nucleic acids are candidates for selection, given the initial target vector, but it does not tell us which nucleic acids will be ultimately chosen from the initial pool of nucleic acids*.

In the analysis below, we have ignored the effects of noncompetitive binding and losses through the support. However, following the analysis given in Levine et al. (2007) using the assumptions in Irvine et al. (1991), these effects can be easily included.

2. Formulation and notation

Although this discussion applies equally to the selection of peptide or nucleic acid aptamers, we frame the underlying chemistry of a single SELEX round in terms of chemical equilibria of nucleic acids. We employ the notation in Table 1 and envisage a pool of N nucleic acids, NA_j for $j \in \{1, 2, \ldots, N\} = \mathcal{N}$ and a pool of M targets T_i for $i \in \{1, 2, \ldots, M\} = \mathcal{M}$. The each target is presumed to be in equilibrium with each nucleic acid:

$$\{T_i : NA_j\} \underset{k_{i,j}}{\overset{k_{-i,j}}{\rightleftharpoons}} Tf_i + NAf_j \quad \text{where } K_{ij} = \frac{k_{-i,j}}{k_{i,j}} = \frac{[NAf_j][Tf_i]}{[\{T_i : NA_j\}]} \tag{1}$$

is the dissociation constant for each of the N nucleic acids binding to the ith target and Tf_i is the available free target of the ith target. In the (idealized) SELEX process, the total bound nucleic acid $\{T : NA\}$ is then separated. The bound nucleic acid is then eluted from the target and subjected to PCR in order to bring the total concentration back to the value of the original pool. The process is repeated with the target (with possibly a different overall concentration but with the same relative proportions of subtarget concentrations, $[T_i]$).

It is probably worth remarking here that the mathematics involved is that of equilibrium statistical mechanics where we are dealing with ensemble averages of molecular species. Each SELEX round can be considered to represent such an equilibrium state.

Define $A_{ij} = 1/K_{ij}$ as the corresponding affinity of the nucleic acid for the target. Define the affinity matrix

$$A = \begin{bmatrix} 1/K_{11} & \dots & 1/K_{1N} \\ 1/K_{21} & \dots & 1/K_{2N} \\ \vdots & \vdots & \vdots \\ 1/K_{M1} & \dots & 1/K_{MN} \end{bmatrix} = \begin{bmatrix} A_{11} & \dots & A_{1N} \\ A_{21} & \dots & A_{2N} \\ \vdots & \vdots & \vdots \\ A_{M1} & \dots & A_{MN} \end{bmatrix}. \tag{2}$$

Denote the rows of this matrix by $\overrightarrow{A}_i$ for $i = 1, \dots, M$ and the columns by $\overrightarrow{A}^j$ for $j = 1, \dots, N$. Associated with such vector quantities are the target and free target concentration vectors: $\overrightarrow{[T]} = \langle [T_1], \dots, [T_M] \rangle$, $\overrightarrow{[Tf]} = \langle [Tf_1], \dots, [Tf_M] \rangle$ as well as the corresponding vectors of percentages (fractions) of each target component:

$$\begin{aligned} \widehat{\Omega} &= \frac{\overrightarrow{[T]}}{[T]} = \left\langle \frac{[T_1]}{[T]}, \dots, \frac{[T_M]}{[T]} \right\rangle = \langle \Omega_1, \dots, \Omega_M \rangle, \\ \widehat{\Omega_f} &= \frac{\overrightarrow{[Tf]}}{[Tf]} = \left\langle \frac{[Tf_1]}{[Tf]}, \dots, \frac{[Tf_M]}{[Tf]} \right\rangle = \langle \Omega_{f,1}, \dots, \Omega_{f,M} \rangle. \end{aligned} \tag{3}$$

These are unit vectors in the L^1 norm (the components are nonnegative and sum to unity).[2]

2.1. *Formulation of the SELEX problem*

There are a number of "conservation" laws for the system:

$$\begin{aligned} [NA] &= \sum_{j=1}^{N} [NA_j], \qquad [T] = \sum_{i=1}^{M} [T_i], \\ \big[\{T : NA_j\}\big] &= \sum_{i=1}^{M} \big[\{T_i : NA_j\}\big] = [NAf_j] \sum_{l=1}^{M} [Tf_l] A_{lj}, \\ [NA_j] &= [NAf_j] + \sum_{i=1}^{M} \big[\{T_i : NA_j\}\big] = [NAf_j] \left(1 + \sum_{l=1}^{M} [Tf_l] A_{lj} \right), \\ [T_i] &= [Tf_i] + \sum_{j=1}^{N} \big[\{T_i : NA_j\}\big]. \end{aligned} \tag{4}$$

[2]The L^1 norm for k-tuple $x = \langle x_1, \dots, x_k \rangle$ of real numbers is denoted variously by $|x| = |x|_1 = \sum_{i=1}^{k} |x_i|$.

Define the fraction F_j of nucleic acid NA_j as $F_j = \frac{[NA_j]}{[NA]}$ and the N unit vector of nucleic acid fractions as $\widehat{F} = \langle F_1, \ldots, F_N \rangle$. For convenience, define

$$D_{j,f} = \sum_{l=1}^{M} [Tf_l] A_{lj} = \overrightarrow{[Tf]} \cdot \overrightarrow{A^j} = \frac{[T : NA_j]}{[NAf_j]} \tag{5}$$

the ratio of bound to free jth nucleic acid. Then the fraction of nucleic acid NA_j bound to the ith target is given by

$$\frac{[\{T_i : NA_j\}]}{[NA]} = \frac{F_j [Tf_i] A_{ij}}{(1 + \sum_{l=1}^{M} [Tf_l] A_{lj})} = \frac{F_j [Tf_i] A_{ij}}{1 + D_{j,f}}. \tag{6}$$

This yields a nonlinear system of equations for the M free target concentrations $[Tf_i]$:

$$\begin{aligned} [T_i] &= [Tf_i] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{(1 + \sum_{l=1}^{M} [Tf_l] A_{lj})} \right) \\ &= [Tf_i] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j A_{ij}}{1 + D_{j,f}} \right) \end{aligned} \tag{7}$$

for $i \in \mathcal{M}$. Summing both sides of the equations in (7) over the free index, we have

$$[T] = [Tf] \left(1 + \frac{[NA]}{[Tf]} \sum_{j=1}^{N} \frac{F_j D_{j,f}}{1 + D_{j,f}} \right). \tag{8}$$

The fraction of bound NA_j to total bound nucleic acid is

$$F'_j \equiv \frac{[\{T : NA_j\}]}{[T : NA]} = \frac{\sum_{i=1}^{M} [\{T_i : NA_j\}]}{[T : NA]} = \frac{[NA]}{[T : NA]} \left(\frac{D_{j,f}}{1 + D_{j,f}} \right) F_j \tag{9}$$

where we have used (6) in the first sum on the right.[3] Summing over both indices in (6), we obtain

$$\frac{[T : NA]}{[NA]} = \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{[\{T_i : NA_j\}]}{[NA]} = \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{F_j [Tf_i] A_{ij}}{(1 + \sum_{l=1}^{M} [Tf_l] A_{lj})} = \sum_{j=1}^{N} \frac{F_j D_{j,f}}{1 + D_{j,f}}. \tag{10}$$

Therefore, $\sum_j F'_j = \sum_j F_j = 1$. The total free nucleic acid is given by

$$[NAf] = \sum_{j=1}^{N} [NAf_j] = [NA] \sum_{j=1}^{N} F_j / (1 + D_{j,f}). \tag{11}$$

[3] In the case of losses through the support and nonselective binding, the term $[T_i : NA_j]$ must be replaced by $a[T_i : NA_j] + b[NA_j]$ where $0 < a, b < 1$. The effect of these parameters is to slow the rate of convergence of the iteration sequence. See Irvine et al. (1991), Vant-Hull et al. (1998) for definitions and Levine et al. (2007) for a detailed analysis and simulations illustrating their effect on the rate of convergence.

Using the notation in (3), (5), and Eq. (10), Eqs. (7), (8), and (9) become

$$F'_j = \frac{[D_{j,f}/(1+D_{j,f})]F_j}{\sum_{l=1}^{N}[D_{l,f}/(1+D_{l,f})]F_l}, \tag{12}$$

$$\begin{aligned}\Omega_i &= \Omega_{f,i}\left(1+[NA]\sum_{j=1}^{N}\frac{F_j A_{ij}}{(1+[Tf]\overrightarrow{A}^{\,j}\cdot\widehat{\Omega_f})}\right)\frac{[Tf]}{[T]}\\ &= \Omega_{f,i}\left(1+[NA]\sum_{j=1}^{N}\frac{F_j A_{ij}}{1+D_{j,f}}\right)\frac{[Tf]}{[T]}, \quad i=1,\ldots,M\end{aligned} \tag{13}$$

and

$$[T] = [Tf]\left(1+[NA]\sum_{j=1}^{N}\frac{F_j\overrightarrow{A}^{\,j}\cdot\widehat{\Omega_f}}{1+[Tf]\overrightarrow{A}^{\,j}\cdot\widehat{\Omega_f}}\right). \tag{14}$$

(Given $\widehat{\Omega_f}$, Eq. (14) has exactly one positive solution $[Tf]$ for each set of F'_js and total target $[T]$ because the right-hand side is strictly increasing in $[Tf]$ and takes values in $(0,\infty)$.)

Notice that, to this point, we have not used amplification (such as by PCR) in our formulation of the SELEX equations. However, when we reformulate Eqs. (9)–(14) as an iterative scheme, we utilize an amplification protocol such as PCR (in the case of nucleic acids) to restore the pool to its original size. Mathematically, this means that the nucleic acid pool size does not change from round to round.

Remark 1. One can also follow the evolution of nucleic acid fraction vectors $\widehat{F}^i = \langle F_{1i},\ldots,F_{Ni}\rangle$ relative to each subtarget defined by

$$F'_{ji} = \big[\{T_i : NA_j\}\big]\Big/\sum_{l=1}^{N}\big[\{T_i : NA_l\}\big] = \frac{F_j A_{ij}}{1+D_{j,f}}\Big/\sum_{l=1}^{N}\frac{F_l A_{il}}{1+D_{l,f}}.$$

This notion is useful if one is to follow the multiple target SELEX by a number of single target SELEX procedures using each of the subtarget species alone in order to fractionate the individual best binders from the pool of those obtained from multiple target SELEX (Vant-Hull et al., 1998, Fig. 2, p. 580). However, we do not do this here because we are only interested in multiple target SELEX.

2.2. *Overall dissociation and association constants, efficiencies*

We consider the overall dissociation and association constants for

$$\{T : NA\} \rightleftharpoons Tf + NAf. \tag{15}$$

When the system is in equilibrium, the dissociation constant for (15) is defined as

$$K_d = \frac{1}{K_a} = \frac{[Tf][NAf]}{[\{T:NA\}]} = \frac{[Tf]\sum_{j=1}^{N} F_j/(1+D_{j,f})}{\sum_{i=1}^{M}\sum_{j=1}^{N}[Tf_i]F_j A_{ij}/(1+D_{j,f})}$$
$$= \frac{\sum_{j=1}^{N} F_j/(1+D_{j,f})}{\sum_{j=1}^{N} F_j(\overrightarrow{A}^j \cdot \widehat{\Omega}_f)/(1+D_{j,f})}. \tag{16}$$

There are two sets of subreactions worth mentioning. The first consists of the N individual reactions of each nucleic acid with the overall target:

$$\{T:NA_j\} \rightleftharpoons Tf + NAf_j. \tag{17}$$

Similarly, at equilibrium, the overall dissociation constant for each reaction is given by

$$K_{d,j} = \frac{1}{K_{a,j}} = \frac{[Tf][NAf_j]}{[T:NA_j]} = \frac{[Tf][NAf_j]}{\sum_{i=1}^{M}[T_i:NA_j]} = \frac{[Tf]}{D_{j,f}} = \frac{1}{\overrightarrow{A}^j \cdot \widehat{\Omega_f}}. \tag{18}$$

The second set of subreactions consists of the M reactions of each target with the overall pool:

$$\{T_i:NA\} \rightleftharpoons Tf_i + NAf. \tag{19}$$

The overall dissociation constant for each reaction at equilibrium is given by

$$\kappa_{d,i} = \frac{1}{\kappa_{a,i}} = \frac{[Tf_i][NAf]}{[T_i:NA]} = \frac{[Tf_i]\sum_{j=1}^{N}[NAf_j]}{\sum_{j=1}^{N}[T_i:NA_j]} = \frac{\sum_{j=1}^{N}[NAf_j]}{\sum_{j=1}^{N}[NAf_j]A_{ij}}$$
$$= \frac{\sum_{j=1}^{N} F_j/(1+D_{j,f})}{\sum_{j=1}^{N} A_{ij}F_j/(1+D_{j,f})}. \tag{20}$$

The overall association constant can be expressed nicely as weighted averages, namely

$$K_a = \frac{1}{K_d} = \sum_{i=1}^{M} \Omega_{f,i}\kappa_{a,i} = \sum_{i=1}^{M} \frac{\Omega_{f,i}}{\kappa_{d,i}} = \sum_{j=1}^{N} \phi_j K_{a,j} = \sum_{j=1}^{N} \frac{\phi_j}{K_{d,j}} \tag{21}$$

where

$$\phi_j = \frac{F_j/(1+D_{j,f})}{\sum_{l=1}^{N} F_l/(1+D_{l,f})} = \frac{F_j - F'_j\sum_{l=1}^{N} D_{l,f}F_l/(1+D_{l,f})}{1-\sum_{l=1}^{N} D_{l,f}F_l/(1+D_{l,f})}.$$

For the reactions (15), (17), and (19), we define overall efficiencies. (In the literature, these are sometimes referred to as *binding probabilities*, e.g. Chen, 2007.) For the first, (15), we define

$$E = \frac{[T]-[Tf]}{[T]} = \frac{[NA:T]}{[Tf]+[NA:T]} = \frac{[NA]}{K_d+[NA]} = \frac{[NA]K_a}{1+[NA]K_a}. \tag{22}$$

For the second, (17), we define

$$E^j = \frac{[T:NA_j]}{[NA_j]} = \frac{D_{j,f}}{1+D_{j,f}} = \frac{K_{a,j}[Tf]}{K_{a,j}[Tf]+1} \tag{23}$$

where we have used the second and third equations in (4) to evaluate the first ratio. This quantity is a measure of the ability of the jth nucleic acid to bind the target. Clearly, for a fixed free target, the larger $K_{a,j}$ is, the greater will be the ability of the jth nucleic acid to bind to the target pool. This is discussed in more detail in Section 4.

For the third, (19), define

$$\begin{aligned} E_i &= 1 - \frac{[Tf_i]}{[T_i]} = \frac{[T_i:NA]}{[T_i]} = \frac{[NA](\sum_{j=1}^N \frac{F_j A_{ij}}{1+D_{j,f}})}{[NA](\sum_{j=1}^N \frac{F_j A_{ij}}{1+D_{j,f}})+1} \\ &= \frac{[NA](\kappa_{a,i}\sum_{j=1}^N \frac{F_j}{1+D_{j,f}})}{[NA](\kappa_{a,i}\sum_{j=1}^N \frac{F_j}{1+D_{j,f}})+1} \end{aligned} \tag{24}$$

where we have used the second and third equations in (4) to evaluate the first ratio. This quantity is a measure of the efficiency of the ith target to bind the nucleic acid pool. This formula tells us that for a given set of nucleic acids and fixed $\kappa_{a,i}$, the target efficiency of each subtarget will achieve its maximum as the total free target converges to zero.

The free target fractions can be expressed in terms of these efficiencies, i.e.,

$$\Omega_{f,i} = \Omega_i \frac{1-E_i}{1-E}. \tag{25}$$

Moreover, an interesting relationship among the efficiencies is given by

$$E = 1 - \frac{\sum[Tf_i]}{\sum[T_j]} = 1 - \frac{\sum[T_i](1-E_i)}{\sum[T_j]} = \frac{\sum[T_i](E_i)}{\sum[T_j]} = \sum_{i=1}^M \Omega_i E_i \equiv \widehat{\Omega}\cdot\overrightarrow{E}. \tag{26}$$

Thus, the overall efficiency is related to the M individual efficiencies of the subtargets in a very geometrically intuitive way.

Finally, we introduce the simplices $\mathcal{S} = \{\widehat{\omega} = \langle\omega_1,\ldots,\omega_M\rangle |\ \omega_i \geq 0, i = 1,\ldots,M$ and $\sum_{i=1}^M \omega_i = 1\}$, and $\mathcal{S}_\mathcal{F} = \{\widehat{f} = \langle f_1,\ldots,f_N\rangle |\ f_i \geq 0, i = 1,\ldots,N$ and $\sum_{i=1}^N f_i = 1\}$. The first simplex is the set of all possible vectors $\widehat{\Omega}$ of target fractions while the second is the set of all nucleic acid fraction vectors.

Because our analysis will be carried out with all targets present, we introduce the open subset of $\mathcal{S}$ namely $\mathcal{S}_0 = \{\widehat{\omega}\in\mathcal{S} |\ \omega_i > 0, i = 1,\ldots,M\}$. The usual physical assumption is that $\widehat{\Omega}\in\mathcal{S}_0$, i.e., the target contains all its components. Therefore, this will be true of the corresponding free target fractions by (25).

From time to time, we will need subsimplices. For example, if $\mathcal{L}\subset\mathcal{N}$, we define the simplex $\mathcal{S}_{\mathcal{F},\mathcal{L}} = \{\widehat{f}\in\mathcal{S}_\mathcal{F} |\ f_i \geq 0$ for $i\in\mathcal{L}$, $f_i = 0$ if $i\in\mathcal{N}-\mathcal{L}\}$.

For each index $j = 1,\ldots,N$ and each vector $\widehat{\omega}\in\mathcal{S}$, we define the vectors (the columns of $\widehat{\omega}^t\cdot A$) as $\overrightarrow{A^j\omega} = \langle A_{1j}\omega_1, A_{2j}\omega_2,\ldots,A_{Mj}\omega_M\rangle^t$. We view the entries of these vectors as the (ω) weighted contribution of each target vector to the affinity of the jth nucleic acid. Because the numbers ω_i, A_{ij} are nonnegative, $|\overrightarrow{A^j\omega}| = \overrightarrow{A}^j\cdot\widehat{\omega}$.

3. The geometry of SELEX

The SELEX iteration process can be summarized and given a geometric interpretation as follows. For each positive number τ_f and each pair of vectors $(\widehat{\Omega}, \widehat{f}) \in \mathcal{S} \times \mathcal{S}_{\mathcal{F}}$, we define $N + M + 1$ functions as follows:

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

$$\Omega_i(\tau_f, \widehat{\omega}, \widehat{f}) = \omega_i \left(1 + [NA] \sum_{j=1}^{N} \frac{A_{ij} f_j}{(1 + \tau_f \overrightarrow{A}^j \cdot \widehat{\omega})}\right) \frac{\tau_f}{\tau}, \quad i = 1, \ldots, M, \tag{27}$$

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

This defines a mapping from $\Re = [0, \infty) \times \mathcal{S} \times \mathcal{S}_{\mathcal{F}}$ into itself. (Notice that $\sum_i \omega_i = \sum_j \tilde{f}_j = 1$.) In the SELEX problem, we are given $(\tau, \widehat{\Omega}, \widehat{f})$ and asked to find $(\tau_f, \widehat{\omega}, \widehat{\tilde{f}})$. Clearly, the direction to take is to solve the first $M + 1$ equations for $(\tau_f, \widehat{\omega})$ and then use these values to compute $\widehat{\tilde{f}}$.

We gain some understanding of the meaning of the final values in the SELEX process by considering the fixed points of the above map. The study of these fixed points amounts to the study of the solutions of the equations satisfied by the limiting values of the iterates. We postpone this discussion until later.

4. The SELEX iteration scheme and its limiting states

4.1. Selection as a function of round number

In terms of an iteration scheme, the single step SELEX problem described in the preceding section as a single algebraic problem is really an iterative process. At the end of the rth round, we have fractions $F_1^{(r)}, F_2^{(r)}, \ldots, F_N^{(r)}$. We give a target vector $\overrightarrow{T}^{(r)} = \langle T_1^{(r)}, \ldots, T_M^{(r)} \rangle$ and solve the M nonlinear equations numerically

$$\left[T_i^{(r)}\right] = \left[T f_i^{(r)}\right] \left(1 + [NA] \sum_{j=1}^{N} \frac{F_j^{(r)} A_{ij}}{(1 + \sum_{l=1}^{M} [T f_l^{(r)}] A_{lj})}\right), \quad i = 1, \ldots, M \tag{28}$$

to compute the free target vector $[\overrightarrow{T f}^{(r)}]$. Then one computes $D_{j,f}^{(r)}$ for $j = 1, \ldots, N$ and computes the fractions for the next round from

$$F_j^{(r+1)} = \frac{[D_{j,f}^{(r)}/(1 + D_{j,f}^{(r)})] F_j^{(r)}}{\sum_{l=1}^{N} [D_{l,f}^{(r)}/(1 + D_{l,f}^{(r)})] F_l^{(r)}}, \tag{29}$$

a cumbersome but easily manageable formula. This alone is a much simpler algorithm than that of Vant-Hull et al. (1998) because one needs only solve M nonlinear equations (28) rather than the much larger nonlinear system used there. Further simplifications

in the numerical procedures are discussed in Section 9.1. A numerical simulation using (28)–(29) is given in Fig. 1.

The fraction of each target component does not change from round to round although we are permitted to vary the total target concentration by dilution. This assumption was adopted so as to be consistent with the constraints of performing SELEX with a single molecular species possessing multiple epitopes and with multiple targets associated with a single cell type. We write $[\overrightarrow{T^{(r+1)}}] = (1 - s_r)[\overrightarrow{T^{(r)}}]$ where $0 \le s_r < 1$ and $[\overrightarrow{T^{(0)}}] = [\overrightarrow{T}]$. The analysis that follows is done in such a way that $[T^{(r)}] \to 0$, i.e. the infinite product $\prod_{r=1}^{\infty}(1 - s_r)$ diverges to zero. The rationale for such a choice is given in Levine et al. (2007).[4] Under these circumstances, it is not hard to see that there are two positive constants d, D such that $d[T^{(r)}] \le [Tf^{(r)}] \le D[T^{(r)}]$. Therefore, the free target concentrations and the total target concentrations converge to zero together.

If $j, l \in \{1, \ldots, N\}$, then

$$\frac{F_j^{(r+1)}}{F_l^{(r+1)}} = \frac{D_{j,f}^{(r)}}{D_{l,f}^{(r)}} \frac{(1 + D_{l,f}^{(r)})}{(1 + D_{j,f}^{(r)})} \frac{F_j^{(r)}}{F_l^{(r)}}. \tag{30}$$

Clearly, the ratio of the fraction of NA_j to that of NA_l will be smaller after one round of SELEX than before the round if and only if $D_{l,f}^{(r)} > D_{j,f}^{(r)}$ (equivalently, $\widehat{\Omega}_f^{(r)} \cdot \overrightarrow{A^l} > \widehat{\Omega}_f^{(r)} \cdot \overrightarrow{A^j}$). We will use this observation to obtain more information about which indices correspond to nucleic acids that are eventually eliminated from the pool and those that survive. To do this, we first need to show that all the (uniformly bounded) sequences $\{F_j^{(r)}\}_{r=1}^{\infty}$ converge to some limiting value, not all of which can be zero. This is done in Appendix B.

4.2. Limiting values as the round number becomes large

Because the sequences $\{F_j^{(r)}\}_{r=1}^{\infty}$ converge, we partition the set of indices $\mathcal{N}$ into a set $\mathcal{L}'$ for which the limit is not zero and its complement, $\mathcal{J}'$ for which the limit is zero. We postpone the determination of these sets for the moment. In the meantime, we let $F_j = \lim_{r\to\infty} F_j^{(r)}$. These limiting values determine the limiting free target fractions.

From (28), we obtain

$$\frac{\Omega_{f,i}^{(r)}[Tf^{(r)}]}{[T^{(r)}]} = \frac{\Omega_i}{1 + [NA]\overrightarrow{A}_i \cdot \widehat{F^{(r)}}}. \tag{31}$$

From Appendix B, the right-hand side has a limit. Therefore, after defining $\mathcal{W}^{(r)} = [Tf^{(r)}]/[T^{(r)}]$ and $\mathcal{W}_i^{(r)} = \Omega_{f,i}^{(r)}\mathcal{W}^{(r)}$, it follows that the limits

$$\lim_{r\to\infty} \mathcal{W}_i^{(r)} \equiv \lim_{r\to\infty} \Omega_{f,i}^{(r)}\mathcal{W}^{(r)} = \frac{\Omega_i}{1 + [NA]\overrightarrow{A}_i \cdot \widehat{F}} = \mathcal{W}_i \tag{32}$$

[4] If the total target concentration is fixed from round to round, many more rounds will be needed to achieve selection in the multiple target case than in the single target case. See Fig. 3, panels (a), (b).

exist. Because $\mathcal{W}^{(r)} = \sum_{i=1}^{M} \mathcal{W}_i^{(r)}$, the limit $\lim_{r\to\infty} \mathcal{W}^{(r)} = \sum_{i=1}^{M} \frac{\Omega_i}{1+[NA]\overrightarrow{A}_i\cdot\widehat{F}} = \mathcal{W}$ exists. Finally, the limiting free-target fractions must be given by $\lim_{r\to\infty} \Omega_{f,i}^{(r)} = \mathcal{W}_i/\mathcal{W} \equiv \Omega_{f,i}$. Since $\kappa_{a,i}^{(r)} \to \overrightarrow{A}_i \cdot \widehat{F}$, it follows that

$$K_a = \lim_{r\to\infty} K_a^{(r)} = \lim_{r\to\infty} \sum_{i=1}^{M} \kappa_{a,i}^{(r)} \Omega_{f,i}^{(r)} = \frac{1}{\mathcal{W}} \sum_{i=1}^{M} \frac{\Omega_i \overrightarrow{A}_i \cdot \widehat{F}}{1 + [NA]\overrightarrow{A}_i \cdot \widehat{F}}. \tag{33}$$

We note that

$$\lim_{r\to\infty} \frac{D_{j,f}^{(r)}}{D_{l,f}^{(r)}} = \lim_{r\to\infty} \frac{K_{a,j}^{(r)}}{K_{a,l}^{(r)}} = \frac{\overrightarrow{A^j}\cdot\overrightarrow{\mathcal{W}}}{\overrightarrow{A^l}\cdot\overrightarrow{\mathcal{W}}} = \frac{\overrightarrow{A^j}\cdot\widehat{\mathcal{W}}}{\overrightarrow{A^l}\cdot\widehat{\mathcal{W}}} \tag{34}$$

where $\overrightarrow{\mathcal{W}} = \langle \mathcal{W}_1, \ldots, \mathcal{W}_M \rangle$.

However, we cannot reverse the argument. That is, suppose the following limits exist:

$$E = \lim_{r\to\infty} E^{(r)} = 1 - \lim_{r\to\infty} \frac{[Tf^{(r)}]}{[T^{(r)}]} = \lim_{r\to\infty} \frac{[NA]K_a^{(r)}}{[NA]K_a^{(r)} + 1} = 1 - \mathcal{W} \tag{35}$$

and

$$\lim_{r\to\infty} \Omega_{f,i}^{(r)} = \Omega_{f,i}. \tag{36}$$

It does not automatically follow that we can uniquely determine the final nucleic acid fractions from this information. We argue as follows: Combining the existence of the limits in (35), (36), with Eq. (25) at each round assures us that the limit

$$E_i = \lim_{r\to\infty} E_i^{(r)} \tag{37}$$

exists. (In fact, the existence of any two of the limits (35), (36), (37) implies the existence of the third.) Therefore, the limit as $r \to +\infty$ of the right-hand side of (31) exists. However, this does not imply that the nucleic acid fraction sequence has a limit because the matrix of affinities is (usually) not of full matrix rank.

4.3. Selection and its connection with asymptotic stability

In order to have selection, i.e., convergence of the SELEX scheme to a unique set of final fractions, *independently of the nucleic acid distribution in the initial pool provided all* $F_i^{(0)} > 0$, we need the notion of global asymptotic stability with respect to such pools. We say that the SELEX process is globally asymptotically stable if for every $\widehat{\Omega}$ there exists a limiting set of nucleic acids $\widehat{F}$ depending only on $\widehat{\Omega}$ such that $\lim_{r\to\infty} |\widehat{F}^{(r)} - \widehat{F}|_1 = 0$ independently of the choice of $\widehat{F}^{(0)}$ (the initial pool with all nucleic acids present).

5. Origin of the SELEX indices

We turn next to the question of determining which nucleic acids will be selected.

5.1. The solution of the SELEX problem at "infinite" dilution ($[NA] \approx 0$)

To motivate the discussion, first consider the scheme (28), (29) when $[NA] = 0$. We assume that the total target concentration is converging to zero in the manner indicated in the discussion following Eq. (29).

Then $[T_i^{(r)}] = [Tf_i^{(r)}]$ so that the fraction of free target at each round is the same as the fraction of total target at that round ($\Omega_{f,i}^{(r)} = \Omega_i^{(r)}$). Moreover, the total and individual efficiencies at each round are zero. (Thus, we expect and do find that the rate of convergence to equilibrium of the SELEX process decreases as the total pool size decreases.) Additionally, $D_{l,f}^{(r)} = [T^{(r)}]\,\overrightarrow{A}^{\,l} \cdot \widehat{\Omega}$, and hence the ratios $D_{j,f}^{(r)}/D_{l,f}^{(r)} = \overrightarrow{A}^{\,j} \cdot \widehat{\Omega} / \overrightarrow{A}^{\,l} \cdot \widehat{\Omega}$ do not depend on the round number. Now let $\mathcal{L} = \{l \in \mathcal{N} \mid \overrightarrow{A}^{\,l} \cdot \widehat{\Omega} = \max\{\overrightarrow{A}^{\,j} \cdot \widehat{\Omega},\ j = 1, \ldots, N\}\}$. Therefore, from (B.3), we see that if $l \in \mathcal{L}$ and $j \notin \mathcal{L}$ the right-hand side of (30) converges to zero, and hence $F_j^{(r)} \to 0$. On the other hand, if $j, l \in \mathcal{L}$, then the above ratios are all unity. Consequently, from (B.1), we see that the indicated products are all unity for every r. *Thus, at infinite nucleic acid dilution, the concentrations of the initial and final pool agree.* The meaning of this result is that for small values of $[NA]$, i.e., $[NA] \ll [T^{(0)}]$, it will take many rounds before the fractional concentrations in the selection process converge to their limiting values. See Fig. 5. There we see that for large pool sizes the overall $K_a^{(r)}$ appears to take its limiting value after less than 100 rounds whereas for small pool sizes, one has to perform many more rounds in order for $K_a^{(r)}$ to be close to the same limiting value.

5.2. The solution of the SELEX problem at finite dilution

When the nucleic acid concentration is positive, the problem becomes mathematically much more intractable than in the case of infinite dilution. Consider, for each round number, the sets $\mathcal{C}_f^{(r)} = \{\overrightarrow{A}^{\,1} \cdot \widehat{\Omega_f}^{(r)}, \overrightarrow{A}^{\,2} \cdot \widehat{\Omega_f}^{(r)}, \ldots, \overrightarrow{A}^{\,N} \cdot \widehat{\Omega_f}^{(r)}\}$ and denote the limiting set by $\mathcal{C} = \{\overrightarrow{A}^{\,1} \cdot \widehat{\mathcal{W}}, \overrightarrow{A}^{\,2} \cdot \widehat{\mathcal{W}}, \ldots, \overrightarrow{A}^{\,N} \cdot \widehat{\mathcal{W}}\}$. Clearly, there is a subset $\mathcal{C}(\widehat{\Omega}, [NA]) \subset \mathcal{C}$ for which all the elements of $\mathcal{C}(\widehat{\Omega}, [NA])$ are equal and exceed all the elements of $\mathcal{C} - \mathcal{C}(\widehat{\Omega}, [NA])$ by some fixed fraction, δ say. Let $\mathcal{L}(\widehat{\Omega}, [NA])$ denote the indices of the elements in $\mathcal{C}(\widehat{\Omega}, [NA])$ and $\mathcal{J}(\widehat{\Omega}, [NA])$ the indices in the complementary set. Then $\overrightarrow{A}^{\,l} \cdot \widehat{\mathcal{W}} \geq (1+\delta)\,\overrightarrow{A}^{\,j} \cdot \widehat{\mathcal{W}}$ where we have written $\widehat{\mathcal{W}} = \overrightarrow{\mathcal{W}}/\mathcal{W}$. It follows by continuity, that for all sufficiently large round numbers r, $|1 - D_{j,f}^{(r)}/D_{l,f}^{(r)}| > \delta/(1+\delta) > \delta/2$ for $l \in \mathcal{L}$, $j \in \mathcal{J}$. Thus, the full series $\sum_{r=1}^{\infty} |1 - D_{j,f}^{(r)}/D_{l,f}^{(r)}|$ diverges. Therefore, $\lim_{r\to\infty} F_j^{(r)} = 0$ when $j \in \mathcal{J}(\widehat{\Omega}, [NA])$.

We do *not* assert that for all $l \in \mathcal{L}(\widehat{\Omega}, [NA])$, $F_l > 0$. That is, the set $\mathcal{L}'$ of Section 4.2 is a (possibly proper) subset of $\mathcal{L}(\widehat{\Omega}, [NA])$.

5.3. A variational characterization of the SELEX indices. The maximal target affinity function, φ, the nonlinear equations satisfied by the final target, and nucleic acid fractions

The observation of the preceding subsection suggests that we define, on the simplex $\mathcal{S}$, the (continuous) convex function

$$\varphi(\widehat{\omega}) = \max\left\{\overrightarrow{A}^{\,j} \cdot \widehat{\omega} \mid j \in \mathcal{N}\right\} \tag{38}$$

for $\widehat{\omega} \in \mathcal{S}$. We call this function *the maximal target affinity function.* We view the numbers $\overrightarrow{A}^j \cdot \widehat{\omega}$ as possible overall association constants for each nucleic acid relative to the target vector, i.e., as numbers $K_{a,j} = 1/K_{d,j}$. Thus, we interpret the value $\varphi(\widehat{\omega})$ as the affinity of each *selected* nucleic acid for the limiting target vector as a function of the individual components of $\widehat{\omega}$. The minimum value of φ is the smallest overall affinity possible for selection to occur. Let $\varphi_{\min}$ denote this value and φ_{Max} the maximum of φ on the simplex. Corresponding to each $\widehat{\omega} \in \mathcal{S}$, there is a unique set of indices $\mathcal{L}(\widehat{\omega}) \subset \mathcal{N}$ for which $\varphi(\widehat{\omega}) = \overrightarrow{A}^j \cdot \widehat{\omega}$ for j in this set.

We denote the number of elements (indices) in $\mathcal{L}(\widehat{\omega})$ by $L_{\widehat{\omega}}$. Let $\mathcal{J}(\widehat{\omega})$ be the complement of this set. Define the level sets of φ as follows. For each positive number $K_a \in [\varphi_{\min}, \varphi_{\text{Max}}]$, the set

$$\mathcal{L}_{K_a} = \left\{ l \in \mathcal{N} | K_a = \overrightarrow{A}^l \cdot \widehat{\omega} = \varphi(\widehat{\omega}) \right\} = \bigcup_{\varphi(\widehat{\omega}) = K_a} \mathcal{L}(\widehat{\omega}) \tag{39}$$

is the maximum possible set of indices that can be selected for this value of $\varphi = K_a$. Thus,

$$\mathcal{L}_0 = \bigcup_{\varphi_{\text{Max}} \geq K_a \geq \varphi_{\min}} \mathcal{L}_{K_a} = \bigcup_{\widehat{\omega} \in \mathcal{S}} \mathcal{L}(\widehat{\omega}).$$

The meaning of this set of integers is that *a nucleic acid, NA_j is selectable for some initial target vector if and only if $j \in \mathcal{L}_0$.*

Just because an index is in $\mathcal{L}_{K_a}$ does not mean that the nucleic acid corresponding to it will ultimately be selected, even if it is present in the initial pool. To see this, suppose $l \in \mathcal{L}_{K_a}$, $F_l^{(0)} > 0$ and $\lim_{r \to \infty} F_l^{(r)} > 0$. Let $\widehat{\omega}$ be the corresponding final free-target fraction vector. This means that $\varphi(\widehat{\omega}) = K_a = \overrightarrow{A}^l \cdot \widehat{\omega}$. Suppose that $j \in \mathcal{L}_{K_a}$. There must be (by definition) another $\widehat{\omega}'$ such that $\overrightarrow{A}^j \cdot \widehat{\omega}' = \varphi(\widehat{\omega}') = K_a$. If $\overrightarrow{A}^j \cdot \widehat{\omega} < \overrightarrow{A}^l \cdot \widehat{\omega}$ then the jth nucleic acid will not be selected.

The graph of φ is convex and made up of closed faces defined as follows: Let $\mathcal{L}$ be an increasingly ordered subset of $\mathcal{N}$ with L elements. The set

$$\Phi(\mathcal{L}) = \overline{\left\{ \left(\widehat{\omega}, \varphi(\widehat{\omega})\right) \ | \overrightarrow{A}^l \cdot \widehat{\omega} = \varphi(\widehat{\omega}), \widehat{\omega} \in \mathcal{S}, \ l \in \mathcal{L} \right\}} \tag{40}$$

where the over line denotes the closure of the set below it (the set together with its limit points) and where the set of vectors $\{ \overrightarrow{A}^l | \ l \in \mathcal{L}\}$ is linearly independent and is called *an L face* of the graph of φ. An L face is called *proper* if the set of indices that describes it is unique. That is, $\Phi(\mathcal{L}) = \Phi(\mathcal{L}')$ implies $\mathcal{L} = \mathcal{L}'$.[5] If an L face is proper, then the set $\mathcal{L}$ is maximal with respect to the defining property. However, a set can be maximal with respect to this property without being unique if the face is not proper.

If precisely two L faces, say $\Phi(\mathcal{L})$ and $\Phi(\mathcal{L}')$ intersect, the intersection is also an L face of the form $\Phi(\mathcal{L}'')$ where $L'' \leq L + L'$ and $\mathcal{L}'' \subset \mathcal{L} \cup \mathcal{L}'$. A similar statement holds if exactly k such L faces intersect.

[5]The standard mathematical notation would be to call an L face an $\max\{M - L, 0\}$ face, but we want an L face to refer to the number of selected nucleic acids if the free target belongs the projection of the L face onto $\mathcal{S}_0$.

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

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

We define certain submatrices of A, the affinity matrix. Suppose $K_a \geq \varphi_{\min}$ and suppose $\widehat{\omega}$ is such that $\varphi(\widehat{\omega}) = K_a$ and belongs to some L face. Let $\mathcal{L}$ be the index set for this L face. Let $A_{\mathcal{L}}$ be the submatrix of A that consists only of those columns whose column index is in $\mathcal{L}$. Let $A_{\mathcal{J}}$ have a similar meaning. *We call the matrix $A_{\mathcal{L}}$ the affinity selection matrix (ASM) for the face with index set $\mathcal{L}$ and the matrix $A_{\mathcal{J}}$ the complement of this matrix (CASM).* We interpret the affinity matrix as an augmented $M \times L$ matrix with an $M \times (N - L)$ matrix, namely $A = [A_{\mathcal{L}}, A_{\mathcal{J}}]$ if we agreed to reorder the columns of A in such a way that the first L columns corresponded to the selected nucleic acids. However, we still use this representation even if the columns of $A_{\mathcal{L}}$ are interspersed among those of $A_{\mathcal{J}}$.

Remark 2. Note that $l \in \mathcal{L}$ is a necessary, but not sufficient condition for the final value F_l to be positive. The actual set of indices, $\mathcal{L}(\widehat{\Omega}, [NA])$ such that $F_l > 0$ will usually be a proper (nonempty) subset of $\mathcal{L}$.

If the only nucleic acids that can be selected are those whose indices belong to $\mathcal{L}$, then from (32) we have

$$\begin{aligned} \mathcal{W}_i &= \frac{\Omega_i}{1 + [NA] \sum_{j=1}^{N} F_j A_{ij}} = \frac{\Omega_i}{1 + [NA] \sum_{j \in \mathcal{L}} F_j A_{ij}} \\ &= \frac{\Omega_i}{1 + [NA] \sum_{j \in \mathcal{L}(\widehat{\Omega}, [NA])} F_j A_{ij}} \end{aligned} \tag{41}$$

where $j \in \mathcal{L}$ is one of the selected nucleic acid indices. Multiplying both sides of this equation by $F_j \geq 0$ (but $F_j > 0$ on $\mathcal{L}(\widehat{\Omega}, [NA])$), summing over $j \in \mathcal{L}$, expressing the dot product as a sum over i and interchanging the order of summation on the left, we find that $\sum_{i=1}^{M} [NA] \Omega_i \overrightarrow{A}_i \cdot \widehat{F} / (1 + [NA] \overrightarrow{A}_i \cdot \widehat{F}) = [NA] \varphi(\widehat{\mathcal{W}}(\widehat{F})) \mathcal{W}(\widehat{F})$ which yields, upon rearrangement, $\mathcal{W}(\widehat{F}) = 1/(1 + [NA] \varphi(\widehat{\mathcal{W}}(\widehat{F})))$. Thus, $K_a[NA]/(1 + K_a[NA]) = E = 1 - \mathcal{W} = [NA] \varphi(\widehat{\mathcal{W}}(\widehat{F}))/[1 + [NA] \varphi(\widehat{\mathcal{W}}(\widehat{F}))]$. Thus, $\varphi(\widehat{\mathcal{W}}(\widehat{F})) = K_a$ as expected. The meaning of this equation is that, *at selection*, i.e., after infinitely many rounds, the overall final affinity of each selected nucleic acid for the target set is the same and has the value $\overrightarrow{A}_{\mathcal{L}} \cdot \overrightarrow{\mathcal{W}}(\widehat{F}) = K_a \mathcal{W} = K_a/(1 + K_a[NA])$ where we have defined the mean target affinities for each target component on $\mathcal{L}$ as the row averages of $A_{\mathcal{L}}$, i.e., the vector $\overrightarrow{A}_{\mathcal{L}}$ has components $\overrightarrow{A}_{\mathcal{L},i} = L^{-1} \sum_{l \in \mathcal{L}} A_{il}$.

Thus, the final fractions must solve a system of $L + M$ nonlinear equations given by

$$\overrightarrow{A^l} \cdot \widehat{\omega} = K_a, \quad l \in \mathcal{L}, \qquad \overrightarrow{A}_i \cdot \widehat{F} = \big(\Omega_i/(\omega_i \mathcal{W}) - 1\big)/[NA], \quad i \in \mathcal{M} \tag{42}$$

where $\mathcal{W} = 1/(1 + [NA]K_a)$ *subject to the constraint that the components of* $\widehat{F}$ *are positive on some nonempty subset* $\mathcal{L}(\widehat{\Omega}, [NA]) \subset \mathcal{L}$ *and otherwise vanish* and where we have set $\omega_i = \mathcal{W}_i/\mathcal{W}$.

It is not hard to show from (42) by summing $(\overrightarrow{A^l} \cdot \widehat{\omega})F_l = K_a F_l$ over l and then summing the equations $\Omega_i = \mathcal{W}\omega_i(1 + [NA]\overrightarrow{A}_i \cdot \widehat{F})$ over i that $\mathcal{W}(\sum_i \omega_i - \sum_l F_l) = 1 - \sum_l F_l$ and, therefore, the constraint conditions $\sum_i \omega_i = 1$ and $\sum_l F_l = 1$ are equivalent. Thus, both hold since the first holds by the definition of $\omega_i = \mathcal{W}_i/\mathcal{W}$. This means that, except for the requirement that the $\omega_i, F_l \geq 0$, we do not need to worry about the constraint conditions as they are satisfied automatically.

We expect, on physical grounds, the set $\mathcal{L}(\widehat{\Omega}, [NA])$ (as a set function of $[NA]$) to be minimal at infinite nucleic acid dilution and to be the set $\mathcal{L}$ at $[NA] = \infty$ (for fixed $\widehat{\Omega}$). The number of elements, $L_{\widehat{\Omega},[NA]}$, in these sets is a piecewise constant, increasing function of $[NA]$.

As $[NA]$ increases, we expect K_a to decrease. (It is not hard to show that $\partial K_a/\partial[NA] \leq 0$.) This suggests that the overall efficiency, $E = K_a[NA]/([NA]K_a + 1)$ should decrease because poorer binders are more likely to bind with the target components. See Fig. 4 for illustrations of these last two statements.

Remark 3. For each L face, $L = 1, 2, \ldots, N$, with index set $\mathcal{L}$, we identify a subsimplex $\mathcal{S}_{\mathcal{F},\mathcal{L}} \subset \mathcal{S}_{\mathcal{F}}$ if and only if the index set defines an L face. Thus, we consider collections of subsimplexes $\{\mathcal{S}_{\mathcal{F},\mathcal{L}} |\ \mathcal{L}$ defines an L face$\}$. Not every index can be in an L face. For example, in Fig. 7, panels (a), (b), the second nucleic acid is not in any index set and in Fig. 7, panels (c), (d), the fifth nucleic acid is not in any index set. The indices $j = 2, 5$ fail to determine the 1 faces. To see this, note that the affinity matrix used to generate the maximal target affinity function, i.e., the quantity $K_a = \varphi(\widehat{\omega})$ that takes the form

$$\begin{aligned} K_a\big([NA]\big) &= (1 - \mathcal{W})/\big([NA]\mathcal{W}\big) \\ &= \left(\sum_{i=1}^{3} \Omega_i A_{ij}/\big(1 + [NA]A_{ij}\big)\right) \Bigg/ \left(\sum_{i=1}^{3} \Omega_i/\big(1 + [NA]A_{ij}\big)\right). \end{aligned}$$

This is always a decreasing function of $[NA]$. It has the value $\sum_{i=1}^{3} \Omega_i A_{ij}$ at $[NA] = 0$ and approaches $1/(\sum_{i=1}^{3} \Omega_i/A_{ij})$ as $[NA]$ becomes large. If the former number is smaller than $\varphi_{\min}$, then j cannot correspond to a $L = 1$ face. Because j is not an $L = 1$ face, we must have that the latter is smaller than $\varphi_{\min}$. Using the data for Fig. 7, in the proper case, for $j = 2$, $\varphi_{\min} \approx 3.07(10)^3\ (\mu\text{M})^{-1}$ while $K_a([NA]) \leq 1.8(10)^3\ (\mu\text{M})^{-1}$. In the improper case, for $j = 5$, $\varphi_{\min} \approx 4.93(10)^3\ (\mu\text{M})^{-1}$ while $K_a([NA]) \leq 2.79(10)^3\ (\mu\text{M})^{-1}$.

6. The connection between the limiting efficiencies and chemical potentials

For readers unfamiliar with the term "chemical potential," we have provided a brief explanation in Appendix A.

When we pass to the limit in the iteration scheme, i.e., to infinite target dilution, the limiting values of the final overall mass fraction ratios for the reactions (15), (17), and (19) have the form: $K_{a,j} = \overrightarrow{A}^{\,j} \cdot \widehat{\omega}$, $\kappa_{a,i} = \widehat{F} \cdot \overrightarrow{A}_i$, and $K_a = \sum_{l\in\mathcal{L}} F_l K_{a,l} = \sum_{i=1}^{M} \omega_i \kappa_{a,i} = \widehat{\omega}^t \cdot A\widehat{F}$, respectively. These formulas imply that the limiting concentrations are equilibrium solutions of the chemical equations at infinite target dilution. Therefore, the chemical potentials μ, μ_i vanish.

The corresponding efficiencies for Eqs. (15) and (19) take the form: $E = [NA]K_a/(1 + [NA]K_a)$, $E_i = [NA]\kappa_{a,i}/(1 + [NA]\kappa_{a,i})$. These formulas as well as the relationship between the chemical potentials and the overall equilibrium constants tell us that any function of the chemical potentials μ, μ_i can be viewed as a function of the overall association constants $K_a, \kappa_{a,i}$ or as a function of the efficiencies E, E_i and the correspondence among the three variable sets is nondegenerate. (See Appendix A.) Define the following function of the individual efficiencies at infinite target dilution:

$$\mathcal{Q}(\overrightarrow{E}) = -RT \sum_{i=1}^{M} \Omega_i \ln\bigl[1/(1 - E_i)\bigr]. \tag{43}$$

This function can be written in several equivalent forms:

$$\begin{aligned}
\mathcal{Q}(\widehat{F}) &= -RT \sum_{i=1}^{M} \Omega_i \ln\bigl(1 + [NA]\overrightarrow{A}_i \cdot \widehat{F}\bigr),\\
\mathcal{Q}\bigl(\overrightarrow{\kappa_a}\bigr) &= -RT \sum_{i=1}^{M} \Omega_i \ln\bigl(1 + [NA]\kappa_{a,i}\bigr) \quad \text{or}\\
\mathcal{Q}\bigl(\overrightarrow{\mu}^{\,a}\bigr) &= -RT \sum_{i=1}^{M} \Omega_i \ln\bigl(1 + [NA]e^{-\mu_i^a/RT}\bigr)
\end{aligned} \tag{44}$$

where R is the gas constant and T is the Kelvin temperature.

When $[NA]$ is very large, $\mathcal{Q}(\overrightarrow{\mu}^{\,a}) \approx \sum_{i=1}^{M} \Omega_i \mu_i^a - RT \ln[NA]$.[6] That is, $\mathcal{Q}$ can be viewed as the (weighted) chemical potential μ^a when the total concentration of target is very small (i.e., at infinite target dilution) and the nucleic acid pool is very large. When this is the case, there is so much nucleic acid present that one can view the nucleic acid pool as interacting with each target in the M subreactions in (19) independently. The fact that the chemical potential is not a finite sum of the individual chemical potentials at moderate values of $[NA]$ is just a reflection of the fact that at moderate concentrations, the nucleic acid pool no longer interacts independently with each target. Therefore, for large concentrations of nucleic acids, seeking minima for $\mathcal{Q}$ subject to a linear constraint such as $\sum F_l = 1$, for example, amounts to seeking minima for the (weighted) chemical potential for the entire system of MN reactions at equilibrium, i.e., seeking the most negative possible (constrained) value of the weighted free energy at equilibrium.

[6] Notice that $0 < \min\{A_{il}\} \le \kappa_{a,i} = \overrightarrow{A}_i \cdot \widehat{F} \le \max\{A_{il}\}$ where the minima and maxima are taken over all nucleic acid indices so that the chemical potentials, μ_i^a, are all bounded.

7. The stability and uniqueness theorems for multiple target SELEX

In the case of single target SELEX, the iterative process always converges to a pool consisting only of a single nucleic acid, the one in the initial nucleic acid pool that binds most tightly to the target.

In the case of multiple target SELEX, the situation is more complicated. The notion of "proper" as defined above replaces the notion that each (set of) nucleic acid(s) in single target SELEX is defined by a unique target affinity. The following statements hold.

Theorem 1. *The target affinity function φ is proper in the sense that we have defined above if and only if for every $\widehat{\Omega}$ of starting target fractions, the SELEX iteration scheme converges to a unique final free target, $\widehat{\omega}$, and a unique set of final fractions $\widehat{F}$ that is independent of the starting pool $\widehat{F}^{(0)} \in \mathcal{S}_{\mathcal{F}}$ as long as this starting vector is an interior point of the simplex $\mathcal{S}_{\mathcal{F}}$. That is, the SELEX process is globally asymptotically stable.*

Theorem 2. *The target affinity function φ is proper in the sense that we have defined above if and only if the chemical potential at infinite target dilution defined by each L face has a unique minimum point.*

The proof of Theorem 1 is given in Appendix C. Theorem 2 is a consequence of the proof of Theorem 1.

Theorem 1 says that, if the maximal target affinity function is improper, i.e., some face is improper, the SELEX process will no longer always converge to a unique set of final nucleic acids for each choice of the initial target vector, $\widehat{\Omega}$ *independently* of nontrivial initial pools of nucleic acids. Rather the outcome will depend on the starting nucleic acid fractions. This loss of uniqueness leads to the possibility, with multiple subtargets, that one can select for more nucleic acids than subtargets. We illustrate these facts in the section on simulations. See Fig. 7 panels (a), (c). Figure 7(a) is constructed from a target affinity function in which all the faces are proper. We see that one cannot select for more nucleic acids than there are target components. However, Fig. 7(c) is constructed from a target affinity function for which there is one improper face. We see that if the initial target fraction vector lies in the quadrilateral labeled $\{1, 2, 3, 4\}$, we can select for four nucleic acids although there are only three components of the initial target vector. In Fig. 9 and Fig. 8, we illustrate the stability statements.

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

The given affinity matrix A defines two functions, one of which, $\varphi : \mathcal{S} \to (0, \infty)$, is convex and continuous on $\mathcal{S}$ and the second, $\mathcal{L} : \mathcal{S} \to P(\mathcal{N})$ where $P(\mathcal{N})$ is the set of all subsets of $\mathcal{N} = \{1, 2, \ldots, N\}$. *For a given free target vector* $\widehat{\omega}$, the function $\mathcal{W} = \mathcal{W}([NA], \widehat{\omega}) = 1/(1 + [NA]\varphi(\widehat{\omega}))$ and unit vectors in $\mathcal{S}$, namely $\widehat{V}^l = \mathcal{W}\widehat{\omega} + (1 - \mathcal{W})\widehat{A^l\omega}$ for $l \in \mathcal{L}$ are defined. The convex hull of this set of vectors was denoted by $\mathcal{H}([NA], \widehat{\omega})$ and is the set from which the starting target fractions must come if and only if the final free target fraction is $\widehat{\omega}$. In order to find the final nucleic acid fraction vector $\widehat{F}$ when $\widehat{\Omega} \in \mathcal{H}([NA], \widehat{\omega})$, we set $F_j = 0$ if $j \notin \mathcal{L}$ and use (C.6) to evaluate the F_j when $j \in \mathcal{L}$.

If φ is proper, then the set $\{\widehat{V}^l \mid l \in \mathcal{L}\}$ is a linearly independent set, i.e., the rank of the Grammian is precisely L and the final nucleic acid fractions are uniquely determined by $\widehat{\Omega}, \widehat{\omega}$. However, if φ is improper and the set $\mathcal{L}$ does not uniquely determine the L face to which $(\widehat{\omega}, \varphi(\widehat{\omega}))$ belongs, then the Grammian will not be invertible and two different starting nucleic acid fraction vectors $\widehat{F}$ will determine two different sets of final nucleic acid fractions. See Fig. 9, panel(a).

In the laboratory, one is concerned with the question: *How can we find the final nucleic acid fractions if we know the starting target fraction vector* $\widehat{\Omega}$, K_a, *and* $[NA]$? If we know the distribution of nucleic acid fractions in the initial pool, we use the SELEX iteration scheme to determine both vector quantities $(\widehat{\omega}, \widehat{F})$ simultaneously regardless of whether φ is proper.

However, consider the laboratory case when the initial pool of nucleic acids is unknown. We proceed as follows:

1. Calculate φ and determine its faces. In principle, this could be a tedious task. However, if M is not too large, it is quite doable. The simplex $\mathcal{S}_0$ can be projected onto the interior of the unit cube in R^{M-1} via the transformation $\omega_1 = 1 - s_1, \omega_2 = s_2(1 - s_1), \ldots, \omega_{M-1} = s_1 \cdots s_{M-2}(1 - s_{M-1}), \omega_M = s_1 \cdots s_{M-2} s_{M-1}$. A rectangular grid can then be imposed on this cube. The pointwise evaluation of φ is then carried out over this grid.
2. Use Eq. (C.5) to compute the convex hull of the set of vectors $\{\widehat{V}^l(\widehat{\omega}) | (\widehat{\omega}, \varphi(\widehat{\omega})) \in \Phi(\mathcal{L})\}$. Call this hull $\mathcal{H}([NA], \Phi(\mathcal{L}))$. Then the simplex $\mathcal{S}$ of target fractions $\widehat{\Omega}$ can be written as

$$\mathcal{S} = \cup\left\{\mathcal{H}\big([NA], \Phi(\mathcal{L})\big) \;\middle|\; \Phi(\mathcal{L}) \text{ is a face of the graph of } \varphi\right\}. \tag{45}$$

3. Suppose that $\widehat{\Omega} \in \mathcal{H}([NA], \Phi(\mathcal{L}))$. Then $\varphi(\widehat{\omega}) = K_a$ for some free-target vector $\widehat{\omega}$ and $\Omega_i = (1 + [NA]\overrightarrow{A}_i \cdot \widehat{F})\omega_i \mathcal{W}$ where $\mathcal{W} = 1/(1 + [NA]K_a)$ and where $F_j = 0$ if $j \notin \mathcal{L}$.
4. If the face in question is proper, the system of L equations (C.2) has one and only one solution. If it is not proper, then there will be a several parameter family of stationary final fraction vectors satisfying (C.2). They will be stable but not asymptotically stable.
5. In the proper case, components of the unique free target are then found from $\omega_i = \Omega_i/(1 + [NA]\overrightarrow{A}_i \cdot \widehat{F})\mathcal{W}$. In the improper case, there will be a several parameter family of free targets corresponding to the family of solutions of (C.2).

9. Simulations

The numerical values used in the simulations are recorded in Tables 2, 3, 4, 5, 6, 7. Before discussing the simulations in detail, a word about units is appropriate. Careful inspection of the formulas in Section 2.1 reveals the following: If all of the terms involving concentrations of all of the nucleic acid and target species, e.g., the values $[X]$ are replaced by a common factor, say $\lambda[X]$ and all of the association constants A_{ij} are replaced by A_{ij}/λ, the resulting formulas are unchanged. The reason for this is that these equations can be rewritten in terms of ratios of concentrations or else as products of concentrations with affinities, both of which are dimensionless quantities.

In Irvine et al. (1991, p. 747), simulations were carried out using a range for K_d between 10^{-9} M to 10^{-7} M (affinities in the range of 10^7 M^{-1} to 10^9 M^{-1}) with a nucleic

acid pool size of 10^{-5} M = 10 μM. Thus, the affinity-concentration product is a dimensionless number of order between 10^2 and 10^4. If we take $\lambda = 10^5$ as the scale factor, this would correspond to the conversion of the values in Irvine et al. (1991) to micromolarity and using association constants in the range between 10^2 $(\mu M)^{-1}$ to 10^4 $(\mu M)^{-1}$ with $[NA] = 1$ μM. (In the laboratory, concentrations of nucleic acid pools are generally in nanomolar to low micromolar range, the concentration of the total target being about the same to an order of magnitude smaller.)

Our goal in the simulations is to explore as best we can, the parameter space in which the dynamics is taking place. To do this as well as to respect the value of the affinity-concentration product above, we took as our initial total target, a value of 1 μM and the total nucleic acid pool as a variable. For most of the simulations, we take the parameter $[NA] = 1$ μM. For the association constant range, our values were in the range 10^2 $(\mu M)^{-1}$ to 10^4 $(\mu M)^{-1}$. These values preserve the affinity-concentration product range used in Irvine et al. (1991).

We sometimes illustrate the effect the nucleic acid pool concentration has on the output by using smaller values. For example, in Fig. 4, we varied this parameter over the range $[10^{-5}, 1]$ in micromolarity. With the choice of affinities in this range, the results we obtain are relatively insensitive to larger values of $[NA]$.

In sum, the reader may disregard the units in the panels and think of them as dimensionless numbers.

9.1. Final fractions using the SELEX iteration scheme

For a given target concentration vector $\overrightarrow{[T^{(r)}]}$ with total concentration $[T^{(r)}]$, with given fraction vector $\widehat{F^{(r)}}$, we first solve the M nonlinear equations (28) to compute the free-target concentration vector $\overrightarrow{[Tf^{(r)}]}$. Rather than use Newton's method to solve (28) as in Vant-Hull et al. (1998), it is much easier (albeit somewhat slower) to use a fixed point iteration method. This obviates the need to calculate the Jacobian matrix. (There is a generalization of Newton's method, the secant method. Its implementation avoids the need to calculate this matrix, but is more cumbersome to program and, while faster than the fixed point scheme, is slower than Newton's method generally.)

We used the zero vector as a first guess, evaluated the right-hand side of (28) to compute a second guess, and repeated the process until the relative error $|[Tf_i^{(r,k+1)}]/[Tf_i^{(r,k)}] - 1|$ was smaller than some specified tolerance. (Here, k is the iteration number for this procedure.) It is not too hard to see that the functions on the right-hand side of (28) are increasing in each of the variables $[Tf_i^{(r)}]$ and that they are bounded above by their numerators so that this iteration scheme must converge. This eliminates the concerns expressed in Vant-Hull et al. (1998, p. 585) concerning the convergence of this iteration scheme.

Then one computes $D_{j,f}^{(r)} = [Tf^{(r)}]\overrightarrow{A^j} \cdot \widehat{\Omega_f^{(r)}}$ for $j = 1, \ldots, N$ and computes the new fractions for the next round from (29). As in Levine et al. (2007), we take $s_r = 1/(r+1)$. It was shown in Levine et al. (2007) for a single target that this choice (or any choice such that $\sum_{r=1}^{\infty} s_r$ is a divergent series) optimizes the target efficiency.

The subfigures in Figs. 1–5 were generated using this iterative approach. In all the figures below, the initial nucleic acid fractions were selected by a random number generator unless otherwise stated.

1a. Figures 1(a), (c), (e), (f). If we use the final free-target vector found by iteration and given in the caption of Fig. 1 to evaluate $\varphi(\widehat{\Omega}_f)$ from its definition, we obtain the value $\varphi(\widehat{\Omega}_f) = 3207.45\ (\mu M)^{-1}$ as well as the set of indices $\{8, 9, 10, 12, 16\}$. To check that this is the minimum value, we computed $\varphi(\widehat{V}_i)$ where $\widehat{V}^i = (\widehat{\Omega}_f + \epsilon\widehat{e}_i)/(1+\epsilon)$ where $\widehat{e}_i$ is one of the standard basis vectors $\widehat{e}_i = \langle \delta_{i1}, \delta_{i2}, \ldots, \delta_{i5} \rangle$ and where $\epsilon > 0$ was small. We found that $\varphi(\widehat{V}^i) > \varphi(\widehat{\Omega}_f)$. Because the graph of φ is convex, this strongly suggests that $\varphi(\widehat{\Omega}_f) = \varphi_{\min} = K_a$. Using the above set of indices, $\widehat{\Omega}_f$ and the starting target fraction vector $\widehat{\Omega}_a = \langle 0.1374, 0.1346, 0.4090, 0.1844, 0.1346 \rangle$, we computed the final nucleic acid fractions corresponding to these vectors from (C.6) and found good agreement. This is confirmed by calculating row and column sums of the classical adjoint of the 5×5 matrix whose columns are indexed by $\{8, 9, 10, 12, 16\}$ and is given in (E.3). This task can be computationally intensive for large matrices. For this matrix, the classical adjoint is given in (E.4). The column vector of row sums is $10^{15}\langle 1.1182, 0.8577, 2.1713, 1.6744, 1.0952 \rangle^t$ and the row vector of column sums is $10^{15}\langle 1.5082, 0.9144, 1.3998, 1.8013, 1.2930 \rangle$. The value for $K_a = 3207.45\ (\mu M)^{-1}$ given in the caption of Fig. 1 was found by iteration. The same value is found by using formula (C.10).

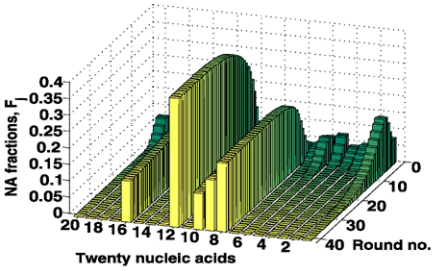
1b. Figures 1(b), (d), (e), (f). For the starting target fraction vector $\widehat{\Omega}_b = \langle 0.2376, 0.1453, 0.1145, 0.2821, 0.2205 \rangle$ we have the set of indices $\{8, 9, 10, 16\}$ with the limiting $K_a = 3256.05\ (\mu M)^{-1} = \varphi(\widehat{\Omega}_f) > \varphi_{\min}$. The limiting NA fraction vector given in the caption of Fig. 1 agrees with the final fractions computed from the formula (C.6). The overall dissociation constant and the overall target efficiency as a function of round number are shown in panel (e), (f) together with the former case in Fig. 1(a).

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

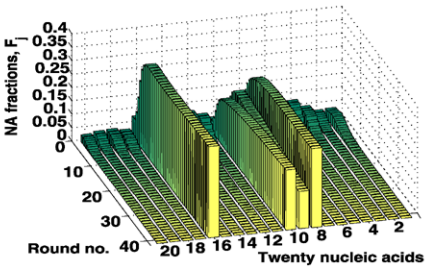
3. Figure 3. We repeated the same calculation as for Figs. 1(a), (b) but did not reduce the total target concentration from round to round. Convergence to the final nucleic acid pools is much slower.

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

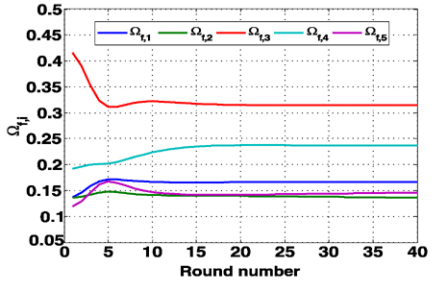
5. Figure 5. We plotted $(pK_a)^{(r)}$ as a function of round number for various pool sizes using the same initial target vector as in Fig. 4. In panel (a), the association constant appears to be decreasing with round number. However, in panel (b), this is not the case when the round number is increased beyond 100, and hence no statement can be made as to the monotonicity of $(pK_a)^{(r)}$. This is a stark difference between single and multiple target SELEX because in the single target case, the overall K_a increases as a function of round number.



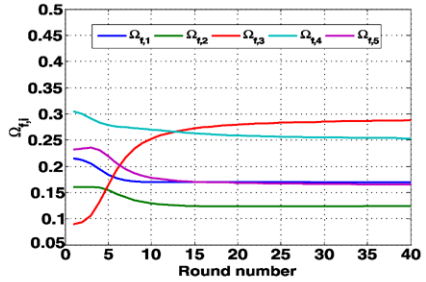
(a) Distribution of NA fractions



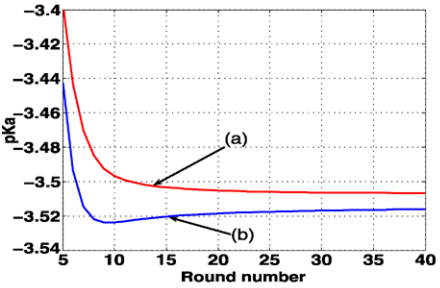
(b) Distribution of NA fractions



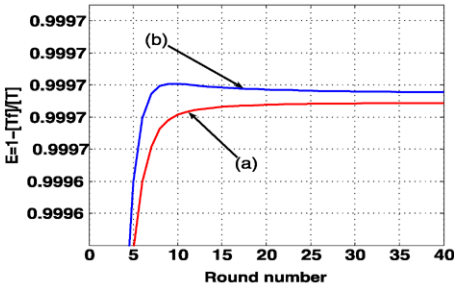
(c) Free target component fractions



(d) Free target component fractions

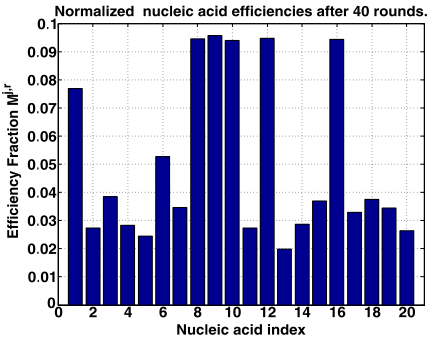


(e) Overall dissociation constant

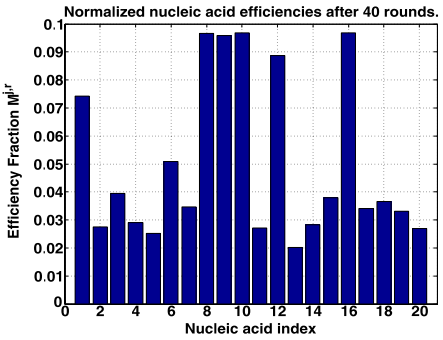


(f) Overall efficiency

Fig. 1 Comparison of two SELEX experiments with different target vector fractions. Here and throughout. $pK_a = -\log_{10}(K_a)$. Panels (a), (b): With $[NA] = [T] = 1$ μM, the index sets for the selected nucleic acids are (a) $\{8, 9, 10, 12, 16\}$ and (b) $\{8, 9, 10, 16\}$ because for indices not in this set, $F_i = 0$. The initial target vectors are (a) $\widehat{\Omega}_a = \langle 0.1374, 0.1346, 0.4090, 0.1844, 0.1346\rangle$ and (b) $\widehat{\Omega}_b = \langle 0.2376, 0.1453, 0.1145, 0.2821, 0.2205\rangle$ with the nonzero components of limiting NA fraction vectors given by $F_a = \langle F_8, F_9, F_{10}, F_{12}, F_{16}\rangle = \langle 0.1956, 0.2794, 0.0498, 0.3843, 0.0908\rangle$ and $F_b = \langle F_8, F_9, F_{10}, F_{16}\rangle = \langle 0.3060, 0.0670, 0.2629, 0.3640\rangle$. These limiting vectors agree with values from the formula (C.6) to at least eight significant figures (not shown). Panels (c), (d): The final free target fractions in cases (a), (b), after 40 rounds, are: (c) $\widehat{\Omega}_f = \langle 0.1617, 0.1241, 0.3139, 0.2420, 0.1583\rangle$ with limiting $K_a = 3207.45\ (\mu\text{M})^{-1}$ and (d) $\widehat{\Omega}_f = \langle 0.1690, 0.1288, 0.2965, 0.2476, 0.1581\rangle$ with limiting $K_a = 3256.05\ (\mu\text{M})^{-1}$. The starting ordinates of each of the curves in panels (c), (d), are the components of the initial target vectors $\widehat{\Omega}_a$ and $\widehat{\Omega}_b$ above. Panel (e): The overall dissociation constant as a function of round number. Panel (f): The overall target efficiency as a function of round number. (Eq. (22).)

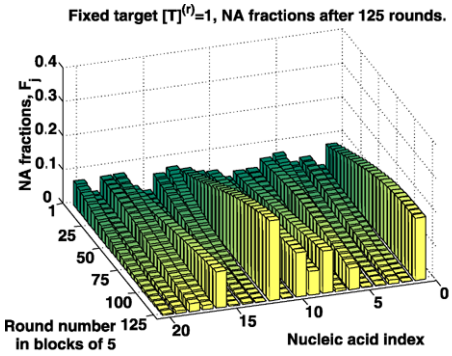


(a) Individual nucleic acid efficiencies, $r = 40$, for Figure 1(a).

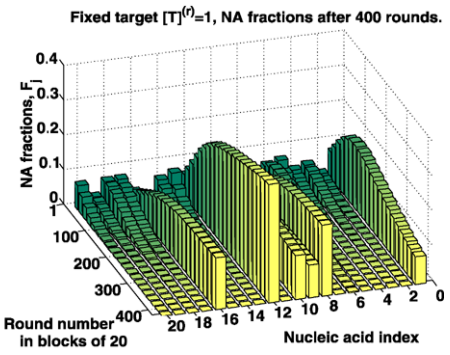


(b) Individual nucleic acid efficiencies, $r = 40$, for Figure 1(b).

Fig. 2 The individual nucleic acid efficiencies after 40 rounds of the SELEX experiments shown in Fig. 1. The bar graphs are not identical because, in Fig. 1(a), selection is for indices $\{8, 9, 10, 12, 16\}$ while, in Fig. 1(b), selection is for indices $\{8, 9, 10, 16\}$.



(a) Selected complexes after 125 rounds



(b) Selected complexes after 400 rounds

Fig. 3 Misleading selection resulting from fixing target. After 125 rounds of the SELEX experiments shown Fig. 1, it would appear from panel (a) that the selected indices are $\{1, 6, 8, 9, 10, 12, 16\}$ whereas from panel (b) after 400 rounds the selected indices would appear to be $\{1, 8, 9, 10, 12, 16\}$. Compare with Fig. 1(a). Thus, the importance of reducing the total concentration of target from round to round in a systematic way is even more pronounced in the multiple target problem.

9.2. Comparison of stationary SELEX solutions with those obtained by iteration. Decomposition of the initial target set

For more than two or three target components, it will be a challenge to the numerical analyst/programmer to write an algorithm that can efficiently test which faces of the maximal target affinity function are proper. Indeed, even for the example in the preceding section, we cannot positively assert that this function is proper. However, in the computations that involve the affinities in Figs. 6 and 7, we do not rely on whether or not the function is proper. In Figs. 8 and 9, we examine the stability properties of a target consisting of three subtargets in order to illustrate the theory discussed in Sections 5–8.

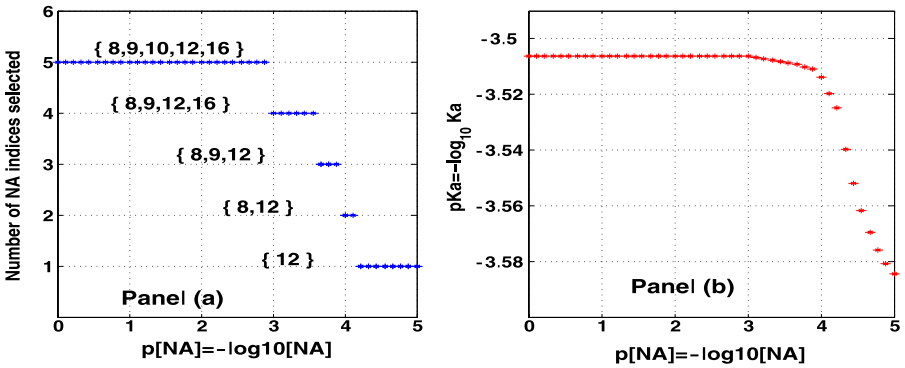


Fig. 4 Variability of selected nucleic acid sets with total pool size. In panel (a), the dependence of the number of selected nucleic acids on the nucleic acid pool size is illustrated. In panel (b), pK_a decreases with decreasing pool size, i.e., K_a decreases with increasing pool size.

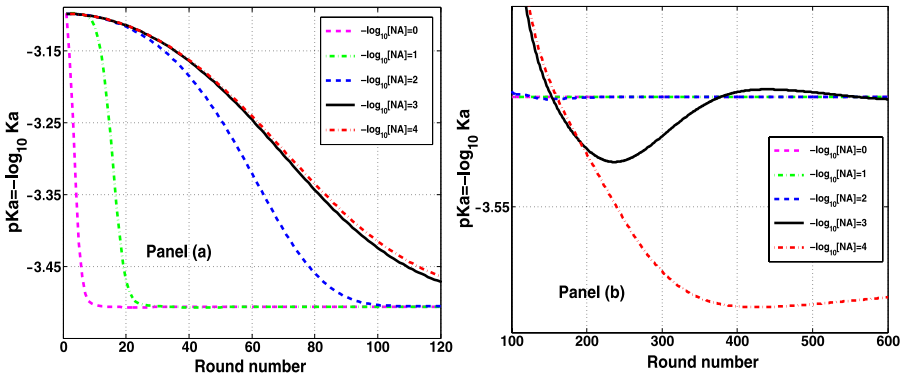


Fig. 5 Nonmonotonicity of pK_a as a function of round number. The left-hand panel traces $pK_a^{(r)}$ for various pool sizes up to 120 rounds. The right-hand panel is a continuation of these curves for an expanded scale on the vertical axis.

The various panels in Fig. 6 represent (a), the graph of the maximal target affinity function, (b), the contour lines of this function, (c), (d), the results of running the SELEX program over a grid in initial target space to partition it into the various hulls that should be obtained independently using the decomposition suggested in the formula (45) and calculated using Eqs. (C.5)–(C.6). We see how the nucleic acid pool size can affect the initial target space. The level sets of φ are drawn in panel (b). Using the SELEX program, the initial target space is partitioned in panel (c) using a value of $[NA] = 10^{-4}$ M and in panel (d) with $[NA] = 1$ μM. In comparing these two panels, note that the size of regions where two nucleic acids can be selected increases at the expense of the regions where only one nucleic acid can be selected and decreases at the expense of regions where three nucleic acids can be selected. Low concentrations of nucleic acids are thus more likely to lead to selection of a single nucleic acid species than are high concentrations.

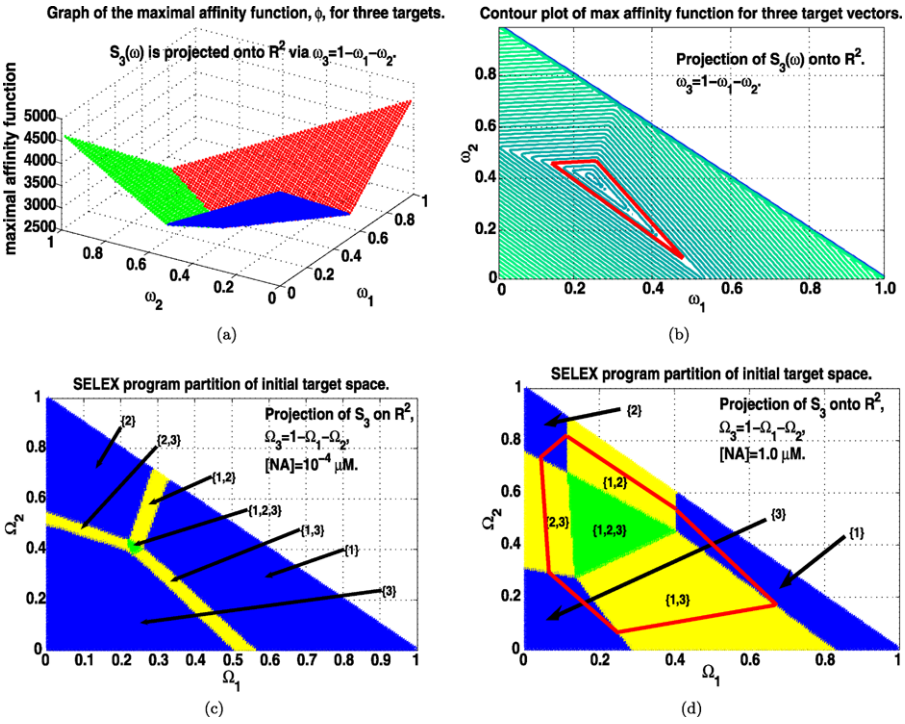


Fig. 6 The maximal target affinity function, its level sets and the effect of pool size on the decomposition of the initial target space. Panel (a) is a graph of the maximal target affinity function when its domain has been projected onto the plane. The panel (c) was generated from a fixed initial pool of nucleic acids $\widehat{F} = \langle F_1, F_2, F_3, F_4, F_5\rangle = \langle 0.3002, 0.0731, 0.1917, 0.1535, 0.2815\rangle$ using the SELEX iteration scheme for 2,000 rounds. (The number of rounds is large because the rate of convergence of SELEX program slows as the nucleic acid pool size decreases.) In panels (c), (d), the indicated regions are labeled with the indices of the nucleic acids that will be selected when the initial target is selected from the interior of the indicated region. The triangle in panel (b) generates the hexagon in panel (d). (There is a corresponding hexagon for panel (c) that is omitted in the interests of clarity.) The initial target vectors $\langle \Omega_1, \Omega_2, \Omega_3\rangle$ were generated by setting $\Omega_3 = 1 - \Omega_1 - \Omega_2$, where for $j = 1, \ldots, 199$, $i = 1, \ldots, 200 - j$ and $\Omega_1 = \Omega_1(i, j) = j/200$, $\Omega_2 = \Omega_2(i, j) = i/200$ or about 20,000 initial target vectors.

In Fig. 7, we illustrate the formula (45) by comparing the results of running the SELEX programs with a partition generated by (45) of the initial target space using the discussion in Section 8 in the two cases φ proper (panels (a), (b)) and φ improper (panels (c), (d)). We considered the following scenarios with three target components and five nucleic acids:

1. The function φ is proper, has a unique minimum point and there is a second point on the graph of φ defined by exactly three intersecting planes. (Panels (a), (b).)
2. The function φ is improper with a unique point common to all four hyperplanes. (Panels (c), (d).)

Panels (a), (c) were generated by the SELEX program for many points in the initial target simplex. Panels (b), (d) were generated using the stationary state algebraic equations discussed in Section 8.

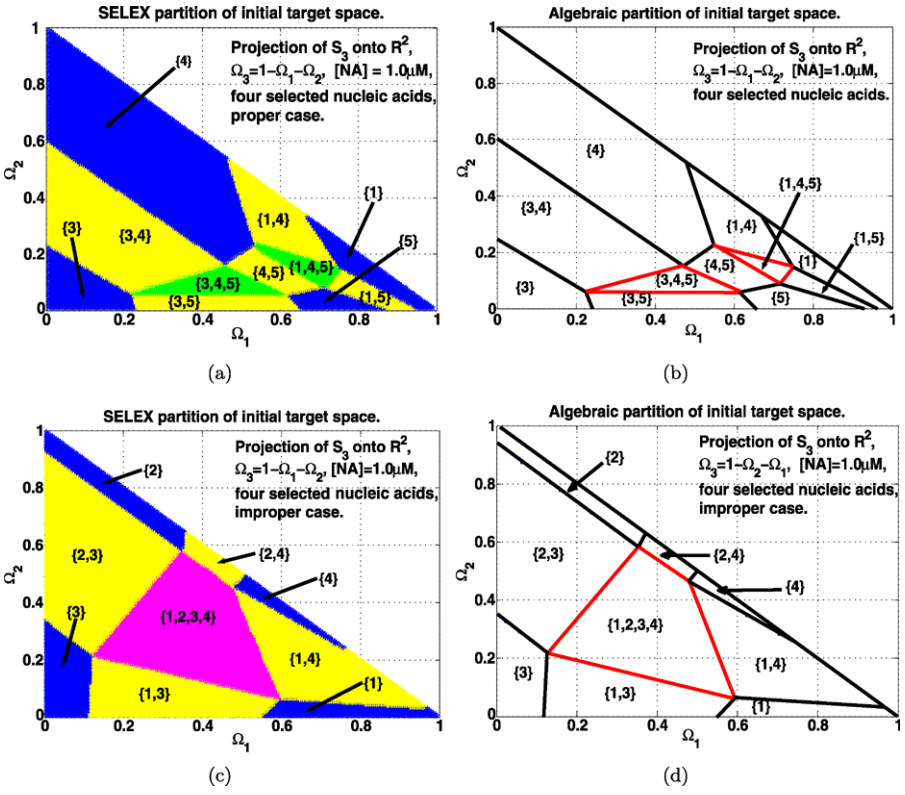
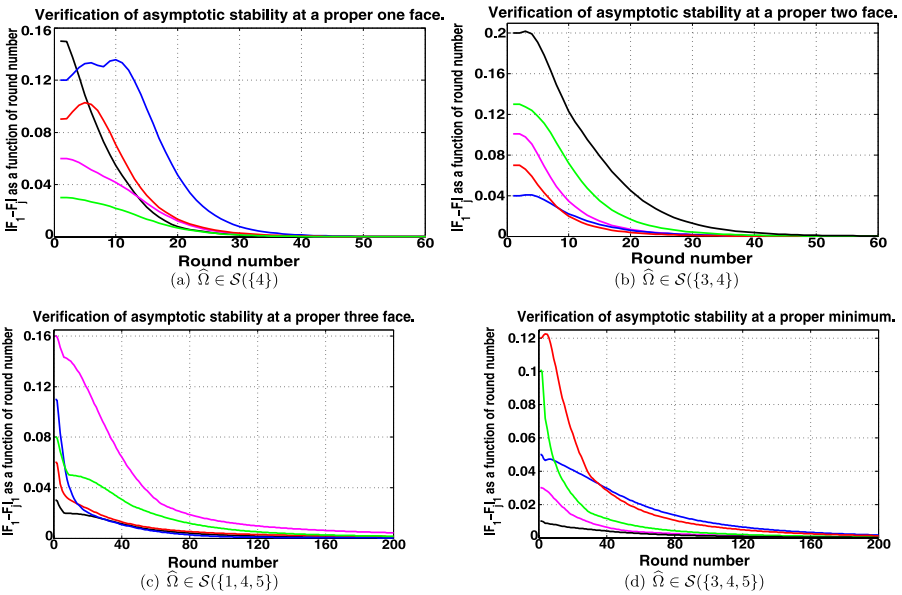


Fig. 7 Comparison of proper and improper case target space decomposition-dynamic and algebraic limiting states. The panels (a), (c), were generated in the same manner as panels (c), (d) in Fig. 6. In all four panels, the subregions are labeled with the indices of the nucleic acids that will be selected when the initial target vector is in the indicated region. We refer to the case illustrated in panels (c), (d) as improper because the minimum of the maximal target affinity function, φ, is defined by the intersection of any three of the four planes that define its graph.

9.3. Stability properties of stationary solutions

To illustrate the asymptotic stability properties of the SELEX process, we consider two cases for which $M = 3$ with several nucleic acids present. In the first case, the maximal target affinity function is proper while in the second case, it is improper.

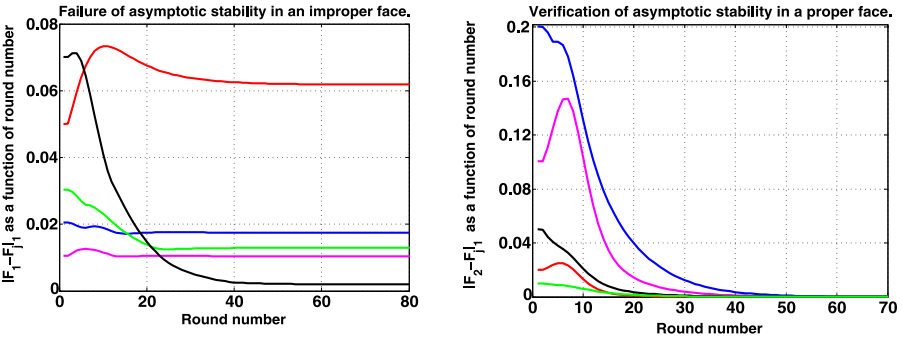
1. In Fig. 8, panel (a) (or panel (b) in Fig. 7), the asymptotic stability of the SELEX scheme was tested for several choices of initial target fractions one from each of the regions labeled $\{4\}$, $\{3, 4\}$, $\{1, 4, 5\}$, $\{3, 4, 5\}$. We label these sets respectively as $\mathcal{S}(\{4\}), \mathcal{S}(\{3, 4\}), \mathcal{S}(\{1, 4, 5\}), \mathcal{S}(\{3, 4, 5\})$. In each of the four regions indicated, a value of the starting target fraction vector $\widehat{\Omega}$ was selected. In the nucleic acid simplex, $\mathcal{S}_{\mathcal{F}}$, six random vectors $\{\widehat{F}_1{}^{(0)}, \ldots, \widehat{F}_6{}^{(0)}\}$ were generated and the one norms, $|\widehat{F}_1{}^{(r)} - \widehat{F}_j{}^{(r)}|_1$, $j = 2, \ldots, 6$, plotted as a function of r, the round number.



(a) $\widehat{\Omega} \in \mathcal{S}(\{4\})$ (b) $\widehat{\Omega} \in \mathcal{S}(\{3,4\})$

(c) $\widehat{\Omega} \in \mathcal{S}(\{1,4,5\})$ (d) $\widehat{\Omega} \in \mathcal{S}(\{3,4,5\})$

Fig. 8 Convergence of SELEX experiments to a unique NA set regardless of the starting NA pool. These figures show that whatever convex hull in Fig. 7, panel (a) (or panel (b)), to which the initial target fraction vector belongs, the SELEX process converges to a unique set of final nucleic acid fractions independently of the distribution of starting nucleic acids in the starting pool. Each of the five curves in each panel correspond to an independent trial for the starting pool of nucleic acids. (The vertical axis notation $|F_1 - F_j|_1$ is shorthand for $|\widehat{F_1}^{(r)} - \widehat{F_j}^{(r)}|_1 = \sum_{i=1}^{5} |F_{i,1}^{(r)} - F_{i,j}^{(r)}|$. Here, $j = 2, \ldots, 6$ for five of the six random starting vectors. The starting ordinates of the norms $|F_1 - F_j|$ for each of the curves in panels (a)–(d) are recorded in the fourth column of Tables 2–5 respectively.)



(a) $\widehat{\Omega} \in \mathcal{S}(\{1,2,3,4\})$ (b) $\widehat{\Omega} \in \mathcal{S}(\{2,3\})$

Fig. 9 The SELEX process is not globally asymptotically stable when the initial target fraction belongs to the convex hull of an improper face (a) and is asymptotically stable when it belongs to a convex hull corresponding to a proper face (b). Each of the five curves in both panels correspond to an independent trial for the starting pool of nucleic acids. (The vertical axis notation $|F_1 - F_j|_1$ is shorthand for $|\widehat{F_1}^{(r)} - \widehat{F_j}^{(r)}|_1 = \sum_{i=1}^{5} |F_{i,1}^{(r)} - F_{i,j}^{(r)}|$. Here $j = 2, \ldots, 6$ for five of the six random starting vectors. The starting ordinates of the norms $|F_1 - F_j|$ for each of the curves in panels (a), (b) are recorded in the fourth column of Tables 6, 7 respectively.)

Table 2 Data for Fig. 8, panel (a). $\widehat{\Omega} = \langle 0.3, 0.6, 0.1 \rangle$

i	$\widehat{F}_i^{(0)}$	$\widehat{F}_i^{(200)}$	$\lvert\widehat{F_1}^{(0)} - \widehat{F}_i^{(0)}\rvert_1$	$\lvert\widehat{F_1}^{(200)} - \widehat{F}_i^{(200)}\rvert_1$
1	$\langle 0.3002, 0.0731, 0.1917, 0.1535, 0.2815 \rangle$	$\langle 0.0000, 0.0000, 0.0000, 1.0000, 0.0000 \rangle$	0	0
2	$\langle 0.3120, 0.0746, 0.1934, 0.1520, 0.2680 \rangle$	$\langle 0.0000, 0.0000, 0.0000, 1.0000, 0.0000 \rangle$	0.0301	0.0000
3	$\langle 0.2846, 0.0587, 0.2093, 0.1539, 0.2935 \rangle$	$\langle 0.0000, 0.0000, 0.0000, 1.0000, 0.0000 \rangle$	0.0600	0.0000
4	$\langle 0.3161, 0.0430, 0.2083, 0.1386, 0.2940 \rangle$	$\langle 0.0000, 0.0000, 0.0000, 1.0000, 0.0000 \rangle$	0.0901	0.0000
5	$\langle 0.3008, 0.1325, 0.1614, 0.1350, 0.2703 \rangle$	$\langle 0.0000, 0.0000, 0.0000, 1.0000, 0.0000 \rangle$	0.1201	0.0000
6	$\langle 0.2542, 0.0947, 0.2451, 0.1481, 0.2579 \rangle$	$\langle 0.0000, 0.0000, 0.0000, 1.0000, 0.0000 \rangle$	0.1500	0.0000

Table 3 Data for Fig. 8, panel (b). $\widehat{\Omega} = \langle 0.15, 0.3, 0.55 \rangle$

i	$\widehat{F}_i^{(0)}$	$\widehat{F}_i^{(200)}$	$\lvert\widehat{F_1}^{(0)} - \widehat{F}_i^{(0)}\rvert_1$	$\lvert\widehat{F_1}^{(200)} - \widehat{F}_i^{(200)}\rvert_1$
1	$\langle 0.3002, 0.0731, 0.1917, 0.1535, 0.2815 \rangle$	$\langle 0.0000, 0.0000, 0.4583, 0.5417, 0.0000 \rangle$	0	0
2	$\langle 0.2991, 0.0833, 0.2015, 0.1536, 0.2625 \rangle$	$\langle 0.0000, 0.0000, 0.4583, 0.5417, 0.0000 \rangle$	0.0401	$0.4(10^{-12})$
3	$\langle 0.3329, 0.0474, 0.1846, 0.1513, 0.2838 \rangle$	$\langle 0.0000, 0.0000, 0.4583, 0.5417, 0.0000 \rangle$	0.0700	$0.3(10^{-12})$
4	$\langle 0.2943, 0.1093, 0.2011, 0.1583, 0.2370 \rangle$	$\langle 0.0000, 0.0000, 0.4583, 0.5417, 0.0000 \rangle$	0.1008	$0.4(10^{-12})$
5	$\langle 0.2561, 0.1227, 0.1708, 0.1660, 0.2844 \rangle$	$\langle 0.0000, 0.0000, 0.4583, 0.5417, 0.0000 \rangle$	0.1301	$1.1(10^{-12})$
6	$\langle 0.2693, 0.1212, 0.2436, 0.1338, 0.2321 \rangle$	$\langle 0.0000, 0.0000, 0.4583, 0.5417, 0.0000 \rangle$	0.2001	$3.6(10^{-12})$

Table 4 Data for Fig. 8, panel (c). $\widehat{\Omega} = \langle 0.65, 0.15, 0.2 \rangle$

i	$\widehat{F}_i^{(0)}$	$\widehat{F}_i^{(200)}$	$\lvert\widehat{F_1}^{(0)} - \widehat{F}_i^{(0)}\rvert_1$	$\lvert\widehat{F_1}^{(200)} - \widehat{F}_i^{(200)}\rvert_1$
1	$\langle 0.3002, 0.0731, 0.1917, 0.1535, 0.2815 \rangle$	$\langle 0.1112, 0.0000, 0.0000, 0.4007, 0.4881 \rangle$	0	0
2	$\langle 0.2958, 0.0880, 0.1912, 0.1433, 0.2817 \rangle$	$\langle 0.1110, 0.0000, 0.0000, 0.4006, 0.4884 \rangle$	0.0301	0.0005
3	$\langle 0.3065, 0.0717, 0.1631, 0.1652, 0.2935 \rangle$	$\langle 0.1105, 0.0000, 0.0000, 0.4007, 0.4888 \rangle$	0.0601	0.0013
4	$\langle 0.2853, 0.1022, 0.2002, 0.1284, 0.2839 \rangle$	$\langle 0.1104, 0.0000, 0.0000, 0.4006, 0.4890 \rangle$	0.0801	0.0016
5	$\langle 0.3315, 0.0841, 0.1432, 0.1470, 0.2942 \rangle$	$\langle 0.1114, 0.0000, 0.0000, 0.4006, 0.4880 \rangle$	0.1100	0.0003
6	$\langle 0.2817, 0.0689, 0.1344, 0.2239, 0.2911 \rangle$	$\langle 0.1091, 0.0000, 0.0000, 0.4008, 0.4901 \rangle$	0.1600	0.0043

Table 5 Data for Fig. 8, panel (d). $\widehat{\Omega} = \langle 0.4, 0.1, 0.5 \rangle$

i	$\widehat{F}_i^{(0)}$	$\widehat{F}_i^{(200)}$	$\lvert\widehat{F_1}^{(0)} - \widehat{F}_i^{(0)}\rvert_1$	$\lvert\widehat{F_1}^{(200)} - \widehat{F}_i^{(200)}\rvert_1$
1	$\langle 0.3002, 0.0731, 0.1917, 0.1535, 0.2815 \rangle$	$\langle 0.0000, 0.0000, 0.3974, 0.4161, 0.1865 \rangle$	0	0
2	$\langle 0.2952, 0.0730, 0.1922, 0.1570, 0.2826 \rangle$	$\langle 0.0000, 0.0000, 0.3974, 0.4162, 0.1864 \rangle$	0.0100	0.0002
3	$\langle 0.3120, 0.0746, 0.1934, 0.1520, 0.2680 \rangle$	$\langle 0.0000, 0.0000, 0.3974, 0.4162, 0.1864 \rangle$	0.0301	0.0003
4	$\langle 0.3028, 0.0558, 0.1855, 0.1759, 0.2800 \rangle$	$\langle 0.0000, 0.0000, 0.3973, 0.4169, 0.1858 \rangle$	0.0500	0.0017
5	$\langle 0.2920, 0.1193, 0.1959, 0.1380, 0.2548 \rangle$	$\langle 0.0000, 0.0000, 0.3975, 0.4158, 0.1867 \rangle$	0.1008	0.0005
6	$\langle 0.3033, 0.0935, 0.2282, 0.1283, 0.2466 \rangle$	$\langle 0.0000, 0.0000, 0.3975, 0.4155, 0.1870 \rangle$	0.1201	0.0012

Table 6 Data for Fig. 9, panel (a). $\widehat{\Omega} = \langle 0.35, 0.4, 0.25 \rangle$

i	$\widehat{F}_i^{(0)}$	$\widehat{F}_i^{(200)}$	$\lvert\widehat{F_1}^{(0)} - \widehat{F}_i^{(0)}\rvert_1$	$\lvert\widehat{F_1}^{(200)} - \widehat{F}_i^{(200)}\rvert_1$
1	$\langle 0.2384, 0.1693, 0.1056, 0.1075, 0.3792 \rangle$	$\langle 0.1334, 0.4464, 0.2506, 0.1696, 0.0000 \rangle$	0	0
2	$\langle 0.2411, 0.1641, 0.1061, 0.1090, 0.3797 \rangle$	$\langle 0.1318, 0.4427, 0.2514, 0.1741, 0.0000 \rangle$	0.0105	0.0104
3	$\langle 0.2440, 0.1637, 0.1048, 0.1122, 0.3753 \rangle$	$\langle 0.1307, 0.4403, 0.2519, 0.1771, 0.0000 \rangle$	0.0205	0.0174
4	$\langle 0.2282, 0.1738, 0.1104, 0.1025, 0.3851 \rangle$	$\langle 0.1353, 0.4509, 0.2497, 0.1641, 0.0000 \rangle$	0.0303	0.0130
5	$\langle 0.2460, 0.1845, 0.0939, 0.0942, 0.3814 \rangle$	$\langle 0.1427, 0.4680, 0.2462, 0.1431, 0.0000 \rangle$	0.0500	0.0618
6	$\langle 0.2391, 0.1875, 0.0744, 0.1237, 0.3753 \rangle$	$\langle 0.1337, 0.4470, 0.2505, 0.1688, 0.0000 \rangle$	0.0700	0.0019

Table 7 Data for Fig. 9, panel (b). $\widehat{\Omega} = \langle 0.1, 0.6, 0.3 \rangle$

i	$\widehat{F}_i^{(0)}$	$\widehat{F}_i^{(200)}$	$\lvert\widehat{F_1}^{(0)} - \widehat{F}_i^{(0)}\rvert_1$	$\lvert\widehat{F_1}^{(200)} - \widehat{F}_i^{(200)}\rvert_1$
1	$\langle 0.2384, 0.1693, 0.1056, 0.1075, 0.3792 \rangle$	$\langle 0.0000, 0.6020, 0.3980, 0.0000, 0.0000 \rangle$	0	0
2	$\langle 0.2327, 0.1685, 0.1156, 0.1074, 0.3758 \rangle$	$\langle 0.0000, 0.6020, 0.3980, 0.0000, 0.0000 \rangle$	0.0201	$0.5(10^{-13})$
3	$\langle 0.2166, 0.1661, 0.1134, 0.1081, 0.3958 \rangle$	$\langle 0.0000, 0.6020, 0.3980, 0.0000, 0.0000 \rangle$	0.0501	$0.3(10^{-12})$
4	$\langle 0.2643, 0.1767, 0.0905, 0.1093, 0.3592 \rangle$	$\langle 0.0000, 0.6020, 0.3980, 0.0000, 0.0000 \rangle$	0.0701	$0.3(10^{-12})$
5	$\langle 0.2378, 0.1562, 0.1545, 0.1090, 0.3425 \rangle$	$\langle 0.0000, 0.6020, 0.3980, 0.0000, 0.0000 \rangle$	0.1009	$0.8(10^{-12})$
6	$\langle 0.1768, 0.1582, 0.0778, 0.2035, 0.3837 \rangle$	$\langle 0.0000, 0.6020, 0.3980, 0.0000, 0.0000 \rangle$	0.2010	$0.8(10^{-11})$

2. In Fig. 9, the asymptotic stability of the SELEX scheme was tested for choices of two initial target fractions, one from each of the regions labeled $\mathcal{S}(\{1,2,3,4\}), \mathcal{S}(\{2,3\})$, in Fig. 7, panel (c) (or panel (d)). Again six choices of initial nucleic acid fraction vectors were selected at random and the five quantities $|\widehat{F}_1^{(r)} - \widehat{F}_j^{(r)}|_1$, were plotted as a function of the round number r. In panel (a), these quantities do not converge to zero, thus illustrating the failure of asymptotic stability at an "improper" face. On the other hand, in panel (b), these quantities do converge to zero.

10. Summary

We have the following statements.

1. We give an algorithm for multiple target SELEX that reduces the computation time and programming labor considerably over the algorithm described in Vant-Hull et al. (1998).
2. The SELEX process always converges to some limiting vector of nucleic acid fractions and some final composition of free target fractions.
3. We give a necessary and sufficient geometric condition on the affinity selection matrix that is equivalent to the statement that for any initial target there is a set of final nucleic acid fractions to which the SELEX iteration scheme must converge *independently of the composition of the initial nucleic acid pool* provided all nucleic acid types are in the initial pool.
4. This geometric condition is the multiple target analog of the condition in single target SELEX that the affinities of the members of a nucleic acid pool to a single target be distinct. It is equivalent to the condition that certain chemical potentials have a unique minimum at infinite target dilution.
5. The geometric relationship between initial target fraction vectors and final free target fraction vectors at infinite target dilution is discussed. This decomposition allows one to determine which nucleic acids are candidates for selection given the initial target vector. When the maximal target affinity function is proper, the set of final nucleic acid fractions is uniquely determined by the overall association constant and the initial target fractions when no member of the initial nucleic acid pool is absent.

Acknowledgements

The authors thank Professors Kahlid Boushaba and Michael Smiley for many useful comments.

Appendix A: The thermodynamic meaning of the chemical potential

We include a brief discussion of the chemical potential for readers who may not be familiar with the concept. Suppose that the overall reactions in Section 2.2 are not in equilibrium (ne). Denote the reciprocals of the ratios of concentrations on the extreme right-hand sides of (16), (18), (20) by $K_a^{\mathrm{ne}}, K_{a,j}^{\mathrm{ne}}, \kappa_{a,i}^{\mathrm{ne}}$. With each of the chemical equations in (15),

(19), there is associated a chemical potential μ, μ_i. (There is also a chemical potential for (17) but that does not concern us here.) These are

$$\mu - \mu^a = RT \ln\big(K_a^{\mathrm{ne}}/K_0\big), \qquad \mu_i - \mu_i^a = RT \ln\big(\kappa_{a,i}^{\mathrm{ne}}/K_0\big) \tag{A.1}$$

where the quantities μ^a, μ_i^a are the chemical potentials at equilibrium see Wall (1958). (The quantity K_0 is a reference value introduced to ensure the nondimensionality of the argument of the logarithm. It is usually taken to be unity in the units of the numerator. We adopt this convention here.) We set $\mu^a = -RT \ln(K_a/K_0)$, $\mu_i^a = -RT \ln(\kappa_{a,i}/K_0)$.

The physical meaning of these is well known to chemists. Consider, for example, the first of these. Then the chemical potential can be written as $\mu = RT \ln(K_a^{\mathrm{ne}}/K_a)$ and represents the change in free energy per mole (chemical potential) in going from a non-equilibrium state to equilibrium. If $K_a < K_a^{\mathrm{ne}}$, the chemical potential (free energy change) will be positive for the reaction as written in (15) and dissociation will be preferred. On the other hand, if $K_a > K_a^{\mathrm{ne}}$, the chemical potential will be negative for the reaction as written. In this case, association will be preferred. Similar remarks apply to the subreactions.

Thermodynamics dictates that the reaction proceeds most readily in the direction of association when the free energy change is at its most negative value or equivalently, when K_a has attained its largest value, or the target efficiency is as close to unity as possible. However, this does not mean that the overall $pK_a = -\log_{10} K_a = -pK_d$ will be a decreasing function of round number. (See Fig. 5.)

Appendix B: Convergence proof for the SELEX iteration scheme

Here, we establish that each of the (uniformly bounded) sequences $\{F_j^{(r)}\}_{r=1}^{\infty}$ converges to some limiting value, not all of which can be zero. This question is closely related to the issue of when certain series converge and when they diverge.

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

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

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

$$\sum_{p=1}^{\infty} \frac{|D_{l,f}^{(p)} - D_{j,f}^{(p)}|}{D_{l,f}^{(p)}(1 + D_{j,f}^{(p)})} = \sum_{p=1}^{\infty} \frac{|1 - D_{j,f}^{(p)}/D_{l,f}^{(p)}|}{(1 + D_{j,f}^{(p)})} \tag{B.2}$$

is divergent. Examining the tail-end of this series and noting that $D_{j,f}^{(p)} \to 0$ uniformly in j, we see that the divergence of the product is equivalent to the divergence of the series:

$$\sum_{p=1}^{\infty} \big|1 - D_{j,f}^{(p)}/D_{l,f}^{(p)}\big|. \tag{B.3}$$

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

To see the latter statement, fix some arbitrary index $l \in \mathcal{N}$. If for all $j \neq l$, the series in (B.3) diverged, then the sequences $\{F_j^{(r)}\}_{r=1}^{\infty}$ would all converge to zero. But then it follows from the normalization condition that the sequence $\{F_l^{(r)}\}_{r=1}^{\infty}$ converges to unity.

Suppose instead that for this l the set of all indices $\mathcal{L}_l$ for which the series in (B.3) is convergent, is not empty. We claim that $\lim_{r\to\infty} F_l^{(r)} > 0$. We write

$$1 = \sum_{j=1}^{N} F_j^{(r+1)} = F_l^{(r+1)} + \sum_{j\in\mathcal{L}_l} F_j^{(r+1)} + \sum_{j\notin\mathcal{L}_l} F_j^{(r+1)}.$$

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

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

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

Appendix C: Proof of Theorem 1

To prove Theorem 1, we need a few preliminary observations based on the system (42).

Suppose we have a solution of the system of equations

$$\overrightarrow{A^l} \cdot \widehat{\omega} = K_a \quad \text{for } l \in \mathcal{L}, \qquad \omega_i = \Omega_i / \left[\mathcal{W}\left(1 + [NA]\overrightarrow{A_i} \cdot \widehat{F}\right)\right] \quad \text{for } i \in \mathcal{M} \tag{C.1}$$

where $\mathcal{W} = 1/(1 + [NA]K_a)$, $\widehat{\omega} \in \mathcal{S}$, $\widehat{F} \in \mathcal{S}_{\mathcal{F},\mathcal{L}}$. Then it is straightforward to see that

$$\mathcal{W} = \sum_{i=1}^{M} \frac{\Omega_i}{1 + [NA]\overrightarrow{A_i} \cdot \widehat{F}} = \frac{1}{K_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il}}{1 + [NA]\overrightarrow{A_i} \cdot \widehat{F}}, \quad \text{for } l \in \mathcal{L}. \tag{C.2}$$

Conversely, if we have a vector $\widehat{F} \in \mathcal{S}_{\mathcal{F},\mathcal{L}}$ that solves the latter system, then we use the second set of equations (C.1) to define $\widehat{\omega}$. Substitution of these values into the second set of equations in (C.2) leads to the first set of equations in (C.1). In either case, $\mathcal{W} = 1 - E = 1/(1 + [NA]K_a)$.

Remark C.1. It can be shown using the Schauder fixed-point theorem that the system (C.2) always has at least one solution in $\widehat{F} \in \mathcal{S}_{\mathcal{F},\mathcal{L}}$ for any nonempty subset of indices $\mathcal{L} \subset \mathcal{N}$ and any $\widehat{\Omega} \in \mathcal{S}$ (Seo, 2010). The idea is to define a continuous map $\overrightarrow{g}$ on the unit N cube by $g_l(\widehat{F}) = 0$ if $l \notin \mathcal{L}$ and $g_l(\widehat{F}) = \frac{F_l}{\mathcal{W}K_a} \sum_{i=1}^{M} \frac{\Omega_i A_{il}}{1+[NA]\overrightarrow{A_i}\cdot\widehat{F}}$ if $l \in \mathcal{L}$. One easily checks that for such l, $0 \leq g_l \leq 1$ so that the map is into the N cube.

Remark C.2. Lagrange multipliers can be used to characterize the above stationary equations. Suppose for the moment that $\mathcal{L}$ is any subset of nucleic acid indices. When viewed as a function on the subsimplex $\mathcal{S}_{\mathcal{F},\mathcal{L}}$, we use the notation $\mathcal{Q}(\widehat{F}) = \mathcal{Q}(\widehat{F}, \mathcal{L})$ for the first form of $\mathcal{Q}$ in (44). Now let, for any set of nonnegative numbers F_l (say $\overrightarrow{F}$, not necessarily normalized to unity),

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

We always have

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

If we define $\mathcal{W}(\overrightarrow{F}) = 1 - E(\overrightarrow{F}) = \sum_{i=1}^{M}\Omega_i/(1+[NA]\overrightarrow{A}_i\cdot\overrightarrow{F})$, $K_a(\overrightarrow{F}) = (1-\mathcal{W})/([NA]\mathcal{W})$ by (C.2) and $\omega_i(\overrightarrow{F}) = \Omega_i/[\mathcal{W}(\overrightarrow{F})(1+[NA]\overrightarrow{A}_i\cdot\overrightarrow{F})]$ for any vector of nonnegative numbers F_l, and compute $\overrightarrow{A}^l\cdot\widehat{\omega}$, we see that

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

Thus, any solution of the system $\partial_{F_l}\mathcal{R} = 0$, $l\in\mathcal{L}$ is an extremal of $\mathcal{Q}$ subject to the constraint that $\partial_\lambda\mathcal{R} = \sum_{l\in\mathcal{L}}F_l - 1 = 0$ if and only if $\overrightarrow{A}^l\cdot\widehat{\omega} = \lambda K_a/E$. *Moreover, the indices will correspond to those for the maximal target efficiency function defined in* (38) *for this extremal if and only if* $\lambda = E$ *and the value of* K_a *is a given number in the range of this function.*

The Hessian matrix for $\mathcal{R}$ (or $\mathcal{Q}$) with respect to $\widehat{F}$ has the bilinear form:

$$\sum_{k,l\in\mathcal{L}}\xi_k\frac{\partial^2\mathcal{R}}{\partial_{F_k}\partial_{F_l}}\xi_l = [NA]^2\sum_{i=1}^{M}\frac{(\overrightarrow{\xi}\cdot\overrightarrow{A}_i)^2\Omega_i}{(1+[NA]\overrightarrow{A}_i\cdot\widehat{F})^2}. \tag{C.3}$$

Therefore, the eigenvalues of the Hessian are all strictly positive if and only if the set of vectors $\{\overrightarrow{A^l}\}$ defining the L face are linearly independent, i.e., the L face is proper. To see this, note that $\sum_{l\in\mathcal{L}}\xi_l\overrightarrow{A}^l = \langle\overrightarrow{\xi}\cdot\overrightarrow{A}_1,\ldots,\overrightarrow{\xi}\cdot\overrightarrow{A}_M\rangle^t$.

This in turn implies that the surface defined by $\mathcal{R}$ over the simplex $\mathcal{S}_{\mathcal{F},\mathcal{L}}$ must be strictly convex if and only if the L face of the maximal target affinity function is proper. Hence, there cannot be more than one critical point in $\mathcal{S}_{\mathcal{F},\mathcal{L}}$, i.e., there is at most one solution of (C.1) in $\mathcal{S}_{\mathcal{F},\mathcal{L}}$ whenever the L face is proper.

The proof of Theorem 2 is simply the statement that an L face is proper if and only if the chemical potential over that face has a unique minimum.

Thus, if φ *is proper, the limiting SELEX solution does not depend upon the composition of the initial pool of nucleic acids but rather only on the value of the limiting overall association constant,* K_a *and the initial target vector* $\widehat{\Omega}$.

Whether or not the face is improper, the initial target fractions Ω_i, the nucleic acid concentration pool size, $[NA]$ and the final fractions are connected through the equations

$$\Omega_i = \frac{\omega_i(1+[NA]\overrightarrow{A}_i\cdot\widehat{F})}{\sum_{j=1}^{M}\omega_j(1+[NA]\overrightarrow{A}_j\cdot\widehat{F})} = \frac{\omega_i(1+[NA]\overrightarrow{A}_i\cdot\widehat{F})}{1+[NA]K_a} \tag{C.4}$$

where only those F_l with l in $\mathcal{L}$ contribute to the indicated dot products. Equations (C.4) in vector form are

$$\widehat{\Omega} = \mathcal{W}\left(\sum_{l\in\mathcal{L}} F_l\left\{\widehat{\omega}+[NA]\overrightarrow{A^l\omega}\right\}\right) \tag{C.5}$$

i.e., the starting fraction vector must be a convex combination of the unit vectors $\widehat{V}^l = \widehat{V}^l(\widehat{\omega}) \equiv \mathcal{W}(\widehat{\omega}+[NA]\overrightarrow{A^l\omega})$.[7] The formula tells us that, if we know the final free target fractions, the final nucleic acid fractions can be uniquely determined from $\widehat{\Omega}$ and K_a *if and only if the vectors* $\{\widehat{\omega}+[NA]\overrightarrow{A^l\omega},\ l\in\mathcal{L}\}$ *are linearly independent* and this is true if and only if the set of column vectors $\{\overrightarrow{A^l\omega}\,|l\in\mathcal{L}\}$ is linearly independent.[8] This in turn is true if and only if $\widehat{\omega}\in\mathcal{S}_0$ and the set of vectors $\{\overrightarrow{A}^{l_1},\ldots,\overrightarrow{A}^{l_L}\}$ that define the columns of the affinity selection matrix $A_{\mathcal{L}}$ is linearly independent.

When this is the case, the final fractions are given by

$$\widehat{F} = \begin{pmatrix} \widehat{V}^{l_1}\cdot\widehat{V}^{l_1} & \widehat{V}^{l_1}\cdot\widehat{V}^{l_2} & \cdots & \widehat{V}^{l_1}\cdot\widehat{V}^{l_L} \\ \widehat{V}^{l_2}\cdot\widehat{V}^{l_1} & \widehat{V}^{l_2}\cdot\widehat{V}^{l_2} & \cdots & \widehat{V}^{l_2}\cdot\widehat{V}^{l_L} \\ \vdots & \vdots & \cdots & \vdots \\ \widehat{V}^{l_L}\cdot\widehat{V}^{l_1} & \widehat{V}^{l_L}\cdot\widehat{V}^{l_2} & \cdots & \widehat{V}^{l_L}\cdot\widehat{V}^{l_L} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Omega}\cdot\widehat{V}^{l_1} \\ \widehat{\Omega}\cdot\widehat{V}^{l_2} \\ \vdots \\ \widehat{\Omega}\cdot\widehat{V}^{l_L} \end{pmatrix}. \tag{C.6}$$

The inverse of the Grammian on the right exists if and only if the vectors $\widehat{V}^i$ are linearly independent. The components of $\widehat{F}$ are nonnegative and sum to unity if and only if $\widehat{\Omega}$ is in the convex hull of the $\widehat{V}^l$'s.

If $\widehat{\omega}$ belongs to an improper face, then we are led to an underdetermined system of linear equations for the nucleic acid fractions. The meaning of this is that we *cannot* determine the final nucleic acid fractions from the stationary system (42) even when we know the final free target vector. That is, when φ is improper, the final nucleic acid fractions will depend, in general, not only upon the initial target fractions *but also upon the initial pool of nucleic acids.*

[7] Because all of the entries of $\widehat{\omega}+[NA]\overrightarrow{A^l\omega}$ are nonnegative $|\widehat{\omega}+[NA]\overrightarrow{A^l\omega}| = 1+[NA]K_a = 1/\mathcal{W}$.

[8] It is not too hard to see that this set is linearly independent if and only if the set $\{\overrightarrow{A^l\omega},\ l\in\mathcal{L}\}$ is linearly independent. For if $\sum_{l\in\mathcal{L}}\lambda_l(\widehat{\omega}+[NA]\overrightarrow{A^l\omega}) = \vec{0}$, then by summing the components these equations we find that $\sum_{l\in\mathcal{L}}\lambda_l(1+K_a)=0$ because $|\overrightarrow{A^l\omega}| = \overrightarrow{A^l}\cdot\vec{\omega} = K_a$. But then $\sum_{l\in\mathcal{L}}\lambda_l[NA]\overrightarrow{A^l\omega} = \vec{0}$. Therefore, if the latter set is linearly independent, so is the former. Likewise, if the former set is linearly independent, and $\sum_{l\in\mathcal{L}}\lambda_l[NA]\overrightarrow{A^l\omega} = \vec{0}$, it again follows that $\sum_{l\in\mathcal{L}(\widehat{\omega})}\lambda_l = \vec{0}$. Consequently, $\sum_{l\in\mathcal{L}}\lambda_l(\widehat{\omega}+[NA]\overrightarrow{A^l\omega}) = \vec{0}$ and the linear independence of the former set implies that the $\lambda_l = 0$.

To establish the claim when the face is improper, we give an example. Suppose $L = 3(= N)$ and $M = 2$ so that we have just two targets. Let $\mathcal{N} = \{l_1, l_2, l_3\} = \{1, 2, 3\}$. Suppose the columns of $A_{\mathcal{N}}$, $\{\overrightarrow{A}^1, \overrightarrow{A}^2, \overrightarrow{A}^3\}$, form a linearly dependent set but any pair of elements in this set are linearly independent.

Suppose there is a single vector $\widehat{\omega}$ such that $\varphi(\widehat{\omega}) = \overrightarrow{A}^1 \cdot \widehat{\omega} = \overrightarrow{A}^2 \cdot \widehat{\omega} = \overrightarrow{A}^3 \cdot \widehat{\omega} = K_a$, but $\varphi(\widehat{\omega}') = \max\{\overrightarrow{A}^1 \cdot \widehat{\omega}', \overrightarrow{A}^2 \cdot \widehat{\omega}'\} > \overrightarrow{A}^3 \cdot \widehat{\omega}'$ when $\widehat{\omega}' \neq \widehat{\omega}$. Then $\mathcal{L}(\widehat{\omega}) = \{1, 2, 3\}$. There are two 1 faces defined by the indices 1, 2 respectively. However the pair $(\widehat{\omega}, \varphi(\widehat{\omega}))$ defines a two face. This two face is not proper because three sets $\{1, 2\}, \{1, 3\}, \{2, 3\}$ define its indices and all are proper subsets of $\{1, 2, 3\} = \mathcal{L}$.

Because the vectors $\overrightarrow{A^1}, \overrightarrow{A^2}, \overrightarrow{A^3}$ are pairwise linearly independent, so are the vectors $\widehat{V}^1, \widehat{V}^2, \widehat{V}^3$. We write $\widehat{V}^3 = \lambda_1 \widehat{V}^1 + \lambda_2 \widehat{V}^2$. (The λ_i' s are found as in (C.6). It suffices to notice that $\lambda_1 + \lambda_2 = 1$ because the components of $\widehat{V}^3$ are nonnegative and sum to unity.) Using $F_1 + F_2 + F_3 = 1$, we have $\widehat{\Omega} = F_1 \widehat{V}^1 + F_2 \widehat{V}^2 + F_3 \widehat{V}^3 = (F_1 + \lambda_1 F_3)\widehat{V}^1 + (F_2 + \lambda_2 F_3)\widehat{V}^2$. Again using (C.6), we find the values of $F_i + \lambda_i F_3 = g_i$ uniquely. Taking the one norm of both sides of this equation, we see that $g_1 + g_2 = 1$ and, therefor, at least one of the g_i is positive. Adding equations, $F_1 + \lambda_1 F_3 = g_1$, $F_2 + \lambda_2 F_3 = g_2$ and eliminating $F_1 + F_2$ between the resulting equation and $F_1 + F_2 + F_3 = 1$, we find that $(\lambda_1 + \lambda_2 - 1)F_3 = g_1 + g_2 - 1 = 0$, and hence F_3 cannot be found from the stationary equations even when $\widehat{\omega}$ is known. Moreover, all the stationary solutions are of the form $\widehat{F}_s(t) \equiv \langle g_1 - \lambda_1 t, g_2 - \lambda_2 t, t\rangle = \langle F_1, F_2, F_3\rangle$ for t in some subinterval of $(0, 1)$, namely that subinterval for which $g_1 - \lambda_1 t \geq 0$ and $g_2 - \lambda_2 t \geq 0$.

If the L face is improper, the Hessian of the chemical potential corresponding to it must have its smallest eigenvalue vanish. Therefore each set of stationary solutions (in our example, the vector family $\{\widehat{F}_s(t)\}$) minimizes the chemical potential at infinite target dilution, i.e., is a realizable thermodynamic state. Thus, with such a state as an initial state, with the given $\widehat{\Omega}$, we obtain another such state as a final state.

Remark C.3. Geometric properties of the family of convex hulls generated by a single free target vector. Let $\mathcal{H} = \mathcal{H}([NA], \widehat{\omega}) \subset \mathcal{S}$ denote the aforementioned convex hull. Suppose the vectors $\widehat{V}^i$ form a linearly independent set. Formula (C.5) has some interesting geometric consequences. First, the dimensionality of the convex hull of the set of unit vectors $\{\widehat{V}^l, l \in \mathcal{L}\}$ is precisely $L - 1$. Therefore, the largest (in dimensionality) sets of initial targets come from those $\widehat{\omega}$ that yield an $M \times M$ SELEX matrix with full rank M, i.e., to those final free targets that correspond to the selection of M nucleic acids.

Second, the diameter of $\mathcal{H}$ is the number $(\mathcal{H}) \equiv \max\{|\widehat{V}^l - \widehat{V}^m| \quad l, m \in \mathcal{L}\} = [NA]\max\{|\overrightarrow{A^l\omega} - \overrightarrow{A^m\omega}| \quad l, m \in \mathcal{L}\}/(1 + [NA]K_a)$ for all indices $i = 1, \ldots, M$. This diameter is an increasing function of $[NA]$ which vanishes at $[NA] = 0$ and has the limiting value $\max\{|\widehat{A^l\omega_l} - \widehat{A^l\omega_m}| \; l, m \in \mathcal{L}\}/K_a$ as $[NA] \to \infty$. We note that the vectors $\overrightarrow{A^l\omega}/K_a = \overrightarrow{A^l\omega}/\varphi(\widehat{\omega}) = \overrightarrow{A^l\omega}/(\overrightarrow{A}^l \cdot \widehat{\omega}) = \overrightarrow{A^l\omega}/|\overrightarrow{A^l\omega}| = \widehat{A^l\omega}$ are unit vectors with positive entries. Also, each vertex $\widehat{V}^l$ to converges to $\widehat{A^l\omega}$ as $[NA] \to \infty$. Moreover, $\widehat{\Omega} = \widehat{\omega}$ as $[NA] \to 0$. That is, the convex hull, $\mathcal{H} = \{\widehat{\omega}\}$.

Third, if we make explicit the dependence of each vertex of the convex hull $\mathcal{H}([NA], \widehat{\omega})$ on $[NA]$ by writing $\widehat{V}^l = \widehat{V}^l([NA])$, then it is easy to see that $\widehat{V}^l([NA]) = \mathcal{W}\widehat{V}^l([0]) + (1 - \mathcal{W})\widehat{V}^l(\infty) = \mathcal{W}([NA])\widehat{\omega} + (1 - \mathcal{W}([NA]))\widehat{A^l\omega}$. Therefore, the family of convex hulls forms an increasing family of sets, i.e., if $[NA], [NA]'$ are two pool concentrations with $0 < [NA] < [NA]'$, $\{\widehat{\omega}\} \subset \mathcal{H}([NA], \widehat{\omega}) \subset \mathcal{H}([NA]', \widehat{\omega}) \subset \mathcal{S}$.

Remark C.4. From (C.6), we see that as $[NA] \to \infty$, $\widehat{\Omega} = \sum_{l\in\mathcal{L}} F_l \overrightarrow{A^l\omega}/K_a = \sum_{l\in\mathcal{L}} F_l \widehat{A^l\omega}$ where

$$\widehat{F} \to \begin{pmatrix} \widehat{A^{l_1}\omega}\cdot\widehat{A^{l_1}\omega} & \widehat{A^{l_1}\omega}\cdot\widehat{A^{l_2}\omega} & \dots & \widehat{A^{l_1}\omega}\cdot\widehat{A^{l_L}\omega} \\ \widehat{A^{l_2}\omega}\cdot\widehat{A^{l_1}\omega} & \widehat{A^{l_2}\omega}\cdot\widehat{A^{l_2}\omega} & \dots & \widehat{A^{l_2}\omega}\cdot\widehat{A^{l_L}\omega} \\ \vdots & \vdots & \dots & \vdots \\ \widehat{A^{l_L}\omega}\cdot\widehat{A^{l_1}\omega} & \widehat{A^{l_L}\omega}\cdot\widehat{A^{l_2}\omega} & \dots & \widehat{A^{l_L}\omega}\cdot\widehat{A^{l_L}\omega} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\Omega}\cdot\widehat{A^{l_1}\omega} \\ \widehat{\Omega}\cdot\widehat{A^{l_2}\omega} \\ \vdots \\ \widehat{\Omega}\cdot\widehat{A^{l_L}\omega} \end{pmatrix}. \quad \text{(C.7)}$$

We rewrite (C.4) as

$$\begin{aligned} \widehat{\Omega} &= \mathcal{W}\left(\widehat{\omega} + [NA]\sum_{l\in\mathcal{L}} F_l \overrightarrow{A^l\omega}\right) = \mathcal{W}\widehat{\omega} + [NA]K_a\mathcal{W}\sum_{l\in\mathcal{L}} F_l \overrightarrow{A^l\omega}/K_a \\ &= \mathcal{W}\widehat{\omega} + (1-\mathcal{W})\sum_{l\in\mathcal{L}} F_l \widehat{A^l\omega} \end{aligned} \quad \text{(C.8)}$$

where $\widehat{F} \in \mathcal{S}_{\mathcal{F},\mathcal{L}}$. Therefore, every starting target fraction vector that can reach $\widehat{\omega}$ can be expressed as a convex combination of a free target vector, $\widehat{\omega}$, and the unit affinity vectors $\widehat{A^l\omega}$ with $l \in \mathcal{L}$.

We illustrate these comments with two examples.

1. Suppose $L = 1$. In this case, $\mathcal{L}$ must be a single positive integer, $l_1 = 1$ say. Then $F_j = 0$ unless $j = 1$ and $F_1 = 1$. The set of initial target fractions that give this single nucleic acid are given by $\widehat{\Omega} = \mathcal{W}(\widehat{\omega} + [NA]\overrightarrow{A^1\omega})$ provided that $\varphi(\widehat{\omega}) = \overrightarrow{A}^1\cdot\widehat{\omega} = K_a > \overrightarrow{A}^j\cdot\widehat{\omega}$ for $j \neq 1$. Thus, there is only one initial target vector that will select uniquely for nucleic acid 1 with overall equilibrium constant K_a. It is easy to see that $\omega_i = \omega_i([NA]) = (1 + K_a[NA])\Omega_i/(1 + [NA]A_{i1}) = [\Omega_i/(1 + A_{i1}[NA])]/[\sum_{j=1}^M \Omega_j/(1 + A_{j1}[NA])]$. The association constant is then

$$K_a = \left(\sum_{i=1}^M \Omega_i A_{i1}/\big(1 + A_{i1}[NA]\big)\right) \Big/ \left(\sum_{j=1}^M \Omega_j/\big(1 + A_{j1}[NA]\big)\right).$$

This formula holds only over the range of total pool size $[NA]$ for which $\mathcal{L}(\widehat{\omega}([NA])) = \{1\}$. As K_a decreases, we experience a jump in this index set when two hyperplanes intersect.

2. Suppose $L = M$ and $M \leq N$. Suppose moreover that $\mathcal{L} = \{l_1, \dots, l_M\} = \{1, \dots, M\}$ and $\{\overrightarrow{A}^1, \dots, \overrightarrow{A}^M\}$ is a linearly independent set. We have $\varphi(\widehat{\omega}) = \overrightarrow{A}^1\cdot\widehat{\omega} = \cdots = \overrightarrow{A}^M\cdot\widehat{\omega} = K_a > \overrightarrow{A}^j\cdot\widehat{\omega}$ for $j \neq 1, \dots, M$. It is not hard to show from Cramer's rule for solving linear systems that

$$\omega_j = K_a \frac{\det\begin{pmatrix} A_{11} & \dots & A_{j-1,1} & 1 & A_{j+1,1} & \dots & A_{M1} \\ A_{12} & \dots & A_{j-1,2} & 1 & A_{j+1,2} & \dots & A_{M2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ A_{1,M-1} & \dots & A_{j-1,M-1} & 1 & A_{j+1,M-1} & \dots & A_{M,M-1} \\ A_{1M} & \dots & A_{j-1,M} & 1 & A_{j+1,M} & \dots & A_{MM} \end{pmatrix}}{\det(A_{\mathcal{L}_M})} \quad \text{(C.9)}$$

for $j = 1, \ldots, M$. The numerator is clearly the column sum of the elements of the jth column of the classical adjoint of the matrix $A_{\mathcal{L}_M}$. Thus, $\widehat{\omega}$ will be well defined if and only if

(a) All the column sums of the classical adjoint of the SELEX matrix $A_{\mathcal{L}_M}$ have the same sign as the determinant of this matrix.
(b) The overall association constant is given by

$$K_a = \det(A_{\mathcal{L}_M})/\Delta A \tag{C.10}$$

where $\Delta A \equiv \sum_{j=1}^{M} \sum_{i=1}^{M} \mathbf{A}_{ij}$ denotes the sum of the cofactors $\mathbf{A}_{ij}$ of $A_{\mathcal{L}_M}$.

Using these results, we compute the final nucleic fractions from (C.6).

Finally, there may be several vertices for the graph of φ, i.e., several M faces. For such M faces, the association constants *do not* depend on the nucleic acid pool size $[NA]$. However, the final nucleic acid fractions, $F_i, i = 1, \ldots, M$, generally do depend upon the nucleic acid pool size, $[NA]$, via the system of equations given in (C.6) because the vectors $\widehat{V}_i$ do.

Appendix D: A special solution of the stationary SELEX equations

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

Another way of formulating this question is to ask whether or not there is a choice of $\widehat{\Omega}$ such that the right-hand side of (C.6) does not depend upon the nucleic pool size although the vectors $\widehat{V}_i$ do depend on it.

Clearly, when $[NA] = 0$, $\widehat{\Omega} \in \mathcal{H}(0, \widehat{\omega})$ for any $\widehat{\omega}$. For other values of $[NA]$, this choice holds for $\widehat{\Omega}$ if and only if

$$\overrightarrow{A}_i \cdot \widehat{F} = K_a \quad \text{for all indices } i = 1, \ldots, M. \tag{D.1}$$

This system arises naturally if we try to minimize the chemical potential in the form given in (43) with respect to the partial energies E_i subject to the constraint that the total (weighted) energy is fixed. We see this as follows. Using Lagrange multipliers to minimize $\mathcal{Q}$ subject to the constraint that $E = \sum \Omega_i E_i$ is fixed yields $E_i = E$. This extreme point is unique and, therefore, $E_i = [NA]\overrightarrow{A}_i \cdot \widehat{F}/(1 + [NA]\overrightarrow{A}_i \cdot \widehat{F}) = E$ and $[NA]\overrightarrow{A}_i \cdot \widehat{F} = E/(1 - E)$. Now suppose that our set of final fractions $\widehat{F} \in \mathcal{S}_{\mathcal{F},\mathcal{L}}$ satisfies (D.1). Then $E = [NA]K_a/(1 + [NA]K_a)$ and $\mathcal{W} = 1 - E = 1/(1 + [NA]K_a)$. Therefore, we must have $\omega_i = \Omega_i/(\mathcal{W}(1 + [NA]K_a))$, and hence $\omega_i = \Omega_i$.

Using Eqs. (D.1) and again invoking Cramer's rule as we did for (42) with $A_{\mathcal{L}_M}$ replacing $\mathcal{A}^t_{\mathcal{L}_M}$, we obtain

$$F_i = K_a \frac{\sum_{j=1}^{M} \mathbf{A}_{ij}}{\det(A_{\mathcal{L}_M})} = \frac{\det(A_{\mathcal{L}_M})}{\Delta A} \frac{\sum_{j=1}^{M} \mathbf{A}_{ij}}{\det(A_{\mathcal{L}_M})} = \frac{\sum_{j=1}^{M} \mathbf{A}_{ij}}{\Delta A}. \tag{D.2}$$

These sum to unity. The set of fractions so obtained must be nonnegative. However, this will be the case if the column sums of the classical adjoint are nonnegative. *Therefore, there is a starting target fraction vector* $\widehat{\Omega}$ *and a final free target fraction vector* $\widehat{\omega}$ *that does not depend on the total pool concentration* [*NA*] *if and only if the SELEX matrix has the property that the row sums and column sums of its classical adjoint all have the same sign as its determinant. In this case,* $\widehat{\Omega} = \widehat{\omega}$.

For any finite value of [*NA*], the efficiency is

$$E_{\text{final}} = 1 - \sum_{i=1}^{M} \mathcal{W}_i = \frac{[NA]\det(A_{\mathcal{L}_M})}{\Delta A + [NA]\det(A_{\mathcal{L}_M})} = \frac{[NA]}{\Delta A/\det(A_{\mathcal{L}_M}) + [NA]}. \tag{D.3}$$

Remark D.5. There is a geometric meaning to the condition that the row sums as well as the column sums of the classical adjoint of the SELEX matrix are positive. This condition is equivalent to the condition that the column sums of the classical adjoint of the matrix $A^t_{\mathcal{L}_M}$ have the same sign as the determinant of this matrix. Suppose, without loss of generality that $\det(A_{\mathcal{L}_M}) > 0$. The M hyperlines that form the one dimensional edges of the surface $z = \varphi(\widehat{\omega}')$ near the minimum $\widehat{\omega}$ are given by the parametric equations (for fixed $j = 1, \ldots, M$) $t = (\omega_i - \omega_{i,m})/\mathbf{A}_{ij}$ for $i = 1, \ldots, M$ where t is the free parameter for the line and where $\mathbf{A}_{ij}$ is defined above. That is, the vector $\overrightarrow{B}_j = \langle \mathbf{A}_{1j}, \ldots, \mathbf{A}_{Mj} \rangle$ is (up to a scalar) the direction vector for the jth edge. The vector $\widehat{I} = \langle 1, \ldots, 1 \rangle / \sqrt{M}$ is the outer normal to $\mathcal{S}$. The condition that the minimum of φ occur at $\widehat{\omega_m}$ is equivalent to the condition that $\overrightarrow{B}_j \cdot \widehat{I} > 0$, i.e., that the row sums of the classical adjoint of $\mathcal{A}^t_{\mathcal{L}_M}$ are all positive. This tells us that the row sums of the classical adjoint $\mathcal{A}^t_{\mathcal{L}_M}$ are positive if and only if the vector $\widehat{\omega}$ whose components are given in (C.9), are well defined and are such that $\varphi(\widehat{\omega}) = \varphi_{\min}$.

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

Remark D.6. When $M = 2$ the classical adjoint of $A^t_{\mathcal{L}_2}$ is $\left[\begin{smallmatrix} A_{22} & -A_{21} \\ -A_{12} & A_{11} \end{smallmatrix}\right]$. Suppose for the moment that $A^t_{\mathcal{L}_2}$ has a positive determinant. Then a necessary and sufficient condition for the row and column sums of the classical adjoint to be positive is that $\max\{A_{12}, A_{21}\} \leq \min\{A_{11}, A_{22}\}$.

The above result becomes

$$\begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \frac{1}{\Delta A} \begin{pmatrix} A_{22} - A_{21} \\ A_{11} - A_{12} \end{pmatrix}$$
$$\text{with } \widehat{F} = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \frac{1}{\Delta A} \begin{pmatrix} A_{22} - A_{12} \\ A_{11} - A_{21} \end{pmatrix}. \tag{D.4}$$

Remark D.7. Consider the case for which the graph of φ has only one vertex. Suppose the formulas (C.9), (C.10) for the components of $\widehat{\omega}$ and K_a are in force. Let $\widehat{\Omega} \neq \widehat{\omega}$ be a vector in $\mathcal{S}$. If $|\widehat{\Omega} - \widehat{\omega}| \geq \text{diam}(\mathcal{H}(\infty))$, then no selection for any of the nucleic acids with indices in $\mathcal{L}$ is possible *for this value of the final free target vector,* $\widehat{\omega}$. If the inequality fails, then whether or not this outcome occurs depends upon the orientation of

the hulls $\mathcal{H}([NA])$ relative to the vector $\widehat{\Omega} - \widehat{\omega}$. If the inequality holds, then there is a smallest value of $[NA]$, say $[NA]_{\min}$ for which it holds and $\widehat{\Omega}$ will be a boundary point of $\mathcal{H}([NA]_{\min})$. This boundary point will belong to some $k < M$ dimensional intersection of the hyperplanes (say those labeled by indices in a subset $\mathcal{K} = \{l_1, \ldots, l_k\} \subset \mathcal{L}_M$) that serve to define the graph of φ. At such a point, k of the SELEX fractions will be positive. As $[NA]$ increases, $\widehat{\Omega}$ will become an interior point of the hulls $\mathcal{H}([NA])$.

Appendix E: Numerical values used to generate the figures

We used the matrix $A = A(1:5, 1:20)$ given in (E.1), (E.2) below: to generate Figs. 1–5:

$$A(1:5, 1:10) = \begin{bmatrix} 822.37 & 618.81 & 521.92 & 984.25 & 759.88 & 1938.0 & 3164.6 & 1623.4 & 4629.6 & 2403.8 \\ 2403.8 & 1091.7 & 1396.6 & 659.63 & 521.92 & 1225.5 & 706.21 & 8620.7 & 4629.6 & 1623.4 \\ 4629.6 & 759.88 & 521.92 & 706.21 & 582.75 & 3164.6 & 984.25 & 550.66 & 1938.0 & 1396.6 \\ 2403.8 & 1225.5 & 3164.6 & 984.25 & 1091.7 & 521.92 & 582.75 & 1396.6 & 4629.6 & 8620.7 \\ 759.88 & 896.06 & 659.63 & 1623.4 & 1225.5 & 1091.7 & 618.81 & 8620.7 & 984.25 & 582.75 \end{bmatrix}, \tag{E.1}$$

$$A(1:5, 11:20) = \begin{bmatrix} 896.06 & 550.66 & 1225.5 & 1396.6 & 496.03 & 8620.7 & 659.63 & 706.21 & 582.75 & 1091.7 \\ 759.88 & 896.06 & 496.03 & 1938.0 & 984.25 & 822.37 & 618.81 & 3164.6 & 582.75 & 550.66 \\ 1091.7 & 8620.7 & 496.03 & 618.81 & 896.06 & 1623.4 & 822.37 & 1225.5 & 2403.8 & 659.63 \\ 896.06 & 706.21 & 659.63 & 822.37 & 1623.4 & 1938.0 & 759.88 & 496.03 & 618.81 & 550.66 \\ 706.21 & 822.37 & 550.66 & 496.03 & 2403.8 & 4629.6 & 3164.6 & 1396.6 & 521.92 & 1938.0 \end{bmatrix}. \tag{E.2}$$

The values for the first row were chosen randomly over the range that corresponds to the range of values used in Irvine et al. (1991). Each of the remaining rows were obtained from the first by doing a random reordering of the values of the first row.

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

$$A_{\mathcal{L}_5} = \begin{bmatrix} 1623.4 & 4629.6 & 2403.8 & 550.66 & 8620.7 \\ 8620.7 & 4629.6 & 1623.4 & 896.06 & 822.37 \\ 550.66 & 1938 & 1396.6 & 8620.7 & 1623.4 \\ 1396.60 & 4629.6 & 8620.7 & 706.21 & 1938 \\ 8620.7 & 984.25 & 582.75 & 822.37 & 4629.6 \end{bmatrix}. \tag{E.3}$$

The classical adjoint of the preceding matrix is

$$\mathrm{adj}(A_{\mathcal{L}_5}) = 10^{15} \begin{bmatrix} -1.1585 & 0.7249 & -0.2051 & 0.0800 & 2.0669 \\ 2.4628 & 4.6087 & -0.0722 & -1.2136 & -4.8713 \\ -1.52016 & -2.1133 & -0.1316 & 3.2899 & 1.8750 \\ -0.5970 & -0.3636 & 2.6343 & -0.2150 & 0.3426 \\ 1.9310 & -1.9990 & -0.0542 & -0.2669 & 1.6821 \end{bmatrix}. \tag{E.4}$$

For Fig. 6, we used the matrix:

$$A = \begin{bmatrix} 4629.6 & 1396.6 & 1623.4 & 3164.6 & 1091.7 \\ 3164.6 & 4629.6 & 1938.0 & 1623.4 & 1225.5 \\ 1091.7 & 1623.4 & 4629.6 & 1225.5 & 2403.8 \end{bmatrix}. \tag{E.5}$$

For Fig. 7 panels (a), (b) and Fig. 8, we used

$$A = \begin{bmatrix} 4629 & 1396.6 & 1623.4 & 3386.5 & 4420.2 \\ 3164.6 & 4629.6 & 1938 & 4800 & 1925.5 \\ 1091.7 & 1623.4 & 4630 & 2445 & 2103.8 \end{bmatrix}. \tag{E.6}$$

For Fig. 7 panels (c), (d) and Fig. 9, we used

$$A = \begin{bmatrix} 9355 & 5529 & 1987 & 7468 & 846.2 \\ 916.9 & 8132 & 3038 & 6451 & 5252 \\ 4993.4 & 990 & 9722 & 931.8 & 2026 \end{bmatrix}. \tag{E.7}$$

References

Chen, C., 2007. Complex SELEX against target mixture: stochastic computer model, simulation and analysis. Comput. Methods Programs Biomed. 8, 189–200.

Chen, C., Kuo, T., 2007. Simulations of SELEX against complex receptors with a condensed statistical model. Comput. Chem. Eng. 31, 1007–1019.

Chen, C., Kuo, T., Chan, P., Lin, L., 2007. Subtractive SELEX against two heterogeneous target samples: numerical simulations and analysis. Comput. Biol. Med. 37, 750–759.

Irvine, D., Tuerk, C., Gold, L., 1991. SELEXION, systemic evolution of nucleic acids by exponential enrichment with integrated optimization by non-linear analysis. J. Mol. Biol. 222, 739–761.

Levine, H.A., Nilsen-Hamilton, M., 2007. Mathematical analysis of SELEX. Comput. Biol. Chem. 31, 11–25.

Seo, Y.-J., 2010. Ph.D. Thesis, Iowa State University.

Sun, F., Galas, D., Waterman, M.S., 1996. A mathematical analysis of in vitro molecular selection-amplification. J. Mol. Biol. 258(4), 650–660.

Vant-Hull, B., Payano-Baez, A., Davis, R.H., Gold, L., 1998. The mathematics of SELEX against complex targets. J. Mol. Biol. 278, 579–597.

Wall, F.T., 1958. Chemical Thermodynamics. Freeman, New York.