analytical chemistry

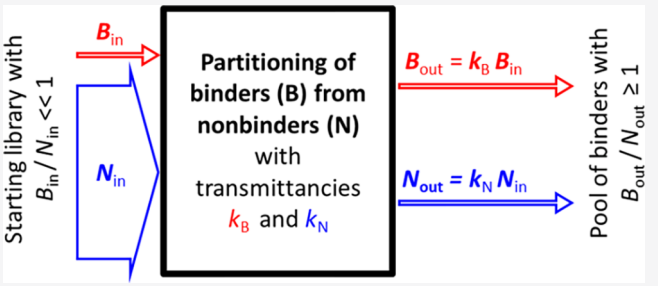
pubs.acs.org/ac Perspective

How to Develop and Prove High-Efficiency Selection of Ligands from Oligonucleotide Libraries: A Universal Framework for Aptamers and DNA-Encoded Small-Molecule Ligands

An T. H. Le, Svetlana M. Krylova, Stanislav S. Beloborodov, Tong Y. Wang, Ryan Hili, Philip E. Johnson, Feng Li, Rakesh N. Veedu, Svetlana Belyanskaya, and Sergey N. Krylov*

Cite This: https://doi.org/10.1021/acs.analchem.1c00601 Read Online

ABSTRACT: Screening molecular libraries for ligands capable of binding proteins is widely used for hit identification in the early drug discovery process. Oligonucleotide libraries provide a very high diversity of compounds, while the combination of the polymerase chain reaction and DNA sequencing allow the identification of ligands in low copy numbers selected from such libraries. Ligand selection from oligonucleotide libraries requires mixing the library with the target followed by the physical separation of the ligand−target complexes from the unbound library. Cumulatively, the low abundance of ligands in the library and the low efficiency of available separation methods necessitate multiple consecutive rounds of partitioning. Multiple rounds of inefficient partitioning make the selection process ineffective and prone to failures. There are continuing efforts to develop a separation method capable of reliably generating a pure pool of ligands in a single round of partitioning; however, none of the proposed methods for single-round selection have been universally adopted. Our analysis revealed that the developers' efforts are disconnected from each other and hindered by the lack of quantitative criteria of selection quality assessment. Here, we present a formalism that describes single-round selection mathematically and provides parameters for quantitative characterization of selection quality. We use this formalism to define a universal strategy for development and validation of single-round selection methods. Finally, we analyze the existing partitioning methods, the published single-round selection reports, and some pertinent practical considerations through the prism of this formalism. This formalism is not an experimental protocol but a framework for correct development of experimental protocols. While single-round selection is not a goal by itself and may not always suffice selection of good-quality ligands, our work will help developers of highly efficient selection approaches to consolidate their efforts under an umbrella of universal quantitative criteria of method development and assessment.



Screening combinatorial libraries of molecules for ligands capable of binding therapeutic targets has become the mainstream approach for hit identification in the early drug discovery process.[1−3] Probability rules suggest that the number of different ligands increases with the increasing structural diversity of the library.[4] Discrete chemical libraries, in which every molecule is synthesized, stored, and tested for target binding separately, are limited to approximately a million different compounds, and they are not a subject of our consideration here.[5,6] Combinatorial mixtures, which are synthesized via the split and mix approach, can provide virtually unlimited diversity. Highly diverse combinatorial mixtures are the sole focus of this work, and accordingly, the term "library" refers to a highly diverse library mixture unless otherwise stated.

Finding ligands in a library requires (1) mixing the library with the target, (2) physically separating target-bound molecules (binders) from target-unbound molecules (non-binders), and (3) identifying the separated binders.[7−9] Increasing library diversity inevitably reduces the number of copies of each unique molecule in the library and, thus, makes binder identification more difficult. The product of physical isolation of binders from a library is a complex mixture of many different binders contaminated by many different nonbinders, each present in a small number of copies.[10] Classic structure-identification tools, such as nuclear magnetic resonance and mass spectrometry, cannot be applied to such complex mixtures with extremely low copy numbers. A solution for

Received: February 8, 2021
Accepted: March 16, 2021

this problem is the use of oligonucleotide libraries in combination with the analytical tools of molecular biology.[7,9] The oligonucleotide moieties of molecules selected from oligonucleotide libraries can be amplified by the polymerase chain reaction (PCR), and their sequences can be identified via DNA sequencing and serve as binder identities. The amplifying power of PCR allows the identification of binders selected even in single copies.

There are two major types of oligonucleotide libraries: random-sequence libraries, which are used for the selection of DNA and RNA aptamers,[11,12] and DNA-encoded libraries (DELs),[13] which are used for selecting small-molecule ligands. Identifying oligonucleotide sequences of selected aptamers facilitates chemical synthesis of their replicas. Sequence information on DNA moieties of binders isolated from DELs is used to deconvolute the structural information on small-molecule binders, and the binders are then synthesized one-by-one without DNA tags. A typical random region in an oligonucleotide library for aptamer selection is 40-nt long resulting in library diversity from 4^{40} to 10^{24} unique sequences. DELs are typically less diverse, but there are examples of DELs composed of as many as 10^{12} different compounds.[13] The enormous sequence and structure diversities in oligonucleotide libraries along with the availability of adequate screening tools makes such libraries very attractive for *in vitro* selection of affinity probes and hits in early stage drug discovery.[14,15]

Efficient physical partitioning of binders from nonbinders constitutes a major challenge for the field of binder selection from oligonucleotide libraries. The abundance of binders in a naïve (unbiased for a specific target) oligonucleotide library may be as low as one suitable binder per million or billion library molecules. Separating the few binders from the overwhelming majority of nonbinders in a single partitioning step is limited by a nonbinder background.[16−19] Consequently, multiple rounds of partitioning are typically used for *in vitro* selection of binders from oligonucleotide libraries.

Multiround selection of binders from oligonucleotide libraries has fundamental drawbacks. Oligonucleotide aptamers are typically obtained by the SELEX process, which involves iterative cycles of binding/partitioning and enrichment by PCR amplification and theoretically enables an unlimited number of rounds.[16−19] In reality, a large number of rounds may not be able to compensate for inefficient partitioning if the sequence bias of PCR toward the amplification of nonbinders is greater than the preference of partitioning for the isolation of binders.[20−22] On the other hand, single-round selection, which may force dealing with single copy numbers, requires high-fidelity polymerases to avoid the loss of binders due to deleterious mutations during amplification.

The limited efficiency of partitioning is even more detrimental in the isolation of binders from DELs. Binders isolated from DELs cannot be amplified by PCR, while every round of partitioning results in binder losses.[13,15,23,24] Thus, a typical binder selection from a highly diverse DEL is limited to three rounds and produces a pool greatly dominated by nonbinders.[25] Secondary screening (typically called "hit validation" in application to DELs), which involves deep DNA sequencing of the enriched pool followed by *in silico* removal of known nonbinders (e.g., binders to the matrix), synthesis of selected DNA-free compounds, and testing them in binding assays, is used as a means of compensating for the inefficiency of partitioning.[13,15,23,24] Sequencing large amounts of DNA is expensive, and the fact that a significant part of the material processed in DNA sequencing is nonbinders makes selection from highly diverse DELs cost inefficient for industrial R&D and largely cost prohibitive for academic laboratories.

It is clear that reliable single-round selection of binders from oligonucleotide libraries could resolve the identified problems of multiround selection. The first report of single-round selection was published in 2005, and notably, the total number of such reports by the end of 2020 is just over 10.[26−36] All of these reports are concerned with the selection of binders from random-sequence libraries rather than DELs. None of the proposed methods became routinely used for single-round selection. A closer look at these reports reveals that there are no common criteria of successful single-round selection. Most studies are qualitative, noncomparative, and insufficiently universal. The current state of this field strongly suggests that none of the proposed methods is sufficiently reliable for routine single-round selection. In our view, the main problem is the lack of a unifying formalism for the field of single-round selection. Without an accepted underlying theory, the field appears to be art rather than science, and the efforts of different developers appear to be largely disconnected. To address this problem, we present a unifying formalism of single-round selection and use this formalism to analyze the previous efforts and guide the future efforts of method developers.

FORMALISM OF SINGLE-ROUND SELECTION

Mathematics of Partitioning. The overall process of binder selection from a highly diverse library includes (i) mixing the library with the target to allow binders to form target−binder complexes, (ii) physically partitioning these complexes (i.e., binders) from the unbound library (i.e., nonbinders), and (iii) identifying the isolated binders. Formal rules which govern binder selection have been previously described[37] and are the basis for the formalism presented here. The efficient physical partitioning is key to the selection process and can be conceptually presented as a filter that passes most binders and stops most nonbinders (Figure 1).

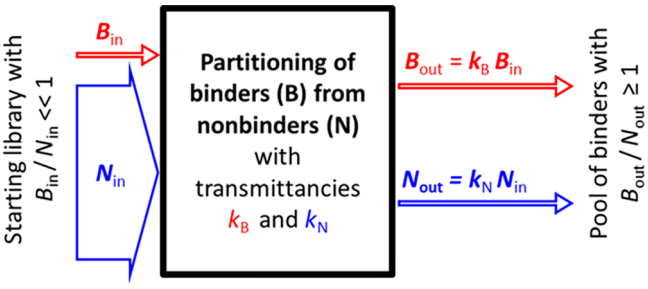


Figure 1. Schematic representation of efficient single-round partitioning of binders from nonbinders in an oligonucleotide library mixture. See text for details.

Partitioning can be described quantitatively using a term of "transmittance" utilized for the characterization of optical filters. Transmittance of partitioning for binders (B), k_B, is defined as the ratio between quantities of binders at the output, B_{out}, and input, B_{in}, of partitioning, respectively

$$k_B = B_{out}/B_{in} \quad (1)$$

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

Accordingly, transmittance of partitioning for nonbinders (N), k_N, is defined as the ratio between the quantities of nonbinders at the output, N_{out}, and input, N_{in}, of partitioning, respectively

$$k_N = N_{out}/N_{in} \quad (2)$$

According to their definitions, k_B is a fraction of binders that passes partitioning, and k_N is a fraction of nonbinders that leaks through partitioning. Both values theoretically range between 0 and 1, and k_B must be greater than k_N. Ideal partitioning is the one with $k_B = 1$ and $k_N = 0$, while in reality $k_B < 1$ and $k_N > 0$. The passage of nonbinders through partitioning creates the nonbinder background in the selection process and contaminates binders at the exit of partitioning. The value of k_N can be used as a quantitative measure of a nonbinder background. The values of k_B and k_N are sufficient to characterize partitioning quantitatively.

We can now link the input and output of partitioning through the values of k_B and k_N. The binder-to-nonbinder ratio at the input of partitioning (the abundance of binders in the starting library), B_{in}/N_{in}, depends on the physical properties and concentrations of the target and the library as well as on the composition, ionic strength, and pH of the buffer. The binder-to-nonbinder ratio at the output of partitioning depends on B_{in}/N_{in}, k_B, and k_N

$$\frac{B_{out}}{N_{out}} = \frac{B_{in}}{N_{in}}\frac{k_B}{k_N} \quad (3)$$

The value of k_B/k_N is the efficiency of partitioning; it theoretically ranges between 0 and ∞. While maximizing k_B/k_N, one must keep in mind the importance of the absolute value of k_B; this practical aspect is discussed below.

Criterion of Successful Single-Round Partitioning. When characterizing single-round selection, one faces a problem of deciding whether the selection was successful. This qualitative decision requires a quantitative criterion of success. Formally, one should choose and justify a certain value Q, which is used as a minimum acceptable value of B_{out}/N_{out} in order to declare the selection successful—if the determined value of B_{out}/N_{out} exceeds Q, single-round partitioning is considered successful. A chosen value Q will define the required efficiency of partitioning in relation to the binder abundance in the starting library

$$\frac{B_{in}}{N_{in}}\frac{k_B}{k_N} > Q \quad (4)$$

We discuss practical considerations for choosing Q below.

Instructiveness of the Formalism. The sole purpose of partitioning is to remove nonbinders from the library while minimizing the loss of binders (similar to isolating gold from ore). Equation 3 clearly shows that partitioning can be improved only by (i) increasing B_{in}/N_{in}, (ii) increasing k_B, and (iii) decreasing k_N. Partitioning is not art, it is science—when introducing any change to partitioning, a developer should clearly understand which of the three parameters will be affected by this change. Some changes may be counterproductive. For example, decreasing k_N is not beneficial if it is accompanied by a proportional decrease in k_B (examples are discussed below). It is also clear that a drastic improvement in partitioning can be achieved by using a library with a greater abundance of binders (higher B_{in}/N_{in}).

Understanding the three key parameters of successful partitioning should help developers to design and optimize their partitioning rationally. The word "optimize" is key as B_{in}/N_{in}, k_B, and k_N depend on many experimental parameters, and maximizing B_{in}/N_{in} and k_B and minimizing k_N requires varying those parameters. In essence, developing a reliable method for single-round selection of binders from oligonucleotide libraries is inevitably a major endeavor which includes ingenuity, rational design, thorough optimization, control-supported proof, and full validation. The efforts must be based on maximizing B_{in}/N_{in} and k_B, minimizing k_N, and, eventually, proving $B_{out}/N_{out} > Q$. Obviously, the formalism must be considered and used in the context of real target−library pairs; this practical aspect is discussed below.

HOW TO DEVELOP AND PROVE SINGLE-ROUND SELECTION

Designing a Comparative Study. Comparativeness is a key requirement for single-round selection studies. Every study should be designed in a way that its results can be quantitatively and conclusively compared to those of the previous and future studies keeping in mind that comparative analysis to the previous studies may be limited due to their nonquantitative nature. To have a comparative study, one should use a target−library pair for which binders have been previously selected and (i) characterize partitioning quantitatively using a mock library with different known B_{in}/N_{in} and (ii) validate partitioning using binder selection from a naïve library followed by comprehensive binder characterization.

Quantitative Characterization of Partitioning Using a Mock Library. A mock library is a mixture containing a known binder and nonbinders with a known binder-to-nonbinder ratio. Mock libraries cannot be used productively for quantitative characterization of multiround selection as cumulative errors (e.g., those of PCR) in multiround selection are very high. Single-round selection is, in contrast, not prone to accumulating errors, and its quantitative characterization can benefit from using a library with known B_{in}/N_{in}.

The mock-library approach for characterization of partitioning is the following. A mock library with different B_{in}/N_{in} is assembled by spiking a known binder into a known nonbinder (with two different sets of PCR primers). Single-step partitioning is applied to each mock library, and the quantities of the binder and nonbinder at the output are determined by quantitative PCR (qPCR) to obtain B_{out}/N_{out} for each value of B_{in}/N_{in}. The values of B_{out}/N_{out} are plotted versus those of B_{in}/N_{in}. The dependence of B_{out}/N_{out} on B_{in}/N_{in} should be linear (eq 3), and the plot is to serve two purposes. First, it helps to identify the linear part of this dependence which corresponds to the range of B_{in}/N_{in} in which B_{out} and N_{out} are quantified reliably by qPCR. Second, the slope of the linear part of the line is the value of k_B/k_N.

The obtained value of k_B/k_N is the characteristic of the quality of partitioning that can be used to compare different partitioning methods. Maximizing k_B/k_N is an approach for optimizing partitioning. Ideally, we would like to use the obtained k_B/k_N to judge if the method is suitable for value single-round selection. However, the required k_B/k_N depends on the target−library pair and can only be roughly estimated after multiple successful single-round selections. In our current view, the required value of k_B/k_N is greater than 10^9;[36] lower values may also work but only for a limited set of targets.[26,28,35,36]

Quantitative Validation of Partitioning. A high value of k_B/k_N obtained for a mock library justifies an effort to validate

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

partitioning in binder selection for the same known target but from a naïve library. A naïve library has an unknown binder-to-nonbinder ratio and uses a single set of PCR primers. The selected binders can only be distinguished from nonbinders in a postselection binding assay. The reliance on a binding assay emphasizes the importance of performing it properly (including all the required controls) and interpreting its results correctly. Binding assays are prone to very large systematic errors. Unlike concentration measurements, the determination of binding constants can lead to errors as large as orders of magnitude. Binding assays deserve separate consideration which is given later in this perspective.

Validation of partitioning is simpler and less expensive to do for a random-sequence library used for aptamer selection. When validated for such a library, the partitioning method can potentially be applied to DELs with physicochemical properties similar to those of the random-sequence library. Although, there are concerns that binding of aptamers and binders from DELs can be affected differently by conditions, for example, buffer composition. The validation strategy for aptamer selection is the following. A single round of partitioning is performed for library–target mixtures at different target concentrations including a control selection with no target. The binder-enriched libraries are amplified by PCR and subjected to a bulk binding assay which can distinguish all binders from all nonbinders. The bulk assay allows the determination of a bulk complex-stability parameter, which is sometimes called the bulk K_d value, but it is better described as an effective concentration, EC_{50}, i.e., target concentration which causes binding of 50% of the library.[38]

The value of EC_{50} is, of course, a function of the concentration of the enriched library and, therefore, is not a true constant. However, it can be used in comparing the results of binder selections performed for different target concentrations. The target concentration of zero should not partition any binders, but binding assays may reveal some nonspecific binding of the product during such selection. This nonspecific binding should, however, have the highest value of EC_{50} of all, and EC_{50} should decrease with increasing target concentration (used in selection) within a certain range. If such behavior is not observed, the results of selection and characterization are inconclusive and likely mean that the selection was not successful. If the dependence of EC_{50} on the target concentration used in partitioning is observed, the partitioning can be further characterized.

The bulk assay can be used to assess the value of B_{out}/N_{out} for selection from a naïve library. These values, of course, will depend on the target concentration used in the selection. One can use the highest obtained value of B_{out}/N_{out} for characterizing partitioning. If this value of B_{out}/N_{out} is greater than the chosen value Q (eq 4), one can declare that the single-round selection was successful.

The results of successful single-round selection can serve several purposes. First, the positive result is a proven achievement on its own. Second, considering that the value of k_B/k_N was determined for the same target using the mock library and B_{out}/N_{out} was determined for the naïve library, one can use eq 3 to estimate B_{in}/N_{in}. There are very limited empirical data for aptamer abundance, and the ability to measure aptamer abundance through single-round selection will be a powerful tool in studies focusing on design of better libraries, such as random-sequence libraries with chemically modified nucleotides.[39−42] Third, successful single-round selection will provide a k_B/k_N value (e.g., obtained with the mock library) that was sufficient for this selection. Such values will serve as reference points for further efforts to develop partitioning approaches for single-round selection.

Above, we showed how our formalism can serve as a general (i.e., without specifying the partitioning method) framework for designing a comparative single-round selection study. Below, we consider the existing partitioning methods, the previous reports of single-round selection, and some pertinent practical considerations through the prism of this formalism.

EXISTING PARTITIONING METHODS

Partitioning on the Surface. There are two major modes of partitioning on the surface: partitioning using a surface-immobilized target (Figure 2a) and partitioning on a physical filter (Figure 2b). The immobilized target approach is more widely used. The target is immobilized on the surface of a solid support, for example, magnetic particles, and allowed to interact with the library in solution. When binders form a complex with the target, they also become immobilized, while the nonbinders remaining in solution can be effectively washed out. The physical filtration approach utilizes filters with a pore size such that the unbound library components predominantly pass through, while target–binder complexes are predominantly retained. Since the sizes of the target–binder complexes and unbound components are on the same order, the heterogeneity of pore sizes results in inefficient partitioning.

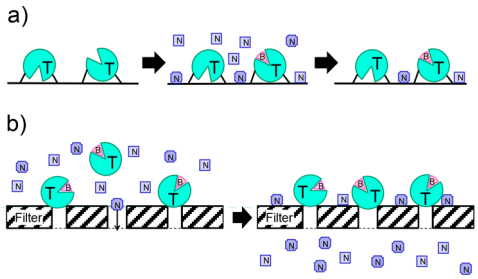


Figure 2. Schematic representation of partitioning on the surface using (a) a surface-immobilized target and (b) a filter that passes through smaller unbound library components (N) but stops larger target–binder complexes (TB). (a, left) A target is immobilized on the surface. (a, middle) The library is incubated with the immobilized target resulting in immobilized TB and some N captured on the surface. (a, right) N is washed out, but a fraction of N is still bound to the surface. (b, left) A target–library mixture is placed on the filter. (b, right) N passes through the pores but also binds nonspecifically to the filter surface.

All methods of partitioning on the surface suffer from the binding of nonbinders to the surface itself.[12,16−19] The values of k_N can be easily measured by finding N_{out}/N_{in} in a target-free selection when the library is allowed to bind to the target-free surface. Measuring k_B is more difficult for several reasons, but the upper limit of efficiency of a partition can be assessed in the assumption of $k_B = 1$. In this case, the efficiency of partitioning equals k_N^{-1}. For partitioning on the surface, the values of k_N have been assessed experimentally and found to be above 10^{-2}.[16,18,19] The value of k_N can be decreased by implementing more stringent wash conditions, but this will also lead to decreasing k_B. Accordingly, it is typically considered that $k_B/k_N < 10^2$ in surface-based partitioning.[16,19]

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

For such partitioning to facilitate successful single-round selection, the abundance of binders in the starting library should be greater than 10^{-2} even for a conservative value of $Q = 1$. As we mentioned above, the scarce data for binder abundancies suggest that they may be as low as 10^{-9}. Accordingly, when surface-based partitioning is the basis for single-round selection, a highly efficient means of reducing the binding of nonbinders to the surface should be identified and used.

Homogeneous Partitioning. Separation of target−binder complexes from nonbinders by capillary electrophoresis (CE) can facilitate homogeneous separation in which the binding of nonbinders to the surface (e.g., to the capillary inner walls) does not decrease the efficiency of partitioning. There are two major modes of CE-based partitioning: nonequilibrium CE of equilibrium mixtures (NECEEM) and ideal-filter CE (IFCE) (Figure 3). In NECEEM, both the target−binder complexes

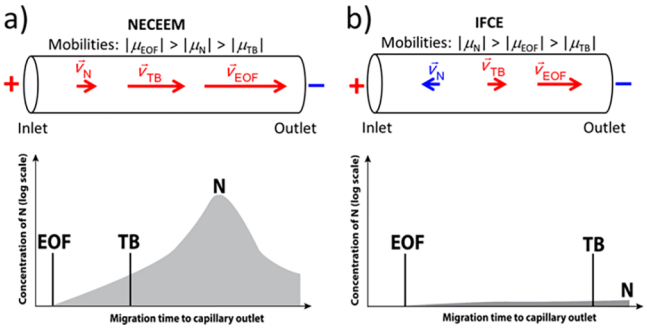


Figure 3. Illustration of homogeneous partitioning by two major modes of CE: (a) NECEEM and (b) IFCE. The top parts show real electrophoretic mobilities (μ) and observed velocities (v) of the electroosmotic flow (EOF), target−binder complexes (TB), and nonbinders (N). The bottom panels show experimental examples of nonbinder backgrounds.

and the unbound library move in the same direction, toward the capillary outlet. The complexes move ahead of the unbound library when an uncoated capillary is used, and the strong electroosmotic flow (EOF) is present (used as an example in Figure 3).[26] When the EOF is suppressed, the unbound library moves ahead of the complexes.[43] In IFCE, the EOF is balanced so that target−binder complexes move to the capillary outlet, while the unbound library moves to the inlet.[36]

In both NECEEM and IFCE, there is a very small fraction of the unbound library that moves to the elution end with a mobility close to that of the target−binder complexes and creates the nonbinder background (see graphs in Figure 3). In IFCE, this fraction is smaller than in NECEEM; the value of k_N in NECEEM and IFCE may be as low as 10^{-6} and 10^{-9}, respectively.[26,36] The value of k_B in NECEEM can be close to unity because even the binders that dissociated from the target during CE separation are separated from nonbinders and collected along with the intact complexes.[26] Therefore, in NECEEM, the efficiency of partitioning is primarily defined by the nonbinder background: $k_B/k_N \approx k_N^{-1}$. In IFCE, the binders that are released from the complexes that dissociate during separation are lost because they move back to the inlet when unbound to the target. As a result, k_B in IFCE can be considerably below unity, and thus, the efficiency of partitioning is affected by both the nonbinder background and loss of binders: $k_B/k_N < k_N^{-1}$. While IFCE is still immature and requires more studies to better understand its potential and limitations, it provides the highest reported efficiency of partitioning. This makes IFCE a good candidate for development of reliable single-round selection. A potential strategy of improving k_B in IFCE is its combination with equilibrium CE of equilibrium mixtures, which is another mode CE-based partitioning, which is not discussed here.[44]

Time as a Parameter of Partitioning. Separation time is a key parameter of every mode of partitioning, and k_B and k_N are functions of separation time. Separation time for partitioning on the surface is defined as the time of washing the binders and nonspecifically bound nonbinders from the surface. Separation time for homogeneous partitioning is the time from the beginning of separation to elution of complexes. The value of k_N decreases with increasing separation time in all modes of partitioning. A longer washing step in surface-based partitioning removes a larger fraction of nonbinders captured by the surface. In homogeneous partitioning, the value of k_N also decreases with increasing separation time as longer separation increases the resolution between zones of binders and nonbinders and reduces the nonbinder background in the zone of binders.

The situation is more case specific for k_B. In surface-based partitioning, k_B decreases with increasing separation time due to complex dissociation during the washing step; in essence, k_B decreases with an effective unimolecular dissociation rate constant of complexes, k_{off}. A similar situation is in IFCE-based partitioning, in which ligands that fell from the target during separation reverse their movement direction and do not reach the capillary outlet. The NECEEM mode of partitioning is different as such binders keep moving toward the capillary outlet and can be collected along with the intact complexes (see the previous section).

In general, k_B decreases with increasing separation time slower than k_N, and therefore, the efficiency of partitioning, k_B/k_N, increases with increasing separation time. However, increasing separation time becomes counterproductive when the total amount of remaining intact target−binder complexes decreases to the limit of quantitation (LOQ) of qPCR. Therefore, choosing separation time is an important practical consideration which depends on a mode of partitioning and the quality of qPCR.

PUBLISHED REPORTS ON SINGLE-ROUND SELECTION

Overview. There are about a dozen published reports on single-round selection from oligonucleotide libraries.[26−36] None of the proposed methods were developed and validated using the general framework described above. Most of the proposed methods were reported without the optimization of k_B and k_N and were not supported by measurements of k_B and k_N. Quantitative characterization of partitioning using a mock library has not been utilized at all. Despite having serious limitations, these reports cumulatively constitute a starting point for future efforts; therefore, we provide their analysis.

There have been successful efforts by synthetic chemists to develop oligonucleotide libraries with increased binder abundancies by using modified oligonucleotides with functionalized protein-like groups.[39−42] None of the single-round selection reports utilized such libraries; instead, unmodified random-sequence oligonucleotide libraries were used. Single-round selection has not been demonstrated for high-diversity DELs either. We do not consider the B_{in}/N_{in} parameter and

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

analyze the proposed single-round selection methods based solely on how improvements in their efficiency of partitioning could be achieved in terms of increasing k_B and decreasing k_N.

One-Round Selection Using Partitioning on the Surface. As we described above, the major drawback of surface-based partitioning is the adherence of nonbinders to the surface of a solid support of a filter. As a result of this nonbinder background, the values of k_N are typically higher than 10^{-2} when using standard washing protocols, which is insufficient for single-round selection even if $k_B = 1$. Thus, the primary goal in surface-based partitioning is to mitigate the effect of the nonbinder background, for example, to decrease k_N hopefully without significantly decreasing k_B. Conceptually, there are two major ways to deal with nonspecific binding to the surface. First, this binding can be suppressed via using antiadhesive materials, and second, the adsorbed compounds can be removed more efficiently than in a standard washing procedure. There are single-round selection reports claiming to suppress nonspecific binding to the surface and/or improve the removal of nonspecifically bound oligonucleotides from the surface.

In 2009, the Soh group published a combination of suppression of nonspecific binding and efficient removal of solubilized nonbinders in a microfluidic platform proposed for single-round selection of aptamers.[29] In this method, the target was immobilized on a magnetic bead coated with carboxylic acid to reduce the nonspecific binding of oligonucleotides to the bead surface. A mixture of these beads and the library was subjected to partitioning in a laminar-flow fluidic system optimized for segregating the beads from the unbound library into separate streams. The authors did not measure k_B but measured k_N which was found to be in the order of 10^{-6}, which is 4 orders of magnitude lower than k_N of conventional surface-based partitioning methods. This group went beyond publishing the proof of principle; the approach was used to attempt one-round selection for other targets. The authors found that other targets required additional rounds of selection along with high-throughput sequencing to identify high affinity aptamers.[45,46] This microfluidic approach is a proven way of high-efficiency partitioning which can potentially be further advanced to facilitate reliable one-round selection of binders from oligonucleotide libraries.

In 2012, Liu and coauthors proposed supplementing the standard washing step with nuclease-catalyzed degradation of oligonucleotides adsorbed onto the surface.[33] A conceptually similar approach was reported by Imashimizu and coauthors in 2018.[35] This approach is restricted to a selection of aptamers from random-sequence oligonucleotide libraries; it is not applicable to DELs. The core concept is that aptamers bound to their protein target are more compact than oligonucleotides bound to the surface, and therefore, the former are more slowly degraded by nucleases than the latter. Accordingly, adding an enzymatic degradation step should decrease k_N to a greater degree than k_B, thus resulting in the increased efficiency of partitioning, k_B/k_N. Since target-bound aptamers are still cleaved by the nuclease, time of the enzymatic digestion step is a key parameter. It should be optimized and precisely controlled to ensure that the unavoidable cleavage of target-bound sequences does not make this step counterproductive.

It appears to us that due to the conceptual simplicity of nuclease-catalyzed degradation of nonbinders, this approach can be fully characterized quantitatively. Even if the enzymatic step proves not to increase the efficiency of partitioning to the level required for reliable single-round selection, this approach can be easily used in the multiround selection of aptamers, thus benefiting the field of aptamer selection. It is worth noting that this approach is not applicable to the selection of binders from DELs as the DNA moiety in all DEL molecules have the same structure and will be susceptible to the same degree of enzymatic degradation for binders and nonbinders.

Our analysis of single-round selection methods revealed that some of them used fractionation, i.e., a collection of multiple fractions from the output of partitioning, as a sole means of improving the efficiency of partitioning. This is a mistaken approach because fractionation cannot improve the quality of partitioning which is defined by the degree of physical separation of binders from nonbinders. Fractionation can assist in identifying fractions with higher binder-to-nonbinder ratios and can, thus, be a useful additional tool if utilized properly.

There are three publications claiming that fractionation can facilitate improvement in partitioning efficiency sufficient for single-round selection. In 2007, Nitsche and coauthors reported partitioning of aptamers on an affinity chromatography column.[27] The column was sliced into segments, and oligonucleotides obtained from individual segments were tested for their affinity to identify the highest affinity aptamer. It was expected that the strongest binders would be found in the beginning of the column, while the weakest ones would be located at the end of the column. However, this hypothesis was not confirmed, further questioning the validity of the proposed fractionation-based approach.

The Çalik[30] and Sotiropoulou[31] groups used fractionation combined with the temperature gradient (2010) and the salt concentration gradient (2012), respectively. The idea was straightforward. Oligonucleotides bound to the target or the surface with low affinity should dissociate at low temperatures and low salt concentrations. Gradually increasing temperature or salt concentration would facilitate progressive removal of stronger bound oligonucleotides. However, higher temperatures and salt concentrations should decrease not only k_N but also k_B. The decrease in k_N and k_B can potentially be proportional, meaning no improvement in the efficiency of partitioning, k_B/k_N. As neither k_N nor k_B were measured in these reports, it is unclear whether the gradients led to any increase in k_B/k_N. Once again, fractionation is only an auxiliary analytical tool which cannot compensate for inefficient partitioning. Similarly, sequencing the entire pool of oligonucleotides with next-generation sequencing (NGS) approaches is also only an auxiliary analytical tool which cannot compensate for inefficient partitioning (see more details on the role of NGS below).

All methods described above can be grouped as a bulk selection approach performed blindly. The availability of advanced imaging tools inspired two groups to utilize microscopy as a means of controlling the partitioning process in an attempt to do single-round selection of aptamers. In 2007, Guthold and coauthors proposed the use of combined atomic force microscopy (AFM) and fluorescence microscopy for selecting a small pool of aptamers in a single round of partitioning.[28] AFM was used to generate a map of target molecules immobilized on the surface. Fluorescent microscopy was then used to identify fluorescently labeled oligonucleotides bound to the target molecules. The bound oligonucleotides were then handpicked with an AFM tip and processed. The strategy of locating and hand picking every binder can

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

drastically reduce k_N but at an inevitable expense of a drastic decrease in k_B. The proof-of-principle work was done using a mock library with a 1:1 binder-to-nonbinder ratio. It is very unlikely that any useful selection can be made by this method.

In 2012, Veedu and coauthors reported the use of fluorescence microscopy to visualize the binding of fluorescently labeled oligonucleotides to the target immobilized on a coverslip.[32] In the framework of our formalism, the visual control was used solely to optimize/control the separation time. The method was bulk selection (no handpicking) as the coverslip was crashed to process the oligos that remained on the surface after washes. While neither k_B nor k_N were measured, it is unlikely that the visual control could markedly change k_B/k_N. Overall, imaging-enhanced selection should be viewed as interesting exploration rather than suggestions for practical development.

Single-Round Selection Using Homogeneous Partitioning by CE. CE-based partitioning separates target−binder complexes from nonbinders in solution based on the differences in their electrophoretic mobilities. There are multiple parameters that can be adjusted and controlled to increase k_B/k_N. In particular, the EOF can be adjusted to control magnitudes and directions of velocities of the unbound library and target−binder complexes. Separation conditions exist for the complexes to move ahead of the unbound library so that any contamination from the nonspecific binding to the inner capillary wall is eliminated.[26] There are also conditions when the library moves ahead of the target−binder complexes.[43] Finally, there are conditions when the complexes and the library move in the opposite directions.[36] If target−binder complexes and nonbinders moved as classic electrophoretic zones with near-Gaussian concentration distributions, the value of k_N could be so low that not a single nonbinder molecule would leak through such partitioning.[47]

In 2005, the Krylov group reported the presence of an unexpected nonbinder background in CE-based partitioning that limits k_N for the NECEEM mode to 10^{-6}−10^{-5}.[26] The value of k_B in NECEEM-based partitioning may be close to unity. NECEEM was shown to facilitate single-round selection for a couple of targets;[26,34] however, in most cases, obtaining a clean pool of binders requires three or four rounds of partitioning.[37,48] Moreover, Mendonsa and Bowser demonstrated that similar efficiency of partitioning (and required number of rounds) can be achieved even if the unbound library elutes from the capillary before the target−binder complexes are collected.[43] In 2019, single-round selection of aptamers was demonstrated with IFCE-based partitioning.[36] The k_N value for IFCE was shown to be as low as 10^{-9}; however, the binders that dissociate during the separation are lost due to their moving back to the capillary inlet. As a result, k_B can be much less than unity, and the effect of the decreased k_B on the efficiency of partitioning can potentially make partitioning efficiencies of NECEEM and IFCE close to each other.

While the nonbinder background in CE-based partitioning has a very complex nature,[49] the values of k_B and k_N depend systematically on a set of CE experimental conditions: the value of the negative charge on the oligonucleotide, the hydrodynamic size of the protein target, the ionic strength and pH of the separation buffer, the length of the sample plug, and the length of the capillary. Studying the influence of all these conditions on k_B and k_N and optimizing these conditions to maximize k_B/k_N are required and will constitute a significant amount of experimental work. Completing this work in a timely fashion may, in turn, require a joint effort of several research groups.

IMPORTANT PRACTICAL CONSIDERATIONS OF SINGLE-ROUND SELECTION

When Is Single-Round Selection Needed? Single-round selection can solve problems associated with multiround selection, but it is extremely challenging as it imposes very high requirements for the efficiency of partitioning. Single-round selection is only needed when addressing fundamental questions such as finding binder abundance in a naïve library. Multiple rounds of partitioning introduce large errors in DNA quantitation, especially when combined with intermediate PCR amplification. Moreover, k_B/k_N can differ from round to round. As a result, accurately back-calculating B_{in}/N_{in} from B_{out}/N_{out} in multiround selection is virtually impossible, and single-round selection is necessary

Any binder selection that aims at simply obtaining good-quality binders does not require single-round selection. An additional round is of course associated with additional costs, but these costs may be completely justified by increasing B_{out}/N_{out}, thus decreasing the cost of secondary screening. To better understand this point, we need to consider the economics of binder identification that follows binder partitioning. In general, binder identification requires (i) sequencing oligonucleotide moieties of the entire pool using NGS, (ii) optional sequence frequency analysis, which can be applicable to naïve libraries with high redundancy of compounds, for example, DELs, (iii) chemical synthesis of a small set of binders limited by economics of both the synthesis and the following validation assay, and (iv) validation of binding, i.e., measuring K_d and k_{off} values for target−binder complexes for all the synthesized compounds.

Steps three and four of the four-step secondary screening require operating with individual binders and are, thus, very resource intensive. Accordingly, an additional step of partitioning can be completely justified if it reduces the amount of nonbinders that need to be sequenced, synthesized, and tested in a binding assay. The cost of secondary screening is different for aptamers and small-molecule binders, and accordingly, a decision on whether or not an additional round of partitioning is justified depends on the circumstances. See further elaboration on the justification of additional rounds in the next section.

Parameter *Q*. The choice of value of Q is important as it defines whether or not single-round selection is successful according to eq 4. While this parameter appears to be arbitrary, there is rational reasoning for setting its value for the challenging selection procedure. Single-round selection can be considered successful if an additional round of partitioning will not significantly decrease the attrition rate in secondary screening. The attrition rate during secondary screening is of critical importance. In addition to the economical aspect, there is a psychological one. If Q is set at a level of 0.1, one accepts *a priori* that the vast majority of resources involved in DNA sequencing will be wasted. This level of wasted resources is likely acceptable for an industrial setting if it is justified economically. However, it may be unacceptable for academic efforts which typically have short timelines defined by durations of specific study programs. We believe that an attrition rate of up to 50% is acceptable, thus setting the minimum Q to unity.

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

Secondary screening (hit optimization) is excessively costly in the case of selection of hits from DELs. Accordingly, any additional round of partitioning that markedly increases B_{out}/N_{out} is justified. However, during binder selection from DELs, the loss of binders (due to both $k_B < 1$ and the inability to amplify small-molecule binders by PCR) typically limits the number of consecutive rounds to three, while the value of B_{out}/N_{out} after four rounds is still very small.[13,15,23,24] In such a case, setting $Q = 0.1$ for single-round selection may be fully justified as it is already higher than B_{out}/N_{out} obtained in three rounds of conventional partitioning currently used (for three rounds, $B_{out}/N_{out} = B_{in}/N_{in} \times (k_B/k_N)^3$).

On the basis of the above, we suggest that developers of single-round selections of aptamers from random-sequence libraries set $Q = 1$ or $Q = 10$. For attempts to achieve single-round selection of binders from DELs, we suggest $Q = 0.1$ or $Q = 1$. Obviously, the higher the value of Q is, the more reliable the result is. A value of Q below 0.1 should be considered as unacceptably low.

Binding Assays. Assessing the quality of binder selection from a naive library is based on the results of secondary screening, which is to confirm formation of target−binder complexes

$$\mathrm{T} + \mathrm{B} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} \mathrm{TB}, \qquad K_d = k_{off}/k_{on} \tag{5}$$

Binding assays are used to measure the rate constants k_{on} and k_{off} and the dissociation constant K_d for this reversible binding process. Binding assays can be classified in a number of ways. For the purpose of this discussion, we split these in three groups: (i) methods relying on physical separation (of B from BT with detection of B both free and target bound or T from BT with detection of T both free and binder bound), (ii) methods relying on spectral separation (resolution), and (iii) methods relying on calorimetry.

Physical separation in binding assays can be surface-based (in biosensoric binding assays, such as surface-plasmon resonance and biolayer interferometry)[50−52] and homogeneous (in binding assays based on CE and transient incomplete separation in a parabolic flow profile).[53,54] Surface-based methods are prone to errors originating from nonspecific binding of the binder to the surface and hindered interaction with the immobilized target. As a result, the obtained constants can be irrelevant to target−binder interaction in a biological context.[55] CE-based binding assays assess binding in solution but are limited to buffers compatible with electrophoresis, i.e., ones with low salt concentration. Advantageously, separating unbound oligonucleotides from protein−oligonucleotide complexes is trivial in CE. A method termed Accurate Constant by Transient Incomplete Separation (ACTIS) is a homogeneous method for finding K_d. It relies on a pressure-driven laminar flow and, therefore, is compatible with any buffer. However, it requires a large difference in diffusion coefficient between TB and B; it is still to be confirmed if ACTIS can be used for measuring K_d values of protein−oligonucleotide complexes. In addition, ACTIS, in its current format, relies on very fast separation and, therefore, is a pseudoequilibrium method that cannot be used to measure k_{on} and k_{off}.

Methods relying solely on spectral separation (i.e., not combined with physical separation), such as mass-spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy, can typically only measure K_d values and not k_{off} or k_{on}. The ionization process in MS can shift the equilibrium in eq 5 and, thus, affect measured K_d.[56] In NMR binding assays, K_d can be measured based on the spectral change of the target macromolecule or ligand (oligonucleotide or DEL in this context) upon the titration of ligand or target, respectively.[57] In target-based NMR, the spectral crowding (broadening and overlap of peaks) observed in large molecules restricts K_d measurements to small macromolecules (<50 kDa). Ligand-based NMR is not limited by the target size; however, transient aggregation and nonspecific binding to the target can lead to false positive binding events and affects the accuracy of K_d measurements. A calorimetric method used for binding assays is isothermal titration calorimetry (ITC).[58] ITC has an advantage that it can also be used to measure k_{on} and k_{off} values in addition to K_d.[59] The accuracy of thermodynamic measurements by ITC can be affected by unavoidable side reactions, such as protein protonation, which also contribute to heat produced in the titration process.[60] In addition, NMR and ITC require high concentrations and large quantities of the target and the binder which may preclude their use in measuring protein−binder complexes. As a result, biosensoric methods, such as surface plasmon resonance and biolayer interferometry, which have very high sensitivity are the most common binding assays used to characterize binders selected from oligonucleotide libraries for protein targets.

All established binding assays are prone to large systematic errors. The value of K_d is more commonly determined than k_{on} and k_{off}, and the results obtained by different methods for the same complex may vary by multiple orders of magnitude.[61] In the absence of reference standards, it is very difficult to assess the accuracy of measurements. For deterministic methods, whose processes can be comprehensively described by a system of partial differential equations, virtual experiments can be used to assess accuracy. We use the term "virtual experiment" to describe computer modeling utilizing a virtual instrument and mimicking the processes which occur in the real instrument very accurately. However, biosensoric methods are hardly suitable for building virtual instruments and running virtual experiments due to difficulties of accurate modeling of processes occurring in the surface of the biosensor. As a result, little is known of their intrinsic inaccuracies.

Homogeneous separation-based methods are deterministic, and their accuracy can be assessed *in silico*. Comprehensive accuracy assessments were performed for K_d measurements by NECEEM and ACTIS. It was found that K_d values obtained with a single measurement by NECEEM are largely unreliable; systematic errors can be multiple orders of magnitude.[62,63] Titration experiments should be performed in NECEEM binding assays to avoid such drastic inaccuracy.[62] ACTIS is based on titration and is intrinsically accurate; virtual experiments show that systematic errors in ACTIS can be below 10%.[54]

Every binding assay likely has a source of inaccuracy that is difficult to trace. Our experience with NECEEM-based binding assays allowed us to identify such a source, which is often disregarded by other researchers. Oligonucleotides can form complexes with counterions and produce false target−binder peaks.[49] This emphasizes the need for all possible control experiments to exclude false-positive results. Considering large inaccuracies of binding assays, we recommend that researchers be more diligent to significant digits used to present K_d values. Reporting K_d values with multiple significant digits makes sense only in a comparative study ranking the binders assessed

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

using the same instrument under identical conditions and when the random error is sufficiently small.

The discrepancies in binding parameters obtained by different groups are rarely noticed and emphasized leading to the continuing accumulation of inconsistent data, misinterpretations, and wrong conclusions. In light of this alarming deficiency, we would like to point the reader's attention to recent work from the De Wael group.[64] The authors demonstrate that ampicillin-binding aptamers claimed to have a nanomolar K_d value that revealed no binding in studies with ITC, native nanoelectrospray ionization mass spectrometry, and ^{1}H-nuclear magnetic resonance spectroscopy (^{1}H NMR). The authors recommend the use of more than one technique to measure binding parameters and call for the standardization of aptamer-characterization approaches. We agree with this proposal and suggest a strategy in which the binding is characterized by two different methods, and if the results are inconsistent, additional methods are used until a conclusion can be made with certainty.

Role of Next-Generation Sequencing. Incorporation of NGS technologies for analysis of binder-enriched pools and identification of candidate aptamers and small molecule binders has been one of the most major changes in binder selection procedures. Traditionally, enriched binder pools were cloned into a plasmid, and only a few hundred individual clones were sequenced and assessed for binding in secondary screening. NGS analysis can yield hundreds of millions of sequence reads, thus enabling a more comprehensive analysis of the whole selection process and providing more options for secondary screening.

In 2011, Glokler and coauthors explored the use of NGS for the analysis of the enrichment process in multiround selection of aptamers.[65] They revealed that the most abundant sequences identified in the final binder-enriched pools are not necessarily the best (highest-affinity) aptamers. The strongest binders were found to accumulate very early during the first three selection rounds but then gradually got replaced by weaker binders in subsequent rounds. This behavior indicates that selection biases, such as high affinity to the partitioning matrix and PCR bias or sequencing bias, might strongly influence the outcome of multiround selection process. The authors suggested minimizing selection cycles to avoid selection biases and the identification of high-affinity aptamer candidates in much earlier rounds with the power of NGS. Another strategy to minimize biases is the skip PCR amplification (applicable to random DNA libraries).[48,66] Although the analysis of sequence enrichment via NGS allows more transparent selection and may exclude unnecessary excessive rounds, it cannot compensate for the poor efficiency of the separation method used.

In general, the optimum number of selection rounds depends on the complexity of the starting library, the nature of the target, the enrichment factor achieved each round, and the number of reads obtained from sequencing. Any given library may contain more binders for some targets and fewer for others. More challenging targets require more rounds of selection even though NGS is employed for binder identification.

The majority of published single-round selection methods utilized Sanger sequencing for identifying high-affinity aptamers in the binder-enriched pools. In 2017, the first application of NGS in the field of single-round selection methods was introduced by the Saito group in a NECEEM-based aptamer selection.[34] Using the MAFFT program, the authors carried out clustering analysis to identify analogous sequences (families). The candidate aptamers were chosen from families composed of the highest number of analogous sequences. In 2018, Imashimizu and coauthors also incorporated NGS to identify aptamer sequences in a single round of nuclease-mediated selection.[35] They used FASTAptamer programs to cluster unique sequences into different families based on the Levenshtein edit distance.[67] The aptamer candidates were obtained by choosing the clustered sequences with the highest frequencies and with the length closest to the original library. It is noted that individual sequences in the starting library are not equally represented due the sequence biases generated by the manufacturing procedure (solid-phase chemical synthesis) of the oligonucleotide library. It is crucial to sequence the starting library to identify any biased sequences, which, in turn, might influence the nucleotide composition of the enriched sequences and their frequencies.

Despite the common use of NGS, the number of available bioinformatic tools for identifying candidate aptamers from large sequencing data obtained from single-round selection is limited. The field of single-round selection of aptamers would benefit immensely from an agreed upon algorithm and metrics for selecting candidate aptamers.

Role of Negative Selection. Negative selection is the removal from the library of binders to the surface in the surface-based partitioning and/or to the nonspecific targets to avoid cross-reactivity of the selected binders with such side targets.[68] Accordingly, there are two major types of negative selection: (i) negative selection against the bare surface[32,69,70] and (ii) negative selection against nonspecific targets.[33,71,72] If the nonspecific targets are immobilized on the surface, then negative selection can remove simultaneously molecules binding the surface and the nonspecific targets. It is important to emphasize that negative selection against the surface aims to remove library components that have affinity to the surface, for example, are specific binders to the surface. Such selections cannot decrease the level of nonspecific adsorption of the library on the surface that is governed, for example, by electrostatic interactions of the oligonucleotides with the surface. Negative selection can either precede the step of positive selection or can be concurrent with positive selection.[32,33,69−72] PCR amplification may not be needed between the steps of negative and positive selection run sequentially.

How will negative selection influence B_{in}/N_{in}, k_B, and k_N? Let us consider negative selection being a part of a single partitioning step along with positive selection. Our formalism in its simplest form depicted in Figure 1 is applicable to this case, and negative selection cannot affect B_{in}/N_{in} but can affect both k_N and k_B. If we assume that specific binders to the surface and/or to the nonspecific targets are not binders to the real target, then the negative selection will decrease only k_N. However, if the negative selection step removes by chance a part of the desirable binders to the target, the value of k_B will also decrease. The degree to which k_N and k_B will decrease, and thus, the resulting change in the efficiency of partitioning, k_B/k_N, will depend on the specifics of selection. We are not aware of studies in which the effects of negative selection on k_B and k_N were quantitatively studied. We know, however, that negative selection is a routine procedure in surface-based partitioning, and while likely improving the efficiency, it does not provide a drastic increase.

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

Choice of Targets and Libraries. Proteins constitute approximately 95% of therapeutic targets,[73] and they are the most common targets for selection of binders from oligonucleotide libraries. Selecting pharmaceutical hits would obviously benefit from using a target protein in its near-native state. This suggests that near-physiological conditions, for example, composition, pH, and ionic strength of the buffers, are preferable for partitioning. Proof-of-principle works on the development of single-round selection should be conducted with target−library pairs for which binders have been previously selected, for example, in a multiround process. When designing a mock library, a binder and a nonbinder of choice should be characterized by measuring their K_d and k_{off} for the target of choice.

It is evident that even if a proposed single-round selection method is proven with our framework (introduced here) for a certain target, a single round may not be enough for selecting binders from the same library for another target. As B_{in}/N_{in} depends on both the target and the library, choosing an alternative library should become a practical consideration. Libraries with modified oligonucleotides are postulated to have greater B_{in}/N_{in} than nonmodified oligonucleotide libraries in selection of aptamers for some targets.[74] So far, the development of single-round selection methods has not utilized the combined optimization of the efficiency of partitioning and the abundance of binders. It would be very interesting to see the implementation of the modified oligonucleotide technology in the field of single-round selection in the future.

CONCLUDING REMARKS

Achieving single-round selection is not a self-justified goal. In the majority of applications, single-round selection is not needed. Two or three consecutive rounds of partitioning are economically acceptable, and there is no proof that they could introduce significant biases in the selection process. Despite that, single-round selection remains attractive as it is an indication of the highest efficiency of partitioning. In this work, we suggest that the logic should be changed. Achieving and proving high efficiency of partitioning must be the goal as it implies that very few rounds of selection (or perhaps a single one) will be enough to obtain a high purity pool of binders. We propose a general framework for designing experiments that aim at the development of high-efficiency partitioning. This framework is based on a simple formalism of partitioning which considers only three universal variables: abundance of binders in the starting library, B_{in}/N_{in}, transmittances of partitioning for binder, k_B, and nonbinders, k_N. It also considers a single criterion for successful selection: the abundance of binders in the resulting binder-enriched library, B_{out}/N_{out}, must exceed a certain threshold value, for example, 1 or 10. We encourage the developers to design their studies and report their results using the framework presented here. We hope that this work will serve as a starting point for a search for standardized strategies for selection of binders from oligonucleotide libraries. Such strategies are to define not only the general framework but details of selection protocols such as all necessary controls. The International Society on Aptamers can serve as a respected forum for such discussions.

AUTHOR INFORMATION

Corresponding Author

Sergey N. Krylov − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada;* orcid.org/0000-0003-3270-2130; Email: skrylov@yorku.ca

Authors

An T. H. Le − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada;* orcid.org/0000-0002-3659-9938

Svetlana M. Krylova − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada;* orcid.org/0000-0002-3291-6721

Stanislav S. Beloborodov − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada;* orcid.org/0000-0003-2377-4845

Tong Y. Wang − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada*

Ryan Hili − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada;* orcid.org/0000-0002-8933-970X

Philip E. Johnson − *Department of Chemistry and Centre for Research on Biomolecular Interactions, York University, Toronto, Ontario M3J 1P3, Canada;* orcid.org/0000-0002-5573-4891

Feng Li − *Department of Chemistry and Centre for Biotechnology, Brock University, St. Catharines, Ontario L2S 3A1, Canada;* orcid.org/0000-0002-2616-5343

Rakesh N. Veedu − *Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University and Perron Institute for Neurological and Translational Science, Perth 6150, Australia;* orcid.org/0000-0003-2192-9532

Svetlana Belyanskaya − *Anagenex, Woburn, Massachusetts 01801, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.1c00601

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the NSERC Canada (Grant STPG-P 521331-2018). The authors thank Dr. Christopher Arico-Muendel (GlaxoSmithKline) for reading the manuscript and providing critical comments on the DEL aspects of this work.

REFERENCES

(1) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5381−5383.

(2) Bleicher, K. H.; Bohm, H. J.; Muller, K.; Alanine, A. I. *Nat. Rev. Drug Discovery* **2003**, *2*, 369−378.

(3) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotechnol.* **2004**, *22*, 568−574.

(4) Li, J.; Murray, C. W.; Waszkowycz, B.; Young, S. C. *Drug Discovery Today* **1998**, *3*, 105−112.

(5) Plunkett, M. J.; Ellman, J. A. *Sci. Am.* **1997**, *276*, 68−73.

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

(6) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. *Br. J. Pharmacol.* **2011**, *162*, 1239−1249.
(7) Vant-Hull, B.; Gold, L.; Zichi, D. A. *Curr. Protoc. Nucleic Acid Chem.* **2000**, *00*, 9.1.1−9.1.16.
(8) Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. *J. Am. Chem. Soc.* **2010**, *132*, 11779−11791.
(9) Ecker, D. J.; Vickers, T. A.; Hanecak, R.; Diver, V.; Anderson, K. *Nucleic Acids Res.* **1993**, *21*, 1853−1856.
(10) Kottani, R.; Valiulin, R. A.; Kutateladze, A. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13917−13921.
(11) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564−566.
(12) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505−510.
(13) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J. W.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L. J.; Van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B. G.; Morgan, B. A. *Nat. Chem. Biol.* **2009**, *5*, 772−772.
(14) Osborne, S. E.; Matsumura, I.; Ellington, A. D. *Curr. Opin. Chem. Biol.* **1997**, *1*, 5−9.
(15) Goodnow, R. A.; Dumelin, C. E.; Keefe, A. D. *Nat. Rev. Drug Discovery* **2017**, *16*, 131−147.
(16) Irvine, D.; Tuerk, C.; Gold, L. *J. Mol. Biol.* **1991**, *222*, 739−761.
(17) Ciesiolka, J.; Illangasekare, M.; Majerfeld, I.; Nickles, T.; Welch, M.; Yarus, M.; Zinnen, S. *Methods Enzymol.* **1996**, *267*, 315−335.
(18) Papoulas, O. *Curr. Protoc. Mol. Biol.* **2001**, *36*, 12.8.1−12.8.9.
(19) Wang, J.; Rudzinski, J. F.; Gong, Q.; Soh, H. T.; Atzberger, P. J. *PLoS One* **2012**, *7*, e43940.
(20) Multer, G. L.; Boynton, K. A. *Nucleic Acids Res.* **1995**, *23*, 1411−1418.
(21) Warnecke, P. M.; Stirzaker, C.; Melki, J. R.; Millar, D. S.; Paul, C. L.; Clark, S. J. *Nucleic Acids Res.* **1997**, *25*, 4422−4426.
(22) Takahashi, M.; Wu, X. W.; Ho, M.; Chomchan, P.; Rossi, J. J.; Burnett, J. C.; Zhou, J. H. *Sci. Rep.* **2016**, *6*, 33697.
(23) Chan, A. I.; Mcgregor, L. M.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2015**, *26*, 55−61.
(24) Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Buhlmann, P.; Supuran, C. T.; Neri, D. *ACS Chem. Biol.* **2011**, *6*, 336−344.
(25) Arico-Muendel, C. GlaxoSmithKline, Personal Communication, 2021.
(26) Berezovski, M.; Drabovich, A.; Krylova, S. M.; Musheev, M.; Okhonin, V.; Petrov, A.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 3165−3171.
(27) Nitsche, A.; Kurth, A.; Dunkhorst, A.; Panke, O.; Sielaff, H.; Junge, W.; Muth, D.; Scheller, F.; Stocklein, W.; Dahmen, C.; Pauli, G.; Kage, A. *BMC Biotechnol.* **2007**, *7*, 48.
(28) Peng, L.; Stephens, B. J.; Bonin, K.; Cubicciotti, R.; Guthold, M. *Microsc. Res. Tech.* **2007**, *70*, 372−381.
(29) Lou, X.; Qian, J.; Xiao, Y.; Viel, L.; Gerdon, A. E.; Lagally, E. T.; Atzberger, P.; Tarasow, T. M.; Heeger, A. J.; Soh, H. T. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2989−2994.
(30) Çalik, P.; Balci, O.; Ozdamar, T. H. *Protein Expression Purif.* **2010**, *69*, 21−28.
(31) Arnold, S.; Pampalakis, G.; Kantiotou, K.; Silva, D.; Cortez, C.; Missailidis, S.; Sotiropoulou, G. *Biol. Chem.* **2012**, *393*, 343−353.
(32) Lauridsen, L. H.; Shamaileh, H. A.; Edwards, S. L.; Taran, E.; Veedu, R. N. *PLoS One* **2012**, *7*, e41702.
(33) Liu, Y.; Wang, C.; Li, F.; Shen, S.; Tyrrell, D. L.; Le, X. C.; Li, X. F. *Anal. Chem.* **2012**, *84*, 7603−7606.
(34) Hirose, K.; Tsuchida, M.; Asakura, H.; Wakui, K.; Yoshimoto, K.; Iida, K.; Sato, M.; Shibukawa, M.; Suganuma, M.; Saito, S. *Analyst* **2017**, *142*, 4030−4038.
(35) Imashimizu, M.; Takahashi, M.; Amano, R.; Nakamura, Y. *Biol. Methods Protoc.* **2018**, *3*, Bpy004.
(36) Le, A. T. H.; Krylova, S. M.; Kanoatov, M.; Desai, S.; Krylov, S. N. *Angew. Chem., Int. Ed.* **2019**, *58*, 2739−2743.
(37) Yufa, R.; Krylova, S. M.; Bruce, C.; Bagg, E. A.; Schofield, C. J.; Krylov, S. N. *Anal. Chem.* **2015**, *87*, 1411−1419.
(38) Berezovski, M. V.; Musheev, M. U.; Drabovich, A. P.; Jitkova, J. V.; Krylov, S. N. *Nat. Protoc.* **2006**, *1*, 1359−1369.
(39) Eaton, B. E. *Curr. Opin. Chem. Biol.* **1997**, *1*, 10−16.
(40) Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E. *J. Am. Chem. Soc.* **2010**, *132*, 4141−4151.
(41) Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T.; Flather, D.; Forbes, A.; Foreman, T.; Fowler, C.; Gawande, B.; Goss, M.; Gunn, M.; Gupta, S.; Halladay, D.; Heil, J.; Heilig, J.; Hicke, B.; Husar, G.; Janjic, J.; Jarvis, T.; Jennings, S.; Katilius, E.; Keeney, T. R.; Kim, N.; Koch, T. H.; Kraemer, S.; Kroiss, L.; Le, N.; Levine, D.; Lindsey, W.; Lollo, B.; Mayfield, W.; Mehan, M.; Mehler, R.; Nelson, S. K.; Nelson, M.; Nieuwlandt, D.; Nikrad, M.; Ochsner, U.; Ostroff, R. M.; Otis, M.; Parker, T.; Pietrasiewicz, S.; Resnicow, D. I.; Rohloff, J.; Sanders, G.; Sattin, S.; Schneider, D.; Singer, B.; Stanton, M.; Sterkel, A.; Stewart, A.; Stratford, S.; Vaught, J. D.; Vrkljan, M.; Walker, J. J.; Watrobka, M.; Waugh, S.; Weiss, A.; Wilcox, S. K.; Wolfson, A.; Wolk, S. K.; Zhang, C.; Zichi, D. *PLoS One* **2010**, *5*, e15004.
(42) Kong, D. H.; Yeung, W.; Hili, R. *J. Am. Chem. Soc.* **2017**, *139*, 13977−13980.
(43) Mendonsa, S. D.; Bowser, M. T. *Anal. Chem.* **2004**, *76*, 5387−5392.
(44) Drabovich, A.; Berezovski, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2005**, *127*, 11224−11225.
(45) Ahmad, K. M.; Oh, S. S.; Kim, S.; Mcclellen, F. M.; Xiao, Y.; Soh, H. T. *PLoS One* **2011**, *6*, e27051.
(46) Cho, M.; Xiao, Y.; Nie, J.; Stewart, R.; Csordas, A. T.; Oh, S. S.; Thomson, J. A.; Soh, H. T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5472.
(47) Kochmann, S.; Le, A. T. H.; Hili, R.; Krylov, S. N. *Electrophoresis* **2018**, *39*, 2991−2996.
(48) Berezovski, M.; Musheev, M.; Drabovich, A.; Krylov, S. N. *J. Am. Chem. Soc.* **2006**, *128*, 1410−1411.
(49) Musheev, M. U.; Kanoatov, M.; Krylov, S. N. *J. Am. Chem. Soc.* **2013**, *135*, 8041−8046.
(50) Polonschii, C.; David, S.; Tombelli, S.; Mascini, M.; Gheorghiu, M. *Talanta* **2010**, *80*, 2157−2164.
(51) Chang, A. L.; McKeague, M.; Liang, J. C.; Smolke, C. D. *Anal. Chem.* **2014**, *86*, 3273−3278.
(52) Lou, X.; Egli, M.; Yang, X. *Curr. Protoc. Nucleic Acid Chem.* **2016**, *67*, 7.25.1−7.25.15.
(53) Krylov, S. N. *Electrophoresis* **2007**, *28*, 69−88.
(54) Sisavath, N.; Rukundo, J. L.; Le Blanc, J. C. Y.; Galievsky, V. A.; Bao, J. Y.; Kochmann, S.; Stasheuski, A. S.; Krylov, S. N. *Angew. Chem., Int. Ed.* **2019**, *58*, 6635−6639.
(55) Nieba, L.; Krebber, A.; Pluckthun, A. *Anal. Biochem.* **1996**, *234*, 155−165.
(56) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 431−441.
(57) Barile, E.; Pellecchia, M. *Chem. Rev.* **2014**, *114*, 4749−4763.
(58) Slavkovic, S.; Johnson, P. E. *Aptamers* **2018**, *2*, 45−51.
(59) Burnouf, D.; Ennifar, E.; Guedich, S.; Puffer, B.; Hoffmann, G.; Bec, G.; Disdier, F.; Baltzinger, M.; Dumas, P. *J. Am. Chem. Soc.* **2012**, *134*, 559−565.
(60) Baranauskiene, L.; Petrikaite, V.; Matuliene, J.; Matulis, D. *Int. J. Mol. Sci.* **2009**, *10*, 2752−2762.
(61) Watzig, H.; Oltmann-Norden, I.; Steinicke, F.; Alhazmi, H. A.; Nachbar, M.; El-Hady, D. A.; Albishri, H. M.; Baumann, K.; Exner, T.; Bockler, F. M.; El Deeb, S. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 847−865.
(62) Kanoatov, M.; Galievsky, V. A.; Krylova, S. M.; Cherney, L. T.; Jankowski, H. K.; Krylov, S. N. *Anal. Chem.* **2015**, *87*, 3099−3106.
(63) Kanoatov, M.; Mehrabanfar, S.; Krylov, S. N. *Anal. Chem.* **2016**, *88*, 9300−9308.

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX

(64) Bottari, F.; Daems, E.; de Vries, A.-M.; Van Wielendaele, P.; Trashin, S.; Blust, R.; Sobott, F.; Madder, A.; Martins, J. C.; De Wael, K. *J. Am. Chem. Soc.* **2020**, *142* (46), 19622−19630.
(65) Schutze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Morl, M.; Erdmann, V. A.; Lehrach, H.; Konthur, Z.; Menger, M.; Arndt, P. F.; Glokler, J. *PLoS One* **2011**, *6*, e29604.
(66) Szeto, K.; Latulippe, D. R.; Ozer, A.; Pagano, J. M.; White, B. S.; Shalloway, D.; Lis, J. T.; Craighead, H. G. *PLoS One* **2013**, *8*, e82667.
(67) Alam, K. K.; Chang, J. L.; Burke, D. H. *Mol. Ther.–Nucleic Acids* **2015**, *4*, e230.
(68) Stoltenburg, R.; Reinemann, C.; Strehlitz, B. *Biomol. Eng.* **2007**, *24*, 381−403.
(69) Haller, A. A.; Sarnow, P. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 8521−8526.
(70) Vater, A.; Jarosch, F.; Buchner, K.; Klussmann, S. *Nucleic Acids Res.* **2003**, *31*, e130−e130.
(71) Wang, C.; Zhang, M.; Yang, G.; Zhang, D.; Ding, H.; Wang, H.; Fan, M.; Shen, B.; Shao, N. *J. Biotechnol.* **2003**, *102*, 15−22.
(72) Shangguan, D.; Li, Y.; Tang, Z.; Cao, Z. C.; Chen, H. W.; Mallikaratchy, P.; Sefah, K.; Yang, C. J.; Tan, W. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11838−11843.
(73) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. *Nat. Rev. Drug Discovery* **2017**, *16*, 19−34.
(74) Gawande, B. N.; Rohloff, J. C.; Carter, J. D.; von Carlowitz, I.; Zhang, C.; Schneider, D. J.; Janjic, N. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 2898−2903.

https://doi.org/10.1021/acs.analchem.1c00601
Anal. Chem. XXXX, XXX, XXX−XXX